PROJECT PERIODIC REPORT: FINAL

COVERING PERIODS 1-3, MAY 2014 – JUNE 2018 (M01-M50)

VERSION 1.0

Date submitted: 17 May 2018

Dissemination Level: PU / PP / RE / CO

WORK PACKAGE: WP1 MANAGEMENT

WORK PACKAGE LEADER: B.L. WILLIAM WONG

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no FP7-IP-608142.

Table of Contents

SECTION 1
PUBLISHABLE SUMMARY:

Project Objectives, Work Progress, Achievements, and Project Management
Period 3, M17 – M50, Sep 2014 – Jun 2018

A. SUMMARY DESCRIPTION OF PROJECT CONTEXT AND OBJECTIVES (4000 CHAR // 3998 CHARS)

INTRODUCTION

The purpose of VALCRI was to develop the next generation criminal intelligence analysis system for European LEAs. Working closely with three European police forces, the project researched and developed at TRL-5, an integrated system of over 75 software components of advanced data processing, analytic and sense-making tools. It includes multiple applications spanning strategic intelligence analysis to tactical intelligence and individual case management.

The VALCRI system was routinely evaluated with project end-users. In the final nine months, it has been evaluated with 214 LEA officers in 50 agencies in 16 countries and 2 international LEAs (Europol and NATO Intelligence Fusion Centre). It is undergoing trials with actual data at the London Metropolitan Police, and the Pasco County Sheriff's Department in Florida. Negotiations are also underway to purchase or licence various VALCRI technologies and non-software outcomes.

VALCRI used a cognitive engineering approach to create a human-technology team that combined advanced concepts of human reasoning and analytic discourse with machine learning and database technologies. The result has been a semi-automated human-mediated semantic knowledge extraction capability that can facilitate and improve investigative sense making and problem solving in crime analysis and criminal investigation in a high ambiguity and constantly evolving environment.

KEY DISTINGUISHING FEATURES

1. SUPPORT HOW ANALYSTS THINK, RATHER THAN WHAT ANALYSTS DO

If VALCRI were designed to mainly support what analysts do, then the system would primarily automate current tasks and workflows. Instead, by designing for how analysts think, the VALCRI system is better able to respond to the variety of sense making, reasoning and inference making and problem solving strategies presented by human analysts.

2. FACILITATE EXPERT INTUITION TO SCIENTIFIC METHOD

In many investigations, analysts are often only presented with fragments of data from which to create an understanding of the situation and to anticipate what might happen. Expert intuition is very useful in generating "hunches", or early, plausible and tentative hypotheses. However, hunches can be error prone and subject to cognitive biases. VALCRI has designed quick ways for analysts to use the scientific method to test their hunches so that they may easily discard it if proven wrong.

3. HUMANS DECIDE, MACHINES DO THE HEAVY LIFTING

VALCRI has been designed so that humans and machines do what each is good at: Humans make decisions under ambiguity; machines are fast at tedious and repetitive task. So, when an analyst instructs VALCRI to "find me more reports like this …", the machine learning-based automation will trawl through large volumes of structured and un-structured data (e.g. free text) to retrieve, triage, collate, thematically analyse the data, and then combines and presents the reports in context of the crime problem being investigated, e.g. Comparative Case Analysis.

4. ETHICS, LEGAL AND PRIVACY BY DESIGN

In many LEA data analytics systems, once a person's data is enmeshed in the system-data networks, that person will continue to be linked to those criminal profiles. Such profiles will be used by the system to predict membership characteristics and to set up alerts for "persons of interest". This can lead to further stops and searches of the person, even though he may be innocent. This interferes with his private life. VALCRI advocates the need for 'computational transparency' as a mitigating approach: make visible the inner workings of 'black box' automated algorithms. A lower TRL prototype has been implemented in VALCRI to investigate how fine grain data access controls may be combined with computational transparency so that analysts and investigators are aware of the provenance of algorithm's computed results and protect the rights of individuals.

5. UP-SKILLING OF ANALYTIC ABILITIES

VALCRI has also identified and addressed varying deficiencies in the abilities of the intelligence analysis community. Some of this have been formalised in a new Master degree level analytics training course at Aston University in Birmingham, in partnership with the West Midlands Police; and some have been formalised into commercial intelligence analysis training packages focusing on analytic reasoning.

6. RESEARCH DATA

Partner AES worked with West Midlands Police to make anonymous three years of actual police data: over 1 million crime reports including structured and un-structured data, and over 6 million ANPR records. This data includes spelling errors, duplicates, similar but different data, and so forth. This dataset has been a crucial enabler.

B. DESCRIPTION OF WORK PERFORMED AND MAIN RESULTS (4000 CHARS // 3999 CHARS)

OBJECTIVE 1: Human Issues Framework DELIVERED

(a) Ethics, Privacy, Law. Comparative analysis of law in Germany, Belgium and UK, led to specification of legal requirements in VALCRI; Evaluated impact of removing ethically sensitive data from data analysis; Developed understanding of Ethics by Design in VALCRI; Set up *Ethics Working Group* in West Midlands Police to assess ethical issues in criminal intelligence analysis; (b) Cognitive bias and sense making. Operationalized insight, imagination, fluidity and rigour, transparency for experimental evaluation; Evaluated visualisation designs for insight, sense-making, cognitive bias, structuring of arguments.

OBJECTIVE 2: Analyst User Interface DELIVERED, INTEGRATED, TRL-5

A suite of AUI tools based around the reasoning workspace developed to orchestrate ML and database capabilities with interaction and visualisation functions to facilitate analytic reasoning and investigative sense

making. The AUI tools include maps and timelines, network evolution, dispersion diagrams, and statistical process charts, with touch-enabled, multiple-coordinated views. It is designed to encourage analysts to ask questions – an important part of sense making and coping with ambiguity.

OBJECTIVE 3: Semantic search and retrieval DELIVERED, INTEGRATED, TRL-5

Semantic search capabilities include an interactive dimension-reduction tool for data exploration and sense making with the Knowledge Generation Model. ML algorithms applied to read and select appropriate texts from crime reports, show feature set and create a first draft Comparative Case Analysis table. Associative Search identifies new associations or links between criminal entities by exploiting information, criminal behavior, modus operandi, geographical and temporal proximity, and associations between unsolved crimes and offenders to generate suspects lists.

OBJECTIVE 4: Crime situation re-construction DELIVERED, INTEGRATED, TRL-5

Developed a method for visual storytelling using argumentation theory to assist with the re-construction of crime situations. Explanations comprising fragments of data can then be formulated into defensible assessments. It enables analysts to record their evolving reasoning during investigations based on inferences from data, visualisations, and can be linked to conclusions through inferential networks.

OBJECTIVE 5: Secure, scalable and distributed architecture DELIVERED, INTEGRATED, TRL-5

The security architecture is implemented through OpenPMF with a Domain Specific Language DLS to configure Attribute and Proximity Based Access Controls (ABAC, PBAC) that translates human readable security policies into machine enforceable code; PET (Privacy Enhancing Technologies) to rapidly anonymise or pseudonymise data so it can be used without compromising privacy; HALA security test-bed set up for High Assurance Logging and Audit method based on a 'Vault' to provide hardware separation.

OBJECTIVE 6: Anonymised dataset DELIVERED, TRL-5, NOT RELEASEABLE

Partner WMP supplied three years of actual fine-grain police data comprising over 6 million crime reports and others, and over 58 million ANPR records. Led by AES, the data was anonymised at a deep level. This dataset was used in the development of the VALCRI system. However, internal tests showed that it was possible to de-anonymise the data. For confidentiality reasons, the data will not be released to the research community.

OTHER RESULTS

Harvester (SPACE). A stand-alone application where police users can search and mark up interesting text in PDF documents, harvest and store in a knowledge base.

Analysts Training Courses. The VALCRI Analytic Reasoning Training Curriculum (TN 13.4) has been developed into commercial courses: i-Intel's 3-day CPD courses in intelligence analysis have been evaluated with 123 LEA officers in 40 agencies in 13 countries; A Master-level Advanced Analyst qualification had been developed by AES, WMP, and Aston University, Birmingham.

Provenance. Recording, playback and state saving features integrated at TRL-5, with advanced analytic provenance being researched (TRL 2-3).

Dissemination and Engagement Activities. Published 119 peer-reviewed publications; VALCRI evaluated with 214 LEA officers from 50 agencies in 16 countries; demonstrated to an estimated audience of 1500 persons at 5 international scientific conferences, 1 EU project event and 5 intelligence events in 6 countries.

User trials. Installed at WMP, LPA, BFP police partner sites for evaluation with anonymised data. Installed at Pasco Sheriff's Office, FL, and the Metropolitan Police, London, for evaluation with actual data.

C. FINAL RESULTS AND IMPACTS
(4000 CHARS // 3996 CHARS)

1. The main outcome is an integrated multi-application criminal intelligence analysis system at TRL-5. Using a cognitive engineering approach, we implemented the concept of a joint cognitive system, demonstrating how mixed-initiative systems can be developed to enable proactive and reactive system behaviours to create a human-machine team. This creates a test-bed for further research: (i) study the impact on operational use of criminal intelligence analysis systems of how the laws and privacy regulations are implemented, (ii) advancement of the semantic search algorithms, (iii) inclusion of formal concept analysis techniques to associative search, (iv) application of hybrid AI techniques to semantic knowledge extraction, (v) investigate alternative methods for storytelling and argumentation to support work with uncertainty, ambiguity and deception, (vi) It will also create opportunities to re-factor the integration platform code to enable plug and play capability, (vii) provide an environment for police to experiment with new methods based on the new VALCRI capabilities, (viii) use behavioural markers for automatic classification of analytic reasoning activities from user interactions with the system.
2. The VALCRI system is not one single application, but a complex multi-application industrial scale system using the following technology stack: Java, Javascript, GWT and ERRAI, Docker containers, RESTful interfaces, Jena/Fuseki RDF triple store, MongoDB, SQL Postgres DB with Elasticsearch, OpenPMF and a Central Authority Service, Graylog, NLP pipeline for concept extraction, ML-based semantic search functions.
3. Training courses have been developed around the analytic reasoning research in VALCRI. These courses are in high demand. New insights about analytic reasoning and new VALRI technologies have created opportunities for new techniques to be developed. By embedding the knowledge into CPD and Master-level courses, opportunities are being created for propagating the knowledge beyond the police intelligence communities.
4. Research into legal, ethical and privacy requirements in Europe has identified key issues and translated them into system design specifications and implementation trade-offs e.g. how to show data or node in a network visualisation graph that may be confidential for security, privacy or ethical reasons?
5. Cognitive engineering research has helped us understand how analysts think. This has enabled us to design how software might facilitate the reasoning in uncertain, ambiguous and deceptive environments through designs that encourage the asking of questions.
6. Partners have implemented different methods for semantic knowledge extraction and associative search. This opens opportunities for new research e.g. computational transparency – how we make the results of black box automated analyses understandable and verifiable by users; computational steering of algorithms such as the use of sub-space clustering methods to discover low frequency but operationally significant events; use VALCRI as a test-bed for investigating hybrid intelligent technologies in a joint cognitive system approach; navigating uncertainty when using the products of such methods given ambiguous and deceptive situations.

7. WMP provided real data that was large and complex enough for developing real systems. The data was anonymised and used to develop the VALCRI prototype system. However, internal evaluations determined that the data could be de-anonymised due to the richness of the data contained in the un-structured text. Therefore the anonymised data cannot be released to the research community as originally planned.
8. Exploitation. A variety of IP has been produced with plans for commercial exploitation and further research. Instead of tying partners down to the usual single exploitation plan, an exploitation agreement was reached for VALCRI that freed partners to exploit the IP they owned as they wish. The 9-point agreement is based on three ideas (a) freedom to commercially exploit IP that is individually owned, (b) freedom to join another partner to create products or services that create commercial value, and (c) profits to be shared only by those who generated the profit.
9. Impact. Most significant is the independent decisions by the Metropolitan Police Service London and the Pasco County Sheriff's Department in Florida to adopt the VALCRI system for trials with actual data. The VALCRI system was installed at both sites. They are in the process of ingesting actual data to solve actual cases. They are not members of the project consortium and are not obliged to adopt nor trial the VALCRI system.

D. PROJECT PUBLIC WEBSITE ADDRESS:

valcri.org

SECTION 2
CORE OF THE REPORT:
Project Objectives, Work Progress, Achievements, and Project Management
Period 3, M17 – M50, Sep 2014 – June 2018

2.1 INTRODUCTION

2.1.1 Project Objectives

The purpose of the VALCRI project is to develop the next generation criminal intelligence analysis system for European LEAs. Working closely with three European police forces, the project researched and developed at TRL-5, an integrated system of over 75 software components of advanced data processing, analytic and sense-making tools. It includes multiple applications spanning strategic intelligence analysis to tactical intelligence and individual case management.

To meet this goal, the following Project Objectives were defined for the project and are detailed below. The results of the project have been summarised against each objective in Section 1.

Objective 1 Human Issues Framework

> To develop a Human Issues Framework that combines various human cognition, bias mitigation, social and legal factors into a single principled framework that developers can use to guide the system design.

Objective 2 Analyst Reasoning Workspace-based Analyst UI

> To develop an advanced, novel interactive visualisation-based Analyst User Interface that is guided by the concept of the Analyst Reasoning Workspace.

Objective 3 Real-Time Semantic Search And Retrieval

> To develop a real-time semantic search and retrieval capability.

Objective 4 Crime Situation Re-Construction

> To develop a crime situation re-construction function that is based on spatial-temporal and network technologies for representing important socio-cultural and organizational constructs that are crucial for understanding of the crime and circumstances, and to then project future possibilities.

Objective 5 Secure And Scalable Distributed Processing Architecture

> To design and develop a secure and scalable distributed processing architecture that is compatible with the requirements of visual analytic dynamic user interaction and analysis.

Objective 6 Anonymised, Realistic, Machine Deployable Dataset

> To develop an anonymised machine-deployable dataset, based on real crimes, that is of adequate size and complexity, and to subsequently develop from that process the ability to create and synthesise data that is good as the real data. This dataset and process will be made available to the research community to advance research in security.

2.1.2 Problems faced by criminal intelligence analysts

The problems faced by criminal intelligence analysts may be described and summarised below, and is illustrated by the keyhole example in Figure 1.

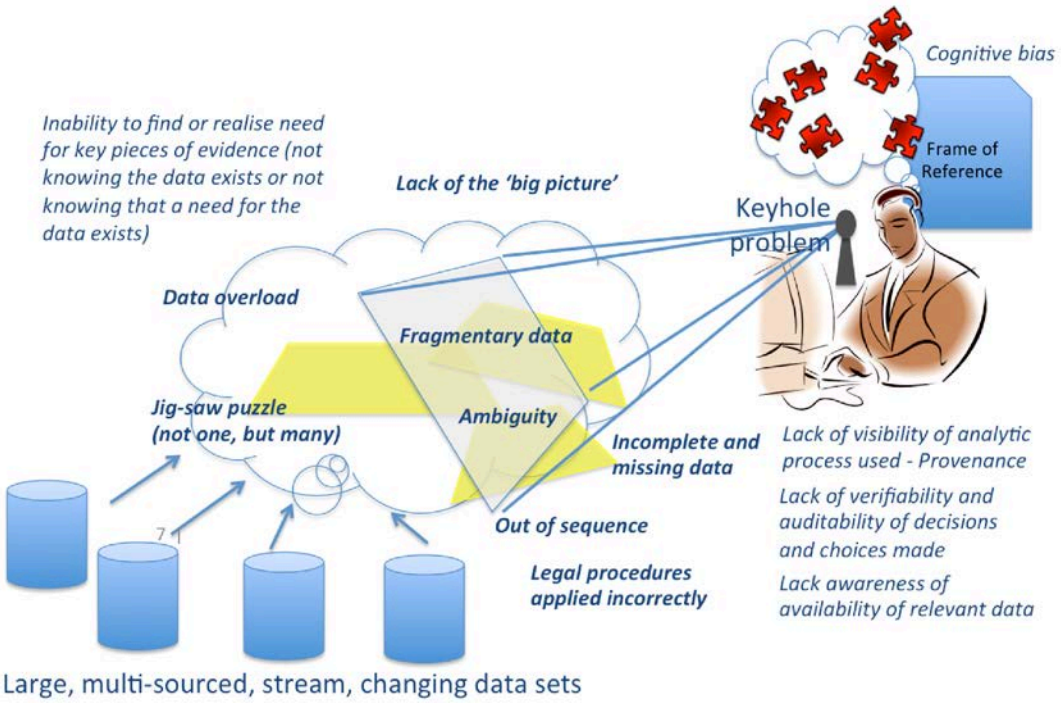


Figure 1. The difficulties faced by criminal intelligence analysts (extended from Wong and Varga, 2012)

- Analysts and investigators have to make sense of large volumes of data that is from multiple sources and heterogeneous; with data that is often missing, out of sequence, deceptive, uncertain and ambiguous.
- Crimes are increasing and changing faster than we can keep up such that "…the planning of terrorist attacks have accelerated from weeks to a handful of days" (Andrew Parker, Director MI5, 2017[1]).
- Fragmentary information. "One of the main challenges we've got is that we only ever have fragments of information, and we have to try to assemble a picture of what might happen, based on those fragments." (Andrew Parker, Director MI5, 2017[1]).
- The failure of imagination, or the failure to figure out and explain how those fragments of information might be connected to explain or anticipate an event or incident.
- Burdensome and repetitive tasks. Analysts need to repeatedly access and re-organise data, and to "… describe inputs-outputs, [assess and decide] what to filter, what data transforms [to make] … so as to make it easier to pick out meta-patterns …" (Gleichauf, 2011[2]). These tasks are difficult to automate because they require human intellectual reasoning.
- Lack of functional integration between tools. Analysts cycle through different pieces of software at a rapid pace, often switching between applications once every 2-3 minutes to gather, process and structure the necessary data (Pallaris & Bielska, 2015[3]), then manually analyzing the digital data (Babuta, 2017[4])
- Police lack advanced tools to trawl and analyse unstructured data (Babuta, 2017[4])

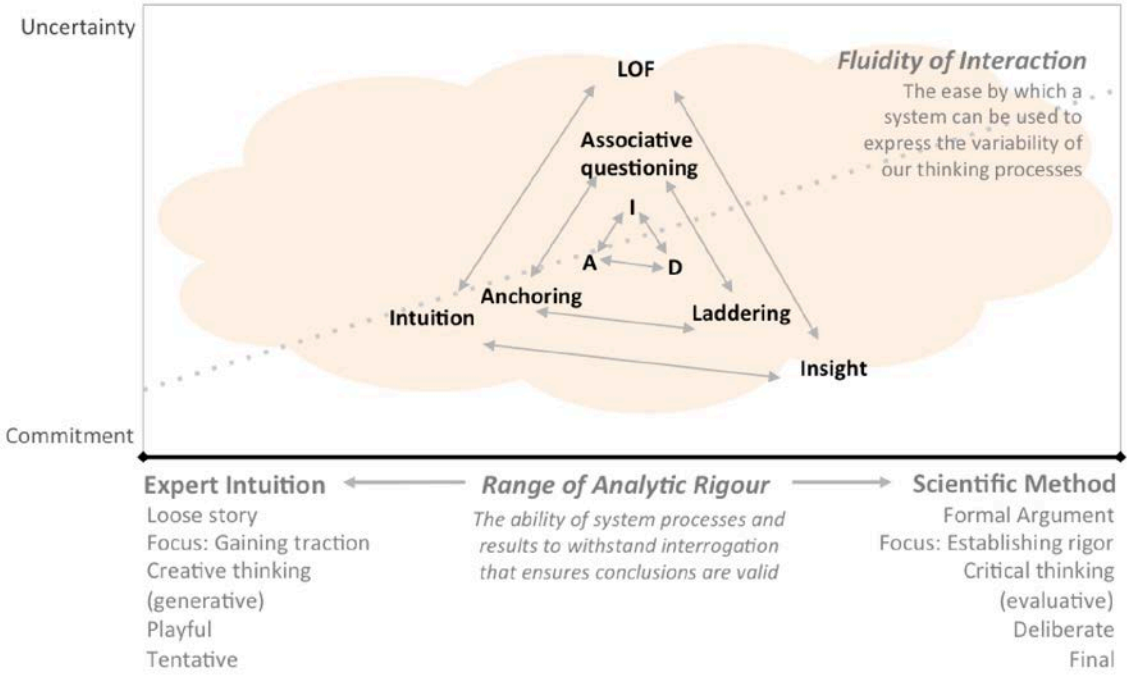
[1] Corera, G. (2017). MI5 boss Andrew Parker warns of 'intense' terror threat. BBC News, 17 October 2017, https://www.bbc.com/news/amp/uk-41655488.
[2] Gleichauf, B. (2011). Beyond Data. IQT Quarterly, 2(4), 2-3.
[3] Pallaris, C., & Bielska, A. (2015). Technical Note 13.1: Zero Point Measurement, Version 1.0, 1 June 2015, VALCRI FP7-608142. 75 pages. Report submitted for Project VALCRI (FP7-IP-608142). 1 June 2015.
[4] Babuta, A. (2017). Big data and policing: An assessment of law enforcement requirements, expectations and priorities. RUSI Occasional Paper, Royal United Services Institute, September 2017, 40 pp.

- Analysts collaborate, but mainly after performing their own analysis and thinking about the problem. There is a need for collaboration while preserving individual analytic reasoning, but there are few tools that facilitate collaborative discussions while preserving the critical individual thinking and reasoning abilities of individual expert analysts (Smallman, 2008[5] and (Rossy & Ribaux, 2014[6])

2.1.3 The VALCRI Concept

Given the nature of the problems faced by analysts and investigators, the VALCRI project developed a next generation criminal intelligence analysis system for European LEAs. Working closely with three European police forces, the project researched and developed at TRL-5, an integrated system of over 75 software components of advanced data processing, analytic and sense-making tools that are intended to facilitate human reasoning and analytic discourse. By being tightly coupled with semi-automated human-mediated semantic knowledge extraction, VALCRI will respond to human analysts in both a proactive and reactive manner, and work with analysts as a human-technology team. It responds and anticipates needs as a Joint Cognitive System (Hollnagel and Woods, 2005[7]). It includes multiple applications that span strategic intelligence analysis to tactical intelligence and individual case management.

The Fluidity and Rigour Model, see Figure 1, or FRM (Wong, 2016[8]), is a model intended for interaction designers. It highlights the variability of analytic reasoning strategies employed by analysts during criminal intelligence and investigative analysis, and describes the range of visualisation and interaction methods needed for criminal intelligence analysis systems.



[5] Smallman, H. S. (2008). JIGSAW - Joint Intelligence Graphical Situation Awareness Web for Collaborative Intelligence Analysis. In M. P. Letsky, N. W. Warner, S. M. Fiore, & C. A. P. Smith (Eds.), Macrocognition in Teams: Theories and Methodologies. Gower House, Croft Road, Aldershot, Hampshire GU11 3HR, England: Ashgate Publishing Company.

[6] Rossy, Q., & Ribaux, O. (2014). A collaborative approach for incorporating forensic case data into crime investigation using criminal intelligence analysis and visualisation. Science & Justice, 54, 146-153.

[7] Hollnagel, E., & Woods, D. D. (2005). Joint Cognitive Systems: Foundations of Cognitive Systems Engineering. Boca Raton, FL 33487-2742: CRC Press, Taylor and Francis Group, LLC.

[8] Wong, B. L. W. (2016). Fluidity and Rigour: Addressing the Design Considerations for OSINT Tools and Processes. In B. Akhgar, P. S. Bayerl, & F. Sampson (Eds.), Open Source Intelligence Investigation: From Strategy to Implementation (pp. 167-189). Cham, Switzeland: Springer International Publishing AG.

Figure 1: Fluidity and Rigour Model.
Legend: Inference making: Induction (I), deduction (D) and abduction (A); User strategies: Anchoring, Laddering, Associative question; Cognitive Acts: Intuition, LOF (Leap of Faith), and Insight

The reasoning strategies invoked by analysts range from making guesses and suppositions that enable storytelling when very little is known, to reasoning strategies that lead to rigorous and systematic evaluation of explanations that have been created through the analytic process. In the FRM, we define fluidity as the ease by which a system can be used to support the variability of thinking strategies expressed in the analytic reasoning process; and by rigour we mean the extent to which analytic methods and processes produce results and conclusions that are valid and can stand up to interrogation.

2.1.3.1 Sources of Variability in Analytic Reasoning

The Law of Requisite Variety, states that "… R's capacity as a regulator cannot exceed its capacity as a channel for variety" (Ashby, 1958[9]). Re-stated, "… the variety of variability of a process to be controlled must be matched or exceeded by the variety of variability of the controlling entity…". This refers to the variety of situations a system designed to control or support a process must be capable of controlling or supporting. The lack of compatibility between the variety of situations a process can produce, and the ability of a controlling system to support or accommodate that variety will invariably lead to system failures or sub-optimal performance. Systems designed to support intelligence analysis need to support not only the observable tasks of information search and retrieval and data analysis, but also the much less observable but crucial thinking and reasoning processes. These are the cognitive processes that determine the logic and how sensible are the narratives created to explain the clues that that present themselves in an investigation.

Analysts make use of various inference making strategies - induction, deduction and abduction - depending upon what data they have, the rules for interpreting the data they are starting with, and the conclusions they would make or would like to make (Wong & Kodagoda, 2016[10]). The early stages of an investigation are often characterised by a lack of information and the need to imaginatively create plausible stories or explanations, such as abductive inferences to initiate possible lines of inquiry. Analysts also practise a mix of critical thinking and storytelling. In this process they elaborate, question, and often reframe and discard explanations (Klein, et al, 2014[11]), with some evolving into stronger, well-justified explanations that are robust enough to withstand interrogation (Rooney, 2014[12]). Wong & Kodagoda (2016) present other aspects of the analytical reasoning process: anchoring, laddering, and posing associative questions. Analysts engage in a process of anchoring to gain traction and initiate inquiry. They then engage in a laddering process where they develop explanations to extend or elaborate their ideas into new understanding. They complement the anchoring and laddering activities by associative questioning to discover what else might exists. Police analysts are taught, for example, the 5WH model – who, what, where, when, why and how – to activate divergent thinking pathways that may lead to un-expected associations; which through intuition, could spur the recognition of un-anticipated patterns (Gerber et al., 2016[13]) across different data sets. Often the problem is not 'joining the dots' – but to imagine more informative ways to connect them to create better insights under information-sparse, uncertain, and ambiguous conditions. This requires some degree of creativity when trying to imagine plausible explanations, and for getting the cognitive traction required for gaining further insight.

We use the x-axis of Figure 1 to illustrate the range of analytic rigour that may be applied to the analytic reasoning process. At the start of an inquiry there is usually very little known about a case. It is therefore of

[9] Ashby, W. R. (1958). Requisite variety and its implications for the control of complex systems. Cybernetica, 1(2), 83-99.

[10] Wong, B. L. W., & Kodagoda, N. (2016). How analysts think: Anchoring, Laddering and Associations Proceedings of the Human Factors and Ergonomics Society 60th Annual Meeting, 19-23 September 2016, Washington, D.C., USA (pp. 178-182): SAGE Publications.

[11] Klein, G. (2014). Seeing what others don't: The remarkable ways we gain insights. London, England: Nicholas Brealey Publishing.

[12] Rooney, C., Attfield, S., Wong, B. L. W., & Choudhury, S. T. (2014). INVISQUE as a tool for intelligence analysis: the construction of explanatory narratives. International Journal of Human Computer Interaction, 30(9), 703-717.

[13] Gerber, M., Wong, B. L. W., & Kodagoda, N. (2016). How analysts think: Intuition, Leap of Faith and Insight Proceedings of the Human Factors and Ergonomics Society 60th Annual Meeting, 19-23 September 2016, Washington, D.C., USA (pp. 173-177): SAGE Publications.

little use to treat information and inferences rigorously as the analyst is still trying to understand what the data means and whether it is sensible to create an argument. The type of thinking and reasoning employed by the analyst at this stage may be characterised as being creative, having to deal with high uncertainty as there are many unknowns and missing data. The need at this stage is to gain traction and to get the investigation started. Analysts engage in the tentative and playful generation of plausible stories and hypotheses that may account for their observations. They tend not to commit to a single explanation and are likely to explore alternatives.

At the high rigour end of the spectrum, the type of thinking and reasoning required may be characterised as 'critical thinking', evaluative, deliberate, and final. As an investigation approaches the closing stages, most of the data required will be known. It is then possible to rigorously structure, organise, or analyse the data, and to make sure that every conceivable logical discussion can be evaluated. By this stage, analysts would have employed a variety of structured analytic techniques (see for example, Heuer and Pherson, 2014[14]) to establish strong and rigorous arguments. Then having done all the analyses and checks – one would be committed to an explanation.

2.1.3.2 Fluidity to interact with the variety of analytic tools

Fluidity is the ease by which a system can be used to support the variability of thinking strategies demonstrated by analysts in the analytic reasoning process (Wong 2016[15]). To achieve this, the interaction and visualisation methods need to enable the analyst to seamlessly transition within and between the tools needed by the different analytical reasoning strategies. Elmqvist et al (2011)[16] has explained that the basic requirement for fluidity is for users to feel that they are directly participating in the interface, where users feel that are able "… to directly 'touch' and manipulate the visualization instead of indirectly conversing with a user interface". Fluidity in a user interface therefore "… involves achieving a sense of immersion, a first-person-ness and direct engagement with the objects and the visualizations" creating an embodiment with the user interface to create a sense of 'being in the flow', directly benefiting analytic performance (Bederson, 2004). Pike et al. (2009) advocates that visual displays must be "embedded in an interactive framework that scaffolds the human knowledge construction process with the right tools and methods to support the accumulation of evidence and observations into theories and beliefs". To achieve this level of interactivity, we also ensure that real-time responses are close to the 100ms recommendation (Kalawsky, 2009[17]). The aim is to create a tight loop between query and analysis to support Niesser's (1976[18]) perception-action cycle to achieve a level of engagement that may be interpreted as a real-time dialogue between the user and the machine. Impediments in the interaction would obstruct the analytic discourse (Dykes, 2005[19]), making the interface frustrating to use, leading to higher cognitive loads, activation of cognitive biases (Munzner 2014[20]), and poorer situation awareness.

2.1.3.3 Requirements For Fluidity And Rigor

In what ways might technology assist in supporting the variability of the analytic reasoning process? Based on a number of studies we conducted: e.g. focus group studies with 20 intelligence analysts (Wong and Varga, 2012[21]); think-aloud studies with analysts and librarians performing simulated intelligence tasks (Rooney et

[14] Heuer, R. J. J., & Pherson, R. H. (2014). Structured Analytic Techniques for Intelligence Analysis (2nd ed.). Los Angeles, CA: SAGE CQ Press.
[15] Wong, B. L. W. (2016). Fluidity and Rigour: Addressing the Design Considerations for OSINT Tools and Processes. In B. Akhgar, P. S. Bayerl, & F. Sampson (Eds.), Open Source Intelligence Investigation: From Strategy to Implementation (pp. 167-189). Cham, Switzeland: Springer International Publishing AG.
[16] Elmqvist, N., Moere, A. V., Jetter, H.-C., Cernea, D., Reiterer, H., & Jankun-Kelly, T. (2011). Fluid interaction for information visualization. Information Visualization, 10(4), 327–340.
[17] Kalawsky, R. S. (2009). Gaining Greater Insight through Interactive Visualization: A Human Factors Perspective. In R. Liere, T. Adriaansen, & E. Zudilova-Seinstra (Eds.), Trends in Interactive Visualization (pp. 119–154). Springer London.
[18] Neisser, U. (1976). Cognition and Reality. New York, NY: W.H. Freeman and Company.
[19] Dykes, A. M. MacEachren, & M.-J. Kraak (Eds.), Exploring Geovisualization (pp. 265–292). Elsevier.
[20] Munzner, T. (2014). Visualization Analysis and Design. Boca Raton: A K Peters/CRC Press.
[21] Wong, B. L. W., & Varga, M. (2012). Blackholes, keyholes and brown worms: challenges in sense making Proceedings of HFES 2012, the 56th Annual Meeting of the Human Factors and Ergonomics Society, Boston, MA, 22-26 October, 2012 (pp. 287-291). Santa Monica, CA: HFES Press.

al, 2014[22]; Kodagoda et al, 2013[23]); and cognitive task analyses with analysts from three major police forces in Europe (e.g. Wong and Kodagoda, 2016[24]; Gerber et al., 2016[25]), we summarise below the key design requirements for fluidity in analytical reasoning.

The tools at the 'loose story' end of the rigour spectrum should be different from the tools supporting more rigorous approaches on the other end of the spectrum. At the 'loose story' end, the tools should enable the analyst to express the creative, generative, chaotic and tentative nature of reasoning by enabling playful experimentation that is needed for one to gain cognitive traction with which to start an idea to pursue a line of investigation. It should facilitate associative and divergent thinking by anticipating and presenting information that might be needed next.

Any software tool should enable analysts to transition fluidly between critical thinking methods, and methods for creative exploration, hypotheses formulation, and storytelling. The tools should enable the analysts to transition seamlessly from early analyses that led to tentative possibilities, to assemble data and ground the explanations so that narratives could be developed into strong arguments. Analysts should also be able to transition fluidly between different forms and assemblies of explanations, outcomes, assessments, and analyses, to uncover other possibilities that may lie within the data.

2.1.3.4 Support How Analysts Think, Rather Than What Analysts Do

If VALCRI were designed to mainly support what analysts do, then the system would primarily support established workflows defined by tasks that analysts currently do, i.e. we automate current tasks. Instead by designing support for how analysts think, the VALCRI system is better able to respond to the variety of sense making, reasoning and inference making and problem solving strategies presented by human analysts. Through our research we found that analysts use thinking and reasoning methods that are similar with those of archaeologists. They use a combination of abductive, inductive and deductive inference strategies depending upon data availability, goals they wish to satisfy, their experience and prior knowledge. They create anchors to start inquiries, elaborate understanding, and search for associations between people, places, activities and information. Sometimes they harness their expert intuition to address issues of ambiguity in order to take leaps of faith that lead to new insights that can be systematically tested with the scientific method.

2.1.3.5 Facilitate Expert Intuition to Scientific Method

In many crime investigations, analysts are often only ever presented with fragments of data from which they have to create an understanding of the situation and to anticipate what might happen in the future. Expert intuition is very useful in creating "hunches", or early and tentative hypotheses, in such information sparse and often ambiguous situations. However hunches by themselves can be error prone and subject to various cognitive biases. Therefore to be effective, VALCRI has designed quick yet powerful ways for analysts to test their hunches using the scientific method. The VALCRI UI has been designed to enable the fluid and low effort transition between expert intuition and the scientific method, so that poorly supported hypotheses can be willingly discarded and new ones formed.

2.1.3.6 Humans decide, machines do the heavy lifting

VALCRI takes the view that technology works best when it augments human cognitive abilities rather than replacing it. In addition, in law enforcement, it must be possible for humans to be held accountable for decisions that could lead to the arrests and imprisonment of individuals. VALCRI has been designed so that

[22] Rooney, C., Attfield, S., Wong, B. L. W., & Choudhury, S. T. (2014). INVISQUE as a tool for intelligence analysis: the construction of explanatory narratives. International Journal of Human Computer Interaction, 30(9), 703-717.
[23] Kodagoda, N., Attfield, S., Wong, B. L. W., Rooney, C., & Choudhury, T. (2013). Using Interactive Visual Reasoning to Support Sense-making: Implications for Design. IEEE Transactions on Visualization and Computer Graphics, 19(12), 2217-2226.
[24] Wong, B. L. W., & Kodagoda, N. (2016). How analysts think: Anchoring, Laddering and Associations Proceedings of the Human Factors and Ergonomics Society 60th Annual Meeting, 19-23 September 2016, Washington, D.C., USA (pp. 178-182): SAGE Publications.
[25] Gerber, M., Wong, B. L. W., & Kodagoda, N. (2016). How analysts think: Intuition, Leap of Faith and Insight Proceedings of the Human Factors and Ergonomics Society 60th Annual Meeting, 19-23 September 2016, Washington, D.C., USA (pp. 173-177): SAGE Publications.

humans and machines do what each are good at: Humans make decisions under ambiguity, while machines perform those many tedious and repetitive tasks, but in an intelligent manner.

VALCRI uses ML techniques to support the search and retrieval of semantically similar data across the different data sets and then orchestrates this with various technologies such as the user interface, database and ontology knowledge-bases, privacy and fine grain data access controls, and secure logging to ensure a fluid and seamless user experience between applications. So, when an analyst instructs VALCRI to "find me more reports like this …", the machine learning-based automation will trawl through large volumes of structured and un-structured data (e.g. free text) to retrieve, triage, collate, thematically analyse the data, and then combines and presents the reports in context of the crime problem being investigated, e.g. Comparative Case Analysis, Semantic Similarity Space, Associative Search, maps and timeline analysis.

2.1.3.7 Ethics, Legal and Privacy by Design

The VALCRI system design had been heavily influenced by the Human Issues Framework or HIF, which has identified important societal and psychological constraints, such as the principles and the law regarding data protection, privacy and the ethical use of data and the system. In many LEA data analytics systems, once a person's profile or data is enmeshed in the system-data networks, that person will always continued to be linked to those criminal profiles because the data was used to create the original profiles, and the profile created will be used for calculating or reinforcing a profile which the system uses to predict membership characteristics and may set up alerts to designate people in that categories as "persons of interest". This can lead to further stops and searches, and increasing their occurrences in the databases. Such issues will affect the rights of innocent persons to a private life and freedom from interference.

Although no agreed solution yet exists, one approach advocated by the VALCRI project is the need for 'computational transparency' as a way to make visible the inner workings of 'black box' automated algorithms that underlie many present day data analytics systems. A lower TRL prototype has been implemented in VALCRI to investigate how fine grain data access controls may be combined with computational transparency so that analysts and investigators are aware of the provenance of algorithm's computed results.

2.1.3.8 Visibility

We are also very cautious to avoid creating automation stovepipes that can lead to the well-known problems associated with 'automation surprise' (Sarter et al., 1997)[26]. This occurs when automation produces unexpected outcomes or fail, causing a loss of awareness and leading to possible loss of control. In automated intelligence analysis systems, this can happen due to algorithmic opacity, or the lack of computational transparency -- i.e. the provision of information in a human-understandable manner that makes visible how the underlying computation produced its recommendations. In VALCRI, we designed and implemented several techniques at lower technology readiness levels (TRL) to demonstrate how computational transparency can be designed into computationally complex intelligence analysis systems.

2.1.3.9 VALCRI as part of a Joint Cognitive System

Rather than just humans working and interfacing better with technology, Joint Cognitive Systems (JCS) characterises a system that combines humans and an artefact within an organization in ways that can "modify its behaviour on the basis of experience so as to achieve specific anti-entropic ends" (Hollnagel & Woods 2005, p22). This requires the cognitive system to be capable of both proactive and reactive behaviours. VALCRI therefore has functions that are anticipatory (e.g. given a crime report, show me more like this crime report), and also proactive (e.g. once a selection of similar reports are presented, and the analyst selects one, VALCRI shows other associations). These functions represent the requisite variety needed to support the variety of search, thinking and reasoning functions performed by analysts in investigative decision making environments. Thus, changes in the human (sub-) system, such as changes in understanding,

[26] Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation Surprises. In G. Salvendy (Ed.), Handbook of Human Factors and Ergonomics (2nd ed., pp. 1926-1943). New York: John Wiley and Sons.

or new intelligence assessments becoming available, drive changes in the technology system to anticipate the retrieval and presentation of information needed by the human. This facilitates an active symbiotic dialogue between two parties, and thus enables VALCRI to provide a new generation of analytic capabilities that might enhance the analytical reasoning of analysts in intelligence and investigative environments.

2.1.3.10 Tactile Reasoning - Interacting with the VALCRI Joint Cognitive Systems

Tactile reasoning is an interaction technique that supports analytical reasoning by the direct manipulation of information objects in the graphical user interface by the end user (Takken & Wong, 2015). As with the alphabet tiles used in the game of 'Scrabble', VALCRI uses tiles to represent items of crime information, which can be directly and freely manipulated, moved and rearranged. Explained as epistemic actions (Kirsh & Maglio, 1994; Kirsh, 1995), this assists the analysts in modifying their work environment in order to support the externalisation of their thinking and reasoning processes, making the complex mental tasks invoked during investigative decision making tractable. The design of the interaction methods and ways in which the information tiles are visualised and laid out are based on a variety of human factors principles, including: Emergent Features, Gestalt, and Proximity-Compatibility.

2.1.3.11 Imagination, Insight, Fluidity and Rigour, and Transparency

Four sensemaking approaches have been explored in the design of a synergistic JCS, in order to create alternative perspectives that can lead analysts to new understandings of their crime data.

a. Encourage Imagination

We define imagination as the ability to creatively and fluently generate new possibilities, ideas or concepts beyond what is presented. In VALCRI, we wish to encourage divergent thinking where analysts can be flexible in their approaches to interpreting data, as described by the Osborne-Parnes Creative Problem Solving model (Fontenot 1993). Our studies show that at the start of investigations, data are sparse, uncertain, ambiguous, filled with unknown missing elements, and thus require a high degree of creativity and imagination to playfully and tentatively generate plausible interpretations and explanations.

b. Enable Insight

Klein explains that "Insights change our understanding by shifting the central beliefs – the anchors – in the story we use to make sense of events … our new understanding can give us new ideas about the kinds of actions we can take; … [and make it] easy for users to shift goals and plans without getting disoriented" (Klein, 2013, p.148). In VALCRI, tools and techniques are needed in the context of the JCS paradigm to design externalisation of sensemaking processes – orchestrating the timing and response type (reactive, proactive) of the underlying machine learning (ML), as well as employing various human factors principles based interaction and visualisation techniques.

c. Engage with fluidity and rigour

Fluidity and rigour are two conflicting design requirements that characterise the nature of intelligence analysis work. By fluidity we mean the ease by which a system can be used to express the variability and uncertainty of our thinking processes; and by rigour we mean the ability of the system processes and results to withstand interrogation (Wong, 2016). Analysts make use of a combination of inference making strategies - induction, deduction and adduction – depending upon what data they have, the rules adopted for interpreting the data, and the premise they start with and the conclusions they would make or would like to make. Analysts are constantly trying to explain a crime situation, re-constructing the situation from pieces of data and from inferential claims, and then carrying out searches or further analysis to find necessary data to back the claims. Their analyses often begin as highly tentative explorations based on very weak data or hunches. As they explore various possibilities, making conjectures, suppositions and inferential claims, they then connect with further data, testing their relevance and significance, elaborate, question, and often reframe and discard their ideas, and eventually build up a story so that it eventually becomes robust enough to withstand interrogation. As greater understanding of the data emerges, analysts then engage in

techniques such as ACH (Analysis of Competing Hypotheses) (Heuer, 1999) and Wigmore charting (Wigmore, 1913; Goodwin, 2000) to critically and rigorously test the claims, data, evidence and supporting arguments.

In VALCRI, we are building tools that will fluidly link the generative, creative, playful and tentative exploration in ways that encourage the exploration of alternatives, appreciation of the context, and the avoidance of pre-mature commitment, with the critical inquiry that leads to deliberate, evaluative, and rigorous explanations.

d. Ensure transparency

A lack of transparency of the crime analysis process makes it difficult to hold public officials accountable for their decisions. In VALCRI, transparency is needed in two areas. The first is *computational transparency*, which considers how computational processes can be made visible to inspection to minimise the effects of automation surprises, and the need to explain assumptions made and data used in the computation of recommendations and probabilities. The second is *analytical transparency*, which considers the Intelligence analysis process which is largely a non-observable artefact of the human mind. Analysts are not practised in describing how they arrive at their decisions. We are therefore exploring ways whereby the process of analytical reasoning may be made visible to inspection by colleagues, trainers and/or supervisors – i.e. the chain of decisions that led to a final outcome. We refer to this as the *conclusion pathway*. One approach is to design methods for externalising the analytic reasoning process so that analyst can articulate their conclusion pathways through visible artefacts which will make the analytic reasoning process open to inspection. Such facilities can help to ensure that analysts are fully accountable.

2.1.4 Summary of Achievements

2.1.4.1 The VALCRI System Prototype

VALCRI has been designed around a knowledge extraction engine which uses machine learning techniques for semantic similarity analysis undertaken in both reactive and proactive modes with the analyst. Crime-related data are stored in two databases: an unstructured database (UDB) for free text fields and video data, and a structured database (SDB) for structured text and data extracted by parsing free-text. A combination of Open Source technologies is being adapted and integrated to undertake varied forms of data analysis across different crime categories and multiple data sets. As described in section 1.2 above, components have already been built for *semantic data mining, associative search,* and *Comparative Case Analysis (CCA)*. Further details of the prototype are described below.

2.1.4.2 VALCRI Technology Readiness Level (TRL)

The majority of the VALCRI prototype will be functionally integrated into a single TRL-5 platform by the extended project end date. The problems addressed by software components developers have proved to be more difficult than anticipated and so the entire system is now being developed primarily at TRL-5, with those components at lower TRLs being made available on separate branches.

2.1.4.3 VALCRI User Interface

The VALCRI user interface (UI) design is based on the concept of tactile interaction, driven by a visual analytic perception-action cycle, guided by the fluidity and rigour model. The design has been further informed by principles and requirements from user practice, human factors and psychology principles, and our own studies of analytic reasoning and sensemaking. The design has been implemented in the GWT (Google Web Toolkit) environment, within which we have developed the Analyst User Interface (AUI). This manages the windowless AUI environment where data records fluidly transition into abstract visual representations on dual screens which can be manipulated to carry out numerous analytical operations. The AUI is further integrated with dynamic visual querying techniques for fast response times across multiple-coordinated and faceted views involving maps and timelines, statistical process charts, crime hotspot analyses, and dispersion diagrams. These tools help the analyst to generate and test the logic of explanations that connect assemblies

of propositions, data and assumptions, structured and presented in ways to facilitate inference making, storytelling, the creation of explanations, and the formulation and testing of hypotheses.

2.1.4.4 Ethical, privacy and legal issues

Studies have been undertaken to compare applicable laws in Germany, Belgium and the UK to determine how legal principles may be implemented within the prototype. These include: purpose limitation, data minimization, the treatment of data subjects, handling of ethically sensitive data, and data storage and deletion. These requirements are being implemented through a fine-grained access control method based on the recently patented OpenPMF (Policy Management Framework) by a project partner. Access rules can be configured for use by the data access software. Other partners are collaborating on PET (Privacy Enhancing Technologies) which can rapidly anonymise or pseudonymise data so it can be used without compromising privacy. At a higher conceptual level, we are currently developing and trialling an *Ethics by Design methodology* for VALCRI, which addresses concrete ethics-related problems encountered in the course of developing the prototype. Following its participation in VALCRI, the West Midlands Police have set up an *Ethics Working Group* to identify and assess ethical issues related to criminal intelligence analysis operations, during both analysis and investigations.

2.1.4.5 Security Test-bed for High Assurance Logging and Audit

A security test-bed has been set up in a Berlin location by partner Object Security to develop and test secure logging method, referred to as the High Assurance and Logging Auditing to create secure crime analysis logs that cannot be tampered with. Object Security has designed the 'Vault' which provides hardware separation through trusted key storage, high performance, trusted crypto operations, trusted mass storage, trusted user I/O, and trusted processing. This permits all system log data to be sent in real time to the Vault from application/middleware, and optionally from kernel modules.

2.1.4.6 Patent

Our partner Object Security has registered a patent with the US Patent and Trademark Office based on research undertaken as part of the VALCRI project. It described a system and method for managing the implementation of policies in an IT system by automatically or semi-automatically generating machine-enforceable rules and/or configurations. This is being adapted to translate European laws and regulations into rules that can guide access to crime-related data in VALCRI.

2.1.4.7 Anonymised Datasets

Three years of police data, comprising over 6 million crime reports, stop and search, stolen property reports, intelligence reports, nominal, and custody reports, and over 58 million ANPR records, have been anonymised at a deep level by VALCRI partner AES from raw data supplied by West Midlands Police. Unlike most publicly available crime data, these are fine-grained, and are being used by VALCRI partners to undertake research, and develop and test the prototype in readiness for operational use by LEAs. Tests are currently under way to determine whether the procedures used to anonymise these data can resist de-anonymisation. At the end of the project, the VALCRI data set will be made available to the broader research community.

2.1.4.8 Development Environment and User Access

The VALCRI software development environment is hosted at three partner locations: London, Linkoping, and Brussels. The primary project source code is stored at Middlesex University, and managed through GitLab. Developers with sufficient machine resources can pull the code from Middlesex and images from Space and run the full stack locally. Resource-limited partners can get some of the images to run locally, and connect to running versions of the other images hosted at Linkoping. All users, whether analysts or non-technical partners, can accessing VALCRI in two ways: use a web browser to access a release version (TRL-5) on a server hosted at SPACE (via VPN access); or download and run it locally on their own machines.

2.1.4.9 VALCRI Deployed at Police locations and Consortium Partners' locations

The VALCRI system prototype, comprising the Analyst Workstation has been deployed in all three police end-user environments so they can learn to use the software in their own time. They will initially use the VALCRI-developed crime dataset, and will migrate to using larger samples of old but real data when appropriate security procedures have been established. This will help them determine what ways VALCRI assists or hinders the criminal intelligence analysis process. For security reasons, the VALCRI prototype will not be connected to any live police systems. The VALCRI system prototype has also been deployed to all other VALCRI partners to enable local familiarisation, and to enable partners to use it for carrying out experiments and studies.

2.1.4.10 Analyst Training Courses

Partners involved in commercial training for intelligence analysis have developed multiple courses. Eight workshops have been run for police analysts. Additionally, a Masters-level (Level-7) Advanced Analyst training qualification has been developed in conjunction with VALCRI police partner, West Midlands Police, and Aston University, Birmingham. The course will include subjects in criminal behaviour, criminal networks, crime linking, crime and criminal profiling, from a critical thinking perspective in the context of data science.

2.1.4.11 Other Dissemination and Engagement activities

The VALCRI system has been evaluated with 214 law enforcement officers in 50 agencies in 16 countries excluding 2 international law enforcement agencies. A subset of these law enforcement agencies participated in 8 events organised by WP13 to evaluate the VALCRI training curriculum developed with 123 law enforcement officers in 40 organisations in 13 countries. In addition to this, VALCRI has been demonstrated at 5 academic international conferences, 1 EU project event and 5 intelligence events in 6 countries, demonstrating VALCRI system to an estimated 1500 persons.

By M48 a total of 119 scientific peer-reviewed publications were written, of which 64 were conference publications, 22 journals articles, 6 book chapters, 8 workshop papers and 19 whitepapers.

2.2 WORK PROGRESS AND ACHIEVEMENTS DURING THE PERIOD M17 – M32

2.2.1 Work Package 1 - MANAGEMENT

Work Package Summary

WP 1	MANAGEMENT
WP Leader	William Wong
Participants	MU, SPACE, AES, XI
Status	COMPLETED
Summary of key outputs	Vision and objectives set for the project Effective communication systems has been reviewed to encourage Work Packages and Sub Groups to maintain regular contact and work across disciplines Regular meetings with members of the External Advisory Board and Independent Ethics Board held Milestones achieved as planned Strong emphasis given to the human issues of the project (security, privacy, ethics, law) Minimised regular Consortium meetings and encouraged smaller Sub Group working meetings
Summary of significant contributions / impact for the project	The key management systems have been instrumental in steering this complex project successfully. The Partners have established good working relationships. The Consortium meetings have allowed innovative ideas from the researchers on the project to be adopted. The inclusion of Technical, User, Interface and Design managers in the Project Management Team has strengthened the scientific and technical management of the project. Regular contact with members of the External Advisory Board and Independent Ethics Board has provided invaluable advice.

Task 1.1	Project Management
Task Leader	MU
Participants	MU, SPACE, AES, XI
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D1.1 ManSystemsReport final (M06) ▪ D1.1 ManSystemsReport v2 Final (M16)
Summary of progress and overall	Vision statement circulated to all partners

contribution to the project	12 consortium meetings and over 60 face to face WP, Sub Group meetings held during the life of the project. Many of these attended by the Project Coordinator and/or Deputy to discuss the key mission and objectives of the project with all partners Regular telephone discussions with Work Package Leaders and Sub Group Leaders Management Team discussed work progress and management of project risks and planning Action plan and work progress discussed at each Consortium meeting Management systems and procedures established for: communication within the Consortium and with the Independent Ethics Board members and the External Advisory Board members; monitoring and reporting progress; quality reviewing project Deliverables; risk management; standards and quality criteria for the Milestones and Deliverables A public website has been established (www.valcri.org), plus two internal websites for Partners to share confidential information, meeting minutes and project news Management Deliverables prepared A technical software repository has been established for version control and software development management A secure communication channel has been deployed for the technical development team A system architecture has been developed to ensure that system components fit together. This has also been enhanced by working in sub groups rather than Work Packages to assist cross cutting cooperation Outcomes from research activities in the project are presented at Consortium meetings so that they can be incorporated into the technical programme of the project if appropriate. Partner XI left the project at end M32, and remaining budget and work was reallocated to MU but in subsequent months XI is disputing the overpayment arising from earlier pre-finance payments.

Task 1.2	Administration and Finance
Task Leader	MU
Participants	
Status	IN PROGRESS
Output and deliverables	Minutes of the start-up meeting
Summary of progress and overall contribution to the project	Administrative support provided to all partners with regular communication and prompt replies to partner questions and concerns

	Cost monitoring of project spend put in place and budget needs assessed – quarterly reporting to the Project Coordinator Financial guidelines disseminated to partner finance officers White papers has been submitted as deliverables (MTR2 10; Final review 7 new) Information tables on the Participant's Portal completed up to date

Issues and Deviations from Planning:

Task	Timing	Status	Deviation / issues / Comments
T1.1	M01-M48	COMPLETED	No main deviation, apart from XI leaving the project in M32 and remaining PM, work relocated to MU.
T1.2	M01-M48	IN PROGRESS	No major deviation, apart from XI leaving the project in M32. Remaining budget and work was reallocated to MU but in subsequent months XI is disputing the overpayment arising from earlier pre-finance payments.

Required Corrective Action

None.

