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.3

Abstract: n/a

Type	Report
Document ID:	CUBIST_D5.2.3
Workpackage:	WP5
Leading partner:	SAP
Author(s):	Frithjof Dau (SAP) Frithjof Dau (SAP) Pavel Mihaylov (ONT) Simon Andrews (SHU) Marie-Aude Aufaure (ECP) Ken McLeod (HWU) Jan Depauw (SAS)
Dissemination level:	PU
Status:	draft
Date:	04 November 2013
Version:	0.96

Versioning and contribution history

Version	Description	Contributors
0.1	First draft with SAP exploitation	Frithjof Dau (SAP)
0.2	HWU Exploitation	Ken McLeod (HWU)
0.3	Extended chapter “consortium-wide exploitation”	Frithjof Dau (SAP)
0.4	Draft for SAP-exploitation added	Frithjof Dau (SAP)
0.5	Revised SAP chapter Added section for ECP	Frithjof Dau (SAP) Marie-Aude AUfuare (ECP)
0.6	Added sections for SAS and SHU	Simon Andrews (SHU) Jan Depauw (SAS)
0.7	Added Section for Ontotext	Pavel Mihaylov (ONT)
0.8	Review comments from FD addressed in SAS and SHU sections	Simon Andrews (SHU) Jan Depauw (SAS)
0.9	Added Section for Innovantage	Dolly Thomas (INN)
0.95	Review comments from KM addressed by SAP, SHU, ECP	

Reviewers

Name	Affiliation
Frithjof Dau	SAP AG
Ken McLeod	HWU

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 in the first two years of the project have been described in the CUBIST deliverables D5.2.1 "Exploitation Report v1" and D5.2.1 "Exploitation Report v2". As described in those deliverables, the consortium has conducted 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 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

According to Gartner's latest CIO (chief information officer) survey1, analytics and business intelligence remains the number one technology priority for 2013. Thus the BI-related results CUBIST are certainly exploitable. However, CUBIST is not only a BI-platform. It does not provide only means to analyse data: it provides means to search for data, navigate data, and

[1] http://www.gartner.com/newsroom/id/2304615

explore data as well. With respect to information structures, it is worthwhile to look at Gartner's report "Top Technology Trends Impacting Information Infrastructure in 2013"2 as well in order to position CUBIST in this area. So, in this section, we shortly recap and adjust recap the key areas with exploitable results of CUBIST, as they have already been discussed in D5.2.1 and D5.2.2.

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 the form of unstructured data, and the importance of incorporating unstructured data in BI is widely accepted. Note that particularly in the CUBIST Innovantage Use Case, the core to be analysed – namely job vacancies - is obtained from scraping job boards on the Web.

   Federating data from unstructured sources is possible with existing BI tools (e.g. with SAP Data Services"), i.e., BI-vendors have already addressed conducting BI over unstructured data. Thus CUBIST results in this area are only moderately exploitable.

2. **Semantically enabled BI**

   In D5.2.1 we argued that traditional BI systems usually store their data in data warehouses, which store structured data (normally) using a dimensional model. This contrasts with 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. But some BI-solutions support so-called "semantic layers", being sets of predefined business objects that represent corporate data in a form that is accessible to business users. These business objects - such as metrics, dimensions, and attributes - shield users from the data complexity of schema, tables, and columns in the data warehouse. Thus traditional BI-solutions partly utilize benefits of semantic technologies (but not all of them, e.g. the schema-last approach of Semantic Technologies (ST), which is considered as a key benefit of ST).

   On the other hand, in the field of Semantic Technologies and semantic repositories, we have a lack of BI-capabilities in the existing applications. Standard BI-means are only partly suited for the graph-based data used in Semantic Technologies, but the novel Visual Analytics means in CUBIST better fit to the paradigms of Semantic

[2] http://www.gartner.com/newsroom/id/2359715

Technologies (compared to the paradigms of relational databases and data warehouses). Thus it is more promising to exploit CUBIST in the form of "bringing BI to ST" instead of "bringing ST to BI".

3. **Comprehensive information access means, advanced visual analytics and qualitative data analysis**

   CUBIST provides comprehensive information means that go beyond traditional BI-solutions. It is possible to search for data (factual search), navigate in data in a textual form or explore data using lode-link-diagrams (explorative search), and analyse the date with visual analytics. This is a distinguishing feature of CUBIST and thus a unique selling point for exploitation activities.

   Next, 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 the cluster hierarchy. The visualizations are graph-based diagrams with nodes representing clusters and lines representing hierarchies. This is the next distinguishing feature of CUBIS: allowing both, meaningful, *qualitative* data analysis and graph-based visualization of that analysis. The current CUBIST prototype offers additional visualizations like tree maps, sunburst diagrams, icicle diagrams and scatter plots. To date, no BI vendor offers Visual Analytics products that are capable of doing similar analytics and visualizations.

In D.5.2.1 and D5.2.2, the exploitable results have been summarized in a diagram. A revised version of the diagram is shown in Fig 1.

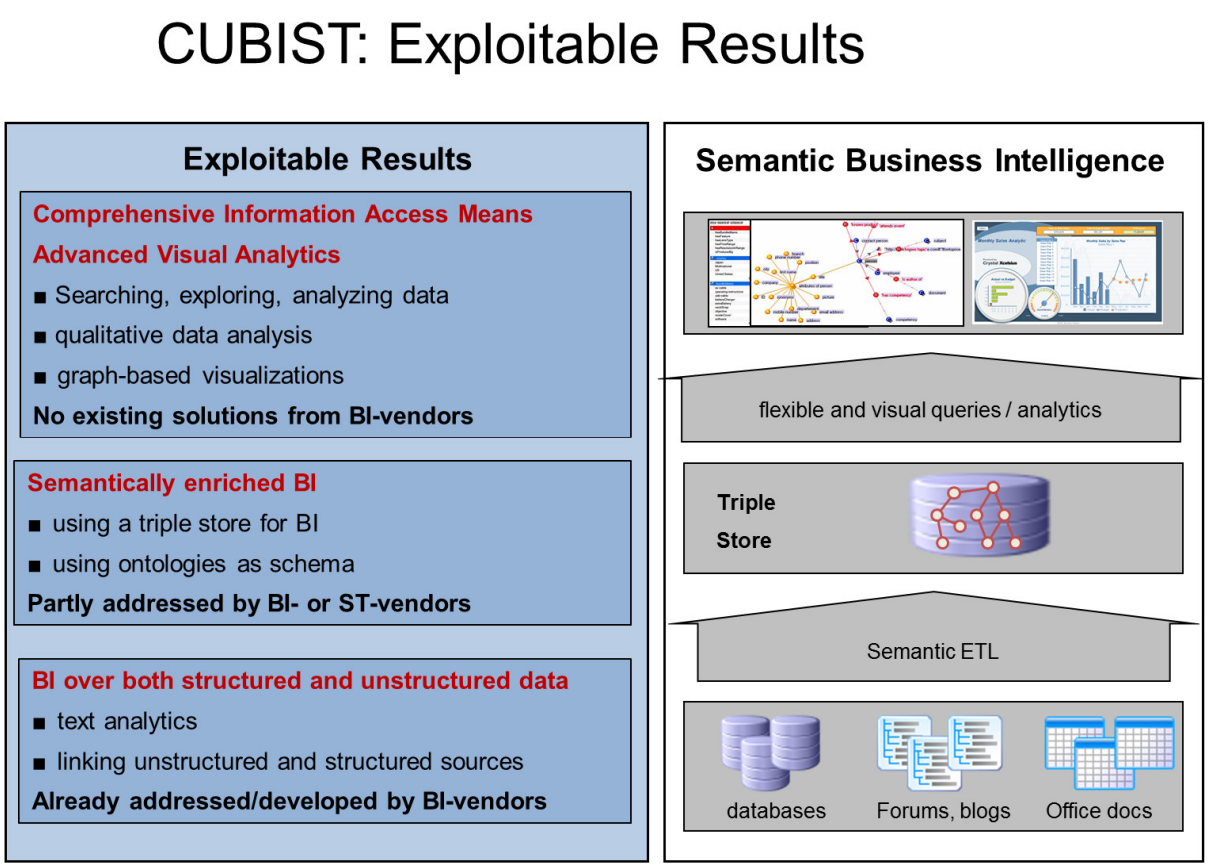


Fig 1: Three core fields for exploitation in CUBIST

As it just has been discussed, and as it turned out during the life of the project, amongst these three areas, the novel Visual Analytics of CUBIST raised the most interest for exploitation, both for CUBIST partners and for parties outside the consortium (e.g. scientific communities or external companies).

2 Joint Exploitation: CUBIST goes Open Source

As already been envisioned in D5.2.2, the CUBIST-consortium decided to publish the CUBIST prototype as open source. In this chapter, we provide the rationale for doing so, the targeted license model, and the current status of the open source process.

