



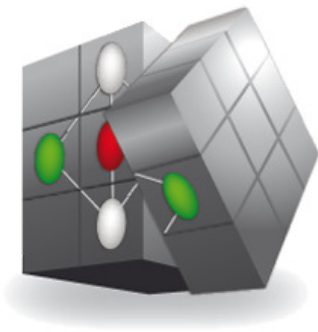
Combining and Uniting Business Intelligence with Semantic Technologies

Acronym: CUBIST

Project No: 257403

Small or Medium-scale Focused Research Project

FP7-ICT-2009-5



cubist

Your Business Intelligence

Duration: 2010/10/01-2013/09/30

CUBIST Exploitation Report v.1

Abstract: n/a

Type	Report
Document ID:	CUBIST D5.2.1
Workpackage:	WP5
Leading partner:	SAP
Author(s):	Frithjof Dau (SAP)
Dissemination level:	PU
Status:	Final
Date:	27 October 2011
Version:	1.0



Versioning and contribution history

Version	Description	Contributors
0.1	First draft with SAP exploitation	Frithjof Dau (SAP)
0.2	Introduction	Frithjof Dau (SAP)
0.3	Incorporated exploitation from HWU and CRSA	Frithjof Dau (SAP)
0.4	Incorporated exploitation from SHU	Frithjof Dau (SAP)
0.5	Incorporated exploitation from SAS	Frithjof Dau (SAP)
0.6	Incorporated review from Ken McLeod	Frithjof Dau (SAP)
0.7	Incorporated exploitation from Ontotext and Innovantage	Frithjof Dau (SAP)
1.0	Received final review from Kenneth M ^c Leod and Cassio Melo. Reviewed version send out again to those partners which have been affected by comments, incorporated their changes.	Frithjof Dau (SAP)

Reviewers

Name	Affiliation
Kenneth M ^c Leod	HWU
Cassio Melo	CRSA



1	INTRODUCTION	4
1.1	EXPLOITABLE RESULTS AND MARKET OVERVIEW	4
1.2	EXPLOITATION CHANNELS AND BENEFICIARIES	6
2	INDIVIDUAL EXPLOITATION PLANS.....	7
2.1	SAP AG	7
2.2	ONTOTEXT.....	9
2.3	SHEFFIELD HALLAM UNIVERSITY.....	11
2.4	CENTRALE RECHERCHE S.A	14
2.5	SPACE APPLICATION SERVICES	16
2.6	HERIOT-WATT UNIVERSITY	20
2.7	INNOVANTAGE.....	22
3	SUMMARY.....	24



1 Introduction

CUBIST is an EC-funded research project which investigates and implements a concept that leverages Business Intelligence to a new level of precise, meaningful and user-friendly analytics of data. CUBIST follows a best-of-breed approach that combines capabilities of Business Intelligence, Semantic Technologies, and Visual Analytics.

In CUBIST, Task 5.2 “Exploitation” will promote and empower exploitation, assessment and broad take-up of CUBIST project results to the target audience and stakeholders. The objective of this task is to develop and execute an exploitation approach for the CUBIST results in relation to the consortium partners. We have started with the following steps:

- First, in order to provide a solid basis for the exploitation of the project results, the initial task is to identify and describe individual exploitable results, and to conduct market examinations for the best use of research results and for creating new business opportunities.
- Second, appropriate means for exploitation, taking into account individual channels per partner, have to be identified. Examples of such channels are internal transfer projects, piloting deployment, improvement of existing product or even new product development.
- Third, beneficiaries of the project’s outcomes have to be detected. This particularly includes the identification of exploitation contacts and stakeholders within the industrial partners for which the results of CUBIST impacts their respective product portfolio, and to identify appropriate courses of action in order to transfer CUBIST results.

The ultimate goal of these activities is to show how the results of CUBIST create a competitive advantage for the participating partners and European businesses.

1.1 Exploitable Results and Market Overview

The first step of our exploitation activities is the identification of exploitable results. In this first step, CUBIST key results will be described. We identified the following three areas with exploitable results:

- 1) **BI over both structured and unstructured data.** Traditional BI only deals with structured data, e.g. coming from ERP systems. This data has a known format and a known position within a data source, like data in a database. It is anyhow estimated that at least 80% of enterprise relevant information comes in form of unstructured data¹, and the importance of incorporating unstructured data in BI is widely accepted². To name two examples:

¹ <http://clarabridge.com/default.aspx?tabid=137&ModuleID=635&ArticleID=551> provides a discussion on where this often-cited number comes from.

² See e.g. Philip Russom: BI Search and Text Analytics, TDWI Best Practices Report, Q2 2007, downloadable at <http://www.teradata.com/assets/0/206/308/96d9065a-0240-44f1-b93c-17e08ae6eacc.pdf>



- Customer feedback about products is obviously very valuable information for enterprises. The feedback is usually given in a purely textual form, e.g. in mails or in web-channels like blogs and forums. *Sentiment Analysis* targets at extracting attitudes or judgements of people with respect to some topic of interest (like a product), and then customer feedback can become a source of information in BI systems.
- Meeting legal compliance requirements, like the Sarbanes-Oxley Act (SOX)³, becomes more prevalent in today's business world. They force enterprises to archive business-relevant information, including for example emails or spreadsheets with process information, and make this information accessible and analyzable. In fact, SOX and similar legal requirements are often named⁴ as a reason for incorporating unstructured information into BI applications and is identified within SAP as a main use case for CUBIST as well.

Many BI vendors have already incorporated text analytics in their products in order to enable BI over unstructured sources as well. For example, SAP BusinessObjects offers Inxight, SAS offers the Enterprise Miner data-mining workbench and the Teragram linguistic-analysis tools, and IBM has incorporated text analytics in SPSS (the IBM SPSS Modeler integrates with IBM's Cognos Business Intelligence software). The results of CUBIST in this area can indirectly improve existing solutions.

