Creating Topic Maps Ontologies for Space Experiments

David Damen, Rani Pinchuk, and Bernard Fontaine

Space Applications Services, Leuvensesteenweg 325, B-1932 Zaventem, Belgium
{david.damen,rani.pinchuk,bernard.fontaine}@spaceapplications.com
http://www.spaceapplications.com

Abstract. This paper outlines the design process that is used to create Topic Maps Ontologies for space experiments in a wide range of scientific disciplines. This design process is implemented in three iterations, one for creating initial Topic Maps Ontologies and two for further refinement. The paper focuses on the first iteration that consists of nine workshops with various providers of space experiment data, each active in a different scientific domain. Furthermore, we report our findings in holding these workshops.

1 Introduction

In Europe, experiments under micro-gravity in the International Space Station are coordinated and executed by User Support and Operation Centers (USOCs). The different USOCs are spread out geographically over Europe and each of them is specialized in one or more scientific disciplines, such as physiology, biology, or fluid science.

Early 2009, an FP7 project called ULISSE (**U**SOCs Know**L**edge **I**ntegration and Dissemination for **S**pace **S**cience and **E**xploration) started [1]. The main aims of ULISSE are:

1. To improve on the return of space experiments through **better valorization, dissemination and exploitation** of the **experiment data**.
2. To create **opportunities for multidisciplinary research** by allowing data from different scientific domains to be correlated.

In order to maximize the possibilities for attaining these goals, the information and data about these space experiments will be described in topic maps. More specifically, the main aim of Topic Maps in ULISSE is to capture information about each experiment's setup, execution and results. To help in understanding this experiment information, the topic maps authored will also describe (relevant parts of) the scientific disciplines of these experiments. As a first step in creating these topic maps, ontologies[1] have to be designed to provide the necessary structure for populating the space experiment topic maps later on.

[1] By ontology we refer to Topic Maps Ontology. A topic map is built from topics that represent specific subjects of the domain of knowledge that the topic map describes.

Maicher, L.; Garshol, L. M. (Eds.): Linked Topic Maps. Fifth International Conference on Topic Maps Research and Applications, TMRA 2009 Leipzig, Germany, November 12–13, 2009 Revised Selected Papers. Leipziger Beiträge zur Informatik. ISBN 978-3-941608-06-1

Table 1. Scientific disciplines covered by the data providers

Scientific discipline	Data provider	Location
Cell Biology	ETH Zurich	Switzerland
Fluid Science	MARS	Italy
Fluid Science	E-USOC	Spain
Material Science	MUSC	Germany
Physiology	CADMOS	France
Physiology	DAMEC	Denmark
Plant Physiology	N-USOC	Norway
Solar Physics	B-USOC	Belgium
Space Health & Medicine, Bed Rest studies	MEDES	France
Space Plasma	SRC-PAS	Poland
Technology	Erasmus-USOC	Netherlands

1.1 Why Topic Maps?

Data providers consist of a large amount of scientists who are not well versed in knowledge management or software engineering, let alone ontologies or UML. Practically, only some of the data providers had previously heard of or worked with ontologies. Because Topic Maps demands for less formalism and training than other techniques, such as OWL or UML, the Topic Maps technology was deemed more accessible and hence likely to be used for communication with our intended audience.

1.2 Who is Involved?

Eleven data providers participate in the project: ten USOCs and one Research Institute. The scientific disciplines covered are listed in Table 1. Space Applications Services took the lead in defining and formalizing the Topic Maps Ontologies in cooperation with these data providers.

2 Topic Maps Ontology Design Process

2.1 Overview of the Design Process

We have adapted the methodology described in [2] for creating Topic Maps Ontologies for space experiments. Four phases have been identified that make up the design process:

1. Preparation
2. Analysis
3. Draft
4. Refinement

In addition, the topic map contains topics that represent the different types, classes, association types, occurrence types and scopes. This second group of topics can be seen as the structure, or the skeleton, of the topic map. This structure is called, in the Topic Maps community, ontology.

Preparation In the preparation phase, each data provider identifies the space experiments he wants to include. He also gathers relevant documentation and data samples, and finds out who the domain experts in his organization are. During this preparation step, Topic Maps is explained to the data provider, as well as the proposed design process.