2.1 Rationale

Apart from individual exploitation plans (see next chapter), the CUBIST consortium decided to release CUBIST as open source, under a permissive license, to the public. There are the following reasons for doing so:

- **Scientific Communities are interested:** CUBIST has been shown at scientific events and disseminated in scientific communities (e.g. the semantic web community or the FCA-community), and members of those communities expressed their interest in the prototype and the opportunity to not only use it, but to develop it further. For example, Frithjof Dau gave an invited talk about CUBIST at the 11th International Conference on Formal Concept Analysis (ICFCA) in Dresden, Germany (May 2013). The feedback was very positive, the audience was impressed by the capabilities of the prototype and expressed interest to use the prototype and to develop it further, if it will be published as open source. Pavel Mihaylov and Marin Dimitrov (Ontotext) have evangelised CUBIST at several conferences related to the Semantic Web and received interest in CUBIST as well. Moreover, some CUBIST partners have propagated CUBIST to third parties related to them. To name an example: HWU is in constant contact with EMAGE, which even resulted in the senior editor of EMAGE, Chris Armit, enthusiastically taking part in the CUBIST evaluation. Other research institutions apart from HWU and EMAGE working on gene expression experiments are likely to become interested into the usage or further development CUBIST as well.
- **Evangelism of FCA-technology and ST-technology:** By releasing CUBIST as open source to the public, the CUBIST consortium aims at a better adoption of solution based on Semantic Technologies or Formal Concept Analysis by the market in the future. The more people use such solutions, the more people will be attracted and accept the novel CUBIST solutions. This in turn will ease the partner-wise exploitation of CUBIST results.
- **Partner-wise exploitation without loss of economic advantage:** As it will be discussed in the next chapters, several partners plan to exploit CUBIST further, for scientific or commercial purposes. Even if CUBIST is released as open source, it should be mentioned that the CUBIST partners do not lose their competitive advantage. The overall code of the prototype is quite extensive and sometimes complex. Thus for any third party interested in adapting the code, it requires to invest

time and resources upfront to understand it. The CUBIST-partners have an in-depth-knowledge of the prototype and the code, which significantly eases any envisioned customizations of the prototype code to partner-specific requirements. Moreover, publishing CUBIST as open source does not prevent any further collaboration between partners, but is eases it, as with respect to licensing no further legal arrangement based on the consortium agreement has to be set up.

With these rationales in mind the CUBIST consortium considers the publication of CUBIST as open source not as win-win-situation for the CUBIST partners and the public audience.

2.1.1 License

The CUBIST consortium discussed the appropriate licensing model for the OS-publication of CUBIST. In order to best enable a partner-wise exploitation, it is clear that a permissive license is needed for the partners. We have discussed a dual-licensing model, e.g. a very permissive license only within the consortium, but a restricted license for the public (e.g. a license which allows to adapt the source further, but only for private or scientific, but not commercial purposes). Such a dual-licensing-model is not feasible for two reasons. As e.g. discussed at SAP internally with open-source-experts and the legal department, a dual-licensing-model is too costly and complex to be established. Secondly, a dual licensing model would require a regular check to determine whether third parties violate the license, which again is too costly to be carried out. Based on these considerations, the consortium targets a single, permissive license. There are different, yet similar permissive licenses, like BSD, Apache 2.0, or MIT. Moreover, there are different places to publish open source software, e.g. SourceForge or GitHub. Based on a recommendation of SAP, we intend to publish the prototype:

- Under the Apache 2.0 license, and
- On GitHub.

In fact, SAP has its own section at GitHub, see https://github.com/SAP.

2.1.2 Current Status of the Open-Source-Publication

Components of three partners have to be individually published. A detailed overview over the different components can be found in D1.2.1 “Reference Architecture”. A short recap of those components and the current status with respect to their publication as open source is given below.,SAP: The contribution of SAP to the prototype is NowaSearch and the Graph Exploration Panel. NowaSearch consists both of the backend and the frontend which enables the user to search for, filter and select entities in the repository by means of faceted search. From NowaSearch, the other services/components (i.e. the Graph Exploration View, FCAService and CUBIX) are called. The Graph-Exploration panel view is that view in the front-end which allows to explore (but not analyse –“analyse” in the BI-understanding) the

content of repository by means of a graph-exploration view, i.e. a node-link-diagram which can be interactively processed.

- SAP, NowaSeach:

  SAP is generally open for publishing open source. For doing so, i.e. for outbounding the software, there exists a process at SAP AG. This process has been started for NowaSearch and the Graph Exploration Panel in August 2013. In this process, several legal checks are conducted, and an approval of midlevel-managers is seeked, before the software is published as open source. During the legal check, IP- and license-related questions are clarified, and SAP-internal stakeholders have to be addressed who have to assess the open-source-request. Those checks are currently, as of writing this deliverable, conducted. Though for the time being, it is likely that the open-source-request will be approved, a final commitment cannot be given yet. A detailed description of the process is provided in Section 3.1.

- SHU, FCAService:

  A first version of the FCAService is published at https://github.com/acesco1/rdf2fca-service under the Apache License 2.0. Updates with additional features are planned still for 2013.

- ECP, CUBIX:

  Cubix v2.1. (front-end snapshot) is published at http://github.com/ksiomelo/cubix under the Apache License 2.0. The code is currently being refactored to make it easy to integrate with partner's platforms. The design, API, and other important information is described at https://github.com/ksiomelo/cubix/wiki .

3 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 and D5.2.2, a section for each partner is provided. Per partner, the exploitation strategy and targets are reiterated, emphasizing significant changes w.r.t. D5.2.2. We provide a recap of important exploitation activities in the first two years of CUBIST as described in D5.1.2 and D5.1.3, as well as new activities in the third year of CUBIST. Finally, we describe the planned exploitation activities after the official end of CUBIST end of September 2013.

3.1 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[3], SAP is the dominant leader in the combined BI market (platforms, CPM suites, and analytic applications and performance management). With annual revenue (IFRS) of €16,22 billion, SAP is the enterprise application software market leader helping companies of all sizes and industries run better. 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 65.000 employees (by the end of Q3/2013) and more than 248,500 customers in 188 countries.

SAP's vision is to help the world run better and improve people's lives. Our mission is to help every customer become a best-run business. We do this by delivering new technology innovations that we believe address today's and tomorrow's challenges without disrupting our customers' business operations: Enterprise mobility will transform consumption of IT; in-memory technology will simplify the IT architecture in the enterprise and drive high-value applications; and the cloud delivery of IT solutions will simplify the consumption of technology and enable business networks.

By leveraging our leadership in applications and analytics and combining them with new technology innovations, we can offer solutions that make our customers run better. To help our customers derive value from their SAP solutions in a fast, cost-effective, and predictable way, we also provide professional services and support.

3.1.1 General Exploitation Strategy

By combining its strengths of applied research and product development, SAP anticipates and explores emerging IT trends and transfers innovations into market-ready prototypes, product enhancements, and new solutions. Research and innovation teams from SAP's R&D area evaluate and prepare business opportunities that inspire both enterprises and end users. Research activities span from co-innovation with customers to collaborative research in consortia with industry and academia. With teams close to markets on all five continents, SAP's research endeavours are a powerful driver of innovation for the benefit of our customers and partners.

[3] 2011 Gartner Report: Market Share Analysis: Business Intelligence, Analytics and Performance Management, Worldwide, 2010

Fig 2 illustrates the invention process for research activities at SAP. Project execution mainly overlaps with phase 3 ("Co-innovative Research"). Any concepts, or artefacts built during the course of the project will be used in the transfer phase. They can lead to completely new technologies, new product ideas or the improvement of existing products. Alternatively, customer pilots and spin-offs are also considered as possible exploitation paths.

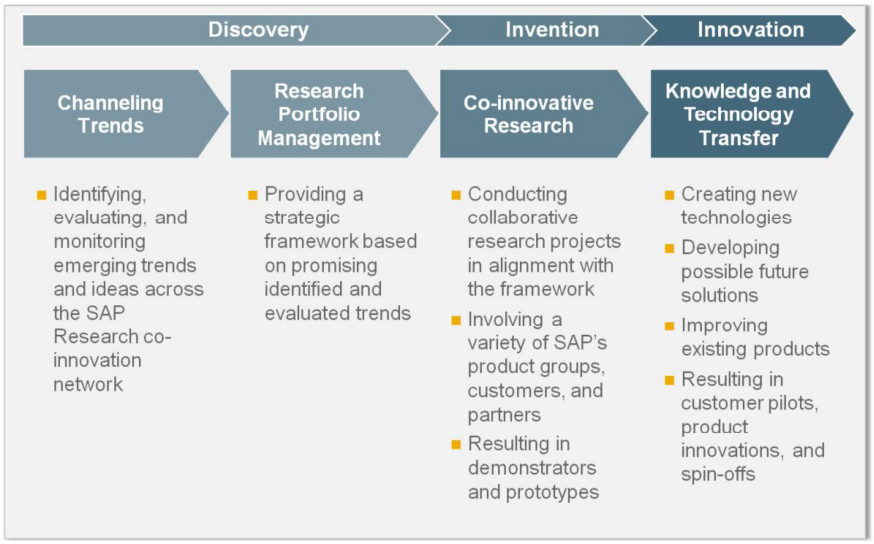


Fig 2 Exploitation Strategy for research activities at SAP

As CUBIST is a research project, it is obvious, that transferability can not be guaranteed. Although the topic of the project was aligned with the corporate strategy during the "Discovery" phases, it might happen that the artefacts produced in CUBIST do not fit into the evolving strategy of the departments to uptake the results or are simply not requested by the market. Thus it is not sufficient to only explore internal exploitation, as this approach might not succeed. For this reason, we now drive on two tracks:

1) On the one had, we have started in Q3/2013 an internal process to publish the SAP-contribution to the CUBIST prototype as open source. As described in the last chapter, we do not expect to loose competetive advantage if we release the prototype as open source. The final decision on the open-source-publication of SAP AGs part of the prototype has not been taken yet. The prototype is currently presented to internal stakehoders, whose assessment will be the basis of the further process. More details of the process are provided in Sec 3.1.4.1.

