CUBIST: Semantic Business Intelligence Supporting Payload Operations

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

and

C. Muller[5]
B.USOC, Belgium

Today, in space control centre operations, much time is spent on transferring, discussing, reviewing, and copying information between the operations partners. The search and replay of operational data for the correct analysis of the on-board situation is also time-consuming. This problem is especially emphasized by the fact that many of the data stores are distributed and have different user interfaces. The current operations show an increasing need for a system providing the operator with a unified interface for fast access to information and analytics, especially during anomalous situations where time constraints are stronger. The FP7 project for Combining and Uniting Business Intelligence with Semantic Technologies (CUBIST) envisions combining Business Intelligence and Semantic technologies to aggregate various sources available to the operators, providing support in analysis and decision making on console. The CUBIST concept will be applied on the data resources of the SOLAR payload, a long-term ESA mission on the International Space Station. CUBIST allows to quickly analyze complex data patterns and to extract essential and relevant information, which is crucial in real-time space operations.

I. Introduction

The Combining and Uniting Business Intelligence with Semantic Technologies project (CUBIST) is a European Union funded research project, under the 7th Framework Program for the topic Intelligence Information Management. It follows 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:

- 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.

CUBIST will allow aggregation from various information sources available. Aggregated data, ready for the BI processing, are 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 will be demonstrated in three use cases from the fields of market intelligence, computational biology and control centre operations.

---

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

The CUBIST project is led by SAP, the technology partners are Ontotext, Sheffield Hallam University and Ecole Central de Paris. The Use Case partners are Innovantage Cardiff, Heriot-Watt University and Space Applications Services, representing the domains of market intelligence, computational biology and control centre operations respectively.

This paper focuses on the application of the CUBIST concept on space operations. As a use case the SOLAR payload is presented. SOLAR is a long-term ESA mission on-board the International Space Station (ISS), operational for more than four years. The mission operator makes use of 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 that are generated every second over large periods of time. Efficient data mining becomes a key point to guarantee the success of the mission and to provide qualitative operations.

In the next section the CUBIST technology will be described in some more detail. In Section III the SOLAR payload is presented and a more detailed description of the different operations tools and data sources is given. The following section describes how the SOLAR use case is implemented in CUBIST, and in Section V a specific event analysis is presented. The conclusions of this research are summarized in a concluding section VI.

II. CUBIST Technologies

Figure 1 provides an overview of the CUBIST architecture:

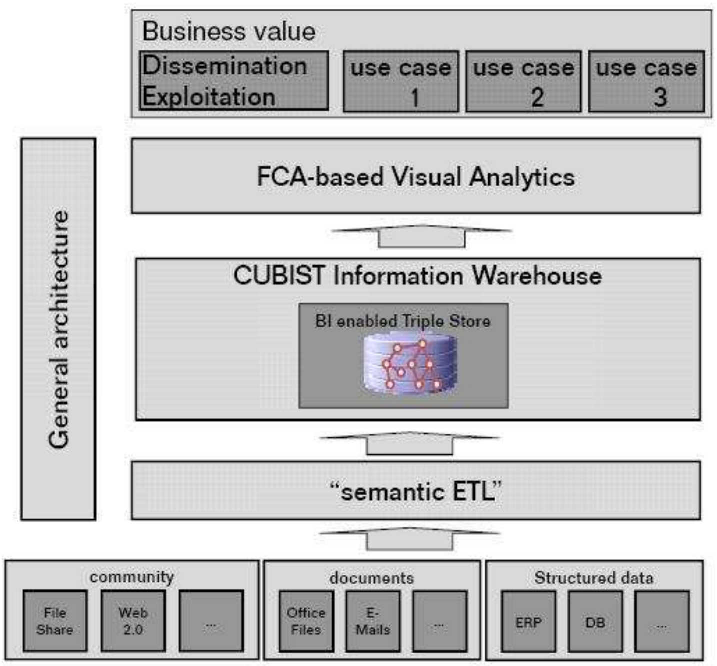

Figure 1: CUBIST General Architecture *(www.cubist-project.eu)*

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). The ETL as known in the classical BI only addresses structured sources such as relational database tables and spreadsheets. The SETL, on the other hand, will need to handle a variety of unstructured data sources such as operations manuals, user logs, daily reports, and web pages that are stored using various formats and standards. One of the main challenges in extracting information from the unstructured data sources is the use of unconstrained natural language, hence the requirement of semantic web and natural language processing technologies. It is the objective of CUBIST to provide comprehensive methods in bringing unstructured data into analytics.

Especially for the space operations, large volumes of data are obtained. In order to safeguard the performance and scalability, CUBIST will employ an 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.

III. SOLAR Payload operations

A. SOLAR Payload

SOLAR (see Fig. 2) is an integrated platform accommodating three instruments complementing each other to allow measurements of the solar spectral irradiance throughout a large part of the electromagnetic spectrum, ranging from the extreme ultraviolet to the near infrared.[1,2]. The SOLAR instruments are mounted on a Coarse Pointing Device (CPD) for Sun pointing and make use of the CPD Common Control and Power Distribution Unit (CU) to get power, to collect, packet and dispatch to ground the

Instruments generated telemetry data and to receive the ground issued telecommands and ISS data. SOLAR was launched together with the European Columbus laboratory in February 2008 and has been operating since that moment.

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 in the premises of the Belgian Institute for Space Aeronomy.

Figure 2: SOLAR Payload *(courtesy ESA)*

B. SOLAR payload operations tools

The B.USOC operators on console are responsible for the control and the command of SOLAR according to the scientist's request. Payload downstream data, or SOLAR telemetry, from the ISS allows real-time monitoring of SOLAR by the operator. This 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. Based on the input from the scientists, B.USOC executes science measurements through telecommanding. This telecommanding consists of structured upstream data which are to be sent in coordination with the Columbus Control Centre (Col-CC). They may contain control structures for shutting down or starting up various modules, as well as uploads of data and scripts.

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. These are to execute activities, to support the on-console operations and the planning processes and to report information. The following provides an overview of the structured and unstructured data sources available to the operator on console.

1. *Structured data sources for SOLAR operations:*

1) Payload Telemetry: the payload telemetry offers the operator an insight in the status of the SOLAR payload. These SOLAR housekeeping data consist of a variety of the SOLAR parameters, such as temperatures, power supply, instrument housekeeping data, pointing device telemetry, etc. These are not only floating point values, but also string, binary readings or integers. SOLAR has been operational for four

years, sending a telemetry packet practically every second, where each telemetry packet contains 343 parameters. For payload data processing two main tools are used:

- The Columbus Decentralized Monitoring and Control Subsytem (CDMCS) is used, whereas another tool, the Unified Synoptics Software (USS) displays real-time payload data. Each of the payload parameters is described in the Mission database, which contains a machine-readable description of the telemetry, including the size of various parameters in telemetry packets and their interpretation from binary to engineering values. This allows the operator on console to identify values out of nominal range.
- For replay or parameter retrieval a software package YaMCS is used (Yet another Mission Control System). This lightweight Mission Control System was developed by Space Applications Services to support long-term mission operations and extends the CD-MCS capabilities. Although the CD-MCS does contain an archiving system, it does not offer the possibility to easily inspect or retrieve data from the archive.[3]

