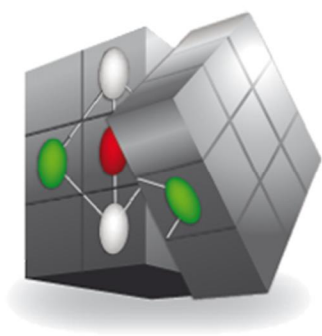Combining and Uniting Business Intelligence with Semantic Technologies

Acronym: CUBIST

Project No: 257403

Small or Medium-scale Focused Research Project
FP7-ICT-2009-5

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

CUBIST Exploitation Report v.2

Abstract: n/a

Type	Report
Document ID:	CUBIST D5.2.2
Workpackage:	WP5
Leading partner:	SAP
Author(s):	Frithjof Dau (SAP)
	Emre Sevinc (SAS)
	Chris Rafferty (Inno)
	Cassio Melo (ECP)
	Ken McLeod (HWU)
	Simon Polovina (SHU)
	Pavel Mihaylov (Onto)
Dissemination level:	PU
Status:	Final
Date:	26 October 2012
Version:	1.0

<Confidential>

Versioning and contribution history

Version	Description	Contributors
0.1	First draft with SAP exploitation	Frithjof Dau (SAP)
0.2	Sections added from SAS and Innovantage	Emre Sevinc (SAS) Chris Rafferty (Inno)
0.3	Section added from ECP and HWU	Cassio Melo (ECP) Ken McLeod (HWU)
0.4	Section added from SHU, finalized SAP section	Frithjof Dau (SAP) Simon Polovina (SHU)
0.5	Section added from Ontotext	Pavel Mihaylov (Onto)
1.0	Incorporated comments by reviewers	Frithjof Dau (SAP)

Reviewers

Name	Affiliation
Constantinos Orphinandes	SHU
Axel Schröder	SAP

<Confidential>

<Confidential>

1 Introduction

CUBIST is an EC-funded research project that investigates and implements the concept that Business Intelligence can be leveraged to provide 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's 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. The first steps towards the exploitation of CUBIST have been described in the CUBIST deliverable D5.2.1 "Exploitation Report v1". As discussed in D5.2.1, the consortium has started with the following steps:

- First, in order to provide a solid basis for the exploitation of the project results, the initial task has been 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 been 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 been detected. This particularly included 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

In this section, we shortly recap the key results of CUBIST as identified in D5.2.1, as they can be exploited in the further conduct of the project. In D5.2.1, 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 estimated that at least 80% of enterprise relevant information comes in form of unstructured data, and the importance of incorporating unstructured data in BI is widely accepted. In D5.2.1, we provided two examples for such needs, namely 1) the need for companies to meet legal compliance requirements, like the Sarbanes-Oxley Act (SOX), where archived business-relevant information like emails or spread sheets with process information have to be analysed, and 2) the need to analyse customer feedback about products, obtained from web-channels like blogs and forums. Note that particularly in the CUBIST Innovantage Use Case, the core data which has to be analysed – namely job vacancies - is obtained from scraping job boards on the Web.

2. **Semantically enabled BI:** In D5.2.1 we have argued that traditional BI systems usually store their data in data warehouses, which are storing structured data and which usually are based on a dimensional model. This contrasts the approach taken in CUBIST, where a triple store is used instead of a data warehouse and an ontology serves as schema.

   To date, no BI-solutions exist which build on top of a semantic repository. With respect to the CUBIST use cases, obviously both above mentioned points are relevant for all of them. Interestingly, the distinguishing information modelling and reasoning features of CUBIST are most prominent in the Innovantage Use Case (to see that, we refer to the different ontologies, as they have been described in the UC deliverables Dx.3.1. "Use Case Prototype" with x=7,8,9).

3. **Advanced visual analytics and qualitative data analysis:** As argued in D5.2.1, 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. 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 for a meaningful clustering of entities along attributes acting on the objects, hierarchically ordering those clusters, and finally visualizing theses cluster hierarchy. The visualizations are graph-based diagrams with nodes representing clusters and lines representing hierarchies.

   The distinguishing features of CUBIST allow both, meaningful, *qualitative* data analysis and graph-based visualization of that analysis. The current CUBIST prototype offers additional visualizations like tree maps, sunburst diagrams, icicles and scatter plots. To date, no BI vendor offers Visual Analytics products which are capable of doing similar analytics and visualizations.