2) The prototype is actively presented and discussed with SAP-internal stakeholders who are likely interested in CUBIST. Those stakeholders comprise of course the stakeholders who evaluate CUBIST from an open-source-perspective, as described above. More stakeholders (e.g. the project manager of a project in which proteins are investigated, a use case which which is closely related to the HWU use case in CUBIST) have been proactively addressed. More details about the stakeholders and the current status of their take on exploiting CUBIST are provided in Sec 3.1.4.2.

Driving those two tracks has a very clear benefit. There is essentially only one reason that can stop the open source publication of the SAP part of CUBIST, namely if one of the stakeholders has a concrete exploitation objective for CUBIST, and for targeting this objective, the CUBIST code will be reused and will thus not be released as open source. So, even though at the present time it can can not be guaranteed that CUBIST results will be transferred into SAP products, it is clear that either the prototype will be relased as open source, or CUBIST results will be transferred into SAP products (or both). Thus the exploitation of CUBIST is garantueed.

3.1.2 Exploitation Targets

As already stated in D.5.2.2, 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.
- 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. and D5.2.2, SAP focused on the advanced visual analytics. This focus has not significantly shifted. In addition to the visual analytics, currently the faceted search engine of CUBIST, backend and frontend, are being investigated by some stakeholders with respect to possible exploitation.

3.1.3 Recap of Activities as Described in D5.2.1 and D5.2.2

In this section, we recap important exploitation activities as described in D5.2.1. and D5.2.2., and we report on the progress of these activities.

3.1.3.1 **Diploma Thesis**

As described in D5.2.2, a Diploma thesis, conducted as a side activity to CUBIST, has been carried out. The thesis focused on the visualization (by means of Hasse-diagrams) and interactive exploration of concept lattices, and resulted in a fully functional and documented Java-library for the FCA-visualization, covering both the overall functionalities needed by any FCA-tool (e.g. the computation of the concept lattice out of the formal context) as well as the implementation of new features. This library has been incorporated into an SAP-owned version of the prototype, replacing the visualizations in the CUBIST prototype, as they were developed by the CUBIST partner ECP. Screenshots of this prototype are provided in the next figures (3 and 4).

The library has been made available within SAP via an SAP-internal code-sharing platform. Moreover, an SAP-internal prototype owned by SAP only has been developed. This prototype is based on a new dataset using SAP-data, namely product-feature-matrices for

BusinessObjects Products. For this prototype, a demo-presentation and video have been produced, which is used for SAP-internal dissemination to attract SAP-internal stakeholders.

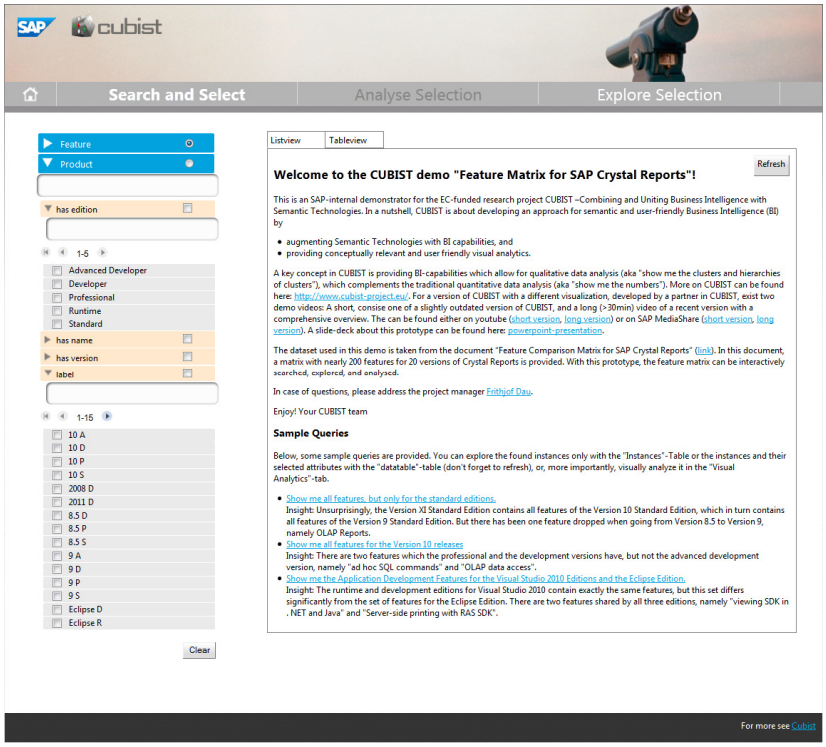


Fig 3: CUBIST SAP Internal Prototype, Startscreen

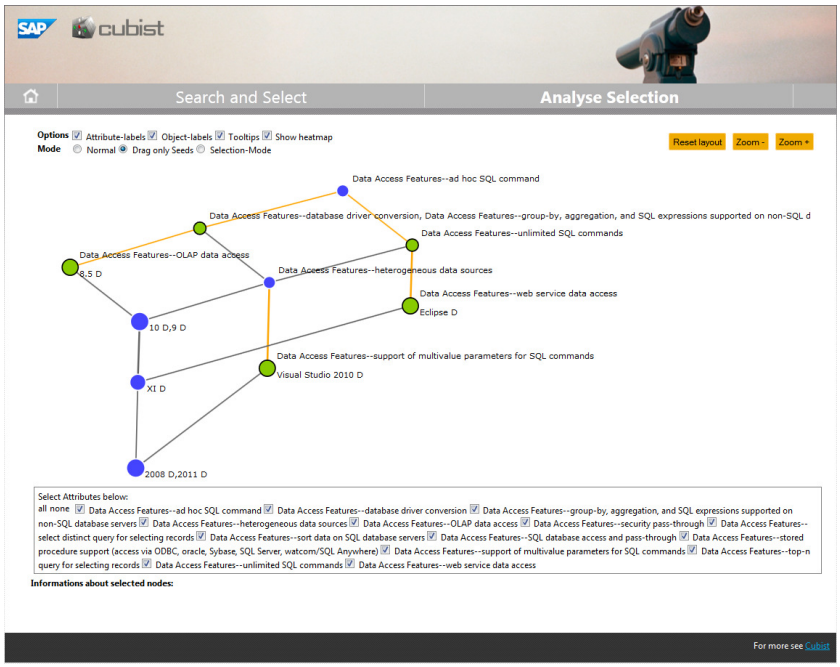


Fig 4: CUBIST SAP Internal Prototype, Example Analytics

3.1.3.2 Bilateral activity between SAP and SAS

In D5.2.2, we described a prospective bilateral activity between SAP and SAS. SAP targeted a dedicated activity in form of a so-called "Customer Engagement Initiative". SAS in turn evaluated this approach within the upper management, and decided not to join such an initiative, but instead suggested to submit a follow-up proposal of CUBIST to the EC. Due to a re-organisation of SAP-Research at that point in time, it was not possible for SAP to commit to a new public funded project, thus SAP and SAS agreed on not following this activity further (instead, SAS will benefit from CUBIST and exploit it further as described in Chapter 3.6.)

3.1.3.3 CUBIST addressed by external companies

As described in D5.2.2., due to the CUBIST dissemination activities (see D6.1.6), CUBIST had been addressed by two external companies which expressed their interest into a collaboration with CUBIST, namely Hatch[4] and Banyas. Unfortunately, in both cases, conducting a further collaboration did not turn out to be feasible for those companies.

3.1.4 New Exploitation Activities in the Third Year

In this section, SAP reports on new exploitation activities that have currently (as of time of writing this deliverable) been started. There are two main types of activities: the publication of (SAP's part of) the prototype as open source, and addressing SAP-internal stakeholders (e.g. product owners of products related to CUBIST) in order to investigate exploitation opportunities. These two activities are related: for publishing CUBIST as OS, an SAP-internal process requires that some stakeholders have to be addressed. We want to re-stress that aiming at the OS publication of the CUBIST prototype and exploring SAP-internal exploitation activities are not conflicting. On the one hand, as we target a permissive OS-license, even an OS-released prototype can still be commercially exploited. On the other hand, it might happen that a SAP-internal CUBIST-exploitation will not be based on the actual *code* of the prototype, but instead on some results of the prototype which have to be (partly) redeveloped to be integrated into SAP software.

In the next subsections, we first discuss the current status of the CUBIST OS release. After this, for specific products and projects, potential and planned exploitation activities are described. This section provides details for the discussion with the stakeholders who have to review the OS request as well.

[4] http://www.hatch.ca

3.1.4.1 **Open Source**

SAP already utilizes open source as a productivity driver and to drive interoperability with open source technologies. This helps SAP to ensure connectivity with well-known open source development environments (such as PHP, Python, Perl, JavaScript, JQuery, Ruby) and by selecting mature open source OS/DB platforms with enterprise-level support this enables SAP customers to significantly lower total cost of ownership (TCO). SAP does not only integrate open source software with SAP's software, but actively participates in various open source projects including Eclipse as well.

As stated in the last chapter, the CUBIST consortium does not want to restrict the exploitation of the CUBIST prototype to CUBIST partners only, as there is already a clear interest in the prototype on the part of scientific communities or third parties related to CUBIST partners. On the other hand, the CUBIST partners will not loose significant competitive advantage when the prototype is released as open source. So the consortium decided to publish the prototype as open source (on GitHub, with the very permissive Apache 2.0 license). This is done independently by each contributing partner (SAP, SHU, ECP).