Analysis. In the analysis phase, the data that is available is analyzed in order to gain an understanding of the space experiments of a data provider and of the scientific discipline he operates in. A list of typical questions is identified which the topic maps authored should be able to answer. The questions will be useful later in the refinement phase.

An example question is “Which carrier was used for experiment X?” To be able to answer this question from a populated topic map about experiment X, the ontology for this topic map will need topic types for *Experiment* and *Carrier*, and an association *performed-on* between them.

Draft In the draft phase, the actual work that leads to an initial Topic Maps Ontology for the space experiments of the data provider begins. This is done in three steps:

1. **Enumerate** the relevant terms: A brainstorming session is organized with the data provider about his space experiments and about his scientific discipline. The focus is on getting a list of words.
2. **Categorize** the relevant terms: topic types, occurrence types, association types or instances are identified among the terms in the word list.
3. **Organize** the relevant terms: The categorized terms are then organized into an initial Topic Maps Ontology that serves as an input to the refinement phase.

Refinement In the refinement phase, time is foreseen to look further into the initial Topic Maps Ontology. Any mistakes are fixed and confusing parts are cleared up by modifying the structure of the Topic Maps Ontology. In this phase the Topic Maps Ontology also undergoes a careful evaluation. This includes checking that the ontology complies with general constraints, e.g. do all association types have role types. In addition, it is checked that the Topic Maps Ontology provides enough structure such that populating it will allow us to answer the questions identified during the analysis phase.

2.2 Implementation of the Design Process

Overview The previously described design process had to be applied to the eleven data providers in a time frame of about five months to be compatible with ULISSE's overall planning. The work has been organized in three iterations:

- **Iteration 1**: For executing the *Analysis* and the *Draft* phases with each data provider in Ontology Workshops (2.5 months).

– **Iteration 2**: For a first *refinement* with the objective to complete the Topic Maps Ontologies of each data provider for all selected experiments (1.5 months).
– **Iteration 3**: For a second *refinement* with the objective to finalize the Topic Maps Ontologies and making sure that all the questions that were identified in the *Analysis* phase can be answered through populated topic maps based on the Topic Maps Ontologies.

Very shortly before Iteration 1, the data providers prepare themselves by deciding which experiment to include, identifying domain experts and gathering relevant documentation.

Ontology Workshops The main task during the first iteration of the Topic Maps Ontology design process is directed towards organizing Ontology Workshops with each of the data providers. Ontology Workshops were held with data providers individually as much as possible, as this would give the best chances in coming up with focused Topic Maps Ontologies. Only when there was a clear overlap in scientific disciplines, an Ontology Workshop was organized with more than one data provider at the same time.

A typical Ontology Workshop takes about three days:

– Day 1:
  1. Half a day is spent on presenting Topic Maps and outlining the Topic Maps Ontology design process (Preparation phase).
  2. Another half day is spent by the data provider for presenting the experiments he wants to cover (Analysis phase). Often, a few representative experiments are chosen to create the initial Ontology. After the presentation, a list of questions related to these experiments is identified.
– Day 2:
  1. With the second day, the Draft phase of the Topic Maps Ontology design is started. Time is spent building a word list of important terms. Afterwards, these words are categorized into Topic Maps concepts.
– Day 3:
  1. On the third day, the draft phase is continued by organizing the categorized word list in simple graphs that illustrate the structure of the Topic Maps Ontology.

After each Ontology Workshop, a few days are spent for creating an extended version of the Topic Maps Ontology as a presentation. A modified version[2] of the Graphical Topic Maps notation (GTM) level 0 proposal [4] is used. The level 0 notation is chosen because it seems more natural for the users who are less familiar with Topic Maps and ontologies. This initial Topic Maps Ontology marks the end of the Draft phase and serves as input to the refinement iterations.

[2] As suggested by Rani Pinchuk to the ISO/IEC JTC1/SC34/WG3 following the publication of GTMalpha: http://www.isotopicmaps.org/pipermail/sc34wg3/2009-February/003973.html

So far, no other workshop is planned for the refinement iterations. Indeed, we expect to be able to manage them through the use of email, phone conferences and, most importantly, the project wiki.