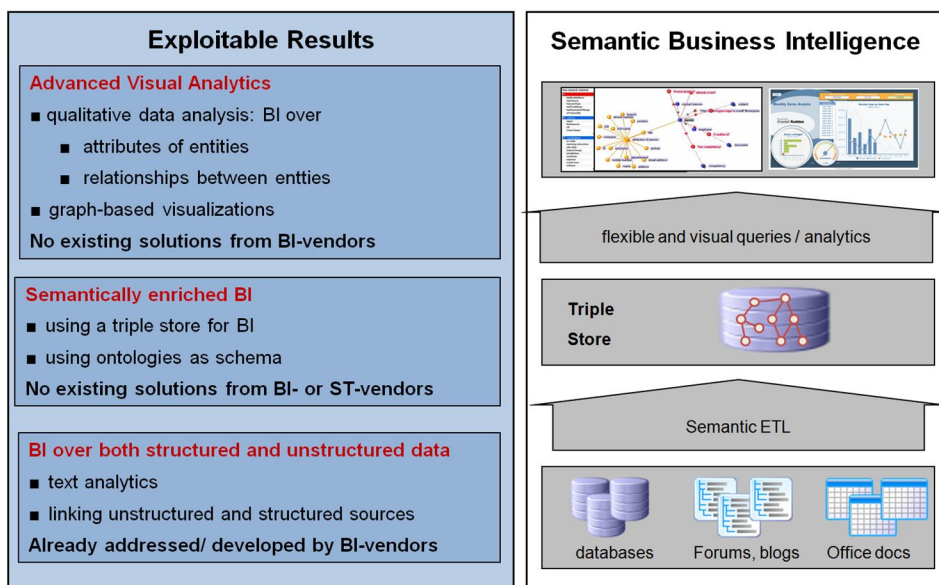In D.5.2.1, the exploitable results have been summarized in a diagram, which is shown in Fig 1.

Fig 1: Three core fields for exploitation in CUBIST

<Confidential>

1.2 Individual Exploitation channels and beneficiaries

Most 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.

1.3 Joint Exploitation

Apart from individual exploitation plans, a wider audience (e.g. the semantic web community or other research institutions apart from HWU working on gene expression experiments) is likely to be interested in CUBIST. The consortium has tentatively decided to publish the final prototype as open source. The appropriate channel and, more importantly, the license model have to be discussed. In Document D5.2.3, we will report about the final decision and the steps which will have been undertaken then.

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. Similarly to D5.2.1, a section for each partner is provided. Per partner, the exploitation strategy and targets are reiterated, emphasizing significant changes w.r.t. D5.2.1. For the initial exploitation activities described in D5.2.1, now their progress and current status are described. Moreover, new exploitation activities are introduced.

2.1 SAP AG

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 2011 Gartner Report[1], 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 61.000 employees (by the end of Q3/2012) and more than 47.800 customers (excludes customers from the acquisition of Business Objects) in over 120 countries.

2.1.1 General Exploitation Strategy of SAP Research

SAP Research is the global technology research and innovation unit of SAP, with a network of more than 20 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.

In the past months, SAP Research and Business Incubation has been reshaped to help accelerate innovation and expand the addressable market for SAP through organic growth. With the goals of thought leadership, business relevance and commercialization, the group is focused on some very exciting topics that will enable SAP to proactively shape the future of business for our customers and drive significant revenue in new markets.

---

[1] 2011 Gartner Report: Market Share Analysis: Business Intelligence, Analytics and Performance Management, Worldwide, 2010

<Confidential>

Based on the probable futures that SAPs customers will be experiencing, the group formed eleven new research areas, including Digital Manufacturing, Trillion Node Network, Social Business Network, and Human Computer Interaction. These areas continuously feed SAP's end-to-end research and commercialization engine with promising ideas. The most promising ones are developed and commercialized by the Global Business Incubator, and then "graduated" into the SAP core.

Recent commercialization successes include: SAP Smart Meter Analytics, a software for utility companies that helps to turn massive smart meter data into powerful insights by providing in-depth customer analysis, advanced segmentation based on energy consumption patterns, and energy efficiency benchmarking; and SAP Supplier InfoNet, a cloud-based solution that uses social media concepts to pool supplier performance data from various sources to benchmark performance, predict risk, and create insights to improve key business processes.

Brand new is SAP Precision Retailing, a cloud-based solution, that enables companies to influence consumer shopping behaviour at the moment of decision by delivering 1-to-1 personalized offers in real-time across multiple channels.

The future innovation pipeline is robust and includes topics that range from 1 to 5 years out. Ideas that are 3-5 years out are worked on by Applied Research, with a focus on technological feasibility and in close cooperation with academic and industrial research partners. Ideas that are 1-3 years out are worked on by Advanced Development, with a focus on taking the most promising ideas, testing their market desirability and turning them into prototypes for future applications.

When it comes to business viability, the Global Business Incubator creates internal start-ups that develop and commercialize new products, 1-2 adjacencies away from the SAP core and viable within an 18 month window. Product ideas that address use cases and pain points that customers have today are developed by Imagineering in close collaboration with customers, and the Co-Innovation Labs (COIL), who drive and facilitate project-based co-innovation with partners.

2.1.2 Exploitation Targets

As already stated in D.5.1, the two major BI aspects of CUBIST which serve as exploitable targets for SAP are:

- BI over data in many information silos, where relevant information is stored as unstructured data.

<Confidential>

- 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 the time of writing D5.2.1, SAP focused on the advanced visual analytics. This focus has not significantly shifted.

2.1.3 Initial Exploitation Activities and Results

In D5.2.1, two initial exploitation activities within SAP (both addressing the advanced visual analytics of CUBIST) have been mentioned. In this section, SAP reports on the progress of these activities.

2.1.3.1 Visualizations for the Business Web

The first activity targeted at incorporating visual analytics features of CUBIST in the SAP Business Web. As described in D5.2.1, two project members of CUBIST had been participating in an SAP internal group, which had developed new approaches for the visualizations in the Business Web.

