SpaceOps Conferences
5-9 May 2014, Pasadena, CA
SpaceOps 2014 Conference

10.2514/6.2014-1867

CUBIST: Implementation and Evaluation of a Semantic Business Intelligence System for Payload Operations

S. Klaï[1], E. Sevinç[2], C. Jacobs[3] and B. Fontaine[4]
Space Applications Services, Belgium

and

C. Muller[5] and D. Moreau[6]
B.USOC, Belgium

The Solar Monitoring Observatory, or in short SOLAR, was among the first Columbus payloads switched on after Columbus installation in February 2008, and continues to perform science until today. The Belgian User Support and Operations Centre (B.USOC, Brussels), responsible for the support of the SOLAR operations, uses, like every mission control centre, heterogeneous sources of information, including structured and unstructured data, for decision making and information tracking. Very large volumes of data are obtained, especially with the SOLAR telemetry data which are generated every second over long periods of time.Today, during real-time operations, the SOLAR operator constructs a mental model of the current operational status. Unfortunately, no software can provide a bird's eye view of the operations, nor combine in a unified application the most important operational information. In an anomalous situation, the first actions consist of bringing all the data together, such as experiment telemetry, user and operations manuals, console logs, configuration status, experiment execution planning, etc. Thus, a lot of time and effort is spent on the retrieval of data for real-time and post-analysis information, prior to the actual analysis. The Combining and Uniting Business Intelligence with Semantic Technologies (CUBIST) project envisions combining the two worlds of Business Intelligence (BI) and Semantic Technologies. The objective of CUBIST is to aggregate various information sources available to operators in mission control rooms using technologies based on semantic web standards. Aggregated data, ready for the BI processing, are expected to provide online support - via an online, web based unified interface - for making better decisions, reveal hitherto undiscovered information and provide supportive evidence in debriefing and decision making processes related to the organisation of space control centre operations.

I. Introduction

The Combining and Uniting Business Intelligence with Semantic Technologies project (CUBIST) was a European Union funded research project, under the 7th Framework Program for the topic Intelligence Information Management. This research project which ran from September 2010 until September 2013, with the final review meeting in November 2013. It followed a best-of-breed approach combining capabilities of Business Intelligence (BI), Semantic Web Technologies and Visual Analytics creating a new type of platform entitled "Semantic Business Intelligence" (SBI). As listed on the project's website (www.cubist-project.eu) CUBIST's main objectives are to:

[1] B.USOC Operator, Space Applications Services, Saliha.klai@spaceapplications.com.
[2] Software Engineer, Space Applications Services, Emre.Sevinc@spaceapplications.com.
[3] B.USOC Operator, Space Applications Services, Carla.Jacobs@spaceapplications.com.
[4] Software Systems Group Manager, Space Applications Services, Bernard.Fontaine@spaceapplications.com.
[5] B.USOC Knowledge Management Responsible, B.USOC, Christian.Muller@busoc.be.
[6] B.USOC Manager, B.USOC, Didier.Moreau@busoc.be.

- support federation of data from unstructured and structured sources,
- persist the federated data in a Semantic Data Warehouse; a hybrid approach based on a BI enabled triple store,
- and provide novel ways of applying visual analytics in which meaningful diagrammatic representations of the data will be used for depicting the data, navigating through the data and for visually querying the data.

The CUBIST objective was also to allow aggregation from various information sources available. This aggregated data, ready for the BI processing, are then expected to provide online support, via a unified interface, for taking better decisions, reveal hitherto undiscovered information and provide supportive evidence in debriefing and decision making processes to the three use cases. The resulting technology stack was demonstrated in three use cases from the fields of market intelligence, computational biology and control centre operations.

The CUBIST project was led by SAP AG (Germany), the technology partners were Ontotext AD (Bulgaria), Sheffield Hallam University (United Kingdom) and Ecole Central de Paris (France). The Use Case partners were Innovantage Systems Ltd (United Kingdom), Heriot-Watt University (United Kingdom) and Space Applications Services (Belgium), representing the domains of market intelligence, computational biology and control centre operations respectively. It should be noted that an integrated approach was applied throughout the project; combining experience and expertise in development and solution management, and the involvement of a user base represented by the consortium use cases were integral parts of this approach and was served as validation points.

This paper focuses on the prototype developed by the consortium for the the use case of the control centre operations, or more specifically the SOLAR payload operations. At first a generic prototype was provided, based on the input of the use cases. At the end of the three year project a SOLAR dedicated prototype was derived from this generic prototype based on the operator's feedback and focusing on the CUBIST features of key importance for this use case. The project concluded with an evaluation fo the CUBIST SOLAR Prototype, using the think-aloud method which was complemented with an interview.