Relevant Publications

Authors	Titles	Year	Conference / Journal	Status

2.2.2 Work Package 2 - REQUIREMENTS AND DESIGN

Work Package Summary

WP 2	REQUIREMENTS AND DESIGN
WP Leader	SPACE
Participants	MU, UKON, LIU, CITY, AES, FHG, OS, INT, XI, WMP, LPA, BFP
Status	COMPLETED
Summary of key outputs	**Deliverables** ▪ D2.2 – Requirement analysis document.pdf (M06) ▪ D2.3_FINAL.pdf (M06) ▪ VALCRI D2.3 System Design v2 Final.pdf (M17) **White Paper** ▪ VALCRI-WP-2017-001 Architecture.pdf **The Semantic media wiki**

	▪ The SMW including component information
Summary of significant contributions / impact for the project	1. Methodology for requirements and system design was prepared. 2. Learning the application domain in detail – first bootstrapping this by preparing quick domain snapshots, then organizing meetings with the end users. 3. Scoping the R&D in VALCRI – through discussions with the end users it was agreed that the project initial scope will be around CCA (Comparative Crime Analysis). 4. Detailing user processes and collection of requirements as epics and user stories. 5. Collecting requirements originating from WP3 activities (Human Issues) and from software quality attributes (security, audit, performance) 6. Organizing the invention workshops and collecting user experience architecture ideas. 7. Technical requirements were detailed. 8. Integration approaches were discussed and agreed. 9. Development and testing environment was designed. 10. Hardware choices were considered. 11. Initial discussion over datasets and ontology development was done (this work continues in the respective WP). 12. Technology enablers were listed. 13. The SMW (the semantic wiki) was set up – this is a collection of key information related to the relevant vision items, requirements, and components of the project (https://docs.valcri.org/smw/index.php/). 14. Overall system design was discussed, and its detailing is going on (using the SMW as a facilitator for that). All these results impact the project by guiding the next steps of the research and development.
Further information	Due to the decision of taking a slightly different approach from the approach originally planned in the DOW, it was a bit less clear how to divide the activities in the first two tasks – this did not have any negative impact beyond the fact that it is difficult to report over them separately in this report. The work on D2.3 was on going some partners including the WP leaders had consumed all the available effort of WP2. To mitigate this, it was suggested to move effort from other WPs.

Task 2.1 Task 2.2	Requirements Analysis Use Cases
Task Leader	SPACE
Participants	MU, UKON, LIU, CITY, AES, FHG, OS, INT, XI, WMP, LPA, BFP
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D2.2 – Requirement analysis document.pdf (M06)

Summary of progress and overall contribution to the project	As mentioned above, due to the organization of the work the border between T2.1 and T2.2 became vague, and therefore they are reported together. The requirement analysis process started with learning the application domain by organizing meetings with the users. In the first weeks of the project, snapshots descriptions of the application domain were prepared and made available to the project partners in order to bootstrap the activities of requirement analysis and design. Later, a more detailed description of the application domain was prepared, encompassing the intelligence model and role, goals of the analysts, techniques used, crime theories and anecdotes. Then, requirements were started to be harvested as epics and user stories, and VALCRI personas were detailed. While the process of collecting the requirements was an eye opener process to many of the partners, it was soon understood that the scope was lost in that process (too many tress to see the forest). As a result pivotal meeting with the end users was organized to define such scope and the result of this meeting was that the scope of the project (at least in its first stages) is the comparison crime analysis (CCA) process. In parallel, invention workshops were organized where the researchers and the users sat together discussing user experience architecture ideas. This process contributed to tasks T2.1 and T2.2 as it allowed to setup shared goals and understandings and to learn better about the problems of the users and their approaches to solve them. It should be noted that these workshops also contributed to T2.3.

Task 2.3	System Design
Task Leader	SPACE FHG (in support)
Participants	MU, UKON, LIU, CITY, AES, FHG, OS, INT, XI
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D2.3_FINAL.pdf (M06) ▪ VALCRI D2.3 System Design v2 Final.pdf (M17) **The Semantic media wiki** ▪ The SMW including component information
Summary of progress and overall contribution to the project	During M1 till M16, a methodology for capturing and documenting the requirements and system design was set up. This methodology provides a natural and practical continuation for the system design document completed in D2.3. In the period after that, during M17 and M39, this methodology was implemented and maintained by means of a Semantic Media Wiki (SMW). It collects the key information related to the relevant vision items of the project, requirements and functional components in the VALCRI software system. The SMW is meant to support the following goals:

	▪ Inform about the relevant project requirements, components and tech enablers. ▪ Connect requirements and technical domains/components ▪ Provide important information for system design (WP2) ▪ Provide important information for component integration (WP12) ▪ Communicate comments and guidelines related to user-stories and components (WP3) ▪ Connect ontologies and their requirements to the overall system requirements (WP10) Beyond updating the SMW, OpenProject and Gitlab's issue management features are in active use in order to manage short-term (or somewhat technical) activities. Furthermore, the way the VALCRI system works from a technical perspective now matured into a stable design and it is now implemented through several working system components. Most of the details of the design were captured into the white paper "Architecture, Development and Testing Environment" for a Visual Analytics-based Criminal Intelligence Analysis System". The work on finding appropriate hardware for running the VALCRI system (aka. The analysts workstation) is now finalised. Our police end users have received a complete hardware kit, as well as the most recent versions of the VALCRI software. This allows our partners to create and work with an isolated, standalone data set in order to evaluate the system. The work in this task started by forming a methodology for the requirements and system design. This methodology was later used as a guideline for designing the SMW as a tool to facilitate its implementation. Work on the technical requirements was done and the detailed results are discussed in the first revision of Deliverable D2.3. This actually bootstrapped much of the work in the different work packages as it encompassed subjects such as data types, data storage, ontologies, provenance, data extraction, data analysis, visual analytics and visualization, user interfaces, security and privacy. The task also included the early discussions regarding hardware choices and technology enablers that the partners considered to use in the different developments were listed. Finally, the task included the design of the development and testing environment

Issues and Deviations from Planning:

Task	Timing	Status	Deviation / issues / Comments
T2.1 + T2.2	M01-M06	COMPLETED	Due to the organisation of the work and additional goals set up by the consortium in the start of the work, the boarder between T2.1 and T2.2 became vague. In addition, as reported above, some activities that were not originally listed in the DOW were added to this task. Beyond that, there were no deviations

			from the planed work. Information available in D2.2 deliverable submission.
T2.3	M01-M39	COMPLETED	The work in this task continued throughout to M39. However, it was possible to achieve substantial results already in the first 6 months and there are no deviations from the planned.

Required Corrective Action

It may be necessary to re-adjust the resources for this Work Package with time from WP12 to reflect the slightly changed Tasks.

Relevant Publications

Authors	Titles	Year	Conference / Journal	Status

2.2.3 Work Package 3 - HUMAN ISSUES

Work Package Summary

WP 3	HUMAN ISSUES
WP Leader	TUGraz
Participants	MU, UKON, KU Leuven, TU Wien, ULD, I-INTELLIGENCE, WMP, STAD ANTWERPEN
Status	COMPLETED
Summary of key outputs	**Deliverables** ▪ D3.4 HumanIssuesFramework v2 Final.pdf **White Papers** ▪ VALCRI-WP-2017-006 Psychology Factors.pdf **Guidelines Checklist** ▪ Feedback on/Review of Knowledge Harvester ▪ Feedback on/Review of CCA Explorer ▪ Feedback on/Review of Linked Space Time Several internal reports/technical notes (see individual tasks)
Summary of significant contributions / impact for the project	The main aim of WP3 is to elaborate the Human Issues Framework (HIF) that brings together research findings from the individual tasks of WP3 in order to inform VALCI's design and specification process. For that purpose, we concentrated on the following: (i) to get a deeper understanding of how analytical thinking and sense-making processes occur in the context of criminal intelligence analysis, how they lead to improvements in insight and

	imagination and how these processes are influenced by cognitive biases; and ii) to refine M16 LEP Guidelines in WP3 to adapt these to the reform of the European data protection framework in the LEA sector (Directive (EC) 2016/680). Outcomes of WP3 work served as input not only for the technical people in terms of informing VALCRI's design but also for WP13 in terms of developing the VALCRI Training Syllabus. Work in M17-M48 of VALCRI built upon and continued the achievements of the first period and led to significant progress on all work package objectives. To deepen our understanding of analyst's work and its underlying processes (i.e. analytical thinking, sense-making and cognitive processes), **meetings with end-users** were held. Based on analyses of interviews and findings from cognitive task analyses conducted with end-users a **conceptual framework for evidential structuring and reasoning** (T3.1) has been elaborated. In addition to that, these analyses served as the basis for the **bias detection framework** elaborated in the context of T3.3. A range of **empirical studies** have been carried out on different VALCRI components and features. Studies have been realised in close cooperation between WP3 and the development work packages, to ensure that i) research interests of both groups were appropriately aligned and covered; and ii) results were fed back to the component development and refinement. In addition to that, results of these studies served as input for the **refinement of** the **design guidelines**. Cooperation with development partners was also expressed through giving **formal feedback** on single prototypes using our guidelines checklists. In Phase 2 of the project, a great emphasis was given on the **operationalisation** of the four design principles: imagination, insight, transparency, and fluidity/rigour. Results of this process (i.e. methods of measurement, instruments) will be used by WP13 to elaborate an evaluation methodology and thus serve as common ground for evaluating the VALCRI system.
Further information	n/a

Task 3.1	Evidential Structuring and Reasoning
Task Leader	MU
Participants	
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D3.4 HumanIssuesFramework v2 Final.pdf **White Papers** ▪ VALCRI-WP-2017-006 Psychology Factors.pdf **Conference papers** Passmore, P. J., Attfield, S., Kodagoda, N., Groenewald, C., & Wong, B. W. (2015). Supporting the externalisation of thinking in criminal intelligence analysis. In

	Intelligence and Security Informatics Conference (EISIC), 2015 European (pp. 16-23). IEEE Selvaraj, N., Attfield, S., Passmore, P., & Wong, B. W. (2016). How Analysts Think: Think-steps as a Tool for Structuring Sensemaking in Criminal Intelligence Analysis. In Intelligence and Security Informatics Conference (EISIC), 2016 European (pp. 68-75). IEEE. Groenewald, C., Wong, B.L., Attfield, S., Passmore. P. and Kodagoda, N., (2017a). How Analysts Think: Managing Uncertainty - How Do Criminal Intelligence Analysts Recognise and Manage Significant Information?. European Intelligence and Security Informatics Conference (EISIC), Athens, Greece, 2017, pp. 47-53. Groenewald, C., Wong, B.L.W., Passmore, P., and Kodagoda, N., (2017b). How Analysts Think: Navigating Uncertainty – Aspirations, Considerations and Strategies. In: 13th International Conference on Naturalistic Decision Making, 20th-23rd June, 2017, Bath, Somerset, UK. Groenewald, C., Wong, B.L., Attfield, S., Passmore. P. and Kodagoda, N., (2017c). How can we Design Tactile Interactive software for Argument Construction in Criminal Intelligence Analysis?. In: 13th International Conference on Naturalistic Decision Making, 20th-23rd June, 2017, Bath, Somerset, UK. Book Chapters and Journal Papers Groenewald, C., Attfield, S., Passmore, P., Wong, B.L.W., and Kodagoda, N. (2017d). A Practical, Hybrid Argumentation Model to Assist With the Formulation of Defensible Assessments in Uncertain Sense-Making Environments. In: Leventakis & Haberfield (eds) Community-Oreinted Policing and Technological Innovations, SpringerBriefs in Policing. Groenewald, C., Attfield, S., Passmore, P., Wong., Qazi, N., Kodagoda, N., (forthcoming). A Descriptive, Practical, Hybrid Argumentation Model to Assist With the Formulation of Defensible Assessments in Uncertain Sense-Making Environments - An Initial Evaluation. Cognition Technology and Work, Special Issue on Naturalistic Decision Making.
Summary of progress and overall contribution to the project	**Development of a Hybrid Argumentation Model for Police Analysts** We have developed an argumentation language specifically for police analysts to externalise their thinking specifically for use in the context of: ▪ The use of data visualisations; ▪ Reasoning from early uncertainty (fluidity) to later stages of increased certainty (rigour); ▪ The socio-technical context within which police analysts work. We began with a review of literature related to evidential structuring and reasoning. The review concluded that a potentially valuable approach would extend upon hybrid argumentation approaches such as Anchored Narrative Theory, which combines argumentation and narrative into a single schema, to include other relations such as thematic grouping and tentative lines of enquiry (Passmore et al., 2015). To further inform the approach, we conducted an in-vivo study of police analyst's sensemaking and reported on 'think-steps' (extensible crime templates) as a primary structuring concept and other considerations arising from the socio-technical work context, such as for example, how requests are communicated and resolved. (Selvaraj et al., 2016). Groenewald (2017a) explored how police analysts recognise and manage significant information, and Groenewald (2017b) looked at how police analysts navigate uncertainty moving from conditions of fluidity (data loosely

	assembled, with high uncertainty and low commitment), to rigour (high certainty and commitment). Together, these studies have informed a hybrid argumentation language presented in the form of a prototype and evaluated in Groenewald et al. (2017c), Groenewald et al. (2017d), and Groenewald et al., (forthcoming).

Task 3.2	Advances in sense-making and insight
Task Leader	TU Wien
Participants	
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D3.4 HumanIssuesFramework v2 Final.pdf **White Papers** ▪ VALCRI-WP-2017-006 Psychology Factors.pdf **Others** ▪ Study on Sense making with a prototype of the VALCRI system (with students and with end-users) ▪ Study on sense making with maps (in cooperation with City University) ▪ Theoretical investigation of sense making strategies ▪ Publications at scientific conferences
Summary of progress and overall contribution to the project	In the last phase of the project we conducted several user studies to identify sense making strategies important for the interaction with the VALCRI system. We especially conducted one user-study with a prototype of the VALCRI system. This investigation was, on the one hand, conducted with students to identify sense making strategies. On the other hand, we also conducted user tests with end users to be able to assess the usefulness of the system. We could identify several usability issues. These results were communicated to the developers, so that they could improve the system accordingly. End users also provided feedback on the design of the system. We also could identify cognitive sense making strategies and how these strategies are supported by the system. These results can help to clarify what makes the VALCRI system superior to other systems. We also discussed theoretical implications of these issues in a separate paper. Another study investigated the differences in sense making in the use of line-up tests, which are based on the police practise for eye witness identification by lining up several suspects next to each other. One question in this regard is which criteria are used to identify the real plot between several decoys. We noticed that the sense making strategies of study participants are similar to those used in other contexts. This is an indication that users adopt similar sense making strategies when working with visualisations, but adapt these strategies to the context in a flexible manner

Task 3.3	Cognitive Bias Mitigation
Task Leader	TUGraz
Participants	TUGraz, UKON
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D3.4 HumanIssuesFramework v2 Final.pdf **White Papers** ▪ VALCRI-WP-2017-006 Psychology Factors.pdf **Internal Working Documents** ▪ Analysis of the VALCRI System with regards to Cognitive Biases ▪ Operationalisation of Cognitive Biases **Publications at scientific conferences** Bedek, M. A., Nussbaumer, A., Hillemann, E. C., & Albert, D. (2017). **A framework for measuring imagination in visual analytics systems**. In J. Brynielsson (Ed.), Proceedings of the European Intelligence and Security Informatics Conference (EISIC 2017), 11-13 September 2017, Athens, Greece. (pp. 151-154). doi:10.1109/EISIC.2017.31 Bedek, M. A., Nussbaumer, A., Huszar, L., & Albert, Dietrich (2017). **Discovering Cognitive Biases in a Visual Analytics Environment**. In Proceedings of the DECISIVe Workshop 2017 at the Vis Conference. Bedek, M. A., Nussbaumer, A., Huszar, L., & Albert, Dietrich (2018, in press). **Methods for Discovering Cognitive Biases in a Visual Analytics Environment**. Springer **Workshops** ▪ **Dealing with Cognitive Biases in Visualisations** (DECISIVe 2017). A workshop held at the Vis Conference 2017
Summary of progress and overall contribution to the project	The primary goal of T3.3 is to deal with cognitive biases in the field of criminal intelligence analysis. This involves the identification of relevant cognitive biases in this field, developing strategies for their detection and mitigation, and the integration of these findings with the VALCRI system. More concretely, work in the last 15 months of the project concentrated on building upon achievements of the last project period and can be divided into three main interrelated working strands: i) empirical research on cognitive biases in visual analytics environments, and ii) elaboration of a cognitive bias discovering framework, and iii) dissemination of the research results and communication with the scientific community. A general overview of this work is given in the White Paper entitled “Cognitive Bias Mitigation in Criminal Intelligence Analysis: Methods, Experiences, and Results from the VALCRI project”.

	Empirical research. The main aim of research done was to refine, to adapt and to validate the VALCRI system as well as the design guidelines that inform the VALCRI system. For that purpose, several empirical studies investigating the effect of and relations between visualisation and cognitive biases. One study investigated the effect of information visualisation techniques on the confirmation bias. A second study investigated the resilience of the VALCRI system against the clustering illusion. Moreover, a serious of studies investigated the effect of visualisations on the anchoring, adjustment, and groupthink cognitive biases. Elaboration of a framework for discovering and mitigation of cognitive biases. This framework consists of three clusters of approaches to measure and detect cognitive biases in the context of a Visual Analytics Environment (VAE). The first approach is called theory-driven since it refers to purely expert-driven methods. One of these methods is based on the definition of design recommendations for interactive data visualisations to reduce the potential effects of cognitive biases. Another expert-driven method is the in-depth analysis of a VAE and the cognitive biases it may induce. The second, empirical approach encompasses behavioural observations, as well as experimental methods to operationalise cognitive biases. We describe methods to measure the Confirmation Bias and the Clustering Illusion. Finally, the third approach refers to data-driven methods that enable a non-invasive measurement of cognitive biases. This method is based on data-mining methods to analyse and interpret user interaction patterns in terms of cognitive biases. Dissemination. The DECISIVe workshop was organised by WP3 and ran at IEEE Vis2017 in Phoenix, Arizona, USA on 2nd October 2017. The workshop, entitled "Dealing with Cognitive Biases in Visualisations", was aimed at participants from a wide range of disciplines, such as information visualisation, visual analytics, software engineering, cognitive psychology and decision science, to explore ways in which cognitive biases have a detrimental impact on users' decision making when using visualisation and analytics tools, and practical ways to reduce their potentially harmful effects. The workshop call attracted and good response and 14 high quality papers were selected to be presented at the workshop, including one from TUGraz team. Interest from the visualisation community in cognitive biases was demonstrated by the good attendance at the presentations and at the practical sessions later in the day, where groups were tasked to invent novel strategies for particular biases. As a result of the successful workshop, we were approached by Springer with regard to producing a book, specifically on cognitive biases in visualisations.

Task 3.4	Legal, Ethical and Privacy Aspects
Task Leader	KUL
Participants	MU, ULD and KUL
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ VALCRI-D3.4-Final.pdf (M06) ▪ D3.4 HumanIssuesFramework v2 Final.pdf (M16)

	Technical Notes ▪ TN3.4.E1_20150826_TechnicalReport_Ethical_MU.doc ▪ TN3.4.L1_20150826_TechnicalReport_Legal_KUL.doc ▪ TN3.4.P1_20150826_TechnicalReport_PrivacyULD.doc **White Papers** ▪ VALCRI-WP-2017-005 Transparency.pdf ▪ VALCRI-WP-2017-012 Roadmap for the Operationalization.pdf ▪ VALCRI-WP-2017-014 Understanding the ethical dilemmas.pdf Finalization of the M16 LEP Guidelines in WP3 to adapt these to the reform of the European data protection framework in the LEA sector (Directive (EC) 2016/680) (2016) Finalization of internal reports from the LEP partners together (KUL, ULD, MU), partially with some tech partners from SEPL and DMO subgroup (FHG, OS, LIU, MU) (2016): ▪ The reform of the European data protection framework and its impact on personal data processing in VALCRI (ULD, KUL, MU) – to assess the relevance of the draft police Directive (in the context of the reform of the European data protection framework in the LEA sector) ▪ Report on Provenance from LEP perspective in VALCRI (ULD, KUL, LIU, MU) ▪ Report on Logging in VALCRI (ULD, OS) ▪ Report on personal data, anonymization and pseudonymization in VALCRI (KUL, ULD, MU, FHG) ▪ Report on international data transfers (Data Sharing Report): applicable EU legal framework for the exchange of intelligence in the EU and list of technical requirements for VALCRI (KUL, ULD) Finalization of the Data Management Policy adapted to Belgian legal framework for the tests of VALCRI prototype (KUL) (2017). Finalization of LEP White Papers (2017): ▪ The Operationalization of Transparency in VALCRI (ULD, MU, KUL) ▪ Roadmap for the Operationalization of Legal and Privacy Requirements in VALCRI Analysis (KUL) ▪ Roadmap for the Resolution of Ethical and Human Rights Issues in Automated Data Analysis and Extraction Computations in VALCRI (KUL, MU) **LEGAL aspects**

	▪ Analysis of the State of the Art: UK and Belgium (2014) (KUL) ▪ Research for various reports included in the Deliverables on topics including national legislation, police ethics, the implementation of the principle of purpose limitation. ▪ End-user workshop on data protection legislation and its requirements for law enforcement system content and functionality at UK Partners in West Midlands Police (2015) ▪ Copyright aspects of the mutualisation of criminal information (KUL) ▪ Criminal procedure and transparency: disclosure obligations (MU, KUL): contribution integrated to the Provenance report and the White paper on the Operationalization of Transparency in VALCRI ▪ Further investigation of the impact of using intelligent computer systems in criminal investigations in the UK in the context of disclosure obligations. This includes the identification of any important issues related to managing relevant material (for disclosure purposes) in these systems and also the possibility of both prosecution and defence counsel having access to these systems as alluded to in The Disclosure Manual of the Crown Prosecution Service at Chapter 5, Section 5.22, which addresses the issue of information recorded on a computer. (MU) ▪ Initial investigation of whether the new Investigators Powers Bill in the UK will have any impact on VALCRI or needs to be addressed in our outputs (MU) **PRIVACY aspects** Besides the work on the above mentioned LEP output, the SEPL group finalized and delivered: ▪ Comparative analysis of purpose limitation in Belgium, UK, Germany and EU (KUL, MU) (2015) ▪ Publication of paper on the benefits and pitfalls of predictive policing through the EISIC Conference (ULD, FHG, MU) (2015) ▪ Conceptual work on mapping the new Police Directive 2016/680 to the LEP guidelines and the data protection goals to develop an evaluation methodology for the VALCRI prototype, including a table comparing data protection requirements with concrete conditions (2016 - 2017) (ULD, KUL) ▪ Research on the legal framework for sharing information between LEAs and intelligence institutions in Germany as contribution to the Data Sharing Report drafted by KUL (chapter 1, section for Germany) ▪ Presentation of the paper "Benefits and Pitfalls of Predictive Policing" at the European Intelligence and Security Informatics Conference (EISIC) in Manchester, UK in September 2015 ▪ Organization of a workshop on Directive (EU) 2016/680, entitled “The Directive for data protection in the police and justice sectors: towards better data protection?”, 1 February 2016, Leuven. Presentation on the topic of “Limiting collection and uses of data by police: wishful thinking under an intelligence-led policing paradigm?” (KUL) ▪ Organization of a panel discussion on Algorithmic Transparency and Accountability in Law Enforcement at the Computer Privacy and Data Protection Conference (CPDP) (2018) (KUL, ULD, MU)

	▪ Publication of an article on the Police Data Protection Directive (EU) 2016/680 and its implications in the Computer Law & Security Review (2017) (KUL) ▪ Publication of an article on the Europol Regulation and Purpose Limitation in the European Data Protection Law Review (2017) (KUL) ▪ Publication of an article on the reformed legal regime for biometric data in the Computer Law & Security Review (2017) (KUL) ▪ Publication of an article on algorithmic transparency and accountability in criminal intelligence profiling in the IEEE EIS Conference Proceedings (2017) (KUL) ▪ Presentation of an article on coding non-discrimination by design in Police Technology at the Next-Generation Community Policing Conference, to be published (2017) (KUL) ▪ Presentation and publication of a paper on addressing ethical challenges of creating new technology for criminal investigations at the Next-Generation Community Policing Conference (2018) (MU) As a conclusion to the work done under T3.4, the SEPL group has conducted an international comparison and analysis of different national guidelines and approaches to data protection impact assessments (DPIA) which will soon be mandatory for law enforcement agencies utilizing new technologies that carry high risks to the rights and freedoms of natural persons. This research has culminated in the drafting of a new methodology to be used by law enforcement agencies and, in particular, VALCRI end-users when conducting a DPIA. This work consists of two reports that can be disseminated separately or as a joint document: ▪ Data Protection Impact Assessments (DPIAs) in the law enforcement sector according to Directive (EU) 2016/680 - A comparative analysis of methodologies (ULD, KUL) (2018) ▪ Data Protection Impact Assessments (DPIAs) in the law enforcement sector according to Directive (EU) 2016/680 – A comprehensive methodology for Law Enforcement DPIAs (ULD, KUL) (2018) (draft currently under revision to get finalised and published within the dissemination WP) **ETHICAL aspects** In addition to contributions to LEP work (on transparency, accountability and provenance as noted above): ▪ Contribution to misuse case development (Malta meeting) ▪ Investigating dimensions of algorithms from an ethical perspective ▪ Development of scenarios for ethical issues identified by IEB (incidental findings, accidental discrimination) ▪ Presentation of an Ethics Tutorial Workshop at European Intelligence and Security Informatics Conference (EISIC) 2017, Athens, September 11-13th, 2017 (MU) Investigating status of IEB issues of concern with MU WP1 (through a number of meetings over the period), and more specifically members of relevant WP's and Sub Group's responsible for different aspects of the VALCRI system (Warsaw October 2016 meeting). Liaising with the IEB to discuss ethical concerns and relay them to the rest of the consortium happened repeatedly and in particular at all VALCRI consortium and group meetings.

Summary of progress and overall contribution to the project	Identification of relevant jurisdictions and legal framework in the areas of privacy and data protection, functioning of the police services, and human rights. Identification and concretization of legal requirements, especially with regard to the reform of the European data protection framework ▪ General relevance of Directive (EU) 2016/680 for VALCRI ▪ Key principles and requirements ▪ Work towards specification of selected and VALCRI-relevant technology approaches for the purpose of relaying legal requirements to the technical partners and assisting the development of the system in a manner mindful of evolving ethical and legal standards. – Logging – Anonymization & pseudonymization – Provenance – Data management and transfers – Accountability – Transparency – Security – General legal compliance ▪ Matching legal developments to LEP guidelines provided in the project ▪ Assessing the relevance of the Council Framework Decision 2006/960/JHA of 18 December 2006 on simplifying the exchange of information and intelligence between law enforcement authorities of the Member States of the European Union for VALCRI: key principles and requirements (KUL) ▪ Evaluating several key ethical and human rights issues such as non-discrimination and the fairness of the trial related to automated data analysis in systems like VALCRI Drafting of full WP3 LEP Guidelines transforming legal requirements into concrete development guidelines and conditions for the VALCRI consortium. Review of further VALCRI prototype components at numerous meetings to provide feedback to the developing partners. ▪ Knowledge Harvester ▪ CCA Explorer (later named HiBrowse) ▪ LinkedSpaceTime ▪ Architecture updates ▪ GrayLog ▪ OpenPMF ▪ SEPL Modules ▪ Data Ingester ▪ Structured and unstructured databases Review of Human Issues Framework drafted by partner I-INT, other Deliverables such as D2.2, and user stories drafted in WP2. Collaborating with technical partners and disseminating knowledge at VALCRI SEPL group meetings, consortium meetings, workshops and other conferences, including Innovation Konstanz 2014, Invention Berlin 2014, Invention Leuven 2014, Designer London 2015, London 2015, Malta January 2016, Berlin SEPL Feb 2017, Berlin May 201, EISIC September 2017, Nice November 2017 and

	NGCP October 2017. This coincides with the discussion of demo components being presented for the purpose of guiding software development and system features with a view on legal and ethical compliance. Concretization of the VALCRI vision item transparency, accountability and general privacy-oriented compliance from LEP perspective as preparatory research work for the relevant data management policies, LEP papers and DPIA guidelines. Concretization of technical requirements for data, process and reasoning provenance (KUL, ULD together with LIU, MU), resulting in further development of provenance ontology and data protection requirements Further work on the access control enforcement concept in VALCRI, together with tech partners and end users (WMP) ▪ Understandings and definitions ▪ Reviewing potential issues and phrasing questions to end users for further refinement ▪ Contributing to the finalization of the SEPL enforcement module in system architecture Draft of a classification of different categories of risk/misuse scenarios plus a table template for adequate description and determination of mitigation techniques. Draft of SEPL related example on user story and correlating specified misuse/risk case based on based on the Guideline G3.4.PL3 (Auditing and Logging), to be fed into the Semantic Media Wiki (SMW). Assessing the newly introduced legal requirement of data protection impact assessment and its relevance for the use of the VALCRI project for the purpose of the dissemination of a general DPIA methodology emphasizing law enforcement work.

Task 3.5	Human Issues Framework
Task Leader	IINT
Participants	TU Graz, MU, WMP, TU Wien, KUL, ULD, UKON, LPA , BFP, WMP
Status	COMPLETED
Output and deliverables	**Technical Notes** ▪ WP13_TN13.3_Training_Syllabus_V1.0.doc ▪ WP13_TN13.5_The First and Second VALCRI Evaluation.doc Our work human issues informed the following outputs and deliverables: ▪ The VALCRI Syllabus developed under T13.3 ▪ The end user workshops held under T13.4 ▪ The evaluation methodology prepared under T13.5 Further, the HIF informed the development of a novel impact model to help analysts address the principal challenges affecting intelligence professionals. This model was presented in our paper, “Redefining the Intelligence Skill Set

	Through the Prism of the Intelligence Analysis Impact Model", presented to the Intelligence Studies Group of the International Studies Association (ISA).
Summary of progress and overall contribution to the project	Our work on the Human Issues Framework (HIF) during M16-M32 built upon research conducted during the preceding period. The framework serves two functions ▪ First, it is used to evaluate the research outputs and guidelines generated by WP3. ▪ Second, the framework's outputs have been used to prepare the various iterations of the VALCRI Training Syllabus, as described in Technical Note TN13.3. This syllabus covers all five domains of the Framework – Organisational, Operational, Informational, Technological and Cognitive – and has been used to develop and deliver the VALCRI Training Workshops. Our work on the syllabus, together with our interaction with End Users, has helped refine our understanding of the issues identified by the HIF. By way of example, we are better aware of how operational issues such as task and workflow management impact the production of intelligence products. Our intention now is to extend the framework to address these issues and to find suitable tools to help law enforcement professionals better address the challenges they confront.

Issues and Deviations from Planning:

Task	Timing	Status	Deviation / issues / Comments
T3.1	M01-M47	COMPLETED	No deviation
T3.2	M01-M47	COMPLETED	No deviation
T3.3	M01-M47	COMPLETED	No deviation
T3.4	M01-M47	COMPLETED	A deviation has been made from the original version of the DOW regarding the creation of: IR3.4.1 Report on requirements in terms of legal, privacy and ethical aspects of technologies under review, frameworks and methods This report has been broken down into a set of smaller reports to reflect more focused on the specific technology areas in VALCRI, as well as to address the respective developing partners responsible for these areas individually. These reports are listed in the Task 3.4 table in the row addressing the summary of progress and overall contribution to the project.
T3.5	M01-M47	COMPLETED	No deviation

Required Corrective Action

None.

Relevant Publications

Authors	Title	Year	Conference / Journal	Status
T. Marquenie & E. Zouave	Speaking Truth to Computational Power: Coding Non-discrimination by Design in Police Technology".	2018	Next-Generation Community Policing Conference	To be published
Penny Duquenoy, Don Gotterbarn, Kai. K. Kimppa, Norberto Patrignani, B.L. William Wong in	Addressing Ethical Challenges of Creating New Technology for Criminal Investigation: The VALCRI Project in George Leventakis and M.R. Haberfeld (Eds),	2018	Societal Implications of Community-Oriented Policing Technology - SpringerBriefs in Policing.	
E. Zouave & T. Marquenie	An Inconvenient Truth: Algorithmic Transparency & Accountability in Criminal Intelligence Profiling	2017	European Intelligence and Security Informatics Conference	published
E. Kindt	Having yes, Using no ? About the new legal regime for biometric data'	2017	Computer Law & Security Review	published
T. Marquenie	The Police and Criminal Justice Authorities Directive: Data Protection Standards and Impact on the Legal Framework"	2017	Computer Law & Security Review	published
F. Coudert	The Europol Regulation and Purpose Limitation: From the 'Silo-Based Approach' to… What Exactly?	2017	European Data Protection Law Review	published
Haider, J., Seidler, P., Kodagoda N., Adderly, R., Pohl, M., and BL Wong	Matrix visualisation, node-link diagram, sensemaking strategies	2016	ACM Conference on Human Factors in Computing Systems (CHI)	submitted
Haider, J., Pohl, M.,	Sensemaking	2016	IEEE VIS VAST Doctoral Colloquium	Technical report
Seidler, P.; Haider, J.; Kodagoda, N.; Pohl, M.; and BL Wong, W	Design for Intelligence Analysis of Complex Systems: Evolution of Criminal Networks	2016	European Intelligence and Security Informatics Conference	published
Nussbaumer, A.; Verbert, K.; Hillemann, E.-C.; Bedek, M.A.; and Albert, D.	A Framework for Cognitive Bias Detection and Feedback in a Visual Analytics Environment	2016	European Intelligence and Security Informatics Conference	Published
Nallini Selvaraj, Simon Attfield, Peter Passmore, William Wong	How Analysts Think: Think-steps as a Tool for Structuring Sensemaking in Criminal Intelligence Analysis	2016	European Intelligence and Security Informatics Conference	Published
William Wong, Ifan D.H. Sheperd, Patrick Aichroth, Sebastian Mann, Rudolf Schreiner, Ulrich Lang, Eva Schlehahn	Benefits and Pitfalls of Predictive	2015	European Intelligence and Security Informatics Conference (EISIC): "Benefits and Pitfalls of Predictive Policing"	published
P. Malaquias	Predictive policing, copyright law, public security exception	2016	SSRN	unpublished
D. Sacha, H. Senaratne, B. C.	The Role of Uncertainty, Awareness, and Trust in Visual Analytics	2016	IEEE Transactions on Visualization and Computer Graphics	published

Kwon, G. Ellis and D. A. Keim				
D. Sacha, I. Boesecke, J. Fuchs and D. A. Keim	Analytic Behavior and Trust Building in Visual Analytics	2016	Eurographics Conference on Visualization (EuroVis)	published
Haider, J.; Pohl, M.; Hillemann, E.; Nussbaumer, A.; Attfield, S.; Passmore, P.; and Wong, B. L. W	Exploring the Challenges of Implementing Guidelines for the Design of Visual Analytics System	2015	Human Factors and Ergonomics Society 59th Annual Meeting	published
Ellis, G.; Dix, A.;	Decision Making, Uncertainty; Cognitive Biases	2015	IEEE VIS	Published

2.2.4 Work Package 4 - ANALYST USER INTERFACE

Work Package Summary

WP 4	ANALYST USER INTERFACE
WP Leader	Chris Rooney, MU
Participants	MU, UKON, CITY, INT, XI
Status	Completed
Summary of key outputs	**White Papers** ▪ VALCRI-WP-2017-002 AUI Thinking Landscape.pdf ▪ VALCRI-WP-2017-009 Provenance3.pdf ▪ VALCRI-WP-2017-010 DVQs.pdf (Dynamic Visual Querying) ▪ VALCRI-WP-2018-018 Representing Uncertainty.pdf **Development Work** ▪ Commercially viable UI Framework based on the Meteor and React libraries for creating a modular and extendable web-based user interface. ▪ Prototype UI Framework based on the Errai and GWT platforms. ▪ Creating prototypes of the Analyst's Workstation, a novel hardware setup for police intelligence analysts. ▪ Connection to Elastic search for providing rapid responses to data queries. ▪ Connection to the SDB for presenting ontological or graph data. **Deployment** ▪ Installed the Analyst Workstations at all three end-user locations. ▪ Installed an Analyst Workstation with the Metropolitan Police, London, and trailed VALCRI with real crime data. ▪ Installed an Analyst Workstation with the Pasco Sherriff's Office, Florida, and trailed VALCRI with real crime data.
Summary of significant contributions / impact for the project	The output of this Work Package is an integrated, commercially viable user-interface platform, which was built upon in Work Packages 5, 6, and 7. By making the platform modular and extendable, different visual components were able to be made by different partners (in different work packages) and still

	integrate in a single user interface. Thus allowing analysts to analyse data, gain understanding and construct hypotheses (see D4.5). As well as a technical platform, this work package also provided a set of interaction, visual, and information visualization guidelines for partners to follow when constructing these prototypes (see D4.5). This is accompanied by a set of proposed techniques for representing both provenance and uncertainty (see D4.6). Over the course of the project, we have. 1. Developed and tested prototype visualization components and published research on their application. 2. Developed a prototype of our integrated user interface to work with the anonymised data made available through Work Package 11. Developed a commercially viable version of the user interface to work with a multitude of real datasets, and that can be deployed operationally.

Task 4.1	User Interface Concept, Strategy and Architecture
Task Leader	MU
Participants	MU, XI
Status	COMPETED
Output and deliverables	**Deliverables** ▪ D4.5 SystemDesignUI Final (M16) **White Papers** ▪ VALCRI-WP-2017-002 AUI Thinking Landscape.pdf
Summary of progress and overall contribution to the project	Through a series of user interface design meetings, we have defined the concept of a 'thinking landscape', and an accompanying design strategy for developing a user interface around this concept. This includes novel approaches such as the use of *active layers*, the *X-Y display* and *interactions with transformable data objects*. Details of this work can be seen in TN 4.1.1.

Task 4.2	Design and Build UI
Task Leader	MU
Participants	MU, XI, CITY
Status	Completed
Output and deliverables	**Deliverables** ▪ D4.5 SystemDesignUI Final.pdf (M16) **White Papers** ▪ VALCRI-WP-2017-010 DVQs.pdf

Summary of progress and overall contribution to the project	After trialling several prototypes, we have developed the Analyst User Interface, a platform where visual components developed as part of Work Packages 5, 6 and 7 can integrate. The software is developed as a native web application and it built upon the Meter and React JavaScript libraries. We have finished the project with a commercially viable system (currently at TRL 6), which includes a complete data pipeline. We have implemented dynamic visual queries that allow for rapid search and querying of the data through a series of interactions with each of the visualisations. Through a thin-client approach, we are able to keep response times to a minimum, keeping in line with the expectations of a Visual Analytics system. Other features developed: 1. Persistent workspaces and bookmarking. 2. Applying complex clustering algorithms to result sets to identify similarities and view similar crimes in a comparative case analysis. 3. Multi-modal interaction through multiple devices such as tablets

Task 4.3	Uncertainty Representation
Task Leader	MU
Participants	MU, CITY, UKON, INT
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D4.6 UIDesignUncertaintyProvenance Final.pdf (M16) **Technical Notes** ▪ WP4_TN.4.3.1_Uncertainty Representation_i-intelligence_Final.doc **White Paper** ▪ VALCRI-WP-2018-018 Representing Uncertainty.pdf
Summary of progress and overall contribution to the project	Data analysis in police intelligence is inherently and uncertain task and, therefore, requires data visualisation methods to address this. We have conducted a literature review of uncertainty representation see Technical Note TN 4.3.1 and, based on this review, have found within the literature in Information Visualization and Cartography design approaches for representing and mitigating uncertainty see Deliverable D4.6.

Task 4.4	Analytic Provenance Representation
Task Leader	MU
Participants	MU, CITY, UKON, INT, XI
Status	COMPLETED

Output and deliverables	**Deliverables** ▪ D4.6 UIDesignUncertaintyProvenance Final.pdf (M16) **Technical Notes** ▪ WP4_TN.4.4.1_Provenance Representation_i-intelligence_Draft_v.2.0.doc **White Paper** ▪ VALCRI-WP-2017-009 Provenance3.pdf
Summary of progress and overall contribution to the project	Technical Note TN4.4.1 provides details on the various definitions of provenance and its relevance to crime analysis. Following on from this we have distinguish between two types of provenance: data provenance and analytic provenance, and suggest techniques for recording and representing both see Deliverable D4.6. From a development perspective, we have implemented workspace persistence. This means we maintain the full state of the workspace at any one time. By saving the changes between states, we can automatically track process provenance, and by saving a complete state, we can save a bookmark.

Task 4.5	Coordinate UI Design
Task Leader	XI
Participants	XI, MU
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D4.5 SystemDesignUI Final.pdf (M16) **Technical Notes** ▪ WP4_TN.4.1.1 UI Design Strategy v2.pdf ▪ WP4_TN.4.1.2 UI Design Touch Interaction.pdf ▪ WP4_TN.4.2.3_InfoVis_Design_Guidelines.pdf ▪ VALCRI TN 4 2 4.pdf
Summary of progress and overall contribution to the project	To maintain a consistent system across the development of independently developed modules, a set of interaction and visual design guidelines have been published by XI, along with a set of information visualisation guidelines provided by CITY. These are to be used by partners involved in Work Packages 4, 5, 6 & 7.

Issues and Deviations from Planning:

Task	Timing	Status	Deviation / issues / Comments
4.1	M01 – M48	COMPLETED	A deviation has been made from the original version of the DoW when XI departed the project in M32 their work was under taken by MU.

4.2	M01 – M48	COMPLETED	A deviation has been made from the original version of the DOW when XI departed the project in M32 their work was under taken by MU.
4.3	M01 – M48	COMPLETED	No deviation
4.4	M01 – M48	COMPLETED	A deviation has been made from the original version of the DOW when XI departed the project in M32 their work was under taken by MU.
4.5	M01 – M48	COMPLETED	A deviation has been made from the original version of the DOW when XI departed the project in M32 their work was under taken by MU.

Required Corrective Action

None.

Relevant Publications

Authors	Title	Year	Conference / Journal	Status
Anslow, C., Rooney, C., Kodagoda, N., & Wong, B. L. W.	Police Analyst Workstation: Towards a Multi-Surface User Interface	2015	Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces, ITS '15, Madeira, Portugal (pp. 307-311). New York, NY: ACM Press.	Published
Celeste Groenewald, Craig Anslow, Junayed Islam, Chris Rooney, and William Wong	Understanding Mid-Air Hand Gestures for Interactive Surfaces and Displays: A Systematic Literature Review.	2016	In Proceedings of the British Human Computer Interaction Conference (BritishHCI), Bournemouth, UK.	Published
Craig Anslow and B. L. William Wong.	Effects of the Display Angle and Physical Size on Large Touch Displays in the Work Place.	2017	In Proceedings of the ACM International Conference on Interactive Surfaces and Spaces (ISS), Brighton, UK,	Published

2.2.5 Work Package 5 - DATA SPACE

Work Package Summary

WP 5	DATA SPACE
WP Leader	UKON
Participants	UKON, CITY, FHG, XI
Status	COMPLETED
Summary of key outputs	▪ Built, evaluated, and adapted software prototypes on thematic visualization, text document visualization, and video/image analysis. ▪ Submitted significant findings to academic conferences. ▪ Integration of components into the VALCRI framework. ▪ Inclusion of security implementations.

Summary of significant contributions / impact for the project	The developed visualization components allow the analysts to explore and investigate crime reports from different perspectives. This covers crime report data about space, time, and the textual description, as well as semantic features that can be extracted from textual descriptions. Additionally, video and image analysis allows the analyst to obtain further details and to investigate unstructured information. From a research perspective, this challenging multivariate and high-dimensional dataset has led to the innovation of novel visual analytics methods, software prototypes, and impactful scientific publications. The software prototypes have been improved and modified based on domain expert feedback during the project period. A major effort has been made to integrate some of these components into the VALCRI framework and to adapt their appearance in the AUI. Inclusions of security features have also been implemented. The analytic results can be interactively adapted by the analyst in the visualizations and the components are ready to be tested in collaboration with other components. Additionally, a provenance gesture has been integrated to capture the current state of the system. All the visualizations will support the crime analysts in their investigations with the goal to improve efficiency and to allow the verification of gained insights from many different perspectives.
Further information	

Task 5.1	Thematic Visualization
Task Leader	CITY
Participants	CITY, MU
Status	COMPLETE
Output and deliverables	**Deliverables** ▪ D5.7 UIDesignDataSpace Final.pdf (M16) **White Papers** ▪ VALCRI-WP-2017-011 Applying Visual Interactive Dimensionality Reduction.pdf **Software**
Summary of progress and overall contribution to the project	We have developed thematic visualization methods that involve a new approach to assessing multiple perspectives concurrently. This kind of concurrent analysis is difficult when data are grouped and split into separate slides or a series of juxtaposed small multiples for comparison. Our design framework allows multiple perspectives to be viewed simultaneously in ways that neither clutter nor overload. It involves the superposition of perspectives, each of which may be represented at three levels of abstraction. By interactively varying the abstraction level, certain perspectives can be brought into, or out of, focus. Our experiments showed the approach to be no less successful than juxtaposed alternatives and open up a new design space for multi-perspective visual analysis.
	The work was presented at IEEE VIS 2015 and EuroVis 2016 and is published in Computer Graphics Forum

	Dykes, J., Rooney, C., Beecham, R., Turkay, C., Slingsby, A., Wood, J. & Wong, W. (2015). Multi-Perspective Synopsis with Faceted Views of Varying Emphasis. Paper presented at IEEE VIS 2015, 25-10-2015 - 30-10-2015, Chicago, USA. http://openaccess.city.ac.uk/12334/ Beecham, R., Dykes, J., Slingsby, A. & Turkay, C. (2015). Supporting crime analysis through visual design. Paper presented at IEEE VIS 2015, 25-10-2015 - 30-10-2015, Chicago, USA. http://openaccess.city.ac.uk/12331/ Beecham, R., Rooney, C., Meier, S., Dykes, J., Slingsby, A., Turkay, C., Wood, J. & Wong, B.L.W. (2016). Faceted Views of Varying Emphasis (FaVVEs): a framework for visualising multi-perspective small multiples. Computer Graphics Forum: the international journal of the Eurographics Association, 35(3), pp. 241-249. http://openaccess.city.ac.uk/14252/

Task 5.2	Textual Document Visualization
Task Leader	UKON
Participants	UKON, CITY, MU
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D5.7 UIDesignDataSpace Final.pdf (M16) **White Papers** ▪ VALCRI-WP-2017-011 Applying Visual Interactive Dimensionality Reduction.pdf **Software**
Summary of significant contributions	We developed and evaluated components to analyse semantic similarities of crime reports based on their modus operandi. ▪ We implemented a sequential pattern mining algorithm to better cover the semantic similarities by maintaining the order of the extracted concepts (WP8 Task 3) ▪ We integrated 3 projection algorithms (PCA / MDS / t-SNE) to visualize semantic similarities of crimes and 3 clustering algorithms (K-MEANS / DBSCAN / Hierarchical Clustering) to create similar groups of crimes. The clustered, projected crimes can be further investigated in the Concept Explorer. ▪ We developed and integrated the Concept Explorer into the VALCRI framework. The Concept Explorer is comprised of several tightly linked components that offer multiple perspectives onto the data- and feature-space: – We developed and integrated the Similarity Space Selector (S^3) component, which iteratively was being simplified due to users' feedback. – We developed and integrated the Sequence Similarity Space Selector (S^4) component that visualizes feature similarities and forms a Hybrid View together with the S^3 component. – We developed and integrated the Crime Cluster Table (CCT) that combines the data- and feature-space in a single view.

	– We developed and integrated the Pattern Selector (PS) component that allows browsing, filtering, and drilling down of the features (sequential patterns). – We developed and integrated the Weight Observer Component (WOC). The WOC is useful for analytic provenance and needed for the Criminal Investigators in order to reason about their analysis process. – We developed and integrated the Sequence Graph Component (SGC), which visualizes the semantic hierarchies of the extracted concept patterns. ▪ We had multiple iterations with the users (LEA) to receive feedback during the development. The feedback rounds were also used to train the users on the Concept Explorer. We successfully published our research on the analysis of high-dimensional data, sequence mining in text data, investigations on crime signature analysis, as well as the whole design process.

Task 5.3	Visual Image/Video Visualization
Task Leader	FHG
Participants	FHG
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D5.7 UIDesignDataSpace Final.pdf (M16)
Summary of progress and overall contribution to the project	The main achievement of this task was the implementation and integration of a graphical user interface for (a) interactive video playback, (b) visualizing pre-analysed/annotated and custom events in videos using synchronized heat maps (incl. different visualization styles) extracted by T8.2 components, (c) navigating a video based on certain events, (d) adding custom event annotations and (e) querying video files by: camera, tag, address and location. Moreover, the video player component was integrated with the UDB (see Task 11.6) for accessing video files and event metadata. Thus, video playback and access to metadata is managed by OpenPMF policies. Results from the service developed in T8.2 were pre-processed on the video and then fed into the unstructured storage. Development included several adaptations of the video player to address overall GUI framework changes. Moreover, we specified and applied a data model for storing event annotations using MongoDB. Several data sets containing video files and related event annotations from different sources have been prepared and have been made available as docker images demonstration the capabilities of the system as a Proof of Concept. In general activities have focused on extending, integrating, and testing the video/image analysis and the unstructured storage components within the VALCRI framework. We added further implementations on handling time view events in media item lists and unstructured storage requests including testing. We developed address search functionality for the unstructured storage service and the client. A *Suggestion Service* has been implemented including client request and server response with the preparation of the unstructured storage

	view by name tags. The inclusion of security implementation (CAS) for video retrieval has been realized. A synchronized workflow between the heat map and video player views has been developed. Further functionality has been added and tested, such as team assignment , dumping the mongo test database and grabbing information of denial access of media items (an object security task). During the entire period, we have had to adapt the video player to continuous AUI framework changes (docker, server, etc.). The client side video player classes have been refactored to better organize injects.

Task 5.4	UI Component Integration and Testing
Task Leader	XI
Participants	XI
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D5.7 UIDesignDataSpace Final.pdf (M16)
Summary of progress and overall contribution to the project	We have built a foundation for efficient and novel interactions with data visualizations using gestures. The summary of report based on comparative study explores several gestural interaction techniques and provide a summary of affordances and intuitiveness of those to various data exploration tasks. From all the different explorations we concluded that one meaningful shape (static gesture) could provide value in the context of provenance. We implemented a one shape recognition gesture for provenance into the VALCRI AUI (GWT framework).

Issues and Deviations from Planning:

Task	Timing	Status	Deviation / issues / Comments
T5.1	M01-M47	COMPLETED	No deviation
T5.2	M01-M47	COMPLETED	No deviation
T5.3	M01-M47	COMPLETED	No deviation
T5.4	M01-M47	COMPLETED	A deviation has been made from the original version of the DOW when XI departed the project in M32 their work was under taken by MU.

Required Corrective Action

Not Applicable.

Relevant Publications

Authors	Title	Year	Conference / Journal	Status
Dominik Jäckle, Florian Stoffel, Bum Chul Kwon,	Ambient Grids: Maintain Context-Awareness via Aggregated Off-Screen Visualization	2015	Eurographics Conference on Visualization (EuroVis),	PUBLISHED

Dominik Sacha, Andreas Stoffel, Daniel A. Keim				
Dominik Jäckle, Hansi Senaratne, Juri Buchmüller, Daniel A. Keim	Integrated Spatial Uncertainty Visualization using Off-screen Aggregation	2015	EuroVis Workshop on Visual Analytics (EuroVA),	PUBLISHED
Leishi Zhang, Chris Rooney, Lev Nachmanson, William Wong, Bum Chul Kwon, Florian Stoffel, Michael Hund, Nadeem Qazi, Uchit Singh, and Daniel Keim	Interactive Spherical MDS for Comparative Case Analysis	2015	Visualization and Data Analysis 2016	PUBLISHED
R. Beecham, J. Dykes, A. Slingsby, C. Turkay	Supporting crime analysis through visual design	2015	IEEE VIS 2015, Posters.	PUBLISHED
F. Stoffel, D. Sacha, G. Ellis and D. A. Keim	VAPD - A Visionary System for Uncertainty Aware Decision Making in Crime Analysis	2015	Symposium on Visualization for Decision Making Under Uncertainty at IEEE VIS 2015	PUBLISHED
Dominik Jäckle, Fabian Fischer, Tobias Schreck, Daniel A. Keim	Temporal MDS Plots for Analysis of Multivariate Data	2016	IEEE Transactions on Visualization and Computer Graphics (TVCG),	PUBLISHED
Dominik Jäckle, Johannes Fuchs, Daniel A. Keim	Star Glyph Insets for Overview Preservation of Multivariate Data	2016	IS&T Electronic Imaging Conference on Visualization and Data Analysis,	PUBLISHED
D. Sacha, L. Zhang, M. Sedlmair, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North and D. A. Keim.	Visual Interaction with Dimensionality Reduction: A Structured Literature Analysis.	2016	IEEE Transactions on Visualization and Computer Graphics (Proceedings of the Visual Analytics Science and Technology),	PUBLISHED
R. Beecham, C. Rooney, S. Meier, J. Dykes, A. Slingsby, C. Turkay, J. Wood, W. Wong	Faceted Views of Varying Emphasis (FaVVEs): a framework for visualising multi-perspective small multiples	2016	EuroVis 2016 and Computer Graphics Forum.	PUBLISHED

W. Jentner, G. Ellis, F. Stoffel, D. Sacha and D. A. Keim.	A Visual Analytics Approach for Crime Signature Generation and Exploration	2016	The Event Event: Temporal & Sequential Event Analysis, IEEE VIS,	PUBLISHED
Nadeem Qazi, Leishi Zhang, Eva Blomqvist, Florian Stoffel, Patrick Aichroth, Christian Weigel	White Paper: Applying Data Science to Criminal Intelligence Analysis	2016	-	PUBLISHED
D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis and D. A. Keim	The Role of Uncertainty, Awareness, and Trust in Visual Analytics	2016	IEEE Transactions on Visualization and Computer Graphics (Proceedings of the Visual Analytics Science and Technology)	PUBLISHED
L. Zhang, C. Rooney, L. Nachmanson, W. Wong, B. C. Kwon, F. Stoffel, M. Hund, N. Qazi, U. Singh and D. A. Keim	Spherical Similarity Explorer for Comparative Case Analysis	2016	IS&T Electronic Imaging Conference on Visualization and Data Analysis	PUBLISHED
D. Sacha, L. Zhang, M. Sedlmair, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North and D. A. Keim	Visual Interaction with Dimensionality Reduction: A Structured Literature Analysis	2017	IEEE Transactions on Visualization and Computer Graphics (Proceedings of the Visual Analytics Science and Technology)	PUBLISHED
D. Jäckle, F. Stoffel, S. Mittelstädt, D. A. Keim and H. Reiterer	Interpretation of Dimensionally-Reduced Crime Data: A Study with Untrained Domain Experts	2017	Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications	PUBLISHED
Dominik Sacha, Wolfgang Jentner, Leishi Zhang, Florian Stoffel, Geoffrey Ellis	Visual Comparative Case Analytics	2017	EuroVis Workshop on Visual Analytics (EuroVA)	PUBLISHED
W. Jentner, M. El-Assady, B. Gipp and D. A. Keim	Feature Alignment for the Analysis of Verbatim Text Transcripts	2017	EuroVis Workshop on Visual Analytics (EuroVA)	PUBLISHED
Florian Stoffel, Wolfgang Jentner, Michael Behrisch, Johannes Fuchs, Daniel Keim	Interactive Ambiguity Resolution of Named Entities in Fictional Literature	2017	EuroGraphics: Computer Graphics Forum	PUBLISHED

J. Buchmüller, W. Jentner, D. Streeb and D. A. Keim	ODIX: A Rapid Hypotheses Testing System for Origin-Destination Data	2017	IEEE Conference on Visual Analytics Science and Technology (VAST Challenge 2017 MC1)	PUBLISHED
D. Jäckle, M. Hund, M. Behrisch, D. A. Keim and T. Schreck	Pattern Trails: Visual Analysis of Pattern Transitions in Subspaces	2017	IEEE Conference on Visual Analytics Science and Technology (VAST)	PUBLISHED
Wolfgang Jentner, Dominik Sacha, Florian Stoffel, Geoffrey Ellis, Leishi Zhang, Daniel A. Keim	Making machine intelligence less scary for criminal analysts: reflections on designing a visual comparative case analysis tool	2018	Springer: The Visual Computer	PUBLISHED

2.2.6 Work Package 6 - ANALYSIS SPACE

Work Package Summary

WP 6	ANALYSIS SPACE
WP Leader	SPACE
Participants	MU, AES, OS, XI
Status	COMPLETED
Summary of key outputs	**Deliverables** ▪ D6.8 UIDesignAnalysisSpace Final.pdf (M16) **White Papers** ▪ VALCRI-WP-2017-003 Data Science.pdf
Summary of significant contributions / impact for the project	▪ Definition of the software system design for allowing semantic search and retrieval capabilities. ▪ Implementation of the design by means of a generic cross-filtering system. ▪ Implementation of numerous search and retrieval strategies accompanied with their dedicated indexing mechanisms. ▪ Runtime switching multiple data sources. ▪ Temporal, geographical and aggregation data representations. ▪ Integration testing for each of the search strategies and critical system functionalities.
Further information	During M1 till M17, this work package was in a very early stage of development. This was justified by the focus and high involvement of the key WP6 partners in WP2, WP4 and WP12. In the period after that, during M17 and M32, the early ideas were transformed into a software design and concretised into functional software components. Most of these components are now at the very basis of the VALCRI system in order to allow semantic search, interactive querying through the visualisations. In M33 to M45 we continued working on the FPA setup (high TRL software deployed in FPA premises and using real data). We were able focus on the searching capabilities of the FPA setup because the functional basis of the

	system was stabilised and ready for use. We have developed temporal, geographical and aggregated data visualisations that allow the result set to be visualised in each of its facets. We also added the ability to switch between different data sources at runtime. This allowed the user to find links between different structured or unstructured data sources.