- 2) **Semantically enabled BI:** Traditional BI systems usually store their data in data warehouses. Data warehouses are designed for storing structured data; they are not well-suited for storing unstructured data as well. Moreover, Most business intelligence data warehouses are based on a dimensional model, where a basic fact table of data is surrounded and linked with other tables holding the dimensions of the fact table (this is called *star-schema* or, in a more general form, *snowflake schema*). In CUBIST, the situation is different:
- As a persistence layer, a triple store instead of a data warehouse is used.
 - As a consequence, an ontology serves as schema.

To date, there exist no BI-solutions which build on top of a semantic repository.

3) **Advanced visual analytics and qualitative data analysis:**

Traditional BI means are usually designed to work with numerical data, thus they provide a quantitative analysis of the data (aka "number crunching") based on mathematical statistics. To some extent, though arguably oversimplified, one can understand BI as acting on lists or tables filled with numbers. In contrast to that, an integral part of CUBIST are advanced visual analytics based on "Formal Concept Analysis" (FCA), which can be understood as a theory which allows

³ <http://www.soxlaw.com/>

⁴ Examples are: PriceWaterhouseCoopers: IT implication of Sarbanes-Oxley: challenge or opportunity, Information Builders: Achieving Compliance – Meeting Today's Reporting Requirements with Business Intelligence Technology (Whitepaper), and Intel: Building an Enterprise Data Warehouse and Business Intelligence Solution (Whitepaper)



for a meaningful clustering of entities along attributes acting on the objects, hierarchically ordering those clusters, and finally visualizing these cluster hierarchy. The visualizations are graph-based diagrams with nodes representing clusters and lines representing hierarchies.

The distinguishing features of CUBIST are that it allows both meaningful, qualitative data analysis and graph based visualization of that analysis. To date, no BI vendor offers Visual Analytics products which are capable of doing similar analytics and visualizations.

The exploitable results are summarized in the diagram below.

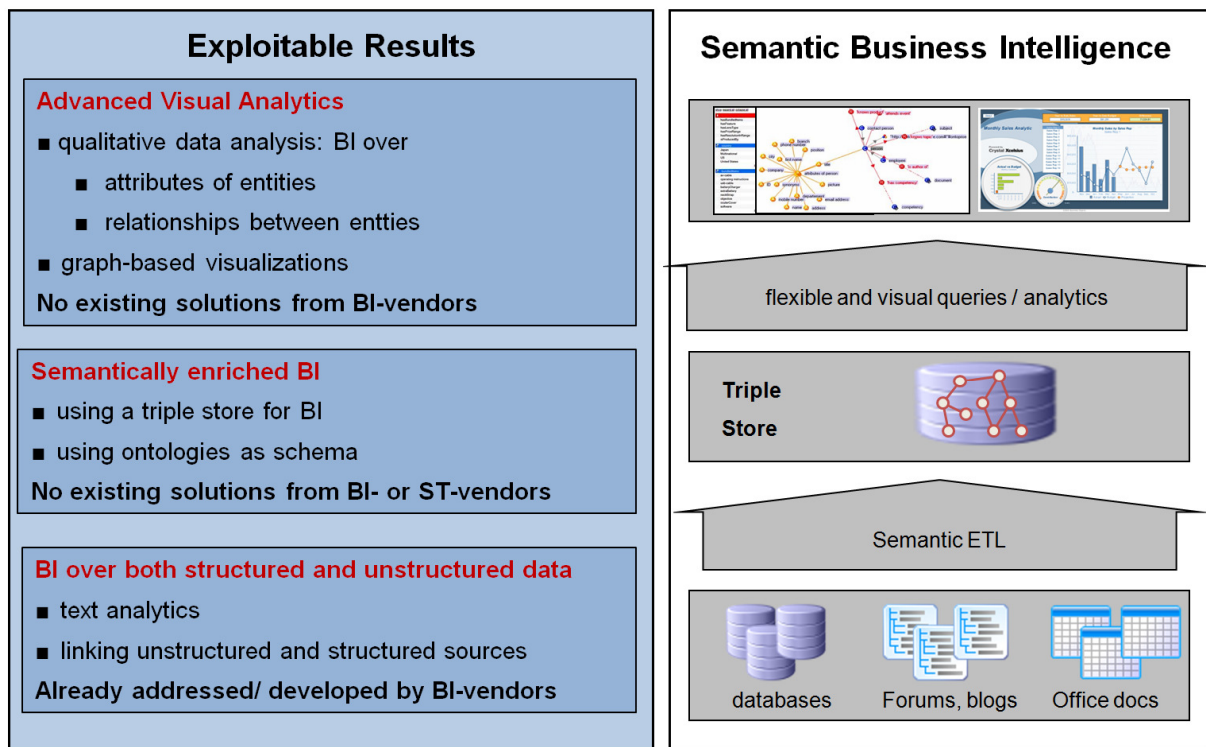


Fig 1: Three core fields for exploitation in CUBIST

1.2 Exploitation channels and beneficiaries

Exploitation channels and beneficiaries heavily depend on the respective partners. For this reason, there is no dedicated section for them in this chapter. Instead, they will be targeted per partner in the next chapter, where the individual exploitation activities per partner are introduced.



2 Individual Exploitation Plans

The following section is dedicated to the intentions of the consortium for the exploitation of the conceptual and technological results from CUBIST. It includes a preliminary exploitation plan for each partner, which makes the business or operating models of each partner explicit and describes their opportunities to exploit CUBIST results. In this way, all partners can look for synergies and opportunities that cross companies and institutes.