In the next section the CUBIST technology and project will be described. In Section II the SOLAR Payload is presented, including the operations concept as well as the processes currently used on console. Section IV discusses the CUBIST SOLAR dedicated prototype, and the user evaluation and result are presented. The conclusion and further dissemination of CUBIST are summarized in the concluding section V.

II. CUBIST Introduction

A. CUBIST Technologies[1] and approach

To extract information from the heterogeneous sources of data a process of enhanced Extraction, Transfer and Load (ETL) will be applied in CUBIST, the so-called Semantic ETL (SETL). Especially for the space operations, large volumes of data are obtained. In order to safeguard the performance and scalability, CUBIST will employ a Resource Description Framework (RDF) triple store and ontology as the backbone for the CUBIST information warehouse. This will reduce the complexity of the integration of heterogeneous data sources by enabling the linking of the facts that will have been federated from structured and unstructured data sources. A layer within the warehouse will integrate the triple store with Formal Concept Analysis (FCA)-based visual analytics. Formal Concept analysis is a mathematical theory of data analysis using formal contexts and concept lattices. Each concept in the hierarchy represents the set of objects sharing the same values for a certain set of properties, so-called attributes; and each sub-concept in the hierarchy contains a subset of the objects in the concepts above it. The FCA concept will be used for the analytics of the data, allowing user friendly visualization of the analysis.

A schematic of the general CUBIST architecture is presented in Figure 1.

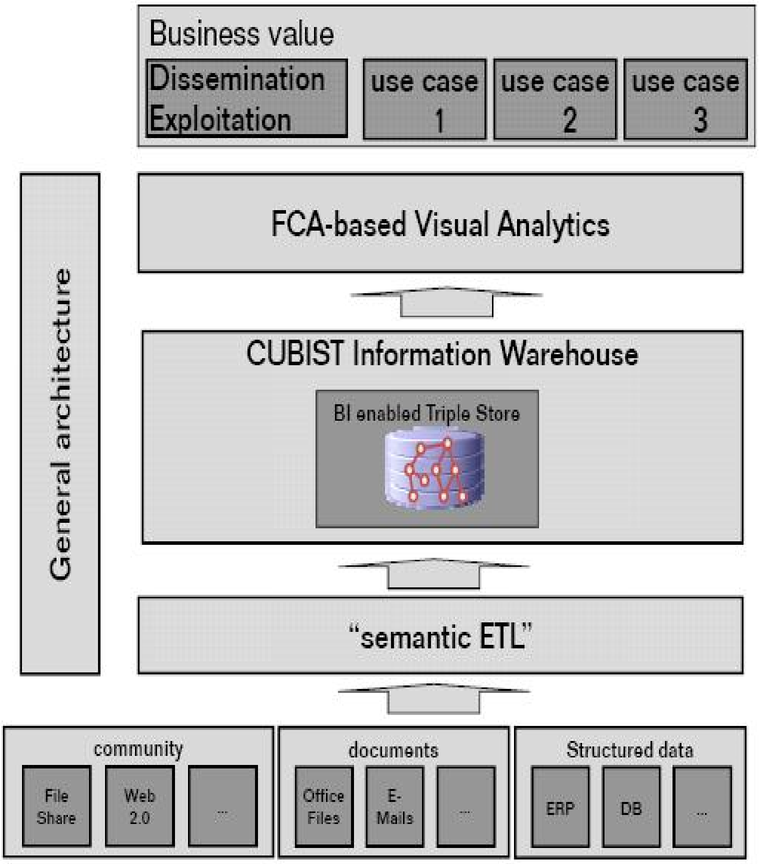

Figure 1: Overview of the CUBIST General Architecture *(source www.cubist-project.eu)*

B. CUBIST for Space Control Centers

As a use case, the involvement of Space Applications Services, with the support of the Belgian User Support and operations Centre (B.USOC), in the CUBIST project was mostly from the end users position, representing the space control centres and mainly focusing on the SOLAR Payload operations. The work had two main objectives:

- To provide structured and unstructured data from the SOLAR Payload operations as an input to the CUBIST system
- To evaluate the benefits of the CUBIST system in the context of the SOLAR operations.
-

Figure provides an overview of the activities with respect to these objectives:

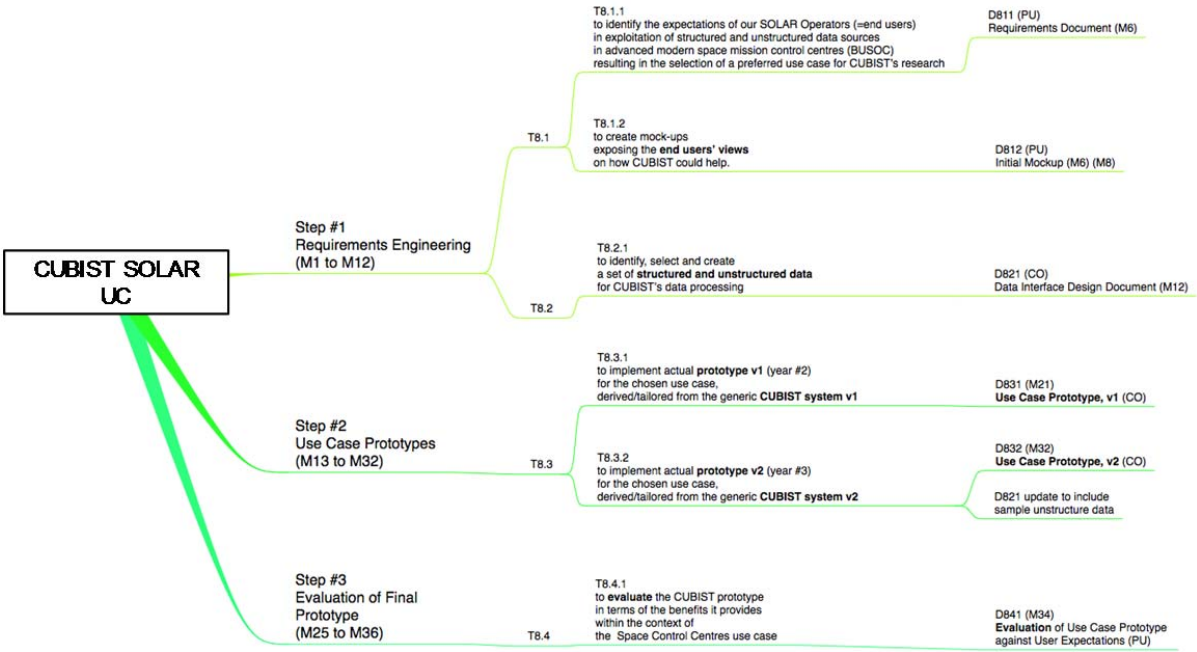

Figure 2: CUBIST SOLAR Operations Use Case

As detailed in the above figure, first a requirements analysis was performed, which was followed by the development of mockups dedicated for the SOLAR Payload Operations Use case[1]. This was to be used as input for the technological partners for the development of the first prototype.

Next to that, the further deployment of the prototype required a representative datapack from the SOLAR operations, for which a detailed description for the Sematic ETL was necessary. Finally, after a first feedback on the generic prototype, a second prototype was developed and evaluated against the initial requirements[2] again.

Obviously, additional support and contribution was provided to the other, more technical facets of the project, such as data integration and the development of the technological architecture.

III. The SOLAR Payload

SOLAR (Figure 3) is an integrated platform accommodating three instruments, each designed to observe the Sun in a specific wavelength range. The SOLAR mission has the aim to provide accurate measurements of the Solar Spectral Irradiance, covering a wide spectral range from the Infra-Red (IR) to the Extreme Ultra-Violet (EUV).

The scientific objectives of the SOLAR mission cover different fields of research, such as solar physics, atmospheric physics, and climatology[3,4]. The SOLAR instruments are mounted on a Coarse Pointing Device (CPD) providing Sun pointing and tracking capabilities. The SOLAR Control Unit (CU) provides power, collects, formats and dispatches to ground the data generated by the instruments in telemetry packets. It receives ISS anciliary data and processes the ground issued

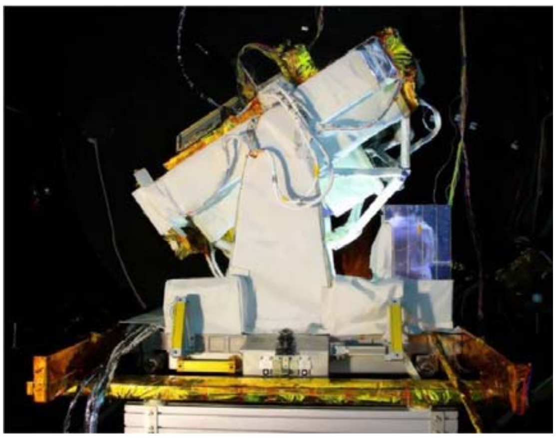

Figure 3 : Side view of the SOLAR Payload

telecommands. SOLAR was launched together with the European Columbus laboratory in February 2008 and has been operating since that time. Since the start of the SOLAR Mission the Belgian User Support Centre (B.USOC) ensures a 24/7 support for the SOLAR operations. B.USOC is located on the premises of the Belgian Institute for Space Aeronomy in Uccle, Belgium.