Task 6.1	Semantic Search and Retrieval
Task Leader	SPACE
Participants	SPACE, AES
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D6.8 UIDesignAnalysisSpace Final.pdf (M16)
Summary of progress and overall contribution to the project	We defined and implemented a software design that allows cross-filtering the VALCRI data over its many different facets: temporal, geographical, relational, etc. We were able to refine and stabilise the design over multiple iterations and end-user deployments. As such, several performance issues and limitations were overcome. The cross-filtering mechanism is a key technology allowing the end-user to interact with a certain visualisation (e.g. zoom in on a map) and allow cross-filtering any other related visualisation and update itself with the correct data. Furthermore, utilising the background work undertaken in T6.2 and working with MU on the associative search we have: ▪ Researched different MO modelling techniques using clustering techniques ▪ Created and analysed criminal networks ▪ Worked on effective visualisations and testing them on the end users In M33 to M45 we have updated and added new visualisations to the FPA setup. In general, three types of visualisations were implemented: geographical, temporal and analytical. We have implemented three new visualisations. If the result set contains geographical information, it will show (1) a contextualized heat map. This means that it performs geographic aggregations and uses the histographic information to highlight hotspots. Furthermore, it will auto-zoom and focus on the relevant areas in the result set. If the result contains temporal information it will show a (2) timeline as well as a (3) time-scatter-plot - a visualisation that relates the count of documents from a day in a week to hours in a day. Each visualisation takes the current search context into account and updates itself accordingly. Under the hood, the cross-filtering mechanism is taking care of most of the heavy lifting. Every time the user alters the query, each of the cards will adapt according to the new result set. In order to tap into more data sources and allow users to find links between both structured and unstructured data sources, we have implemented the ability to switch back and forth between multiple data sets. This allows users to perform the same searches (and their associated resulting visualisations) on each of those data sources.

Task 6.2	Criminal Profiling Based Search
Task Leader	AES
Participants	SPACE
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D6.8 UIDesignAnalysisSpace Final.pdf (M16)
Summary of progress and overall contribution to the project	By examination of the offender records and their physical descriptions we have developed and tested: ▪ Fuzzy age band schema ▪ Fuzzy height schema ▪ Crime time bands ▪ Crime day bands (single day, overnight and long weekend bands were created and tested) ▪ Criminal profile clustering ▪ Theft from Motor Vehicle console crime The visual templating system, as outlined in D6.8 has been implemented. These templates allow criminal profiling through a generalised UI framework, supported by the available concepts in the knowledge base. There were several meetings with LIU and UKON. After testing the outputs, this information was passed to LIU for inclusion into the ontology. The crime time and day bands are directly included in the VALCRI system. After testing the outputs, this information was passed to LIU for inclusion into the ontology.

Task 6.3	Natural Language Interfaces
Task Leader	SPACE
Participants	SPACE
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D6.8 UIDesignAnalysisSpace Final.pdf (M16)
Summary of progress and overall contribution to the project	After investigating the question answering system, developed in the past by SPACE, it was decided to put this system on hold in order to be able to focus on task 6.1. This makes sense because the cross-filtering mechanisms serve as the basis for incorporating the natural language processing capabilities needed for this task. Because the users' feedback suggests that the semantic search provides the features they are looking for, we are still considering how to proceed with this system. In parallel, we are working with partners parsing modus operandi from free text to establish CCA table(s). Using the cross-filtering architecture at its core, we were able to considerably expand the searching capabilities in M33 to M45. The user can enter any kind of free text query into the LPA setup and the system will optimise its search

	strategies accordingly. In concreto, we have implemented search strategies that include, but are not limited to: smart case searching, strict, partial and fuzzy searching and code searching (e.g. phone numbers, document ID's, vehicle number plate, partial matching). The strategies were selected according to the specific operational needs of the LPA analysts. These searching capabilities are enabled by a set of indexing strategies that boost runtime search performance. We have implemented custom analysers, tokenizers and filters that map incoming into a suitable representation for free text querying

Task 6.4	UI component integration and testing
Task Leader	XI
Participants	XI, SPACE, AES
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D6.8 UIDesignAnalysisSpace Final.pdf (M16)
Summary of progress and overall contribution to the project	Using the cross-filtering mechanism, as developed in T6.2, we were able to put the envisaged interactive working principles of the UI into practice. The visualisations, as developed in WP4, were interconnected and integrated using this mechanism. Evaluations being recorded in WP13. In support of searching capabilities, as developed in T6.3, we have extended our test suite with tests that operate based on probabilities rather than facts. The results of integrating probability tests give a good indication whether a set of features have evolved correctly over each release of the software and can still be considered as useful. We have integrated probability tests in our testing pipeline by ingesting specific sets of data that mimic use-cases defined by analyst end users on which the testing pipeline performs a specific set of search and analysis features. Each probability test must pass a certain probability score for the pipeline to succeed.

Issues and Deviations from Planning:

Task	Timing	Status	Deviation / issues / Comments
T6.1	M07-M47	COMPLETED	No deviation
T6.2	M07-M47	COMPLETED	No deviation
T6.3	M07-M47	COMPLETED	No deviation
T6.4	M07-M47	COMPLETED	A deviation has been made from the original version of the DOW when XI departed the project in M32, their work and PM were under taken by MU.

Required Corrective Action

See table above.

Relevant Publications

Authors	Titles	Year	Conference / Journal	Status

2.2.7 Work Package 7 - HYPOTHESIS SPACE – USER INTERFACE

Work Package Summary

WP 7	HYPOTHESIS SPACE – USER INTERFACE
WP Leader	MU
Participants	MU, CITY, AES, XI
Status	COMPLETED
Summary of key outputs	**Deliverables** ▪ D7.9 UIDesignHypothesisSpace Final.pdf (M16) **White Papers** ▪ VALCRI-WP-2017-009 Provenance3.pdf
Summary of significant contributions / impact for the project	▪ Design of provenance capture, analysis and replay system. ▪ Prototype of a situational awareness tool using superimposed views. ▪ Redesign of a Statistical Process Control charts and development of a standalone stool for viewing SPC data geographically. This tool has been deployed at West Midlands Police and is being used to analyses real police data. ▪ Developed an award-winning method of communicating new visualization designs to domain-expert analysts through *Dynamic Design Documents*
Further information	N/A

Task 7.1	Situation Re-construction
Task Leader	CITY
Participants	MU XI AES
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D7.9 UIDesignHypothesisSpace Final.pdf (M16)

Summary of progress and overall contribution to the project	Our interactive interfaces for situation reconstruction combine geo-spatial and temporal visualization techniques to inform situational awareness and provide means for developing and structuring narrative through which this can be assessed and shared in the context of uncertainty. We have re-designed **Statistical Process Control** (SPC) charts, as used by crime analysts at West Midlands Police (WMP), for a series of decision-making and reporting tasks to supports the spatial and historical analysis of signals in crime data in ways that current techniques do not. Our work involved characterising the data and analysis tasks associated with crime monitoring through SPC charts and redesigning a family of graphics through which these tasks can be achieved. Through our engagement with analysts at WMP we have: developed a new approach to engaging with collaborators that may be used in applied visualization work more widely; proposed a new visual encoding of design characteristics; established **Dynamic Design Documents** as a means of explaining, justifying, exploring and evaluating candidate designs with analysts in the workplace. The methods support ongoing dialogue - between designers and analysts and amongst analysts. We have applied the notion of **Literate Visualization** to the generation of Dynamic Design Documents that provide interactive prototypes and design exposition as a means of engaging with users in applied contexts. These integrate live coding input, rendered visualization output and textual narrative to support the generation of data graphics and textual narrative at low authoring cost. The DDDs can be structured and validated according to design schema and enable narratives to branch in ways that support alternative designs, explanations and design views.
Publications	This work has been presented at, submitted to and received awards from the leading visualization conferences in Computer Science. Rooney, C., Beecham, R., Dykes, J. & Wong, W. (2017). Dynamic Design Documents for supporting applied visualization. (InfoVis Best Poster). Paper presented at IEEE VIS 2017, 01-10-2017 - 06-10-2017, Phoenix, USA. http://openaccess.city.ac.uk/view/creators_id/roger=2Ebeecham=2E1.html Rooney, C., Beecham, R., Dykes, J. & Wong, W. (submitted an 2018). Statistical Process Control Charts: A Visualization Makeover for Crime Analysis. IEEE Transactions on Visualization and Computer Graphics. Wood, J., Kachkaev, A. & Dykes J. (submitted to IEEE InfoVis 2018). Design Exposition with Literate Visualization. IEEE Transactions on Visualization and Computer Graphics

Task 7.2	Structuring and Explaining Arguments
Task Leader	MU
Participants	XI AES
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D7.9 UIDesignHypothesisSpace Final.pdf (M16)

Summary of progress and overall contribution to the project	Created the back end system infrastructure to facilitate the interaction of components thus facilitating searching and reasoning throughout differing abstract levels. Sub components were tested with end user partners. Created a set of interactions called 'connecting the dots' that allows the analysts to collate several pieces of evidence into a group. This group can be used by analysts as they see fit, examples include a shoebox of relevant evidence or a temporal narrative of activity. Created the ability to annotate analysis and publish them as reports, to either be presented to decision makers, or contribute to future analysis as a supporting argument.

Task 7.3	**Conclusion Pathways**
Task Leader	MU
Participants	XI AES
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D7.9 UIDesignHypothesisSpace Final.pdf (M16) **White Papers** ▪ VALCRI-WP-2017-009 Provenance3.pdf
Summary of progress and overall contribution to the project	Developed a provenance system in the core of the VALCRI system, capturing analyst interactions. Analysts can undo/redo their actions and make bookmarks or mark decision points in their analysis. Further to this, we have developed a provenance visualisation system for exploring analyst's interactions. The view shows a tree of activity, allowing meta-analysts to see avenues of investigation and false leads. The view can be filtered to show a subset of actions, and clicking on a node in the tree jumps to that point in the analysis - arranging visualisations and data filters as they were at that point in time.

Issues and Deviations from Planning:

Task	Timing	Status	Deviation / issues / Comments
T7.1	M07-M47	COMPLETED	A deviation has been made from the original version of the DOW when XI departed the project in M32 their work was under taken by MU.
T7.2	M07-M47	COMPLETED	A deviation has been made from the original version of the DOW when XI departed the project in M32 their work was under taken by MU.
T7.3	M07-M47	COMPLETED	A deviation has been made from the original version of the DOW when XI departed the project in M32 their work was under taken by MU.

Required Corrective Action

None.

Relevant Publications

Authors	Title	Year	Conference / Journal	Status
Celeste Groenewald, Junayed Islam, Craig Anslow, Chris Rooney, William wong	Understanding 3D Mid-Air Hand Gestures with Interactive Surfaces and Displays: A Systematic Literature Review	2016	Proceedings of the 30th International BCS Human Computer Interaction Conference (HCI 2016)	PUBLISHED
Chris Rooney, Neesha Kagadoa, Craig Anslow, William Wong	Multi-device, interaction	2015	EICS 2015 Workshop on Cross-Device User Interfaces (XDUI 2015)	PUBLISHED
Islam, Junayed and Anslow, Craig and Xu, Kai and Wong, William and Zhang, Leishi	Towards Analytical Provenance Visualization for Criminal Intelligence Analysis	2016	Computer Graphics and Visual Computing (CGVC)}	PUBLISHED

2.2.8 Work Package 8 - DATA EXTRACTION

Work Package Summary

WP 8	DATA EXTRACTION
WP Leader	FHG
Participants	MU, SPACE, UKON, AES, FHG
Status	COMPLETED
Summary of key outputs	▪ Integrated textual analysis component with multi language support ▪ Associative Search module incorporated in VALCRI Framework ▪ First CNN model for VALCRI object detection ▪ Joint white paper on WP8 and WP10 work ▪ Close interaction with SEPL and VIZ groups
Summary of significant contributions / impact for the project	The outputs of the work package (esp. textual and visual analysis components) allow more efficient search and analysis of unstructured content

Task 8.1	Data Analysis Framework
Task Leader	FHG
Participants	FHG
Status	COMPLETED
Output and deliverables	**Deliverables**

	▪ D8.10 SystemDesignDataAnalysisFramework Final.pdf (M16)
Summary of progress and overall contribution to the project	D8.10 included a draft re orchestration issues, with an outline of relevant high-level requirements and some technology candidates that could be used to address these requirements. Within the 2nd phase of the project, however, it became clear that there was no need to support complex workflows, for two reasons: (1) there are only few analysis modules, and as became evident during requirement analysis, they do act independently from each other; (2) within this project phase, there was an agreement on using ONE structured database (SDB) which now serves as a common integration point, thereby supporting interaction between analysis components. Respective efforts were therefore moved into WP11 PET (in which activities went significantly beyond what was planned in the DoW). The design of asynchronous communication of visualization and analysis VALCRI services also wr.t. the security model was carried out. This is directly related to the implementation work of that VALCRI service in T8.2.

Task 8.2	Visual Content Analysis
Task Leader	FHG
Participants	FHG
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D8.11 DataAnalysisDataExtraction Final.pdf (M16) **White Papers** ▪ VALCRI-WP-2017-003 Data Science.pdf **Software** ▪ Software components (video analysis service) in the VALCRI repository
Summary of progress and overall contribution to the project	Within discussions with end-users in the second phase of the project, object detection became a priority, and it was therefore decided to develop a new object detection based on convolutional neural network (CNN) that can quickly identify objects of specific classes (person, car etc.) within a video. IN addition, methods to track the results of such detectors over time were also investigated. This was achieved by using a state of the art single shot deep learning object detection system (based on the YOLOv2 architecture), training it for VALCRI-specific concepts (person, car, etc.) and a Kalman Filter based tracking approach in order to keep track of an object once it has been detection. Although the system is capable to detect and track the region of multiple objects within a video, for VALCRI it was decided to count the occurrences of objects and display them as heat map along with the video player (see WP5). In order to facilitate integration within the VALCRI system, a service /w a simple interface for the video analysis tools was provided, which was adapted several times due to changes in the overall system design (e.g. the original ActiveMQ messaging implementation had to switch to a REST-based approach). The

	service can be easily extended with plugins for other classification and object detection tasks which makes it ready for new algorithms that might come up in the future. Metadata formats for low-level native annotations (e.g. motion information, regions in video frames) were specified, and considering that they were of low-level nature (not to be used for end-user search) and to be independent from the SDB discussions within the project, these annotations (XML or json) are stored within the unstructured storage (UDB) provided by WP11. In order to efficiently keep up with changing VALCRI development cycles and architectural changes (esp. w.r.t. to deployment), a VirtualBox development image was created and maintained that helps developers to dive into VALCRI development much faster. While not being exclusively relevant to T8.2, it helped to speed up development in general

Task 8.3	Textual Document Analysis
Task Leader	UKON
Participants	UKON
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D8.11 DataAnalysisDataExtraction Final.pdf (M16) **White Papers** ▪ VALCRI-WP-2017-003 Data Science.pdf Software Software to analyse textual documents
Summary of progress and overall contribution to the project	We developed the following key components to analyse unstructured textual documents: ▪ Data pipeline for integrating different data sources containing unstructured data. ▪ Unified handling of external text analysis components to enrich and annotate unstructured data, such as part of speech taggers or sentiment determiners. ▪ Completed descriptions of analysis components so that the analysis can be composed out of steps that can be configured and put together independently from each other. ▪ Completed multi-language support for each analysis component. ▪ Integrated basic data provenance framework that recorded changes per analysis step as textual diffs. ▪ Extended and adapted concept lists (lexicons) to fit better to the VALCRI data. ▪ HTTP-based stateless REST API to configure the analysis and concept lists, as well as to execute the textual document analysis on the fly. ▪ Integrated the textual document analysis in a Docker container for easy deployment and use.

	▪ Performance optimization that the textual document analysis REST API can be utilized in interactive systems with ad-hoc text data analysis. Most of the items have been subject to informal feedback collected from the experts in the VACLRI consortium and numerous iterations incorporating the feedback. Besides the textual document analysis framework, we were experimenting with various other analysis methods (shallow parsing, various different chunkers, parse tree traversal-based information extraction) for additional enrichments of the text document annotations. In order to incorporate more data sources from the FPA, SPACE has included an OCR/Text module into the ingester that can convert most common data sources (MS Office formats, PDF and common image formats) into plain text so the user can start searching through them.

Task 8.4	Financial And Travel Data Analysis
Task Leader	AES
Participants	AES
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D8.11 DataAnalysisDataExtraction Final (M16)
Summary of progress and overall contribution to the project	▪ Worked with the WMP Fraud Squad to elicit requirements. ▪ Using the 1.8 million synthetic finance records produced by SPACE in T11.3 and created processes to add meta data to support the requirements. ▪ Created a series of smaller data sets to undertake visualisation testing. ▪ Inserted fictitious data into the smaller data sets to support one of the one of the requirements to test visualisations. ▪ Developed a series of visualisations to support the analysis. ▪ Using the developed tool there is an ability to detect a range of financial fraud within a network of connected people. ▪ Created a user manual to support the software. ▪ Supported Linkoping in the analysis of real time travel analysis using the ANPR data

Task 8.5	Associative Search (**NOTICE:** This task is not an official task of the VALCRI DoW. Yet, since the topic and the work done differ significantly from the original **T8.3** where the resources had been allocated, we decide to separate these topics)
Task Leader	MU
Participants	MU, XI, AES
Status	COMPLETED
Output and deliverables	**White Papers** ▪ VALCRI-WP-2017-003 Data Science.pdf

Summary of progress and overall contribution to the project	▪ Associative module has been incorporated in VALCRI Framework using VALCRI Data via SQL Server. ▪ Three levels of Associations have been established from the given data consisting of crime and nominals. ▪ Between Crimes ▪ Between Crimes and offenders ▪ Between Offenders and offenders ▪ Temporal spatial and Modus operandi based interactive clustering is performed for grouping the search results based on the user specified feature vector. ▪ Visualisation is achieved through clustering using MDS and D3 in two separate layers ▪ The Aggregated view offers doughnuts view to find out the clusters of solved unsolved crimes and associated offenders. ▪ The Detail view shows offender card including his crime hot spots and crime network in Spatial Temporal Knowledge graph. A prototype for Associative search was developed and presented to end user in May 2017 berlin, Germany. It included ▪ Multi-level association model for extracting co-offender network and plausible suspect list for given offender and unsolved crime respectively ▪ A hierarchical knowledge graph. Widget was used to visualize the plausible suspect list and co-offender network

Issues and Deviations from Planning:

Task	Timing	Status	Deviation / issues / Comments
T8.1	M01-M47	COMPLETED	No deviation
T8.2	M01-M47	COMPLETED	No deviation
T8.3	M01-M47	COMPLETED	No deviation
T8.4	M01-M47	COMPLETED	No deviation
T8.4	M01-M47	COMPLETED	This task is not an official task of the VALCRI DOW. Yet, since the topic and the work done differ significantly from the original T8.3 where the resources had been allocated, we decide to separate these topics.

Required Corrective Action

None.

Relevant Publications

Authors	Title	Year	Conference / Journal	Status
N. Qazi, B.L. William Wong,	Interactive Knowledge Discovery Scheme for Crime Pattern mining and Community Detection	2018	KDD 2018,London 2018	Submitted

N. Qazi, B.L. William Wong,	A Human Centred Data Science Approach towards Crime matching	2018	to Journal of Information Processing and Management	Submitted
N. Qazi, B.L. William Wong	Contextual visualization of crime matching through similarity clustering and Bayesian analysis	2018	Chapter in book titled : Social Media Strategy in Policing	Accepted
N. Qazi, B.L. William Wong	Behavioural Tempo-spatial Knowledge Graph for Crime matching through Associate Questioning and Graph Theory	2017	IEEE, European Intelligence, and Security Informatics Conference EISIC,2017,Geece)	Published
Nadeem Qazi, Leishi Zhang, Eva Blomqvist, Florian Stoffel, Patrick Aichroth, Christian Weigel	White Paper: Applying Data Science to Criminal Intelligence Analysis	2016	-	Published
N. Qazi, B.L. William Wong, Neesha Kodagoda and Rick Adderley	Associative Search through Formal Concept Analysis in Criminal Intelligence Analysis	2016	SMC 2016	Published
N. Qazi ,William Wong	Semantic Based Image Retrieval Through Combined Classifiers of Deep Neural Network and Wavelet Decomposition of Image Signal	2016	EUROSIM Congress on Modelling and Simulation	Published
L. Zhang, C. Rooney, L. Nachmanson, W. Wong, B. C. Kwon, F. Stoffel,N. Qazi, U. Singh and D. A. Keim	Spherical Similarity Explorer for Comparative Case Analysis	2016	Electronic Imaging Conference on Visualization and Data Analysis	Published
W. Jentner, G. Ellis, F. Stoffel, D. Sacha and D. A. Keim	A Visual Analytics Approach for Crime Signature Generation and Exploration.	2016	IEEE VIS 2016 Workshop	Published
F. Stoffel, D. Sacha, G. Ellis and D. A. Keim.	VAPD - A Visionary System for Uncertainty Aware Decision Making in Crime Analysis	2015	IEEE VIS 2015 Workshop	Published

2.2.9 Work Package 9 - EVENT DETECTION

Work Package Summary

WP 9	EVENT DETECTION
WP Leader	LIU
Participants	LIU, SPACE
Status	COMPLETED
Summary of key outputs	▪ Evaluation of RDF Stream Processing (RSP) systems and their limitations with respect to application in VALCRI use-cases ▪ Development of RSP-SPIN, an extension to express RSP query templates in a format compatible with that used to in the structured storage

	▪ Demonstration of how RSP-SPIN can be used to transition between RSP dialects ▪ Architecture for distributing streaming data across multiple stream processing systems, using ActiveMQ to handle inter-component communication. ▪ CO-060: RSP-SPIN
Summary of significant contributions / impact for the project	▪ RSP functionality is exposed to other system components based on prepared query templates, which are exposed in a dynamically generated REST API. This hides a lot of the complexity from clients, such as details about the particular RSP engine or language employed under the hood. To internally stream data between components the architecture leverages ActiveMQ but it also supports Apache Kafka

Task 9.1	Requirements and technology survey
Task Leader	LIU
Participants	SPACE
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D9.12 DataAnalysisEventDetection Final.pdf (M16)
Summary of progress and overall contribution to the project	There are several different approaches to stream processing, and an increasing number of these are aimed at supporting large-scale data processing and distributed processing. We have evaluated a number of RDF Stream Processing systems with respect to the expressivity required for the VALCRI use-cases (US-041.1, US-037.1, US-037.1, US-19.1). Two systems, C-SPARQL and CQELS, were found to have the expressive power required for these use-cases. Supporting other types of stream processing techniques can help scale data processing further, which is necessary to manage high-velocity streams (e.g. thousands of events per second). Importantly, for these systems to interact with RSP reliably they need to respect two principles: strict ordering of events, and one-time delivery semantics. To communicate high-velocity streams reliably in a distributed context ActiveMQ, which is supported in the current RSP service component (CO-055), does not scale very well. A more scalable approach is to instead use Apache Kafka as the messaging service, which is supported by many of the high-level stream processing systems.

Task 9.2	Semantic complex event processing - method development
Task Leader	LIU
Participants	SPACE
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D9.12 DataAnalysisEventDetection Final.pdf (M16) CO-060: RSP-SPIN

Summary of progress and overall contribution to the project	Developing queries in a streaming context is time-consuming and error prone. To make queries reusable and easier to maintain we extended SPIN (SPARQL Inferencing Notation) to support the representation of RSP queries, thus allowing us to represent the RSP query templates in a format compatible with that used to represent the query templates of the structured database (see WP10 for details). The service for executing RSP queries will be aligned with *CO-059: Structured Storage REST API*, allowing a similar level of access control to be implemented without any additional overhead.

Task 9.3	Method integration, method interactions and technical evaluation
Task Leader	LIU
Participants	SPACE
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D9.12 DataAnalysisEventDetection Final.pdf (M16)
Summary of progress and overall contribution to the project	The two RSP systems, C-SPARQL and CQELS, were found to scale to the VALCRI use-cases given the current ANPR data volume and velocity. They provide different expressivity and will therefore be supported in parallel. They also provide the means to incorporate knowledge from the structured database, although aspects such as refresh rates have yet to be explored. The security implementation of VALCRI currently complicates matters, since it blocks access for non-user initiated requests. As a next step we will therefore require certificate support for accessing CO-059. A major limitation of how time is represented in current RSP systems is that it is bound to individual RDF triples. This means that expressing complex events is often problematic if they are constructed in multiple processing steps. The upcoming standardized query language, RSP-QL, is used as a reference point to future-proof the system, and RSP-SPIN allows us to seamlessly transition into either CQELS or C-SPARQL in the meantime. We have made some initial evaluations of common stream processing systems (Samza, Spark Streaming, and Apache Storm). While these systems cannot take the place of RSP they can facilitate pre-processing, filtering, and aggregation of data, as well as open up a host of data analysis techniques and machine learning. Following the evaluation we've chosen to evaluate Spark Streaming further. To manage the communication between the RSP systems and Spark Streaming we have experimented with a communication pipeline based on Kafka, which is to be used for intra-component communication (i.e. not with the VALCRI system as a whole).

Issues and Deviations from Planning:

Task	Timing	Status	Deviation / issues / Comments

T9.1	M01-M47	COMPLETED	Minor deviations from the planning due to the lack of clear system requirements, leading to a more generic study than what was planned initially.
T9.2	M01-M47	COMPLETED	Minor deviations from the planning due to the lack of clear system requirements, leading to more generic system components than what was planned initially.
T9.3	M01-M47	COMPLETED	Minor deviations from the planning due to the lack of clear system requirements, which makes it difficult to technically evaluate and test the components against such requirements initially.

Required Corrective Action

Lack of clear system requirements: this has been an overall issue in the project, and is not dealt with on WP9 level specifically. However, a requirements management system was put into place and requirements were being entered, to assist with the minor deviation during project phase.

Relevant Publications

Authors	Title	Year	Conference / Journal	Status
Robin Keskisärkkä	Representing RDF Stream Processing Queries in RSP-SPIN	2016	Proceedings of the ISWC 2016 Posters & Demonstrations Track co-located with the 15th International Semantic Web Conference (ISWC-2016)	PUBLISHED
Robin Keskisärkkä	Query Templates for RDF Stream Processing	2016	Stream Reasoning Workshop 2016 Collocated with the 15th International Semantic Web Conference (ISWC 2016)	PUBLISHED

2.2.10 Work Package 10 - ONTOLOGY LIBRARY

Work Package Summary

WP 10	ONTOLOGY LIBRARY
WP Leader	LIU – Henrik Eriksson
Participants	MU, LIU, AES, FHG, OS, SPACE
Status	COMPLETED

Summary of key outputs	**Deliverables** ▪ D10.13 DataAnalysisModels Final.pdf **White Paper** ▪ VALCRI-WP-2018-019 Integration of Streaming Data.pdf **Other** ▪ Ontologies to represent and reason over the VALCRI datasets ▪ Methodology extensions and tool adaptations for ontology development, alignment, evolution and maintenance in the VALCRI setting, and in the criminal intelligence domain in general ▪ Conversion of VALCRI datasets into RDF on the fly, i.e. components for ingestion and transformation of legacy data formats ▪ Setup and maintenance of the main VALCRI structured data storage solution (SDB), including a sSwitch from Virtuoso to Jena Fuseki ▪ REST API to access the structured database using query templates ▪ Partially populated query template library
Summary of significant contributions / impact for the project	▪ Structured storage solution (SDB) for the overall VALCRI system, including conversion of VALCRI datasets into RDF on the fly, REST API to access the structured database using query templates (including a partially populated template library), and Elasticsearch integration for search performance ▪ Ontologies to represent, access and reason over the VALCRI datasets, including a component-based ontology architecture that supports easy replacement of data sources, and methods and tools for creating new or evolving existing ontologies ▪ Support for SEPL requirements in both the data storage solution, REST API, and ontologies ▪ Conversion of VALCRI datasets into RDF on the fly ▪ Switch from Virtuoso to Jena Fuseki ▪ REST API to access the structured database using query templates ▪ Partially populated query template library
Further information	N/A

Task 10.1	Data models
Task Leader	SPACE
Participants	LIU, FHG, OS
Status	COMPLETED

Output and deliverables	▪ Crime domain ontology (version 1.1) ▪ VALCRI crime data ontology (version 3.1) ▪ Crime profiles ontology (version 1.0)Crime classifications ontology (version 1.0) ▪ Provenance ontology (version 1.10) ▪ Finance ontology (version 1.0) ▪ Information extraction ontology (version 1.0) ▪ Annotations ontology (version 1.0) ▪ Unstructured data ontology (version 1.0) ▪ Structured storage REST API (CO-059): - uUpdated to support the new choice of storage back-end ▪ Integration of SDB and Elasticsearch, for better search performance ▪ CO-061: CSV-to-RDF Ingester
Summary of progress and overall contribution to the project	The set of ontologies had to modified and complemented to represent the anonymized VALCRI datasets that are now available. Main ontology development work was performed by LIU. Importantly, we now distinguish between concepts and properties that belong to the more general crime domain, in contrast to those that should be regarded as specific to the VALCRI data (i.e. mainly data originating from WMP). This is an important contribution since it allows the system to be used with a different set of data sources, without completely replacing all the data models. Originally RDF datasets were generated in VALCRI using a freely available visual mapping tool (Open Refine). This tool to facilitates a manual one-time export of data. LIU was requested to instead provide "ingesters" that could produce these RDF data mappings on the fly, based on a set of mappings. Ingesters have now been implemented for almost allmost VALCRI datasets (CO-061: CSV-to-RDF Ingester). To consolidate the various storage solutions that were used in the project the DAR (Design and Architecture special interest group) we conducted a new requirements collection effort and a review session of possible replacements for the initial structured storage (SDB) Virtuoso. We evaluated severalSeveral alternatives were then evaluated based on thean updated list of storage requirements, mainly by LIU and SPACE, and it was found that Jena Fuseki was the best alternative when taking into account RDF/OWL support, licencing, scalability, reasoning, and quad support. Changing the structured storage backend from Virtuoso to Jena Fuseki required a major update in the structured storage REST API (CO-059), implemented by LIU. The new version integrates SPIN (SPARQL Inferencing notation) to allows users to execute queries based on templates represented in the structured storage, similar to how stored procedures are used in relational databases. The service is also integrated with OpenPMF, which lets additional access control to be configured based on the exposed query templates. The use of this APIis has major benefits with respect to security (e.g. reduced injection risk) and usability (e.g. developers are not required to write and maintain SPARQL queries, and the underlying query format is hidden from other system components). The query templates also allow queries to be reused and make it possible to update templates as the ontology evolves, often without requiring any changes at all in the dependent components.

	During M17 and M32 considerable effort was spent on indexing part of the data in a way that the data model remains intact and to transparently accommodate for faceted search and free text querying through those relations. This was done mainly by LIU, in collaboration with MU and SPACE. These functionalities are key enablers for the visual analytical features of the VALCRI system, as they require very low- latency responses from the datae store. We were able to use part of the WP11 data set to experiment with, but most importantly, we were able to validate and test various strategies together, in close collaboration with the Federal Police in Antwerp using an offline, isolated installation. In the context of working with real data, during M33 and M45 SPACE has investigated how data related to crimes and crime reports is captured and structured in the so called ISLP (Integrated System for Local Police) database. In cooperation with the local police of Antwerp we have tried to understand how the various bits of information ("entities") are related to each other and how they are stored, used and queried. In this context it became clear that the semantic interpretation of certain entities may differ or evolve, sometimes to represent completely new concepts. In addition, the system encourages a dichotomy of users: information providers and information users. Currently the ISLP software favours data entry with little restrictions, making coherent efficient and effective use of that data challenging. Even though time constraints did not allow us to pursue this further, ideally we wanted make suggestions and/or improvements to the system to bring these main classes of users closer together.

Task 10.2	Provenance model
Task Leader	MU
Participants	LIU
Status	COMPLETED
Output and deliverables	▪ Demo for saving and retrieving states in the UI using SQL ▪ Provenance ontology (version 1.0) ▪ Switch to quad representation in the structured storage to support data provenance on entity/document level ▪ Template-based data access for supporting the enforcement of provenance-based restrictions
Summary of progress and overall contribution to the project	MU implemented a demo showing how states in the UI could be stored and retrieved using an SQL database, while tracking an analyst's actions. An ontology was created based on the storage requirements, and the appropriate query templates for storage and retrieval were added to the template library. The focus of this ontology was storage and retrieval of analytical provenance. In cooperation with the SEPL subgroup LIU discussed the data provenance requirements and concluded that triple based provenance tracking would not be required. Instead data provenance annotations could be done on an entity or document level (e.g. on the level of a crime or nominal participation). This prompted us to switch from a triple based representation to a quad based one, i.e. named graphs, since this allows us to easily annotate anything that is captured within a graph with provenance information.

	For actually expressing annotations on data the W3C standard PROV-O was used as a basis, but was extended by LIU to support the kinds of provenance needed in VALCRI (according to what has been specified by the SEPL group), which includes support for the enforcement of legal, ethical and privacy requirements in the form of access restrictions, anonymization, or deletion of data under certain criteria. Another important component to support this is the API solution, including SPIN templates, developed in T10.1, since this enables to take the annotations into account when accessing data.

Task 10.3	Criminal Profile Ontology Development
Task Leader	AES
Participants	LIU
Status	COMPLETED
Output and deliverables	Crime profiles ontology (version 1.0)
Summary of progress and overall contribution to the project	Summary of progress and overall contribution to the project ▪ In developing the first version of the criminal profile ontology and laying the framework for future development we have had several F2F meetings (Linköping, Warsaw, and Middlesex). The current version of the ontology exemplifies parts of these discussions, but has also been further extended with additional aspects of a criminal profile. ▪ The above was extended into the further development of the ontology. ▪ AES Developed and tested, Fuzzy age band schema, Fuzzy height schema, Crime time bands, Crime day/overnight/weekend time bands and passed to LIU for inclusion into the ontology, all which is included in the latest version. ▪ AES dDeveloped, extracted and tested a set of vehicle console theft crime entities and passed to LIU for inclusion into the ontology, all which is included as a test case example in the latest ontology version.

Task 10.4	Ontology Evolution
Task Leader	LIU
Participants	
Status	COMPLETED
Output and deliverables	▪ Requirements for user support for ontology alignment ▪ Requirements for user support for ontology evolution ▪ Methodology for integrating debugging and aligning into ontology development methodology. ▪ Quality audit of VALCRI ontologies
Summary of progress and overall contribution to the project	In our use of the term ontology evolution we include methods for updating the ontologies and their connections to their ontologies, thereby including ontology debugging, completion and alignment.

	An important aspect that has not received much attention yet is how to help users to debug, complete, align and in general evolve ontologies. In this respect we have identified requirements for user support for ontology alignment and for visualization for ontology evolution in general. We also studied how state-of-the-art ontology engineering systems satisfy these requirements. The requirements will be used in implementations for VALCRI. We have also shown how ontology alignment, completing and debugging can and should be integrated in ontology development methodologies. This can be used in the development of the VALCRI ontologies. Using this methodology and our previously implemented tools we did an audit of the quality of the VALCRI ontologies and suggested possible improvements.

Issues and Deviations from Planning:

Task	Timing	Status	Deviation / issues / Comments
T10.1	M01-M47	COMPLETED	Work has been delayed due to lack of clear system requirements in the initial stages.
T10.2	M01-M47	COMPLETED	Work was delayed for at least 5 months, due to recruitment problems in the initial stages.
T10.3	M01-M47	COMPLETED	Work has been delayed due to lack of clear system requirements in the initial stages.

Required Corrective Action

Lack of clear system requirements: this has been an overall issue in the project, and is not dealt with on WP10 level specifically. However, a requirements management system was put into place and requirements were being entered, to assist with the minor deviation during project phase.

Recruitment issues: At the initial stages of the project a person being recruited for this task was delayed. Although it was difficult to recover time lost, was not disruptive for other activities, hence was not critical for the success completion of the project.

Relevant Publications

Authors	Title	Year	Conference / Journal	Status
Valentina Ivanova	Fostering User Involvement in Ontology Alignment and Alignment Evaluation	2018	PhD Thesis, Linköping University	PUBLISHED
Karl Hammar	Content Ontology Design Patterns: Qualities, Methods and Tools	2017	PhD Thesis, Linköping University	PUBLISHED
Karl Hammar, Valentina Presutti	Template-based Content ODP Instantiation	2017	*Advances in Ontology Design and Patterns,* IOS Press	PUBLISHED
Zlatan Dragisic	Completion of Ontologies and Ontology Networks	2017	PhD Thesis, Linköping University	PUBLISHED
Zlatan Dragisic, Valentina Ivanova, Huanyu Li, Patrick Lambrix	Experiences from the Anatomy track in the Ontology Alignment Evaluation Initiative	2017	Journal of Biomedical Semantics	PUBLISHED

Patrick Lambrix, Rajaram Kaliyaperumal	A Session-based Ontology Alignment Approach enabling User Involvement	2017	Semantic Web Journal	PUBLISHED
Valentina Ivanova, Benjamin Bach, Emmanuel Pietriga, Patrick Lambrix	Alignment Cubes: Towards Interactive Visual Exploration and Evaluation of Multiple Ontology Alignments	2017	Proceedings of the International Semantic Web Conference (ISWC 2017)	PUBLISHED
Valentina Ivanova, Benjamin Bach, Emmanuel Pietriga, Patrick Lambrix	Alignment Cubes: Interactive Visual Exploration and Evaluation of Multiple Ontology Alignments	2017	Proceedings of the ISWC 2017 Posters & Demonstrations Track, CEUR Workshop Proceedings	PUBLISHED
Keskisärkkä, R.	Representing RDF Stream Processing Queries in RSP-SPIN	2016	Proceedings of the ISWC 2016 Posters & Demonstrations Track co-located with the 15th International Semantic Web Conference (ISWC-2016), Kobe, Japan, October 17-21, 2016, CEUR Workshop Proceedings.	PUBLISHED
Keskisärkkä, R.	Query Templates for RDF Stream Processing	2016/10	Proceedings of Stream Reasoning Workshop. 2016 October 17th-18th, 2016, Kobe, Japan. Collocated with the 15th International Semantic Web Conference (ISWC 2016), CEUR Workshop Proceedings, 2016.	PUBLISHED
Valentina Ivanova	Applications of Large Displays: Advancing User Support in Large Scale Ontology Alignment,	2016-10	Proceedings of the Doctoral Consortium at the 15th International Semantic Web Conference (ISWC 2016)	PUBLISHED
Patrick Lambrix, Zlatan Dragisic, Valentina Ivanova, Craig Anslow	Visualization for Ontology Evolution	2016-10	Visualization and Interaction for Ontologies and Linked Data, 2nd International Workshop, co-located with ISWC 2016	PUBLISHED
Zlatan Dragisic, Valentina Ivanova, Patrick Lambrix, Daniel Faria, Ernesto Jiménez-Ruiz, Catia Pesquita	User validation in ontology alignment	2016-10	Proceedings of the International Semantic Web Conference (ISWC 2016)	PUBLISHED
Karl Hammar	Ontology Design Patterns in WebProtégé	2015-12	Proceedings of the ISWC 2015 Posters & Demonstrations Track co-located with the 14th International Semantic Web Conference (ISWC-2015), CEUR Workshop Proceedings	PUBLISHED

Karl Hammar, in Hitzler, P., Gangemi, A., Janowicz, K., Krisnadhi, A., and Presutti, V.,	Quality of Content Ontology Design Patterns	2016	Ontology Engineering with Ontology Design Patterns, IOS Press	PUBLISHED
Eva Blomqvist, Karl Hammar, Valentina Presutti, in Hitzler, P., Gangemi, A., Janowicz, K., Krisnadhi, A., and Presutti, V.	Engineering Ontologies with Patterns – The eXtreme Design Methodology	2016	Ontology Engineering with Ontology Design Patterns, IOS Press	PUBLISHED
Zlatan Dragisic, Patrick Lambrix, Eva Blomqvist	Integrating Ontology Debugging and Matching into the eXtreme Design Methodology	2015	Proceedings of the 6th Workshop on Ontology and Semantic Web Patterns (WOP 2015)	PUBLISHED
Agnese Chiatti, Zlatan Dragisic, Tania Cerquitelli, Patrick Lambrix	Reducing the Search Space in Ontology Alignment Using Clustering Techniques and Topic Identification	2015	Proceedings of the 8th International Conference on Knowledge Capture	PUBLISHED

2.2.11 Work Package 11 - RESEARCH DATA

Work Package Summary

WP 11	RESEARCH DATA
WP Leader	AES
Participants	AES, SPACE, FHG, MU
Status	COMPLETED
Summary of key outputs	1. Seven anonymised data sets delivered to the project; 1.1 million crime reports, 1.4 million nominal records (victims, offenders and suspects), 57,000 burglary free text modus operandi records associated with one of the crime years, 200,000 base intelligence records, 185,000 custody records, 1.2 million incident reporting records and 1.8 million financial transaction records.. 2. Five anonymised data sets have been passed to WMP for examination prior to project release; 57,400 free text MO notes relating to burglary offences, 1.8 million records of property that has been stolen and/or damaged as a result of the crime, 47,000 records containing one or more offender descriptions, 60,000 stop and search records and 200,000 intelligence free text additions to the original data. 3. We have iteratively updated a set of master look up tables that provide anonymised data; Crime numbers, Nominal (people) reference numbers, Crime reference numbers, Forenames, Surnames, Streets, Towns, Districts and vehicle registration numbers. Where appropriate the anonymised data represents bad spellings and close matching e.g.

	Original Anonymised Lesley Ashley Leslie Ashlie Lessley Asshley 4. Processes for validating the anonymised data sets have been devised Implemented algorithms and technologies in three different areas of PET. For detailed information on the concepts and algorithms, see D2.3, D11.14, and SEPL White Paper: (a) selective face encryption in h.264 Videos; (2) re-identification analysis tool to determine the overall and individual risk of person re-identification in data; (3) anonymization and pseudonymization techniques to be used in combination with the re-identification analysis tool
Summary of significant contributions / impact for the project	It is impossible to obtain real Police data containing personal information on which to undertake research and development. Therefore the anonymised data sets provide a realistic base on which to test all development algorithms and components and facilitate relevant research. The anonymisation will be refined to provide a set of Police data that will be released to the general research community within the EU.

Task 11.1	Data Requirements
Task Leader	AES
Participants	SPACE, FHG, MU
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D11.14 DataAnalysisSyntheticData Final.pdf (M16)
Summary of progress and overall contribution to the project	A set of Police data with the right combination of attributes to support VALCRI research and development has been secured, anonymised and validated

Task 11.2	Data Analysis for the Generation of Synthetic Data
Task Leader	AES
Participants	SPACE, FHG
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D11.14 DataAnalysisSyntheticData Final.pdf (M16)
Summary of progress and overall contribution to the project	Our End User partners have been interviewed with a view to securing financial data formats for credit card and bank account investigations. A methodology has been developed, but not yet tested, that will permit the generation of synthetic financial data.

Task 11.3	Data Synthesis
Task Leader	SPACE
Participants	AES
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D11.14 DataAnalysisSyntheticData Final.pdf (M16)
Summary of progress and overall contribution to the project	A software application was developed to generate around 100M of financial transactions over a period of a few months. The tool uses the Actor model to simulate complex interactions between a set of entities (actors) over a period of time. .The tool is configurable through a config file, allowing the user to define different types of actors (people, organisations, utility companies, etc.) In order to obtain an even more convincing synthetic data set, the resulting data was further edited by AES in order to relate the data entries to the entries in the exiting WP11 SPACE has developed a data generator for data similar to Flux24. Flux24 is a data set that federates police reports over all the local police forces in the region of Antwerp. The data generator was developed in order to have a realistic enough dataset to trigger many of the issues we discovered in our software installed in the Federal Police premises. This way, we could debug and test our software before each version deployment in the Federal Police. The dataset features realistic statistics triggering each of the visualizations, and it is big enough to run stress tests. The Flux24 generator has been used for reproducing bugs, testing the visualizations and stress testing the system. Textual data, as big as 700MBytes which represents police reports of a full year, could be generated in few minutes.

Task 11.4	Data Anonymisation
Task Leader	FHG
Participants	AES, FHG
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D11.14 DataAnalysisSyntheticData Final.pdf (M16) **White Papers** ▪ VALCRI-WP-2017-007 Research Data.pdf ▪ VALCRI-WP-2017-004 Security PET.pdf

Summary of progress and overall contribution to the project	1. Seven anonymised data sets delivered to the project; 1.1 million crime reports, 1.4 million nominal records (victims, offenders and suspects), 57,000 burglary free text modus operandi records associated with one of the crime years, 200,000 base intelligence records, 185, 000 custody records, 1.2 million incident reporting records and 1.8 million financial transaction records. 2. Five anonymised data sets have been passed to WMP for examination prior to project release; 57,400 free text MO notes relating to burglary offences, 1.8 million records of property that has been stolen and/or damaged as a result of the crime, 47,000 records containing one or more offender descriptions, 60,000 stop and search records and 200,000 intelligence free text additions to the original data. 3. We have iteratively updated a set of master look up tables that provide anonymised data; Crime numbers, Nominal (people) reference numbers, Crime reference numbers, Forenames, Surnames, Streets, Towns, Districts and vehicle registration numbers. Where appropriate the anonymised data represents bad spellings and close matching e.g. Original Anonymised Lesley Ashley Leslie Ashlie Lessley Asshley 4. FHG: Implemented algorithms and technologies in three different areas of PET. For detailed information on the concepts and algorithms, see D2.3, D11.14, and SEPL White Paper: - selective video face encryption components (face detection and subsequent selective encryption of h.264 videos; C/C++) - Re-identification analysis components (statistical analysis of person-related data sets to determine the overall and individual re-identification risk, and to provide parameters for respective anonymization/pseudonymization; JavaScript, in order to allow client execution before access to the db is gained) - anonymization and pseudonymization components (based on filtering, permutation, generalization; JavaScript, in order to allow client execution before access to the db is gained)
Relevant Publications	Anonymisation of data White Paper, see VALCRI-WP-2017-007 Research Data SEPL White Paper, see 2016-12-22-WhitePaper_SEPL_TECH.pdf Selective Face Encryption in H.264 Encoded Videos: Gerhardt, C.; Aichroth, P. & Mann, S. Selective Face Encryption in H.264 Encoded Videos. In IEEE Visual Communications and Image Processing, 2017.

Task 11.5	Data Validation
Task Leader	AES
Participants	SPACE, FHG, MU
Status	COMPLETED

Output and deliverables	**Deliverables** ▪ D11.15 SyntheticDataCreation Final.pdf (M16)
Summary of progress and overall contribution to the project	All data sets that have been released have been validated through the training and evaluation processes. As the data is anonymised real Police data there will be naturally occurring links between data sets. These have been tested. Fictitious case data have been inserted into the data sets with known links in order to test components and algorithms. These were deleted prior to final release. SPACE has participated and had an important role in the de-anonymisation activity that was a part of this task. Using different techniques, SPACE found the correct year of the anonymised MOSAIC dataset and found examples in the data that can be related to open source information. For example, SPACE revealed the real name of a specific offender in the UK for whom specific confidential insights from the MOSAIC data set could be seen. As a result of this work, the anonymization of the data was improved – and part of the data was omitted to avoid the risk of exposing confidential data.

Task 11.6	Content Aggregation Storage Components
Task Leader	FHG
Participants	AES, SPACE, FHG
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ VALCRI D2.3 System Design v2 Final.pdf (M17) ▪ D11.16 UnstructeredDataStorageAndAggregation Final.pdf (M16)
Summary of progress and overall contribution to the project	The main achievements of this task are the implementation of the "Unstructered Data Storage" (UDB) service and the implementation of a CLI-application for content ingestion. The UDB provides the following key functionalities: ▪ REST-API for CRUD operations for binary data using MongoDB's GridFS as persistence layer ▪ Metadata storage and retrieval (using GridFS's metadata fields or a custom relationship mechanism) ▪ **Security: Fully integrated OpenPMF-support for policy-based access control** ▪ **Privacy: Integration of the "Face Encryption"-component for video files** ▪ Basic search capabilities ▪ Support for media fragment delivery (ffmpeg-based) for server-side transcoding and partial (spatial/temporal) delivery of video files The CLI-application has been actively used for ingestion of video files and associated event annotations into the storage sub-system (see D2.3 v2 for related work flows). Both UDB and the CLI-application has been used for **realizing EPIC-042: A/V Event Analysis** (see Task 5.7 for further details).

Issues and Deviations from Planning:

Task	Timing	Status	Deviation / issues / Comments
T11.1	M01-M47	COMPLETED	No deviations
T11.2	M01-M47	COMPLETED	This task consumed fewer person months than allocated. This is due to our Policing partners providing a large amount of real data for the project.
T11.3	M01-M47	COMPLETED	This task consumed fewer person months than allocated. This is due to our Policing partners providing a large amount of real data for the project.
T11.4	M01-M47	COMPLETED	This task consumed far greater number of person months than was allocated. This is due to our Policing partners providing a large amount of real Police data for the project. AES were the only partner authorised to anonymise the WMP Police data and SPACE the Belgium Police data.
T11.5	M01-M47	COMPLETED	No deviations from the planning; The work in this task continued for many months beyond M16. This task required the output from tasks 11.3 and 11.4.
T11.6	M01-M47	COMPLETED	No deviations

Required Corrective Action

T11.4 will consume a greater number of person months by AES as they are solely responsible for anonymising the WMP Police data. Unused person months from the other tasks will be consumed in T11.4.

Relevant Publications

Authors	Titles	Year	Conference / Journal	Status

2.2.12 Work Package 12 - SCALABLE & SECURE DISTRIBUTED PROCESSING & INTEGRATION

Work Package Summary

WP 12	SCALABLE & SECURE DISTRIBUTED PROCESSING & INTEGRATION
WP Leader	OS
Participants	MU, SPACE, UKON, LIU, CITY, AES, FHG, OS
Status	COMPLETED
Summary of key outputs	The work package has the following key outputs: ▪ Development of an integration architecture ▪ Implementation of a complete build framework ▪ Development of a first security, data protection and privacy Domain Specific Language ▪ Development and implementation of concepts for advanced access control, in order to define and enforce security policies driven by operational

	requirements and regulations ▪ A web based graphical user interface for high level policy and system definition, policy and attribute refinement, generation of low level policies in various formats and human readable documentation, ▪ Multi user support for the VALCRI system, including login and authentication ▪ Secure unstructured storage (UDB) ▪ Logging and auditing ▪ High Assurance Logging and Auditing (HALA) with separated domain for storage of logging information (requires additional hardware ▪ First prototype of High Assurance Logging and Auditing (SDB)
Summary of significant contributions / impact for the project	The outcomes enable the developers of the VALCRI system: ▪ build their individual components in a uniform way ▪ assemble them to an integrated system The outcomes enable the security administrator of the VALCRI system: ▪ to define high level security and data protection policies ▪ to generate low level enforceable security configurations ▪ to generate human readable descriptions of the high and low level security policies e.g. for security auditing purposes ▪ to enforce simple, initial security policies for the SDB ▪ to log relevant events both to a standard logging system and to a protected storage not accessible to the system administer (High Assurance Logging and Auditing) ▪ to analyse logged events, e.g. for security incidents The outcomes enable the data protection auditor: ▪ to analyse and assess the human readable policy representations for compliance ▪ to analyse and audit log files ▪ HALA ensures that log files cannot be modified by the organisation's system administrator The outcomes will allow the police organisations and individual users to ▪ protect their data based on fine grained and high level operational policies ▪ detect misuse The outcome will enable the data protection authorities: ▪ To define and enforce high level data protection and privacy polices ▪ To audit that the system is used in accordance with the data protection regulations ▪ To detect misuse by the organisation running the system ▪ ▪ FHG: For integration aspects related to SEPL + UIUX + DMO integration, relevant flows and component communication and component overviews were specified.
Further information	OS got an US patent granted which includes considerable contributions from VALCRI research

Task 12.1	Definition of Requirements for S2DP
Task Leader	OS
Participants	MU, SPACE, UKON, LIU, CITY, AES, FHG, OS
Status	COMPLETED
Output and deliverables	**White Papers** ▪ VALCRI-WP-2017-004 Security PET.pdf Semantic Wiki content
Summary of progress and overall contribution to the project	FHG, OS: Updates to SEPL requirement in the SMW
Relevant Publications	SEPL white paper, see 2016-12-22-WhitePaper_SEPL_TECH.