Of course, a company like SAP does neither consume nor publish open source software without carefully investigating whether this will lead to any unwanted legal or commercial implications. Thus both for the inbound process (consuming OS) and outbound process (publishing OS), there are dedicated processes to follow. For the release of SAP's contribution to the CUBIST prototype as OS, this process has been started in August 2013.

During this process, a legal review is carried out. In this review, amongst others, the following important checks are conducted:

- Does the intended contribution include business application functionality? Is the intended contribution differentiating or non-differentiating? Is there a (potential) functional overlap with existing SAP products? In order to get an answer to these questions, SAP internal stakeholders, e.g. product owners of products related to the intended contribution, have to assess the contribution.
- The proposed license is to be checked. Particularly, the intended contribution, i.e. the actual software code, is validated against the usage of third-party-software and their licenses.

If this review is successfully passed, mid-level managers are asked to approve the OS-request. Once after this final step is conducted, SAP will publish its part of the CUBIST prototype as open source.

As of writing this deliverable, the legal review is ongoing. As CUBIST clearly covers business application functionality, the SAP person responsible for OS and affiliated lawyers have named three internal stakeholders – the product owner of the SAP graph store, a SAP expert in Semantic Technologies, and a product manager for SAP BI analytic tools – who

have to review the OS request. All of them have been contacted and meetings have been conducted where CUBIST and the OS request have been discussed. All of them generally support the OS request and will approve it once some final IP-related questions have been clarified. More details are provided in the next section. Thus though the legal review is not passed yet, chances are very good that CUBIST will successfully be approved in this review.

3.1.4.2 Addressing Stakeholders of SAP Products and Projects

In this subsection, for all CUBIST-related, SAP internal products/projects (respectively the corresponding stakeholders, e.g. product owners) possible and planned exploitation activities are described. All stakeholders are addressed with the following information about CUBIST:

1) The project website and the latest CUBIST videos.
2) Dedicated presentations, focusing on aspects relevant for the OS assessment.
3) Two different CUBIST prototypes, namely the prototype as it has been developed during the project by the consortium, and the SAP-internal prototype which has been developed during a side-activity to CUBIST.

In total, we have addressed the stakeholder of a CUBIST related research project, product owner of three related SAP products, and an expert for Semantic Technologies, as described next.

1) **ProteomicsDB**[5] is a joint effort of the Technische Universität München (TUM) and SAP AG. The term "Proteomics" refers to the large-scale study of proteins, particularly their structures and functions, and ProteomicsDB is a project dedicated to expedite the identification of the human proteome and its use across the scientific community. In the project a proteomics repository for the human proteome is implemented, which is using SAP HANA: SAPs leading column-based in-memory database. The project is in its first year and will run another two years. Even though ProteomicsDB does not use a semantic repository, it is using semantic technologies, e.g. well-known ontologies from the biological domain like the Gene Ontology (GO)[6]. Moreover, the overall goal, context and even the used data structures of this project are obviously closely related to the HWU case in CUBIST (e.g. both for genes and proteins, biologists investigate where and how strongly they are expressed: in tissues for genes and in cells for proteins). We have addressed the project manager of ProteomicsDB and have presented CUBIST to her. For ProteomicsDB, the graph exploration view and the qualitative visual analytics features of CUBIST are of high

[5] https://www.proteomicsdb.org/

[6] http://www.geneontology.org/

interest. It is planned to take over some of the CUBIST-results into ProteomicsDB, and it is currently discussed which results are the most important ones and how the corresponding software components of CUBIST have to be adapted in order to reuse them in ProteomicsDB.

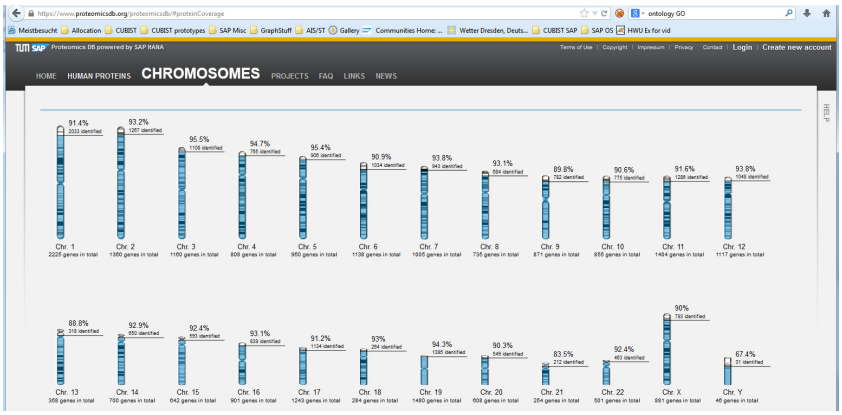


Fig 5. Screenshot of the Current ProteomicsDB Prototype

2) **SAP Active Information Store:** As of HANA SP6, which was released end of June 2013, HANA features a relational database engine, a text processing engine and a brand new component, namely a graph-store called "Active Information Store (AIS)"[7]. CUBIST has been presented in 09/2013 to the product owner (PO) of the AIS, who is one of three stakeholders who has to assess CUBIST in the open-source-process. With respect to the exploitation of CUBIST, the PO of AIS has a vital interest in the graph-exploration-panel of CUBIST, which allows exploration of a semantic repository in a visual, interactive and graph-based manner. This component has been developed by SAP and incorporated into the CUBIST prototype during an early stage of the project. The concepts and functionalities of this component have been thoroughly scrutinized and apply generally to any graph-based data. That is, from a conceptual and methodological point of view, this component is state-of-the-art. Technically, the graph-visualisation is realized with a JavaScript-library called "JavaScript InfoVis Toolkit (JIT)"[8], This toolkit has some restrictions (e.g. a fixed set of layouts for graph-based visualizations). As of end/2013, a different library strongly dominates the field of JavaScript-based visualizations, namely D3.JS (Data Driven Documents). This library is even used in SAP products for information visualizations. So, even though

[7] See e.g. http://www.btw-2013.de/proceedings/The%20Graph%20Story%20of%20the%20SAP%20HANA%20Database.pdf for a recent introduction.

[8] http://philogb.github.io/jit/

the graph-exploration-component of CUBIST is conceptually up-to-date, it is from a technical point of view slightly outdated. With this in mind, the outcome of the discussion with the PO of the AIS is two-fold:

a. With respect to the publication of CUBIST as open-source, this stakeholder does not see significant obstacles, and he supports the OS request.

b. With respect to exploitation, it is planned to redevelop the graph-exploration-view of CUBIST using d3.js. This redeveloped view will be integrated into the AIS workbench, which is the main administrative tool accessing the AIS and which is part of the overall HANA Studio. A mockup is shown in Fig 6.

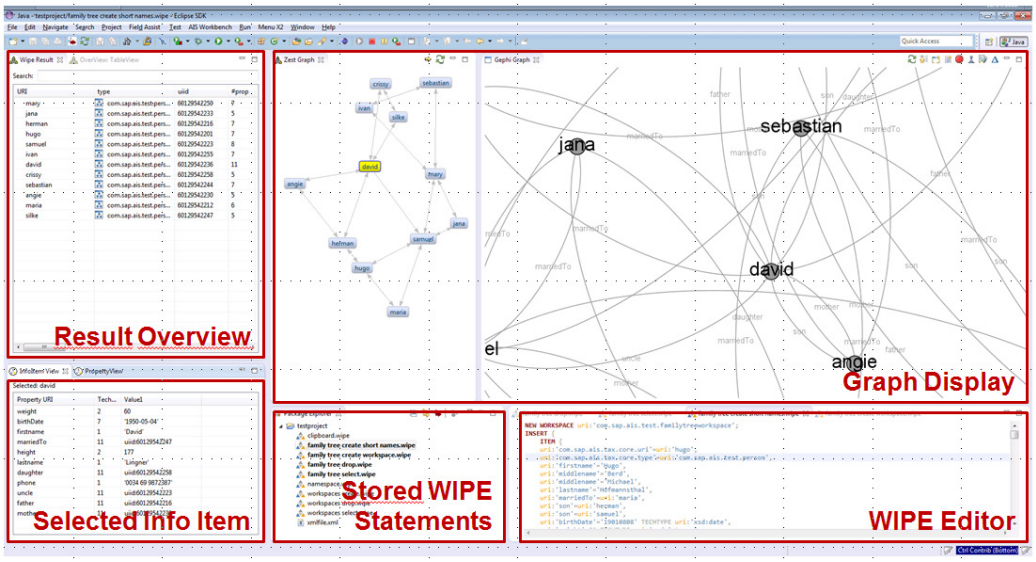


Fig 6: Graph-Exploration in the AIS Workbench

3) **SAP Semantic Technologies:** CUBIST has been presented to a dedicated SAP expert in Semantic Technologies is the second of three stakeholders who has to assess CUBIST in the open-source-process. Acting as a "hub-expert" within SAP, this stakeholder has no dedicated interests in the prototype on its own. Instead, he will disseminate CUBIST further. With respect to the publication of CUBIST as open-source, this stakeholder does support the publication of CUBIST as open source.

4) **SAP BI Analytic Tools:** CUBIST is currently presented to different stakeholders in the field of SAP BI tools. Amongst them is the third stakeholder who has to assess CUBIST in the open-source-process, namely the solution management lead of a SAP BI Analytics Tool. So far, three stakeholders have been addressed, a fourth will be addressed soon. Though showing interest in the graph-based visualization capabilities of CUBIST, none of the stakeholders addressed so far objected to the publication of CUBIST as open source.