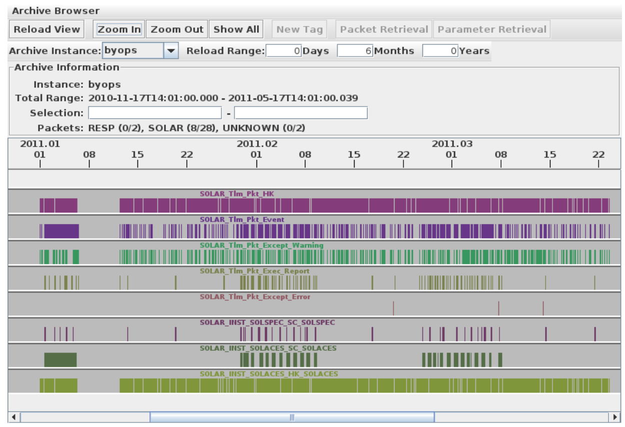

Figure 3: YaMCS GUI showing different SOLAR packages

2) The On Board Short Term Plan (OSTPV) View provides an overall view of the ISS timeline for the upcoming two weeks. It is a common tool for both crew and all ground support personnel and depicts the schedule for all ground and on-board activities, duration of activities, on-board procedures to be executed, stowage notes, day-night cycle, ISS attitude changes, South Atlantic Anomaly fly-over and other elements related to the ISS operations.

2. *Unstructured data sources for SOLAR Operations:*

1) Columbus Support Tools: this is a set of secured web-based tools developed by the Industrial operations Team (IOT) to support and track the on-going Columbus operations.
It allows access to
    - The International Procedure Viewer (IPV) for managing and displaying the database of onboard and ground procedures (Operations Data Files)
    - Columbus Electronic Flight Note (CEFN): a tool for written communication, review, processing and tracking. It is utilized for all coordination of written communications among the operations teams.
    - Console Logs: short time-stamped brief narrative of events kept by each console position
    - Joint Operations Interface Procedures (JOIP): outlines these operational interfaces and guidelines; it documents the real-time operational interfaces and procedures used to exchange information.
2) System Problem Report database and Anomaly reporting Tool: online issue tracking system for ground and onboard issues about Columbus, its payloads and control centres.
3) Payload Operations Data File (PODF) providing a pre-defined sequence of commands and control checks in order to operate the payload. For European payloads in Columbus, all operations, in nominal situation, are done by the on console position or crew member following a PODF.
4) The Daily Operations Report (DOR) which provides a summary of the operations that day and is provided to all stakeholders
5) SOLAR Predictor Tool: an in-house developed web-based tool serving to generate DORs, implement configuration control and support the short-and long term planning[4].
6) Local bugs database: a database running on Bugzilla used for the B.USOC operators to enter bug reports to the software running inside B.USOC

7) Documentation tools such as eRoom and B.USOC wiki allowing fast retrieval of documents, user manuals, tips and tricks

With the current large amount of data collection, located in different stores and not correlated, practice shows that a lot of time and effort is spent on the retrieval of data prior to the actual analysis, which can have an impact on the science measurements.

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: not only 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. These recovery actions require some kind of verification and validation via existing procedures (PODFs), agreements with engineering or supporting payload documentation.

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.

IV. Research area for CUBIST

1. General overview

The main tools for the on console operator, monitoring the SOLAR payload, are the USS Displays, showing the SOLAR real-time telemetry or housekeeping data (see Figure 4 for an example).

As a requirement from the use case this was identified as the entry point into the CUBIST system for the support of SOLAR operations. The CUBIST system will be visualized by a web-based main page, allowing the operator to quickly search through the available data of which some can already be predefined in an earlier stage. In the light of the use of FCA, for SOLAR operations, all parameters become attributes, while states are objects. To define a dataset and create the so-called formal context for the analytics, the operator can easily flag or unflag attributes and identify the formalization for each parameter: for example indicate the formalization for temperatures below a certain value or above.