Task 12.2	Joint Data Model and Interface Definition
Task Leader	OS
Participants	OS, LIU, FHG
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D12.17 JDMInterfaceDefinition Final.pdf (M16)
Summary of progress and overall contribution to the project	▪ OS, LIU, FHG: Started the definition of model elements required for SEPL ▪ OS, FHG: Interface definitions for SEPL components

Task 12.3	S2DP Tool Chain and Runtime Design and Implementation
Task Leader	OS
Participants	OS, FHG, SPACE, LIU
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D12.18 SystemIntegration Final.pdf (M16)
Summary of progress and overall contribution to the project	▪ SPACE, FHG, OS: Development of an overall component integration approach ▪ ALL: setup of local dev environment; participation in regular integration PhC ▪ FHG: integration of unstructured storage component (prototype 1) ▪ OS, FHG: Draft of a Domain Specific Language for Privacy and Data Protection (with contributions from ULD) ▪ OS: Further development/improvement of OpenPMF policy modelling UI ▪ UI web interface

	▪ Policy/system modelling support ▪ Model transformations for policy and attribute refinement ▪ Wildcard rule filling ▪ OpenPMF policy modelling UI wizard infrastructure ▪ UI wizards for importing XACML and TCPdump ▪ UI multi user support and authentication ▪ XACML support (partially funded by VALCRI) ▪ Policy natural language text editor (partially funded by VALCRI) ▪ Identity importer ▪ Various exporters ▪ OS: OpenPMF policy enforcement runtime infrastructure ▪ Runtime policy transformers scripting capability ▪ Runtime value caching support ▪ Add MAC and IP to the PEP API ▪ Java-servlet runtime adapter ▪ Improvement /update of ActiveMQ PEP (developed in a previous project) ▪ Fixes of many issues ▪ Access control for Postgresql database (SDB) ▪ Logging and auditing ▪ Graylog based logging (including graylog docker image) ▪ Implementation of High Assurance Logging and Auditing ▪ Reading events from host syslog ▪ PCIe communication between host and sensor(VC707) ▪ Separation of sensor and storage station, connected over protected fiber optics link ▪ On the storage station (ZC706 running Linux), getting the events from the fiber optics link and storing it locally using syslog ▪ OS: VALCRI system integration ▪ Multi user support ▪ dockerised multi-user VALCRI branch ▪ migrate/re-implement multi-user support from old development to new development ▪ User log in and authentication ▪ securing VALCRI pipe-line top to bottom (TLS + CAS-based authentication of pipeline components) ▪ VALCRI deployer script (for SpaceApps and others) ▪ OpenPMF/SDB/Security ▪ FHG, OS: adaptation/preparation of UDB storage service for OpenPMF ▪ OS: Support to other partners: ▪ Docker toolbox ▪ Support to others to get them working with virtual pipeline ▪ OS: Other ▪ Fixes in Ecore to XSD generator ▪ Got US patent which included considerable contributions from VALCRI granted

Task 12.4	VALCRI System Integration and Validation
Task Leader	SPACE
Participants	

Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D11.16 UnstructeredDataStorageAndAggregation Final.pdf (M16) ▪ D12.17 JDMInterfaceDefinition Final.pdf (M16) ▪ D12.18 SystemIntegration Final.pdf (M16)
Summary of progress and overall contribution to the project	▪ SPACE: Development of a complete build system for the consortium during its initial stages. ▪ MU: A second build system was later developed to support the changing needs of the VALCRI system. ▪ FHG: coordination of SG-SEPL activities in regular, biweekly PhC, and in several SEPL workshops (two SEPL-only workshops in Berlin); compilation of SEPL Tech White Paper (FHG, OS, ULD); collection of SEPL docs in OwnCloud; communication with IEB ▪ FHG: The storage component for unstructured data has been updated in order to support the OpenPMF/CAS-based authentication and authorization approach. In order to support integrated testing of the storage component, relevant parts have been made available as Docker containers (storage-mongodb, storage-service), as well as docker images. Moreover, the client for using the service which is available in the Valcri Nexus (de.fraunhofer.idmt.commons: file-storage-client:1.4.4) has been updated accordingly. The "unstructerd data database" (aka UDB) is part of the ongoing vertical integration process involving video player components (UI) and SEPL components (OpenPMF, CAS). The components are described in detail in the updated version of Deliverable D11.16. ▪ SPACE: Setup of a Docker repository for the consortium ▪ SPACE: Integration of docker into the build system in order to automate building docker images. ▪ SPACE: mprove windows compatibility for the build system. ▪ SPACE: Preparing the deployment of the VALCRI system and exposing it to the partners through VPN. During M33 and M45 in WP12, SPACE has focused primarily on runtime re-configurability of the ingester. SPACE has implemented a simple DSL in order to provide enough flexibility to cover a multitude of data ingestion use cases. The DSL exposes 17 components that are configurable and inter-connectable in order to form a one-off or real-time data processing pipeline. These components cover, amongst other things, the following topics: ▪ ingesting specific data formats (relational databases, CSV, JSON, Word etc.) ▪ manipulation of the incoming data: joining related documents, splitting or joining fields and field formatting ▪ image to text conversion ▪ scheduling operations ▪ file systems crawling ▪ bulk data processing capabilities ▪ incremental backup

	▪ geocoding In parallel to our work on the ingester, we continued our support and overall maintenance of the VALCRI server that hosts most nexus artefacts. Summary: Most of the partner components were integrated into the VALCRI system by MU lead at TRL 5-6. VALCRI prototype validation has been carried out which has been described in WP13. Two components of the VALCRI system were deployed at WMP by MU and CITY which accepts real data. While components which were at lower levels of TRL were left as standalone to demonstrated early research concepts. While components at developed at higher TRL7-9 by SPACE were not integrated to the VALCRI system but deployed as a separate system at LPA.

Issues and Deviations from Planning:

Task	Timing	Status	Deviation / issues / Comments
T12.1	M01-M47	COMPLETED	No deviation
T12.2	M01-M47	COMPLETED	This task (and the related deliverable) lost most of its importance, because the system will not be developed as a Service Oriented Architecture, but as an information centric system with a database as central means of integration.
T12.3	M01-M47	COMPLETED	Based on the requirements, we put a much stronger focus than planned on logging and auditing, especially its assurance.
T12.4	M01-M47	COMPLETED	There was a deviation from the DOW. This is due to VALCRI components being at different TRL levels. MU lead the main system integration work for the VALCRI system. Due to this did not affect the outcomes.

Required Corrective Action

None.

Relevant Publications

Authors	Titles	Year	Conference / Journal	Status

2.2.13 Work Package 13 - TRAINING AND EVALUATION

Work Package Summary

WP 13	TRAINING AND EVALUATION
WP Leader	IINT

Participants	MU, TUG, AES, SPACE, LIU, XI, WMP, LPA, BFP
Status	COMPLETED
Summary of key outputs	WP13 generated the following results over the course of the project: A Zero Point Measurement that generated insights into the tools, workflows and cognitive strategies used by analysts to complete routine tasks. The results of this exercise have been used to support work on WP3 (e.g. with regard to sensemaking and bias mitigation) and WP4 (e.g. with regard to the management of uncertainty and analytic provenance). A detailed syllabus to support the training and development of criminal intelligence professionals in Europe. A formal program for the training and development of UK police analysts to be offered via a UK university, developed in collaboration with one of the project's end users. A structured approach to evaluating the VALCRI system and its components over the duration of the project. This approach was amended as necessary to accommodate new metrics and a continuously expanding set of VALCRI features. Three formal evaluations of the VALCRI system. These recorded a consistent improvement in end user feedback on the VALCRI system. The evaluations also identified opportunities for the continued development of the VALCRI system. Eight end user workshops. These were used to a) evaluate the research conducted by WP3 on sensemaking, evidential reasoning and bias mitigation in intelligence analysis; b) validate the training modules defined in the VALCRI Syllabus; and c) showcase the VALCRI System to a wider End User community.
Summary of significant contributions / impact for the project	WP13 has made the following contributions to the VALCRI project: ▪ Through the development of the VALCRI Syllabus, WP13 has contributed to an evolving debate on the training and development of intelligence analysts. ▪ Through the End User workshops, WP13 has introduced the VALCRI project to a wider network of potential end users, particularly in the fields of law enforcement and national security. Through our formal and informal evaluations, WP13 has contributed to the refinement and development of the VALCRI System and its components.

Task 13.1	Zero Point Measurement (ZPM)
Task Leader	IINT
Participants	AES, SPACE, MU, TUGraz
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D3.4 HumanIssuesFramework v2 Final.pdf (M16) ▪ VALCRI D13.19 Final.pdf (M18)

	Technical Notes ▪ WP13_TN13.1_Zero-Point-Measurement_V1.0.doc ▪ WP13_TN13.5_Evaluating_VALCRI_Methodology_and_End_User_Feedback.doc The ZPM generated the following outputs: ▪ Four close observation exercises / evaluations of criminal intelligence analysts at work in Birmingham and Antwerp. ▪ Structured interviews with six criminal intelligence professionals by both WP3 and WP13 partners. These interviews were recorded for future reference if needed. ▪ The results of the ZPM informed the development of the evaluation metrics used in WP13 for Technical Note TN13.5 The results of the ZPM can be seen in Technical Note TN13.1, Deliverable D13.19. Further, after analysing the physiological monitoring of LPA and BFP personnel, AES wrote a report on the identified stress levels and delivered it to both parties.
Summary of progress and overall contribution to the project	The ZPM provided an invaluable opportunity to observe and engage end users in their place of work. Consortium members could examine how analysts cycle through phases of uncertainty and understanding, and how they engage in sensemaking and evidential reasoning under routine operating conditions. The results of TN13.1 informed i-intelligence's contribution to WP3, WP4 and WP13, specifically our research on the human issues shaping criminal intelligence analysis.

Task 13.2	Training Concept and Pedagogic
Task Leader	TUGraz
Participants	AES
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ VALCRI D13.19 Final.pdf (M18) **Technical Notes** ▪ WP13_TN13.2_Training-Pedagogical-Concept-V1.0.doc **White Papers** ▪ VALCRI-WP-2017-008 Training.pdf Task 13.2 generated the following outputs: ▪ A Technical Note TN13.2 on the state-of-the-art in adult education theory, detailing relevant instructional design modules, psycho-pedagogical principles for learning, and recommended approaches to the training and development of criminal intelligence professionals. ▪ A psycho-pedagogic training module detailing *what* to learn, *how* to learn it, and *who* should support the learning process. ▪ Our work on the Pedagogic Concept can be found in Technical Note TN13.2, as well as the Deliverable D13.19.

Summary of progress and overall contribution to the project	Technical Note TN13.2 elaborates a psycho-pedagogical model that provides a basis for the professional development of criminal intelligence analysts. The model detailed a number of adult learning theories that was used to inform: ▪ The various iterations of the VALCRI Training Syllabus ▪ Subsequent training modules developed over the course of WP13 ▪ Where applicable, the development of the VALCRI components

Task 13.3	Training Syllabus/ Module Outline
Task Leader	IINT
Participants	AES
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D13.19 Final.pdf (M18) **Technical Notes** ▪ WP13_TN13.4_End-User Workshops.doc ▪ WP13_TN13.3_Training_Syllabus_V1.0.doc ▪ WP13_TN13.3_Training_Syllabus_V2.0.doc ▪ WP13_TN13.19_VALCRI_Traning_Syllabus_Module_Outlines.doc **White Papers** ▪ VALCRI-WP-2017-008 Training.pdf **Main Conference Presentations** Title: "Recasting the Intelligence Curriculum: Lessons from Two EU Projects". Event: CEPOL Science and Technology Conference, Budapest, Hungry, 29 November 2017 Title: "The Intelligence Analyst as Internal Consultant". Event: The Australian Institute of Professional Intelligence Officers (AIPIO) Intelligence Conference, Hobart, Australia, 22-24 August 2017 Title: "Recasting the Intelligence Curriculum: Lessons from Two EU Projects". Event: The Second Annual Conference of IAFIE's EUROPE, Athens, Greece, 22-24 June 2017 Roundtable Discussion. Event: The Five Eyes Analytic Training Workshop, Harrison, VA, USA, 29 February - 2 March 2016 Title: "A Holistic Training Program for Law Enforcement Analysts". Event: International Crime and Intelligence Analysis Conference, Manchester, United Kingdom, 25-26 February 2016 **Published Conference Papers** Bielska, Aleksandra, and Pallaris, Chris. "Redefining the Intelligence Skill Set by Use of the Intelligence Analysis Impact Model". Paper presented at the International Studies Association Annual Convention, Baltimore, MD, United States, February 2017. Bielska, Aleksandra, and Pallaris, Chris. "Toward an Understanding of Uncertainty in Intelligence Analysis". Paper presented at the International Studies Association Annual Convention, San Francisco, CA, United States, April 2018. **Other Publications** Bielska, Aleksandra, and Pallaris, Chris. "The Psychology of Intelligence Analysis: Where Are We and Where Should We Be?". White paper submitted to the National Academies of Sciences, Engineering, Medicine, July 2017.

	Bielska, Aleksandra, and Pallaris, Chris. "Addressing the Internal Challenges to Intelligence Work". Journal of Mediterranean and Balkan Intelligence, Vol. 10, No. 2 (2017), pp. 89-102. The VALCRI Syllabus Version 1 of the VALCRI Syllabus was published as VALCRI D13.19. It was elaborated further in TN13.3_V1.0. In addition to the above we developed a list of modules for feedback and review which we labelled TN13.19. This was circulated to security and law enforcement professionals for their feedback and review. TN13.19 served as a working document which was updated per end user feedback, including from the VALCRI workshops held between Autumn 2016 and Winter 2017. The final list of recommended modules was published in the Technical Note TN13.19. This provides: ▪ An index of recommended modules for the training and development of law enforcement analysts ▪ An index of modules for the training analysts on the use of the VALCRI system The final version of the full Training Syllabus was published in the Technical Note TN13.3 V2. **Training Courses** Further, together with WMP, AES created a training syllabus for Police intelligence analysts. This has been ratified by the UK National Analysts Working Group and accepted by Aston University as part of its Distance Learning Apprenticeships. The modules comprise: Advanced Intelligence Professionalisation: 1. Mining Police Data - Theory and Practice 2. Critical Thinking 3. Risk Management / Prioritisation 4. Statistics Essentials 5. Data Mining Tools and Algorithms 6. Profiling Communities and Neighbourhoods 7. Criminal Behaviour 8. Linking Crime and People Networks 9. Crime and Criminal Profiling 10. Open Source Intelligence 11. Cyber Criminality 12. Ethics and Data Protection 13. Influencing and Negotiation 14. Case Study Exercises Finally, i-intelligence has used the Syllabus to define a series of professional development programs for security and law enforcement analysts. These programs will be tested, evaluated and launched in the second half of 2018.
Summary of progress and overall contribution to the project	The VALCRI Syllabus reflects the evolution of our research on the training of intelligence analysts and criminal intelligence professionals. The document represents the current state-of-the-art, as well as recommended approaches put forward by academics and professionals working in the field. Indeed, over the course of the project, WP13 staff attended several academic

	and practitioner forums in Europe and the US to present the syllabus and soliciting input from a wider End-User audience. Their inputs informed our work on the Human Issues Framework in WP3, as well as the development and delivery of the End User Workshops discussed in TN13.4. Finally, the Syllabus has informed the exploitation activities of AES and i-intelligence, with both organisations developing educational programs to support the training and development of law enforcement analysts.

Task 13.4	End User Training
Task Leader	IINT
Participants	AES, MU, SPACE, TUGraz
Status	COMPLETED
Output and deliverables	**Technical Notes** ▪ WP13_TN13.4_End-User Workshops.doc **White Papers** ▪ VALCRI-WP-2017-008 Training.pdf ▪ VALCRI-WP-2018-016 The VALCRI Workshops Key Findings.pdf ▪ WP13 organised eight end user workshops under the auspices of WP13 and WP14 (three more than originally planned). These workshops were held between Fall 2016 and Winter 2017 at the following venues: ▪ The Scenario Planning and Strategic Foresight Workshop held in London, UK, from 31 October to 2 November 2016 ▪ The Strategic Foresight Workshop held at West Midlands Police in Birmingham, UK, from 6 to 9 November 2016 ▪ The Business Process Analysis Workshop held at West Midlands Police in Birmingham, UK, from 6 to 8 March 2017 ▪ The Analytic Techniques Workshop that took place in Dublin, Ireland, on 10 March 2017 ▪ The Strategic Foresight Workshop that took place in Lisbon, Portugal, from 13 to 15 March 2017 ▪ The Open Source Intelligence and Strategic Foresight Workshop that took place in Vilnius, Lithuania, form 19 to 20 September 2017 ▪ The Open Source Intelligence Workshop that took place in Los Angeles, CA, from 25 to 27 September 2017 ▪ The Strategic Foresight Workshop that took place in Paris, France, from 23 to 24 November 2017 The workshops gathered 123 participants from 40 different law enforcement and security agencies, including representatives from 13 countries excluding European and international law enforcement agencies: ▪ Belgium ▪ Czech Republic ▪ Denmark ▪ France ▪ Ireland

	▪ Lithuania ▪ Norway ▪ Portugal ▪ Romania ▪ Spain ▪ Switzerland ▪ United Kingdom ▪ United States The results of these workshops, and details of the attending agencies, are captured in Technical Note TN13.4 and the attendant white paper "The VALCRI Workshops: Key Findings on the VALCRI System and the Training of Law Enforcement Analysts".
Summary of progress and overall contribution to the project	The VALCRI workshops succeeded in: ▪ Showcasing the VALCRI system and soliciting feedback for its improvement ▪ Evaluating selected modules from the VALCRI Syllabus. Emphasis here was given to those skills that support strategic and operational intelligence analysis, process analysis and improvement, and open source intelligence ▪ Testing and validating the research conducted by WP3 on sensemaking, evidential reasoning and bias mitigation ▪ Identifying the training needs of law enforcement analysts using structured surveys ▪ Identifying opportunities for the exploitation of the VALCRI project The workshops confirmed that end Users need better tools to support the collection, analysis and processing of intelligence data. Such tools can be technical in nature (e.g. VALCRI) or cognitive (e.g. the many structured analytical techniques that can be used to measure and evaluate the quality of one's thinking). Either way, WP13 has demonstrated the importance of taking a holistic approach to law enforcement capability development. The workshops were well received by participants, with feedback commenting on the value of structured methodologies to support critical and creative thinking, insight generation, and the search for novel policy prescriptions to today's policing challenges. The workshops also provided invaluable feedback on the VALCRI system. This feedback confirmed the potential of the VALCRI system to support strategic, tactical and criminal intelligence analysis, as well as to improve sensemaking, collaboration and intelligence reporting. A statistical assessment of end-user feedback is also provided in Technical Note TN13.4.

Task 13.5	Prototype Testing and Evaluation
Task Leader	IINT
Participants	AES, SPACE, TUGraz
Status	COMPLETED
Output and deliverables	**Deliverables**

	▪ D13.20_V2.0_Final.pdf (M18) **Technical Notes** ▪ WP13_TN13.5_Evaluating_VALCRI_Methodology_and_End_User_Feedback.doc **White Papers** ▪ VALCRI-WP-2017-015 Evaluating the VALCRI System.pdf ▪ VALCRI-WP-2018-016 The VALCRI Workshops Key Findings.pdf WP13 conducted three formal scientific evaluations of the VALCRI system: 1. Krakow - June 2015 2. Berlin - May 2017 3. Nice – November 2017 These evaluations sought to: ▪ Determine whether and how VALCRI enhances analytic insights ▪ Determine whether and how VALCRI supports imagination, insight, fluidity and rigour, and transparency ▪ Benchmark end users' performance with successive iterations of VALCRI system ▪ Understand how law enforcement analysts think and reason during routine analytic tasks ▪ Identify areas for further research and development The results of the first evaluation were captured in Deliverable D13.20. The results of the second and third evaluations were captured in Technical Note TN13.5 and the attendant white paper, "Evaluating the VALCRI System: A Summary of End User Feedback and Recommendations". In addition to the formal evaluations listed above, the VALCRI system has been subject to a series of informal evaluations at the various consortium meetings, at the end user workshops detailed in Task 13.4, and at various dissemination activities held under the auspices of WP14. The feedback generated from all of these testing and evaluation activities were shared with the project's developers to support the evolution of the VALCRI system and its various components.
Summary of progress and overall contribution to the project	The end users from West Midland's Police and the federal and local police forces in Antwerp gave VALCRI a positive assessment. Indeed, there was a marked improvement in the feedback given to evaluators over the course of the project. End users commended VALCRI for its multipurpose utility and its ability to support descriptive, prescriptive, and predictive analysis. They also noted its ability to support sensemaking, collaboration and intelligence reporting. End users also spoke positively of VALCRI's interface and its look and feel, noting that the workstation and the system's many visualisation options provided a dynamic "canvas" with which to work. End users did express some reservations on the "complexity" of the VALCRI system. However, they added that their assessment would likely change with more training and greater familiarity with the interface. Overall, end users agreed that: ▪ VALCRI would enable them to accomplish tasks more quickly ▪ Using VALCRI would improve their job performance ▪ Using VALCRI in their job would increase their productivity

	▪ Using VALCRI would enhance their effectiveness on the job ▪ Using VALCRI would make it easier for them to do their job ▪ VALCRI would benefit them in their job

Issues and Deviations from Planning:

Task	Timing	Status	Deviation / issues / Comments
T13.1	M01-M47	COMPLETED	No deviations from the official project plan. The task was initiated and completed as planned.
T13.2	M01-M47	COMPLETED	No deviations from the official project plan. The task was initiated and completed as planned.
T13.3	M01-M47	COMPLETED	No deviations from the official project plan. The task was initiated and completed as planned.
T13.4	M01-M47	COMPLETED	No deviations from the official project plan. The task was initiated and completed as planned.
T13.5	M01-M47	COMPLETED	No deviations from the official project plan. The task was initiated and completed as planned.

Required Corrective Action

None.

Relevant Publications

Authors	Titles	Year	Conference / Journal	Status

2.2.14 Work Package 14 - DISSEMINATION AND EXPLOITATION

Work Package Summary

WP 14	DISSEMINATION AND EXPLOITATION
WP Leader	TUW
Participants	MU, SAS, UKON, LIU, CITY, KUL, AES, TUGraz, FHG, TUW, OS, ULD, INT, XI, WMP, LPA
Status	COMPLETED
Summary of key outputs	**Deliverables** ▪ D14.22.pdf (M18) - Exploitation Strategy Report **Technical Note** ▪ VALCRI_Exploitation.doc **IPs developed during the VALCRI project** ▪ VALCRI IP list - Consortium.pdf

	Public Website ▪ www.valcri.org **LEA dissemination & awareness raising** ▪ Contacts with 214 law enforcement officers in 50 agencies in 16 countries excluding European and international law enforcement agencies
Summary of significant contributions / impact for the project	▪ Dissemination ▪ Awareness Raising ▪ Exploitation ▪ Contribution to Standards

Task 14.1	Academic Dissemination
Task Leader	TUW
Participants	MU, SAS, UKON, LIU, CITY, KUL, AES, TUGraz, FHG, TUW, OS, ULD, INT, XI, WMP, LPA
Status	COMPLETED
Output and deliverables	Publications and presentations at conferences
Summary of progress and overall contribution to the project	A total of 119 peer reviewed papers were written. (please see Annex B) 64 conference papers 22 Journal 6 book papers 8 workshop papers 19 White papers

Task 14.2	Dissemination to Target Audience
Task Leader	AES
Participants	MU, SAS, UKON, LIU, CITY, KUL, AES, TUGraz, FHG, TUW, OS, ULD, INT, XI, WMP, LPA
Status	COMPLETED
Output and deliverables	
Summary of progress and overall contribution to the project	Dissemination to the target audiences comprises activities to include, publicising and discussing the VALCRI system: VALCRI demonstrations and evaluations to potential LEA: 214 intel analysts in 50 LEA agencies in 16 Countries excluding European and International law enforcement agencies.

	A subset of the above events was used to trail the VALCRI Training courses and do evaluations: 123 intel analysts in 40 LEA agencies in 13 countries. VALCRI demoed at 5 academic conferences, 1 EU project and 5 intel events: 11 events in 6 countries, demonstrating VALCRI to an estimated 1500 persons.

Task 14.3	Project Website
Task Leader	SPACE
Participants	MU, SAS, UKON, LIU, CITY, KUL, AES, TUGraz, FHG, TUW, OS, ULD, INT, XI, WMP, LPA
Status	COMPLETED
Output and deliverables	www.valcri.org
Summary of progress and overall contribution to the project	The project website was put in place at the beginning of the project. It is continuously adapted, and new developments in the project are announced on the website.

Task 14.4	Raise awareness of privacy, ethical and legal issues
Task Leader	KUL
Participants	MU, SAS, UKON, LIU, CITY, KUL, AES, TUGraz, FHG, TUW, OS, ULD, INT, XI, WMP, LPA
Status	COMPLETED
Output and deliverables	**Technical Notes** ▪ TN3.4.E1_20150826_TechnicalReport_Ethical_MU.doc ▪ TN3.4.L1_20150826_TechnicalReport_Legal_KUL.doc ▪ TN3.4.P1_20150826_TechnicalReport_PrivacyULD.doc **White Papers** ▪ VALCRI-WP-2017-005 Transparency.pdf ▪ VALCRI-WP-2017-014 Understanding the ethical dilemmas.pdf ▪ VALCRI-WP-2017-012 Roadmap for the Operationalization.pdf ▪ VALCRI-WP-2018-017 Lessons Learnt and Impact.pdf Working extensively with our End User partners, their legal & ethical teams and the IEB to ensure that any concerns are raised and suitably answered. Present research findings to the consortium through deliverables, workshops and presentations at group meetings. Liaise with the ethics group of the Society of Data Miners to obtain a wider awareness of issues and responsibilities. Take inspiration from the VALCRI project as part of the educational and teaching activities by the partners involved. Publish findings in journal or conference papers and participate in workshops or panel discussions.

Summary of progress and overall contribution to the project	The majority of the T14.4 dissemination activities are described in detail under TN3.4 (E1,L1,P1). To avoid redundancy, this section will not repeat all of the white papers, reports and published articles. Aside from the above, the privacy, ethics and legal group within VALCRI has disseminated the knowledge gained in several ways. Workshops, such as those on the adoption of Directive 2016/680 and its impact on policing at KUL and others on the ethics of developing advanced analytical systems at the EISIC Conference, have been held during the project. The published papers and VALCRI's research findings have repeatedly been presented and discussed at a variety of different conferences, including EISIC and NGCP. Panel discussions on the legal and ethical side of advanced analytical systems such as VALCRI have been organized in cooperation with other EU H2020 projects at conferences such as CPDP. A video interview was conducted on the SEPL progress during the November 2017 consortium meeting in Nice, France. Expert discussions and debates have been hosted in close collaboration with similar European initiatives on cybersecurity, data protection and research in the area of law enforcement, as illustrated by the Dublin Experts Workshop of 4/7/2017 held by TITANIUM, ASGARD, DANTE, VALCRI and VOX-POL. The findings from the research conducted VALCRI have therefore been communicated to technical partners within the project through numerous Deliverables and presentations at group and consortium meetings, as well as to external groups and experts to raise awareness and share knowledge on legislative developments and measures to approach system development in the area in a legally compliant and ethical manner. Furthermore, the research findings obtained through the VALCRI project have also been indirectly disseminated to students at the law faculty of KU Leuven University by building upon the research done on data protection, privacy, transparency, cybersecurity and others by professors and teaching staff involved in the project and teaching. This contributed to the raising of awareness on highly contemporary debates on privacy and security or transparency and accountability. In similar fashion, the ethics components of the VALCRI research have been disseminated under the topic "Technologies for Law Enforcement" over the course of the three years of the project to students (100+) in the department of computer science partaking in the ethics and professionalism module at Middlesex University. In addition to the above, a series of blog entries at the KU Leuven CiTiP Blog were dedicated to VALCRI throughout the project: ▪ "The Europol Regulation and Purpose Limitation: From the 'Silo-Based Approach' to… What Exactly?", part 1-2, KU Leuven CiTiP Blog, 18-20 April 2017. ▪ "Data analytics in a police context – addressing legal issues in VALCRI", KU Leuven CiTiP Blog, 14 March 2017. ▪ "Countering Algorithmic Discrimination in Profiling", KU Leuven CiTiP Blog, 1 June 2017. Finally, it is worth repeating the research done for the drafting of the VALCRI Data Protection Impact Assessment Methodology. This is explained in detail in the T3.4 section of this document. While the research supporting this document was conducted under WP3.4, the methodology and international comparison

	of approaches and DPIA standards is completed under WP14. The methodology and comparative analysis can be disseminated separately or jointly to VALCRI end-users, other law enforcement agencies and the general public.

Task 14.5	Contribution to standards
Task Leader	SPACE
Participants	
Status	COMPLETED
Output and deliverables	Amending a standard through the different standardization bodies is a long and unpredictable process. Therefore, planning this activity is a challenging task. Adding to this concern is the fact that it takes much time and effort while it only represents a very small portion of the activities in Work Package 14. The most important standard relates to the Topic Maps standard. Topic Maps is an international industry standard (ISO/IEC 13250:2003) for knowledge representation and information integration. It provides the ability to store, together with the data, complex meta-data that represents the semantics. During M33 to M45, SPACE has approached different players (including big players) in the domain of security in Belgium and beyond seeking to cooperate in business for the police in Belgium.
Summary of progress and overall contribution to the project	Development of standards

Task 14.6	Exploitation
Task Leader	FHG
Participants	MU, SAS, UKON, LIU, CITY, KUL, AES, TUGraz, FHG, TUW, OS, ULD, INT, XI, WMP, LPA
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D14.22.pdf (M18) **Technical Notes** ▪ VALCRI_ExploitationPrinciples_25Nov2017.doc **White Papers** ▪ VALCRI-WP-2017-013 Commercial Exploitation in Research Projects.pdf
Summary of progress and overall contribution to the project	9 exploitation principles were developed in addition to the CA. Raising awareness in the consortium re IPR issues and exploitation options in several meetings.

	The VALCRI system has been demonstrated at 17 various events and to EU, UK, USA, Australia, Hong Kong Policing organisations. There have been high level discussions with a commercial Police IT provider regarding possible commercialisation. MU, AES and FHG have organised the discussion about IPR issues and exploitation options the VALCIR system in several project meetings. SPACE has worked extensively with LPA to develop a functional intelligence tool.

Task 14.7	Security Scrutiny
Task Leader	WMP, LPA
Participants	BFP
Status	COMPLETED
Output and deliverables	**Deliverables** ▪ D1.1 ManSystemsReport final.pdf (M06)
Summary of progress and overall contribution to the project	The process of security scrutiny is supposed to ensure that no sensitive material concerning the partners from police forces is being published. A procedure concerning security scrutiny is put into place which works efficiently.

Issues and Deviations from Planning:

Task	Timing	Status	Deviation / issues / Comments
T14.1	M01-M48	COMPLETED	No deviation
T14.2	M01-M48	COMPLETED	No deviation
T14.3	M01-M48	COMPLETED	Slight deviation from original lead, MU took the lead of the design and development outward facing website and TUW maintained the updates.
T14.4	M01-M48	COMPLETED	No deviation
T14.5	M01-M48	COMPLETED	No deviation
T14.6	M01-M48	COMPLETED	No deviation
T14.7	M01-M48	COMPLETED	No deviation

Required Corrective Action

None.

Relevant Publications

Authors	Titles	Year	Conference / Journal	Status

S. Schlehahn, T. Marquenie, E. Kindt	Data Protection Impact Assessment Guidelines	2018	Undecided	In progress
Wong, B.,L.	How Analysts Think (?): Early Observations	2014	Intelligence and Security Informatics Conference (JISIC), 2014IEEE Joint (pp. 296-299). IEEE.	ACCEPTED / PUBLISHED
Pohl, M. et al	Sensemaking and Cognitive Bias Mitigation in Visual Analytics	2014	Intelligence and Security Informatics Conference (JISIC), 2014 IEEE Joint (pp. 296-299). IEEE.	PUBLISHED
Rooney, C. et al	INVISQUE as a Tool for Intelligence Analysis: The Construction of Explanatory Narratives	2014	International Journal of Human-Computer Interaction, 30 (9). pp. 703-717.	PUBLISHED

2.3. ALIGNMENT VIA SUBGROUPS (SG), ACTIVITIES IN THE REPORTING PERIOD

Within VALCRI, sub groups were formed to align across work packages. They are described in the following sections, including a description of their relevant activities within the reporting period. The respective efforts are fully described in the related WP.

2.3.1 SG1: DEA (data extraction and analysis)

Goal	coordinate work related to data extraction (unstructured->structured data) and data analysis, and related technologies
Lead	MU, FHG, UKON
Participants	MU, UKON, MU, FHG, AES, CITY, SPACE, LIU
Related WP	WP05, WP06, WP08
Description of activities	The data extraction group (DEA) developed data extraction components that process crime data see WP08 and visualization components that display the result in human-interpretable visual forms see WP05. The group held regular teleconferences as well as conversations via emails and face to face meetings. The subgroup made good progress and achieved the following: ▪ Robust text processing and concept extraction methodologies ▪ The similarity space selector component were refined with a weight observer component added to record the analysis provenance, and a sequential pattern miner to look for crime patterns ▪ The associative search component refined and completed ▪ Visual content analysis component refined and completed ▪ The above mentioned components are integrated within the VALCRI system ▪ A number of publications are published in international journals and conferences see 2.2.5 and 2.2.8. few are recently submitted and expected to published soon The data extraction components and the visualization components developed by WP05 and WP8 were geared together to form a visual analytics pipeline that forms part of the analytics backbone of the VALCRI system. End user evaluation sessions were carried out though the lifetime of the project and the DEA components had multiple iterations and the last evaluation received positive feedback.

2.3.2 SG2: DMO (data management and ontologies)

Goal	coordinate work related to data management and ontologies, (incl. modelling, storage, querying, data import/export) and related technologies
Lead	Eva/LIU + Rudi/OS
Participants	LIU, OS, SPACE, FHG, AES, ULD, IINT, MU
Related WP	WP10, WP06, WP07, WP09, WP11, WP12
Description of activities	At the initial stages of the project there was some setback identifying a storage solution to support all the project requirements which were unfolding. The

	System Design and Architecture group continued to review the storage requirements and identified a solution as part of T10.1. For the structured data storage Apache Jena Fusekias and Elasticsearch which is based on Lucene was identified. For the unstructured data storage MongoDB was identified. All solutions had the capability of implementing the identified SEPL requirements.

2.3.3 SG3: DAR (design and architecture)

Goal	coordinate work related to system design & architecture, and related tools; consideration of exploitation aspects for architecture
Lead	SPACE
Participants	SPACE, FHG, OS, LIU, MU, AES, IX
Related WP	WP02, WP03, WP04, WP05, WP06, WP07, WP08, WP09, WP10, WP11, WP12, WP14
Description of activities	▪ Organised multiple meetings in order to consolidate and refine the overall system architecture ▪ Addressed multiple IEB remarks in the overall system ▪ Updated system architecture documentation ▪ A uniform data ingestion framework was developed ▪ Uniform deployment strategy was implemented

3.3.4 SG4: RCC (requirements consolidation and concretization)

Goal	coordinate work related to requirements consolidation and concretization, and related tools
Lead	MU, AES
Participants	SPACE, MU, AES, WMP, LPA, BFP, ULD, FHG
Related WP	WP02, WP03
Description of activities	▪ During multiple meetings the Semantic Media Wiki (SMW) was updated taking into account the latest developments of the individual partners. ▪ Organised multiple evaluation sessions for the VALCRI system in order to gather feedback from end-users. ▪ In order to make the VALCRI system available to all consortium members the system was deployed though a secure VPN connection.

2.3.5 SG5: SEPL (security, ethics, privacy, legal)

Goal	
Lead	Patrick Aichroth, FHG + Rudi Schreiner, OS + Eva Schlehahn, ULD
Participants	FHG, OS, KUL, MU, ULD, AES, IINT
Related WP	WP03, WP11, WP12

Description of activities	Regular SEPL-subgroup teleconferences and face to face meetings were organise to align work from a legal, ethical, privacy and security technological perspectives. Creation of definition documents: ▪ Personal data ▪ Justification, Proportionality, and Necessity ▪ Purpose Limitation SEPL requirements alignments, extension of requirements see WP02. Specification of SEPL /UIUX /DMO related component communication and flows (see WP12 and WP2) and communication with UIUX and DMO for system integration of SEPL technologies were completed. Overall SEPL process descriptions and related technological "toolbox" definitions to address SEPL requirements see WP12 was improved. Development and Improvement of key technologies to address SEPL issues: MDS, access control, logging, PET, re-identification analysis, video face encryption see WP11+12. Joint presentation of SEPL goals and overall technological approach in meetings, organization of several specific SEPL workshops in general meetings, and two specific SEPL meetings see WP12 Investigation of the national differences in interpretation and correlating practices in relation to purpose limitation in the context of the legal data protection framework. This was especially done so far with regard to the UK specifics. SEPL-subgroup meetings in Malta (Nov 2014), and in Birmingham (May 2015, as a follow-up of the workshop on data limitation and purpose in the UK) see WP3 Alignment of a comparative analysis of purpose limitation in Belgium, UK, Germany and EU see WP3 Alignment of the 1st privacy DSL see WP12 Several SEPL white papers see WP3 and WP12

2.3.6 SG6: UUC (UX/UI and cognitive aspects)

Goal	Coordinate work related to user experience, interfaces, visualization (incl. related cognitive aspects), and related technologies
Lead	MU, TUGraz
Participants	MU, CITY, IX, UKON, TUGraz, IINT, TUW, SPACE, AES, FHG
Related WP	WP04, WP05, WP06, WP07
Description of activities	The main priority of the UI/UX group is to coordinate between the human issues in sense-making and argumentation, user interface design, and the technical development of the Analyst User Interface. We routinely organised regular teleconferences and face to face meetings. WP03, 04, 05, 06 & 07 are all related to the design and development of the Analyst User Interface and the UI/UX group coordinates research and

	development across these work packages. The Reasoning Workspace comprises of three conceptual spaces needed by an intelligence analyst to work with and exploit information in large data sets. These three spaces are split across WP 05, 06, & 07. They sit on top of a technical UI architecture that is designed and developed in WP04. They are supported and guided by research in sense-making and argumentation conducted as part of WP03. Our first approach to interaction between the WP03 psychologists and the UI designers and developers was through the publication of a series of guidelines on how to design VALCRI such that it would mitigate biases and support sense-making. This has further evolved into closer collaborations between the two groups, where researchers and developers worked together to create and refine components through user evaluation.

2.3.7 SG7: Training and Evaluation

Goal	Coordinate work related to training and evaluation, and related tools
Lead	i-intelligence
Participants	IINT, AES
Related WP	WP13
Description of activities	The Subgroup on Training and Evaluation under took the following activities:

	▪ VALCRI system evaluations: - Scientific evaluations for publication purposes: Organised and coordinated the three end user evaluations in Krakow (June 2015), Berlin (May 2017) and November (2017). Nine end users in each evaluation session. - Evaluation with LEA: The VALCRI system has been evaluated with 214 law enforcement officers in 50 agencies in 16 countries excluding European and international law enforcement agencies. These generated end user feedback on the VALCRI system. ▪ A subset of the law enforcement agencies participated in 8 events evaluating the VALCRI Training curriculum developed with: 123 law enforcement officers in 40 organisations in 13 countries. These workshops were used to develop training materials for the training and development of law enforcement professionals. Evaluated specific modules from the VALCRI syllabus, and to validate the importance of structured analytic methodologies to support sensemaking, evidential reasoning and bias mitigation. ▪ Organised informal evaluations of the VALCRI components at successive Consortium Meetings. ▪ Organised a series of meetings (physical and virtual) with End Users to discuss the various iterations of the VALCRI Syllabus, as well as their professional training needs. ▪ Discussed ways in which the results of WP13 can be sustained and / or commercialised after the end of the project. These discussions have resulted in a formal program on police analysis to be given via Aston University in the UK. They have also resulted in a series of continuing professional development programs to be launched in the second half of 2018. ▪ Generated successive drafts of the VALCRI Syllabus and submitted these for review to End Users within and beyond the VALCRI project. This includes analysts working in other spheres such as national security, disaster management, and business intelligence. Our objective here was to generate as many constructive inputs as possible. ▪ Presented the VALCRI syllabus at: - The International Crime and Intelligence Analysis Conference held from 25 to 26 February 2016 in Manchester, England - The Five Eyes Analytic Training Workshop held from 29 February to 2 March 2016 at the James Madison University in Harrison, VA, USA - The International Studies Association Annual Convention held from 22 to 25 February 2017 in Baltimore, MD, United States - The Second Annual Conference of IAFIE's EUROPE held from 22 to 24 June in Athens, Greece - The Australian Institute of Professional Intelligence Officers (AIPIO) International Conference held from 22 to 24 August 2017 in Hobart, Australia - The CEPOL Research and Science Conference held from 28 to 30 November 2017 in Budapest, Hungry

2.4 PROJECT MANAGEMENT

In Period 1 (M1-M16), we reported on the set up of all necessary management systems and communications channels. All these have been set up and are being successfully used. In Period 2 (M17-M32) we reported on a number of significant issues concerned with the integration of the respective components and how the HIF and SEPL would influence the design of VALCR. In Period 3 (M33 – M44/M50), the emphasis was on two areas: integration and testing of the software, and demonstration and evaluation with the view to exploit the VALCRI outcomes. Issues in Period 3 were generally concerned with ensuring that the remaining funding was consumed in a productive and responsible manner. Most problems were generally resolved internally, sometimes with assistance from the EC, and with the EC in the loop particularly when significant decisions were being made.

2.4.1 Scientific Coordination

By the start of Period 3, most applications and functions were ready for integration. We attempted to integrate as many components as possible. The results have been pleasing and have been almost able to demonstrate the full set of VALCRI capabilities. Some significant outcomes are:

1. An integrated multi-application system of over 75 software components. This system is stable and was used extensively in various demonstrations to numerous LEA officers.
2. A re-factoring of the VALCRI software, from versions used in Period 2 to versions more likely able to accommodate police actual data, e.g. user interface - changing from the GWT framework to a mainly Javascript based environment.
3. We changed the database to a Postgres SQL database with ElasticSearch capabilities to ensure continuity to commercial exploitation.
4. While much has been learnt, much of the ontology efforts were not to be integrated into the VALCRI system, but will be demonstrable as a stand-alone entity.
5. DPIA or Data Protection Information Audit is a tool jointly developed between KUL and ULD
6. Completed the development of the assembly box functions.

2.4.2 Administrative and financial stability

In Period 2, partner XI agreed to withdraw from active participation in the FP7 VALCRI project. Having helped us establish the modern, flat design VALCRI look and feel which were further developed by MU and partners, XI believed it was time to focus their limited resources on other business opportunities. The European Commission was kept aware of all proceedings. In their new role, XI continued to be a partner but would not be expected to contribute to further activities of VALCRI. In line with the GA and CA, all partners who need to use any IP generated by XI, will be able to continue to freely use it. XI would no longer have voting rights on the Project Board. Partner XI left the project at end M32, and remaining budget and work was reallocated to MU but in subsequent months XI disputed the overpayment to them arising from earlier pre-finance payments. The Commission has been kept aware of these more recent developments.

One of the follow-up actions from MTR-1 and MTR-2 was to monitor partner spending to ensure that large un-used funding would not accumulate at the end of the project, especially if it that the funding could have been better used in delivering software at a higher TRLs or other exploitable outcomes that have real impact on policing. In Period 2 we instituted a 3 monthly informal reporting by all partners to the Project Coordinator to monitor spending and to also ensure that the spending are was sensible. This was continued into Period 3. Partners would report their information in the same format as is currently practised using the Form Cs on the EU Portal. In this way, the Project Coordinator has been able to review each partner's work area, and to discuss with them how we could more closely align their research or development efforts to achieve the ideals of the VALCRI concept, and whether at their given spend rate, would they be likely to end up have

with an under-spend. If an under-spend was likely, corrective actions could be taken early to either increase their outcomes or the possibility for re-allocation of their funding.

2.4.3 Exploitation: Phase 4 Plans and Project Extension

Given the progress we were making in the first few months of Period 2, we believed it might be possible to alter the project's plans to create a separate pathway for developing a commercially viable exploitable track we tentatively referred to as Phase 4 (NB: there are only 3 phases in the project as described in the DOW). It was envisaged that Phase 4 would productise a number of core VALCRI components so that each component may be very near a commercially saleable state by the end of the project. Early financial estimates suggested that the under-spend reported in Period 1 could be used to fund this separate exploitation pathway, and together with a no-funding extension of 6-months, that it should have been possible to also continue with the development of the remaining of the project components at the target TRL-5.

In the remaining six months of Period 2, it became very clear that Phase 4 will not be achievable. Progress on the technical development and technical integration became much slower than expected. A number of hold-ups include the delays in the selection of a database that would meet both the research needs of partners while at the same time being open sourced while not requiring commercial partners to reveal their source code. This has since been resolved, settling for the open source graph database OrientDB (Apache license) instead of the triple store Virtuoso database, in combination with ElasticSearch. This would have meant that if we had proceeded with Phase 4, then we would only have a suite of individual components. Police end-users and the External Advisory Board have constantly reminded us that the real power of VALCRI is in the functional integration of the components that works in an operationally integrated manner. Other delays include the finalisation of the design for how the semantic concepts would be extracted, presented, and manipulated by an analyst to be used in "find me more like this" operations, in the automated population of the Comparative Case Analysis tables, and others; the design of the associative search function; the finalisation of how security and data access can be implemented and the dynamics of interactions with other system components and user selection of data; and the completion of the data pipelines connecting the databases through the middleware, analytics and computation layers, through to the applications at the user interface.

We have since returned to the original plan as indicated in the DOW and with an additional 6 months to complete the work. Given the new situation and circumstances, the project focused on developing a functionally and operationally integrated TRL-5 system prototype, with a small number of components at TRL-4, or -3 and -2. The architecture accommodated incremental and modular addition (although some effort would still be required to enable new components to interface). But the longer term plan was for the main backbone to be modular such that new components can be relatively easily connected and integrated. In this way, we ensure that the VALCRI project will produce robust platform at TRL-5 that has been trialled in police environments. We are currently operationally trialling this – with the current VALCRI system prototype deployed in a stand-alone configuration at the West Midlands Police, Antwerp City Police, and at the Belgian Federal Police. Furthermore, all partners have full access to the VALCRI system prototype for their free use in current and future research activities, demos, and even further development should they have the funding to do so.

We have also demonstrated VALCRI to 50 European and International law enforcement agencies. Whilst all police feedback emphasise the integration aspect and are very keen on operationally using the VALCRI system, the main problem is still that of resources to take VALCRI across the "valley of death" from TRL-5 to TRL-9.

The Independent Ethics Board, or IEB, was a body proposed by us in the project proposal to collaborate with the Consortium by guiding our work and the resulting project outcomes to give some assurance that (i) the processes by which we developed VALCRI are ethically sound, and (ii) the products of VALCRI would similarly be ethically sound. In Period 1, we set up the coordination procedures, the points of contact between the IEB and the Consortium, and to establish a very open working relationship. In Period 2 we further developed

the working procedures to include an Issues Management system to keep track of the ethical issues raised by the IEB and how we followed-up to resolve the issues. All the issues raised have been documented in this Issue Management system, a simple spread-sheet describing the issues and the follow-up actions taken, and the major ones have been highlighted in their second IEB Report to the EC, provided at the end of Period 2.

Although the idea of ethics is not new, implementing in development processes as well as incorporating ethics by design into the technologies we develop, has proven extremely challenging. This is not due to a lack of willingness to follow-through on implementing IEB recommendations, but largely due to a lack of understanding how technology might be designed to prevent some of the potential ethical pitfalls. Other debates include whether it is at all possible to adopt IEB recommendations to "hard code" ethical safeguards into the software. In several cases, we have argued that hard coding safeguards simply make the software no longer practical for use and commercially not exploitable.

The IEB has submitted a separate and independent report (IEB Final Report on VALCRI Project, 2018). Please see Annex D.

In summary, our approach to addressing the IEB concerns is through the Ethical Issue Management system and by deliberately making our processes open and transparent to the IEB: they were invited to attend any of our meetings, and we have specially invited them to all our 3-4 monthly consortium meetings. We developed a method where we use a number of specific case studies to analytically reason about how the situation might cause ethical issues to emerge, and to role-play how the system will respond with data to those problems. One key problem is that ethical problems are highly context dependent, making the assessments subjective, and conclusions potentially limited. Regardless, we continued to pursue the resolution of the ethical issues raised as fertile grounds for research.

ANNEX A
ADVANCES AND GROUND BREAKERS

ADVANCES AND GROUNDBREAKING SCIENCE AND TECHNOLOGIES

In this Annex, we document what we believe are some of the significant advances and ground-breaking science and technologies that have occurred during the course of the project. Science and technology is interpreted broadly to include the wide range of disciplines that have been applied to the research and development of VALCRI, and includes advances we have made in disciplines that have informed the development of the VALCRI science and technology, such as legal, ethical, privacy, psychology, and training. In the description of the significant advances and ground-breaking science and technology we sought to explain in about 500 words or less, the significant outcome. We then provide a brief explanation of why we have considered it a significant advancement or a ground-breaker in terms of the following:

1. Created an S&T that is entirely new.
2. Improved an existing S&T to a significant (ideally measurable) degree.
3. Proved – e.g. by operational experimentation -- what was previously only a hypothesis concerning a particular S&T.
4. Rejected an established view or conventional wisdom about a particular S&T.
5. Developed a new way of thinking about an S&T and/or its foundational principles.
6. Developed a new way of evaluating an S&T.

We then provide a list of publications or other materials that can be used to justify our claims. Finally, for each AGB, we also provide a brief statement of the impact that the work reported has had or will have in the near future.

AGB1. The Analyst's User Interface: How Analysts Think

B.L. William Wong, Neesha Kodagoda, Chris Rooney, and Patrick Seidler
Middlesex University

Description of the significant Advances or Ground-breaking Science or Technology

The VALCRI Analyst's User Interface (AUI) is the user interface software that orchestrates how police analysts and investigators access and use the VALCRI functions during investigative and intelligence analysis. The AUI includes both the visualisation and the interaction methods. The design of the AUI has been based on two parts: (i) How Analysts Think, a description of the analysts' reasoning and problem solving during investigative sense-making, and (ii) the Thinking Landscape design of the interaction and visualisation of the User Interface to support how analysts' think during investigative sense-making.

In this report, we focus on our findings of How Analysts Think.

We took a cognitive engineering approach and through a series of cognitive task analysis interviews with police intelligence analysts and investigators, we discovered that the process of intelligence analysis is about investigative sense-making in complex systems environments.

The criminal intelligence and investigative analysis process addressed by VALCRI covers a range of tasks from strategic intelligence to tactical intelligence and individual case support. Strategic intelligence often found with performance analysis and is largely based on statistical analysis of large historical datasets in relation to geography to determine crime trends, patterns, and hotspots. Tactical intelligence and individual case management require the collation of specific pieces of data from which to explain a criminal situation and to justify police action.

The analyst's working environment particularly challenging because of ambiguity and uncertainty, incomplete, out of sequence data from multiple sources, comprising structured and unstructured data, stored amongst large volumes of data, distributed across many different databases. Analysts are also under significant time pressures, with the nature of crimes constantly evolving. The outcomes from their analyses are often emergent, rather than determinant.

Analysts and investigators only ever have fragments of data about an incident from which to construct an explanation, a supposition, or a plausible hypothesis that can guide the process of investigative sense-making and inference making. Analysts engage in a very broad variety of thinking strategies and analytic reasoning activities involving a range of analytic rigour from: (a) the use of expert intuition where their thinking is highly playful, tentative and creative, generating suppositions may be used to create plausible hypotheses that can guide subsequent inquiry, leading to (b) more rigorous analyses using formal and critical thinking strategies and structured analytic techniques, applying the scientific method to systematically test if these plausible hypotheses are correct. Figure 1 shows the range of analytic rigour, and the triple triangle of sense-making illustrating how the inference making process A – I – D (Abductive – Inductive – Deductive) inference making, co-occurs with reasoning strategies for anchoring, associative questioning and laddering; and the higher cognitive functions that facilitate the use of expert intuition to make sense of fragmentary pieces of data through which leaps of faith enable analysts to arrive at insights. Then guided by argumentation theory, analysts will assemble the data fragments to create explanatory narratives that explain a criminal situation and provide a basis for police to justify their decisions.

This process becomes the basis for how an analyst might employ their expert intuition in combination with the need for analytic rigour through the application of the scientific method. Analysts will formulate hypotheses based on incomplete and ambiguous fragments of data, deepen their inquiry, and then apply critical thinking and the scientific method to test those hypotheses. To encourage low commitment to any hypotheses and thereby discourage certain forms of human biases, the tools must enable fluid interactions to fluidly move between expert intuition and the application of the scientific method.

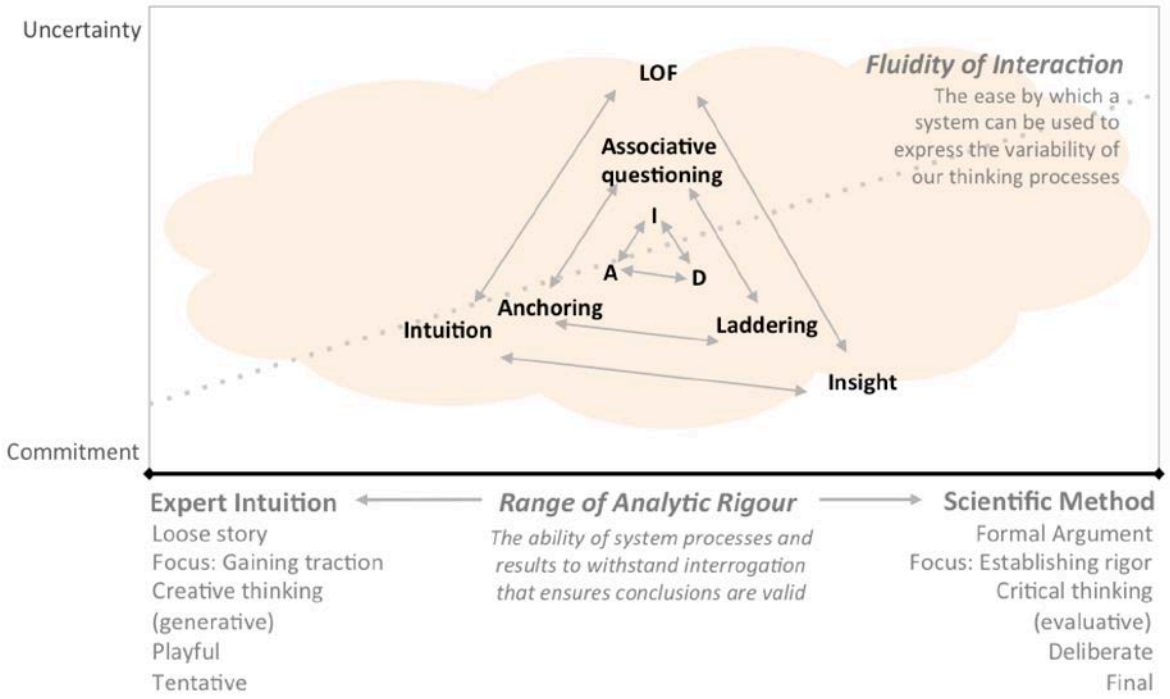


Figure 1. How Analysts Think (Wong et al, 2018)

The importance of differentiating between 'what analysts do' and 'how analysts think': Focusing on 'what analysts do' (such as the Pirolli and Card model, and the Attfield and Blandford model), will lead to designs that reflect the workflows associated with processes such as collection, indexing, retrieval, collation, analysis, shoe-boxing, presentation and communication. While important, they fail to capture the key challenges that analysts face: Having to make inferences and drawing conclusions based on fragmentary data that is out of sequence, ambiguous and uncertain.