5) **SAP NetWeaver Enterprise Search:** Finally, it is planned to address the product owner of SAP Enterprise Search. SAP NetWeaver Enterprise Search is SAP's search solution providing unified, comprehensive, and secure real-time access to enterprise data and information from within and outside of a company. The search returns both structured data (business objects) and unstructured data (HTML files, presentations, and documents) from SAP systems and other search providers, and allows direct access to the associated applications and actions. Some notions of the CUBIST Search-and-Select-functionalities namely the usage of facets, and the usage of relationships between entities for navigating from one entity to another, loosely resemble concepts in SAP NetWeaver Enterprise Search (facets correspond to "categories", and the functionality "related actions" has some resemblance to "Navigate in Data" in CUBIST). Thus it is planned to present the Search-and-Select as well as the Navigate-Data-functionalities to the product owner of SAP Netweaver Enterprise Search in order to discuss and investigate possible exploitation.

3.1.5 Further Planned Exploitation after Project End

As it can be clearly seen from the last section, the exploitation activities in CUBIST will continue.

1) First and foremost, the open source request has to be processed further in order to finalize it. At the present time none of the stakeholders that have been addressed so far had significant objections against the publication of the CUBIST prototype as open source. Thus there is a clear tendency that the open source process at SAP will come to a positive end.
2) Secondly, the exploitation activities for ProteomicsDB and AIS have to be continued in order to successfully integrate CUBIST results into the ProteomicsDB prototype respectively the AIS.
3) Finally, it is planned to further disseminate CUBIST within SAP in order to enable more exploitation.

3.2 Ontotext

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

3.2.1 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 our 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 and web mining platforms and its expertise in the area of RDF-ising legacy enterprise data sources.
    c) Experience with RDB2RDF tools within WP2.
    d) Experience with various data integration and link detection frameworks (Ontotext's own IdRF, as well as Silk and Limes) within WP2.
    e) WP9 will improve Ontotext's job market analysis platform, providing BI for the UK recruitment market (originally a joint collaboration between Ontotext and Innovantage, now developed as a separate Ontotext product).

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.

3.2.2 Exploitation activities and results

The main exploitation opportunities for Ontotext come from technology transfer from CUBIST into existing Ontotext products and solutions. The main opportunities with respect to the four main directions outlined in the previous section are:

1) Analysis of applying semantic technologies (namely RDF databases, semantic ETL approaches, text and web mining) in the context of the three use cases has been performed. 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's products and solutions. Work in the use case work packages provided some interesting requirements that were not foreseen during the initial planning. 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) Results from work in WP2 and WP3 have already been transferred into our products or applied in current industrial projects:

   a) Contributions from Ontotext to the open source framework Sesame 2.6.x and 2.7.x releases in the area of SPARQL 1.1 support and federated queries.

   b) Expertise gained with Google Refine (one of the open source tools that was 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.

   c) IdRF has been successfully used for a proof-of-concept project whose aim is matching CVs to job vacancies across the whole European Union.

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 and D2.3.2, namely:

   a) Performance and query optimisation improvements related to SPARQL 1.1 querying

   b) Better handling of SPARQL 1.1 UPDATE statements

   c) Improvements to the full-text search functionality

   d) More scalable RDF Rank implementation

   e) New index compression scheme

   f) Redesigned and improved plug-in architecture

   g) Various general performance improvements

   h) New OWLIM Workbench with more features

i) Support for OWLIM on the cloud

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

7) Feedback from using our new Identity Resolution Framework (IdRF) in the recruitment domain, which proved useful for another project.

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

3.3 Sheffield Hallam University

Sheffield Hallam is an innovative and responsive university committed to the belief that high-quality inspirational teaching and applied research can transform individuals, organisations and communities. The university's Corporate Plan 2008–2013 sought to identify clear priorities in relation to the student experience, the development of the Academic Portfolio, research base and the efficiency of the university's operations. The plan was refreshed in 2011 so that the institution could achieve greater focus in these priorities and set out a clear statement of its ambition. The University's vision for 2015 is to be an established top 50 UK university, known for the quality of our teaching and learning, with a particular commitment to education for employment. Its research strength will be concentrated in specific subject areas, such as Intelligent Information Management as CUBIST typifies, and will be world-class. By 2015, the university will be known for the quality and range of its partnerships with the professions, business, industry and government.

3.3.1 General Exploitation Strategy

Sheffield Hallam University (SHU) is actively committed to fundamental and applied research, and its dissemination and wider exploitation. SHU has 18 Research Centres in 3 Research Institutes, and several specialised research groups. Many are of national and international renown featuring key research platforms such as materials science, art and design, sports science and engineering, computing and information management, biomedicine, and economic and social research. The Research Assessment Exercise (RAE) 2008 demonstrated that 68 per cent of the university's research activity was of international standing, and continues to be ranked in the top 10 post-1992 Universities for research quality. 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.

3.3.2 Exploitation Targets

SHU continues to exploit CUBIST through publicising the project in a variety of industry sources. These exploitations are underpinned by publication of CUBIST in learned, peer

reviewed journals and conferences. CUBIST feeds into SHU's undergraduate and graduate programmes. SHU has 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 also has 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 the university's PhD students, who will continue to build upon CUBIST in their research 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, as described below. Where commercial or contract research opportunities arise with industry, and as outlined in SHU's general exploitation strategy, these will also be exploited. One example is a presentation of CUBIST to the financial services industry, and government regulators given CUBIST's potential added BI impact. The impact of CUBIST is included in the University's UK Higher Education REF (Research Excellence Framework) submission (http://www.shu.ac.uk/research/achievements/).

3.3.3 Recap of Activities as Described in D5.2.1 and D5.2.2

As outlined in those previous documents, SHU had publicised CUBIST that included "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 the 10th International Conference on Formal Concept Analysis (ICFCA) in May 2012, Leuven, Belgium) continued this activity, again involving SHU in the form of Simon Andrews as one of the workshop chairs.

A paper by SHU on a new FCA algorithm will appear, subject to corrections, in 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, and Enterprise Systems.

SHU exploited CUBIST results to attract and recruit new PhD students. Three PhD students have passed their confirmation of proposal stage, and one has since been successfully awarded their PhD. 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 the now successful student, who is also an experienced SAP Consultant, is "Discovering the Hidden Knowledge in Transaction 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 at SHU were exploited to inform and feed into the ePOOLICE and ATHENA bids. These proposals fall 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 as a Capability Project. Both of these proposals have been successful and have now either commenced (ePOOLICE, since January 2013) or about to commence (ATHENA).

3.3.4 New Exploitation Activities in the Third Year

The recap section above also reflects exploitation that has continued or been realised in the third year, as already described in that section. Further activities include the 3rd CUBIST workshop (CUBIST-WS 2013) that took place in Dresden in May 2013, in conjunction with the 11th International Conference on Formal Concept Analysis, ICFCA 2013. Selected papers from these workshops are to be published in the International Journal of Intelligent Information Systems, thus fulfilling the publication exploitation objective of SHU as well as the dissemination of CUBIST. Again the workshop was organised by SHU in conjunction with SAP Research, as was the journal publication. Much other activity was focussed on completing the project, with the view that major exploitation would occur after the project's end as described next.

3.3.5 Further Planned Exploitation after Project End

In addition to the on-going exploitations above, SHU anticipates significant exploitation in the following two activities

3.3.5.1 Exploitation of the FCA 'fault tolerance' approach

It is clear to SHU that there is great potential for the development of the 'fault tolerance' approach to FCA (Pensa and Boulicaut, 2005), where concepts are accepted even when a

small number of relationships are absent from the closed set of objects and attributes. In CUBIST the approach was developed both theoretically (Dau, 2013) and practically (Andrews and McLeod, 2011) but not fully realised in the final prototype. The potential is not only for approximation of concepts to manage scale and complexity of results, but also in predictive analysis. As shown in Andrews and McLeod (2011), it was possible to predict the results of experiments where there was evidence of missing relations in large sets of gene expression data. The opportunity to develop this approach further for large scale data analytics will be taken in the European FP7 ATHENA project (Andrews et al, 2013), grant agreement nº FP7-SEC-2012-313220. In ATHENA, Formal Concept Analysis will play an important role in the analysis of social media during crisis situations. It will provide better situational awareness of Member States' Emergency Services. In such situations, where information is likely to be partial and inaccurate, a fault tolerance approach to situational awareness will facilitate faster and more effective responses from Emergency Services. To implement the approach for the large amount of information likely to be involved and in a timely manner for fast response, the In-Close concept mining algorithm developed in CUBIST (Andrews, 2011) will be implemented with a fault tolerance adaptation.

3.3.5.2 **Exploitation of the FCA-based classification method**

Classification of information is a well-known problem and there are many methods to do this, including FCA-based approaches where the clustering of objects according to their attributes leads naturally to a categorisation of objects. Whilst attribute associations add richness to the classification, beyond that provided by tree-based approaches, the normally large number of formal concepts in a data set means that FCA-based approaches can be slow and produce complex results. The development of the high-performance concept mining algorithm In-Close in CUBIST (Andrews, 2011) gives an opportunity to address these issues and this will be exploited by SHU in the development of new, high performance classification techniques and tools. Funding has been secured for this as part of the European FP7 ePOOLICE project (grant agreement nº FP7-SEC-2012-312651). The aim of ePOOLICE is to develop a system to scan the Internet to detect Organised Crime OC). For this a classification system based on indicators of OC is being developed that will harness the new implementation of In-Close that implements the new classification technique. Because indicators may be of a variety of types (categorical, continuous, geospatial, temporal, etc.), a means of classification is required that is broad enough to encompass this variety. To this end, the CUBIST development of FcaBedrock (Andrews et al, 2011) for acquiring and scaling a wide range of data for FCA will be exploited and further enhanced to form a key component in the ePOOLICE system.