Today, due to the above mentioned internal restructuring process, the team's focus has shifted towards a similar internal project that now works on visualization in real time scenarios. Intermediate results from the previous transfer partially influenced an upcoming product in the field of on-demand web based BI software. However, the activity around SAP Business Web had to be stopped due to those organizational changes.

2.1.3.2 Diploma Thesis

The second activity was the beginning of a Diploma thesis, conducted as a side activity to CUBIST, aiming at developing a prototype which will be owned by SAP only. For the time of writing this deliverable, the thesis has been finished, and the targeted prototype has been developed.

The thesis focused on the visualization (by means of Hasse-diagrams) and interactive exploration of concept lattices. Layouting Hasse-diagrams in a way that it is *always* ensured that the diagrams become appealing, is still a topic of research. In the thesis, the positioning of nodes for generating additive-line-diagrams has been investigated. Next, a heat map has been developed, which shows the user "good" and "bad" positions when nodes are moved. Finally, it has been scrutinized on how the concept lattices changes when attributes are removed or added. This enabled to develop a user interface where, after changing the set of attributes, the originating concept lattice is *smoothly transitioned* into the new lattice. This is a feature not provided by any FCA-tool so far. As the thesis was a mathematical thesis, all described features have been thoroughly and formally investigated.

The core contribution of the diploma thesis was the provision of a fully functional and documented Java-library for the FCA-visualization, covering both the overall functionalities

<Confidential>

needed by any FCA-tool (e.g. the computation of the concept lattice out of the formal context) as well as the implementation of the new features described above. This library has been incorporated into an SAP-owned version of the prototype, replacing the visualizations in the CUBIST prototype, as they are developed by the CUBIST partner ECP. Screenshots of this prototype are provided in the next figures.

The library has been made available within SAP via an SAP-internal code sharing platform. Moreover, the prototype has been deployed on an SAP-internal server and is thus ready to be tested by interested users. To better promote the library, the following steps will be conducted next:

1) SAP has already developed a new dataset for the prototype using SAP-data, using product-feature-matrices for BusinessObjects Products.
2) Once this data is incorporated into the prototype, a demo-presentation and demo-video will be produced.
3) SAP will finally address the product owner of visualizations in BusinessObjects-Products directly.
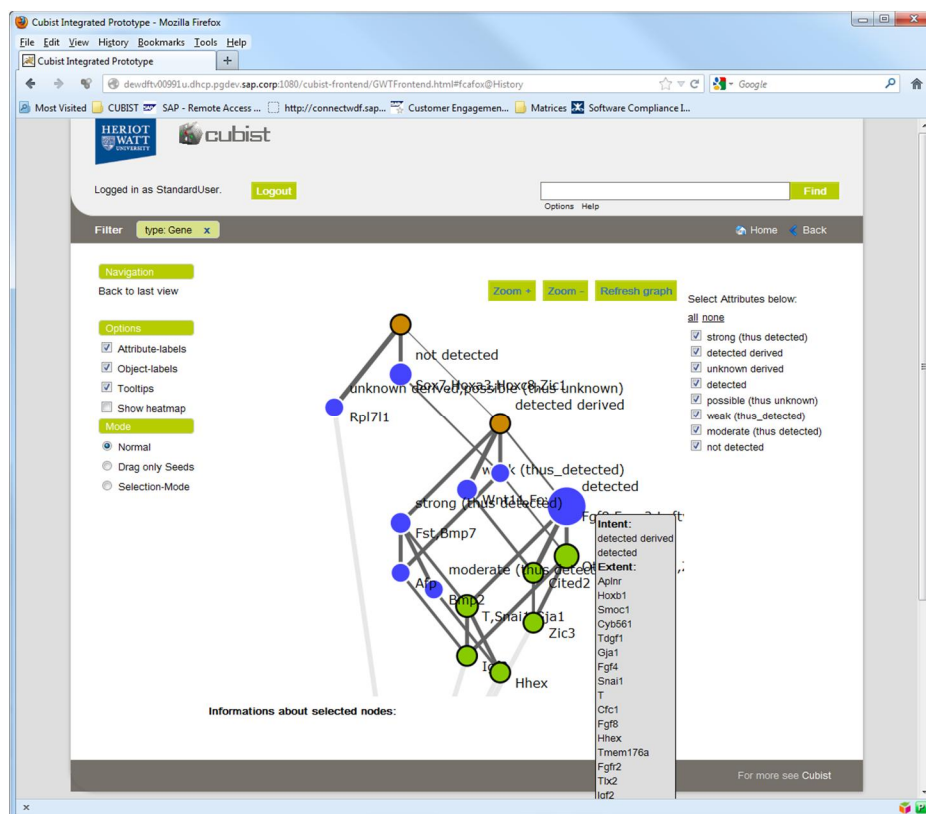