Explanation of why it is a significant advancement or ground-breaking

Almost all intelligence analysis systems support the workflows associated with information search and retrieval of data, and the analyses that need to be performed on the data. However, few systems, if any, have combined UI technologies with ML and database technologies to support sense-making and the range of analytic reasoning from the use expert intuition to the application of the scientific method.

Published papers or patents to prove the above claims

Gerber, M., Wong, B. L. W., & Kodagoda, N. (2016a). How analysts think: decision making in the absence of clear facts. Adaptation of the RPD model and the decision ladder to analysts' decision making *Proceedings of the 7th European Intelligence Security Informatics Conference, EISIC 2016, on Counterterrorism and Criminology, 17-19 August, 2016, Uppsalla, Sweden* (pp. To be published): SAGE Publications.

Gerber, M., Wong, B. L. W., & Kodagoda, N. (2016b). How analysts think: Intuition, Leap of Faith and Insight *Proceedings of the Human Factors and Ergonomics Society 60th Annual Meeting, 19-23 September 2016, Washington, D.C., USA* (pp. 173-177): SAGE Publications.

Qazi, N., Wong, B. L. W., Kodagoda, N., & Adderley, R. (2016). Associative Search through Formal Concept Analysis in Criminal Intelligence Analysis *Proceedings of 2016 IEEE International Conference on Systems, Man, and Cybernetics • SMC 2016 October 9-12, 2016, Budapest, Hungary*: IEEE Press.

Rooney, C., Attfield, S., Wong, B. L. W., & Choudhury, S. T. (2014). INVISQUE as a tool for intelligence analysis: the construction of explanatory narratives. *International Journal of Human Computer Interaction, 30*(9), 703-717.

Takken, S., & Wong, B. L. W. (2015). Tactile reasoning: Hands-on vs. Hands-off - what's the difference? *Cognition, Technology & Work, 17*(3), 381-390. doi:10.1007/s10111-015-0331-5

Wong, B. L. W. (2013). *Fluidity and Rigour - Designing Visual Analytics for the Demands of Intelligence Analysis. Keynote Presentation.* Paper presented at the NATO IST-116 Symposium on Visual Analytics, Defence Academy of the United Kingdom, Shrivenham, UK, 28-30 October 2013.

Wong, B. L. W. (2014). How analysts think (?): Early observations *Proceedings of the IEEE Joint Intelligence and Security Informatics Conference, The Hague, The Netherlands, 24-26 September 2014* (pp. 296-299): IEEE Press.

Wong, B. L. W. (2016). Fluidity and Rigour: Addressing the Design Considerations for OSINT Tools and Processes. In B. Akhgar, P. S. Bayerl, & F. Sampson (Eds.), *Open Source Intelligence Investigation: From Strategy to Implementation* (pp. 167-189). Cham, Switzeland: Springer International Publishing AG.

Wong, B. L. W., & Kodagoda, N. (2015). How analysts think: Inference making strategies *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting, 26-30October 2015, Los Angeles, USA* (pp. 269-273): SAGE Publications.

Wong, B. L. W., & Kodagoda, N. (2016). How analysts think: Anchoring, Laddering and Associations *Proceedings of the Human Factors and Ergonomics Society 60th Annual Meeting, 19-23 September 2016, Washington, D.C., USA* (pp. 178-182): SAGE Publications.

Wong, B. L. W., Seidler, P., Kodagoda, N., & Rooney, C. (2018). Supporting variability in criminal intelligence analysis: From expert intuition to critical and rigorous analysis. In G. Leventakis & M. R. Haberfeld (Eds.), *Societal Implications of Community-Oriented Policing Technology* (pp. In Press). Cham, Switzerland: Springer International Publishing AG.

Impact

The “How analysts think” model of sense-making was the blue-print for what VALCRI would be designed to do. The VALCRI user interface design has attracted significant interests from police forces around the world. By designing the User Interface so that it supports the ‘human decide, machines lift’ principle, we have developed a system that facilitates human reasoning and analytic discourse, by being tightly coupled with semi-automated human-mediated semantic knowledge extraction. This has become a significant product differentiator in a market occupied with many similar functional capabilities.

AGB2. The Analyst's User Interface: The Reasoning Workspace and Thinking Landscape

B.L. William Wong, Neesha Kodagoda, Chris Rooney, and Patrick Seidler
Middlesex University

Description of the significant Advances or Ground-breaking Science or Technology

The VALCRI Analyst's User Interface (AUI) orchestrates how police analysts and investigators access and use the VALCRI functions during investigative and intelligence analysis. The AUI includes both the visualisation and the interaction methods. The design of the AUI has been based on: (i) How Analysts Think, a description of the psychological and human factors concepts that underlie human reasoning and problem solving during investigative sense-making, and (ii) the design of the interaction and visualisation of the User Interface to support how analysts' think during investigative sense-making.

Imagination, insight, fluidity and rigour, transparency

In this report, we focus on our how the Analysts' User Interface was designed to Support How Analysts Think:

The Reasoning Workspace, a design concept comprising three workspaces: a Data Space where the analysts may see what is stored in the database and to join up data to create the potential for making associations; an Analysis Space where machine learning (ML) and other computational methods are employed to manage large data sets, find possible interesting and relevant data based on similarity and associative characteristics; and a Hypothesis Space to create a 'play space' where analysts can create, disconnect, and re-create links between different pieces of evidence (in a non-legal context) to enable storytelling, sense-making, decision making, inference making, using various cognitive strategies. We emphasise the use of ML and related techniques to support the principle that 'humans decide, machines lift'. VALCRI uses ML to help find interesting and potentially relevant data from amongst data distributed across multiple databases.

Tactile Reasoning is an interaction technique that supports sense-making by the direct manipulation of information objects in the user interface. When pieces or fragments of information can be freely moved, manipulated, grouped and re-arranged in a visuo-spatial manner, new meanings or relationships may be discovered. VALCRI has implemented this concept of tactile reasoning in the AUI.

Representation Design. In representing physical systems such as nuclear power plants, important functional relationships existing between physical components and higher order goals and constraints are known a priori. These functional relationships can be mapped to visual representations that are designed before the system is commissioned. Such visual representations then enable operators of the system to control the performance of the physical processes in relation to more abstract goals such as profitability or system integrity. In intelligence analysis systems, such a priori relationships do not exist. In fact, such relationships – the narrative that explains the connections between pieces of evidence – needs to be constructed as the data or evidence becomes available in the context of the changing situation. The VALCRI user interface enables an analyst to assemble and construct such relationships between fragments of information to create explanatory narratives.

The Law of Requisite Variety explains that for a system to work, it must possess the functions needed to control or support the variety of behaviours inherent in the processes it was intended to control or support. A lack of compatibility will lead to brittle systems or failure or sub-optimal performance. Systems designed to support intelligence analysis need to support not only the observable tasks of information search and retrieval and data analysis, but also the much less observable but crucial thinking and reasoning processes in the formulation of hypotheses. These are the cognitive processes that determine the logic and how sensible are the narratives created to explain the clues that present themselves in an investigation.

Fluidity and Rigour. Fluidity refers to the ease by which a system can be used to support the variability of thinking strategies expressed in the analytic reasoning process; and by rigour we mean the extent to which analytic methods and processes produce results and conclusions that are valid and can stand up to interrogation. The interaction and visualisation tools are required to support the application of expert intuition and its fluid transitions to and from the use of scientific methods to test and evaluate hypotheses.

Explanation of why it is a significant advancement or ground-breaking

Almost all intelligence analysis systems support the workflows associated with information search and retrieval of data, and the analyses that need to be performed on the data. However, few systems, if any, have combined UI technologies with ML and database technologies to support the analytic reasoning processes invoked when using and transitioning between expert intuition to scientific method, and back.

In user evaluations of the AUI and the system with over 123 police analysts, has been that the AUI is intuitive and that it helps them think through the problems.

Published papers or patents to prove the above claims

Wong, B. L. W., Seidler, P., Kodagoda, N., & Rooney, C. (2018). Supporting variability in criminal intelligence analysis: From expert intuition to critical and rigorous analysis. In G. Leventakis & M. R. Haberfeld (Eds.), *Societal Implications of Community-Oriented Policing Technology* (pp. In Press). Cham, Switzerland: Springer International Publishing AG.

Impact

The VALCRI user interface design has attracted significant interests from police forces around the world. By designing the User Interface so that it supports the 'human decide, machines lift' principle, we have developed a system that facilitates human reasoning and analytic discourse, by being tightly coupled with semi-automated human-mediated semantic knowledge extraction. This has become a significant product differentiator in a market occupied with many similar functional capabilities.

AGB3. Insight into challenges of using sophisticated software in criminal investigations and meeting disclosure obligations with regard to the production of "relevant material" under CPIA 1996

Carlisle George
Middlesex University

Description of the significant Advances or Ground-breaking Science or Technology

Advances in technology has seen the application of sophisticated (and intelligent) software systems being used in criminal investigations. These technologies are not simple tools that produce predictable calculations. Instead, they carry out sophisticated computational analyses on various sources of data to produce new data that can amount to "relevant material". This "relevant material" has the potential of being used as evidentiary material by the prosecution in criminal trials and hence disclosable to the defence (under the Criminal Procedure and Investigations Act 1996, s. 23(1)).

Since the repeal of section 69 of the UK Police and Criminal Evidence Act 1984, under the Common Law rule there is a presumption that a computer producing an evidential record was working properly at the material time and therefore that record is real evidence and admissible. This presumption is however, rebuttable if there is evidence to the contrary.

The design of sophisticated software tools in VALCRI has focused on many issues that are relevant to the production of such "new data" (potential evidentiary material). Two examples are (i) understanding and managing cognitive bias embedded in software and (i) implementing provenance mechanisms to keep track of user activities and the generation of new data. Both examples illustrate intractable challenges that can be faced when software systems have computational complexity.

The development of VALCRI has shed new light on the need to consider to what extent designers of sophisticated and intelligent systems used in criminal intelligence analysis need to consider how such systems fit in terms of meeting criminal procedure requirements (such as disclosure obligations) especially in common law legal systems.

The potential use of the outputs of sophisticated computational analyses as evidentiary material in criminal trials raises some concerns, especially with regard to the extent to which such software systems can be subject to scrutiny and the extent to which their records can be challenged by a defence team. The later may require an understanding of specialized systems and computational issues in other to investigate competence/errors, an ability for such systems to keep very efficient provenance records, an ability to explore such systems and an ability to understand the management of issues such as cognitive bias, among others.

Explanation of why it is a significant advancement or ground-breaking

The insights are ground-breaking because the design of software and use of techniques such artificial intelligence and machine learning, increasingly moves us towards a total "black box" approach to computations and the outputs of systems/machines. Further such outputs (data) may be seen as "uncontestable" and infallible. As humans rely more and more on machines and computer systems it may become impossible to challenge data generated by some software systems (because of computational complexity and access to such systems), yet such data have the potential to be use as evidentiary material. For example an output based on probabilistic reasoning from machine learning may become persuasive if used as "circumstantial" evidence.

Current criminal procedure requirements in relation to the use of evidence in criminal trials may not have been carefully considered in light of the ability of machines to learn, reason, make decisions and produce potential "evidentiary material" based on complex analyses. There is an increasing reliance on machines as an integral part of decision making yet, the computational aspects of these machines are rarely being considered. It is possible that software systems can be programmed to have biases, or even develop biases based on input data or machine learning algorithms. It may be impossible to test for these biases or correct them. It may also be impossible to have a comprehensive accounting of the provenance of data and system activities. These concerns raise important questions about the impact of new technologies on the criminal justice system.

Published papers or patents to prove the above claims

N/A

Impact

No impact yet

AGB4. Shewmaps - Gridded Geographical Summaries Of Multiple Spc Charts

Chris Rooney[1], Roger Beecham[2], Jason Dykes[2], William Wong[1]
[1] Middlesex University
[2] City University of London

Description of the significant Advances or Ground-breaking Science or Technology

We re-design Statistical Process Control (SPC) charts, as used by crime analysts at West Midlands Police (WMP), for a series of decision-making and reporting tasks. In undertaking this 'make-over' of an established graphical idiom we characterise the data and analysis tasks associated with crime monitoring through SPC charts and contribute a new approach to engaging with collaborators that may be used in applied visualization work more widely. The approach involves explicit visual evaluation of candidate designs against visual design principles with collaborating analysts in their intended context of use. It is achieved through a new visual encoding of design characteristics and dynamic design documents to explain, justify, explore and evaluate designs.

Explanation of why it is a significant advancement or ground-breaking

Our family of designs supports the spatial and historical analysis of signals in crime data in ways that current techniques do not. Through concise composite summaries, we can show current signals and trends, signal history, and process history in a small space, allowing, in the case of WMP, 174 neighbourhoods to be analysed concurrently - something that analysts at WMP could only do before by looking at 174 separate SPC charts.

Published papers or patents to prove the above claims

Rooney, C., Beecham, R., Dykes, J. & Wong, W. (2018) Statistical Process Control Charts: A Visualization Make-over for Crime Analysis. Under review: IEEE TVCG.

Impact

We have deployed Shewmaps at West Midlands Police as a standalone system. Analysts are able to upload their own CSV files and perform real data analysis.

AGB5. A Descriptive, Practical, Hybrid Argumentation Model to Assist With the Formulation of Defensible Assessments in Uncertain Sense-Making Environments

Celeste Groenewald, Simon Attfield, Peter Passmore, and B.L. William Wong
Middlesex University

Description of the significant Advances or Ground-breaking Science or Technology

Police crime analysts work with crime data in order to draw conclusions about factors such as potential perpetrators and causes of crime. Analysis is intended to support ongoing police work and is coordinated with the work of investigation teams, decision makers and prosecutors. Analysts' reasoning can by necessity be complex and conclusions can be based on many factors (Kodagoda and Wong, 2016), with varying levels of certainty. Task switching means that analysts need to reorient as they return to partially complete analyses. There is value in allowing analysts to visually record their evolving reasoning as a resource for reflection, orientation, audit and communication.

We have developed a visual argumentation language specifically designed for this purpose. It enables police analysts to record their evolving reasoning when conducting Visual Analytics with systems such as VALCRI. Argumentation schemes are subsumed under explicit terms of reference and can be structured in terms of the crime schema's that analysts have been observed to use. Assertions can be created based on inferences from data, including visualisations, and these can be linked to conclusions through inferential networks. Assertions and conclusions can be marked with varying levels of certainty and arguments can be linked to tasks, subtasks and external requests for information.

Explanation of why it is a significant advancement or ground-breaking

The novelty of this language lies in its ecological utility and the fact that it has been explicitly informed by empirical studies of how analysts think and work. For this reason it intentionally deviates from traditional, 'purist' argumentation schemes in the interests of practical value. Studies that have informed the language were conducted under VALCRI and include:

Gerber, M., Wong, B. W., & Kodagoda, N. (2016). How analysts think: intuition, leap of faith and insight. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 60, No. 1, pp. 173-177). Sage CA: Los Angeles, CA: SAGE Publications.

Groenewald, C., Wong, W.B.L., Attfield, S., Passmore, P., Kodagoda, N. (2017). How Analysts Think: Navigating Uncertainty – Aspirations, Considerations and Strategies. Proceedings of the 13th International Conference on Naturalistic Decision Making. Bath, United Kingdom.

Groenewald, C., Wong, W.B.L., Attfield, S., Passmore, P., Kodagoda, N. (2017) How Analysts Think: How do Criminal Intelligence Analysts Recognise and Manage Significant Information? Proceedings of the European Intelligence and Security Informatics Conference (EISIC) 2017, Dekelia Air Base, Attica, Greece. IEEE.

Selvaraj, N., Attfield, S., Passmore, P., & Wong, B. W. (2016). How Analysts Think: Think-steps as a Tool for Structuring Sensemaking in Criminal Intelligence Analysis. In Intelligence and Security Informatics Conference (EISIC), 2016 European (pp. 68-75). IEEE.

Wong, B. W. (2014). How analysts think (?): Early observations. In Intelligence and Security Informatics Conference (JISIC), 2014 IEEE Joint (pp. 296-299). IEEE.

Wong, B. W., & Kodagoda, N. (2016, September). How analysts think: Anchoring, Laddering and Associations. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 60, No. 1, pp. 178-182). Sage CA: Los Angeles, CA: SAGE Publications.

Published papers or patents to prove the above claims

Groenewald, C., Attfield, S., Passmore, P., Wong, B.L.W (in press) A Descriptive, Practical, Hybrid Argumentation Model to Assist With the Formulation of Defensible Assessments in Uncertain Sense-Making Environments. Springer Briefs in Policing.

Groenewald, C., Attfield, S., Passmore, P., Wong, B.L.W., and Kodagoda, N. (2017) A Practical, Hybrid Argumentation Model to Assist With the Formulation of Defensible Assessments in Uncertain Sense-Making Environments, Next Generation Community Policing Conference (NGCP), 25th-27th October 2017, Hereklion, Crete, Greece.

Impact

No impact yet.

AGB6. Interactive human-centered extraction of temporal-spatial and behavioral associations for crime analysis

Nadeem Qazi and B.L. William Wong
Middlesex University

Description of the significant Advances or Ground-breaking Science or Technology

This research has opened doors towards the use of text mining in analyzing the unstructured crime documents particularly focusing towards search methodologies. A relatively new concept of search different than semantic and keyword search is introduced and named it as Associative search. This research is an attempt to encapsulate all three dimensions of a crime scene (i.e modus operandi, spatial and temporal) in a single envelope. It uses NLP to extract the frequent terms from the crime reports through vector space models and represent crimes as a bag of these terms.

The proposed associative search like the linkage analysis elicits associations or links between the connected criminal entities based on criminal behavior, geographical and temporal proximity. It establishes associations between crimes and criminal identifying criminal communities that have similar tempo-spatial and modus operandi characteristic. On the other hand, it maps associations between unsolved crimes and offenders to generate the suspect list.

An Interactive human-centered data mining pipeline coupled with interactive visualization is developed to implement associative search in facilitating the crime analysis process in general and crime matching process in particular. The proposed pipeline is an integration of interactive data mining incorporating clustering algorithms, Bayesian reasoning, and graph theory to visualize similarities in crimes, offenders network and plausible suspect lists for unsolved crime and plausible unsolved crime list for a suspect. The developed framework integrates machine learning algorithms and human as a team where the machine does the heavy lifting of computation and human does the decision making. It facilitates human reasoning and analytic discourse for intelligence analysis through a semi-automated human-mediated semantic knowledge extraction capability. Thus this framework enables analysts to interact directly and interactively with ML models so they can integrate domain knowledge into the analysis process.

Another contributing research output is clustering analysis for categorical data lacking ground truth. The various clustering algorithms are evaluated to find the effect of criminal attributes on clustering.

Another aspect of this work is the development of interactive visualization that facilitates the sense making the process of crime analysis. A dynamic 2D crime cluster space is created to present the crime scene Key process indicators (KPIs) in coordinated views along with the iconic graphic. Multidimensional scaling allows visualizing the underlying hidden relationship between crime KPIs. Another important aspect is the visualization of this association in the form of knowledge graph.

This similarity based association are also evaluated through Bayesian theory in order to examine how the crime pattern of the suspect in one crime can be used in another for making any type of association. This is implemented through measuring prior and posterior probabilities of a suspect in committing a crime employing crime pattern characteristic as evidence. This thus facilitates crime matching process through visualizing all the plausible similarities and also assists in determining the likelihood of the given suspect in committing an unsolved crime. The proposed visualization aims to assist in hypothesis formulation reducing computational influence in the decision making of criminal matching process.

Explanation of why it is a significant advancement or ground-breaking

The police analysts follow a search based on associative questionings for establishing associations among the criminal to discover and reconstruct crimes scene. It needs a searching mechanism other than semantic and

keyword-based search. The proposed search mechanism incorporates associative questioning and not only elicit the association but also visualized them in a manner assist to find out how crime entities that appear to be unrelated at the surface, are actually linked to each other. The human-centered machine learning algorithms for unsupervised clustering algorithms with dynamic feature vector selection and visualization for the intelligent crime analysis may think to be stepping stone towards interactive criminal analysis. The framework is not automatic and brings human in the loop for decision making. The important characteristic of the presented visualization porotype is that it enables the analyst to make assessment rather than a recommendation.

This association based mechanism is also applicable in other fields especially in medical text mining to visualize diseases symptoms patients etc.

Published papers or patents to prove the above claims

Interactive Knowledge Discovery Scheme for Crime Pattern mining and Community Detection submitted in KDD 2018 to be held in London August 2018.

N. Qazi, William Wong ”Contextual visualization of crime matching through similarity clustering and Bayesian analysis ” is accepted as a chapter for publication in Social Media Strategy in Policing (from cultural intelligence to community policing) book to be published by Springer

N. Qazi, William Wong ”Behavioural Tempo-spatial Knowledge Graph for Crime matching through Associate Questioning and Graph Theory” (IEEE, European Intelligence, and Security Informatics Conference EISIC,2017,Greece)

Nadeem Qazi, Malachy McElholm & Liam Maguire (2017) A Model-View-Controller (MVC) architecture for contextual visualization of task-based multi-dimensional energy KPIs in a manufacturing process, published in International Journal of Ambient Energy (Taylor & Francis) May 2017 DOI: 10.1080/01430750.2017.1310135

N. Qazi, B.L. William Wong, Neesha Kodagoda and Rick Adderley,”Associative Search through Formal Concept Analysis in Criminal Intelligence Analysis”, Presented in 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2016) OCTOBER 9-12 2016 Budapest.

N. Qazi, William Wong “Semantic-Based Image Retrieval Through Combined Classifiers of Deep Neural Network and Wavelet Decomposition of Image Signal”, Presented at IEEE 2016 9th EUROSIM Congress on Modelling and Simulation Oulu, Finland 12-16 September 2016, will be published in IEEE Conference Proceeding

L. Zhang, C. Rooney, L. Nachmanson, W. Wong, B. C. Kwon, F. Stoffel,N. Qazi, U. Singh, and D. A. Keim.Spherical Similarity Explorer for Comparative Case Analysis (2016), Electronic Imaging Conference on Visualization and Data Analysis

Impact

Not Known.

AGB7. Aggregated Visualization of Elements outside of a Visualization Viewport (Off-Screen)

Juri Buchmüller, Dominik Jäckle, Daniel A. Keim, Bum Chul Kwon, Dominik Sacha, Hansi Senaratne, Andreas Stoffel, Florian Stoffel (alphabetic order)
University of Konstanz, Germany

Description of the significant Advances or Ground-breaking Science or Technology

In abstract data visualization, navigational and spatial context is key for efficient exploration of large data spaces. State-of-the-art techniques commonly provide zoom in (drill-down) features, but then the analyst loses spatial context. With aggregated visualization of elements outside of the viewport (off-screen), the spatial context can be preserved to a certain extent.

Explanation of why it is a significant advancement or ground-breaking

This family of techniques is a significant improvement for visual data exploration in large data spaces.

Published papers or patents to prove the above claims

Jäckle, D., Stoffel, F., Kwon, B. C., Sacha, D., Stoffel, A., Keim, D. A. (2015). Ambient Grids: Maintain Context-Awareness via Aggregated Off-Screen Visualization. In Eurographics Conference on Visualization (EuroVis)-Short Papers. The Eurographics Association.

Jäckle, D., Senaratne, H., Buchmüller, J., Keim, D. A. (2015). Integrated spatial uncertainty visualization using off-screen aggregation. In EuroVis Workshop on Visual Analytics (EuroVA). The Eurographics Association.

Impact

no impact yet

AGB8. Interactive Machine Learning for Crime Data Analysis

Michael Behrisch1, Geoffrey Ellis1, Johannes Fuchs1, Wolfgang Jentner1, Daniel A. Keim1, Dominik Sacha1, Florian Stoffel, Leishi Zhang2 (alphabetic order)
1University of Konstanz, Germany
2 Middlesex University

Description of the significant Advances or Ground-breaking Science or Technology

Currently, the analysis of crime data is mostly a manual task. Visualization and interaction techniques have been developed to allow analysts to utilize the latest techniques from machine learning and automated data analysis.

Explanation of why it is a significant advancement or ground-breaking

The researched techniques act as an enabler of automated data analysis and machine learning for a new application domain.

Published papers or patents to prove the above claims

Jentner W., Sacha D., Stoffel F., Ellis G., Zhang L., Keim D. A. (2018). Making Machine Intelligence Less Scary for Criminal Analysts: Reflections on Designing a Visual Comparative Case Analysis Tool. In The Visual Computer (2018): 1-17.

Stoffel, F., Jentner, W., Behrisch, M., Fuchs, J., Keim, D. (2017). Interactive ambiguity resolution of named entities in fictional literature. In Computer Graphics Forum (Vol. 36, No. 3, pp. 189-200).

Impact

Will be an integral component of a commercial product (VALCRI spinoff).

AGB9. Interactive Dimensionality-Reduction to foster Data Exploration and Sense Making

Geoffrey Ellis1, Dominik Jäckle1, Daniel A. Keim1, Sebastian Mittelstädt1, Harald Reiterer1, Dominik Sacha1, Florian Stoffel1, Leishi Zhang2 (alphabetic order)
1 University of Konstanz, Germany
2 Middlesex University, UK

Description of the significant Advances or Ground-breaking Science or Technology

Combines automated dimensionality reduction techniques and interaction for an effective support of data exploration and sense making tasks.

Explanation of why it is a significant advancement or ground-breaking

Typical data exploration systems require manual handling of data attributes. Interactive dimensionality-reduction does not require the analyst to select interesting attributes. It allows a comprehensive view on the dataset without any preconditions/filtering.

Published papers or patents to prove the above claims

Jäckle, D., Stoffel, F., Mittelstädt, S., Keim, D. A., & Reiterer, H. (2017, February). Interpretation of dimensionally reduced crime data: A study with untrained domain experts. In 12th International Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017) (pp. 164-175).

Sacha, D., Jentner, W., Zhang, L., Stoffel, F., Ellis, G. (2017). Visual Comparative Case Analytics. EuroVis Workshop on Visual Analytics (EuroVA).

Impact

Will be an integral component of a commercial product (VALCRI spinoff).

AGB10. Modelling the visual analytic process

Geoffrey Ellis, Daniel A. Keim, Bum Chul Kwon, Hansi Senaratne, Dominik Sacha, Andreas Stoffel, Florian Stoffel (alphabetic order)
University of Konstanz, Germany

Description of the significant Advances or Ground-breaking Science or Technology

Visual analytics is a complex interaction between machine and human. This work provides a terminology and process model and illustrates how uncertainties propagate though visual analytic systems and how user's awareness and trust affects knowledge construction.

Explanation of why it is a significant advancement or ground-breaking

Provides a framework for discussing visual analytics systems, which is highly beneficial in both evaluating existing systems and in designing new systems that can aid the user in better decision making.

Published papers or patents to prove the above claims

Sacha, D., Senaratne, H., Kwon, B. C., Ellis, G., & Keim, D. A. (2016). The role of uncertainty, awareness, and trust in visual analytics. *IEEE transactions on visualization and computer graphics*, *22*(1), 240-249.

Sacha, D., Stoffel, A., Stoffel, F., Kwon, B. C., Ellis, G., & Keim, D. A. (2014). Knowledge generation model for visual analytics. *IEEE transactions on visualization and computer graphics*, *20*(12), 1604-1613.

Impact

The Knowledge Generation Model for Visual analytics has been widely cited since its publication and has formed the basis of several other process models in the field.

Summary Explanation

In summary, collectively UKON has achieved ground-breaking advances in the combination of data and feature space that bridges gaps between machine learning and human reasoning. Data analysis, visualization, and interaction facilities are combined in novel ways to enable analysts to interact with techniques such as unstructured data analysis, dimensionality reduction, clustering, and pattern mining to foster sense-making in the area of comparative data analysis. The techniques are combined and visualized in a fashion such that users without any or little expertise in data science are capable of exploiting these techniques and their visualizations for comparative case analysis.

On the application level, the unique combination of techniques researched by UKON enables analysts to continue to use methods that they already are familiar, enriched with automated clustering and visualization techniques to indicate groups visually, their differences/commonalities, as well as outliers possibly influencing the comparison results. The research and development activities happened in the domain of criminal data analysis were developed in a user-driven fashion. In consequence, the results enable criminal data analysts to quickly recognize correlations, groups, commonalities, and differences for criminal cases that are subject to an on-going investigation, which is highly relevant and useful not only for cases that are hard to assess manually.

The lessons learned during this design process impacts the scientific community beyond the scope of criminal investigation as the techniques themselves as well as their unique combination can be adapted and used in many other problem domains.

Outreach

Because of the VALCRI project, UKON was able to establish a stable and fruitful collaboration with a German LEA on state-level (State Office for Criminal Investigation of North Rhine-Westphalia). Within the scope of the VALCRI project, the collaboration enabled the LEA to contribute feedback to the VALCRI prototype. Beyond VALCRI, a number of mutual onsite visits were happening, eventually leading to the first “Konstanz Summer School for Criminal Data Analysis” in June 2017, were a case analyst and a professor for criminology from the partnering LEA were involved. Additionally, UKON was able to attend several meet-ups and workshops (4th Crime Mapping Conference 2016 in Munich, 21st European Police Congress 2018) to advertise VALCRI and its benefits for LEAs.

AGB11. Provenance models, methods and system architecture to support legal, ethical, and privacy requirements when dealing with police data

Eva Blomqvist, Henrik Eriksson, Robin Keskisärkkä, Olaf Hartig
Linkoping University

Description of the significant Advances or Ground-breaking Science or Technology

In VALCRI we have explored various ways to deal with the legal, ethical and privacy requirements when it comes to managing police data using semantic web technologies, such as RDF/OWL/SPARQL. On one hand we have used the existing standards, such as PROV-O (W3C recommendation) to model various aspects of provenance that then support the enforcement of legal, ethical and privacy requirements, e.g., in the form of access restrictions, anonymisation, or deletion of data under certain criteria. This in itself is not groundbreaking, merely an application of existing standards. However, to actually enforce and manage criteria based on such a model, there is no current standard.

In VALCRI we have therefore developed an architectural approach, using special API:s, a template library, and the data models themselves, which together constitute the overall framework for actually using the provenance data to enforce restrictions (such as per-statement or per-graph access restrictions) or actions (such as data deletion or anonymisation) over the data. This novel architecture is an advancement in the field of applying semantic web technologies to real world problems, and relies on a new SPIN representation that also encompasses continuous queries for streaming data (http://w3id.org/rsp/spin). Currently we are also experimenting with how to best (e.g. most efficiently) support these requirements using the underlying RDF model, or RDF*, a proposed extension of RDF.

Explanation of why it is a significant advancement or ground-breaking

In the semantic web field, work on provenance, security and data protection, for instance, has so far mostly been about how to model these aspects in the data. No well-established solutions exist for actually working with these models, managing data and enforcing restrictions. This makes our work important from an architectural and method perspective, i.e., advancing the state of the art in how to actually use provenance models, and other data annotations, to solve real-world requirements pertaining to legal, ethical, and privacy aspects. Additionally, there are some well-known drawbacks of using RDF, such as the inability to add annotations or attributes directly on statements, without workarounds that instead explode in terms of data volume and reduce performance. Although the proposal of RDF* as an alternative model (by a colleague at LIU) is not related to VALCRI, the project provides a great opportunity to tests its benefits and limitations, since the data in VALCRI displays exactly the characteristics that RDF* was developed to support. Results of testing RDF* on VALCRI data will therefore be the first real-world test case for the new RDF* language, and will hopefully (experiments are still ongoing) bring interesting insights into its benefits and drawbacks.

Published papers or patents to prove the above claims

Papers describing the template model and API:s (although focusing mostly on the streaming aspects, in VALCRI the same principles are also used for "static" data):

Keskisärkkä, R. Representing RDF Stream Processing Queries in RSP-SPIN. In Proceedings of the ISWC 2016 Posters & Demonstrations Track co-located with the 15th International Semantic Web Conference (ISWC-2016), Kobe, Japan, October 17-21, 2016, CEUR Workshop Proceedings, 2016.

Keskisärkkä, R. Query Templates for RDF Stream Processing. In Proceedings of Stream Reasoning Workshop. 2016 October 17th-18th, 2016, Kobe, Japan. Collocated with the 15th International Semantic Web Conference (ISWC 2016), CEUR Workshop Proceedings, 2016.

Paper about RDF* (note, this work was not supported by VALCRI, but current experiments not yet published are applying it and evaluating it for VALCRI data):

Olaf Hartig **Foundations of RDF⋆ and SPARQL⋆ (An Alternative Approach to Statement-Level Metadata in RDF).** AMW 2017

Impact

Template model for streaming data is being included in a reference implementation of the RSP-QL language for querying streaming data on the web (https://github.com/streamreasoning/yasper), being developed as a result of the W3C RSP community group (https://www.w3.org/community/rsp/) work.

Other parts have no impact yet, especially since experimental results are not yet published, but we expect that both the comparison between classical RDF and RDF* modelling, as well as the architecture in itself could have quite some impact on practical use of these technologies in the future.

AGB12. Alignment Cubes – a tool for comparing and evaluating ontology alignments

Valentina Ivanova, Patrick Lambrix, (Emmanuel Pietriga, Benjamin Bach)
Linköping University

Description of the significant Advances or Ground-breaking Science or Technology

AlignmentCubes is a tool which allows for interactive visual exploration of several ontology alignments and thus supports alignments' evaluation at different levels of detail. These are some use cases in which AlignmentCubes will be helpful:

- Selecting, combining and fine tuning alignment algorithms and tools;
- Matchers development;
- Ontology alignment evolution;
- Validating and debugging of ontology alignments and reference alignments;
- Collaborative ontology alignment.

The tool is publicly available and can be downloaded from:

http://www.ida.liu.se/~patla00/research/AlignmentCubes/

Explanation of why it is a significant advancement or ground-breaking

Ontology alignment is an area of active research where many algorithms and approaches are being developed. Their performance is usually evaluated by comparing the produced alignments to a reference alignment in terms of traditional measures such as precision, recall and F-measure. These measures, however, only provide an overall assessment of the quality of the alignments, but do not reveal differences and commonalities between alignments at a finer-grained level such as, e.g., regions or individual mappings. Furthermore, reference alignments are often unavailable, which makes the comparative exploration of alignments at different levels of granularity even more important. Making such comparisons efficient calls for a "human-in-the-loop" approach, best supported through interactive visual representations of alignments. Our tool Alignment Cubes is the first tool that supports interactive exploration of multiple ontology alignments at different levels and complements the traditional measures.

Published papers or patents to prove the above claims

Ivanova V, Bach B, Pietriga E, Lambrix P, Alignment Cubes: Towards Interactive Visual Exploration and Evaluation of Multiple Ontology Alignments, 16th International Semantic Web Conference, LNCS 10587, 400-417, Vienna, Austria, 2017. https://doi.org/10.1007/978-3-319-68288-4_24

Ivanova V, Bach B, Pietriga E, Lambrix P, Alignment Cubes: Interactive Visual Exploration and Evaluation of Multiple Ontology Alignments, 16th International Semantic Web Conference Posters and Demos, CEUR Workshop Proceedings Volume 1963, Vienna, Austria, 2017.

Impact

After discussions with the organizers of the Ontology Alignment Evaluation Initiative, a yearly event for the evaluation of ontology alignment systems, the Alignment Cubes tool is now made available on their web site. Organizers of the different tracks can use the tool while evaluating and preparing the discussion section in the final report of the event. Participants can use the tool to evaluate and discuss the results they obtained.

AGB13. Requirements for user involvement for ontology alignment systems

Valentina Ivanova, Zlatan Dragisic, Patrick Lambrix, (Johan Åberg, Daniel Faria, Ernesto Jimenez-Ruiz, Catia Pesquita)
Linköping University

Description of the significant Advances or Ground-breaking Science or Technology

We have identified requirements for user involvement in ontology alignment. The requirements are organized into requirements regarding user interfaces and requirements regarding infrastructure and algorithms. The former category has three

Subcategories: manipulation, inspection and explanation. Those in the manipulation category include actions for transforming the mapping suggestions in an alignment. Those in the second subcategory cover a broad set of actions for inspecting the ontologies and alignments. The explanation category includes services for presenting information to the user. The infrastructure and algorithms category includes various requirements that arise from the growing the size and complexity of the ontologies, alignments and alignment problems.

Explanation of why it is a significant advancement or ground-breaking

The growth of the ontology alignment area in the past ten years has led to the development of many ontology alignment tools. The progress in the field has been accelerated by the Ontology Alignment Evaluation Initiative (OAEI) which has provided a discussion forum for developers and a platform for an annual evaluation of their tools. The number of participants in the OAEI increases each year, yet few provide a user interface and even fewer navigational aids or complex visualization techniques. Some systems provide scalable ontology alignment algorithms. However, **for achieving high-quality alignments user involvement during the process is indispensable**.

Nearly half of the current challenges in the field are directly related to user involvement. These include explanation of matching results to users, fostering the user involvement in the matching process and social and collaborative matching. Another challenge aims at supporting users' collaboration by providing infrastructure and support during all phases of the alignment process. All these challenges can be addressed by providing user interfaces in combination with suitable visualization techniques.

The demand for user involvement has been recognized by the alignment community and resulted in the introduction of the OAEI Interactive track in 2013.

Our contributions provide the first guidelines on what is required for introducing user involvement for ontology alignment systems.

Published papers or patents to prove the above claims

VALCRI publication:

Dragisic Z, Ivanova V, Lambrix P, Faria D, Jimenez-Ruiz E, Pesquita C, User validation in ontology alignment, *15th International Semantic Web Conference*, LNCS 9981, 200-217, Kobe, Japan, 2016. https://doi.org/10.1007/978-3-319-46523-4_13

Earlier:

Ivanova V, Lambrix P, Åberg J, Requirements for and Evaluation of User Support for Large-Scale Ontology Alignment, *12th Extended Semantic Web Conference - ESWC 2015,* LNCS 9088, 3-20, Portoroz, Slovenia, 2015. https://doi.org/10.1007/978-3-319-18818-8_1

Impact

We have used the requirements to evaluate our and other existing systems and have provided feedback to the community. For the future we expect developers of ontology alignment systems to use our guidelines as a checklist to be able to build better systems.

AGB14. RepOSE plugin for Protégé – a tool for completing ontologies

Zlatan Dragisic, Patrick Lambrix
Linköping University

Description of the significant Advances or Ground-breaking Science or Technology

The RepOSE plugin for Protégé allows for using the RepOSE functionality for detecting and repairing missing is-a structure within the Protégé ontology development environment.

When adding concepts to an ontology, it suggests additional is-a relationships as well as instances for the newly added concept. It checks for logical problems and when adding new information to the ontology, it shows the developer the inferred knowledge.

Explanation of why it is a significant advancement or ground-breaking

Developing ontologies is not an easy task and often the resulting ontologies (including their is-a structures) are not complete. In addition to being problematic for the correct modelling of a domain, such incomplete ontologies also influence the quality of semantically-enabled applications. Incomplete ontologies, when used in semantically-enabled applications, can lead to valid conclusions being missed.

RepOSE is a system for repairing missing is-a structure in ontologies. It is based on a formalization of the repairing problem as an abductive reasoning problem for which algorithms were developed for a specific kind of solutions to the problem, so called skyline-optimal solutions. RepOSE is the first system that deals with repairing missing is-a structure in ontologies in a more advanced way than just adding the missing relations.

The plugin enables the use of RepOSE functionality via Protégé, one of the most used ontology development tools.

Published papers or patents to prove the above claims

We have not published any paper about the plugin yet. However, there have been several publications related to the RepOSE system.

Earlier papers (not plugin):

Lambrix P, Wei-Kleiner F, Dragisic Z, Completing the is-a structure in light-weight ontologies, Journal of Biomedical Semantics 6:12, 2015.

Lambrix P, Ivanova V, A unified approach for debugging is-a structure and mappings in networked taxonomies, Journal of Biomedical Semantics 4:10, 2013.

Lambrix P, Liu Q, Debugging the missing is-a structure within taxonomies networked by partial reference alignments, Data & Knowledge Engineering 86:179-205, 2013.

Impact

The tool has been used for the VALCRI ontologies. Earlier versions of RepOSE were used for Anatomy track in the Ontology Alignment Evaluation Initiative and work for the Swedish National Food Agency. Currently, it is used in work for the Swedish Veterinary Agency.

AGB15. Advances in Ontology Design Pattern (ODP) usage methods (the eXtreme Design methodology) and support tooling.

Karl Hammar, Zlatan Dragisic, Patrick Lambrix and Eva Blomqvist
Linköping University

Description of the significant Advances or Ground-breaking Science or Technology

Development and evaluation within the VALCRI project has advanced the state of the art with regard to ODP usage methodology, specifically, the eXtreme Design (XD) methodology. Firstly, ontology debugging and matching has been integrated into the methodology. When integrating ODP-based ontology modules into one coherent ontology, it is not uncommon that modelling defects or inconsistencies arise, e.g., missing *is-a* or other relations. Ontology debugging approaches have traditionally been used on finished ontologies – the integration of these approaches into an ODP-based development methodology allows the developer to catch and remedy errors earlier. Secondly, a template-based ODP instantiation approach and corresponding algorithm has been developed within the project. This approach enables developers to "stamp out" copies of ODPs for reuse in their own ontology project, without having to use the potentially confusing *owl:imports* directive to reference and import remote ODPs as-is. This new method is also implemented in a WebProtégé tool plugin that is freely available (https://github.com/hammar/webprotege). Thirdly, the XD methodology has been extended with support for member roles and associated responsibilities in an ontology engineering project, support for the reuse of non-ODP ontology resources, and support for adaptation of projects to non-optimal project contexts.

Regarding support tooling for ODP use, advances have been made in terms of the quality of ODPs and their documentation and the metadata that ODPs need to display or encode in order to be successfully used (by both humans and machines). This work indicates trade-offs that may need to be made, e.g. between performance efficiency when reasoning over resulting ontologies, and the learnability of those ontologies, or the interoperability of ODPs versus the usability of those same ODPs, etc. These trade-offs often arise from underlying structural differences in the ODPs being used, so tooling that is aware of such matters can aid developers in the selection and application of ODPs. Further, work within the VALCRI project has increased the quality of ODP search engines (precision and recall) over established methods (as also implemented in the online search service https://github.com/hammar/XdpServices). This advance works on the basis of indexing and matching search queries against commonly occurring ODP metadata, and by exploiting the knowledge that ODPs tend to be generic solutions and that their implementations therefore include generic terms/concepts, whereas developers tend to pose questions that include quite a lot more specific terminology.

Explanation of why it is a significant advancement or ground-breaking

To the best of our knowledge, exploiting ontology debugging technology to improve quality already in the initial development phase of an ontology project, is not something that has been done before. Template-based ODP instantiation, while discussed as an idea in literature previously, has never been implemented by way of an ODP-independent generic algorithm in the way our work does.

The advances regarding roles and responsibilities for XD, ODP quality and documentation, and ODP search engines, are more incremental in nature. They take the next natural steps in adapting or improving upon existing technologies or ideas, sometimes by borrowing techniques from neighbouring fields. While there is certainly novelty to this work also that takes it beyond the state-of-the-art in its respective field, that novelty lies primarily in the application of ideas, rather than the generation of entirely new ones.

Published papers or patents to prove the above claims

Dragisic, Z., Lambrix, P., & Blomqvist, E. (2015). Integrating ontology debugging and matching into the extreme design methodology. In *6th Workshop on Ontology and Semantic Web Patterns (WOP 2015), Bethlehem, Pensylvania, USA, October 11, 2015*

Hammar, K., & Presutti, V. (2017). Template-based Content ODP Instantiation. In *Advances in Ontology Design and Patterns.* IOS Press

Hammar, K. (2017). *Content Ontology Design Patterns: Qualities, Methods, and Tools* (Vol. 1879). Linköping University Electronic Press.

Impact

The work relating to ODP support tooling forms the basis for two research grant applications (one submitted, one presently under development) which aim to explore and exploit these ideas further. The expressed goal of these applications includes to deliver near-product-quality software built on this work over the coming 24 months.

AGB16. RSP-SPIN Service and RDF Stream Processing architecture

Robin Keskisärkkä
Linköping University

Description of the significant Advances or Ground-breaking Science or Technology

Semantic Web technologies are used to represent and integrate the structured data in VALCRI, but this technology has historically focused primarily on data that is static or slowly changing. While modern RDF stores scale well with respect to the size of data, performance degrades as the rate of incoming data increases. Data streams are in principle unbounded and data becomes obsolete or loses relevance over time. New processing paradigms are therefore required to handle streaming data. For example, the ANPR data in VALCRI can be processed in a streaming fashion to identify patterns continuously rather than in batch.

RDF Stream Processing (RSP) systems provide an extension to the Semantic Web technology stack to support integration of streaming and static RDF data. Although the current RSP systems may appear very similar they are based on slightly different underlying semantics, provide different expressiveness given their respective query languages, and have distinct scalability characteristics. RSP standardization efforts are currently ongoing, and the RSP W3C Community Group is in the process of defining a new unified query language (RSP-QL).

RSP-SPIN was developed as an extension to SPIN based on this new query language, allowing RSP-QL queries to be represented as parameterized query templates. While no RSP engine currently supports RSP-QL it can be used to represent queries in some of the most popular RSP engines, including CQELS, C-SPARQL, and SPARQLstream, and the REST-SPIN API enables seamless transition between the different query languages.

In VALCRI, RSP functionality is exposed to other system components based on prepared query templates, which is exposed as a dynamically generated REST API. This hides much of the complexity from the users, including details about the particular RSP engine implementation being employed and the RSP language in use. The service API leverages ActiveMQ to internally stream data between components in a low-latency middleware.

Explanation of why it is a significant advancement or ground-breaking

There are several advances in this work with respect to state-of-the-art in this area. First, RSP-SPIN allows reusable RSP query templates to be expressed as RDF. These queries can be translated on the fly into some of the most common query languages. Second, the query templates are used to form the basis of an RSP architecture that is configurable directly from the service layer, which allows multiple RSP implementations to be used in parallel. Finally, the service layer allows access-control and security restrictions to be implemented based query templates. This is, to the best of our knowledge, the first work within the RSP domain that uses query templates to expose RSP functionality.

Published papers or patents to prove the above claims

Keskisärkkä, R. Towards Semantically Enabled Complex Event Processing. Linköping University, Department of Computer and Information Science. Licentiate thesis. 2017.

Keskisärkkä, R. Representing RDF Stream Processing Queries in RSP-SPIN. In: Proceedings of the ISWC 2016 Posters & Demonstrations Track co-located with the 15th International Semantic Web Conference (ISWC-2016), Kobe, Japan, October 17-21, 2016, CEUR Workshop Proceedings, 2016.

Keskisärkkä, R. Query Templates for RDF Stream Processing. In: Proceedings of Stream Reasoning Workshop. 2016 October 17th-18th, 2016, Kobe, Japan. Collocated with the 15th International Semantic Web Conference (ISWC 2016), CEUR Workshop Proceedings, 2016.

Keskisärkkä, R. and Blomqvist, E. Supporting Real-Time Monitoring in Criminal Investigations. In: ESWC 2015 Satellite Events, Revised Selected Papers, May 31–June 4, 2015, Portorož, Slovenia.

Keskisärkkä, R. and Blomqvist, E. Sharing and Reusing Continuous Queries – Expression of Interest. In: RDF Stream Processing Workshop Collocated with the 12th Extended Semantic Web Conference (ESWC 2015), May 31–June 4, 2015, Portorož, Slovenia.

Impact

No impact yet.

AGB17. REST-SPIN and Query Template Library

Robin Keskisärkkä, Henrik Eriksson, Eva Blomqvist
Linköping University

Description of the significant Advances or Ground-breaking Science or Technology

The Semantic Web technology stack is widely recognized for its ability to integrate and model complex data. RDF data forms the backbone of VALCRI's structured database, which uses ontologies and Linked Data principles to organize information. The standard interface for exposing data over the Web (i.e., HTTP) is to use SPARQL endpoints. The main advantage of SPARQL endpoints, as opposed to application-specific protocols, is that they form a universal method for issuing queries and receiving results, in a way analogous to JDBC for SQL database access from Java. However, SPARQL endpoints are very general, which means that application developers often have to build queries through string manipulation, which results in programs that are highly coupled with the underlying data model. In VALCRI, the data model is not only complex but it also evolves over time, causing the string manipulation strategy to scale poorly.

System architectures often introduce additional layers for handling data access for this very reason, while additionally providing data encapsulation and decoupling from application code. For example, a separate application server can act as the middle ground between the user-interface frontend and the data store. This approach helps structure the data-communication paths and provides data isolation. However, the development and maintenance of such intermediate layers can be costly and error prone, especially in the face of changes to the underlying data model.

In VALCRI, REST-SPIN provides an architectural pattern and model for generating APIs for accessing the structured database using query templates. The API is generated from a library of parameterized query templates, which forms a layer in front of VALCRI's structured database. This lets client components access data through a high-level API that is generated dynamically. This way of exposing data access is similar to the way in which some relational databases expose data through stored procedures; however, the implementation as a service layer adds an extra level of control to the operations performed on the database and decouples it from the specific database.

The layer of abstraction added by REST-SPIN allows access to data without requiring analysts and developers to have detailed knowledge of the specific data model, storage technology, or query language. While the preparation of query templates requires someone with in-depth understanding of these details, the parameterization of queries help make queries generalizable. For example, the patterns for retrieving crime reports are typically similar, even if the specific filter requirements differ, and a single query template can often be used in more than a single context. An important aspect of this is that the efforts invested in developing, validating, and optimizing queries can be leveraged even if the query library is managed completely independently from the VALCRI application code.

The service layer maintains compatibility with established techniques and formats used on the Semantic Web. Since it is not coupled with any specific RDF store it can be configured to work with any endpoint that supports SPARQL and SPARQL Update.

The service represents query templates using SPIN, which is the de facto standard for representing queries as RDF. Templates are defined on top of queries, allowing the query templates to be tagged and organized, while exposing the structure and content of the queries in a machine-readable format, rather than "hidden" as part of query strings.

Prior to the execution of a template, input values are validated against the template's parameter constraints and checked for potential query injections, which adds an extra layer of security. The service is also integrated with OpenPMF, which lets additional access control to be configured based on the exposed query templates.

Explanation of why it is a significant advancement or ground-breaking

There are currently three main methods for exposing RDF data using web services. The first is to expose SPARQL endpoints and allow users and client applications to submit any SPARQL query with minimal restrictions. The second method is to allow HTTP lookup of resources, which exposes a selection of the available data in relation to a particular resource or an entire dataset. The third method is to use dedicated APIs to access the data. However, each of these methods has drawbacks. Exposing SPARQL endpoints directly does not allow for tailored access restrictions, while inefficient or unresponsive queries are common due to users submitting computationally heavy or inefficient queries. Allowing only HTTP lookup requires the entire processing task (i.e., graph traversal to find the data needed) to be managed by the user. Dedicated APIs instead allow data-access restrictions and can support users in accessing data without requiring them to have detailed knowledge of the data model or underlying technology. However, creating good APIs is nontrivial and commonly involves information about the underlying data structure to be hardcoded.

REST-SPIN allows user-defined APIs to be generated dynamically based on a set of query templates, which are exposed according to best practices for REST services. The service builds on Semantic Web standards, supports content negotiation, and can be used to implement access-control on the level of query templates. The service exposes a basic administrator interface for adding, removing, and changing query templates to reflect changes in the underlying data model or in response to new data requirements.

Published papers or patents to prove the above claims

"Generation of REST APIs from SPARQL Query Templates" (*to be published*)

Impact

REST-SPIN provides secure access (using CAS and OpenPMF) to the structured data in VALCRI, for both retrieval and persistence of structured data. It is also used to provides query templates for indexing data in Elasticsearch, which exposes a portion of the structured data to the analyst interface. Many queries that drive the analyst interface are issued primarily against Elasticsearch, while complex queries and data persistence is managed using REST-SPIN. In particular, graph queries (e.g., all crimes associated with a particular nominal within two degrees of separation) are issued against the structured storage, as well as all queries that require access to all available data.

AGB18. Advances in the understanding of design and interpretation of data dense interactive graphics in crime analysis.

Roger Beecham[1], Alex Kachkaev[1], Jason Dykes[1], Chris Rooney[2], Aidan Slingsby[1],
Cagatay Turkay[1], Jo Wood[1], B.L. William Wong[2]
[1] CITY, University of London
[2] Middlesex University London

Description of the significant Advances or Ground-breaking Science or Technology

Our advances in the understanding of visualization in applied and experimental contexts include: theoretical frameworks, experimental results, new graphical methods, new ways of engaging with users in the visualization design process and software prototypes for crime analysis.

We established a conceptual framework and software prototypes for superimposing small multiple graphics of related data with different themes. This allows us to combine multiple perspectives on crime data, such as where, when and how crimes of different types occur in datasets of recorded crime, in interpretable ways. Our Faceted Views of Varying Emphasis allow analysts to relate these aspects of large crime data sets coherently and were shown to be effective in our experiments.

Crime data are frequently mapped as choropleths - maps showing counts or ratios for areas in which crimes are recorded. We tested and then modelled the effects of spatial autocorrelation (the degree to which phenomena are geographically related) and geometric configuration (the shapes, sizes and numbers of units in which geographic phenomena occur) on people's abilities to perceive differences in choropleth maps. Our models show that abilities to detect differences between maps are dependent upon autocorrelation and geometry and that these effects are predictable to an extent. This understanding of our visual abilities can inform both the way we map and our use of maps.

Our work with police analysts has resulted in a new approach through which visualization designers can engage with collaborators in exploratory visualization design. This new method is based upon the concept of Dynamic Design Documents - living documents that combine data, design and explanation that can be consumed and used by analysts in their place of work. We found this approach engendered a different level of engagement to that achieved through existing means of explanation and feedback elicitation. Our work shows that Dynamic Design Documents partially address certain intractable deficiencies common to evaluation in information visualization design and could be an effective focus for learning, with likely positive effects on the designs produced. The approach can be used in applied visualization work more widely than crime analysis. We continue to develop the technology through which the documents are created and expect to release a software framework to support their generation.