2.1 SAP AG

2.1.1 General Exploitation Strategy of SAP Research

SAP is the world's leading provider of business software (SAP defines business software as comprising enterprise resource planning and related applications) and according to a recent Gartner Report⁵, SAP is the dominant leader in the combined BI market (platforms, CPM suites, and analytic applications and performance management), owning nearly a quarter of the market. SAP offers applications and services that enable companies of all sizes and in more than 25 industries to become best-run businesses. SAP has more than 47,800 customers (excludes customers from the acquisition of Business Objects) in over 120 countries.

SAP Research is the global technology research and innovation unit of SAP, with a network of 19 research locations worldwide. By exploring emerging IT trends, SAP Research significantly drives innovation for SAP and its ecosystem. Activities span from collaborative research with academic partners to co-innovation with industry partners and customers. The best validated results and technologies are further developed into prototypes and potential business opportunities within SAP.

Using the fundamentals of applied research, SAP Research explores promising ideas and turns them into prototypes, with the ultimate goal of creating product enhancements and new solutions that benefit our stakeholders and customers. One of the core projects in 2011 and beyond that drives the vision of SAP Research is the *Business Web*⁶ – a cloud infrastructure that is optimized for mobile devices and allows businesses and their employees to collaborate easily, to access data from systems and smart items, and to carry out business processes anywhere and at any time. SAP Research will contribute innovations from our various research topics to help make this vision a reality.

⁵ 2011 Gartner Report: Market Share Analysis: Business Intelligence, Analytics and Performance Management, Worldwide, 2010

⁶ <http://www.sap.com/corporate-en/our-company/innovation/research/businessweb/index.epx>



The business model of SAP Research is based on co-innovation through collaborative research. SAP Research currently has a global community of more than 800 partners from industry, academia, and governments, as well as SAP customers. In addition to contributing to external projects (bi- and multi-lateral, publicly funded, and grants), the researchers collaborate with SAP's development and field organization on internal projects transferring and implementing the research results.

In summary, SAP Research significantly contributes to SAP's product portfolio and extends SAP's leading position by identifying and shaping emerging IT trends through applied research and corporate venturing.

2.1.2 Exploitation Targets

As BI is a main facet of CUBIST and as SAP is a leader in the BI market, it comes as no surprise that that SAP has decided to focus on two major BI aspects of CUBIST as exploitable targets. These aspects are:

- BI over data in many information silos, where relevant information is stored as unstructured data
- advanced visual analytics, with a focus on the combination of “classical” BI charts for quantitative data analysis and FCA-based information visualization for qualitative data analysis

For both aspects, key stakeholders within SAP have been contacted, who expressed their vital interest in CUBIST and agreed to support the CUBIST exploitation activities.

For the time of writing this document, we are focusing on the advanced visual analytics mentioned in the previous section and interacting with two stakeholders of SAP BusinessObjects, one of them being responsible for the visualizations of SAP BusinessObjects products. In what follows we give a brief description of SAP BusinessObjects and relevant products.

SAP BusinessObjects is a division of SAP which offers a broad portfolio of BI tools and applications designed to help you optimize business performance by connecting people, information, and businesses across business networks. Below, we list those three tools which are mostly related to the CUBIST aims:

- *SAP BusinessObjects Explorer* is a data discovery application that allows you to retrieve answers to your business questions from corporate data quickly and directly. Through searches, one can find relevant data that is held within consistent, meaningful datasets known as Information Spaces. To find your answers, one can filter and drill through Information Spaces and view only the data you are interested in via advanced visualizations or charts. One can then able to perform visual analysis to attain the information you are looking for quickly in the most appropriate format.



- *SAP BusinessObjects Explorer software, accelerated version*, combines intuitive information search and exploration functionality with the high performance and scalability of in-memory analytics, so one can empower its organisation and extend the reach of BI to all of the business users. With immediate insights into vast amounts of data from anywhere in the organisation, one can explore business at the speed of thought – and improve your ability to make sound, timely decisions.
- *SAP BusinessObjects Web Intelligence* is an ad hoc query, reporting, and analysis tool that is used to create queries or use existing reports, format retrieved information and perform analysis to understand trends and root causes. SAP BusinessObjects Web Intelligence offers business users self-service access, ad hoc reporting, and intuitive analysis across heterogeneous sources, so you can turn data into actionable insights.

2.1.3 Initial Exploitation Activities and Results

Two initial exploitation activities within SAP have already been started, both addressing the advanced visual analytics mentioned in the previous section.

As already stated, the Business Web is a core project of SAP research which spans over several teams. Both stakeholders for advanced visual analytics that have been mentioned in the last section are driving the Business Web, e.g. the stakeholder responsible for visual analytics in SAP BusinessObjects is product owner of the visual analytics in the Business Web as well. Two project members of CUBIST are participating in the corresponding SAP internal group, which will develop new approaches for the visualizations in the Business Web. In a mid-term perspective, it is expected that insights of the visual analytics related research in CUBIST will feed into the visualizations of the Business Web.

Secondly, as a side activity to the core CUBIST activities, a diploma thesis, which starts in Oct. 2011 will explore and investigate specific aspects of the combination of classical and FCA-based visualizations, and a corresponding prototype will be developed. This prototype will be owned by SAP only. It will be used as an SAP-internal demonstrator for advanced visual analytics and will be presented to and exploited by the visual analytics stakeholders.

2.2 Ontotext

2.2.1 General Exploitation Strategy & Targets

Ontotext has a product portfolio of solutions for semantic data integration, scalable semantic data warehouses, and text and web mining platforms. The CUBIST project is very well aligned with the strategic areas of interest of the company and many of the outcomes of the projects (in particular within WP2 and WP3) will be directly transferred into the existing products of the company.