3.3.5.3 **References**

Andrews, S., Orphanides, C. and Polovina, S.: Visualising Computational Intelligence through Converting Data into Formal Concepts. Book chapter in Bessis, N. and Xhafa, F. (eds.): Next Generation Data Technologies for Collective Computational Intelligence.

Springer Engineering Book Series, Vol 352, pp. 140-165. Springer-Verlag, 2011. ISBN: 978-3-642-20343-5. DOI: 10.1007/978-3-642-20344-2

Andrews, S., Yates, S., Akhgar, B., Fortune, D.: The ATHENA Project: Using Formal Concept Analysis to Facilitate the Actions of Responders in a Crisis Situation, in Strategic Intelligence Management, pp. 167-180, Elsevier: Butterworth-Heinemann, 2013.

Andrews, S.: In-Close2, a High Performance Formal Concept Miner. In: Andrews, S., Polovina, S., Hill, R. and Akhgar, B. (eds.): Conceptual Structures for Discovering Knowledge - Proceedings of the 19th International Conference on Conceptual Structures (ICCS) 2011. LNAI 6828. pp. 50-62. Springer - Berlin, 2011.

Andrews, S. and McLeod, K.: Gene Co-Expression in Mouse Embryo Tissues. In: Dau, F. (ed.) 1st CUBIST Workshop, at ICCS 2011, Derby, UK. CEUR Workshop Proceedings, Vol. 753, pp. 1-10. ISSN: 1613-0073

Dau, F., An Implementation for Fault Tolerance and Experimental Results, In: Proceedings of the 3rd CUBIST (Combining and Uniting Business Intelligence with Semantic Technologies) Workshop, CEUR Workshop Proceedings, Vol. 1040, pp21-30, ISSN 1613-0073, 2013.

Pensa, R. G., Boulicaut, J-F.: Towards Fault-Tolerant Formal Concept Analysis.
In: Banidini, S., Manzoni, S. (eds.) AI*IA 2005, LNAI 3673, pp. 212{223, Springer-Verlag, Berlin Heidelberg (2005)

3.4 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 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 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.

3.4.1 General Exploitation Strategy

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.

3.4.2 Exploitation Targets

The exploitation targets have not changed since D.2.2, the main motivation for CRSA/MAS laboratory is to transfer the state of the art knowledge in data mining, semantics technologies and visualisation to industry. The key techniques employed by CRSA 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 amounts 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 relational databases, semantic expansion from unstructured data sources and visual querying based on semantic and topological features of the graph.

3.4.3 Recap of Activities as Described in D5.2.1 and D5.2.2

Following the exploitation activities described in D.5.2.1 and D.5.2.2, we 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. The main results are outlined below.

Early Interviews and Prototypes

In the beginning of the first year, we conducted interviews with three CUBIST use cases attempting to find patterns within their data: users conducting market intelligence, bioinformatics, 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. 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 confidence.

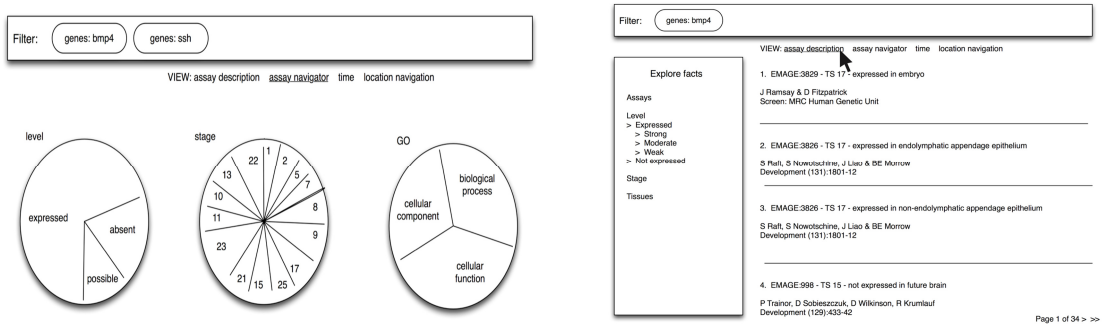


Figure 1. Prototypes designed by the users from the biological use case (HWU).

Cubix: A Visual Analytics Tool for Formal Concept Analysis

A web-based visual analytics tool, called Cubix, was developed to implement the visual analytics techniques investigated for each use case. Typical uses of Cubix include semantic data analysis and pattern detection, dependency analysis, entity implication, 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 was developed and continuously refined with the actively involvement of our three users groups and their use cases.

Association Rules Analytics

A four-month long internship was offered by CRSA focusing 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 *association rules* in Cubix. The visualizations are: A traditional matrix view 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.

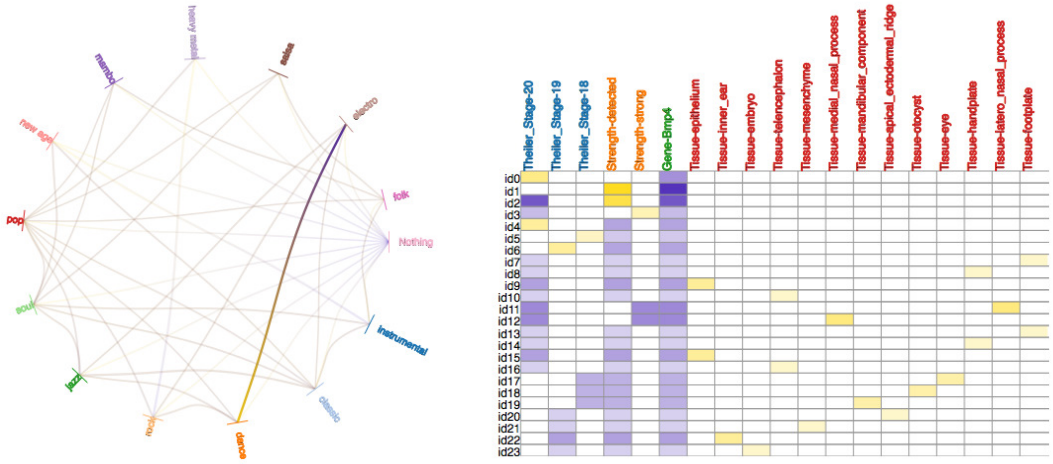


Figure 2. A radial visualisation for the Association Rules. b) Matrix view.

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 was 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[9].

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

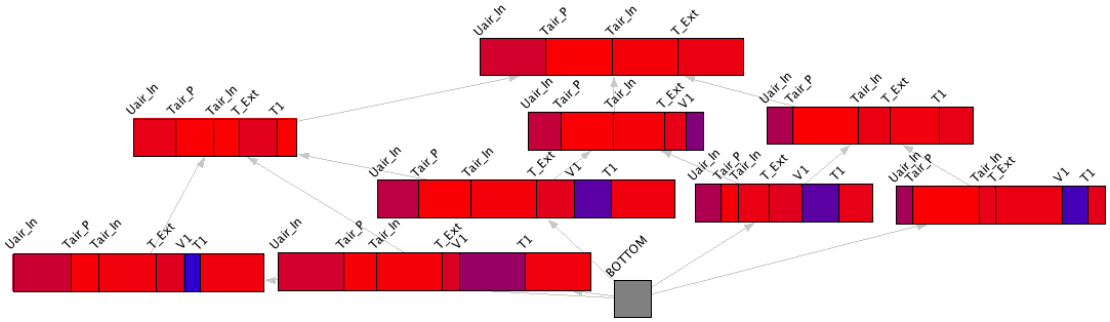


Fig 7: One of the visualisations created for multi-valued concepts. Each box is an attribute in the concept. The box colour indicates the attribute value in the continuum (from blue to red) and the box width represents the length of the interval.

3.4.4 New Exploitation Activities in the Third Year

In the third year, CRSA concentrated efforts on a novel distributed approach for mining formal concepts over data streams. This approach is motivated by the Space Application Services use case: How to predict anomalies from real-time telemetry data? The proposed platform computes and maintains closed itemsets incrementally and can return the current computation in real time on user's request. It is comprised of several components that carry out the computation of concepts from a basic transaction, filter and transform data, store and provide analytic features to visually explore data. Currently, CRSA is carrying out tests with different prediction methods based on formal concepts. The idea is to provide a warning interface able to detect anomalies before they happen.

3.4.5 Further Planned Exploitation after Project End

CRSA will follow up experiments with SAS and HWU concerning the analytic interface. CRSA has also published the visual analytics tool, Cubix, under a permissive open-source license[10]. This way CRSA believes it can attract a community of practitioners and developers that in turn, will lead to a broader adoption. Another candidate project is an open source Javascript library designed to facilitate the development of visual analytics interfaces. This library conveys many techniques, design patterns and solutions to common problems in the visual analytics design.

[10] www.github.com/ksiomelo/cubix

3.5 Heriot-Watt University

Heriot-Watt University (HWU) is a technological university that began as the Edinburgh School of Arts in 1821. HWU was granted its Royal Charter, and thus became a university, in 1966. That was also the year in which HWU became the first university in Scotland to offer a degree in Computer Science. Within the HWU Computer Science department, the BISEL research lab undertakes the work on the CUBIST project using data generated by EMAP[11].

EMAP is the eMouse Atlas Project located within the Biomedical Systems Analysis (BSA) Section at the Medical Research Council Human Genetics Unit (HGU)[12]. 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/abnormal development, thereby gaining important insights into basic biological mechanisms.