The Dynamic Design Documents have enabled us to develop a series of interactive graphics for crime analysts at West Midlands Police (WMP). Presenting and testing candidate designs in this way has enabled us to generate a family of Statistical Process Control Charts to support analysts in addressing recurring criminal intelligence tasks. By applying principles of information visualization to this well-established chart type in order to aid interpretation and by adding interaction and complimentary graphics, we have developed a series of graphics that are being used in VALCRI prototypes by analysts at WMP. Explaining design decisions with a clear rationale, enabling analysts to interact with data through the design candidates and capturing feedback from analysts through Dynamic Design Documents was key to achieving this suite of task-specific graphics.

Explanation of why it is a significant advancement or ground-breaking

The conceptual framework for superimposing small multiple graphics is the first documented description of this approach to visualization. In defining the design space for superimposed small multiples we open up opportunities for others to design and test multi-perspective visual approaches. These can be used for data sets that vary in in terms of geography, time and attributes beyond crime analysis.

Our ability to see differences in maps with measurably different characteristics has not been modelled previously, and we do so in some detail. Knowing how this works allows us, for the first time, to account for likely omissions in visual detection and construct a visual equivalent of statistical power for geospatial data. The results compliment those generated in recent years for correlation visualization, providing an empirical basis for improving the construction of visual line-ups for maps and developing theory to inform geospatial tests of graphical inference. Our paper on this issue received an Honorable Mention at IEEE VIS in 2016, an accolade reserved for the research that is of particular quality in terms of novelty, rigor and relevance. Less than 5% of submissions at the World's leading visualization conference receive this award.

Our development of Dynamic Design Documents compliments existing process models for visualization design by providing new technology to support communication between users and designers in ways that have been called for in the academic literature. It builds upon approaches that use data in the design process by supporting close dialogue between designers and analysts through data as designs develop. Viable methods to achieve this level of dialogue and interaction have not been reported in the visualization literature previously.

These contributions are all sufficiently new to science to be documented in publications in (and submissions to) the primary journals in visualization - the academic discipline dedicated to understanding the use of computer–based systems that use visual representations of data to help people carry out tasks more effectively. The results have informed and been adopted in VALCRI software and can be used in and applied to many visualization settings.

Published papers or patents to prove the above claims

Beecham, R., Dykes, J., Meulemans, W., Slingsby, A., Turkay, C. & Wood, J. (2016). Map LineUps: effects of spatial structure on graphical inference. IEEE Trans. on Visualization and Computer Graphics, 23(1), pp. 391-400. doi: 10.1109/TVCG.2016.2598862

Beecham, R., Rooney, C., Meier, S., Dykes, J., Slingsby, A., Turkay, C., Wood, J. & Wong, B.L.W. (2016). Faceted Views of Varying Emphasis (FaVVEs): a framework for visualising multi-perspective small multiples. Computer Graphics Forum: the international journal of the Eurographics Association, 35(3), pp. 241-249. doi: 10.1111/cgf.12900

One other paper submitted.

Rooney, C., Beecham, R., Dykes, J. & Wong, B.L.W. (submitted). Statistical Process Control Charts: A Visualization Makeover for Crime Analysis, IEEE Trans. on Visualization and Computer Graphics.

Two other papers are in production and expected to be submitted to journals by end of March 2018 and May 2018 respectively.

Impact

The Map LineUps paper (Beecham et al., 2016) received an Honorable Mention at the World's leading visualization conference – IEEE VIS (Baltimore, 2016).

AGB19. Furthering understanding and implementation of legal developments in the field of privacy and data protection and assisting in development of innovative legally compliant police technologies.

Fanny Coudert, Audry Delvaux, Thomas Marquenie, Ruben Roex, Peggy Valcke, Frank Verbruggen, Joyce Verhaert
KU Leuven CiTiP (KUL)

Description of the significant Advances or Ground-breaking Science or Technology

The accomplishments of KUL CiTiP relate to the general improvements in understanding and implementation of the recently adopted European Union framework on data protection in the area of law enforcement and criminal justice. During the first phase of the project, particular attention was paid to the analysis of these at the time ongoing reforms of data protection law. As the changes to the legal framework were revealed and completed during the project, KUL had the opportunity to follow these developments up close and collaborate with other research partners to be at the forefront of the analysis, understanding and implementation of the new European ruleset before integrations at the national level took place. At this stage, the ground-breaking advancements relate to the interdisciplinary collaboration with partners of technical, ethical, security and privacy-related backgrounds to conduct an analysis of substantial legal reforms and lay the groundwork for the development of future technologies in the sector of law enforcement and criminal justice in Europe. Additionally, KUL focused on the topics of data sharing and international transfers of data, criminal (procedure) law and police protocol, and fair trial safeguards. As such, the legal, ethical and privacy-related achievements culminated in several reports communicated to technical partners as well as general analyses of the legal framework that were translated into concrete LEP guidelines. The results of this research can be found in VALCRI Deliverables and Internal Reports mentioned below.

In the second phase of the project, the lessons drawn from the initial analysis of the legal framework were expanded upon and put into practice for the specifics of the VALCRI project. In cooperation with other partners such as ULD, MU, FHG, OS and LIU from the Security, Ethics, Privacy and Legal (SEPL) subgroup, several reports and white papers were drafted to explore the impact and consequences of the applicable legislation on the VALCRI project, and to research and present possible solutions and techniques to develop the system in a legally compliant manner. These reports include the analysis and impact assessment of the European data protection reforms as well as topics like data provenance, logging, anonymization, international data transfers, police protocol and data management policies for the VALCRI project. In the subsequent White Papers, more attention was paid to operationalizing transparency and LEP-conditions, addressing incidental discrimination, assessing fair trial concerns, working towards algorithmic accountability, and resolving certain issues relating to automated data processing in advanced systems like VALCRI. Additionally, KUL and the other partners within the SEPL subgroup collaborated with the technical developmental teams to incorporate legal norms within the VALCRI system. The technical achievements to which were contributed consist of, among others, the logging system, SEPL enforcement modules, access control measures, transparency techniques, and OpenPMF translations of rules into a machine-readable format.

In the third, final and currently ongoing phase of the project, methodological guidance is being drafted for the execution of data protection impact assessments (DPIA). Following the national transpositions of Directive (EU) 2016/680 with a due date in May 2018, these assessments will be legally required from law enforcement agencies (LEAs) using new technologies. As such, KUL is collaborating with ULD to draft guidelines for the VALCRI end-users that will enable them to successfully assess its impact in light of data protection law. These guidelines consist of three parts. The first covers an international comparison of different approaches to privacy impact assessments within Europe. This section analyses the French,

German, British, Belgian and Spanish methodologies to distil key elements and lay a groundwork for the development of new guidelines. The second consists of these new guidelines aimed specifically at law enforcement agencies. Finally, the third section contains a brief executive summary and a step-by-step overview of the necessary aspects of conducting such a DPIA. Furthermore, the general outline of the DPIA guidance is planned to be disseminated to the general public and the wider audience of LEAs for further use and standardization in a field that currently lacks comprehensive guidance on this topic.

Explanation of why it is a significant advancement or ground-breaking

The abovementioned research is ground-breaking for several reasons. First, the fact that the European data protection reforms took place during the VALCRI project and directly influence its outcome allowed KUL and the other SEPL consortium partners to be among the first to explore the impact and practical implications of Directive (EU) 2016/680. This Directive is the first to regulate data protection for law enforcement agencies and criminal justice authorities for domestic processing activities and includes a variety of new legal requirements with considerable technical and practical implications. In this light, the research delivered during the project was innovative by identifying the consequences of novel legal conditions and determining concrete and technical solution approaches to achieve legal compliance, as well as by looking at select national situations and policies on data protection and police law to assess their relevance and interplay with the European changes.

Second, KUL contributed to the development of new and innovative police technologies in a legally compliant way respectful of human rights and societal interests. Through the reports, white papers and publications mentioned later in this form, KUL and the larger SEPL group investigated the obstacles faced by these new technologies and sought to provide technical partners with guidance on how these advanced analytical systems can be developed to be fully effective and highly practical while still conserving human rights interests and containing necessary and important safeguards against misuse and potentially negative outcomes of system processes. While more research beyond the scope of the VALCRI project is still needed, it is our intention that the VALCRI research disseminated through the reports, white papers and the soon-to-be released DPIA guidelines shall continue to be available to the wider public and provide resources and lessons to other developers and technologists in similar situations.

Third, considerable amounts of research were conducted during the VALCRI project that contributed to the very recent and currently ongoing debate on the topic of algorithmic transparency and accountability. As a growing body of research is shedding light on the possible pitfalls, downsides and unintended consequences and advanced analytical systems, the mitigation of bias, discrimination and inaccuracies is at the current forefront of the debate on AI and automated decision-making in both Europe and the USA. This is of particular concern in the area of public governance, criminal justice and law enforcement, as the potential negative impact caused by the use of flawed technologies is undeniably significant in this field. As such, KUL has paid particular attention to the development of solution approaches for transparency, accountability and auditability in VALCRI and similar LEA systems in general. Through publications and public presentations on the topic (see below), its research conducted through VALCRI is among the first to present certain mitigation techniques and suggest technical solutions in line with the new data protection legislation.

Published papers or patents to prove the above claims

F. Coudert, Workshop: The Directive for data protection in the police and justice sectors: a significant step towards modern EU data protection?", 1 February 2016, KU Leuven University Belgium, available at: https://www.law.kuleuven.be/citip/en/calendar/item/old/workshop-lea-data-protection-directive.

F. Coudert, "The Europol Regulation and Purpose Limitation: From the 'Silo-Based Approach' to… What Exactly?", *KU Leuven CiTiP Blog*, 18-20 April 2017, available at: https://www.law.kuleuven.be/citip/blog/the-europol-regulation-and-purpose-limitation-from-the-silo-based-approach-to-what-exactly-part-i/ and https://www.law.kuleuven.be/citip/blog/the-europol-regulation-and-purpose-limitation-from-the-silo-based-approach-to-what-exactly-part-ii/.

F. Coudert, "The Europol Regulation and Purpose Limitation: From the 'Silo-Based Approach' to… What Exactly?", *European Data Protection Law Review*, Vol. 3, Issue 3, 2017, available at: https://edpl.lexxion.eu/article/EDPL/2017/3/6.

F. Coudert, C. George, E. Schlehahn & S. Mann, "Report on definitions and understandings of personal data, anonymization, and pseudonymization, based on Directive (EU) 2016/680", 2016.

F. Coudert, C. George & E. Schlehahn, "VALCRI International Data Sharing Transfers Report", Task 3.4 Legal and Ethical Aspects Analysis of European Regulatory Framework, 2016.

A. Delvaux, F. Verbruggen, R. Roex & J. Verhaert, "Legal Aspects" in *D3.4.1 Human Issues Framework Version 1.0*, 14 November 2014.

E. Kindt, "Having yes, using no? About the new legal regime for biometric data", *Computer Law & Security Review*, Article in Press, 2017, available at: https://www.sciencedirect.com/science/article/pii/S0267364917303667.

P. Malaquias, "Predictive Policing and the Need for a Public Security Exception under Copyright Law: The Belgian Example", *SSRN*, 2016, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2737456.

T. Marquenie, "The Police and Criminal Justice Authorities Directive: Data Protection Standards and Impact on the Legal Framework", *Computer Law and Security Review*, Vol. 33, Issue 3, 2017, available at: https://www.sciencedirect.com/journal/computer-law-and-security-review/vol/33/issue/3.

T. Marquenie, "Data analytics in a police context – addressing legal issues in VALCRI", *KU Leuven CiTiP Blog*, 14 March 2017, available at: https://www.law.kuleuven.be/citip/blog/data-analytics-in-a-police-context-addressing-legal-issues-in-valcri/.

T. Marquenie, F. Coudert, P. Duquenoy, P. Paudyal, "Roadmap for the Resolution of Ethical and Human Rights Issues in Automated Data Analysis and Extraction Computations in VALCRI", *VALCRI White Paper Series*¸ 2017.

T. Marquenie & F. Coudert, "Roadmap for the Operationalization of Legal and Privacy Requirements in VALCRI Analysis", *VALCRI White Paper Series*, 13 February 2017, available at: http://valcri.org/valcri/security-and-privacy-technologies-in-valcri/.

T. Marquenie & F. Coudert, "VALCRI - Data Management Policy in Belgium", 2017, unpublished.

T. Marquenie & E. Zouave, Participation at and co-organization of the Joint ASGARD, DANTE, TITANIUM, VALCRI, VICTORIA, and VOX-Pol Workshop on Data Protection Issues
in the Context of EU-funded Big Data Research Projects in the Security Domain, Dublin City University, 4 July 2017.

T. Marquenie & E. Zouave, "Speaking Truth to Computational Power: Coding Non-discrimination by Design in Police Technology", (presented at) *Next-Generation Community Policing Conference* (currently unpublished), 2017.

E. Schlehahn, P. Duquenoy, P. Paudyal, C. George, F. Coudert, A. Delvaux & T. Marquenie, "The Operationalization of Transparency in VALCRI", *VALCRI White Paper Series*, 1 January 2017, available at: http://valcri.org/valcri/white-paper-the-operationalisation-of-transparency-in-valcri/.

E. Schlehahn, A. Delvaux & P. Malaquias, "VALCRI Legal, Ethical and Privacy Guidelines", 2015.

E. Schlehahn, F. Coudert, A. Delvaux, C. George, P. Duquenoy, K. Xu, C. Anslow, E. Blomqvist, "Report on Provenance from LEP Perspective in VALCRI", 2016.

E. Schlehahn, F. Coudert, C. George, "The reform of the European legal data protection framework in the police and justice sectors and its relevance for the processing of personal data in the context of criminal investigation and intelligence in VALCRI", Task 3.4 Legal and Ethical Aspects Analysis of European Regulatory Framework, 2016.

A. Vedder, L. Naudts & T. Marquenie, Panel Discussion "Algorithmic Transparency and Accountability in Law Enforcement", organized by EU Projects VALCRI, DANTE & VICTORIA, hosted by A. Vedder, E. Schlehahn, J. Klerx, P. Duquenoy, C. Svanberg & E. Zouave, *CPDP Conference*, 26 January 2018, available at: http://www.cpdpconferences.org/assets/CPDP2018_PROGRAM_FINAL.pdf.

F. Verbruggen & A. Delvaux, "Crime Analysis and the Storage and Deletion of Personal Data: Holding the Belgian Legislation to the Europol 'Standard'" in *D3.4.2 Human Issues Framework Version 2.0*, 28 August 2015.

E. Zouave & T. Marquenie, "An Inconvenient Truth: Algorithmic Transparency & Accountability in Criminal Intelligence Profiling", *IEEE Conference Proceedings – Intelligence and Security Informatics Conference (EISIC)*, 2017, available at: http://ieeexplore.ieee.org/document/8240764/.

E. Zouave & T. Marquenie, "Countering Algorithmic Discrimination in Profiling", *KU Leuven CiTiP Blog*, 1 June 2017, available at: https://www.law.kuleuven.be/citip/blog/countering-algorithmic-discrimination-in-criminal-profiling/.

E. Schlehahn, T. Marquenie & E. Kindt, "Data Protection Impact Assessments (DPIAs) in the law enforcement sector according to Directive (EU) 2016/680 - A comparative analysis of methodologies", currently in progress.

Impact

The concrete impact from KUL's research in VALCRI is threefold. First, the research on the European data protection reforms has resulted in publications on and detailed analyses of the new Directive 2016/680. Its findings have contributed to the larger debate surrounding the new ruleset and its implementation. Among others, KUL's research has been requested by Greek scholars involved in the drafting of the upcoming Greek law to implement the new European rules on police data protection. Second, the identification of initial solution approaches and technical techniques to achieve legal compliance and mitigate human right concerns have helped shaped the project and contributed to the earliest adoption and realization of the novel legal framework applicable in this field. Third, it is the expectation that the data protection impact assessment (DPIA) guidelines and methodology shall be distributed and adopted to end-users and LEAs working with new technologies in a way that may contribute to the national compliance with Directive (EU) 2016/680 and improve data protection practices among European police actors.

AGB20. Toolset for Model-Driven Security and Privacy Protection

Patrick Aichroth, Sebastian Mann, Jens Hasselbach Fraunhofer IDMT (FHG)
Rudolf Schreiner, Ulrich Land, Karel Gardas, ObjectSecurity (OS)

Description of the significant Advances or Ground-breaking Science or Technology

Based on the requirements that resulted from the very productive collaboration between legal and technical partners within the SEPL subgroup, the project resulted in the development of several innovative technical SEPL solutions developed by OS and FHG: (a) a Model-Driven Security (MDS) framework based on OpenPMF, (b) OpenPMF-based access control implementations for VALCRI's structured database (AC4SDB) and unstructured database (AC4UDB), (c) high-assurance logging and auditing (HALA) and (d) Privacy-Enhancing Technologies, namely selective video encryption (VFE) and adaptive anonymization & pseudonymization (AAP):

- **Model-driven security and OpenPMF** (provided by OS) are used to edit and translate high-level, human-readable policies to low-level enforcement rules, thereby ensuring auditability, flexibility, and extensibility for the overall security and privacy protection approach based one common policy
- **Implementations** for fine-grain access control for the structured database (provided by OS), and access control for unstructured data such as video (provided by FHG), both of are controlled via OpenPMF / policies
- **High-assurance logging** (provided by OS) a FPGA-based implementation, ensures a strong separation of the user/admin and logging domain, to avoid log modifications and support ex-post auditing and misuse detection
- Components for **selective video encryption**, currently implemented for h.264 videos, which allow the selective encryption and decryption of individual regions within video material without affecting other parts of the video. Combined with video face detection, they can be used e.g. for semi-automatic privacy protection of CCTV video material
- Components for **adaptive anonymization and pseudonymization**, which use a definition of the functional requirements for data analysis, and apply a statistical re-identification analysis to identify information that is (potentially) person-identifying, thereby deriving an "optimal" anonymization approach for a given dataset and context, which is then processed using a set of anonymization and pseudonymization algorithms, to be used for privacy-aware data import and export.

Explanation of why it is a significant advancement or ground-breaking

To our knowledge, the aforementioned tools and approaches all went beyond the state-of-the-art: Model-driven security approaches and respective OpenPMF-based implementations for structured and unstructured data are new for the LEA domain, and to the best of our knowledge this has been the very first time that they have been used for to also support and enforce privacy (and to some extent, ethics) requirements. Similarly, the project also delivered the first high-assurance logging implementation of this kind, which in addition complements the policy-driven approach based on OpenPMF. Finally, the h.264 selective encryption and its combination with face detection is a new approach, as are the tools for adaptive "optimized" anonymization and pseudonymization.

Published papers or patents to prove the above claims

- Patents and patent applications US8856863B2, US14466382, US15656393
- Chr. Gerhardt, P. Aichroth, S. Mann; "Selective Face Encryption in H.264 Encoded Videos", IEEE Visual Communications and Image Processing (VCIP), 2017

Impact

While some of them will require further R&D efforts for improvement, the aforementioned technologies provide a power toolset to deal with various security and privacy challenges related to data confidentiality, auditing, content sharing, and consideration legal, privacy, and ethical requirements in the LEA domain, and potentially many other domains. Most importantly, it is not a set of technologies with "hard-coded" rules, but based on a policy-driven approach that allows extension and modifications, also considering nation and regional differences regarding requirement.

AGB21. Methods for Discovering and Mitigating Cognitive Biases in a Visual Analytics Environment

Name of person(s) to attribute the advances:
Partner Organisation:

Description of the significant Advances or Ground-breaking Science or Technology

The development and application of new knowledge and information technologies leads to a situation where analysts have to constantly make sense of large amounts of unstructured data. Visual Analytics methods provide substantial support in their reasoning and judgement process. However, these mental processes can also lead to thinking errors, so-called cognitive biases. The VALCRI project has addressed the problem of cognitive biases (1) by identifying relevant cognitive biases in the context of criminal analysis, (2) by developing methods to discover and measure cognitive biases, and (3) by developing approaches to mitigate cognitive biases.

In a structured process eight cognitive biases have been identified that are specifically relevant to criminal analyses in the context of visual analytics environments. This process was based on a classification of cognitive biases developed in the RECOBIA project, which resulted in 288 different cognitive biases. By taking into account the results of an extensive literature review and the VALCRI requirements analysis, eight cognitive biases could be identified as most relevant. These are the confirmation bias, anchoring and adjustment effect, clustering illusion, framing effect, availability heuristic, base rate fallacy, selective perception, and group think.

In order to discover, if a cognitive bias is involved in the reasoning process, the particular cognitive bias needs to be operationalised. In VALCRI several operationalisation methods have been developed. First, empirical experiments have been designed and implemented that can detect the confirmation bias and the clustering illusion. Second a general method for operationalising cognitive biases has been elaborated that are based on patterns of cognitive processes. This method is based on both theoretical elaborations of cognitive processes and empirical observations in task-based studies. These method has been tried out in an evaluation study and serves as a basis for the operationalisation of further cognitive biases. Third, a data-driven method has been elaborated that takes into account user interactions to detect the occurrence of cognitive biases while analysts use a VA system and solve tasks. This method is based on the comparison of biased and unbiased interaction patterns.

In order to mitigate cognitive biases, a focus was put on the development of design guidelines for visual analytics systems (instead of training or prompting the users). For example, this set of guidelines include the recommendation that data should be presented from different perspectives (visualisation techniques, multiple views), which causes the the analyst to think differently. Another example concerns the recommendation to include uncertainty information, which causes to deeper reflect the information. Due to

limited resources, only few of them could be taken up by the VALCRI system development. In addition to the guidelines, an evaluation method has been proposed how to critically evaluate if a system supports bias mitigation. This method suggests to let experts critically analyse each tool and feature of a system with regards to each relevant cognitive bias. Such an analysis results in a matrix of tools and cognitive biases. A preliminary analysis of the final system has been undertaken, which demonstrages strenghts and weaknesses of the VALCRI system.

Explanation of why it is a significant advancement or ground-breaking

The concern of cognitive biases is a well known problem in psychology and the intelligence sector (and other fields of science and practice). Though a vast body of literature exists that deals with cognitive biases, most of it treat cognitive biases on a theoretical level. The work in VALCRI included several steps forward towards methods for measuring and mitigating cognitive biases.

In terms of measuring / discovering cognitive biases, to the best of our knowledge only one method is known, which is the Selective Exposure Experiment to measure the confirmation bias. Thus our elaborated methods extend the use of state-of-the-art of measuring cognitive biases on several dimensions. First a new empirical method has been developed to measure the Clustering Illusion bias, which was not available in literature before. Second, the method to measure cognitive biases through a classification of cognitive processes and assigning them in a structured observation constitutes a new approach in this field. This provides a basis for developing further cognitive bias operationalisations. Third, the data-driven approach outlines a method to detect cognitive biases based on user interactions with a visual analytics system. All these methods outlines new directions how cognitive biases can be measured, consisting of empirical studies, observations through a person, and automatic observations through a logging system.

In terms of mitigating cognitive biases methods and guidelines for mitigating cognitive biases have been elaborated. Though these guidelines are based on existing ideas in literature, the innovation lies in in the translation of these ideas to the design of visual analytics components. Furthermore, the tool-bias analysis method provides a new approach to critically evaluate a visual analytics system according their potential inducements and mitigation of cognitive biases. This method allows for both formative and summative assessment of a VA system. To our knowledge this also constitutes a new way of evaluating a system with respect to its mitigation capabilities of cognitive biases.

Due to time and resource limitation, these new methods have not been fully exploited. Many of them have been tried out in a preliminary way. However, they have not been fully applied on the project.

Published papers or patents to prove the above claims

Bedek, M., Nussbaumer, A., Huszar, L., & Albert, D. (2017). Discovering cognitive biases in a visual analytics environment. In Proceedings of the Workshop on Dealing with Cognitive Biases in Visualisations, held at the VIS 2017 conference. Retrieved from
http://decisive-workshop.dbvis.de/?page_id=555

Nussbaumer, A., Verbert, K., Hillemann, E. C., Bedek, M. A., & Albert, D. (2016). A framework for cognitive bias detection and feedback in a visual analytics environment. In J. Brynielsson & F. Johansson (Eds.), Proceedings of European Intelligence and Security Informatics Conference (EISIC 2016) (p. 148-151). IEEE. doi:10.1109/EISIC.2016.038

Hillemann, E.-C., Nussbaumer, A., & Albert, D. (2015). The Role of Cognitive Biases in Criminal Intelligence Analysis and Approaches for their Mitigation. In Proceedings of the European Intelligence and Security Informatics Conference (EISIC 2015), pp. 125-128. IEEE. doi: 10.1109/EISIC.2015.9

[three publications will be submitted soon]

Impact

No impact could be observed so far. We expect that the methods will be taken up by future projects and visual analytics implementations.

AGB22. Sense-making in intelligence analysis and development of a framework of recommendations for the design of such systems

Johanna Haider-Doppler1, Margit Pohl1, Neesha Kodagoda2, Patrick Seidler2
1Vienna University of Technology
2Middlesex University London

Description of the significant Advances or Ground-breaking Science or Technology

We achieved a significant advance in analysing the sense-making and reasoning processes users engage in when they interact with visual analytics systems. We were able to provide a detailed description of these processes and we could identify several such sense-making strategies. We tried to identify fairly general cognitive strategies and how they are used. Nevertheless, these cognitive strategies are still based on intelligence analysis as an activity. We also related the strategies we found in empirical investigations with theoretical approaches from the scientific literature. An important foundation for such research are studies in everyday thinking and reasoning, naturalistic decision making and sense-making. While there is already a growing body of research in this area, such processes are still not very well understood. Especially in the area of intelligence analysis which is an open-ended and explorative process there is still too little research to inform the design of visualisation systems. In addition, we developed a framework of guidelines or recommendations for the design of visual analytics systems. This framework of guidelines is based on the scientific literature in this research area. It was successfully applied in the design of the VALCRI system.

Explanation of why it is a significant advancement or ground-breaking

There is almost no research concerning these issues. There is a lot of research concerning cognitive biases. Such research focuses on the negative aspects of reasoning processes and how to overcome them. In contrast to that, we focus on the question how successful reasoning and sense-making processes work. Research indicates that sense-making processes in general work fairly well and that cognitive biases are less common in practice than the research in this area suggests. Users usually have a lot of background knowledge helping them to use efficient strategies to solve problems. They often use context information to make sense of the data they have to analyse. In the long run, such research can also help to develop a comprehensive model of open-ended and explorative reasoning processes. Such a model does not exist so far. Especially the research in cognitive biases has not been able to develop such a model.

This kind of research is especially important to support designers to design usable and useful systems. When we know what kind of strategies users adopt (e.g. when they use verification strategies) we are able to support such strategies with specific visualisations (e.g. multiple views). So far, such decisions have been made based on the intuition of the designers. Research in that area can help to develop an empirical foundation for such design decisions.

Published papers or patents to prove the above claims

M. Pohl, J. Haider: "Sense-making Strategies for the Interpretation of Visualizations-Bridging the Gap between Theory and Empirical Research";
Multimodal Technologies Interact 2017, 1 (2017), 3; 21 S.

J. Haider, M. Pohl, E. Hillemann, A. Nussbaumer, S. Attfield, P. Passmore, W. Wong: "Exploring the Challenges of Implementing Guidelines for the Design of Visual Analytics Systems"; Human Factors and Ergonomics Society 2015 International Annual Meeting, Los Angeles; 26.10.2015 - 30.10.2015; in: "Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting - 2015", (2015), S. 259 - 263.

J. Haider, M. Pohl, C. Pallaris, W. Wong: "Supporting Sense-Making and Insight Processes in Visual Analytics by Deriving Guidelines from Empirical Results";

International Summer School on Visual Computing 2015, Rostock, Germany; 17.08.2015 - 21.08.2015; in: "Proceedings of the Internation Summer School on Visual Computing 2015", (2015), ISBN: 978-3-8396-0960-6; S. 59 - 68.

J. Haider, P. Seidler, M. Pohl, N. Kodagoda, R. Adderley, W. Wong: "How Analysts Think: Sense-making Strategies in the Analysis of Temporal Evolution and Criminal Network Structures and Activities"; HFES 2017, Austin, Texas, USA; 09.10.2017 - 13.10.2017; in: "Proceedings of the HFES 2017 conference", SAGE, 61/1 (2017), ISBN: 978-0-945289-53-1.

S. Kriglstein, J. Haider, G. Wallner, M. Pohl: "Who, Where, When and with Whom? Evaluation of Group Meeting Visualizations"; Vortrag: 9th International Conference on the Theory and Application of Diagrams (Diagrams 2016), Philadelphia, PA, USA; 07.08.2016 - 10.08.2016; in: "Diagrammatic Representation and Inference", Springer, Lecture Notes in Computer Science (2016), ISBN: 978-3-319-42332-6; S. 235 - 249.

P. Seidler, J. Haider, M. Pohl, N. Kodagoda, R. Adderley, W. Wong: "Design for Intelligence Analysis of Complex Systems: Evolution of Criminal Networks"; European Intelligence and Security Informatics Conference EISIC 2016, Uppsala, Schweden; 16.08.2016; in: "European Intelligence and Security Informatics Conference Proceedings 2016", Uppsala, Sweden (2016).

Impact

Based on this research it is possible to develop more refined design recommendations how to develop systems that support intelligence analysts in their daily work. These recommendations can guide developers how to design efficient systems. We provide some first results in that area.

AGB23. Analysis of the new data protection framework for the Law Enforcement Agencies (LEA) sector in Europe to determine technical and organisational solution approaches for achieving legal compliance.

Eva Schlehahn, Marit Hansen, Harald Zwingelberg, Daniel Deibler, Julia Victoria Pörschke, Andreas Schlisske
Unabhängiges Landeszentrum für Datenschutz (ULD, Engl. Independent Centre for Privacy Protection), Schleswig-Holstein

Description of the significant Advances or Ground-breaking Science or Technology

During the first half of the project, the reform of the European data protection law was underway. By May 2018, the Directive (EU) 2016/680 determines the relevant legal framework for the processing of personal data in the EU LEA sector. Its national implementations in the individual EU Member States are still mostly pending with open outcomes. The new legal preconditions are very novel in the LEA field, which means that ULD's work is initial groundwork of legal and practical analysis, predictably reaching far beyond VALCRI. ULD worked in close, interdisciplinary collaboration with other researchers in the project to see how the data protection requirements can be realized in VALCRI within a holistic approach across the professions (e.g. technical, ethics, and criminal law).

Moreover, the input of the project's Ethics Board was taken into account as well.

The findings of the research were captured in internal reports from the LEP (Legal, Ethical, Privacy) partners together (KUL, ULD, MU). Furthermore, some tech partners (namely FHG, OS, LIU, and MU) from the SEPL (Security, Ethical, Privacy, Legal) and DMO subgroups were involved in this work as well to support and give advice during the development.

These reports are:

- The reform of the European data protection framework and its impact on personal data processing in VALCRI (ULD, KUL, MU)
  - to assess the relevance of the Directive 2016/680 for VALCRI
  - to identify issues + solution approaches with regard to the UK police end users and the factual and legal impact of the Brexit situation
  - to identify issues + solution approaches with regard to the current reform of the EU export control policy framework affecting potential dual-use items
- Report on Provenance from LEP perspective in VALCRI (ULD, KUL, LIU, MU)
  - to identify and recommend approaches to achieve transparency, chain of custody and auditability of the VALCRI prototype and individual components
  - to identify provenance ontology requirements from LEP perspective
  - to map LEP requirements to solution suggestions of the technical partners to cover data, process and reasoning provenance
- Report on Logging in VALCRI (ULD, OS)
  - to identify and recommend approaches to realize Directive (EU) 2016/680 requirements for logging and auditability in VALCRI
- Report on personal data, anonymization and pseudonymization in VALCRI (KUL, ULD, MU, FHG)
  - to identify and assess beyond-state-of-the-art advances of data protection by design and by default approaches in VALCRI

- Report on international data transfers (Data Sharing Report)
    - to analyze the applicable EU legal framework for the exchange of intelligence within EU and to identify the resulting technical requirements for VALCRI (KUL, ULD)

In progress is a methodology guidance for data protection impact assessments (DPIA) for police analytics systems in compliance with Directive 2016/680. The goal is to provide this guidance to the VALCRI police end users as a DPIA tool when deploying a system like VALCRI. A further intention is to disseminate the DPIA guidance to a wider audience of potential LEAs to equip them with data protection expert knowledge in this field.

Explanation of why it is a significant advancement or ground-breaking

The VALCRI project started already while the reform of the European data protection framework was still in progress. The anticipation and identification of relevant requirements has supported an utmost early adoption of the new legal preconditions and helped determine potential technical solution approaches for some of these. Therein, selected and VALCRI-relevant technology approaches were analysed from LEP perspective to determine specific functional requirements, which could be addressed by the technology partners in the project. These approaches were mainly:

- Logging
- Anonymisation & pseudonymization
- Provenance
- Access control enforcement

However, more research work would be needed beyond VALCRI to facilitate fully functional, effective and compliant solutions in an integrated police analytics system prototype.

To enable this for the future, ULD conducted groundwork by doing a conceptual mapping of diverse elements as listed below:

Mapping conducted of:

<table>
<tr><td>the LEP Guidelines drafted in VALCRI</td><td>to -></td><td>the requirements of Directive 2016/680</td></tr>
<tr><td colspan="3">➔ To prove that the LEP guidelines drafted at the beginning of the VALCRI project correctly anticipated the legal requirements of the new legal framework which was still underway at this time.</td></tr>
<tr><td>The requirements of Directive 2016/680</td><td>to -></td><td>abstracted data protection goals</td></tr>
<tr><td colspan="3">➔ Those data protection goals are used as evaluation criteria for data protection impact assessments in the German Standard Data Protection model (SDM). The SDM is officially acknowledged by the German data protection authorities and proposed by ULD and others as basis for a DPIA methodology. The SDM is listed as one possible DPIA approach by the Article 29 Working Party and being used as such in Germany and beyond.</td></tr>
<tr><td>the data protection goals</td><td>to -></td><td>concrete technical and organizational realization measures</td></tr>
<tr><td colspan="3">➔ To bridge the gap between law and technology and to show in which way legal concepts and principles may be realized via concrete measures.</td></tr>
</table>

Furthermore, mapping of all of the above	to VALCRI-specific	(1) o User Stories o Use Cases o Misuse Cases o Components o Technology Enablers

The above mentioned methodology guidance for data protection impact assessments (DPIA) builds upon this previous conceptual work. Once finalized, it may equip LEA as data controllers with the needed knowledge how to conduct a DPIA in compliance with Directive 2016/680 when they intend to deploy police analytics system like VALCRI. Moreover, the DPIA conducted on this basis may support the implementation of suitable technical and organizational data protection measures and enable the LEA's to show compliance with the legal preconditions.

Published papers or patents to prove the above claims

Most of the publications made to guide the development processes in VALCRI were internal, i.e. confidential as not to disclose sensitive information about VALCRI components or police end user's work, systems, tools and strategies. This non-public work was already listed above under section 2.

Notable public papers involving ULD are the LEP White Paper '*The Operationalization of Transparency in VALCRI*' (ULD, MU, KUL) addressing the corresponding VALCRI vision item Transparency and the essay '*Benefits and pitfalls of predictive policing*' (ULD, FHG, OS, for the EISIC 7-9 September 2015, Manchester, UK).

Currently in progress is a public guidance on data protection impact assessments for the use of police analytics systems in compliance with Directive (EU) 2016/680, which ULD crafts together with KUL.

Impact

- Identification and concretization of legal requirements of Directive (EU) 2016/680 and related issues specifically in the domain of police analytics systems
- Determination of initial solution approaches in technology and identification of further research potentials in that area -> fosters an early adoption of the new legal framework applicable for the LEA sector
- Development of a DPIA methodology -> distribution to LEA end users may positively affect compliance with Directive (EU) 2016/680 and corresponding implementing Acts on national level

AGB24. The Master Analyst: A Reference Curriculum for Law Enforcement Professionals

Aleksandra Bielska, Cristina Viehman, Chris Pallaris
i-Intelligence, Zurich

Description of the significant Advances or Ground-breaking Science or Technology

The Reference Curriculum provides a detailed index of modules to support the training and development of criminal intelligence professionals. These modules are intended to enhance individual and organisational performance and include such disciplines as: analytic reasoning, problem solving, human cognition, the management of the analytic function, legal, ethical and privacy considerations, and research and investigative skills. For every module we have defined the practical and theoretical knowledge to be covered.

- The Curriculum is intended for use by:
- Organisations wishing to improve staff training
- Universities looking to develop applied training programs
- Individual analysts hoping to identify gaps in their own knowledge and abilities

The Curriculum's structure and composition reflects our discussions with the VALCRI project's end-users, as well as other security and law enforcement practitioners. It looks to address existing gaps in intelligence training, while also extending analysts' knowledge of the state-of-the-art in analytic reasoning. It is also intended to alert analysts to the ecology of knowledge needed to operate effectively in contemporary security environments.

Finally, the Curriculum is intended to support the development of what we refer to as "Master Analysts". Such analysts can work both in an analytic and advisory capacity and can address challenges that are internal or external to the organisation. Put differently, the Master Analysts is a reflexive practitioner, one capable of analysing an issue, as well as how that issue should be analysed. We posit that expanding the scope of an analyst's training is critical to developing a workforce that can serve in a multi-role function and generate greater value for their organisation.

Explanation of why it is a significant advancement or ground-breaking

Traditional approaches to intelligence training are built around the Intelligence Cycle, a five-step process model that focuses on requirements planning, collection, processing, analysis and dissemination. This model fails to reflect the range of activities that analysts engage in on a daily basis.

The Reference Curriculum looks to address these shortcomings by arguing that intelligence work spans five separate but interconnected domains:

- Organisational: Tasks pertaining to organisational development, environmental analysis, strategic planning, etc.
- Operational: Tasks pertaining to the management of policies, processes and projects
- Informational: Tasks pertaining to the use and management of information, including data quality, records management, etc.
- Technological: Tasks pertaining to the use and management of one's IT tools, including analytic software
- Cognitive: Tasks pertaining to the cognitive dimensions of analytical work

Acknowledging the scope of activity allows us to better prepare analysts for the work they will do, not just as the start of their career but at every stage thereafter. The Reference Curriculum extends the state of the art by broadening the range of skills and literacies to be taught to criminal intelligence professionals. Second, it addresses the human factors and organisational dynamics that impede or enable analytic performance; such factors are routinely ignored by existing intelligence programs. Third, it encourages analysts to transcend the narrow definitions of their role and find ways to provide greater value to their organisations. Finally, it provides detailed listings on the theories, models and frameworks that analysts can use to enhance individual and organisational performance.

Published papers or patents to prove the above claims

Bielska, Aleksandra, and Pallaris, Chris. “The Psychology of Intelligence Analysis: Where Are We and Where Should We Be?”. White paper submitted to the National Academies of Sciences, Engineering, Medicine, July 2017.

Bielska, Aleksandra, and Pallaris, Chris. “Redefining the Intelligence Skill Set by Use of the Intelligence Analysis Impact Model”. Paper presented at the International Studies Association Annual Convention, Baltimore, MD, United States, February 2017.

Bielska, Aleksandra, and Pallaris, Chris. “Addressing the Internal Challenges to Intelligence Work”. Journal of Mediterranean and Balkan Intelligence, Vol. 10, No. 2 (2017), pp. 89-102.

Bielska, Aleksandra, and Pallaris, Chris. “Toward The Understanding of Uncertainty in Intelligence Analysis”. Paper to be presented at the International Studies Association Annual Convention, San Francisco, CA, United States, April 2017.

Impact

We held a series of workshops with law enforcement practitioners during the last 18 months of the VALCRI project. These workshops were intended to test our assumption that analytic performance could be improved if we gave analysts training in those disciplines that do not fit into the traditional intelligence cycle. The feedback generated from these workshops confirmed our hypothesis and the value of developing a Reference Curriculum. The analysts who participated felt a renewed sense of purpose and a greater awareness of their value and potential contribution. As one analysts commented: "...We are more than the [organisation] currently allows us to be. We can add a lot more value than we currently do." Participants were grateful to discover tools to help them address the impediments to analytic performance and improve their ability to meet stakeholder requirements.

These findings of these workshops have since informed our approach to the training and development of analysts working in other fields. We are working to develop new training programs for clients in both the public and private sector that reflect the systems approach advocated by the VALCRI curriculum.

AGB25. Visual Analytics for Federal Police of Antwerp

Nick Evers, Christophe Vandenberghe, David De Weerdt, Rani Pinchuk
Space Applications Services (SPACE)

Description of the significant Advances or Ground-breaking Science or Technology

While the VALCRI platform aims at complex and advanced Visual Analytics software that is intended to help police analysts, it cannot face, within the given project duration and budget, two main challenges:

Bringing the software to high TRL, allowing it to be useful in the day to day activities: the VALCRI team includes great researchers and software engineers but a research project that aims for flexible and dynamic development cannot aim for high TRL. The flexibility - allowing to try out ideas - is a core principle in a research environment. This flexibility, however, challenges its robustness in the field. Integration tests are a good example of this. On the one hand they would increase the overall stability a lot but, on the other hand, they would make the code very rigid and difficult to change and thus hamper research.

Deep integration with the data and infrastructure of police forces and addressing their specific needs: one of the biggest challenges of any information technology project is to engage with the end-user's (usually legacy) data. Not only is the data organized in different ways (that are not trivial to understand), many times it has various different semantics which influence the end user needs and the software itself. Moreover, understanding the semantics of the data is, many times, a very demanding task.

To overcome these issues and with the goal of pursuing commercial exploitation, SPACE engaged with the Federal Police of Antwerp, aiming to implement a high TRL software that is still in line with the overall VALCRI ideology but taking into account the real data, Flux24, and its specific challenges.

Flux24 is a data set that federates police reports over all the local police forces in the region of Antwerp. In practice, Flux24 provides dump of the latest changes in each of the police zones every day at 6AM in a MS Access format.

The work for the Federal Police of Antwerp entailed many face to face meetings with the police analysts. By following the Lean Software Development approach, we were able to focus on the most pressing problems related to the real data.

Figure 2 - The Lean Software Development approach applied to the user interface (UI). Over each iteration from left to right, the UI got more (re)defined

The high TRL software developed includes:

- A configurable ingester, which ingests the Flux24 data. Additionally, it can cope with unstructured data sources such as PDFs and most other common text formats and it extracting text from images using OCR (Optical Character Recognition).
- A user interface that includes several inter-related analytical components:
- Fact card - showing aggregations and counts of attributes from the data resulting into factual information about different facets of the data.

- Suggestion card - machine suggestive aggregations based on tf-idf (term frequency-inverse document frequency) intending to reflect the most important words in the corpus of the result set.
- Timescatter plot - a visualization that shows the relation between the day of the week and the moment or hour of the day. It aggregates police reports over both dimensions and reveals hot-spots in time.
- Timeline card - provides a temporal aggregation over the selected time range.
- Map - geographically aggregates police reports using automatic geocoding of addresses in the reports in order to reveal hot spots on a map. It also automatically realigns and animates the map boundaries and zoom level to focus on the search results.
- Police analysts can export a result set and save queries

Each of the cards updates itself according to the overall search context as defined in the search bar. The search input supports various search strategies that are derived from day-to-day problems and needs. It includes smart-case searching, fuzzy searching, code searching and others.

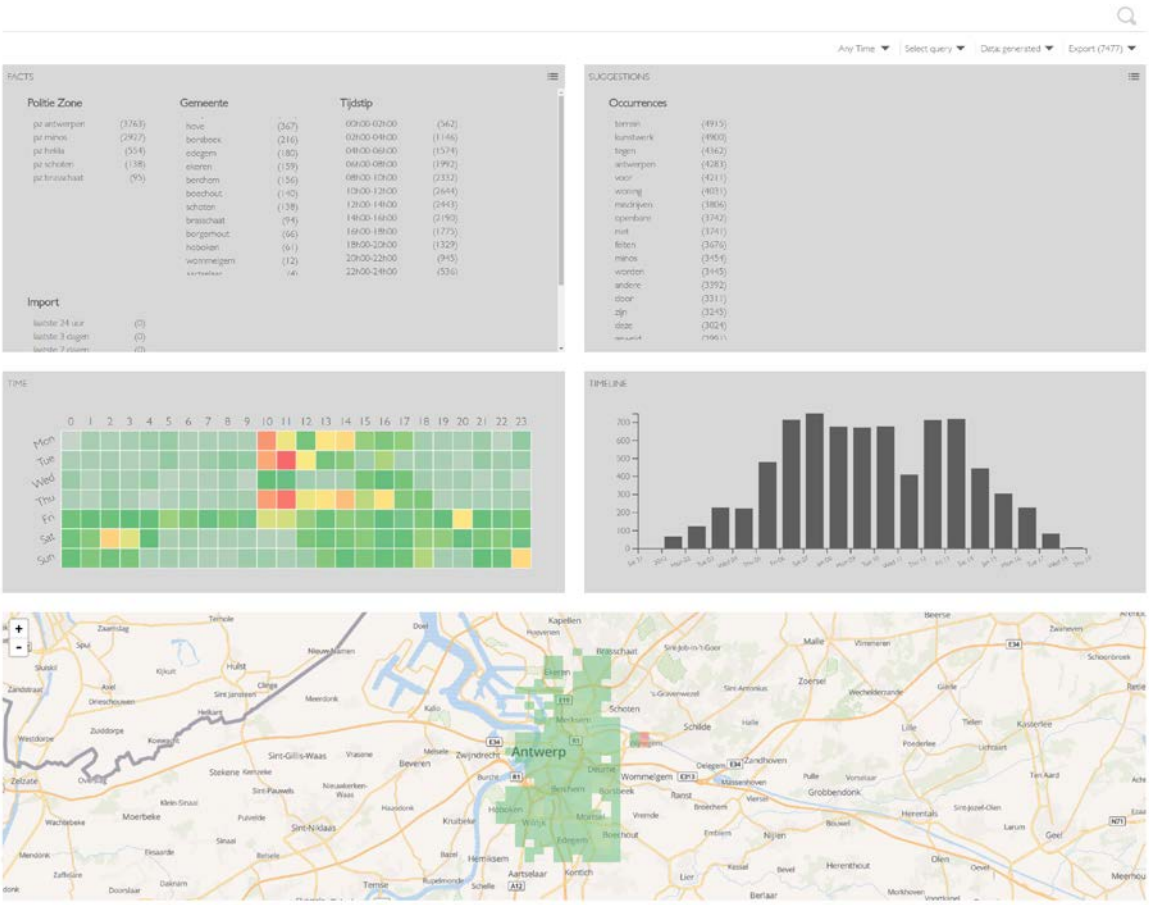


Figure 3 - The UI for the Federal Police of Antwerp (over faked realistic generated data)

To allow the Federal Police Analysts to use the software, security facilities had to be added to the software. This included encryption, authentication, firewall integration and usage auditing.

In order to support the development of the software and its testing, a generator allowing to generate large amount of faked yet realistic data has been developed as well. In addition, a fuzzy integration testing environment has been set up.

Explanation of why it is a significant advancement or ground-breaking

Some of the ground breaking achievements related to this activity are described separately. We, therefore, detail here only about the Configurable ingester and the lean approach.

Ingester

All of the analytical capabilities of the system rely on the proper pre-processing of the incoming data. One often overlooks or underestimates the value of the of the system components that perform this task. The ingester was shaped and remodelled many times over each of the development sprints. Each time its flexibility was challenged by the complex pre-processing steps that had to be applied to the Flux24 data. Over time, the ingester modularised into a generic and reusable system component that could run any analytical processing we offered it. Moreover, each of those components can be reconfigured, enabled, disabled or rewired simply without changing a single line of code.

Lean approach

In most research projects done under the framework programs, the user requirements are done in a separated WP and as a well-defined phase in the start of the project. In VALCRI we have followed the this very same approach. However, later on, SPACE engaged with the Federal Police of Antwerp in a different process - the lean approach. Interestingly, we found that the two approaches not only led to different needs, but also to a very different level of interaction with the end users during the on-going development.

Two key principles were followed. The first principle is the lean approach itself. This way of working allowed us to expose interesting features and carefully develop them together with the users from the stage of a paper mock-up and up to the high TRL final implementation. The short development cycles (ten days per sprint) embedded a feedback stage, planning stage, development stage, quality assurance stage and release stage (see the figure below). Each sprint is designed to keeps the user's need in the focus of development whilst allowing the user to adapt the requirements to his real needs.

Figure 4 - Development cycle stages

The other key principle is running the software over real and relevant data. In many occasions police analysts communicated to us that a certain application or component looks "interesting" or "exciting" but they cannot give their real opinion over that software without seeing in it the data they use every day. In some cases, the police analysts are not even sure if their data could be applied to the given component. This is mainly due to the very different data structures used by different law enforcement organizations. The lean approach bridges this gap as it allows presenting ideas from the research into working software that operates on the data end-users are familiar with.

Published papers or patents to prove the above claims

Not applicable

Impact

In real-life cases, the Police Analyst immediately found the information he was looking for. The details of this cannot be disclosed due to their sensitive nature.

Potential for commercial exploitation is being explored.

AGB26. Fuzzy integration tests over changing data

Nick Evers, Christophe Vandenberghe, David De Weerdt, Rani Pinchuk
Space Applications Services (SPACE)

Description of the significant Advances or Ground-breaking Science or Technology

When performing analysis over real-time data, the analyst has to be provided with the freedom to investigate and navigate through the data depending on the current context. In turn, the context is highly dependent on the cases being studied and what questions the analyst is trying to answer. Therefore, the system has to support a large variety of possible investigation approaches and techniques. Analysts are faced with two main problems:

1. The data is inherently incomplete and can only be seen as a subset of the whole domain. The missing links are most often the answers the analyst is looking for.

2. The semantic evolution of the data ingested by the system. Each piece of ingested data is a snapshot of an event and its context as perceived by its author. Its meaning over time may have been lost or changed.

For these reasons, analysts require a flexible set of techniques to analyse and navigate through the data, such as partial/strict matching, smart-case mating, code matching (e.g. phone numbers), fuzzy searches, suggestive linking, ...

In essence, many of these features are non-deterministic over time. As a result, it is challenging to prove their correctness using standard testing techniques. That is why we have integrated and are actively developing a testing framework that operates based on probabilities rather than facts. Using probabilities allows a feature and/or a set of features to be tested for correctness and usefulness.

So far, we have integrated probability tests in our testing pipeline by ingesting specific sets of data that mimic use-cases defined by analyst end users on which the testing pipeline performs a specific set of search and analysis features. Each probability test must pass a certain probability score for the pipeline to succeed. Currently, we are researching and developing a way of generating meaningful data in real-time to be used as input data set for the probability tests. The challenge is to generate data over time that remains of significant relevance to represent the domain in which the system is used. The big advantage of this process is to allow continuous evaluation of the system and determine if the analysis features provided by the system are still useful.

Explanation of why it is a significant advancement or ground-breaking

Traditional tests validate whether a certain feature is correct or not. These tests prevent specific features from being broken during further development. The difference with probability testing is that they attempt to assess if a feature and/or a combined set of features is still valuable for an analyst to use during investigation.

In other words, certain features may evolve and adapt over time and in the end not provide the same outcome for an analyst when used. However, such features can still provide results that help the analyst to find answers. Traditional tests are monochromatic and would consider a feature that does not provide the same result over time as being broken. The probability tests are more opinionated and try to assess the relevance of such features over time.

The results of integrating probability tests give a good indication to development whether a set of feature have evolved correctly and can still be considered as useful. Therefore, probability tests can be used to

determine when the system needs to be adapted, specially when it has been operationally active for a long period of time and there is no more active development.

Published papers or patents to prove the above claims

Not applicable

Impact

Being able to consistently test different use-cases while the system evolves is a mandatory requirement to avoid regression and provide stability to the system. Using a fuzzy integration test framework improves development as it dramatically decreases the required involvement of end users to test and validate the system.

AGB27. Generators of faked yet realistic data.

Nick Evers, Christophe Vandenberghe, David De Weerdt, Rani Pinchuk
Space Applications Services (SPACE)

Description of the significant Advances or Ground-breaking Science or Technology

In order to develop the VALCRI platform as well as its associated software components, realistic data must be available.

Much anonymised data has been provided internally for the project. This data originated from West Midlands Police and was anonymised by AES.

However, the anonymised data did not include financial data, which meant that research related to this topic was difficult to conduct. In addition, this anonymised data could not be used for developing the high TRL software for the Antwerp forces as it has a structure and semantics that is very different from the real data used by these forces.

In support of the research and development activities part of the project two additional realistic yet synthetic datasets were created:

Financial data. The goal here was to generate realistic financial transactions data represented by bank accounts logs of people or companies. The requirement was to automatically generate much “noise” data, and then introduce in this generated data, manually, entries that relates to the rest of the available anonymised data. The noise data, however, cannot be completely random. A police analyst looking into the data expects financial transactions that make sense. Daily activities where, for example, a person makes a purchase in a supermarket in his home town should be part of the background noise. Malicious transactions that were embedded had to be chosen very carefully in order to not be too conspicuous. For example, purchases done at impossible hours or in an unreasonable amount of money also would immediately attract attention. Therefore, complex constraints were taken into account when implementing and when generating the data.

Data similar to Flux24: Flux24 is a data set that federates police reports over all the local police forces in the region of Antwerp. Here, the goal for having the dataset was very different. Whilst we could test the software developed on real data at the premises of the Federal Police of Antwerp on a dedicate, standalone machine, the data cannot be copied or moved. As a result, some bugs were a challenge to fix because we simply could not reproduce these issues without the real data. Therefore, we created a data set that was realistic enough to trigger many of the issues we discovered in the software. The dataset features realistic statistics triggering each of the visualizations, and is big enough to run stress tests.

Explanation of why it is a significant advancement or ground-breaking

The financial data generator was built using system of actors. Actors are software objects which encapsulate state and behaviour. They communicate exclusively by exchanging messages, which are placed into the recipient’s mailbox, and run concurrently. The overall resulting software model is conceptually similar to a group of people sending messages to each other, where each individual follows a certain behavioural pattern. In our system of actors, there were actors that each represent a person or a business. Configuration files were designed for the different types of actors allowing to configure a different behaviour for each actor type.

The overall system included also transaction actors that process financial transactions and introduce realistic delays, much like real transaction delays in modern banks. The dataset can be derived from the log files of the transaction actors.

This design allowed to configure many interesting constraints in the system. For example:

Each organization has some employees, and the salaries to these employees have to be paid every month.

Each entity must have an address, and transactions between geographically close entities are favoured. This feature was implemented using Nominatim.

Each actor should be able to correctly select the times to do different activities - a person can visit the supermarket only during its opening times - which means that we have to configure the activity hours and days, and include also holidays.