In particular, the potential exploitation opportunities for Ontotext emerging from CUBIST can be summarised as:



1. Application of current technologies and solutions from Ontotext to a new set of use cases in various domains.
2. Analysis of current limitations of our technologies & solutions in the context of CUBIST use cases and gathering of requirements for extensions & improvements.
3. Transfer of CUBIST results into the company product portfolio, in particular:
 - a. WP3 provides important results related to Ontotext's RDF database (OWLIM).
 - b. WP2 provides results of interest for Ontotext's text mining platform (KIM) and its expertise in the area of RDF-izing legacy enterprise data sources.
 - c. WP9 will improve Ontotext's JOCI platform (a joint collaboration between Ontotext and Innovantage, providing BI for the UK recruitment market).
4. Opportunities for collaboration with CUBIST use case partners beyond the lifetime of the CUBIST project itself, if the respective organisation needs expertise and solutions improving its current business processes & information systems with Semantic Technologies.

2.2.2 Initial Exploitation Activities and Results

Since the main exploitation opportunities for Ontotext come from technology transfer from CUBIST into existing Ontotext products & solutions, the main opportunities will emerge after M30 when the infrastructure workpackages of interest (mainly WP2 and WP3) will be completed.

Nonetheless, since the work in WP2 and WP3 has already started and the initial deliverables have been finalised, there have been several exploitation opportunities and activities within the M1-12 period of the CUBIST project with respect to the four main directions outlined in the previous section:

1. Initial analysis of applying semantic technologies (namely RDF databases, semantic ETL approaches, text mining) in the context of the three use cases has been performed. Such "proof of concept" applications of Ontotext technologies provide very important feedback about the advantages and more importantly the limitations of our technologies and products (see next section).
2. New requirements for extending Ontotext products and solutions. Work in the use case WPs during the first year of the project already provided some interesting requirements that were not foreseen during the initial planning for the project. For example the requirements for extending the OWLIM database with 3D spatial querying capabilities provides an interesting direction for research & development (currently OWLIM supports only geo-spatial queries). Additional requirements for analytical pre-processing of the data will most probably result in extending the built-in



graph processing capabilities of OWLIM (the “RDF Priming” and “Spreading Activation” features described in CUBIST D3.1.2).

3. Work in WP2 and WP3 is still in its initial stages but nonetheless some results have already been transferred into our products or applied in current industrial projects:
 - a. Contributions from Ontotext to the open source Sesame 2.6.0 release in the area of SPARQL 1.1 support and federated queries.
 - b. Expertise gained with Google Refine (one of the open source tools that will be used for the Semantic ETL within CUBIST) has been successfully applied in a project for a big US media publisher, where the major challenges was data integration of legacy data sources with external Linked Open Data sources.
4. Ontotext has been discussing opportunities for collaboration with HWU which go beyond the scope of the CUBIST project, since the two organisations are active in the area of Life Sciences (using ontologies & Linked Data for integration of bio-medical databases) and there is a good fit between the technological needs of HWU and the technological capabilities of Ontotext.

2.3 Sheffield Hallam University

2.3.1 General Exploitation Strategy of Sheffield Hallam University

Sheffield Hallam University (SHU) is actively committed to fundamental and applied research, and its dissemination and wider exploitation. SHU has 16 Research Centres and Institutes, two Yorkshire Forward Centres of Industrial Collaboration, and several specialised research groups, many of national and international renown featuring key research platforms such as materials science, art and design, sports science and engineering, biomedicine, and economic and social research. This reputation is reflected in the fact that we are one of the leading new universities in the UK, building on our previous successes in the 2001 Research Assessment Exercise. SHU drives and promotes cross-disciplinary research, such as encouraging traditional science disciplines to bring together artists, designers, healthcare professionals and industrial partners into their research activities. And through our portfolio of European and international research projects, we are partnering with academic research institutes, industry, SMEs, public, voluntary and community sectors, and not least the end users.

SHU’s Cultural, Communication and Computing Research Institute (C3RI)’s Computing and Communication Research Centre (CCRC) is a key factor in this overall commitment. The CCRC’s Conceptual Structures Research Group will exploit the outputs of CUBIST by producing learned scientific publications, combining expertise with its other funded projects, marketing newly acquired expertise to industry on a consultative basis, promoting SHU



courses, expanding overseas computing and information systems markets, and by increasing the research profile of SHU through participation in further related projects.

2.3.2 Exploitation Targets

At the outset, SHU had exploited CUBIST through publicising the project in a variety of industry sources. Having established these channels and others, SHU will continue to exploit CUBIST in this manner. These exploitations will be underpinned by publication of CUBIST in learned, peer reviewed journals and conferences. CUBIST will also be fed into SHU's undergraduate and graduate programmes. SHU will have an advantage over its competitors in the student market by publicising the fact that leading edge European research is informing and featuring in the curricula of its computing courses. It will also have an added outcome in that these students will take their knowledge of CUBIST into their careers, thereby exploiting the project as they apply CUBIST results to the benefit of their employers. This channel will also extend to our PhD students, who will continue to build upon CUBIST in their research in turn and its consequent exploitation. The experiences and benefits of CUBIST will also be integrated into future projects, also capitalising on its relationship with SHU's other funded projects, particularly the Odyssey project that was also funded under FP7. Where commercial or contract research opportunities arise with industry, and as outlined in SHU's general exploitation strategy, these will also be exploited. The impact of CUBIST will also be included in the University's forthcoming UK Higher Education REF (Research Excellence Framework) submission.

2.3.3 Initial Exploitation Activities and Results

The industry sources where SHU has publicised CUBIST to date include "CUBIST project aims at better Semantic Web search" (kntheiet.org), "Semantic Web Meets BI In New Project Whose Partners Include SAP, Sheffield Hallam University, Ontotext" (semanticweb.com), "Researchers win funds for semantic business intelligence" (zdnet.co.uk), "Sheffield scientists lead £4m semantic web search project (computerweekly.com), "Connecting To The Semantic Web" (businesscomputingworld.co.uk), "UK researchers tap semantic web for BI innovation" (c3.co.uk), and "€4m web project looks set to open up access to hidden knowledge" (Headlines 17, SHU's Services for Business).