Tools The tools for ontology modelling have been chosen based on the following criteria:

1. Prior user experience and knowledge.
2. Ease of use in creating and editing of content.
3. Tool maturity.

There already exists a number of complex tools for modelling Ontologies [5, 6, 7, 8], but they do not fit well in an environment with end-users who are new to domain knowledge modelling. In order to have a streamlined and efficient process, the following tools were chosen:

1. Microsoft Excel: for collecting and post processing terms.
2. IHMC CMapTools [3]: for grouping terms and ad-hoc ontology modelling. The IHMC CMapTools software was used to visually create the initial Topic Maps Ontology. However, the CMapTools software lacks several essential features for fully modelling a Topic Maps Ontology, i.e. concepts such as Scopes, Role Types and Occurrence Types could only be handled through annotations on the structure. Nevertheless, it was still very useful to create the overall structure of the Ontology by focusing mostly on Topic Types and Associations.
3. Microsoft PowerPoint: for drawing Topic Maps Ontology in a simplified GTM notation.

2.3 Methodology Modifications

Compared to the methodology defined in [2], the following changes were made:

1. [2] lists an "end-user" phase concurrently with the analysis phase whose purpose is to learn what information the end users want from the Topic Maps based application. This phase is reduced to identifying relevant questions for verifying the Topic Maps Ontologies. It is also made part of the analysis phase instead of being a wholly separate phase.
2. [2] lists the purpose of the draft phase as creating an initial picture of what the ontology might look like. We made the draft phase a part of the Ontology Workshops. Data providers can immediately see the steps that lead to the creation of the draft of their initial Topic Maps Ontology.
3. [2] lists an interaction design phase to agree on the user interface of the application and to ensure that the ontology can support that user interface. This interaction design phase was removed. Separate work packages in the ULISSE project are tasked for defining tools and the accompanying user interfaces based on the topic maps that are authored. The refinement steps were used to verify that the ontologies can support the functionality that is required.

4. The large amount of data providers and the short timeframe available means that it is not possible to spend a lot of time teaching them about notations and tools they are not familiar with. Therefore, the Topic Maps Ontologies are simply drawn in Microsoft PowerPoint using a simplified GTM, instead of a more specific tool (such as an UML tool mentioned in [2]).

3 Results

For each workshop with a data provider, several results at different levels of detail are obtained. Afterwards, common results across workshops are merged together in order to get a "*General Ontology of Space Experiments*" next to the domain specific Ontologies for each scientific discipline.

3.1 Ontology Workshops

The draft phase of a workshop consists of several tasks (see Table 2) that subsequently lead to an ontology at a finer level of detail.

First, relevant terms are being collected in a brainstorming session by entering them into an Excel sheet. The list of words is sorted afterwards, and duplicate entries are removed. In the next step, the collected words are entered into a Concept Map and grouped together based on the experience of the domain expert (this also includes further explanation of the terms whenever necessary). For each group, a taxonomical structure is derived ("is-a" relationships or type-instance/supertype-subtype associations in Topic Maps terminology), see Figure 1 for an example.

The separate taxonomical structures are then put in relation to each other using typed associations. Further refinement of the ontology is done after the workshop by the Domain Modelling Expert. The concept map is transformed into a simplified graphical Topic Maps notation and the missing Topic Maps constructs (scopes, occurrences, roles) are added.

Table 2. Tasks in the Draft phase of a Workshop

Task	**Tool**	**Result**
Enumeration (identifying terms)	Microsoft Excel	Word List
Categorization (grouping terms)	CMapTools	Concept Map
Organization (modelling taxonomical relationships)	CMapTools	Concept Map
Organization (modelling ontological relationships)	CMapTools	Concept Map
Organization (transforming into Topic Maps Notation, including Scopes, Roles, Occurrences)	Microsoft PowerPoint	Ontology in simplified GTM notation