Regular transactions e.g. water bills, electricity bills, insurances, and of course monthly salaries were included.

The Flux24 generator was initially based on the statistics of real data provided by the police force (e.g. length of fields). In order to generate text, word frequency analysis was made over relevant text (police reports, as well as relevant public documents). From this word frequency list, random text was generated that is statistically realistic. In order to generate random addresses, a reverse geocoder was set up, throwing digital darts on the map and generating the addresses accordingly. The areas that were to be hit with a higher probability can be configured to allow imitating realistic geographical crime statistics. Also the timing of the events can be configured.

This kind of reconfigurability is especially useful when testing the time scatter visualization. For example, we have statistically embeded the word "hello" amongst the thousands of fake crime reports. As a result, the time scatter plot displays the word "hello" and confirms that the software operates as intended.

The Flux24 generator has been used for reproducing bugs, testing the visualizations and stress testing the system. Textual data, as big as 700MBytes which represents police reports of a full year, could be generated in few minutes.

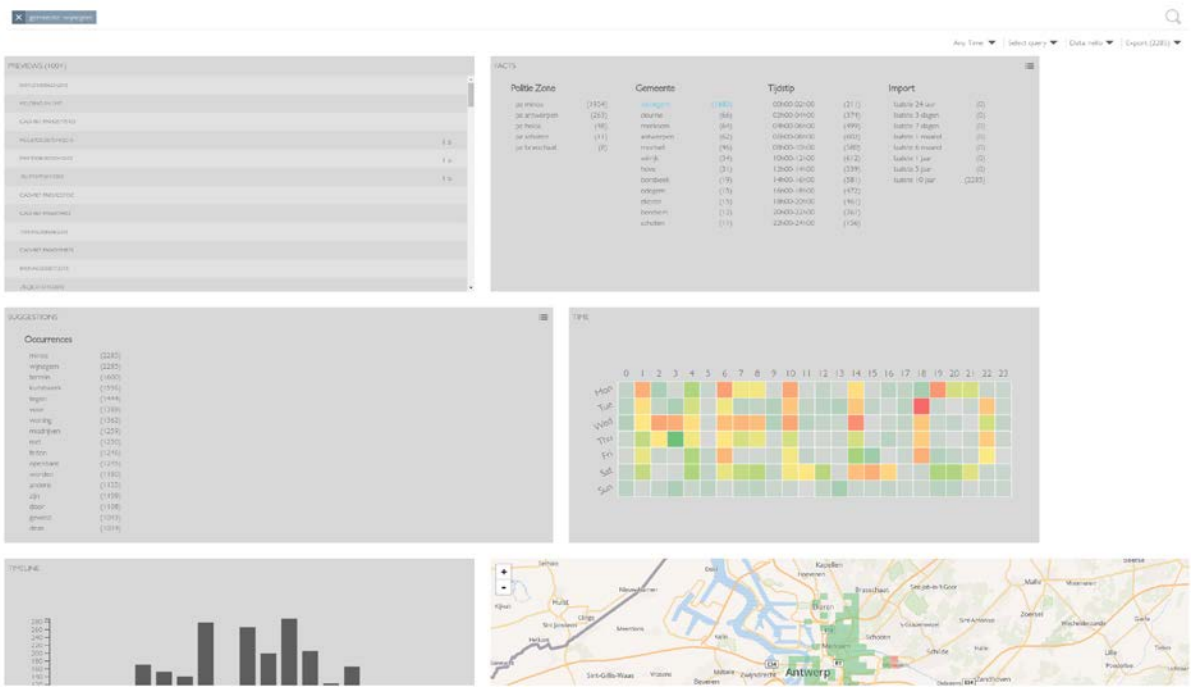


Figure 5 - A screenshot showing where timescatterplot displays the word "hello" as statistically encoded into the synthetic Flux24 data set

Published papers or patents to prove the above claims

Not applicable

Impact

The financial dataset was used for the relevant research conducted in the project.

The Flux24 datasets were used to develop and test the software components developed for the Federal Police of Antwerp. The data set allowed pushing TRL of our software.

AGB28. Enhancing VALCRI's operational efficacy by evaluating a dominant LPA data source - ISLP - for integration.

Nick Evers, Christophe Vandenberghe, David De Weerdt, Rani Pinchuk
Space Applications Services (SPACE)

Description of the significant Advances or Ground-breaking Science or Technology

Of fundamental importance to the long-term success and applicability of the outcome of an R&D project such as VALCRI is its relevance to real-life policing. This to allow for a proper assessment of what additional efforts are needed to make the project outcomes truly useful in the real world. To this end, efforts have been undertaken to gain a clear understanding of the policing domain, the police departments structure, the intra- and interdepartmental information flows and their accompanying IT and data landscape, chains of (quality) control, and the relationships, both technically and organizationally with other police zones, of the police forces of Antwerp, Belgium.

This activity resulted in non-straightforward but interesting and crucial findings. As an example, the way a chain of police reports on a single incident is built results in a need to merge chains of reports into a single coherent picture in order to efficiently gain a full understanding. Also, no matter the amount of (r)evolution done in policing and analysis activities, it has to be taken into account that other governmental agencies, such as justice, are unlikely to (be able to) follow the same innovation and integration pace for various reasons (legislation, infrastructure, budgets, ...)

Based on the understandings above, a close look was taken at ISLP (Integrated System for Local Police) which is the main IT system for storing policing data used by the local police throughout the whole of Belgium. ISLP is elaborate but relatively old technology and had to, since its inception more than 30 years ago, evolve with ever-changing policing practices. At no point in time all police zones in Belgium shared an identical way of working as they operate quite autonomously. Consequently, at no point in history a complete from-scratch overhaul has been performed on the ISLP system to have it aligned with all zones. The complexity and its technical rigidity offset against the wide-spread current use make the ISLP system a prime candidate for further investigation on how ground-breaking new developments can be aligned with it. To this end, a detailed understanding of the technology was pursued with the ultimate intention to evaluate the ramifications of integrating new ground-breaking R&D with ISLP.

Explanation of why it is a significant advancement or ground-breaking

The first version of the root predecessor of the ISLP database and its schema have been designed and implemented over 30 years ago. Operational constraints and requirements have changed fundamentally since then but the ISLP system has never received an equally fundamental overhaul. Instead, the semantics but not the implementation of the database have been redefined when needed. Because of this, hooking into the ISLP system to wire it as the data source of new applications is very challenging. This is substantiated by the following consequences which are the result of the implementation and update/-grade procedures used for keeping ISLP operational up to current modes of usage.

The ISLP system, both the database and the end user application built on top of it, use highly-domain specific terminology which requires a thorough understanding of the domain. This domain is very complex and very specific which makes that only domain experts will be able to sufficiently convey the intricacies involved in understanding why certain things are as they are. These domain experts are typically police officers or directly supporting the police departments. They are not database experts and cannot themselves relate the domain knowledge to database design and implementation choices. In addition, the documentation detailing the database schema (above 400 tables) is large (above 600 pages) and complex but still incomplete, outdated

and at times incorrect as fields and tables have changed semantics, sometimes completely unrelated to the original intent, making the database field names nonsensical with respect to its current use.

As the ISLP database is nationally deployed, cloned and used, updating the schema non-disruptively is very hard. Instead, the ISLP software itself that provides the main user interface to the database, contains most of the required updates and peculiarities. These include both hardcoded but closed-source definitions, translated and stored as numerical codes only, and various interface and usability constraints and capabilities which undoubtedly have organically grown with good purpose but result in undesired side effects in the database. In particular, the latter has no clear consistent data typing, no normalisation, null-values can happen everywhere, seemingly (semantically) unchangeable values can change and so on. Because of its nation-wide scope, the ISLP database is designed to support the three national languages of Belgium (Dutch, French and German) necessitating many translation tables, thus increasing the query complexity.

The main semantic constructs stored in the database are uniquely identified using the combination of up to four identifier fields requiring many join queries to link various tables together. This, together with insufficient indexing severely impacts the performance that can be reached. No additional indexing or even data views can be added as the database used is an Intersystems Cache system which is a hybrid object/relational datastore, providing only an SQL facade with limited functionality.

The main achievement accomplished here is to get an initial understanding of the relevant aspects as discussed above, which is a precursor to be able to use this specific real-life police database.

Published papers or patents to prove the above claims

Not applicable

Impact

None yet.

AGB29. Development and testing environment

Nick Evers, Christophe Vandenberghe, David De Weerdt, Rani Pinchuk
Space Applications Services (SPACE)

Description of the significant Advances or Ground-breaking Science or Technology

As a consortium of 18 partners, 10 of which contribute various software components, there is a strong need for a development and testing pipeline in order to ensure the interoperability and cooperation between each of the components. It is fruitless to have powerful components that do not collaborate once the developed system is installed. Therefore, early and systematic system integration is and has been a key principle to the success of VALCRI.

For this purpose, each new component or every change to an existing component in the VALCRI system must pass through a chain of processing stages where each stage deals with a different integration aspect of the component in question. This chain of processing stages is more commonly referred to as the development and testing pipeline.

The pipeline provides system integration by attaining the following fundamental goals:

- Promote collaboration.
- Welcome new components and technology initiatives.
- Provide fast feedback for failure and success.
- Guard the functionality of each of the components
- Continuously challenge and evolve the interfaces between the components.

All development and code is synchronised though a GIT Source Control Management (SCM) (https://git-scm.com/) platform. The SCM is used to keep track of changes in the source code. It is accompanied by a GitLab (https://about.gitlab.com/) installation that provides a user interface that allows managing the user accounts and the individual code repositories.

In order to compile and build the code available from the SCM, a custom build system was developed using Gradle (https://gradle.org/). Gradle provides a Groovy (http://www.groovy-lang.org/) Domain Specific Language (DSL) and is used to define and automate all the details for managing and building the VALCRI system. The Gradle setup is also accompanied with a Nexus (http://www.sonatype.org/nexus/) component or artefact repository which hosts all the compiled, binary VALCRI system components.

For each change in the SCM, the source code is automatically rebuilt and all tests are run. In order to do this, a Jenkins (https://jenkins-ci.org/) installation was put into place. Jenkins continuously polls for changes in the SCM. Upon each change, it runs a build and test cycle over the full code base. Upon any failure, it will send an email to the VALCRI developer mailing list informing all technical partners about the failure. JUnit (http://junit.org/) was selected as the means to define unit tests and Selenium to run integration tests.

Getting the VALCRI system up and running while provisioning it with the correct configuration and environment is very complex. Even for small applications this can quickly become complicated and time consuming as each application has its own unique set of run-time dependencies and configuration files. These run-time dependencies are different from the compile-time dependencies that are resolved by the build system. In essence, the dependencies of the build system are comprised of the internal VALCRI software components depending on each other and the software libraries they are built from. Even though the development and testing pipeline (refer to D2.3 version 1) is able to weave together all those compile-time dependencies into one deployable and executable unit, this binary still has run-time dependencies to an operating system, networking and hardware facilities, logging, identity management, etc. These

dependencies are usually off-the-shelf services that are managed outside VALCRI. Yet, they need to be available and configured properly in order for the VALCRI system to work and remain secure.

In VALCRI we have chosen Docker (https://www.docker.com/) as the primary tool to solve this problem. As nicely described on their website, Docker is a platform for building, shipping and running the complete environment for distributed applications. It allows you to package your application and all its run-time dependencies into a standardized unit of deployment: a container. Each container can thus wrap a VALCRI software component in a complete filesystem that contains everything it needs to run: the executable binary, system tools, system libraries - anything you can install and execute on a server.

Explanation of why it is a significant advancement or ground-breaking

First integration was done only 6 months after the project started.

Even though the development team was distributed over different organizations in different countries, and was heterogeneous (e.g. research students, software engineers, ...), all managed to integrate their components to the overall system.

Dozens of components were integrated using the development and testing environment as its back-bone.

Its implementation is reused by: (1) each of the partners, (2) each of the components in the system and (3) it shows potential beyond the project.

The testing framework for the high TRL implementation for the FPA covers numerous integrations tests that guarantee the working state of the software and the semantic output of its search abilities.

Published papers or patents to prove the above claims

Not applicable

Impact

Integration is achieved very early in the VALCRI project lifetime and has contributed a lot to its success. Taking into account the challenges the project has faced, its rich and complex facets and in comparison to similar EU-projects, this is note-worthy.

ANNEX B
PUBLISHED PAPERS

SUMMARY OF PUBLISHED PAPERS

Total Published	**119**
Conference papers	64
Journal papers	22
Book chapters (springer Series)	06
Workshop papers	08
White papers	19
Phd Topics within the consortium	15

Conference papers - 64

1. Bedek, M. A., Nussbaumer, A., Hillemann, E. C., & Albert, D. (2017) A framework for measuring imagination in visual analytics systems In J. Brynielsson (Ed.), Proceedings of the European Intelligence and Security Informatics Conference (EISIC 2017) (pp. 151–154). IEEE. doi: 10.1109/EISIC.2017.3
2. Bedek, M. A., Nussbaumer, A., Huszar, L., & Albert, Dietrich (2017) Discovering Cognitive Biases in a Visual Analytics Environment. In Proceedings of the DECISIVe Workshop 2017 at the Vis Conference.
3. Beecham, R., Dykes, J., Slingsby, A. & Turkay, C. (2015) Supporting crime analysis through visual design.IEEE VIS 2015, 25-10-2015 - 30-10-2015, Chicago, USA. http://openaccess.city.ac.uk/12331/
4. Bielska,A., and Pallaris, C. (2017).Redefining the Intelligence Skill Set by Use of the Intelligence Analysis Impact Model International Studies Association Annual Convention, Baltimore, MD, United States, February .
5. Blomqvist,E., Keskisärkkä,R. (2015) Semantic web technologies for near real-time monitoring in criminal investigations TAMSEC 2015, 24-25 Nov 2015, Krista, Sweden
6. Buchmüller,J., Jentner,W.,Streeb,D., and Keim,D.A ODIX: (2017) A Rapid Hypotheses Testing System for Origin-Destination Data IEEE Conference on Visual Analytics Science and Technology (VAST Challenge 2017 MC1) 2017/10 PUBLISHED
7. Buchmüller., Juri., Jäckle., Dominik., Stoffel, F,. Keim, D. (2016) SpaceCuts: Making Room for Visualizations on Maps Eurographics Conference on Visualization (EuroVis), Shortpapers
8. Chiatti, A.,Dragisic, Z., Cerquitelli ,T., Lambrix ,P. (2015) Reducing the search space in ontology alignment using clustering techniques and topic identification 8th International Conference on Knowledge Capture, October 7-10, Palisades, NY, USA DOI 10.1145/2815833.2816959
9. Dominik,S., Ina,B.,Fuchs, Johannes; Keim, D.A (2016) Analytic Behavior and Trust Building in Visual Analytics Eurographics Conference on Visualization (EuroVis), Shortpapers 2016/06 PUBLISHED
10. Dragisic, Z., Ivanova ,V., Lambrix, P., Faria ,D., Jimenez-Ruiz, E., Pesquita ,C. (2016) User validation in ontology alignment, 15th International Semantic Web Conference, LNCS 9981, 200-217, Kobe, Japan, 2016. 10.1007/978-3-319-46523-4_13
11. Dykes, J., Rooney, C., Beecham, R., Turkay, C., Slingsby, A., Wood, J. & Wong, (2015) W Multi-Perspective Synopsis with Faceted Views of Varying Emphasis IEEE VIS 2015, 25-10-2015 - 30-10-2015, Chicago, USA. http://openaccess.city.ac.uk/12334/
12. Eriksson,H., Blomqvist,E., Keskisärkkä,R. (2017) Ontologies for Intelligence Analysis. CEPOL 2017 Research and Science Conference, 28-30 November 2017, Budapest, Hungary.
13. Gerber, M., Wong, B. L. W., & Kodagoda, N. (2016) How analysts think: decision making in the absence of clear facts. Adaptation of the RPD model and the decision ladder to analysts decision making Conference EISIC 2016, 17-19 August, 2016, Uppsalla, Sweden
14. Gerber, M., Wong, B. L. W., & Kodagoda, N. (2016). How analysts think: Intuition, Leap of Faith and Insight HFES 2016 19-23 September 2016, Washington, D.C., USA
15. Groenewald, C., Wong, B.L.W., Attfield, S., Passmore, P., Kodagoda, N. (2017). How Analysts Think: How do Criminal Intelligence Analysts Recognise and Manage Significant Information? Proceedings of the European Intelligence and Security Informatics Conference (EISIC) 2017, Dekelia Air Base, Attica, Greece. IEEE 2017.
16. Groenewald, C., Wong, B.L.W., Attfield, S., Passmore, P., Kodagoda, N. (2017). How Analysts Think: Navigating Uncertainty – Aspirations, Considerations and Strategies Proceedings of the 13th International Conference on Naturalistic Decision Making. Bath, United Kingdom. (p56-64)
17. Groenewald, C., Wong, B.L.W., Attfield, S., Passmore, P., Kodagoda, N. (2017). How can we design Tactile Interactive software for Argument Construction in Criminal Intelligence Analysis? Proceedings of the 13th International Conference on Naturalistic Decision Making. Bath, United Kingdom. (p322-328)
18. Groenewald,C., Islam,J., Anslow, C.,Rooney,C.,Wong,B.L.W Understanding Mid-Air Hand Gestures for Interactive Surfaces: A Systematic Literature Review, British HCI 2016 conference, Bournemouth, United Kingdom, DOI:10.14236/ewic/HCI2016.43.
19. Haider, J.; Pohl, M.; Hillemann, E.; Nussbaumer, A.; Attfield, S.; Passmore, P.; and Wong, B. L. W. Exploring the Challenges of Implementing Guidelines for the Design of Visual Analytics Systems. Conference HFES 2015
20. Hammar, K. (2015). Ontology Design Patterns in WebProtégé. Conference In Proceedings of the ISWC 2015 Posters & Demonstrations Track co-located with the 14th International Semantic Web Conference (ISWC-2015), Betlehem, USA, October 11, 2015. CEUR Workshop Proceedings.
21. Hillemann, E. C., Nussbaumer, A., & Albert, D. (2015). The role of cognitive biases in criminal intelligence analysis and approaches for their mitigation. European Intelligence and Security Informatics Conference (EISIC), 7-9 September 2015, Manchester, UK. (pp. 125-128). IEEE. doi:10.1109/EISIC.2015.9

22. Hillemann, E. C., Nussbaumer, A., Bedek, M., & Albert, D. (2016). The what, the how and the who – A psycho-pedagogical training model for professional development in criminal intelligence analysisIn L. Gómez Chova, A. López Martínez & I. Candel Torres (Eds.), ICERI2016 Proceedings, 14-16 November 2016, Seville, Spain. (pp. 7867-7873). IATED.
23. Hillemann, E., Nussbaumer, A., & Albert, D. (2015). Visualisations and their effect on cognitive biases in the context of criminal intelligence analysis Proceedings of the International Conferences Interfaces and Human Computer Interaction 2015, Game and Entertainment Technologies 2015, and Computer Graphics, Visualization, Computer Vision and Image Processing 2015, 22-24 July 2015, Las Palmas de Gran Canaria, Spain. (pp. 313-315).
24. Islam,J., Wong,B.L.W, (2017), Behavioural Markers: Bridging the Gap Between Art of Analysis and Science of Analytics In Criminal Intelligence, Conference European Intelligence and Security Informatics Conference (EISIC) 2017, Dekelia Air Base, Attica, Greece, ISBN: 978-1-5386-2385-5, DOI: 10.1109/EISIC.2017.30.
25. Islam,J.,Anslow,C., Xu,K., Wong,B.L.W (2018), Uncertainty of Visualizations for SenseMaking in Criminal Intelligence Analysis, EuroRV3: 20th EG/VGTC Conference on Visualization, EUROVIS (2018), Brno, Czech Republic (Accepted).
26. Islam,J.,Anslow,C., Xu,K., Wong,B.L.W, Zhang,L (2016)Towards Analytical Provenance Visualization for Criminal Intelligence Analysis, Conference Computer Graphics & Visual Computing (CGVC) 2016, Bournemouth University, United Kingdom, ISBN 978-3-03868-022-2, DOI: 10.2312/cgvc.20161290.
27. Ivanova, V., Bach ,B., Pietriga E., Lambrix ,P. (2017) Alignment Cubes: Interactive Visual Exploration and Evaluation of Multiple Ontology Alignments, 16th International Semantic Web Conference Posters and Demos, CEUR Workshop Proceedings Volume 1963, Vienna, Austria, 2017.
28. Jäckle,D .,Stoffel,F.,Kwon,B.C.,Sacha.D, Stoffel,A.,Keim.D.A (2015)Ambient Grids: Maintain Context-Awareness via Aggregated Off-Screen Visualization. Eurographics Conference on Visualization (EuroVis)
29. Jäckle,D .,Stoffel,F.,Mittelstädt.S, Keim,D.A, Reiterer,H., (2017) Interpretation of Dimensionally-Reduced Crime Data: A Study with Untrained Domain Experts Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications 2017
30. Jäckle,D., Fuchs,J., Keim,D.A (2016) Star Glyph Insets for Overview Preservation of Multivariate Data Electronic Imaging Conference on Visualization and Data Analysis, 2016/02
31. Jäckle.D., Hund.M., Behrisch.M., Keim.D.A and T. Schreck(2017) Pattern Trails: Visual Analysis of Pattern Transitions in Subspaces IEEE Conference on Visual Analytics Science and Technology (VAST) 2017/10 PUBLISHED
32. Jentner,W., El-Assady,M., .,Sacha.D., Jäckle,D., and Stoffel,F., (2016) Dynamite: Dynamic Monitoring Interface for Task Ensembles IEEE VAST Challenge 20162016/10 PUBLISHED
33. Jentner,W., Ellis,G., Stoffel,F.,Sacha.D, Keim.D.A (2016) A Visual Analytics Approach for Crime Signature Generation and Exploration The Event Event: Temporal & Sequential Event Analysis IEEE VIS, 2016/10 PUBLISHED
34. Keskisärkkä,R (2016) Representing RDF Stream Processing Queries in RSP-SPIN. ConferenceIn Proceedings of the ISWC 2016 Posters & Demonstrations Track co-located with the 15th International Semantic Web Conference (ISWC-2016), Kobe, Japan, October 17-21, 2016, CEUR Workshop Proceedings, 2016.
35. Keskisärkkä,R, Blomqvist,E (2015) Supporting real-time monitoring in criminal investigations ESWC conference 31 May-4 June 2015, Portoroz, Slovenia
36. Kriglstein, S.; Haider, J.; Wallner, G.; and Pohl, M. (2015) Who, Where, When and with Whom? Evaluation of Group Meeting Visualizations. Conference on the Theory and Application of Diagrams 2016, Philadelphia, 2016.
37. Marquenie,T., & Zouave,E., 2018 Speaking Truth to Computational Power: Coding Non-discrimination by Design in Police Technology Next-Generation Community Policing Conference,2018
38. Nussbaumer, A., Verbert ,K.,Hillemann, E.C., Bedek ,M.A., and Albert,D. (2016) A Framework for Cognitive Bias Detection and Feedback in a Visual Analytics Environment Conference EISIC 2016, 17-19 August, 2016, Uppsalla, Sweden
39. Passmore, P.J., Attfield, S., Kodagoda, N., Groenewald, C. and Wong, B.L.W. (2015). Supporting the Externalisation of Thinking in Criminal Intelligence Analysis. In Intelligence and Security Informatics Conference (EISIC), 2015 European (pp. 16-23). IEEE.
40. Paudyal,P, Duquenoy,P Barn,B., Wong,B.L.W (2016) Balancing security and privacy in development of interactive technology for mediating criminal intelligence analysis BritishHCI, 2016, accepted
41. Paudyal,P., Rooney,C.,Kodagoda,N.,Wong,B.L.W,Duquenoy,P,Qazi,N(2017) How the Use of Ethically Sensitive Information Helps to Identify Co-offenders via a Purposed Privacy Scalea Pilot Study, EISIC, 2017, published
42. Paudyal,P., andWong,W.B.L (2018) Algorithmic Opacity: Making Algorithmic Process Transparent through Abstraction Hierarchy HFES, 2018, accepted
43. Paudyal,P., Wong,B.L.W., Kodagoda,N., Rooney., Duquenoy,P ,(2017) Algorithmic Transparency: Use of ethically sensitive information via a purposed privacy scale for decision-making in criminal intelligence NGCP, 2017, paper presented

44. Phong,N.,Kai,X.,Walker, Rick and Wong, B.L. W (2014) SchemaLine: timeline visualization for sensemaking. DOI 10.1109/IV.2014.14 IEEE 18th International Conference on Information Visualisation (IV), 9-14 Nov 2014, Paris, France
45. Pohl,M.,Winter,L-C.,Pallaris,C.,Attfield,S., Wong,B.L.W (2014) Sensemaking and Cognitive Bias Mitigation in Visual Analytics DOI 10.1109/JISIC.2014.68 JISIC 24-26 Sept 2014, The Hague, The Netherlands
46. Qazi,N., Wong,B.L.W (2016) Semantic-Based Image Retrieval Through Combined Classifiers of Deep Neural Network and Wavelet Decomposition of Image Signal Conference at IEEE 2016 9th EUROSIM Congress on Modelling and Simulation Oulu, Finland 12-16 September 2016, will be published in IEEE Conference Proceeding
47. Qazi,N., Wong,B.L.W (2018) submitted Interactive Visualization for Crime Pattern and Offender Network Detection to be held in London August 2018.
48. Qazi,N., Wong,B.L.W,2017 Behavioural Tempo-spatial Knowledge Graph for Crime matching through Associate Questioning and Graph Theory European Intelligence and Security Informatics Conference (EISIC) 2017, Dekelia Air Base, Attica, Greece,
49. Qazi,N., Wong,B.L.W,Kodagoda,N.,Adderley,R (2016) Associative Search through Formal Concept Analysis in Criminal Intelligence Analysis IEEE International Conference on Systems, Man, and Cybernetics (SMC 2016) OCTOBER 9-12 2016 Budapest
50. Sacha ,D., Senaratne, H. , Kwon, B. C. Ellis, , G. Keim , D. A. (2015) The Role of Uncertainty, Awareness, and Trust in Visual Analytics IEEE Transactions on Visualization and Computer Graphics (Proceedings of the Visual Analytics Science and Technology) 2015/08 PUBLISHED
51. Sacha ,D., Zhang,L.,Sedlmair,M., Lee,J.A.,Peltonen,J.,Weiskopf,D.,North,S.C., Keim.D.A (2016) Visual Interaction with Dimensionality Reduction: A Structured Literature Analysis. IEEE Transactions on Visualization and Computer Graphics (Proceedings of the Visual Analytics Science and Technology), 2016/08
52. Sacha ,D.,Boesecke,J., Fuchs,j., and Keim , D. A. (2016) Analytic Behavior and Trust Building in Visual Analytics. Eurographics Conference on Visualization (EuroVis) - Short Papers, The Eurographics Association, DOI: 10.2312/eurovisshort.20161176, 2016.URL: https://diglib.eg.org/handle/10.2312/eurovisshort20161176
53. Schlehahn,E., Aichroth,P., Schreiner,R., Mann,s. (2015) Benefits and pitfalls of predictive policing EISIC 7-9 September 2015 Manchester, UK
54. Seidler, P.; Haider, J.; Kodagoda, N.; Pohl, M.; and BL William,(2016) Design for Intelligence Analysis of Complex Systems: Evolution of Criminal Networks. EISIC 2016, 17-19 August, 2016, Uppsalla, Sweden
55. Selvaraj, N., Attfield, S., Passmore, P., & Wong, B. W. (2016). How Analysts Think: Think-steps as a Tool for Structuring Sensemaking in Criminal Intelligence Analysis. In Intelligence and Security Informatics Conference (EISIC), 2016 European (pp. 68-75). IEEE.
56. StoffelF., Jentner,W., Behrisch,M., Fuchs,J., Keim,D.A (2017) Interactive Ambiguity Resolution of Named Entities in Fictional Literature EuroGraphics: Computer Graphics Forum 2017/04
57. StoffelF.,Sacha,D., Ellis,G., and Keim,D.A (2015) VAPD - A Visionary System for Uncertainty Aware Decision Making in Crime Analysis Symposium on Visualization for Decision Making Under Uncertainty IEEE VIS 2015 PUBLISHED
58. Wong, B. L. W. (2016) Fluidity and Rigour: Designing for Intelligence Analysis. Open Source Intelligence Investigation: From Strategy to Implementation (pp. To be published): Springer. 2016
59. Wong, B. L. W., & Kodagoda, N. (2016). How analysts think: Anchoring, addering and Associations HFES 19-23 September 2016, Washington, D.C., USA
60. Wong, B. L. W., Seidler, P., Kodagoda, N., & Rooney, C. (2018). Supporting variability in criminal intelligence analysis: From expert intuition to critical and rigorous analysis. In G. Leventakis & M. R. Haberfeld (Eds.), Societal Implications of Community-Oriented Policing Technology (pp. In Press). Cham, Switzerland: Springer International Publishing AG.
61. Wong, B.L. William (2014) How Analysts think (?): early observations Intelligence and Security Informatics Conference (JISIC), 24-26 Sept 2014 The Hague, The Netherlands IEEE Joint (pp. 296-299). IEEE.
62. Zhang,L., Rooney,C., Nachmanson,L., Wong,B.L.W.,Kwon,B.C., Stoffel,F.,Qazi,N., Singh,U and Keim.D.A 2016 Spherical Similarity Explorer for Comparative Case Analysis (2016), Electronic Imaging Conference on Visualization and Data Analysis
63. Zhang,L., Rooney,C., Nachmanson,L., Wong,B.L.W.,Kwon,B.C., Stoffel,F.,Qazi,N., HundM., Singh,U., Nachmanson,L Keim,D (2015) Interactive Spherical MDS for Comparative Case Analysis IEEE VIS 25-30 Oct 2015, Chicago, USA
64. Zouave,E., Marquenie,T.,(2017) An Inconvenient Truth: Algorithmic Transparency & Accountability in Criminal Intelligence Profiling European Intelligence and Security Informatics Conference,2017

Journal papers - 22

1. 1. Bedek, M. A., Nussbaumer, A., Huszar, L., & Albert, Dietrich (2018) Methods for Discovering Cognitive Biases in a Visual Analytics Environment Journal Springer (submitted)
2. Beecham, R., Dykes, J., Meulemans, W., Slingsby, A., Turkay, C. & Wood, J (2017) Map LineUps: effects of spatial structure on graphical inference Journal IEEE Transactions on Visualization and Computer Graphics.2016 DOI 10.1109/TVCG.2016.2598862
3. Beecham, R., Rooney, C., Meier, S., Dykes, J., Slingsby, A., Turkay, C., Wood, J. & Wong, B.L.W. (2016) Faceted Views of Varying Emphasis (FaVVEs): a framework for visualising multi-perspective small multiples. Journal Computer Graphics Forum: the international journal of the Eurographics Association, 35(3), pp. 241-249.
4. Bielska, A., and Pallaris, C. (2017). Addressing the Internal Challenges to Intelligence Work Journal of Mediterranean and Balkan Intelligence, Vol. 10, No. 2 (2017), pp. 89-102.
5. Dragisic,Z., Ivanova,V., Li,H., Lambrix, P.(2017) Experiences from the Anatomy track in the Ontology Alignment Evaluation Initiative Journal Journal of Biomedical Semantics 8:56, 2017. 10.1186/s13326-017-0166-5
6. E. Kindt ,(2017) Having yes, Using no ? About the new legal regime for biometric data, Journal Computer Law & Security Review
7. F. Coudert,(2017)The Europol Regulation and Purpose Limitation: From the 'Silo-Based Approach' to… What Exactly? Journal European Data Protection Law Review
8. Haider,D.,Pohl,J.,Beecham,M.R.,Dykes, J (2018) Understanding Map Comparison: Strategies for Detecting Difference n Map Line-up Tasks. Journal IEEE Transactions on Visualization and Computer Graphics. (submitted)
9. Islam, J., Anslow, C., Xu, K., Wong, W., and Zhang, L. (2016).Towards Analytical Provenance Visualization for Criminal Intelligence Analysis. Journal In Turkay, C. and Wan, T. R., editors, Computer Graphics and Visual Computing (CGVC).
10. Islam,J., Xu,K.,Wong,B.L.W (2018) Analytic Provenance for Criminal Intelligence Analysis Journa Chinese Journal of Network and Information Security, Beijing, China, 2018, pages: 18-33, ISSN: 2096-109X, CN:10-1366/TP.
11. Jäckle,D.,Fischer,F.,Schreck,T.,Keim,D.A (2016) Temporal MDS Plots for Analysis of Multivariate Journal Data IEEE Transactions on Visualization and Computer Graphics (TVCG), 2016
12. Jentner,W., Sacha,D., Stoffel,F., Ellis,G.,Zhang,L.,Keim,D.A (2018) Making machine intelligence less scary for criminal analysts: reflections on designing a visual comparative case analysis tool Springer: Journal The Visual Computer 2018/02 PUBLISHED
13. Lambrix ,P., Kaliyaperumal,R. (2017) A Session-based Ontology Alignment Approach enabling User Involvement Journal Semantic Web Journal 8(2):225-251, 2017. 10.3233/SW-160243
14. Marquenie,T (2017)The Police and Criminal Justice Authorities Directive: Data protection standards and impact on the legal framework,Computer Law & Security Review,Pages 324-340
15. Nguyen, P. H., Xu, K., Wheat, A., Wong, B. L. W., Attfield, S., & Fields, B. (2016). SensePath: Understanding the Sensemaking Process through Analytic Provenance. Journal IEEE Transactions on Visualization and Computer Graphics, 22(1), 41--50. doi: http://doi.org/10.1109/TVCG.2015.2467611
16. Qazi,N., Wong,B.L.W, (2018) 2018 A Human Centered Data Science Approach towards Crime matching submitted to Journal of Information Processing and Management
17. Rooney, C., Beecham, R., Dykes, J. & Wong, (2018) W. Statistical Process Control Charts: A Visualization Makeover for Crime Analysis. Journal IEEE Transactions on Visualization and Computer Graphics. ((submitted an 2018))
18. Rooney, C., Attfield, S., Wong,B.L.W; Choudhury, S (2014) INVISQUE as a tool for intelligence analysis: the construction of explanatory narratives. DOI 10.1080/10447318.2014.905422 Journal International Journal of Human-Computer Interaction, 30 (9). pp. 703-717, 2014
19. Sacha,D.,Zhang,L., Sedlmair,M.,Lee,J.A.,Peltonen,J.,Weiskopf,D.,North,S.C.,Keim,D.A (2016)Visual Interaction with Dimensionality Reduction: A Structured Literature Analysis. Journal IEEE Transactions on Visualization and Computer Graphics (Proceedings of the Visual Analytics Science and Technology), DOI: 10.1109/TVCG.2016.2598495, 2016.URL: http://ieeexplore.ieee.org/document/7536217/
20. Sacha,D., Senaratne,H.,Kwon,B.C.,Ellis,G., Keim,D.A (2016) The role of uncertainty, awareness and trust in visual analytics Journal IEEE Transactions on Visualization and Computer Graphics (Proceedings of the Visual Analytics Science and Technology), 22(01), to appear, 2016
21. Wong, B.L. W., Kodagoda, N., Attfield,S (2015) How Analysts Think: Inference Making Strategies Human Factors and Ergonomics Society 2015 Annual Meeting, 26th - 30th October, Los Angeles. (Accepted - In Press)
22. Wood, J., Kachkaev, A. & Dykes J. 2018 Design Exposition with Literate Visualization IEEE Transactions on Visualization and Computer Graphics. 2018 submitted

Book Chapters – 06

1. Duquenoy,P., Gotterbarn,D., Kimppa,K.K.,Patrignani,N., Wong,B.L.W (2018) Addressing Ethical Challenges of Creating New Technology for Criminal Investigation: The VALCRI Project in George Leventakis and M.R. Haberfeld (Eds), Book chapter Societal Implications of Community-Oriented Policing and Technology. Springer International Publishing 2018. pp31-38.
2. Groenewald, C., Attfield, S., Wong, B.L.W, Passmore, P., Qazi, N., Kodagoda, N. (2018). A Descriptive, Practical, Hybrid Argumentation Model to Assist With the Formulation of Defensible Assessments in Uncertain Sense-Making Environments – An Initial Evaluation. Book chapter in Special Issue of Cognition, Technology and Work: Naturalistic Decision Making: Navigating Uncertainty in Complex Sociotechnical Work. To be released in 2018.
3. Groenewald, C., Wong, B.L.W, Attfield, S., Passmore, P., Kodagoda, N. (2018). Descriptive, Practical, Hybrid Argumentation Model to Assist With the Formulation of Defensible Assessments in Uncertain Sense-Making Environments. In Book chapter In Haberfeld, M.R. Briefs in Policing, Springer. ISSN: 2194-6213. To be released in 2018.
4. Islam,J., Wong,B.L.W (2017), Analytic Provenance as Constructs of Behavioural Markers for Externalizing Thinking Processes in Criminal Intelligence Analysis Book chapter Next Generation Community Policing (NGCP) Conference 2017, 25-27 October, Heraklion, Crete, Greece, Springer Book_463110 (Chapter 10), ISBN 978-3-319-89293-1.
5. Keskisärkkä, R. and Blomqvist , (2015 published 2015 E. Supporting Real-Time Monitoring in Criminal Investigations. Book chapter The Semantic Web: ESWC 2015 Satellite Events: ESWC 2015 Satellite Events, Revised Selected Papers, May 31–June 4, 2015, Portorož, Slovenia.
6. Qazi,N., Wong,B.L.W (2018) Contextual visualization of crime matching through similarity clustering and Bayesian analysis Book chapter accepted as a chapter for publication in Social Media Strategy in Policing (from cultural intelligence to community policing) book to be published by Springer Accepted.

Workshop papers - 08

1. Dragisic. Z., Lambrix P., Blomqvist E. (2015) Integrating Ontology Debugging and Matching into the extreme Design Methodology 6th Workshop on Ontology and Semantic Web Patterns, CEUR Workshop Proceedings Volume 1461, Bethlehem, Pennsylvania, USA, 2015.
2. Ellis,G., Dix ,A., (2015)Decision making under uncertainty in visualisation? IEEE VIS 2015 Workshop
3. Jäckle,D., Senaratne,H.,Buchmüller,H.,Keim,D.A (2015) Integrated Spatial Uncertainty Visualization using Off-screen Aggregation EuroVis Workshop on Visual Analytics (EuroVA) 2015/05
4. Jentner,W., El-Assady,M., Gipp,B.,Keim,D.A (2017) Feature Alignment for the Analysis of Verbatim Text Transcripts EuroVis Workshop on Visual Analytics (EuroVA) 2017/04 PUBLISHED
5. Keskisärkkä, R. (2016) Query Templates for RDF Stream Processing. Proceedings of Stream Reasoning Workshop. 2016 October 17th-18th, 2016, Kobe, Japan. Collocated with the 15th International Semantic Web Conference (ISWC 2016), CEUR Workshop Proceedings, 2016.
6. Keskisärkkä,R., Blomqvist., (2015) Towards the use of RDF stream processing engines for event enrichment from social media streams SAFE workshop, ISCRAM 2015 24-27 May , Kristiansand, Norway
7. Lambrix ,P., Dragisic, Z., Ivanova ,V., Anslow ,C. (2016) Visualization for Ontology Evolution, 2nd International Workshop on Visualization and Interaction for Ontologies and Linked DataCEUR Workshop Proceedings Volume 1545, 54-67, Kobe, Japan, 2016.Workshop
8. Sacha ,D., Jentner,W., Zhang,L., Stoffel,F., Ellis,G (2017) Visual Comparative Case Analytics EuroVis Workshop on Visual Analytics Workshop EuroVA) 2017/04 PUBLISHED

White papers - 19

1. Adderley R. (2017) The Art and Science of Anonymising Data [White paper],VALCRI-WP-2017-007.
2. Aichroth P, Mann S, & Schreiner R. (2017). Security and Privacy Technologies in VALCRI. ([White paper],VALCRI-WP-2017-004).
3. Bielska A, & Pallaris C. (2017). Evaluating the VALCRI System: A Summary of End User Feedback and Recommendations. ([White paper],VALCRI-WP-2017-015).
4. Bielska A, & Pallaris C. (2017). Improving Professional Training in Criminal Intelligence Analysis. ([White paper],VALCRI-WP-2017-008).

5. Bielska A, & Pallaris C. (2018). The VALCRI Workshops: Key Findings on the VALCRI System and the Training of Law Enforcement Analysts. ([White paper],VALCRI-WP-2018-016).
6. Dykes J. (2018). Representing Uncertainty in VALCRI - A Proposal Uncertainty ([White paper],VALCRI-WP-2018-018).
7. Evers N, David D. W, Vandenberghe C, & Pinchuk R. (2017). Commercial Exploitation in Research Projects - Development of High TRL Visual Analytics-based Criminal Intelligence Analysis Software ([White paper],VALCRI-WP-2017-013).
8. Hillemann E.-C, Bedek M. A, Nussbaumer A, Pohl M, Haider J, Attfield S, Ellis G. (2017). Psychological Factors in Criminal Intelligence Analysis:What they are why they are important and how to deal with. ([White paper],VALCRI-WP-2017-006).
9. Islam J, Craig A, Xu K, & Wong B. L. W. (2017). Analytical Provenance for Criminal Intelligence Analysis ([White paper],VALCRI-WP-2017-009).
10. Keskisärkkä R, Blomqvist E, & Eriksson H. (2018). Integration of Streaming Data and Background Knowledge in VALCRI. ([White paper],VALCRI-WP-2018-019).
11. Marquenie T, & Coudert F. (2017). Roadmap for the Operationalization of Legal and Privacy Requirements in VALCRI Analysis. ([White paper],VALCRI-WP-2017-012).
12. Martin E, Martin J, Todd S, Lowe C, De Wachter J, Linden I. V, Wong B.L.W (2018). Lessons Learnt and Impact on Policing and Intelligence Analysis. ([White paper],VALCRI-WP-2018-017).
13. Pinchuk R, Evers N, & Vandenberghe C. (2017). Architecture Development and Testing Environment for a Visual Analytics-based Criminal Intelligence Analysis System. ([White Paper],VALCRI-WP-2017-001).
14. Qazi N, Zhang L, Blomqvist E, Stoffel F, Aichroth P, & Weigel C. (2017). Applying Data Science to Criminal Intelligence Analysis. ([White paper],VALCRI-WP-2017-003).
15. Rooney C, & Seidler P. (2017). Applying Dynamic Visual Queries to Crime Intelligence Analysis. ([White paper],VALCRI-WP-2017-010).
16. Sacha D, Jentner W, Zhang L, Stoffel F, Ellis G, & Keim D. (2017). Applying Visual Interactive Dimensionality Reduction to Criminal Intelligence Analysis. ([White paper],VALCRI-WP-2017-011).
17. Schlehahn E, Duquenoy P, Paudyal P, George C, Coudert F, Delvaux A, & Marquenie T. (2017). The Operationalisation of Transparency in VALCRI. ([White paper],VALCRI-WP-2017-005).
18. Todd S. (2017). Understanding the ethical dilemmas of implementing Data Driven Insights for West Midlands Police. ([White paper].VALCRI-WP-2017-014).
19. Wong B.L.W, Rooney C, & Neesha K. (2017). Analyst User Interface:Thinking Landscape as Design Concept. ([White paper],VALCRI-WP-2017-001).

PhD Theses Titles Associated with Project VALCRI

Middlesex University

1. Celeste Groenewald: A descriptive, practical, hybrid argumentation model to assist sense-making in uncertain environments, such as with criminal intelligence analysis. (In process)
2. Junayed Islam: Visual Analytics for Geo-Spatial Temporal Re-Construction of Situations In Criminal Intelligence Analysis (In process)
3. Pragya Paudyal: How can we make the Algorithmic Process Transparent for Ethical Decision-making in Criminal Intelligence Analysis? " (In process)

Linköping University

4. Valentina Ivanova: Fostering User Involvement in Ontology Alignment and Alignment Evaluation
5. Karl Hammar: Content Ontology Design Patterns: Qualities, Methods and Tools
6. Zlatan Dragisic: Completion of Ontologies and Ontology Networks
7. Robin Keskisärkkä: Towards Semantically Enabled Complex Event Processing (Licentiate)

TuGraz

8. Michael Bedek: The application of formal concept analysis for learner modeling
9. Eva Catherine Hillerman: "Visualisierungsformen und ihr Einfluss auf kognitive Verzerrungen beim Entscheidungsprozess" (“Kinds of Visualizations and their Effect on Cognitive Biases in Decision Processes“)

TU Wien Vienna University of Technology

10. Johanna Haider: Supporting Sensemaking Models for Visual Analytics

University of Konstanz.

11. Wolfgang Jentner: Visual Pattern Analytics for Event Sequences
12. Dominik Sacha: "Knowledge Generation in Visual Analytics : Integrating Human and Machine Intelligence for Exploration of Big Data"
13. Florian Stoffel: "Transparency in Interactive Feature-based Machine Learning: Challenges and Solutions"

KU Leuven

14. Fanny Coudert: “The purpose specification principle in the Area of Freedom, Security and Justice: towards renewed data protection principles for information-based practices in the field of Security”
15. Thomas Marquenie: "Algorithmic analysis and emerging AI in law enforcement practices and the criminal justice system: Data protection as an answer to fair trial concerns". (Commencing post project)

ANNEX C
VIDEOS

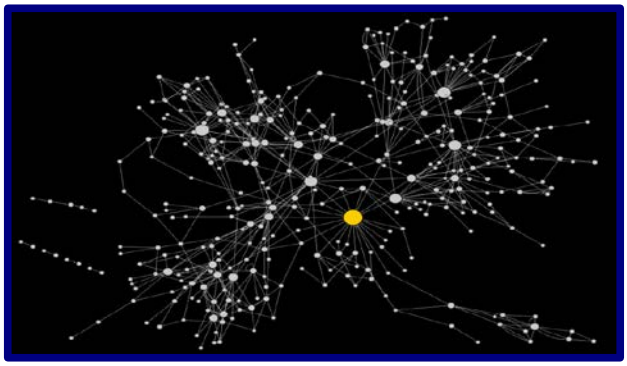
VALCRI Video List

1. What is VALCRI? - High-level overview of the project as a whole. (3:41)
2. VALCRI UI - A detailed look at the VALCRI user interface. (~5:00)
3. Data Analysis and Extraction - How concepts are mined. (5:27)
4. Data modelling - How we use evolving ontologies to model the data.
5. (4:49)
6. Security, ethics, legal and privacy - How these issues are
7. enforced/mitigated in VALCRI. (2:18)
8. Human Issues Framework - Understand bias, sensemaking, and how analysts
9. think. (~5:00)
10. Analyst Feedback - What our experts thought of VALCRI (5:09)

ANNEX D
IEB Report and IEB Risk Assessment

Final Report of the Independent Ethics Board on the VALCRI Project

Final Report of the Independent Ethics Board on the VALCRI Project



30 April 2018

Prepared by the Independent Ethics Board (IEB) members:
Kai Kimppa, Norberto Patrignani

The project VALCRI
(Visual Analytics for sense-making in Criminal Intelligence Analysis)
is funded from the European Union Seventh Framework Programme (FP7/2007-2013),
European Commission Grant Agreement N° FP7-IP-608142,
awarded to Middlesex University and partners.

Project Coordinator: Professor B.L. William Wong
Deputy Project Coordinator: Professor Ifan Shepherd

Content

1. Introduction

This document is the third and last report of the Independent Ethics Board (IEB) of the Project VALCRI. Its objectives are: to shortly introduce the goals of the VALCRI project, to highlight the main ethical issues involved with this kind of technology applications, and to provide the comments, from the IEB point of view, about the situation at the end of the project, with respects to the questions and advices provided by the IEB since the beginning of the project to the researchers involved with the project.

2. The VALCRI project

"VALCRI will be a leading edge research-based criminal intelligence analysis system:

(i) It will automatically or with human intervention analyse and intelligently extract semantically meaningful information from aggregated data sets,

(ii) Data will be aggregated from across multiple and mixed format sources,

(iii) It will support the analysts in the process of thinking and reasoning for situation reconstruction, discovery and insight generation,

(iv) It will enable the fluid and playful creation and assembly of tentative explanations of crimes that can evolve into rigorous arguments;

(iv) VALCRI will incorporate designs that protect against the weaknesses of human cognitive biases and abuses that may arise from accidental (human error) or deliberate violations of ethical, legal and privacy principles."

(Source: http://www.valcri.org/introduction-to-valcri/)

3. The IEB

The VALCRI IEB is composed of experienced ethics advisers who can help identify and deal with several ethical concerns issues of the VALCRI project. These advisers have no conflict of interest with the partners in this project nor with its proposed user community. IEB members are external to the host institution and operate totally independently in forming their assessment using information they researched and information provided by the project.

Each member of the IEB has been working in the field of ICT ethics for more than ten years, has written extensively in the field, and has done work on ethics for International Professional computing organizations such as the International Federation for Information Processing (IFIP) and for national and local professional organizations. The members are:

- PhD Kai K. Kimppa, Vice Chair of and Finnish National Representative to IFIP SIG 9.2.2: Special Interest Group on Framework on Ethics of Computing, Member (and former Secretary of six years) of IFIP WG 9.2: Social Accountability and Computing, Vice Chair of and National Representative to IFIP TC 9 ICT and Society, Chair of

FIPA (Finnish Information Processing Association) Ethics Working Group, kakimppa@utu.fi.

- Professor Norberto Patrignani, Senior Associate Lecturer of "Computer Ethics" at Graduate School of Politecnico di Torino, Italian National Representative to IFIP TC 9 ICT and Society, norberto.patrignani@polito.it.

3.1. The role of the IEB in the VALCRI Project

The role of IEB in an EU project is formally described as: *"The 'European Parliamentary Technology Assessment' network(*) ensuring that the artefacts developed are socially and ethically desirable, perform reliably and to specification, and that any potential risks, dual uses, or other issues are identified and mitigated before release."* (see European Parliamentary Technology Assessment network,

http://www.eptanetwork.org/ Accessed 27/1/2017).

The IEB has worked independently and not on behalf of any organization to help identify and make sure that the development team deals correctly with ethics issues that may arise from this project both during development and in production. IEB have worked to assist the project team in applying the highest ethical standards to the project.

In accordance with the ToR IEB have worked to determine if they comply with ethical principles and applicable international, EU and national law (in particular, EU Directive 95/46/EC).

http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:01995L0046-20031120&qid=1399635560429&from=EN

This is one of the formal IEB reviews of the VALCRI product and process up to TRL 6, a simulated operational environment and an INTEGRATED Prototype system.

The IEB has been involved in the project from the beginning, participating and providing guidance, aware and potential difficulties. Initially there seemed to be some distractions and primary focus on the complex technical issues of the system. Most participants have begun to understand the project support nature of the IEB work and the VALCRI team has even established a technical Security, Ethics, Privacy and Law (SEPL) support group to help incorporate IEB concerns into the project. For the most part the VALCRI team are addressing (more or less) concerns IEB has raised.

IEB main general concerns are:

- communications across multiple teams and through technology Web site has been difficult but along the project is has been addressed;

- language barrier, not all partners are native English speakers and different profession understand terms differently, or do not understand each other's terms as well.

3.2. Interactions with VALCRI team

IEB has reported concerns and suggestions to the project staff on a regular basis, on ethics concerns as they arise. Initially it was not always clear that IEB is an essential part of project management structure and why it was not invited to all meetings. These concerns have, however, been addressed along the life of the project.

3.2.1. Project meetings participation by IEB

To maintain an overview of the work, the IEB has met with the Project team throughout the course of the project and indicated potential problems that might arise and how they can be addressed. The IEB has attended and reported at:

- Kick-off meeting, Middlesex University, London, UK (May 2014)
- Barcelona meeting (October 2014)
- Linkoping meeting (February 2015)
- Krakow meeting (June 2015)
- Graz meeting (September 2015)
- Malta meeting (February 2016)
- Vienna meeting (June 2016)
- Warsaw meeting (October 2016)
- Berlin (May 2017)
- Nice (November 2017)

and (now planned) the last project meeting in

- London (June 2018).

At each of these meetings the IEB had board meetings, attended development team meetings and gave a summary report making suggestions about things that needed to be addressed in the project.

3.2.2. Other interactions

The IEB had other interactions like:

- email communications with key members of the project (Ifan Shepherd, I.Shepherd@mdx.ac.uk, was indicated as the point of contact with IEB)
- telephone conference with Ifan Shepherd
- the VALCRI team has used the IEB board at docs.valcri.org to provide IEB with documents specifically targeted for IEB
- Feedback Forum https://internal.valcri.org/projects/valcri/boards.

4. Main ethics issues

The VALCRI project, like most of the projects based on BigData collection, analysis, and visualization (see fig.1), presents many ethics issues.

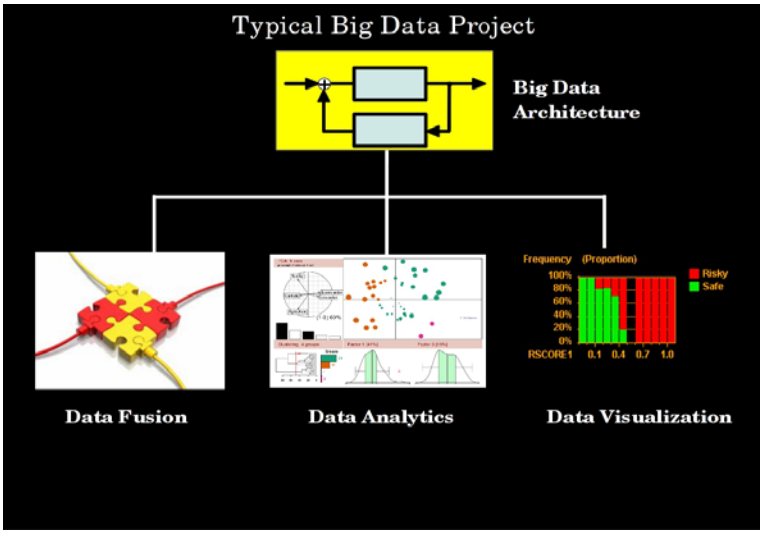


fig.1 – Typical Big Data Project

The main ethical issues are usually known as: *human beings involvement*, *protection of personal data*, *potential misuse / abuse*:

– *Human beings involvement*

Of course, due to the nature of the project, human beings will be involved since a lot of data will be related to physical persons. In this case the typical requirement of "informed consent" will be difficult to apply, not only because the subjects are not aware about the investigation, but also because the amount of data visualised in the visual analytics system - the system does handle a variety of data in large amounts;

– *Protection of personal data*

The VALCRI project involves personal data collection, processing, analysis, and visualization. This data used in the investigations, will be sensitive personal data (related e.g. to health, sexual life-style, ethnicity, political opinion, religious, or philosophical convinction), biometric information, and of course involves tracking or observation of participants, including not only data collected during the investigation but mostly further processing of previously collected personal data (the so-called secondary use);

– *Potential abuse / misuse*

The VALCRI system, once implemented will be a powerful tool that not only risks the malevolent, or criminal use by non-authorized people but also to be improperly used by authorized people. It is a basic principle of *ethical design* that a system installs safeguards against misuse and abuse by rogue users who may gain access to it. The project

should embed software measures to prevent dual use/misuse, whether accidental or on purpose.