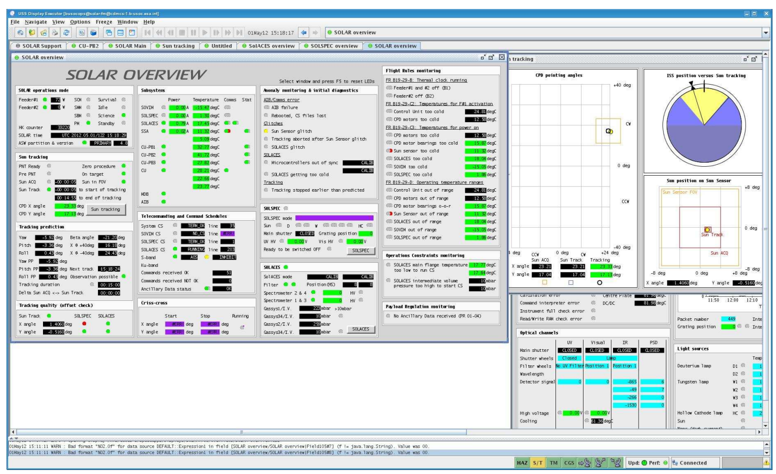

Figure 4: SOLAR USS Displays

After the selection of the population, the lattice visualization can be shown to the operator in different ways. All visualizations will allow manipulation:

- Zoom and easy search options in the lattice
- Adding or removing attributes
- Narrowing the visualization to a certain node, viewing a single event of relations
- Visualization through scaling

2. Preliminary design examples

The screenshots below show some important aspects of the preliminary designs of the CUBIST graphical user interface, in which some analytics are defined and then run on the telemetry data[5,6]:

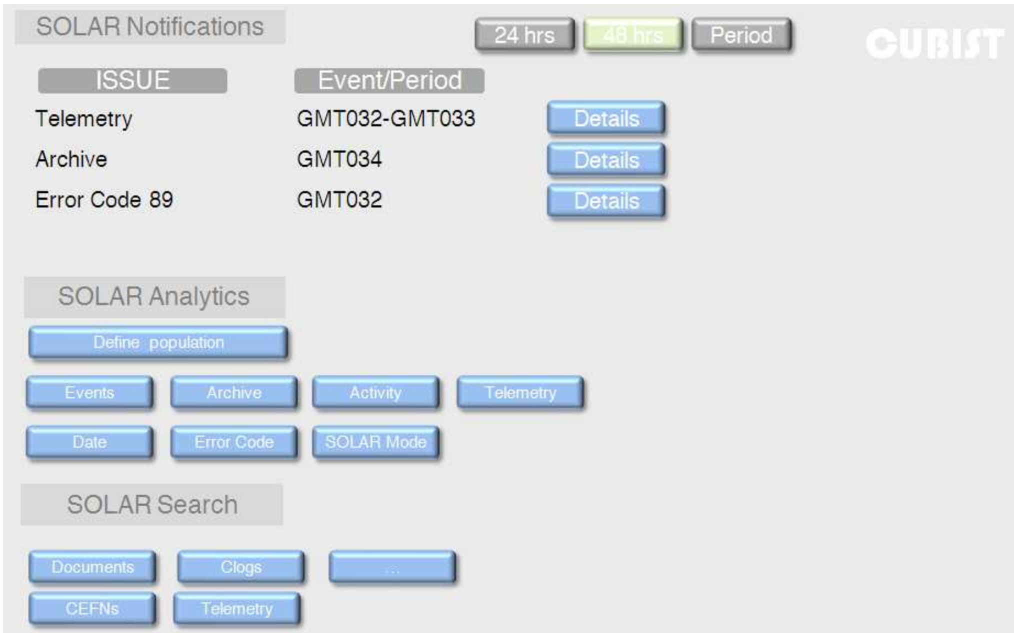

Figure 5: Example CUBIST main page as per user requirements

Figure 5 shows the main page of CUBIST, which is a web-based application. It not only provides the user with a simple interface to various analytics, but also hosts an easy access to quickly querying the unstructured data sources, such as Console Logs, CEFN, etc.