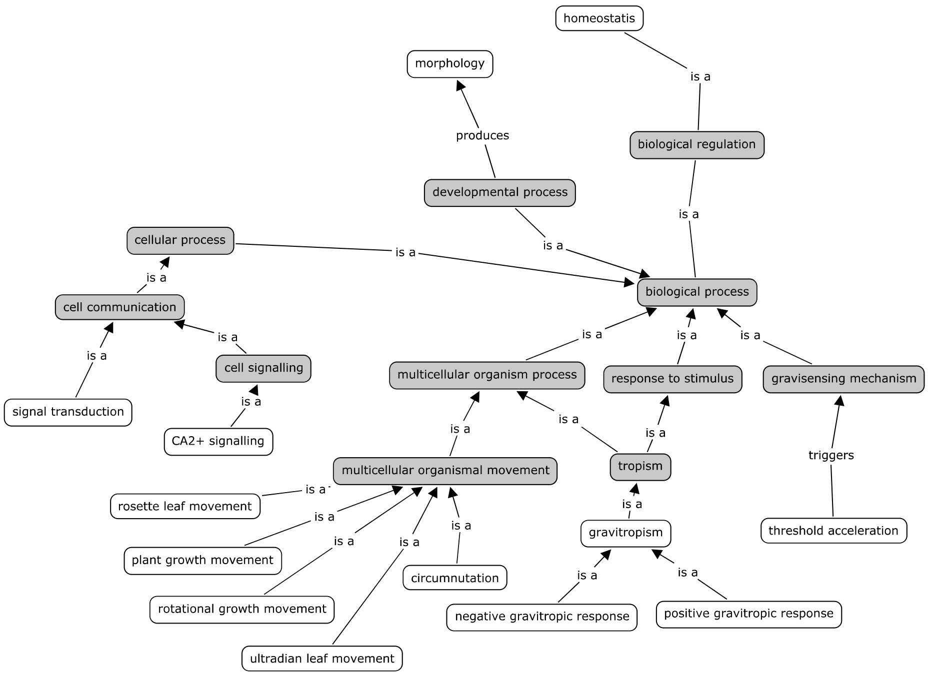

Fig. 1. Concept Map showing parts of the Space Biology domain

3.2 General Ontology for Space Experiments

In addition to the domain-specific ontologies, a "*General Ontology for Space Experiments*" is created by extracting common concepts from the other ontologies (see Figure 2). This process is done manually, but allows automatic merging of general and domain-specific ontologies afterwards by using common subject identifiers.

3.3 Ontology Statistics

Some statistics have been collected in Table 3 on the size of the Topic Maps Ontologies while they were being developed.

4 Conclusions and Future Work

This paper has presented an effective process applied for modelling ontologies for space experiments in a number of scientific domains using Topic Maps. In total, nine workshops have been conducted and the results of these workshops have been turned into a "*General Ontology for Space Experiments*" complemented by several smaller domain specific Ontologies. For creating the ontologies, existing standard applications and tools were chosen in order to efficiently streamline the