Fig 2: Default layout for genes and levels of expressiveness in Theiler Stage 10. For a selected node, nodes above and below are highlighted. Line thickness visualizes similarity between concepts; node size visualizes the concept size.
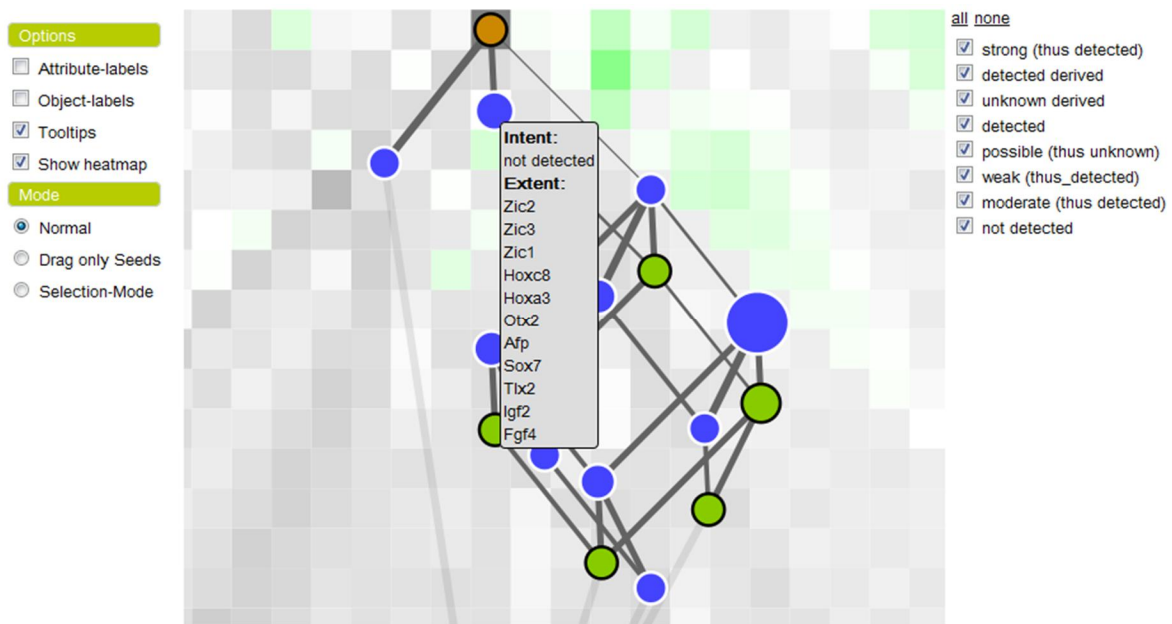
<Confidential>

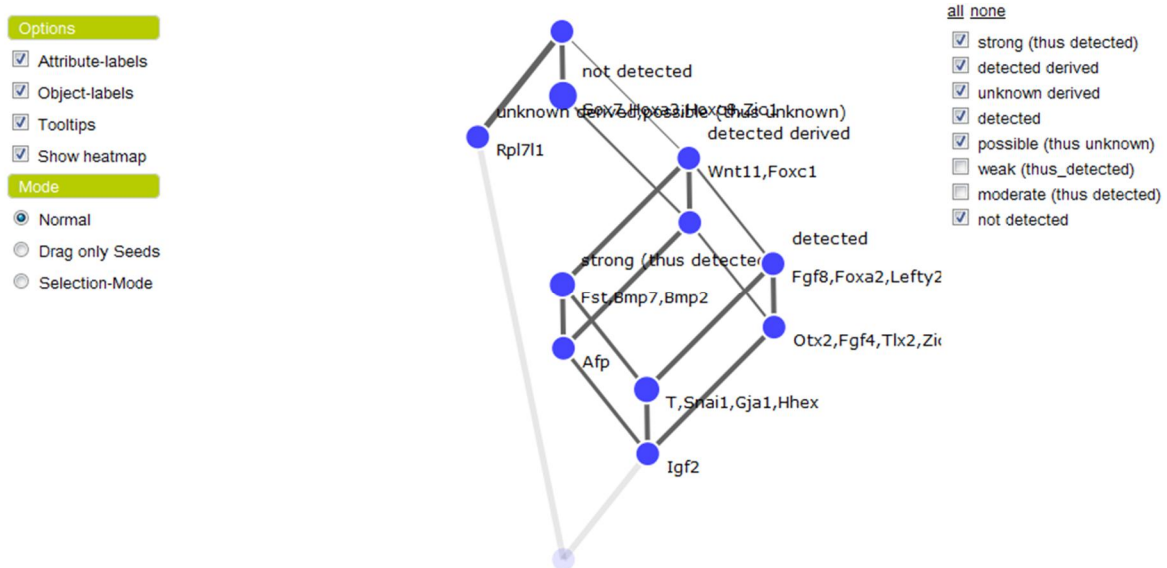**Fig 3: Using a heat map to position nodes. Node labels are deactivated.**

Fig 4: Concept lattice after removal of two attributes.

2.1.4 New Exploitation Activities

In this section, SAP reports about new exploitation activities which have currently (as of time of writing this deliverable) been started.

2.1.4.1 Bilateral activity between SAP and SAS

The CUBIST partner Space Application Services develops mission control system and monitors payloads (equipment) in the space and aerospace market. The main means for

<Confidential>

monitoring is payload data, being a couple of hundred parameters which are sent to the control centers once per second. One of the SOLAR instruments, SOVIM, had a breakdown in October 2008, and SAS aims with CUBIST at analysing SOVIMs Housekeeping data in order to detect patterns in the parameters which indicated the breakdown before it actually happened. Of course, SAS wants to use such detected patterns in the real-time monitoring to prevent such failures in the future. Anyhow, the real-time-monitoring of data is outside the scope of CUBIST.