The B.USOC SOLAR operators on console are responsible for the control and the command of SOLAR according to the science requests of the scientists[5,6]. Payload downstream data, or SOLAR telemetry, from the ISS allows real-time monitoring of SOLAR by the operator. These so-called housekeeping data provide the overall status of the payload, consisting of SOLAR temperatures, instrument temperatures, voltages, current consumption, operational status, execution reports, etc.

To successfully perform the SOLAR operations, B.USOC runs a set of hardware and software tools and utilities common to all USOCs. Moreover, in the light of this being a long-term mission, an additional set of tools and applications has been developed by B.USOC to support the SOLAR specific operations[6,7,8]. These are used to execute activities, to support the on-console operations and the planning processes and to report information.

The following provides a brief overview of the structured and unstructured data sources available to the operator on console[5]:

1. Unstructured data sources:
    a. Columbus Operations Support Tools, providing access to the a number of tools required for operations such as console logs, electronic flight note system, onboard and ground procedures for executing activities, interface procedures.
    b. Systen Problem Report database and Anomaly Reporting Tool, the online issue tracking system for ground and onboard issues
    c. Local bugs database: a database running Bugzilla used for the B.USOC operators to enter bug reports to the software running inside B.USOC.
    d. Documentation tools such as eRoom, Alfresco Documentation management System and the BUSOC wiki allowing relatively fast retrieval of documents, user manuals, ground procedures, tips and tricks, etc.
2. Structured data sources:
    a. Payload Telemetry consisting of floating point values, strings, binary reading or integers. These telemetry packets are generated and sent to ground every second, where each telemetry packet contains 343 parameters.
    b. The ON Board Short Term Plan providing an overall view of the ISS timeline including relavent timing indication of external events such as the day-night cycle, ISS attitude changes, South Atlantic Anomaly fly-over and other elemets related to ISS operations.

Today, for on-going operations, the SOLAR operator continuously constructs a mental model of the current operational status. Unfortunately, no software provides a bird's eye view of the operations, nor combines the most important operational information in a single screen. Especially in an anomalous situation, the first action is to bring all the data together: mostly the experiment telemetry, but also background information such as user and operations manuals, console logs, configuration status, experiment execution planning, etc providing the full picture of the occurrence. Together with the payload developer, the operator makes a first assessment and proposes, if possible, a way forward to resume science as soon as possible.

In an anomalous situation, yet again a lot of time and effort is spent on the retrieval of data for real-time and post-analysis information, prior to the actual analysis. During such an event, the operator strives to find a kind of signature of the anomaly in order to have the required supportive evidence that the anomaly is a reoccurrence or a new event. This data mining is today done through actual replays of the telemetry or retrieval of specific parameters list of a short period around the occurrence.

Currently, post analysis of anomalies are performed rather manually. With the present data sources and tools, to retrieve the list of occurences is done through searching through the actual textual anomaly reports. With the time tags listed in those, the retrieval of parameters or replay can be performed, for each of these occurrences listed in the anomaly reporting tool. Furthermore the post analysis in requirements fulfillements or identification of correlations between sets of or single parameters is enormously time consuming.

IV. CUBIST for SOLAR operations

A. CUBIST SOLAR Prototype

The CUBIST Prototype for the SOLAR operations has been derived and tailored from the second enhanced version of the general, hence use-case independent, CUBIST prototype.

In general, the CUBIST prototype aimed at satisfying different kind of information needs:

1. **Factual Search:** Finding and exploring specific information, like a single fact or a list of objects which satisfy specific, information-need-relevant, criteria.
2. **Explorative Search:** Explore facts and entities when there is no precise information need, but the need to inform oneself about some facets of the domain of discourse
3. **Visual Analytics:** (Visually and interactively) analyze aggregated information to get a more condensed view on (subsets) of the data.