The first CUBIST workshop at 19th Annual Conference on Conceptual Structures (ICCS), Derby, UK in July 2011 was an early exploitation vehicle and included the workshop's publicly available peer-reviewed proceedings. The workshop was initiated by SHU, and organised with SAP Research. A second CUBIST workshop is planned for 2012 to be held in conjunction with ICFCFA 2012 (6 - 10 May 2012, Leuven, Belgium). A call for papers for this is currently being disseminated.

There will be a paper submitted by SHU on a new FCA algorithm to the International Journal on Information Sciences (an exploitation activity related to D4.2.1: High performance formal



concept miner, v1, M18), and another paper from SHU on Discovering Knowledge in Data using Formal Concept Analysis in the Int. Journal of Knowledge Discovery and Data Mining (an exploitation activity related to D4.3.1: Large scale/novel FCA visualization tools, v.1, M18).

CUBIST has been included in the curricula of SHU's Computing courses, in the following modules at undergraduate or graduate level: Smart Applications, Multiprocessing and Parallel Systems, Enterprise Systems. With the planned Corporate Masters with SAP, CUBIST will be rolled into the curricula of Sheffield Business School. A feature of that course is that its students will be consultants from industry, who will be able to exploit their acquired knowledge of CUBIST.

SHU is exploiting CUBIST results to attract and recruit new PhD students. Currently three PhD students have passed their conformation of proposal stage, and one has passed the research confirmation stage, the next stage of which is the submission of the thesis. The titles of three PhDs are "Appropriating Data from Structured Sources for Formal Concept Analysis", "Dealing with inconsistent and incomplete data in a semantic technology setting" and "Generating Formal Concepts for Large Contexts". The title of one student, who is also an experienced SAP Consultant, is "Discovering the Hidden Semantics in Enterprise Data through Formal Concept Analysis". This student anticipates the direct exploitation of CUBIST in Enterprise Systems, notably given that 65-70% of the world's transactions run through a SAP system (as stated by SAP's co-CEO Bill McDermott, in the influential Forbes business magazine).

CUBIST results (and those of Odyssey) at SHU are being exploited to inform and feed into other research funding bids, namely:

MINERVA: The proposal for this project has been submitted under a Department Homeland Security (U.S.A) Human Factors/Behavioural Sciences Division funding opportunity. The project aims to apply state of the art models of online radicalisation behaviour to the development of software tools to support the law enforcement agencies in the identification, prevention and contesting of terrorist communication strategies and ideologies.

ATHENA/e-POOLICE: This project proposal is currently being written under EU FP7 Topic SEC-2012.6.3-1: Developing an efficient and effective environmental scanning system as part of the early warning system for the detection of emerging organised crime threats and is a Capability Project.

Xanalys Delve: A platform for the collection, processing, analysis, and management of Public and Social Network data. This is a project proposal submitted under the UK HM Government's INSTINCT/SBRI programme of funding.



2.4 Centrale Recherche S.A

2.4.1 General Exploitation Strategy of Centrale Recherche S.A.

École Centrale Paris (ECP) was founded in 1829 as the first major engineering school to train engineers in the early days of industry in France. Today, ECP trains managers for industry, a role which gives it a unique position among the major French engineering schools. To promote its research activities which are heavily geared towards industry, ECP created a Private Limited Company in 1986, Centrale Recherche S.A (CRSA). CRSA operates as a commercial interface with industry, promoting the expertise of ECP research laboratories to potential customers interested in contracting-out research work.

The Applied Mathematics and Systems (MAS) Laboratory of École Centrale Paris (France) is focused in mathematical and computing tools and methods for the analysis, design and exploitation of complex systems. It is structured around projects including Scientific Engineering and Visualization, Information Processing and Information Systems. The MAS laboratory now employs 100 people (teachers, researchers, PhD students and engineers) and is linked to two teaching units within the École Centrale Paris (the Applied Mathematics Department and the Computer Science and Telecommunications Department), in particular for projects dedicated to initiation of research. The MAS laboratory has also developed partnerships for several Masters Degrees with various universities.

The MAS laboratory focuses on the methods, mathematical algorithms, and information systems related to the analysis, conception and development of complex systems, especially through techniques of modelling, simulation and optimization. It seeks to develop original techniques and tools, both in their content and their application. The Business Intelligence team was recently created as part of an academic chair (*SAP Business Objects*) and aims to research and cover diverse aspects of the BI processes. The BI domain, and thus the BI research team, spans many fields such as high-performance computing, visualization and user experience, and data processing and management. CRSA's interaction and visualization techniques are designed and centred around user experience and are evaluated for effectiveness and appropriateness by users.

In CUBIST, CRSA is involved in designing the user interface and developing new semantic visualizations.

2.4.2 Exploitation Targets

The main motivation for CRSA/MAS laboratory is to produce relevant knowledge in data mining, semantics and visualization. The main axes of research of the Business Intelligence team are:



- Formal Concept Analysis – FCA provides an intuitive understanding of generalization and specialization relationships among objects and their attributes in a structure known as a *concept lattice*. This structure can provide reasoning for classification and clustering, implication discovery and rule learning. CRSA is leveraging FCA for Business Intelligence by highlighting interesting patterns of data. In particular, the analysis of multi-valued contexts in the case of industrial conception and measuring e-reputation in online social networks.
- Adding a semantic dimension to both structured and unstructured data.
- Graph manipulation and analysis – CRSA investigates new methods to extract, manipulate and query graph data. This includes the extraction of graphs from relational databases, semantic expansion from unstructured data sources and visual querying based on semantic and topological features of the graph.
- Visual Analytics – Visual analytics is a key field the understanding and analysis of large amount of information. Apart from CUBIST, CRSA is currently supporting another research with interactive dashboards for BI (in partnership with SAP Business Objects).