An SAP-internal thought-leadership-project investigates on the basis so-called Complex Event Processing (CEP), using SYBASE[2]-products, real-time-detection of patterns in streaming data. As this perfectly fits into the mission of SAS, SAP has addressed SAS with the proposal to conduct a bilateral project in order to conduct CEP on the SAS data. This initiative started in August 2012. A first call between SAS and SAP has been conducted on September 6th. In this call, both demos for CEP have been given to SAS, and possible legal frameworks between SAP and SAS have been discussed. The first feedback from SAS is quite promising, and currently, SAS evaluates internally the participation in a bilateral activity within the upper management.

Obviously, the current efforts count as exploitation activity both for SAP and SAS.

2.1.4.2 **CUBIST addressed by external companies**

Due to the CUBIST dissemination activities (see D6.1.6), CUBIST has been meanwhile addressed by external companies which expressed their interest into a collaboration with CUBIST.

The first company is Hatch[3], being an 11.000-employee company situated in Canada. It "supplies engineering, project and construction management services, process and business consulting and operational services to the mining, metallurgical, energy and infrastructure industries". Hatch is SAP customer and starting to implement their system architecture based on RDF. The approach of semantically enabled BI looks very appealing to them, as they are looking for a solution to conduct BI on data from triple store and regular DWs. A first contact has been made by an SAP Architect of Hatch end of August 2012, thus it is for the time being unclear whether and how collaboration between Hatch and CUBIST will be carried out.

Secondly, CUBIST very recently (middle of September 2012) has been addressed by a Belgian business analysis company called Banyas. They like to develop an idea of automated business scenario identification by reading the processes, running in the ERP system, and are very attracted by the visual analytics features of CUBIST. Similarly to Hatch, only a first

---

[2] SYBASE has been acquired by SAP in 2010; http://www.sybase.com/

[3] http://www.hach.ca

contact has been made for the time being, and the further conduct of a prospective collaboration is not clear yet.

2.2 Ontotext

Ontotext has a product portfolio of solutions for semantic data integration, scalable semantic data warehouses, and text and web mining platforms.

2.2.1 General Exploitation Strategy & Targets

The CUBIST project is very well aligned with the strategic areas of interest of the company and many of the outcomes of the project (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 their technologies and solutions in the context of CUBIST use cases and gathering of requirements for extensions and 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-ising 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).

   d) Experience with RDB2RDF tools within WP2.

   e) Experience with various link detection frameworks (Sil and Limes) within WP2.

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 and information systems with Semantic Technologies.

2.2.2 Exploitation Activities and Results

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

<Confidential>

Nonetheless, since the work in WP2 and WP3 started early on, there have been several exploitation opportunities and activities within the M1-M24 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.

2) New requirements for extending Ontotext products and solutions. Work in the use case WPs during the first and the second year of the project already provided some interesting requirements that were not foreseen during the initial planning for the project. For example the requirement for extending the OWLIM database with 3D spatial querying capabilities provides an interesting direction for research and 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 not complete 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 challenge 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 and 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.

5) Various improvements to our OWLIM RDF database in the context of D2.3.1, namely:

   a) performance improvements related to SPARQL 1.1 querying

   b) improvements to the full-text search functionality

   c) more scalable RDF Rank implementation

   d) new index compression scheme

   e) re-designed and improved plug-in architecture

<Confidential>

f) various general performance improvements

6) Practical experience with RDB2RDF tools (WP2), which can be applied directly in our commercial projects.

7) Practical experience and benchmarking of various link identification frameworks (Silk, Limes) in WP2, which are relevant to our product lines for text mining and Linked Data management.

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 it is one of the leading new universities in the UK, building on their 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. Through their portfolio of European and international research projects, they 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

<Confidential>

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 their 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. The second CUBIST workshop in 2012 held in conjunction with ICFCA 2012 (6 - 10 May 2012, Leuven, Belgium) continued this activity, again involving SHU in the form of Simon Andrews as one of the workshop chairs.

A paper was 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 International Journal of Distributed Systems and Technologies (IJDST) (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.

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 was submitted under a Department Homeland Security (U.S.A) Human Factors/Behavioural Sciences Division funding opportunity. The project aimed 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. Although the proposal reached the final rounds it was sadly not accepted.

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. Both of these proposals have been successful and are now at the negotiation stage.

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.

Ecole 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 that 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

<Confidential>

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

Recalling CRSA's exploitation targets in D.5.1, the main motivation for CRSA/MAS laboratory is to produce relevant knowledge in data mining, semantics and visualisation. Hence its principal exploitable targets are:

- Formal Concept Analysis – 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.

- Big Data and Visualisation – Big Data and Visualisation are two essential aspects for the processing and understanding of large amount of data in modern companies. CRSA has been investing efforts on the research and development of distributed, computationally-intensive graph manipulation and querying algorithms.

- Graph manipulation and analysis – CRSA investigates new methods to extract, manipulate and query graph data. This includes the extraction of graphs from

<Confidential>

relational databases, semantic expansion from unstructured data sources and visual querying based on semantic and topological features of the graph.