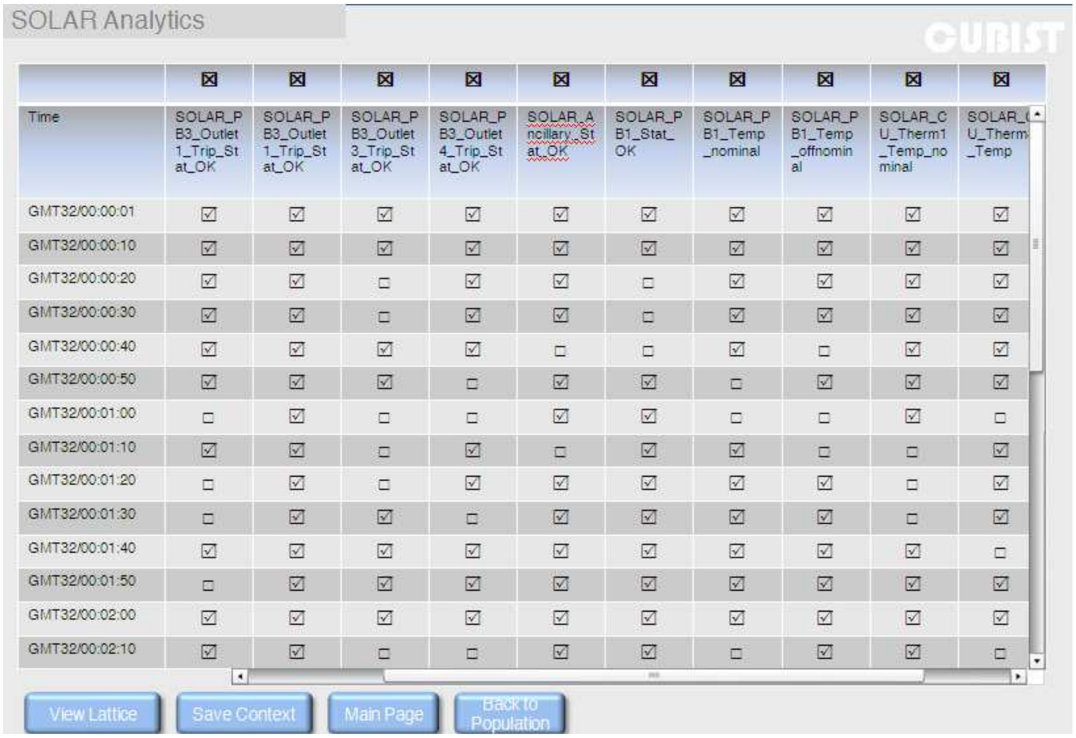

Figure 6: CUBIST population selection

The formal context overview presented in Figures 6 allows the operator to further manage it by removing attributes (or SOLAR parameters) which does not seem relevant for the analysis. The user will also be able to restrict the selection of objects having specific attributes. In case the user has defined the population, the CUBIST system will allow saving this selection as a predefined selection and adding it to the main page. This predefined set will not include the restrictions set initially, but it will remember the grouping and selected parameters and will be included in the list on the CUBIST Main page, if requested by the user.