As briefly mentioned in Section III, SOLAR has been operational for more than six years, sending one telemetry packet every second or so. Over a year, this represents approximately $3x10^7$ packets. Each telemetry packet contains 343 parameters, of which 44 parameters do not change at all or very rarely. Among the others, 135 have binary readings, such as ON and OFF. Others have readings that span between 3 and 2 $x10^6$ distinct values. The data released to CUBIST consortium partners for the development of the CUBIST prototype covers 30 days between September 26, 2008 and October 25, 2008[9]. This datapack resulted in approximately 500million RDF triples. It should be noted that the prototype was initially build on the structured data with the intention to extend this with the unstructured data. Unfortunately, due to lack of resources and based on the initial use case feedback, this has not been included in the final prototype

At the end of the second year of the project the first Use case prototype was evaluated by an experienced SOLAR operator who also had previous end-user experience with FCA. The resulting report stressed some desired enhancements, mostly in the area of usability and responsiveness and including the proposal of some additional features. On the other hand, there were other general features in the prototype which were of limited value to the SOLAR operations. The main reason lied in the fact that the semantics for the structured data is fairly simple and not well suited for the SOLAR data, hence, all features related to that did not give any added asset to the SOLAR operator.

Following provides an overview of the key features of the SOLAR Operations CUBIST prototype:

1. Search and Select

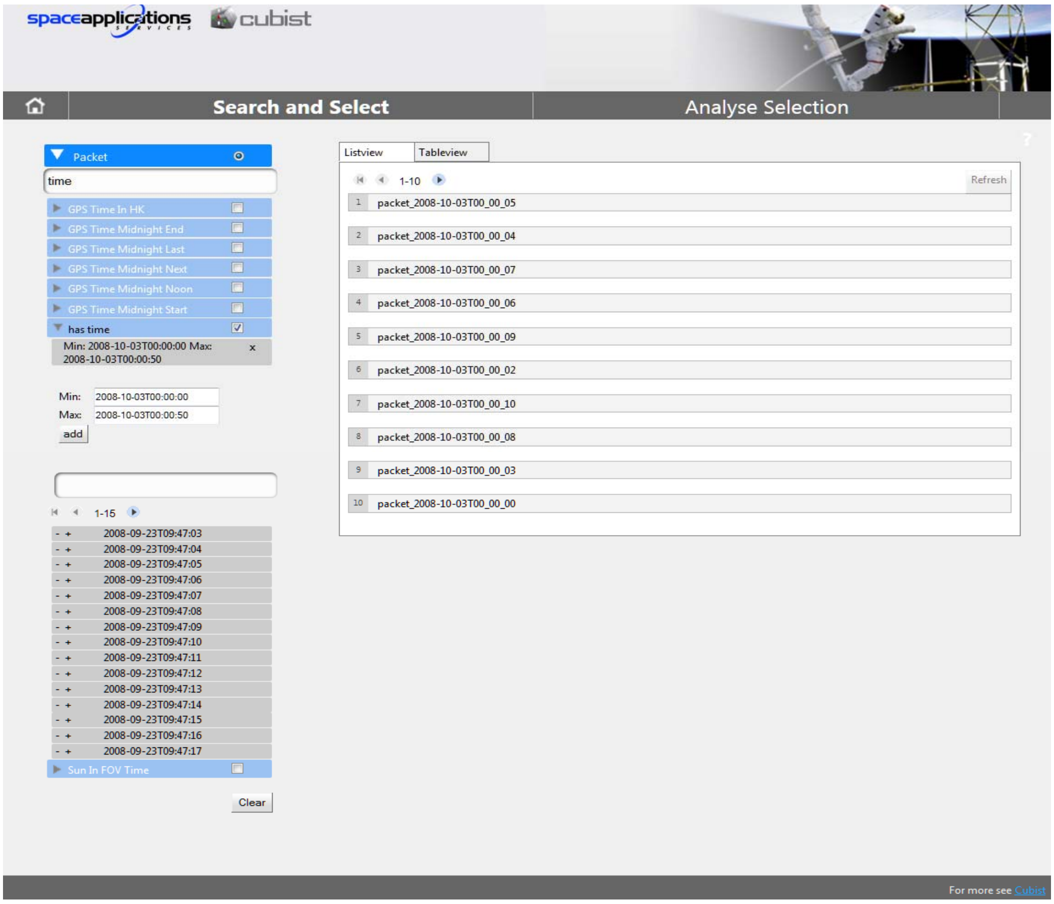

Figure 4: Search and Select panel

The search and Select panel is the starting point of the CUBIST prototype User Interface (see Figure 4). It allows the operator to select the required parameters for further analysis. The auto-complete facilitates the browsing through the parameters, by typing part of the string, the list of parameters will be limited to those having that specific string in the name. Moreover the “Search and Select” does not only provide the option to select the parameters, but also to select specific values or a range within the values. The RDF-caching provides faster visualization and will show for pre-programmed strings such as ON/OFF, the possible values. For such parameters, the operator can then choose to be only interested in specific cases. Moreover, for floating point values or integers, there is the possibility to select only those parameters within a certain range by providing a minimum and maximum. A special parameter has been added for the SOLAR case which is the time parameter, hence, allowing the operator to look for specific events in an indicated period of time. The list view provides the listing of the parameters with time stamp that matches the selection.