Although CRSA is implied mostly in visual analytics, the CUBIST project intersects three of these axes almost entirely, fostering synergy among different competencies in similar contexts.

2.4.3 Initial Exploitation Activities and Results

Preliminary results from CUBIST are encouraging for further development on visual analytics and FCA. As a first step we conducted interviews with three CUBIST use cases attempting to find patterns within their data: users conducting market intelligence, computational biology, and space control centre operation monitoring. All users had in common large amounts of data within which they wanted to answer 3 types of questions: frequent pattern detection, anomaly detection and pattern comparison. An example of frequent pattern detection is “*during the first stage of a mouse embryo development what are the genes expressed together most often*“. An example of anomaly detection is “*what are the sensor and telemetry logs of a space load on the International Space Station that may be related to a specific instrument malfunction*“. And a pattern comparison question would be “*Are the jobs available in Liverpool similar to those in Manchester*”?

Based on the nature of each use case analysis, we have implemented several alternatives for the visual exploration of concept lattices through searching, filtering sub-selection of concepts and attributes; visual display of related attributes and their implication confidences. A web-based visual analytics prototype was developed to implement all the functions mentioned, called *CUBIX*. Typical uses of *CUBIX* include semantic data analysis and pattern detection, anomaly detection, comparisons, information classification, and knowledge



discovery. *CUBIX*'s workflow allows users to carry out an analysis starting from a real data set, converting it into a formal context, simplifying the context to make it manageable, and visualizing the result as a concept lattice. The tool is currently being developed and continuously refined with the active involvement of our three users groups and their use cases.

To illustrate the features of our first version of *CUBIX* to our use case partners, we have used a sample of a traffic accident dataset occurred in the UK in 2006 and conducted an analysis for most common causes between different combinations of attributes (e.g. road surface, weather, accident severity, etc). FCA allowed us to answer pattern identification questions such as “*What can be considered the main causes of ‘accident severity: serious’?*” or “*how many accidents are ‘light conditions: darkness’ and ‘accident severity: serious’?*”.

Other subsequent studies and applications from CUBIST are:

- Use of CUBIX to analyse content interaction in social networks;
- Visualisation of multi-valued contexts in a case of air plane cabin conception (in partnership with CSCL);
- Dynamic FCA from live data streaming.

2.5 Space Application Services

2.5.1 General Exploitation Strategy of Space Applications Services

Space Applications Services (aka SAS) is a leading provider of system and operations engineering as well as software engineering expertise in the field of **space** and **aerospace**. SAS also applies these capabilities to other industrial application domains such as the **security** market.

- As a **software engineering** company, SAS has long experience in a wide range of practices and techniques, from the collection and analysis of requirements from user groups, to the implementation of software to support operations, including Knowledge Management software and other precursor infrastructures for the next generation of software used in space mission control centres (known as “User Support and Operations Centres”, USOC) in Europe.
- As an **operations engineering** company, SAS has practical knowledge of and expertise in dealing with large quantities of mission data, of structured and unstructured types, accumulated over a long period of time. SAS has Ground Controllers involved in system administration of operation centres.

The company can take responsibility for the complete project life cycle from establishing customer needs, designing and selecting the most appropriate technology, assembling the best



team to implement the project solution and, in some cases, operating the system for the customer.

SAS' expertise includes:

- Space system engineering, specification, operations engineering, integration, training and software development from the earliest phases of spacecraft and mission concept definition to on-orbit operations.
- Software Engineering: design and development of monitoring and control systems, decision support systems, distributed control for fixed and mobile robots in structured and unstructured environments.
- Facilities Operations: operation of payloads and complete missions at operations facilities.
- Research and Development: establishment of methods and processes for collaborative multi modal human-computer interaction; development of knowledge management systems.

Through its participation to European and National **research** projects (such as this FP7 CUBIST project), SAS investigates, pioneers and matures new ideas, concepts, technologies, algorithms, services and solutions (e.g. via the set up of prototypes, such as here the CUBIST prototype), which can then be folded into the company's **industrial** activities in order to:

- Enhance our service and solution offerings to our existing customer base.
- Increase our business capabilities within our prime business areas.
- Expand our business to new market areas.

2.5.2 Exploitation Targets

SAS' prime customers are the European Space Agency (ESA), national space agencies such as CNES and DLR, as well as prime aerospace contractors such as Thales Alenia Space and Astrium. All of these are either developing and/or operating ground-based control centres such as mission control centres, mission operation centres, user operations support centres, satellite operations centres, etc. These solutions are required for a multitude of mission types, e.g. earth observation missions, scientific missions, manned missions (Columbus/ISS), navigation and surveillance systems (EGNOS, Galileo), etc. All of these systems require the processing of a huge number of extremely large and heterogeneous datasets containing both structured and unstructured data.

Among all these possible centres, SAS' first exploitation targets for application of CUBIST's Semantic Business Intelligence technology are the space control centres of the European Space Agency known as User Support Operations Centres. The USOC especially chosen first is the Belgian User Support and Operations Centre (B.USOC), that is operated a.o. by



operators of Space Applications Services, who will first provide their user requirements and will afterwards evaluate the successive versions of the CUBIST's Semantic Business Intelligence platform.

Accordingly, SAS intends to exploit and adapt the CUBIST system prototype first for operators at B.USOC, and use this pioneering work to afterwards respond to the needs of other USOCs (the Dutch USOC, where SAS has operators, is a good candidate) and finally the needs of other European aerospace stakeholders. For each of these exploitation targets, the adaptation of the CUBIST system prototype will require some additional development to bridge the requirements that were not implemented in the CUBIST prototype and to respond to the customization needs for each specific Space Control Centre environment.