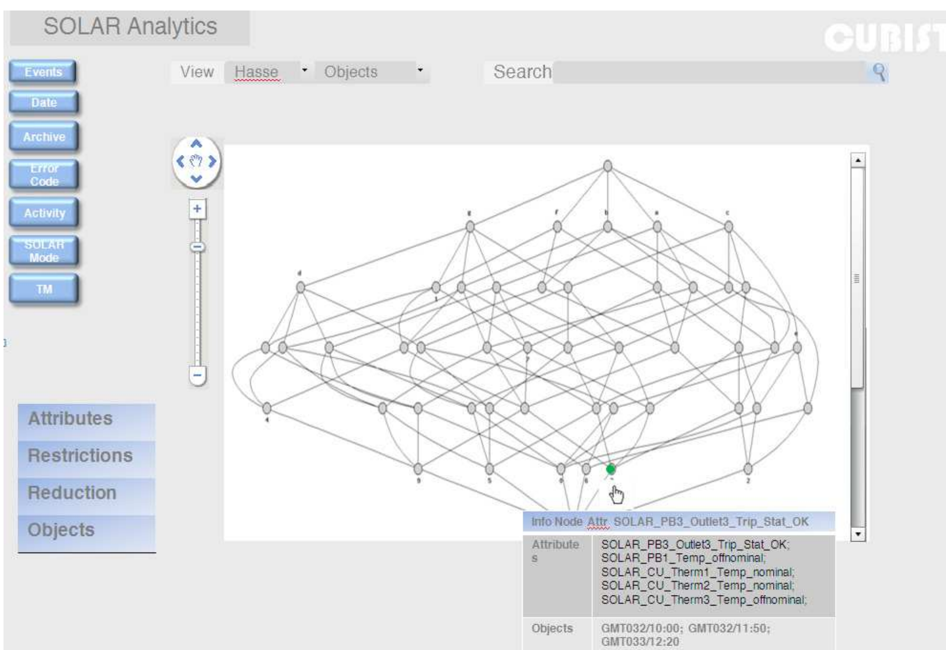

Figure 7: CUBIST lattice creation

Finally, Figure 7 shows the visualization aspects of Formal Concept Analysis, as a Hasse diagram. The CUBIST system lets the operator to zoom in or out of the diagram, filter out some nodes based on selections and easily see the details of a node in the same screen, simply by clicking on the node.

The following subsections provide an overview of typical scenarios where the SOLAR operations will benefit from the CUBIST system.

A. Data mining

Throughout the years, for many different reasons, such as training or anomaly analysis, data mining has been an important aspect of the operator's job. It currently consists of the manual replay or retrieval of parameters. The operator searches through console logs, DORs and activity schedules, to look for similar situations in the past and then manually retrieves or replays the data. The CUBIST project would provide a system enabling 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
- external factors: beta angle range, thruster events, South Atlantic Anomaly, etc.

One of the advantages of CUBIST will be the flexibility of conducting such searches in a unified data store that relies on a semantic RDF data store; without such unification, the amount of development and implementation effort to have a similar affect generally proves to be unfeasible.

B. Anomaly analysis

Especially during long term missions, anomaly re-occurrence is bound to happen. The SOLAR payload experiences certain anomalies, of which the exact cause is not yet known. As mentioned in the previous sections, anomaly resolution is currently done through data mining and literal replays of telemetry. The CUBIST project will correlate the occurrences and might reveal certain patterns with the SOLAR payload or even with factors outside the payload, retrieved from the other data sources. This will then result in workarounds or early indication of the upcoming failure, permitting preventive action to avoid the anomaly or to quickly recover from it mitigating the loss of science.

C. Forensic analysis

Some failures that occurred in the past have had a significant impact on the planned science measurements. To date, it is still not known whether some parameters values, or even a combination of different parameters, are a precursor for the failure. The objective of CUBIST is to find patterns and propose them to the operator for further analysis with the end goal to have a prediction of future failures.

V. Formal Concept Analysis on SOLAR

For the use case of Space Control Centers, B.USOC has provided the CUBIST consortium with an extraction of non- restricted data sources limited in time. This extraction consisted of a set of housekeeping data as well as an example of each of the unstructured data sources.

For a first trial and in the light of familiarization with the data for the CUBIST technological partners on the one hand and FCA for the operators on the other, a limited set of SOLAR parameters was ingested in the FCA Bedrock[7] tool. This tool was developed by one of the consortium partners, SHU with the specific aim to convert large data sets, such as those in SOLAR operations, into formal contexts. With the project being in the early phase, the dataset was limited to a certain number of events, more specific typical parameters were provided from a failure where the Analogue Interface Board suddenly switches off. This is a known anomaly for SOLAR, which requires a power cycle of the payload to resume science.

For the CUBIST exercise, 5 seconds from each failure was taken with the initial status of SOLAR and the issued errors as attributes. Figure 7 shows the visualization of the formal context created for those events.