2. Scaling Panel

The prototype provide the ability to scale the dataset which was selected. By browsing from the Search and Select to the “Analyse Selection” Panel the operator is first guided to the Scaling panel.

Especially for the analysis with parameters consisting of floating point values scaling is a critical aspect before proceeding with the visual analytics. In order to render the next step of analysis and visualization, the data often needs to be put in a more discrete form; grouped into bins so that it becomes a manageable and understandable visualization.

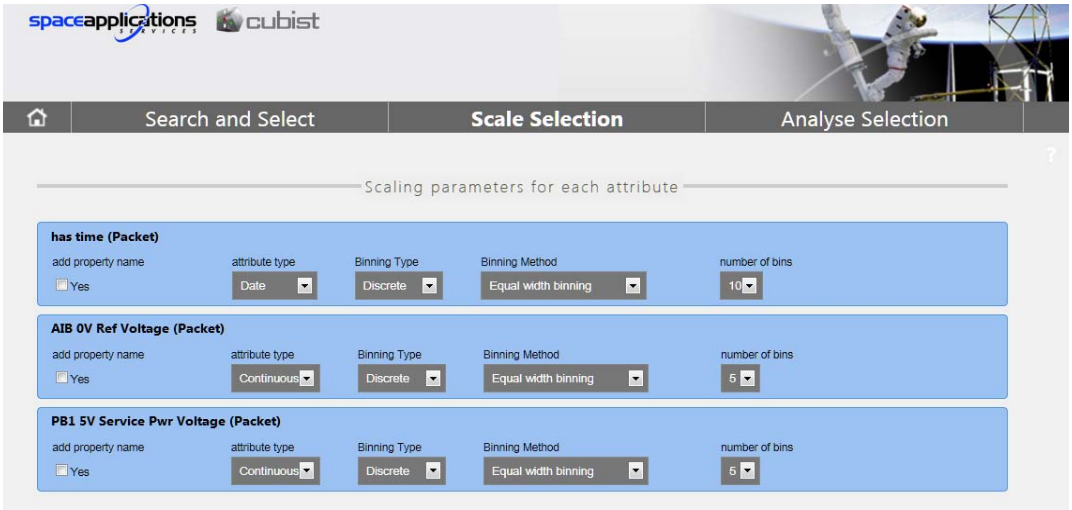

Figure 5: Scale Selection Panel

The feature either allows automated binning by indication of the number of bins, or manual bins. For time parameters specifically there is also the option to group the parameters per bins of 10seconds or as indicated by the operator. The scaling panel is shown in Figure 5.

3. Analysis

In CUBIST a graphical workbench called Cubix provides the visual analytics. Cubix offers a range of visualisations methods including Hasse diagrams, co-occurrence bar chart and matrix view(see Figure 6).