Although CRSA is implied mostly in Visual Analytics in CUBIST, it nevertheless contributes systematically to the back-end data mining algorithms such as the "*tree-extraction from concept lattices method*" which helps navigating through large conceptual data, or the progressive clustering of similar formal concepts, which provides more compact visualisations.

2.4.3 Exploitation Activities and Results

Following the exploitation activities described in D.5.2.1, CRSA evaluated existing and new visualisations with each of the use cases. In the integrated prototype, new features were added while others, less useful, were removed. Overall these activities have broadened the scope of use of our methods and tools, resulting in five research publications, one integrated prototype for CUBIST and a number of potential applications. Some distinctive results of this endeavour are described below.

2.4.3.1 Association Rules Analytics

A four-month long internship was offered by CRSA focused on the research of new analytic features for Association Rules (AR). Having ended in August 2012, three new visualizations combined with statistics and charts were developed to enable progressive exploration of the *ruleset*. A traditional matrix view (see Fig 5 b)) where each rule is displayed in a row and the concerned pairs of attribute-value in columns. The confidence of each rule can be measured by the opacity of each cell. The second visualization is a radial graph showing how pairs of attribute-value imply to each other. The confidence of a rule is represented by the thickness of the connecting line. Finally, a bubble-like graph visualization displays premises and conclusions as connected bubbles with the concerned attribute-value pairs inside each bubble (see Fig 5 a)).
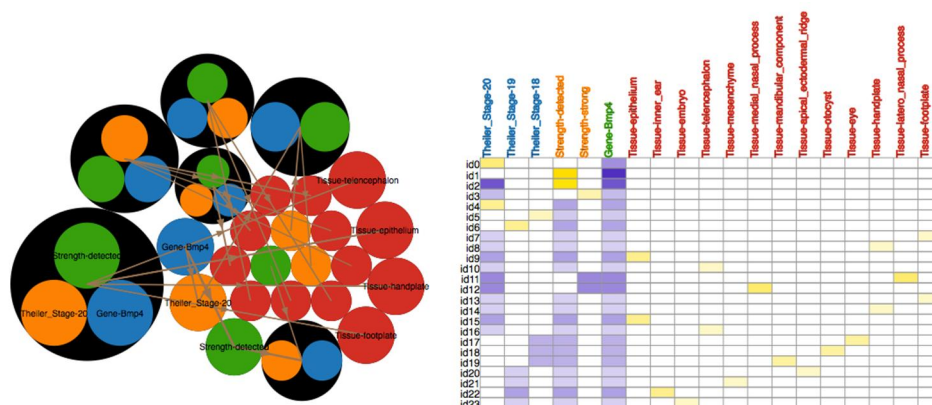

Fig 5: a) A "Bubble" visualisation for the Association Rules. b) Matrix view.

<Confidential>

2.4.3.2 Conceptual analysis of complex system simulation data for decision support: Application to aircraft cabin design

Some of the visualisations created for CUBIST proved to be useful in other domains, such as *Complex System Design*. In a joint research with the University of Tours, a method for dealing with continuous attributes in FCA (e.g. temperature) and corresponding visualisations were proposed. The approach takes advantage of the use of Similarity-based Formal Concept Analysis (SFCA) to classify, visualize, and explore simulation data in order to help system designers to identify relevant design choices. In contrast with traditional FCA which takes as input a binary table of objects and attributes, SFCA uses a similarity measure to group multi-valued attributes in their corresponding concepts. The approach was tested on an aircraft cabin design case study, which concerns the simulation of different configurations of the ventilation system to study the passengers' comfort in the cabin[i].
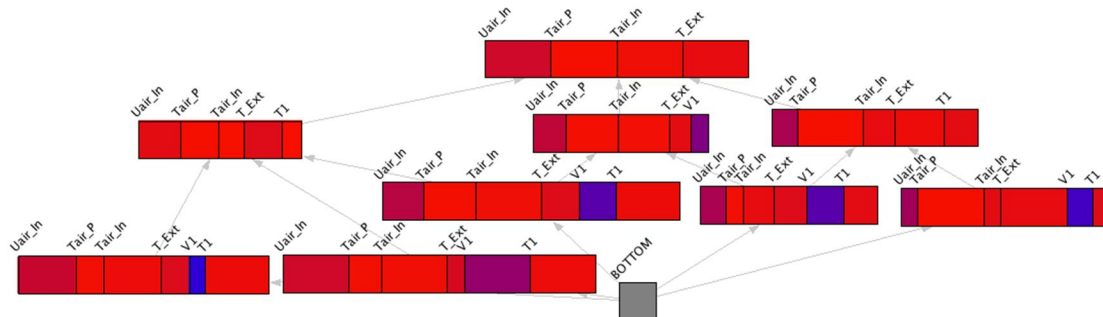

Fig 6: One of the visualisations created for multi-valued concepts. Each box is an attribute in the concept. The box colour indicates the attribute value in the continuum (from blue to red) and the box width represents the length of the interval.

2.4.3.3 Analysis of Dynamic Communities in Social Networks using Conceptual Measures: Application to Business Process Management

Another interesting outcome of the research done in CUBIST was the extension of the conceptual measures, used to simplify concept lattices, to the Business Process Management domain. The measures determine conceptual proximity between people and enterprise resources in a case of market analysis. The temporal evolution of this conceptual measure was analysed, and it provided significant insights on trends and market behaviour. The approach has been exemplified with a case study on Twitter[4] with an emphasis on content dynamics within user communities[ii]. This study revealed promising research perspectives for future work.