An initial exploitation objective in mind is to enhance the situational awareness of the operators, by means that provide support to decisions during space missions.

Another exploitation objective enabled by CUBIST research is the deployment of solutions that will increase the cost-efficient utilisation of the vast quantity of unstructured data (mainly technical documents) that are predominantly available within the development and operations life cycle of these ground-based systems. The space industry is indeed extremely document-centric. Lots of past formal documents have to be used as input for software development first and for running the operations afterwards (e.g. documents about applicable standards, or manuals of operation software).

Finally, afterwards, Space Applications Services intends through further self-developed enhancements and deployment of CUBIST methods and solutions on a mix of structured and unstructured data, to further address the following additional exploitation targets:

- Within the aerospace domain:
 - Improved usability and efficiency e.g. via the identification and solving of problems related to consistency, completeness, determinism.
 - By integrating novel concepts such as NLP (Natural language Processing) and NL-QAS (Natural Language Question Answering Systems) to develop and deploy improved end-user querying solutions.
 - (Automatic) generation of new content/documents from the existing (structured and/or unstructured) data.
- Business areas requiring enhanced decision support facilities that are relying on a mix of structured and unstructured data such as network control systems (utilities: electricity, gas, water), transport networks (combined multimodal air-road-rail-fluvial-sea-fluvial transport facilities), etc.



2.5.3 Initial Exploitation Activities and Results

At this early specifications stage, there is no CUBIST software yet. However, a number of activities performed so far in the project (Task 8.1 “Requirements Analysis” and Task 8.2 “RDF Modelling of Data Sources”) have already enabled early exploitation of results, and/or have open the door to new types of exploitation not previously [easily] achievable: we review them hereafter.

D811 – Requirements Document

- This task has covered the gathering of dedicated end-user requirements associated to the needs (a.o. for situational awareness) of ground controllers when they monitor and control scientific instruments attached to the International Space Station, in a space control centre environment. This exercise has allowed capturing requirements that are nowadays already taken into account and incorporated in the system requirements analysis and proposal for design enhancements of our next generation of space control centres.
- A preferred Use Case has emerged from the operators: CUBIST helping and identifying the early signs of failure of one of the SOLAR payload instruments. In turn, this idea of post-mortem analysis of an instrument part of a platform still in space is new and is further attracting the interest of B.USOC deciders: this provides a positive impact on the image of the company in terms of its capability to bring innovations, and encourages deciders in supporting new SAS’ initiatives such as the idea of building the “USOC of the Future”.

D812– Initial Mock-up

- A first important and very concrete achievement has been done by obtaining access to real data from a space control centre (with all necessary authorizations) and by making it available outside the control centre for further study. This has been done thanks to the support and collaboration of the B.USOC, who has provided the “CUBIST Space Data Pack” (v1.0.0), which is a structured data set derived from the SOLAR telemetry archive, over a period of time where a failure of one of the instruments of the SOLAR platform happened. This data set and its associated documentation have allowed the technological partners of CUBIST to get familiar with the space Use Case. Scripts that were elaborated for creating this researcher-friendly version of the SOLAR telemetry archive are now part of the B.USOC library of scripts, and help them to provide similar data provision services, for any other interested researcher. A second version of the “CUBIST Space Data Pack”, covering the same period of time, but with more refined data (and therefore more elaborated analysis) has been initiated.



- Following the pioneering example of this “CUBIST Space Data Pack”, more and more data sets are now planned to be extracted and released soon, at a faster rate (more easily) with the help of the B.USOC, covering not only structured data but also unstructured data, such as PDF/Word documents found in eRooms, console logs, System Problem Reports and other reports such as Daily Operations Reports. We can say here that the CUBIST project is facilitating the access to otherwise quite inaccessible data.
- The initial mock-up exercise has allowed our operators to get familiar with possible ways of using Formal Concept Analysis to support situational awareness.

D821 – RDF Modelling of Data Sources

- This deliverable provides a technical inventory of structured and unstructured data sources to be ultimately used for exercising the “situational awareness” use case. This inventory report now provides a basis for enabling further exploitation of the same data in other projects related to the “USOC of the Future”.

Proper exploitation of the CUBIST software can and will start only once the [initial version of the] CUBIST prototype will have been developed. This generic prototype will then be adapted for tests and validation against the dedicated Space Control Centres use case setup by SAS.

2.6 Heriot-Watt University

Heriot-Watt University (HWU) is a technological university that began initially as the Edinburgh School of Arts, which was founded in 1821 and was granted its Royal Charter as a University in 1966. Computer Science has been taught at Heriot-Watt since the 1960s. The Computer Science department’s BISEL research lab undertakes the work on the CUBIST project using data generated by EMAP.

EMAP is the Edinburgh Mouse Atlas Project located within the Biomedical Systems Analysis Section at the Medical Research Council Human Genetics Unit (HGU). The remit of the HGU is to undertake basic and strategic research to obtain a molecular and cellular understanding of genetic factors implicated in human disease and normal and abnormal development, thereby gaining important insights into basic biological mechanisms.

EMAP has pioneered the atlas-based concept for spatio-temporal databases in biomedicine and in particular the Edinburgh Mouse Atlas (EMA) and associated databases, e.g., EMAGE.

The BISEL lab at HWU has a long-standing close association with EMAP, with Dr Burger being also part of the EMAP team, and participates in knowledge transfer activities, particularly relating to semantic web related technologies and spatio-temporal and distributed biomedical data.



2.6.1 General exploitation strategy

EMAP provides services to the life science community: the Edinburgh Mouse Atlas, and Edinburgh Mouse Atlas of Gene Expression (EMAGE) databases. These services offer a use case that drives the research undertaken by EMAP. Additionally, through the close relationship HWU has built up with EMAP, the use case features prominently in HWU's research too.