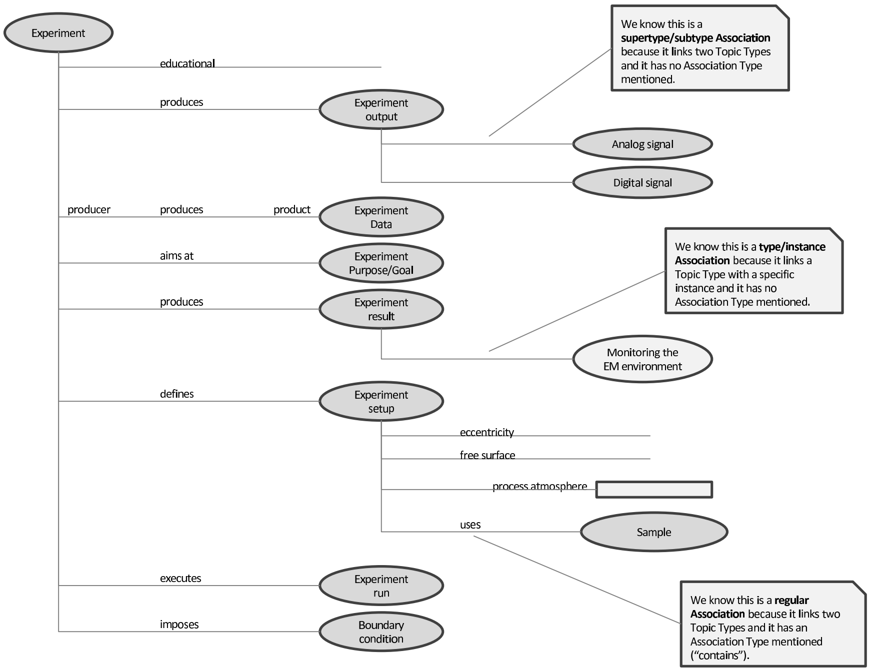

Fig. 2. Part of the General Ontology in simplified GTM notation

Table 3. Statistics on the size of five of the topic map ontologies

Topic map ontology	**#(topic types)**	**#(occurrence types)**	**#(association types)**	**#(role types)**
General Space Experiments	165	83	42	52
Cell Biology	71	3	5	10
Fluid Science	42	7	27	16
Plant Biology	63	3	4	8
Space Weather	51	5	17	24
Material Science	25	3	4	2
Physiology	70	4	11	12

process and allow the domain experts to further refine the results themselves, without Knowledge Engineers, in later stages. Using graphical Topic Maps notation to assist the process of domain knowledge modelling turned out to be a great success, as the domain experts quickly grasped the important concepts.

From the experience we got in the workshops with multiple data providers/ domain experts, we recommend the development of an integrated ontology modelling tool based on Topic Maps technology that supports the previously described process. Existing tools [5, 6, 7, 8] offer a lot of functionalities but suffer in usability – especially for domain experts who are not knowledge engineers. In addition, the existing tools do not allow focusing on a specific stage within the ontology design process.

Indeed, each phase and task focuses on different aspects and has different needs in order to allow a domain expert to work efficiently (e.g. in the "Enumeration" step the expert enters terms into a list, while in further tasks it is more convenient to visualize the terms in a 2D view for structuring). Therefore, several key requirements have been extracted:

1. Support different views to reflect different stages in the ontology modelling process (e.g. Column view for brainstorming, 2D-Graph view for structuring).
2. Ability to hide certain details of the ontology based on the current stage (e.g. hide all occurrences while structuring).
3. Operations that often occur need to be performed with very few mouse clicks or keyboard input (e.g. creating a "is-a" relationship between two topics).
4. Easy collaboration together with a versioning/diff mechanism in the refinement phase would be highly desirable.

An implementation of such a tool would further increase the efficiency of the process in modelling Topic Maps Ontologies by domain experts who are not familiar with knowledge modelling.

5 Acknowledgments

This work has received funding from the European Commission through the Seventh Framework Programme (FP7/2007-2013) under the Space Theme, under grant agreement n°218815 within the ULISSE Project (USOCs KnowLedge Integration and Dissemination for Space Science Experimentation).

References

1. ULISSE project website, http://www.ulisse-space.eu
2. Garshol, L.M.: Towards a Methodology for Developing Topic Maps Ontologies. In: Maicher, L.; Sigel, A.; Garshol, L.M. (eds.) TMRA 2006. LNCS (LNAI), vol. 4438. Springer, Heidelberg (2007)
3. IHMC CMapTools website, http://cmap.ihmc.us/conceptmap.html
4. Thomas, H.; Redmann, T.; Pressler, M.; Markscheffel, B.: GTMalpha – Towards a Graphical Notation for Topic Maps, Maicher, L.; Garshol, L. M. (eds.): Subject-centric computing. Fourth International Conference on Topic Maps Research and Applications, TMRA 2008, Leipzig, country-regionGermany, October 16-17, 2008, Revised Selected Papers. (Leipziger Beiträge zur Informatik: XII), ISBN 978-3-941152-05-2
5. TopBraid Composer, http://www.topquadrant.com/products/TB_Composer.html
6. Protégé OWL Editor, http://protege.stanford.edu
7. Onotoa Topic Maps Schema Editor, http://onotoa.topicmapslab.de
8. Altova SemanticWorks, http://www.altova.com/products/semanticworks/semantic_web_rdf_owl_editor.html