---

[4] http://www.twitter.com

Another branch of the research team is currently working on influence analysis and patterns of information dissemination in Twitter using FCA, among other techniques.

2.4.4 New Exploitation Activities

As part of the final year of the project, CRSA will increasingly concentrate efforts on the integrated prototype and evaluation with use cases. In turn, new research will be needed to address the challenges it may encounter. CRSA already started scientific collaboration with some use case patterns. CRSA resumes the initial perspectives for exploitation next year:

- Extension of CUBIST to other domains like the ones aforementioned;

- A research on influence analysis on job-seeking data in social network (in collaboration with Innovantage);

- Exposition of CUBIST to digital innovation catalogues, conferences, seminars and data mining classes;

- Open sourcing of non-proprietary software developed by CRSA;

- A new internship offer will be available from Mars 2013 to work on the integrated prototype on CUBIST.

2.5 Space Application Services

2.5.1 General Exploitation Strategy of Space Applications Services

Space Applications Services (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

<Confidential>

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 facilities serve a multitude of mission types, e.g. earth observation missions, scientific missions, manned missions (Columbus/ISS), navigation and surveillance systems (EGNOS, Galileo), etc., and they 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[5] known as User Support Operations Centres. The USOC especially chosen first is the Belgian User Support and Operations Centre (B.USOC) that is operated by operators of Space Applications Services, who first provided their user requirements and will afterwards evaluate the successive versions of the CUBIST's Semantic Business Intelligence platform.

At the time of writing D5.2.1, SAS focused on the utilization of CUBIST's advanced integrated visual analytics for fast and correct root cause finding of SOLAR instrument errors. This focus has not significantly shifted since, so to summarize:

- 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. 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. Finally, 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 exploitation targets within the aerospace domain such as improved usability, integrating novel concepts and semi-automatic generation of new documents.

2.5.3 Initial Exploitation Activities and Results

At this stage, a number of activities already performed so far in the project (Task 8.1 "Requirements Analysis", Task 8.2 "RDF Modelling of Data Sources", and Task 8.3 "Use Case Prototype, v1") allowed early exploitation of results, and/or have open the door to new types of exploitation not (easily) previously achievable:

D8.1.1 – 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

---

[5] http://www.esa.int

<Confidential>

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.

As mentioned in D5.2.1, 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 decision makers: 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".

- 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 extended with the full package of unstructured data from that same period or useful for the Use case. This extended second CUBIST Space Data Pack provides the complete package of structured and unstructured data from a limited period for an operator on console, allowing the technological partners full transparency for the use case and the development of the prototype.

- Further analysis of and support to the Space Control Centres use case and the review of the initial prototype from the technological partners, have provided the B.USOC more insight in the possibilities of CUBIST and the further use of it on-console for near-real-time analysis. Therefore the creation of a third space data pack has been initiated, covering data from re-occurring anomalies from real-time operations. This with the intention to use CUBIST for helping and identifying underlying patterns in the telemetry, quick identification of a known anomaly and perhaps indicating possible root-causes, but also to further familiarise the operators on-console with the CUBIST for the day-to-day operations.

<Confidential>

D8.2.1 – 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".

D8.3.1 – Use Case Prototype v1

- Following the deliverables above, this deliverable started with the initial process of converting the structured part of the Space Data Pack into RDF format so that it could be stored in the OWLIM triple store. With the help of Ontotext, an automatic Talend process had been designed that took the structured telemetry data and the associated ontology into account and that can could be run in a batch fashion for telemetry data covering other periods. This was an important step because it now provides the opportunity to evaluate the utilization of CUBIST's advance visual analytics not only for a single instrument failure, but also for other failures from other time periods.

- Once the structured telemetry data had been stored in the triple store, the generic CUBIST prototype had been customized to reflect the SAS look and feel with the help of SAP.

- In the third phase, a few preliminary analytics have been described by SAS operators working at B.USOC and these have been converted to corresponding SPARQL queries.

- Finally those queries had been integrated into the use case prototype and provided on the web user interface.

2.5.4 New Exploitation Targets

When SAP took an initiative in offering SAS the opportunity to evaluate the feasibility of utilizing Complex Event Processing (CEP) for detecting patterns for error prediction in telemetry data, which in a sense is very high dimensional time series data, SAS expressed its interest to learn more about it. SAP has addressed SAS with the proposal to conduct a bilateral project in order to conduct CEP on the SAS data. This Initiative started in August 2012; a first call between SAS and SAP has been conducted on September 6. In this call, two demos for CEP have been given to SAS, and possible legal frameworks between SAP and SAS have been discussed. The first feedback from SAS is quite promising, and currently, SAS evaluates internally the participation in a bilateral activity within the upper management.

2.6 Heriot-Watt University

2.6.1 General exploitation strategy of 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 (BSA) 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.

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 spatial-temporal and distributed biomedical data.

2.6.2 Exploitation targets

During the initial year of CUBIST, as reported in D.5.1, HWU concentrated on two targets:

- Semantic representation of spatial-temporal biomedical images/data; and,
- The application of FCA tools for the exploration of biomedical data.