Both HWU and EMAP are primarily research focussed academic intuitions, as such, both endeavour to publish rather than generate commercial activities.

Where applicable, research outcomes will be used to enhance the services offered by EMAP: either providing new functionality or upgrading existing capabilities.

Whilst the EMAP use case provides a focal point for exploration, any new insights into the potential benefits of the technologies and tools developed as part of CUBIST will be communicated and promoted to the wider biomedical research community.

2.6.2 Exploitation targets

HWU are conducting two lines of research within the CUBIST project, thus there are two possible outcomes that may be exploited: 1) semantic representation of spatio-temporal biomedical data and 2) the application of FCA tools for the exploration of biomedical data.

A semantic representation of the spatial information held within EMAGE potentially allows for new ways of EMAGE's and other biomedical spatial data to be mapped to other resources with a spatial model of the mouse, for example, the Allen Brain Atlas (www.brain-map.org). We are specifically considering the following communities for our explorations in this area. The W3C Semantic Web Health Care and Life Sciences Interest Group (<http://www.w3.org/2001/sw/hcls>) and bio2rdf communities. The latter in particular have already been looking at the semantic representation of microarray gene expression data (non-spatial). Complementing that with a representation of our in-situ gene expression data (spatial) would open interesting integration and analysis possibilities. We will consider loading both, microarray and in-situ, data into a CUBIST warehouse. We are also looking at possible contributions to the International Neuroinformatics Coordination Facility (www.incf.org), which has a great interest in brain-related atlas data integration and analysis. (Dr Burger is a member of a number of INCF task forces engaging in the development of a framework to share gene expression data between resources that store spatially orientated data for mouse brain.) Finally, because of the involvement of HWU and MRC-HGU in the EU's RICORDO project (www.ricordo.eu), the final potential target for our explorations is the VPH (Virtual Physiological Human) community and its image data related resources. The VPH link is particularly relevant for the possible outreach to the clinical images informatics community.



A proven FCA tool that provides an efficient mechanism for browsing or manipulating the underlying data would be extremely useful for EMAGE. That resource already provides a number of mechanisms for working with the raw data, and is always interested in more powerful techniques. A web-based version of the tool generated by CUBIST, or perhaps an evolution of the tool generated, could be placed upon the EMAGE homepage, thus providing another route into the EMAGE data. Again, any positive experiences on the specific EMAGE use case will be of interest to the wider communities described above.

2.6.3 Initial Exploitation Activities and Results

HWU and SHU are investigating the value of FCA for the life science community. Whilst this activity is primarily research focused, discussions with the EMAGE editorial team ensure that they are able to monitor progress. Initial results of this collaboration were published at the 1st CUBIST workshop, and were positively received by the EMAGE team. However, further work focusing on the validation and verification of the results obtained through FCA is required.

HWU have started an exploration of the semantic representation of EMAGE's spatial data. This work is in an embryonic stage; as such there are no published papers or other results currently visible.

2.7 Innovantage

2.7.1 General Exploitation Strategy

Innovantage is the leading provider of labour market intelligence in the UK. Vacancy information is collected from publicly available sources on the Internet including 154 job boards and over half a million UK based corporate sites. In excess of 1.5 million vacancy advertisements are collected each month.

This data underpins Innovantage's main product Insight, a web-based portal that significantly improves the performance of recruitment agencies, reducing the time taken to locate leads, the number of leads and lead quality.

Innovantage has also created a data warehouse of over 36 million vacancy advertisements since 1st January 2009. This unique database allows Innovantage to:

- Publish labour market reports that identify trends and market gaps.
- Generate bespoke reports on behalf of individual clients to analyse market penetration in comparison to competitors.

These reports are, presently, manually created and published periodically, Innovantage is building a web based tool that will allow clients to customise and generate such reports on demand. This is the initial step to create high level business intelligence tools, part of the longer term strategy to create actionable intelligence from web sourced data collected in other



vertical markets. Innovantage believes that CUBIST will make a significant contribution towards these goals and provide a unique selling point.

2.7.2 Exploitation Targets

Innovantage is planning to migrate its existing Insight product to an RDF data structure, supported by a triple store database. It is envisioned that moving to a database using an RDF structure will allow a variety of improvements to the existing product:

- The early stages of CUBIST have already influenced Innovantage's roadmap with its potential applications. It is expected that results on the next year's effort will heavily affect the continued evolution and refinement of the roadmap.
- Exploitation of linked data. As the use of semantic technologies becomes more prevalent on the web, this new generation of Insight will be able to capture and store link data contained within vacancy advertisements.
- Enhanced search capabilities. The migration to RDF structures will allow new search capabilities to be developed within the product. Such as semantically enhanced job title searching, semantically enhanced skills searching and matching.
- Advanced analysis of historic vacancy information. Innovantage is currently developing a new product to provide clients with a suite of tools to analyse and generate reports from its unique database of over 30 million job advertisements. It is hoped that the semantic business intelligence capabilities developed during the CUBIST project will enhance this analytical environment. Formal concept analysis would provide an additional facet to the suite of tools available to the client that should aid the discovery of hidden relationships within the data.



<Confidential>



3 Summary

This document reports on initial exploitation activities in CUBIST. Most exploitation activities and results naturally arise at the late phase of a research project, and the current exploitation plans are in a provisional state. As CUBIST is in its first year, we expect more concrete results in the second and mostly in the third year of the projects. This document is an “Interim” which will be refined subsequently throughout and according to the course of the project. There will be two more, extended versions of this document, namely D5.2.2 “Exploitation plan / report, v.2” and D5.2.3 “Exploitation plan / report, v.3”, which are due in M24 and M36, respectively.