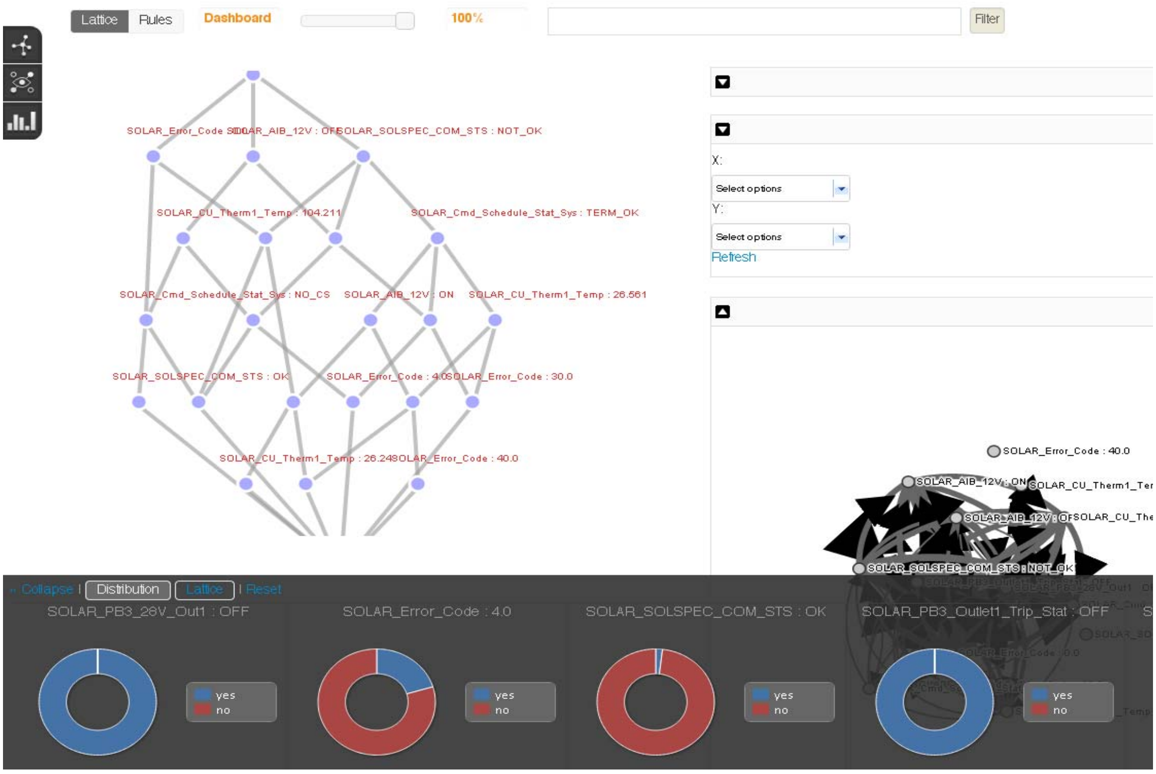

Figure 6: CUBIST Analysis Vizualisation -- CUBIX

In Figure 6, the middle graph represents the FCA visualization which can be changed following the operators need. By default it generates the Hasse diagram lattice, besides this one, the user can choose for other FCA based visualizations such as Matrix vizualisation, Sankey, Tree or Sunburst. Furthermore, the appearance of the diagram can be customized through the different features such as the Metrics and the filter. By applying those, the diagram can for example either show correlations comparison between events by the resizing the connection lines or the number of events by the resizing the nodes. One can also directly zoom in or out in the lattice or filter the lattice to a partial set of the selected data based on the filtering pane below the diagram.

Besides the FCA visualizations, some typical Business Intelligence features are consultable on the right hand side of the user interface.

B. User Evaluation

1. User Evaluation Conduct

For all use cases within the CUBIST project the main goals were:

1. Providing an assessment of the of system's functionality extent. In particular, it has to be investigated whether CUBIST meets the requirements of the use case.
2. Answering the question on how the CUBIST user interface affects the user. It has to be

investigated how efficient and effective users can deal with CUBIST, as well as the user's attitude towards the system.

3. Identify specific problems of the prototype. Such problems might be unexpected results, and causes of user confusion, and negative aspects of the design of the user interface.

The evaluation from this perspective should be focused on rather specific questions and tasks.

The evaluation method was investigated within the consortium. It was clear that it should cover both qualitative as quantitive measurements and focusing on high-level-information. The given constraints such as available resources, the fact that the evaluation was conducted at the end of the project and the geographical spead of the users in the use cases, were also taken into account. Therefore it was concluded to utilize a combination of the following methods:

- the task-based thinking aloud method, this will inform both about low-level-task accomplishment in CUBIST as well as about high-level benefits of CUBIST,
- interviews, which will deepen the high-level evaluation of CUBIST, and
- questionnaires, which will serve for collecting quantitative measures, both for lowlevel and high-level information in the evaluation.

For the SOLAR use case two operators were selected to conduct the user evaluation. Both had a several years of experience in the SOLAR operations. Only one of the test persons had been involved in the CUBIST project, hence, had some background knowledge of the CUBIST technologies. While the other operator received a tutorial video (www.cubist-project.eu) as well as an introduction in FCA.

The evaluation sessions consisted of the execution predefined tasks related to analysis of the SOLAR data. The sessions were conducted through a teleconferencing system allowing the technological partners to follow remotely. During this session, which also included recording the video and voice, the think-aloud method was employed: the test users described what they were trying to accomplish and what she was currently thinking. The observer did not providew comments or interpretations at this stage.

Following the completion of predefined tasks using the CUBIST prototype, an interview followed immediately. The interview was also fully recorded. Predefined open ended questions were answered by the test user and this completed the major part of the evaluation. A short questionnaire was also sent to the test user to be filled in. Transcribing the interview and collecting the answers to the questionnaire, whose results can be seen in the following sections, completed the evaluation.

2. User Evaluation result

From both users some key points were raised. The first appreciation was the unified interface, unfortunately the systems performance could be improved, especially in an operations environment.