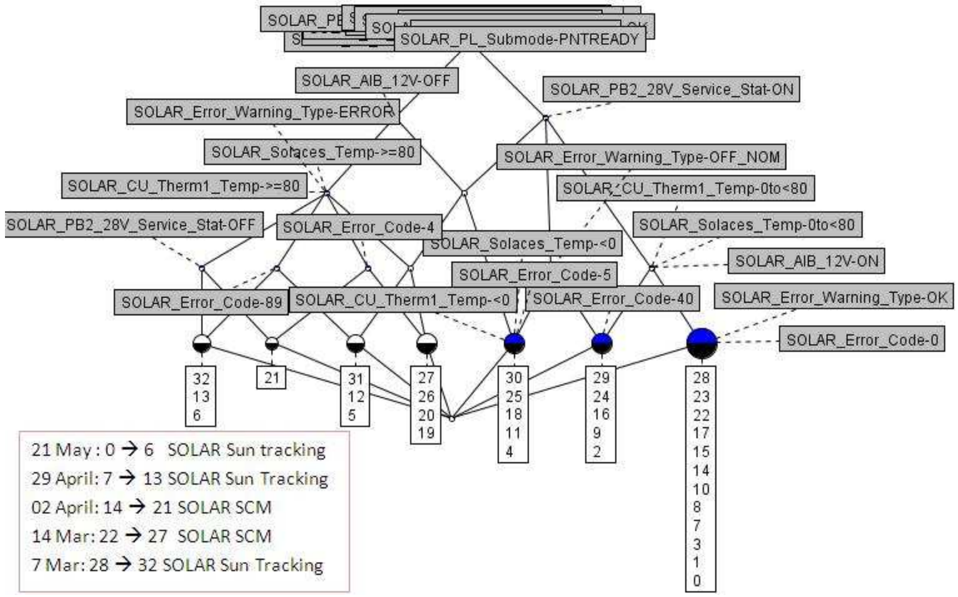

Figure 8: AIB Failure via FCA BedRock Tool

The objects appear as numbers, from 0 to 32, and they correspond with one line of the parameter extraction, and thus one second of the failure. Line 0 to line 6 is the occurrence of 21 May, line 7 to 13 is the occurrence of 29 April and so on. All occurrences show two identical initial errors, namely SOLAR Error Code 0 and the SOLAR Error Code 40. After these initial errors, the following errors are different depending on the initial situation as indicated in the legend of Figure 7.

From an operator point of view, there is a clear correlation between the initial status of SOLAR and how SOLAR reacted to the anomaly. When SOLAR was in an initial status of Sun Tracking, different errors were issued than when SOLAR was in the stable Set-up and Configuration mode, providing more insight in the actual anomaly. The FCA lattice visualization immediately indicates this pattern, while manual replay did not allow putting the sequence of events next to each other in such a way.

VI. Conclusion

In this paper the CUBIST project has been presented. This new approach in BI yields a quick and detailed analysis of complex data. As a proof of concept, the CUBIST ideology has been applied on spacecraft operations. The ESA SOLAR payload is a long-term mission on-board the ISS and is operated by the B.USOC. Especially for long-term mission in a complex environment as the ISS, an overwhelming amount of structured and unstructured data is collected over the years. A project as CUBIST allows for a fast and accurate data mining and allows to link specific events with each other in an organized way. This in its turn permits an accurate analysis of anomalies and failures. As an operator it is important to recognize precursors of possible failures, and to handle quickly and precise in order to guarantee the science output. Although CUBIST is a research project, the experience and skills the consortium will gain from the project will allow the further development of a promising tool for space operations.

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.

References

[1]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.

[2]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.

[3]Sela, A, YaMCS-Mission Control Software, *Proeeding of SpaceOps 2012*

[4]Sela, A, SOLAR Mission Tools, *Proeeding of SpaceOps 2012*

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

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

[7]S. Andrews, C. Orphanides, FcaBedrock, A Formal; Context Creator, *Proceedings of ICCS 2010*, Kuching, Malaysia. LNAI 6208, Springer-Verlag (2010)