3.5.1 General Exploitation Strategy

The BISEL lab at HWU has a long-standing association with EMAP, with Dr Burger having worked for both BISEL and EMAP. BISEL participates in knowledge transfer activities particularly relating to semantic web related technologies, spatial-temporal and distributed biomedical data.

As an academic institution, HWU (BISEL) entered CUBIST with the aim of generating knowledge that could be transferred to EMAP and the wider biomedical world.

3.5.2 Exploitation Targets

As reported last year in D522, following feedback from the potential user community, HWU spent the final year of CUBIST concentrating on:

- The semantic representation of spatial-temporal biomedical images/data; and,
- The visualisation of gene expression query results.

3.5.3 Recap of Activities as Described in D5.2.1 and D5.2.2

Throughout the life of the CUBIST project HWU have been involved in a range of exploitation activities. This section summarises those that occurred within the first two years of the project:

[11] http://www.emouseatlas.org/

[12] The HGU is now part of the Institute of Genetics and Molecular Medicine, which is part of the University of Edinburgh.

- HWU presented the CUBIST project and its results to our partners within EMAP on numerous occasions. Likewise, HWU undertook several presentations within their own academic department.
- HWU wrote a number of research papers (see relevant Dissemination Reports for details).
- Knowledge of semantic spatial representations of images generated during CUBIST was transferred to the EU project Ricordo[13]. Ricordo (now finished) was tasked with delivering recommendations concerning volumetric space to the Virtual Physiological Human[14] (VPH) network of excellence.
- Additionally, knowledge of semantic spatial descriptions was transferred to the International Neuroinformatics Co-ordindating Facility (INCF) taskforce on Digital Brain Atlasing.

3.5.4 New Exploitation Activities in the Third Year

HWU exploited CUBIST in the final year of the project in the following ways:

- Following interest from users of the biomedical resource (EMAGE) at the centre of this use case, HWU further explored the semantic visualisation of gene expression query results. In particular, two existing CUBIX visualisations (sunburst and icicle) were specialised and deployed within a standalone prototype. The aim of this prototype was to learn how valuable such visualisation were and how they could be implemented for maximum usability and usefulness. Evaluation results from within the EMAGE community were positive. Consequently, this work is being tested with other biomedical resources (for example the GenitoUrinary development resource GUDMAP[15]).
- Knowledge of semantic spatial descriptions gained during the CUBIST project was used to write the PHenoImageShare proposal. The project secured funding from the UK's Biotechnology and Biological Sciences Research Council (BBSRC). It is a collaboration between EMAP, HWU and the European Bioinformatics Institute[16] (EBI) that aims to semantically describe mouse phenotype images and develop a mechanism to share those descriptions with the broader community.

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

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

[15] http://www.gudmap.org

[16] http://www.ebi.ac.uk

- HWU continued to publish and present CUBIST at a range of academic venues including a poster at the prime international conference ISMB 2013[17] (see D517 "Dissemination Report" for details), and at an internal HGU seminar.
- Dr Burger continues to brief the INCF Digital Brain Atlasing task force on the discoveries relating to semantic spatial descriptions developed during CUBIST. He attended a Task Force meeting in Stockholm, Sweden, on August 29-31, 2013, and agreed on further collaborations with other INCF members to enhance brain atlas data integration using semantic spatial descriptions. The objective is to include this work in the INCF's Digital Atlasing Infrastructure (DAI) framework and then seek further US and/or EU funding, with the help of INCF, to deploy DAI internationally.

3.5.5 Further Planned Exploitation after Project End

The following exploitation activities are planned for the future:

- The semantic descriptions of biomedical images prototyped in CUBIST are to be developed in the UK funded project PHenoImageShare. The CUBIST draft representations will be evolved further through the inclusion of a broader range of use cases and their differing requirements.
- Dr. Burger is a member of the INCF's taskforce on Digital Brain Atlasing. This is an on-going commitment that will persist beyond the span of CUBIST. The knowledge gained from work undertaken in CUBIST, and subsequently PHenoImageShare, will guide the future development of the taskforce's efforts.

Within the third year of CUBIST, CUBIX visualisations were specialised to the HWU Use Case. Following an expression of interest from both the EMAGE and GUDMAP projects, it is hoped that this work can be developed further in order to generate real world tools for the community.

[17] http://www.iscb.org/ismbeccb2013

3.6 Space Application Services

3.6.1 General Exploitation Strategy

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 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 in 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 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.

3.6.2 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[18] known as User Support Operations Centres. The USOC 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 for CUBIST 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

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

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.

3.6.3 Recap of Activities as Described in D5.2.1 and D5.2.2

During the previous 2 reporting periods SAS did perform a number of activities (Task 8.1 "Requirements Analysis", Task 8.2 "RDF Modelling of Data Sources", and Task 8.3 "Use Case Prototype, v1") that allowed for some early exploitation of results, and/or to 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 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 have 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 created 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.

- 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 with more insight in the possibilities of CUBIST and the further use of it on-console for near-real-time analysis. It is the intention to use CUBIST for helping to identify 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.

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 was 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 was customized to reflect the SAS look and feel with the help of SAP.
- In the third phase, SAS operators working at B.USOC described a few preliminary analytics, which were converted into corresponding SPARQL queries.
- Finally those queries had been integrated into the use case prototype and provided on the web user interface.

3.6.4 New Exploitation Activities in the Third Year

During the 3rd year SAS has worked further on the development of its prototype under task T8.3 in preparation of the exploitation actions targeted. Doing so, SAS has focused on the consolidation of the exploitation capabilities associated to its close collaboration with B.USOC and the prototype work done on the Space use case. The use case results will be mainly targeted towards all other USOCs and Control Center environments.

3.6.5 Further Planned Exploitation after Project End

3.6.5.1 Enhancement of YAMCS

The current Mission Control Systems (MCS) delivered by ESA to USOCs, such as CD-MCS (Columbus Decentralized MCS) are not particularly suited for the operations of payloads. To respond to this need, SAS has developed an open source Mission Control and Archiving System called YAMCS as a set of tools that complements the current MCS. A primary reason for developing YAMCS was the need to support multiple parallel commanding/monitoring chains, as the USOCs typically need to control several payloads. The CD-MCS system does not offer the possibility to easily inspect/retrieve data from an archive and to assess archive completeness. As there are multiple places on the data path from the payload to the USOC where data can be temporarily stored and lost, there was a need for developing an efficient mechanism for detecting and retrieving missing data.

Based on the deployment of YAMCS a set of new needs have been put forward by the user such as the challenge of rapidly processing real time data before it is stored in an archive system, i.e. moving the processing of data closer to the data source. Once the data is stored in an archive, off-line data analysis and processing can be very challenging. For this reason SAS intends to incorporate capabilities offered by visual analytics technology developed in CUBIST for making sense of vast quantities of data. Today humans have major difficulties in identifying, describing and manipulating/processing data, and in perceiving and understanding what it is that the data really means. Operators in any control center, not just in the space domain, face similar problems.

One particular problem appearing in the space domain, probably more than in other domains, is that data is often received out of order and is incomplete, thus any analysis technique must be robust enough to deal with incorrect and incomplete data. The challenge is to enable a human operator or engineer to describe the event in a way that is straightforward and meaningful for him and which allows the archive system to find the event data and all relevant associated information, and present the information in a useful manner.

Although the raw data size typically applicable for mission operations amounts to a few gigabytes per day for a small satellite, all these parameters are practically unique and a combination of any number of them can potentially quickly ramp up. Typical estimations of data volumes for a single spacecraft mission are between 10 TB and 200TB, depending on the effectiveness of the algorithms used to process the data.
Through the integration of dedicated solutions emanating from CUBIST (especially the visualisation tools and FCA solutions), SAS intends to enhance YAMCS to provide a holistic approach to the monitoring, control, data analysis and investigation of the performance of complex systems.

3.6.5.2 Improved Space Weather Environment (SWE) solutions

The study of space weather encompasses a broad range of research domains like solar physics, space sciences, atmospheric sciences, engineering, etc. The interest in space weather is ever increasing, as our modern society is more and more dependent on space-born technologies. Despite the huge efforts done in this field over the last decennium, the complex processes driving the Sun-Earth interaction are still not fully understood.

The effects of Space Weather on several of the modern technology infrastructures on which we rely together with their impact on humans in air and space have been the subject of world-wide scientific and technical consideration over the last decade. These effects cover a broad range of sectors such as radio communication problems, effects on synthetic aperture radar systems, GPS and the future European Galileo systems, aircraft crew and passenger radiation risks, power grids, etc.

Present day research has demonstrated that:

- Adverse Space Weather poses a non-negligible threat to humans and modern technological systems and assets on the ground, in the air and in space.
- Methods to model some aspects of space weather have been developed, although their performance needs to be improved.
- Prediction of the behaviour of various Space Weather related physical parameters is possible and has in some cases indeed been achieved. However, for many of these physical parameters the prediction accuracy is insufficient to allow the transition into reliable operational services.