In the following the main ethical issues will be detailed using this simple model for IEB concerns (see fig.2).

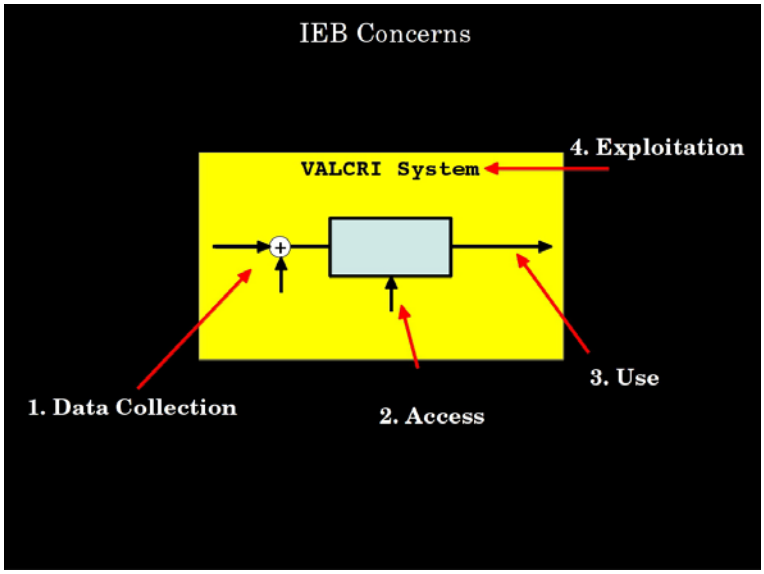


fig.2 – IEB concerns

4.1. Data Collection

The typical requirements, in a research environment using significant amount of personal data, are related to the *informed consent*, the justification of the use of such data and the *approvals of competent authorities*. In the case of VALCRI project, the IEB assumes that these are not applicable in this environment. It is also important to underline that 'data has a context' so data need to be related to *metadata*.

Data has also a dimension of *reliability / uncertainty* that could compromise the quality of human judgements. Also *data provenance* is important since it could be the result of previous analysis steps. After the data collection it is fundamental to annotate it. The annotation of the collection requires noting the varying reliability of the data before the analysis and interpretation of the results.

So the IEB *has asked the VALCRI team to comply with the following:*

- *Compliance*

to confirm that they comply with national and EU legislation about the procedures that will be implemented for data collection, storage, protection, retention and destruction;

- *Metadata*

to ensure, where applicable, that the metadata (source, time, location, etc.) of previously collected data is stored in order to provide the appropriate context to data;

- *Data Reliability / Uncertainty*

failure to address "data uncertainty" means that critical elements like the quality of judgements, the ability to trace the reasoning process is further compromised by the failure to note the differing levels of competences of the analysts;

- *Data Provenance*

data provenance is already a part of being an analyst, but the system should track data provenance for ethical auditing of decisions (data has history, context, provenance means you know what have been done to data).

4.2. Access

Of course the potentiality of VALCRI system requires the highest level of *security* measures provided by technologies and related processes in order to prevent its use by *non-authorized people*.

There is also a more critical dimension related to the access to this kind of systems: the potential misuse / abuse of authorized people of the VALCRI system: the risk of using VALCRI as an "automatic decision system" and the use (by *authorized people*) for purposes different from legitimate investigations.

In the first case the main measures should be in the definition of the mission of the system and in the software processes. IEB recommends a general statement that "**VALCRI is a Decision Support System**" and that the software should be treated as a "decision support system" and not an "automatic decision system". This is in accordance with the principles of Directive 95/46. Not making judgements automatically: preventing crimes by "mitigating measures" initiated by "trigger events" should not be automated but the final judgement should be always in the hands of a human, human beings should be informed of all the steps that software have detected by "digesting" big data. This constraint is consistent with the strict limitations on automatic decisions of Article 7 - Council Framework Decision 2008/977 JHA.

In the second case it is critical to discourage authorized people to use **the VALCRI system for purposes different from legitimate investigations.** The system should be developed in two concentric circles, the inner circle is the system itself, the outer circle is the "meta-system" (Secure-Logs) that logs all access and store events (see fig.3).

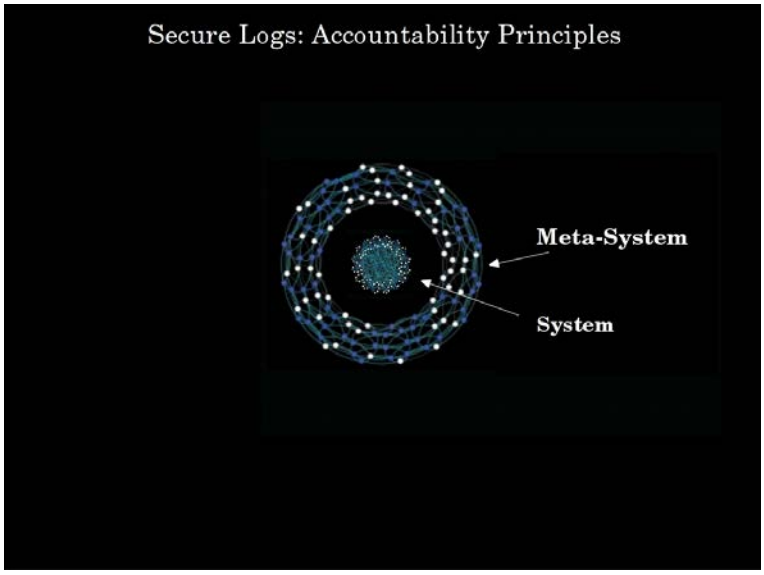


fig.3 - Secure Logs

In this way it will be ensured that the system is for "Authorized Persons only" by means of "Strong Authentication" technologies and the "Accountability Principles" will apply by logging:

- All Accesses (Identity, Level of Competence) and Operations and
- All Changes to the System.

All Logs should be Time-stamped, Encrypted, and kept in a Secure Place (a different database). Logs are Read-Only, and Opening Logs requires "4-Hands-Authorization" (the joint agreement of two appropriate people)

So the IEB *has asked to VALCRI team to comply with the following:*

- ***Security***

to maximise the security measures and protections to prevent malevolent, or criminal use by *non-authorized people*;

- ***Misuse***

to avoid that the system is used (by *authorized people*) as an (automatic) decision system (delegation to technology);

- ***Abuse***

to adopt the appropriate measures to prevent the improper use by *authorized people* (secure logs: accountability principles).

4.3. Use

The analysis of Exabytes of data related to human beings can raise another ethics issue that should be taken into account: *incidental findings*. If, during the investigation and consequent analysis of data, the system highlights personal, familiar, medical or psychiatric or other conditions that are unrelated to the

undergoing investigation, but that can have serious consequences on the persons involved, will they be informed? This ethics issue in medical research is one of the most serious problems but also in this scenario, where a powerful analytical system is processing an amount of data that is beyond the human capability of control, it needs to be addressed in some way.

In systems like VALCRI it is necessary to recall that "computer output is not truth", to this purpose it is very important the visibility of the logic reasoning. Computer output is not truth but it is the result of *uncertain human input* and *calculations*. The following points should be considered: human domain, cognitive bias and hypothesis and predictions have a tendency to become fact. So computers do not provide the answer, the systems facilitate multiple organizations of data as Decision Support System. There needs to be additional reasons for pursuing combining groups of facts into a hypothesis and prediction. So it is important to have "traceable reasons":

- record all changes of reasoning
- others can learn from what you did
- no one can question your decisions
- facilitates formal review
- addressing the "computer verifiability problem": the fact that "we can't process all the data from a computer" does not entail blind trust
- common language problem: suggest use of operational definitions
- need to include judgement/justification in using data
- use known standards related to quality, accuracy, reliability of input data
- relate the data to the context which colours the understanding of the data. The context needs to be recorded as part of the logic trail.

It is also important for transparency that the visibility of the reasoning process of the analysis is available: the situation reconstruction process is visible, exposing its underlying assumptions to the analysts, open to inspection by co-workers (independent reviews).

Another important aspect is the minimization of cognitive bias risks (e.g. stigmatization of particular segments of society).

The role of Chief Data Officer (or similar) is also essential for this kind of systems (e.g. who is responsible in case of a break?). It is recommended to have a CDO in any situation where VALCRI is used.

For systems like VALCRI the training phase is fundamental and many ethical issues could be mitigated with the proper materials and training.

So the IEB *has asked the VALCRI team to comply with*:

- Incidental findings

to provide an incidental finding policy including the potential outcome and the related actions;

- *Visibility of the reasoning process*

to provide the visibility of situation reconstruction process, exposing its underlying assumptions to the analysts, open to inspection to co-workers and that will be enable independent reviews.

- *Cognitive bias*

to provide measures for minimizing the risks of cognitive bias;

- *Chief Data Officer*

to recommend the identification of a Chief Data Officer in any situation where VALCRI is used;

- *Training*

to provide the proper documentation and training (that could mitigate many of the ethical issues) to users.

4.4. Exploitation

A complex system like VALCRI requires a detailed Risk Assessment effort. This assessment can be modelled on the ethical risks identified in deliverable D3.4. Since the VALCRI project involve the development of technologies or the creation of information that could have severe negative impacts on human rights (e.g. privacy, stigmatization, discrimination) this risk assessment is critical.

From some points of view the VALCRI system can be seen as a powerful "data-correlation technology" so the exploitation plan should specify that the system is for a particular user group (police intelligence agencies) only. Also due to the power of the system and its potential negative impacts on society, IEB strongly recommend that post-project exploitation plan is consistent with the ethical principles and aspirations contained in the proposal and consequent plan agreed with the EU Commission.

So the IEB *has asked the VALCRI team to comply with*:

- Risk assessment

to define a risk assessment to be undertaken before the final delivery of the system;

- Exploitation Plan

to ensure that the post-project exploitation plan is consistent with the ethical principles and aspirations contained in the proposal and consequent plan agreed with the EU Commission.

5. Answers to ethics issues

In order to check the status of the interactions between the IEB and the VALCRI team in the following are reported the main ethics issues and the main related questions posed by the IEB team, and the answers the IEB has received. For each of the identified ethics issue is reported the requirement and the answer received.

5.1. Data Collection

- Compliance

to confirm that they comply with national and EU legislation about the procedures that will be implemented for data collection, storage, protection, retention and destruction;

Answer: The compliance with national and EU legislation of the procedures for data collection, storage, protection, retention and destruction has been confirmed by the VALCRI team. It will be based on policies of the organization as delegated to the organization's Data Protection Officer. Also, the VALCRI team has demonstrated PET, Privacy Enhancing Technologies and techniques in the lower TRL components.

- Metadata

to ensure, where applicable, that the metadata (source, time, location, etc.) of previously collected data is stored in order to provide the appropriate context to data;

Answer: The VALCRI team confirms the inclusion of metadata where available. The metadata can be interrogated through the data space.

- Data Reliability / Uncertainty

failure to address "data uncertainty" means that critical elements like the quality of judgements, the ability to trace the reasoning process is further compromised by the failure to note the differing levels of competences of the analysts;

Answer: The VALCRI team confirms that all changes to the system are recorded (including the original item and what it was changed to, the date of change, who changed it, and the analysts level of competence). Colour codes have been implemented for characterizing the different "quality" of data, but have not yet been integrated into the system.

- Data Provenance

data provenance is already a part of being an analyst, but the system should track data provenance for ethical auditing of decisions (data has history, context, provenance means one knows what have been done to data).

Answer: The VALCRI team confirms that the logging capability includes also data provenance (in a limited way). The main system has a full secure data access logging, analytic provenance is captured and logged into the main system, but not displayed. The reasoning

provenance is captured and displayed in a rudimentary way, on the main system. The analytic reasoning is captured and displayed in Version 1.

5.2. Access

- *Security*

to maximise the security measures and protections to prevent malevolent, or criminal use by *non-authorized people*;

Answer: The VALCRI team confirms that the best security technologies (High Assurance Logging and Auditing, HALA; OpenPMF) have been included into the system for access control.

- *Misuse*

to avoid that the system is used (by *authorized people*) as an (automatic) decision system (delegation to technology), as the VALCRI system is a decision *support* system;

Answer: In one VALCRI document IEB was able to find that: *"VALCRI is a system that will help law enforcement agencies (LEAs) in decision making, therefore transparency of the process is vital. The system should be transparent so that LEAs know why specific data was collected. In addition to this, the system will be handling some personal data. There should be transparency about the handling of personal information to users. VALCRI will make sure that the whole process is transparent, how the LEAs came to certain decisions so that LEAs will be accountable for decisions made and the ways in which certain pieces of data were used in the decision-making process."*

The IEB main concern, that the system should be used as a Decision Support System and not an (automatic) decision making system needs to be addressed more firmly and clearly in the specifications and in software constraints.

The IEB recognises that the VALCRI design team have worked towards a system that augments rather than replaces human decision making; also, the system design and investigative situations require that investigator to search for data / evidence to support the claims that are made.

The VALCRI team confirms the inclusion of recommendations and warnings in the training material.

- *Abuse*

to adopt the appropriate measures to prevent the improper use by *authorized people* (secure logs: accountability principles).

Answer: This logging capability is the most critical requirement for systems like VALCRI. The VALCRI team confirms that the logging is ensured, all logs are time-stamped, encrypted, and kept in a secure place (a different database that is not accessible even to the administrator of the system). Logs are read-only, but the requirement that opening logs needs a "4-Hands-Authorization" (the joint agreement of two appropriate people) is

left to local configurations (end user environment). It is also impossible a dump/export of data.

5.3. Use

- *Incidental findings*

to provide an incidental finding policy including the potential outcome and the related actions;

Answer: The VALCRI team has not provided any incidental finding policy, although the IEB has been informed on that that at least among the *current* police forces (especially WMP) there are policies on how to handle these kinds of situations. In other situations this may not be true, however. Handling this has been suggested to be included in the training material, which is not an optimal solution from the IEB's perspective.

- *Visibility of the reasoning process*

to provide the visibility of situation reconstruction process, exposing its underlying assumptions to the analysts, open to inspection to co-workers and that will be enable independent reviews.

Answer: The VALCRI team confirms that the reasoning process of the analyst is partially logged and is open to inspections.

- *Cognitive bias*

to provide measures for minimizing the risks of cognitive bias;

Answer: Multiple tools are being developed so that the analyst sees the situation from various perspectives that lessen the likelihood. The training material includes warnings about this. Less strong connections will be used, so that things not connected are not connected accidentally. Also, some connections which do at first glance look like they are relevant, the system can suggest might be, which could prevent bias.

How does cognitive bias activate (initiate) ethical problems or issues? Origins: personnel originally biased already (e.g. cultural biases). Actions: work on removal of PII from data analysis, associative search fan changed from thick dark lines to dotted lines, associative search fan: additional data that would not normally be used to show other possible avenues.

Although a group is working on this, IEB still considers the implementation to be an OPEN ISSUE

- *Chief Data Officer*

to recommend the identification of a Chief Data Officer in any situation where VALCRI is used;

Answer: OPEN ISSUE

- Training

to provide the proper documentation and training (that could mitigate many of the ethical issues) to users.

Answer: The VALCRI team confirms that training material includes also many ethical recommendations, e.g. how to recognise the occurrence of an ethical issue, and what to do about it?

5.4. Exploitation

- Risk assessment

to define a risk assessment to be undertaken before the final delivery of the system;

Answer: IEB still have not seen a risk assessment for the whole project from the ethics perspective (for example the risk of discrimination need to be addressed). This is critical. The VALCRI team has informed the IEB that they hope to have this done by June 25th. Nonetheless, currently IEB still considers this to be an OPEN ISSUE.

- Exploitation Plan

to ensure that the post-project exploitation plan is consistent with the ethical principles and aspirations contained in the proposal and consequent plan agreed with the EU Commission.

Answer: At the time of writing the VALCRI team is still investigating possible exploitations of the project results. There are still many unknowns like:

- Restrictions from European authorities, National authorities, etc. (some of the critical components are developed in specific countries);

- The system will be at TRL5 (technology **validated** in relevant environment) by the end of the project. As the project moves forward, the validation level needs to be upgraded from TRL5.

- Who will take care of taking the complete system to the market? With what kind of funding? New version with other data than West-Midlands Police - funding from MDX to go further, in advance of that doing testing with the Metropolitan Police and Pasco, sheriff's office in Florida. Testing how readily VALCRI can take their real data in the system. If these work, route to market available. If not, then limitations. Also other companies have been approached by the project (Unisys, for example), still no end result however. So from the perspective of the VALCRI project there is a need to emphasize to the EU that IEB do not currently know in whose hands the system might eventually end up in.

- Will the VALCRI be available only "inside the wall" of a law enforcement agency? Or will be available also via Software-as-a-Service (SaaS)? Currently it seems that the police forces only want the system "inside the wall", however. Many of the partners are especially sensitive about to whom VALCRI system should be sold.

6. Concluding remarks

After some initial difficulties in IEB-VALCRI interaction the IEB can confirm that most of the ethical issues raised along the life of the project have been accepted by the VALCRI team. The goal of an approach "Ethics-by-Design" is still far but some progress has been reached. For example the establishment of the SEPL team in the VALCRI project (set up some time into the lifecycle of the project), as a technical Security, Ethics, Privacy and Law (SEPL) development group, is an important help to incorporate IEB concerns into the project.

Unfortunately it seems like the ethical issues still have been given a secondary place with respect to the technical developments.

Some issues are still open like: the lack of holistic view of the system (it is not clear when the first integration of the system will be completed). Main developments of the project are still undergoing so many of the questions and issues raised by the IEB are still open.

The IEB recognises the importance of the inclusion of a dedicated work package (WP3) to Human Issues. In particular the deliverable D3.4 "Human Issues Framework", where: *"The framework combines human cognition, bias mitigation, social and legal factors into a single principled framework that developers can use to guide and specify the system design. It will identify the ethical, legal and privacy issues that the new technology must address or trade-off, and examine how technology might hinder human sense-making and the activation of cognitive biases*". And in the first release of the deliverable 3.4: *"... The Human Issues Framework brings together these diverse human issues into a single framework that can be used in a practical way for designing the VALCRI system."*

In a way the work of the IEB is strongly connected with this WP3. The aim of the project has been to make a system as little misusable as possible.

Ethics-by-Design framework will be needed:

- Research in designing for ethical consideration is at the bleeding edge

- There are no known *specific* methods for translating ethics issue into design

- Experts needed to identify the ethics issues and how they manifest themselves in practical ways, translate them in tangible problems that computer software designers can build (specifications).

As the project has been pushing boundaries, and because there are no ready frameworks, what is needed in future similar projects is an "ethics designer" or "ethics specifier" who can take the ethical questions as they arise and then specify them in a format that the implementers can understand and create. A good example in VALCRI has been the IEB's suggestion to introduce a secure logging functionality in order to address the ethics issue of potential abuse by authorized personnel. IEB suggests that such position will be included in future projects

with similar novel topics that have ethical questions that need to be resolved during the project.

VALCRI Risk Assessment for IEB-related issues

			1 = Low; 5 = High		1 = Lo risk; 10 = Hi risk	
	IEB Requirement	**Consequences**	**Likelihood**	**Impact**	**Risk Score**	**Summary of Actions Taken in VALCRI**
DATA COLLECTION						
Compliance	Compliance with national and EU legislation about the procedures that will be implemented for data collection, storage, protection, retention and destruction.	Negative impact on uptake of VALCRI within the EC.	3	3	6	SEPL team (ULD and KUL) have carried out a comparative analysis of the legal framework across EU countries with the more rigorous and less rigorous privacy and data protection legislation and regulations and has incorporated the legal requirements into software data protection and access through OS's OpenPMF and HALA.
Metadata	Ensure, where applicable, that the metadata (source, time, location, etc.) of previously collected data is stored	Anaysts deprived of data context.	3	4	7	Meta data for context is important but not sufficient for helping analysts understand the situation and truth. VALCRI is designed such that if the meta data is available, it will be read into VALCRI system
Data Reliability / Uncertainty	Address data uncertainty, including critical elements such as the quality of analyst judgements, the ability to trace the reasoning process, and the differing levels of competences of the analysts.	Makes it difficult to trace the evolution of analysts' decision making.	3	3	6	VALCRI has: (1) adopted 5 x 5 x 5 data reliability framework; and (2) implemented an analytic provenance recording function to trace 'conclusion pathways'.
Data Provenance	Track data provenance. (Data has history, context, provenance means knowledge of what has been done to data.)	Unable to undertake ethical auditing of analyst decisions.	2	2	4	VALCRI has implemented logging system for all actions and is recorded in the secure GrayLog system. An additional provenance recording system to track analysts' decisions has also been implemented but at a much less mature level

ACCESS						
Security	Maximise the security measures and protections to prevent malevolent, or criminal use by *non-authorized people.*	Absence of such measures could lead to compromise of the system from without.	2	2	4	VALCRI has implemented a password-based security log-in system to protect it from non-authorised use. Secure encapsulation of VALCRI and its data communications between nodes and ports has been demonstrated in a stand-alone configuration of VALCRI using the the HALA (High Assurance Logging and Audit) system.
Abuse	Adopt appropriate measures to prevent improper use by *authorized people* (secure logs: accountability principles).	Absence of such measures could lead to compromise of the system from within.	2	2	4	VALCRI has implemented a detailed logging system (GrayLog) for tracking all actions by authorised users. GrayLog is also imutable and cannot be edited by the system administrators.
Misuse	Ensure that the system is not used (by *authorized people*) as an (automatic) decision making system (delegation to technology).	Absence of such measures could lead to the severe erosion of public trust, and to significant problems in legal proceedings.	2	2	4	VALCRI is and has been designed so that "humans decide, machines do the heavy lifting". In this way,the role of 'black box automation' such as the Machine Learning semantic similarity searching functionsare designed for VALCRI to augment than replace decision making by making it possible for the human analysts to quickly retrieve potential records and present them for inspection to ascertain if they can or should be included in one's decision making or assessments. VALCRI does not have an automated decision making capability.

USE						
Incidental findings	Provide an incidental findings policy including the potential outcome and the related actions. Incidental findings can lead to (i) positive benefits such as faster apprehension of perpetrators, or (ii) negative outcomes that lead to discovery of data and relationships that reveal situations that should be left unknown.	Absence of an incidental findings policy can lead to un-regulated behaviours on making spurious associations appear real, i.e. 'discovering links where there are none'; or encouraging searches across security or organsiational boundaries, creating people profiles that can be mis-used or abused.	3	4	7	VALCRI is designed to help analysts discover criminal relationships they could not discover easily before. Unintended and incidental discoveries of relationships requires knowledge in the heads of analysts. At a basic security level, VALCRI has implemented standard access-notification protocols in the security and data access system using OpenPMF. If classified data has been accessed, and depending on its sensitivity, the data may be presented or withheld, the requestor notified if the data is available, available but not displayable, not notified that the data exists; and the owner of the data may or may not be notified. Just like the Data Protection Policy and Procedures, Incidental Findings Policies should be developed and implemented by the organisation, and not by the VALCRI project.
Visibility of the reasoning process	Provide visibility for the situation reconstruction process, exposing its underlying assumptions to the analysts, and being open to inspection to co-workers, so that it will enable independent reviews.	Absence of such visibility may compromise the ability to undertake independent reviews.	2	4	6	In VALCRI the situation re-construction process is supported by the capability to collate and assemble data, analyses and reports into unique sequences that form the basis for explanatory narratives. These assemblages are designed to to enable visual inspection by colleagaues and show the conclusion pathways and to display the data and reports to show the grounds for which claims are made, the conditions, assumptions, and qualifiers, and the respective backing data to support the claim.

Cognitive bias	Provide measures for minimizing the risks of cognitive bias.	Absence of such measures could lead to analysts behaving in cognitively biased ways.	2	2	4	Cognitive bias cannot be totally eliminated. Experts by virtue of their expertise are automatically biased towards search for or use of certain critical cues. VALCRI has attempted to address the more common availability and confirmation biases by adopting the design approach of showing and revealing the data being sought and by joining data.
Chief Data Officer	Recommend the identification of a Chief Data Officer in any situation where VALCRI is used.	Absence of a Chief Data Officer could lead to misuse of data (e.g. absence of purpose limitation).	1	5	6	The appointment of a 'chief data officer' or 'data protection officer' is a decision made by the organisation, not by the VALCRI project.
Training	Provide proper documentation and training that could mitigate many of the ethical issues among users.	Absecnce of proper documentaiton and training could lead to both the under-use and mis-use of VALCRI functionality, with potentially serious knock-on effects.	3	3	6	VALCRI has developed a set of 'how to use VALCRI' user manuals to ensure end-users will know how and where to access various functions. In addition, plans are under-foot to investigate how police use VALCRI when presented with real data develop new analytic strategies for optimising VALCRI during investigative sense making and problem solving.

EXPLOITATION						
Exploitation Plan	Ensure that post-project exploitation plans are consistent with the ethical principles and aspirations contained in the proposal and consequent plan agreed with the EU Commission.	Absence of a post-project exploitation plan to ensure consistency with ethical principles can result in the mis-use or abuse of the technology.	3	2	5	VALCRI consortium partners have agreed on a 9-point post-project exploitation plan which adheres to EC policies on ethical use, dual-use including sale to countries with repressive regimes, while ensuring that the exploitation plan is not so restrictive as to stop or hinder the commercial exploitation, forcing partners to leave the exploitable technolgy on the shelf.

NOTE: The impact and likelihood values refer to risk of the consequences arising if the project does not put in place suitable measures in relation to specific issue raised by the IEB.

ANNEX E
EXPLOITATION PRINCIPLES AND COMMERCIAL IP LIST

Technical Note 14.6

VALCRI: Exploitation Principles and Commercial IP List

VERSION 1.0

Date submitted: 14 May 2018

Dissemination Level: PU / PP / RE / CO

WORK PACKAGE: WP14 Dissemination and Exploitation

WORK PACKAGE LEADER: Margit Pohl

Report Prepared by: Ifan Shepherd (MU), William Wong (MU), Neesha Kodagoda (MU)

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no FP7-IP-608142.

Programme:	FP7 Theme 10: Security
Grant Agreement No.	FP7-IP-608142
Project Acronym:	VALCRI – Visual Analytics for sense-making in Criminal Intelligence Analysis
Project Co-ordinator:	Middlesex University – Professor B.L. William Wong

Deliverable No.	D 14.6
Document Title:	VALCRI: Exploitation Principles and Plans
Version:	1.0
Status:	Final
Document Manager:	Ifan Shepherd
Organisation:	MU
Date:	14 May 2018
Filename:	14.6 – Exploitation Plan V3

Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	**X**
RE	Restricted to a group specified by the consortium participants (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Keywords:

VALCRI Evaluation, End User Feedback

SECTION 2

Annex E:

EXPLOITATION PRINCIPLES AND COMMERCIAL IP LIST
Period 3, M01 – M48, May 2014 – May 2018

Table of Contents

EXPLOITATION PRINCIPLES AND COMMERCIAL IP LIST

This report makes an assessment on Intellectual Property (IP) origination within the life of the project and their individual and joint ownership. The CA and Deliverable 14.2 do not cover all IPR regulations that might be useful for commercial exploitation purposes. Also the CA does not cover agreement ending the lifetime of the project. To maximise the potential of VALCRI being exploited individually or jointly the following steps were undertaken.

1. VALCRI Commercial IP Exploitation Principles developed
2. Explored partner intent in exploiting IP commercially
3. Partner statement on commercially exploiting IP
4. IP origination and individual or joint ownerships

Section 1 - VALCRI – Commercial IP Exploitation Principles

A set of 9 commercial IPs were agreed by the consortium on the 25 November 2017 following a consortium meeting. Please see attached Annex E, Appendix A.

Section 2 - Partner intent in exploiting commercial IP/ other research purposes

Apart from the law enforcement agencies (LEAs), we have captured to the best of our knowledge partners commercial IP during the project lifetime and their intent to commercially exploit the IP individually and/or jointly. We have also captured if partners intent to exploit knowledge gained for future research.

Out of the 13 partners, 9 partners have developed exploitable IP. Of these, 6 directly plan to exploit their commercial IP. One partner plans to exploit their IP based on commercial IP exploitation principle 2. The intent of partners are described in section 3.

The commercial IP and their individual or joint ownership are listed in section 4 to the best of partners knowledge.

No	Partner	Commercial IP	Intent to exploit commercial IP	Intent to exploit for research purpose
1	AES	IP listed	Intend to exploit	Intend to exploit
2	CITY	IP listed	VALCRI Commercial IP Exploitation Principles (P2)	Intend to exploit
3	FHG	IP listed	Intend to exploit	Intend to exploit
4	TUGraz	No commercial IP		Intend to exploit
5	IINT	IP listed	Intend to exploit	
6	KUL	No commercial IP		Intend to exploit
7	LIU	IP listed		Intend to exploit
8	MU	IP listed	Intend to exploit	Intend to exploit
9	OS	IP listed	Intend to exploit	
10	SPACE	IP listed	Intend to exploit	Intend to exploit
11	TUW	No commercial IP		Intend to exploit
12	UKON	IP listed		Intend to exploit
13	ULD	No commercial IP		Intend to exploit

Section 3 - Partner **statement on commercially exploiting IP**

Partner	Statement on Intent to exploit commercial IP	Statement on intent to exploit for research purpose
AES	▪ AES will continue to build and refine the analyst training course and work with one or more universities to gain accreditation. ▪ AES will continue to develop and refine the Financial Explorer and, in the process of integrating it into their data mining workbench tool, Authority Miner®, develop if from its current TRL4 to TRL9.	▪ AES intend to utilise the anonymised data set within other research projects on condition that WMP provide written agreement. The data sets will be enhanced to accommodate differing research questions.
CITY		Plans to exploit joint IP based on VALCRI Commercial IP Exploitation Principles (P2) and intent to exploit for future research work and publications.
FHG	Fraunhofer IDMT's preferred business model and cooperation form are ▪ Direct technology marketing of research results via licensing of software components and patents to commercial customers for integration, or indirect technology marketing via integration partners ▪ Contractual R&D or subcontracting for requirement analysis and system design, prototype development, evaluation, and consulting ▪ Collaborative and joint R&D in the aforementioned domains We are interested in and depend on using and extending our technological know-how, contributing to beyond-state-of-the-art activities, and we have a strong interest in long-term cooperation with partners. Within VALCRI, Fraunhofer IDMT's primary R&D domains and results are related to visual analysis, unstructured data storage (including support for OpenPMF from ObjectSecurity) and Privacy-Enhancing Technologies (PET), the latter including adaptive anonymization tools for data import and export, and h.264 video face encryption. In addition, we provided a requirement and system design tool (RDT) for the project. Corresponding with our aforementioned strategic goals, we intend to further improve (via R&D) and license the aforementioned results in various domains, including LEA, critical infrastructure protection, Smart Cities and IoT applications, all of which include the need for visual analysis, security and privacy solutions. As	

	for the RDT, we plan to use it as a tool for future collaborative projects. Beyond individual exploitation, we also see a huge potential in further cooperation with other VALCRI partners, e.g. with ObjectSecurity, ULD, and KUL in the domain of developing security and privacy solutions. VALCRI has provided us a unique opportunity for respective interdisciplinary cooperation within the SEPL group, which has worked very well, and we are very interested to continue this cooperation beyond the project lifetime.	
TUGraz		The **Cognitive Bias Discovering and Mitigation Framework** (CBDMF) is the main contribution of TUGraz in the VALCRI project that will be exploited after the project end. This framework includes psychology-based methods for (1) designing and evaluating a visual analytics environment regarding its cognitive bias mitigation capabilities, (2) methods for operationalisation and measuring cognitive biases, and (3) methods for automatically detecting potentially occurred cognitive biases during the use of a visual analytics environment. The TUGraz team aims to undertake further research in the field of cognitive biases and decision making based on the results achieved in the VALCRI project. Advancing research results and applying them in other fields is a key method to make the work performed in VALCRI sustainable. Overall, the exploitation and sustainability planing follows the **future research interests** of the TUGraz team (Cognitive Science Section) on Decision Making Support: "Research on decision making is based on cognitive processes, competence models, decision heuristics, and cognitive biases. These underlying models and theories are applied for creating new technical concepts that are included in decision support systems. Particularly, user interface designs, the tracking of user behaviour, and feedback techniques are influenced by this research stream. This impacts the user by providing support for valid and effective decisions and the avoidance of cognitive biases. Research on decision making is predominantly applied in the Secure Society domain, but is

		not limited to this area." (http://cognitive-science.at/research/research-agenda/) The **short- and mid-term exploitation plan** mainly targets further research and development of the CBDMF in the context of submitted and future research projects. Currently, two research endeavours are actively pursued in the context of submitted and probable future research projects. ▪ Through the emergence and use of digital and social media, new problems arise. For example, digital echo chambers refer to situations where users seek for confirmation of the own opinion in more or less closed groups of users and content. Another problem concern false information and fake news distributed through new media. Both examples refer to a lack of computer literacy of the users in combination with news and social platforms that do not prevent these problems. Research on discovering and mitigating cognitive biases contribute to the solution of these problems. For example, the design of user interfaces and information presentation can prevent users from thinking errors and wrong opinions. The CBDMF developed in VALCRI will be used as basis for further research in this area. ▪ The automatic detection of cognitive biases while users actively work with a computer system is currently not possible. The CBDMF proposes a methodology how to detect cognitive biases by taking into account user interaction data and operationalised biases. Future research endeavors address the implementation of an indicator that warns user if they run into the danger of a cognitive bias. This has important impact on the quality of knowledge workers and also societal value for end users. The **long-term exploitation plan** includes the aim to commercialise knowledge, methods, and implementation of the CBDMF. The TUGraz team has contacts to business incubators that are specialised on

		the commercialisation of research outcomes. For example, Science Park (http://sciencepark.at/en/) is a business incubator owned by the major universities in Graz that specifically supports the creation of start-up companies around research outcomes. Potential business models that can be built around the CBDMF include the consultancy in all areas where cognitive biases play a negative role, evaluation of technology and its use regarding the danger of cognitive biases, design of user interfaces that prevent cognitive biases, and the implementation of a cognitive bias indicator.
IINT	i-intelligence plans to undertake the following exploitation activities: ▪ The material generated for the VALCRI workshops will be integrated into our training and capacity building programs for public and private sector organisations. This material includes the slides generated on strategic foresight and operational intelligence analysis. ▪ The VALCRI Syllabus (presented in Technical Note TN13.3) will inform future course development efforts. We intend to develop a range of training courses and professionals development programs based on the feedback generated from the project's end users and the participants at the various VALCRI workshops. ▪ The research undertaken on the Human Issues Framework (HIF) will be used to support our consulting practice. Specifically, we have identified a range of strategic and operational challenges using "Five Architectures of Intelligence" model. We plan to address these challenges through the development of new tools and the extension of our consulting concept. ▪ Finally, we intend to continue our research on specific HIF challenges such as uncertainty management and operational process improvement. This research will be used to improve our toolkit as intelligence consultants.	
KUL		We may use the IP as background in other research projects and in teaching.

LIU		We may use the IP as background in other research projects and in teaching.
MU	MU is exploring to exploit the VALCRI IP developed: ▪ MU has undertaken research to identify LEA market size for UK, Europe and USA to attract bridging funding to stabilise and improve current IP it has developed. ▪ The known market size has been used to attract potential funding organisations post-June 18 (end of project) to continue with VALCRI activities. ▪ MU is also working closely with potential LEAs to deploy versions of VALCRI at their premises to draw interest and show capability.	We may use the IP as background in other research projects and in teaching.
OS	ObjectSecurity plans to exploit the VALCRI results in three different configurations: 1. ObjectSecurity plans to contribute security and data protection to a future commercial VALCRI offering of a group of VALCRI partners and, possibly, other organizations, like investors. 2. We plan to exploit the security and data protection related results with VALCRI SEPL partners provide generic security and data protection tools. 3. 3. We are already using or plan to use our security and data protection results, especially OpenPMF and fine grained access control for databases, by ourselves. This includes for example the protection of data in a supply chain risk analysis tool, the protection of healthcare systems, an Intelligent Transport System, environmental data acquisition systems and UAV/satellite data management systems.	
SPACE	SPACE was and is busy trying to exploit commercially the knowledge and software it developed during the project. At this stage, SPACE concentrates on the Belgian LEA market, and is specifically in contact with Federal and Local police forces. Amongst other activities, SPACE also approached other potential LEA customers and large players in order to develop a higher commercial potential.	
TUW		We intent to exploit for future research work and publications.
UKON		We may use the IP as background in other research projects and in teaching.

ULD		As a data protection authority and governmental institution, we do not intend to commercialize any results (i.e. design ideas, software, code, components) and are only interested in the publication/ dissemination of good practice suggestions.

Commercial IP origination and individual or joint ownerships

For the purpose of this exercise we are referring to software code and other embodiments such as training materials as commercial IP only. In total 72 commercially exploitable IP were developed during the lifetime of the project.

Please see list of commercial IP developed by partners during the lifetime of the project to the best of the consortiums knowledge. The listed IP and who owns the listed IPs have been circulated and among the consortium partners to verify individual or joint ownerships. Currently, partners have no conflicts on the listed individual or joint ownership of IP, apart from SPACE which could not immediately see any potential conflict. However, it is preferred to leave the discussion of joint ownership to specific cases, because IP list descriptions were insufficient to judge the exact meaning of each item.

Please see attached Annex E, Appendix B.

Annex E, Appendix 1 - VALCRI – Commercial IP Exploitation Principles

VALCRI – Commercial IP Exploitation Principles

This document sets out the principles by which VALCRI partners will undertake the commercial exploitation of VALCRI-related IP. It supplements the Consortium Agreement in relation to significant exploitation issues to which it pays little or no attention.

1. Because a single exploitation plan for the VACLRI-related IP held by all consortium partners is likely to be unworkable, restrictive and legally complicated, plans will be devised by individual and/or groups of IP-holding partners, focussing on any relevant forms of IP (software, consultancy, training, publications, etc.). Partners may draw up more than one plan, and may involve relevant non-VALCRI partners. The individual plans will be assembled into an edited collection for submission to the EC.
2. Exploitation plans will be based on a clear understanding of what items of IP (foreground and sideground) have been created during the project, and which partners own these items. Gaining access to IP owned by other partners will be resolved through negotiation.
3. The anonymised datasets from WMP may not be included in any exploitation plan.
4. Partners may adopt, individually or jointly, whatever form of business model (sale, leasing, service, consultancy, spin-off, startup, etc.) they believe is best suited to the successful exploitation of their IP. They will agree not to pursue anti-competitive (i.e. monopolistic) behaviour, as defined in Article 101 of the TFEU.
5. All partners, whether individually or in groups, will make every effort to ensure that in exploiting any IP that involves the analysis of data, they will take into consideration all relevant legal and ethical requirements, for example the protection of individual's fundamental rights, such as the right to privacy and personal data protection, and the EC's relevant export regulations. If resources are acquired for the conversion of the VALCRI prototype into an operational crime analysis system, additional efforts will be taken to ensure adherence to all relevant legislation, for example in relation to purpose limitation, the security of personal data, the potential of system misuse, or the handling of incidental findings.
6. The successful exploitation of IP by individual partners or groups of partners is likely to require additional resources (time, effort, funding, personnel, etc.) beyond those provided by the VALCRI project. It will be the responsibility of partners seeking to exploit their IP to acquire any necessary resources required to ensure successful exploitation.
7. Any benefits from IP exploitation activities will accrue to the individual partners or groups of partners who own the IP and to those who invest in the exploitation activities.
8. The exploitation plans will credit the FP7 project funding provided by the EC, and will acknowledge the role of the VALCRI consortium in generating the IP.
9. Plans should indicate a shared intention to maximise the benefits to EU citizens from exploitation, as required by the Commission. In order to provide the EU with notification of such benefits, the collected set of exploitation plans will be submitted to the Project Coordinator by no later than the 1st of April 2018, so they can be included in the Final Project Report.

25 November 2017

Annex E, Appendix 2 - Commercial IP origination and individual or joint ownerships

VALCRI -- IP List (15 May 2018)

For the purpose of this exercise we are referring to software code and other embodiments such as training materials as commercial IP only.
To the best of the consortium's knowledge please see below commercial IP generated during the lifetime of the project and the commercial IP owners and the intent to exploit.

No	Frontend (UI)/ Backend/ Other	Software Part	Commercial IP owned by	Discussed with	Description	Integrated within the VALCRI system?	TRL	Is this VALCRI IP going to be exploited individually or jointly?	If shared VALCRI IP: has agreement been arranged on how IP will be used jointly to exploit?
1	Model	Map Line-ups Framework		CITY	A model of capabilities for making inferences from mapped geospatial data.				
2	Model	Map Line-ups Strategies		CITY TUWIEN	A series of six sensemaking strategies for those making judgments in challenging map line-up tasks				
3	UI	313 - Uncertainty Representation & Visualization		CITY MU AES	Framework for uncertainty representation including a pragmatic model with which uncertainty can be recorded (313), a straightforward process for establishing levels of qualitative uncertainty (APT), and an effective means of visualizing levels of uncertainty captured by the model through appropriate encodings and interaction.	No	TRL3		
4	UI	DDD - Design Framework and Schema	CITY	CITY	Authoring environment for creating and validating Dynamic Design Documents. Based on the litvis framework for literate visualization, it uses a high-level declarative language for generating visualizations and narrative mark-up - making it east to	No	TRL3		

					produce, explain and obtain feedback on visualisation prototypes.				
5	UI	DVQ - Applying Dynamic Visual Queries to Crime Intelligence Analysis		CITY MU	Combining data visualization as input and output to ElasticSearch queries to provide a fast and fluid interface for querying large datasets.	No	TRL2		
6	Backend	UDB /w OpenPMF interface	FHG	FHG	Storage and retrieval of unstructured data + metadata, with support for OS's OpenPMF		TRL3	FHG	
7	Backend	PET: Adaptive Anonymization	FHG	FHG	Reidentification analysis and adaptive anonymization for data import and export		TRL3	FHG	
8	Backend	Video analysis	FHG	FHG	video processing service for object detection (standalone binary version for pre-processing and demonstration purposes)		TRL4	FHG	
9	Backend	PET: Video face encryption	FHG	FHG	Video face encryption to avoid unauthorized person identification e.g. when CCTV recordings are transmitted		TRL3	FHG	
10	UI	Video player	FHG	FHG MU	Show videos which are pre-processed with metadata can be viewed and make new annotations.	Yes	TRL4	FHG	
11	Training	Analyst Training Material	IINT	IINT	Training material to support the training and development of criminal intelligence professionals. This includes slides, templates and other learning materials to support descriptive, predictive and prescriptive reasoning for both strategic and operational intelligence purposes.		TRL8	IINT	
12	Backend	Ontologies for VALCRI	LIU	LIU	A set of top-level ontologies defined on OWL/RDF for the VALCRI domain and application	Yes	TRL4	LIU	
13	Backend	Ontology alignment	LIU	LIU	Method and system for aligning ontologies	Yes	TRL4	LIU	

14	Backend	Ontology backend communication for data analysis	LIU	LIU	System for communication between R and data storage backend, including SPARQL endpoints and data analysis	Yes	TRL4	LIU	
15	Backend	Query templates for ontologies	LIU	LIU	System for SPARQL query templates	Yes	TRL4	LIU	
16	Backend	Query templates for streaming data	LIU	LIU	System for streaming data queries and templates, including management and organisation of templates	Yes	TRL4	LIU	
17	Frontend	Visualisation of data from ontology-based data analysis	LIU	LIU	Experimental visualisation in R of ontology-based data obtained from backend using SPARQL queries	No	TRL4	LIU	
18	Backend	Agnostic, Multi-DB Ingestion Engine	MU	MU	Through configuration files, the ingestion engine is able to crate searchable documents from the ingestion of different police datasets. The data can come from a variety of different sources (e.g., cvs, sql dumps), and be ingested into Mongo, Elastic or Postgres.	Yes	TRL5	MU	
19	Backend	Build and CI pipeline	MU	MU		Yes	TRL5	MU	
20	Backend	UI Framework	MU	MU XI CITY	A UI framework for integrating multiple visualisation components and enabling multiple groups of visualisations to synchronise with a query system to show multiple representations of the same data. The framework allows for multiple users and tracks user interactions - capturing analytic provenance. It is developed using the Meteor.js and Reacts frameworks.	Yes	TRL5	MU	
21	Backend	VALCRI Data REST API	MU	MU LIU	Multi DB Query engine for VALCRI - a middleware for retrieving multiple data types though a series of filters and aggregations.	Yes	TRL5	MU	
22	UI	Argumentation	MU	MU	Narrative and storytelling reasoning thinking space – A view which will allow the analyst to capture reasoning provenance, uncertainty propagation and to	Yes	TRL2	MU	

					develop a narrative based on argumentation.				
23	UI	Associative search	MU	MU AES	Identifying possible list of suspects for open crimes based on closed crimes, proximity to location, temporal and MO using a tree visualization.	Yes	TRL5	MU	
24	UI	Bubble chart	MU	MU	Identifying possible list of suspects for open crimes based on closed crimes, proximity to location, temporal and MO using a bubble visualization.	No	TRL2	MU	
25	UI	Context Cards	MU	MU	Shows more contextual information on an event/ entity and associated information e.g. based on time or space.	Yes	TRL5	MU	
26	UI	Co-offender networks - Harm	MU	Vienna MU AES	The social network of events and associated co-offenders over time and harm.	No	TRL2	MU	
27	UI	Data as objects in transposable spaces (DOTS)	MU	MU	DOTS allow re-structuring and re-organizing of events/ entities from one visual representation to another.	Yes	TRL5	MU	
28	UI	DDD - Dynamic Design Documents	MU	CITY MU	A means of explaining, justifying, exploring and evaluating candidate designs with analysts in the workplace to support ongoing dialogue - between designers and analysts and amongst analysts.	No	TRL3		
29	UI	Donut clusters	MU	MU	Revels more information about the crimes, e.g. if they are open and closed crimes along with their associated victims and defendants (semantically zoom way).	Yes	TRL5	MU	
30	UI	Entity networks	MU	MU	Associations entities and how they are connected (Graph View)	Yes	TRL3	MU	
31	UI	FAVVEs - Faceted Views of Varying Emphasis	MU	CITY MU	A design framework for visualizing multi-perspective small multiple graphics simultaneously in ways that neither clutter nor overload.	No	TRL2		

32	UI	List Browser	MU	MU	Shows a list of event/ entities, along with Overview over features extracted from different events.	Yes	TRL5	MU	
33	UI	Location	MU	MU	Shows geographical information surrounding events/entities, which are represented using multi-layers (NUP boundary, ANPR, Camera).	Yes	TRL5	MU	
34	UI	Provenance	MU	MU	Show provenance of work carried out during analysis.	Yes	TRL3	MU	
35	UI	Schema Line	MU	MU	Examine information manually in chronological order and to identify temporal patterns and relationships.	Yes	TRL5	MU	
36	UI	Shewmap - Statistical Processing Chart - Line graph	MU	CITY MU	Rule based identification of abnormal behaviour for pattern and signal detection based on process history.	Yes	TRL6	MU	
37	UI	Table	MU	MU	Overview over features extracted from different events.	Yes	TRL5	MU	
38	UI	Time	MU	MU	Shows aggregated temporal information surrounding events using a bar chart.	Yes	TRL5	MU	
39	UI	Time Gradient	MU	CITY MU	A diverging colour scheme used to show uncertainty associated with events in criminal intelligence analysis. Incidents that see the greatest change in gradient colour are those for which there is greatest uncertainty.	No	TRL2	MU	
40	UI	Timeline	MU	MU	Shows events belonging to entities in chronological order.	Yes	TRL5	MU	
41	UI	Type	MU	MU	Allows any unique characteristics of events to be aggregated.	Yes	TRL5	MU	
42	UI	Shewmap - Statistical Processing Chart	MU	CITY MU	Rule based identification of abnormal behaviour for spatial patterns in the signal detection based on process history selectable for neighbourhoods.	Yes	TRL6	MU	

43	Backend	Crypto service	OS	OS	The crypto service is responsible for generating X509-based files which are used by the security libraries and tools for secure communication based on TLS between dynamically spawn users' containers. The service naturally complements Docker containers life-cycle management service.			OS	
44	Backend	Docker containers life-cycle management	OS	OS	The Docker container life-cycle management is a generic service with is able to spawn service related and user related Docker containers based on the accessing user credentials. The configuration of the service is very generic and capable of supporting multi-containers cooperation and waiting on start-up of user defined container.			OS	
45	Backend	GlassFish Policy Enforcement Point	OS	OS	The GlassFish Policy Enforcement Point is responsible for connecting GlassFish application server to OpenPMF. It is responsible for automatic security policy updates and for security policy violations notifications.			OS	
46	Backend	Logging	OS	OS	The basic VALCRI logging is based on Graylog which is connected to the OpenPMF logging facility. It will be used in the pilots. The advanced logging is based on HALA with custom developed FPGA solution for secure logging.			OS	
47	Backend	OpenPMF policy authoring and low level policy generators	OS	OS	The OpenPMF policy authoring tool is usable in its GUI form for high-level policy definition. It is capable of invoking various low-level policy generators which generates low-level security policies for secured technology stacks. In VALCRI case that			OS	

					means for PostgreSQL and GlassFish/UDB.				
48	Backend	OpenPMF runtime	OS	OS	The OpenPMF runtime is responsible for security policy and policy enforcement points management. It is responsible for storing low-level policies. Notification of PEPs about policy changes. Notification of interested parties/services about policy violations.			OS	
49	Backend	PostgreSQL Policy Enforcement Point	OS	OS	The PostgreSQL Policy Enforcement Point is responsible for connecting PostgreSQL DB engine to OpenPMF. It is responsible for automatic security policy updates and for security policy violation notifications.			OS	
50	Backend	Privacy Ontology	OS	OS	An ontology for the description of privacy regulations.			OS	
51	Backend	Single-sign on VALCRI UI	OS	OS	The VALCRI login and single sign-on is based on Central Authentication Service which is modified for VALCRI needs			OS	
52	Backend	User management	OS	OS	The VALCRI user management is based on LDAP standard with additional services for user attributes synchronizations between LDAP, OpenPMF stack and OpenPMF PEPs (PostgreSQL PEP).			OS	
53	Backend	High TRL Configurable ingester	SPACE	SPACE	Part of Visual Analytics for Federal Police of Antwerp - the software developed and deployed by SPACE at the FPA. The ingester was tested on Flux24 data. It can also cope with unstructured data sources such as PDFs and most other common text formats and it extracting text from	No	TRL8	SPACE	

					images using OCR (Optical Character Recognition).				
54	Backend	High TRL server implementation	SPACE	SPACE	Part of Visual Analytics for Federal Police of Antwerp - the software developed and deployed by SPACE at the FPA. A server implementation that allows end-users to cross filter over its various different facets.	No	TRL8	SPACE	
55	Backend	ISLP querying	SPACE	SPACE	Enhancing VALCRI's operational efficacy by evaluating a dominant LPA data source - ISLP - for integration the initial work done.	No	TRL8	SPACE	
56	Data	Financial data generator	SPACE	SPACE	The generator generates realistic financial transactions data represented by bank accounts logs of people or companies.	No	TRL8	SPACE	
58	Data	Flux24 data generator	SPACE	SPACE	Flux24 is a data set that federates police reports over all the local police forces in the region of Antwerp. The generator generates a data set in the structure of Flux24, that is realistic enough to trigger many of the issues we discovered in the high TRL software working on the real Flux24 data. The dataset features realistic statistics triggering each of the visualizations, and is big enough to run stress tests.	No	TRL8	SPACE	
59	Integration	Development and testing environment	SPACE	SPACE	The integration platform was used in the early VALCRI prototype, but now is solely used by the high TRL software developed by SPACE. It promotes collaboration, welcomes new components and technology	No	TRL8	SPACE	

					initiatives, provides fast feedback for failure and success, guards the functionality of each of the components and continuously challenges and evolves the interfaces between the components.				
60	Integration	Environment for integration tests	SPACE	SPACE	Environment that provides integration tests over changing data.	No	TRL8	SPACE	
61	UI	High TRL user interface	SPACE	SPACE	Part of Visual Analytics for Federal Police of Antwerp - the software developed and deployed by SPACE at the FPA. A user interface that includes several inter-related analytical components, searching/filter functionalities and data export.	No	TRL8	SPACE	
62	Backend	UKON Data Processing: Clustering	UKON	-	1. transforms output of the projection for clustering 2. applies the clustering algorithms (k-Means, DBSCAN, hierarchical clustering – from external libraries) 3. transforms the output back into a common data structure	Yes	TRL5		
63	Backend	UKON Data Processing: NLP	UKON	-	transforms input into a suitable data structure 1. loads descriptions of processing, concatenates different NLP components (from external libraries) accordingly, loads corresponding ML-Models (from external libraries) 2. - iteratively applies NLP components (from external libraries) and converts their output as required 3. transforms the output back into a common data structure	Yes	TRL5		
64	Backend	UKON Data Processing: Projections	UKON	-	- creates distance matrices where needed 1. transforms input data for projection algorithms (PCA, MDS, t-SNE – from external libraries)	Yes	TRL5		

					2. applies the projection algorithm 3. transforms the output back into a common data structure				
65	UI	Concept Graph	UKON	UKON	Feature relationships based on the shared events.	Yes	TRL5		
66	UI	Crime cluster table (CCT)	UKON	UKON	Crime cluster table comprises of crime cases and clusters based on prominent features or sequence of features to support Comparative Case Analysis.	Yes	TRL5		
67	UI	Pattern Selector	UKON	UKON	Crime pattern similarity (features) based on the shared crimes is shown in the pattern selector.	Yes	TRL5		
68	UI	S3 - Similarity Space Selector	UKON	UKON	A two-dimensional embedding of the crime similarities showing the clustering based on extracted features.	Yes	TRL5		
69	UI	S4 - Sequence Similarity Space Selector	UKON	UKON	Another two-dimensional embedding of the crime pattern similarity (features) based on the shared crimes.	Yes	TRL5		
70	UI	WOC - Weight Observer component	UKON	UKON	Provenance of feature interaction weighting Tracked and configurations are shown.	Yes	TRL5		
71	Backend	UKON Data Processing: Sequential Pattern Mining (SPM)	UKON	AES	Provides methods to mine sequential patterns in feature space.	Yes	TRL5		
72	Backend	UKON Data Processing: Unstructured Data Abstraction	UKON	AES	Encapsulates unstructured data for data analysis and provides input and output functionality.	Yes	TRL5		
73	Backend	UKON Data Processing: Unstructured Data Analysis	UKON	AES	Provides methods to analyse unstructured data.	Yes	TRL5		