However, in the second year, the focus shifted to:

- Semantic representation of spatial-temporal biomedical images/data.

Following feedback from the potential user community, HWU will spend the final year of CUBIST concentrating on:

- Semantic representation of spatial-temporal biomedical images/data; and,
- The visualisation of analytical query results.

2.6.3 Initial exploitation activities and results

As an academic institution, HWU's two primary goals are the publication of papers and the transfer of knowledge.

In addition to writing a number of papers (see D5.1.6. Dissemination Report v.2 for details), a further three papers are currently being written with a view to publication next year.

HWU routinely brief EMAP on the continued successes of the CUBIST project. It is hoped that some ideas, and perhaps technologies, will eventually be implemented in one of EMAP's

two main products (EMA and EMAGE). This is unlikely to happen until after the ideas have been fully evaluated. Accordingly, such a transfer will probably occur after the conclusion of the CUBIST project.

Dr Burger participated in an internal HWU seminar in which he briefed colleagues on the full range of his activities, including CUBIST.

Additionally, HWU will shortly take part in a BSA seminar. HWU will present to biologists, bioinformaticians and computational biologists at the HGU. Some of those present will be from EMAP, but research staff from other BSA projects will form the majority of the audience. For many, this will be their first introduction to CUBIST.

Knowledge of semantic spatial representation, gained through the efforts of CUBIST, is being transferred into another EU project – Ricordo[6]. There are no standards for describing biological space. However, Ricordo is tasked with delivering recommendations concerning volumetric space to the Virtual Physiological Human[7] (VPH) network of excellence. Many of the lessons learnt during CUBIST are relevant to this, and Dr Burger is ensuring they feature in the Ricordo report.

Dr Burger is a member of the International Neuroinformatics Co-ordinating Facility (INCF) taskforce on Digitial Brain Atlasing[8]. During their recent meeting in September, Dr Burger highlighted the importance of semantic spatial descriptions. His comments were greeted positively, and there was general agreement that investigations should take place within the existing INCF atlasing framework.

2.6.4 New exploitation activities

In the second year HWU mainly concentrated upon their initial targets and activities in order to ensure successful completion. Additionally, HWU are exploring the possibility of collaborating with the HGU, and the European Bioinformatics Institute (EBI) to adapt and extend some of the EBI's existing tools. In particular, the knowledge of semantic spatial representations will be used in order to ensure the tools, and the data they produce, are spatially aware.

Furthermore, HWU will shortly commence work on a new activity. There is considerable interest from EMAP in the visualisation tool (CUBIX) developed within CUBIST. Accordingly, HWU will pursue this avenue of research, with the aim of developing a more

---

[6] http://www.ricordo.eu/

[7] http://www.vph-noe.eu/

[8] http://incf.org/programs/atlasing

generic (not confined to FCA) set of visualisations. This shall be the second of HWU's two core exploitation activities in the third year.

2.7 Innovantage

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.

2.7.1 General Exploitation Strategy

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.

<Confidential>

- **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.

2.7.3 Initial Exploitation Activities and Results

Although it is expected that the integration of CUBIST will take place once the development project has completed, Innovantage are keen to demonstrate the potential of CUBIST of existing and new clients.

Since the initial version of this document was written Innovantage has been contracted to supply data for a new product being developed by a major UK based job board. Innovantage will be demonstrating the CUBIST prototype to senior executives within this company.

2.7.4 New Exploitation Activities

Innovantage has been developing a classical BI reporting tool. This product, 'Insight Analytics', has already been sold to a number of job boards and two of the UK's largest recruitment consultancies.

This product offers the user access to a number of predefined reports appropriate to their vertical market, e.g. a market penetration report for job board operators, or active recruiters report for recruitment consultancies.

The CUBIST prototype will be demonstrated to selected, 'friendly', clients to acquire product feedback and to gauge interest.

It is envisioned that CUBIST will be integrated into the existing Analytics product to provide an alternative viewpoint. The diagram shows an existing Analytics product with a representation of how CUBIST may be integrated:

<Confidential>

Fig 7: Analytics product with a link into the Formal Concept Analysis powered by CUBIST.

Reports within 'Insight Analytics' operate standard 'pickers' that allow users to select criteria, such as geographic region, discipline, number of postings, range, etc., from which the report is generated. The pickers are consistent across reports, i.e. the same fields are offered to the user no matter which report is selected. In the CUBIST prototype system the selection options are derived from the ontology. While this is powerful in allowing the prototype system to seamlessly handle the different used case ontology's, the very generic nature of this mechanism means the users must understand which options are appropriate and will lead to comprehensible results. Before CUBIST could be offered to Innovantage users, the criteria offered would have to be controlled by standard 'pickers'. The image provided above shows standard pickers used within the current Insight Analytics product.

<Confidential>

3 Summary

This document reports on our 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. This document is an "Interim", following D.5.2.1 "Exploitation plan / report, v.1", and to be refined subsequently throughout and according to the further conduct of the project. There will be another, extended version of this document, namely D5.2.3 "Exploitation plan / report, v.3", which is due in M36.

---

[i] Published in the *Concept Lattices and Applications 2012 – (CLA 2012)*

[ii] Published in the *The 5th Workshop on Business Process Management and Social Software (BPMS2'12)*