The Search and Select component was very positively evaluated, it was praised for its user friendliness and easy searching capabilities for specific events. This data mining feature would be of great value within the operations environment. This would enable the possibility for automatic searches for identical or similar events in the past based on a combination of different types of input such as

- payload parameters: e.g. specific temperature behavior, specific errors issued by SOLAR

- payload activities: a specific activity for one of the instruments or the platform

The Scale Selection component is extremely usefull for the SOLAR operations as mostly it is only in a specific range of a parameter that the interest lies. It was noted that the automated binning is useful for beginners, but for more expert users more granular binning should be considered. This additional feature is currently not available in the CUBIST prototype.

The evaluation of the Analyse Selection component is two-fold. The visualization only becomes useful once the operator is familiar with FCA and its different visualizations. It is clear that in-depth training and clear understanding in FCA is a mandatory investment to be able to maximize the return of the visualization. Especially Hasse diagrams with filtering tools, in terms of FCA analysis, and co-occurrence matrix (classic BI) are considered very valuable tools for error analysis by the operators .

Generally, compared with the traditional tools currently used on console, a system like the CUBIST prototype, could facilitate a significant number of typical analysis tasks performed today. It allows for a fast and accurate data mining, and indicates correlations, if any, in the searched dataset. This in turn permits an accurate analysis of anomalies and failures. There where classic BI provides an overall analaysis of the data, the CUBIST technology allows looking for those events with minimal support but high confidence, which are the typical signature of unexpected behavior.

V. Conclusion

At the end of the project the technological partners decided to share the experience and skill gained by the consortium by releasing the overall CUBIST protype as open source on GiHub under the Apache 2.0 License (www.cubist-project.eu). The prototype consists of three components developed by the partners SAP, SHU and ECP which are published independently. At Space Applications Services, further exploitation of the prototype is certainly intended. At first the focus lies on the consolidation of the explotation capabilities to the close collaboration with B.USOC and the prototype work done on the SOLAR Payload operations. As explained in ref. 7[7,] a set of tools complementing the existing Mission Control System has been developed, which is named YAMCS. Through the integration of dedicated solutions emanating from CUBIST, especially the visualization tools and FCA soluations, it is intended to enhance YAMCS to provide a holistic approach to the monitoring, control, data analysis and investigation of the performance of complex systems, such as payload operations.

VI. Acknowledgments

This work has been carried out within the CUBIST project. CUBIST is funded by the European Commission under the 7th Framework Programme of ICT, topic 4.3: Intelligent Information Management. The authors would like to thank the CUBIST consortium for the good work and progress in the CUBIST project.

The authors would also like to thank the ESA USOC Management, the B.USOC Management and the Space Applications Services Management for their close cooperation and for providing the budget and personnel for the B.USOC.

VII. References

[1]Klaï S., Sevnic E., Fontaine B., Jacobs C. and Muller C., "CUBIST: Semantic Business Intelligence Supporting Payload Operations", *Proceedings of SpaceOps 2012,* AIAA 2012-1279934, Stockholm, Sweden, June 2012

[2]S. Klaï, S. A. Mikhailian A., Ursik M., Initial Mockup (Space Control Centers), Issue 1.2, CUBIST-SA-D812, June 2011

[3]Thuillier, G., Fr ohlich, C., and Schmidtke, G., Spectral and Total Solar Irradiance Measurements on Board the International Space Station," Utilisation of the International Space station 2, edited by A. Wilson, Vol. 433 of ESA Special Publication, 1999, p. 605.

[4]Schmidtke, G., Fr ohlich, C., and Thuillier, G., ISS-SOLAR: Total (TSI) and spectral (SSI) irradiance measurements," *Advances in Space Research*, Vol. 37, 2006, pp. 255-264.

[5]Brantschen S., De Smet L, Michel A, SOLAR Payload Operations: Achieving Flexibility to Support a Long term Science Mission, *Proceeding of SpaceOps 2010,* AIAA 2010-1951, Huntsville April 2010.

[6]Diaz A., Klaï S., Jacobs C., Van Hoof D., Sela A., Karl A., Michel A., This N., Moreau D., SOLAR Predictor: A Knowledge Management Tool Supoprting Long-Term Console Operations, *Proceedings of SpaceOps 2014,* May 2014

[7]Sela, A, N. Mihalache, D. Moreau, YaMCS-Mission Control Software, *Proceeding of SpaceOps 2012,* AIAA 2012-280790, Stockholm, Sweden, June 2012

[8]Sela, A, Michel A., The SOLAR Mission Tool – an integrated ops tool, *Proceeding of SpaceOps 2012,* AIAA 2012-280958, Stockholm, Sweden, June 2012

[9]A. Mikhailian, J-M. Wislez, B. Fontaine, S. Klai Requirements Document (space Control Centers), Issue 2.9.1, CUBIST-SA-D811, April 2011