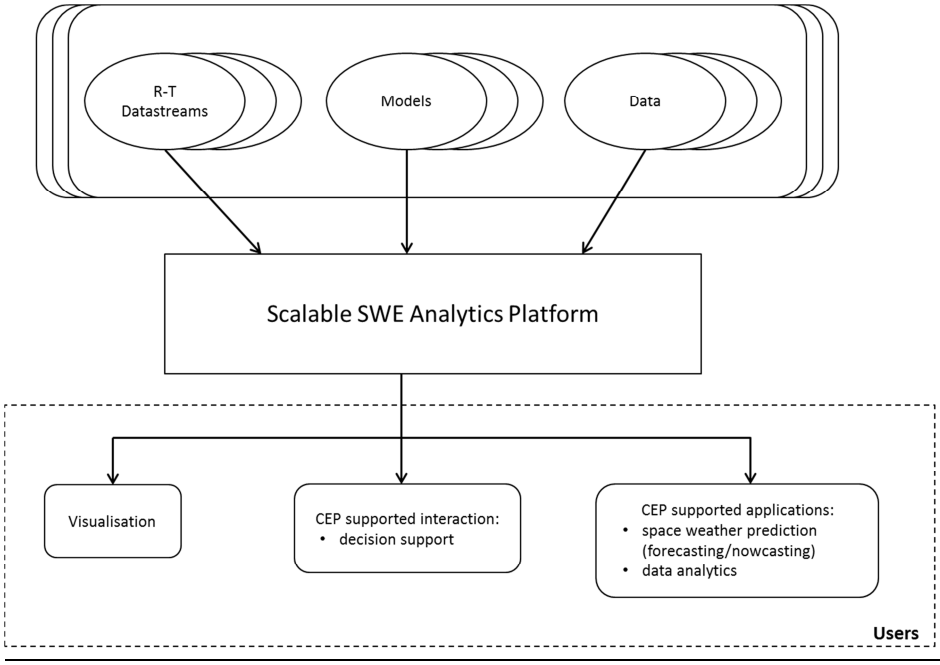
In order to monitor the space weather an armada of spacecraft and ground based stations are providing relevant information. These data are essential for obtaining a better insight in the drivers and effects of space. Relevant data include solar images in different wavelengths, measurements of the solar electromagnetic spectrum in different wavelengths (X-ray, EUV, UV, visible, radio), in-situ solar wind data, geo-magnetic data, galactic cosmic ray data, proton and electron flux, etc. Besides the data from currently operational instruments, there exists also a large archive of historical data of past solar and atmospheric missions (e.g., Helio, Ulysses). A primary objective of the space weather community is to set up operational tools for nowcasting and forecasting purposes to support several applications affected by space weather. Therefore, much of the research efforts are systematically spent in the development and/or the improvement of advanced space weather modelling techniques/solutions that are suitable for operational implementation and progressively enhance the prediction capabilities for the specific application areas.
Spacecraft operators have a need for the situational awareness of space weather so that they understand what is happening to their spacecraft and can quickly analyse anomalies. For this purpose they require long-term access to data and to models of spacecraft environment and effects in order to correlate spacecraft effects with environmental conditions. A future need

will be to extend the spectrum and range of data available for such their needs, and to further improve the data analysis techniques through use of advanced data analysis and processing techniques.

Given the end-user stakeholders requirements and the current activities with respect to the improvement of the current space weather prediction models there is a strong need for:

- Innovative data collection, storage and representation methods.
- Metadata and data ontologies for the federation of a highly increased number of diverse and large (distributed) data sources, where some of these data sources are populated through dynamic feeds of extremely large numbers of data streams generating thousands of events per seconds (hence trillions of records).
- Development of ontology based (semantic) virtual observatories to support the ingestion of relevant knowledge models to facilitate knowledge handling and knowledge discovery on data sets.
- Advanced real-time interaction, visualisation, data analysis and decision support techniques/solutions to correlate anomaly data related to the respective application domains, extending the availability and use of data sources of anomaly recordings.
- Development of operational global data assimilation schemes using coupled space weather models and large amounts of diverse data.

Relying on its current involvement in SWE activities, SAS intends to respond to the above stakeholder needs through the set up of a Scalable Data Analytics platform (see below diagram) based on Complex Event Processing (CEP) techniques that would further incorporate a set of the solutions coming out of CUBIST for the visualisation and the data analytics.



3.6.5.3 Long-Term Data Preservation (LTDP)

As part of its involvement in the past FP7 Space project ULISSE (USOCs Knowledge Integration and Dissemination for Space Science Experimentation) SAS has developed a rich client application (ScienceCast) that serves as a single, integrated interface for describing, publishing and accessing space data experiments. While the current implementation of ScienceCast has made space experiment data more reusable by making it more accessible and better understood, although it did not address the need for long-term data preservation in the space science domain required to improve the efficiency, reduction of errors and the capture of relevant information to increase the scientific return of payloads. SAS intends to investigate the applicability of the CUBIST results and where appropriate integrate these to develop enhanced LTDP solutions for the space market.

3.6.5.4 Smart Cities / Urban Management solutions

Urban centers depend on a large number of urban infrastructure systems and services, such as healthcare, transportation, utilities, city administration and public safety and security. Efficiently managing, controlling and maintaining the urban infrastructure associated to the fast growing urbanization is a huge challenge for European cities and cities worldwide.

To address this challenge, novel technology for next generation control centers for urban management have to be implemented. Next generation control centers can improve monitoring, decision making and coordination of actions by all services involved. The benefits are multiple, ranging from improved environmental monitoring and sustainable city development, over optimized resource allocation and traffic optimization, to better management of critical events. In the very near future, lack of data is no longer the bottleneck for smart city applications. To the contrary, the amount of available data will quickly surpass the limits of what can be handled with the current technical and administrative infrastructure of cities. The following core concepts need to be integrated for the development of these advanced control centers for urban management:

- Intelligent systems for interactive data analysis.
- Efficient and natural interaction with technology.
- Open platform for data delivery and processing.

The data delivery addresses the following features:

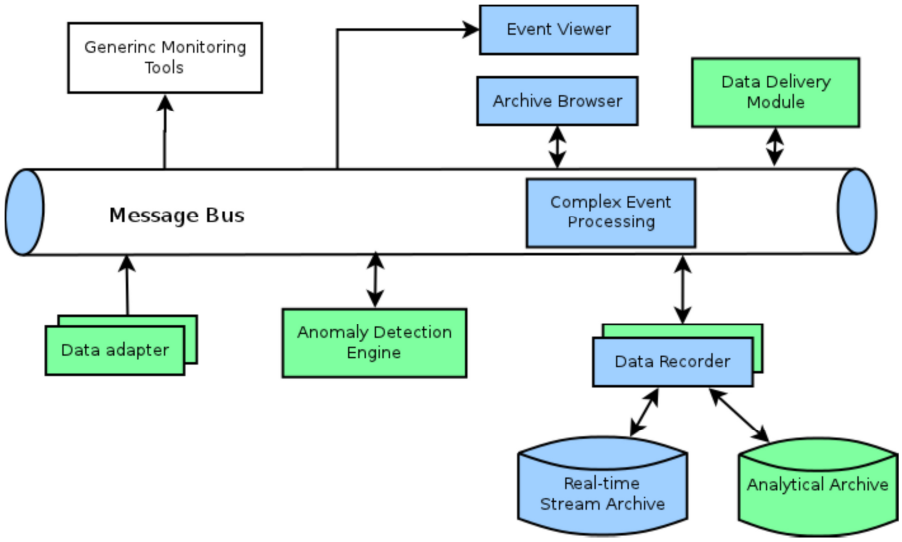
- Processing of heterogeneous data (both structured and unstructured) comprising data of different nature, such as sensor readings and video streams, but also the increasingly important data from the social Web.
- Supporting participatory sensing where citizens provide data, e.g., using mobile phones or social Web tools.

- Automating the processing of data by applying techniques such as complex event processing (CEP).

With respect to the data delivery it will be required to address the following challenges:

- Providing intuitive data analysis tools and visualizations, also addressing mobile operators in the field.
- Supporting the work processes in control centers, including collaborative work processes.

SAS intends to develop a platform for the data delivery and processing as specified below.



To achieve this SAS intends to incorporate a number of CUBIST results relating to visualisation and data analytics into the above platform.

-

3.7 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.

3.7.1 General Exploitation Strategy

This data underpins Innovatage's main product Insight, a web based portal that significantly improves the performance of recruitment agencies, reducing the time taking to locate leads, the number of leads and the quality of leads. Innovantage has also created a data warehouse of over 36 million vacancy advertisements since 1st January 2009, this unique data allows Innovantage to do things such as publish labour mark reports and generate bespoke reports on behalf of clients. These reports are manually created and publish.

3.7.2 Exploitation Targets

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

- **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**. During the life of the project Innovantage have been (separate from CUBIST) 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 was hoped that the semantic business intelligence capabilities developed during the CUBIST project would 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.

3.7.3 Recap of Activities as Described in D5.2.1 and D5.2.2

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.

3.7.4 New Exploitation Activities in the Third Year

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.

3.7.5 Further Planned Exploitation after Project End

Innovantage will continue to work with Cubist, using it along aside out current analytic product.

4 Summary

This deliverable described the exploitation deliverables in CUBIST, focusing on activities in the third year and planned further exploitation. The exploitation activities are passed on two pillars:

- One the one hand, the overall CUBIST prototype is planned to be published as open source under the Apache 2.0 License. Two of three partner contributions (SHU and ECP) have already been published their components, the SAP-component undergoes currently an SAP-internal approval process. Releasing the prototype to the public aims at the evangelism of core technologies (ST- technology and FCA-technology) used in CUBIST for BI applications, and will meet the interest expressed by different scientific communities. The CUBIST partners do not believe that they will lose competitive advantage.
- On the other hand, we have reported different partner-wise exploitation activities, both in commercial (SAP, ONTO, SAS) or research- and teaching-oriented (SHU, ECP, HWU) environments. These activities will continue after project end.