Project Number: **RI-312579**

Project Acronym: **ER-flow**

Project Full Title:

Building an European Research Community through Interoperable Workflows and Data

Theme: **Research Infrastructures**
Call Identifier: **FP7-Infrastructures-2012-1**
Funding Scheme: **Coordination and Support Action**

Deliverable D5.3
Requirements for domain semantic data and workflow description

Due date of milestone: 31/08/2013

Actual submission date: 31/08/2013

Start date of project: 01/09/2012

Duration: 24 months

Lead Contractor: University of Westminster
Dissemination Level: PU
Version: 1.0

ER-flow is supported by the FP7 Capacities Programme under contract no RI-312579

1 Table of Contents

2 List of Figures and Tables

2.1 Figures

None

2.2 Tables

3 Status and Change History

Status:	Name:	Date:	Signature:
Draft:			n.n. electronically
Reviewed:			n.n. electronically
Approved:	Gabor Terstyanszky		n.n. electronically

Table 1. Deliverable Status

Version	Date	Pages	Author	Modification
0.1	21/05/13	All	Montagnat	Document layout
0.2	27/05/13	Sections 5 and 7	Montagnat, Olabarriaga	Introduction and questionnaire
0.3	30/05/13	Appendix	Cerezo	Questionnaire and process
0.4	07/06/13	Section 6	Montagnat, Cerezo	Introduction to semantic technologies
0.5	28/06/13	Appendix	Cerezo	MoSGrid and DRIHM answers
0.6	16/07/13	Appendix	Cerezo	Inserted all community answers
0.7	19/07/13	Section 7	Montagnat	Completed methodological description
0.8	9/08/13	Sections 8 and 9	Montagnat Cerezo	Data analysis
0.9	12/08/13	All	Olabarriaga	Review, minor edits, inserted table 4
0.10	19/08/13	All	Terstyanszky	Review, minor edits
1.0	20/08/13	All	Montagnat	Integrated internal review comments from G. Terstyanszky, G. Sipos and K. Eigelis

Table 2. Deliverable Change History

4 Glossary

CGI	Coarse-Grained Interoperability
CNRS	Centre National de la Recherche Scientific (French National Centre for Scientific Research)
DCI	Distributed Computing Infrastructure
DOI	Digital Object Identifier
DRIHM	Distributed Research Infrastructure for HydroMeteorology
EGI	European Grid Infrastructure
FGI	Fine-Grained Interoperability
FITS	Flexible Image Transport System
FMA	Foundational Model of Anatomy
GEMLCA	Grid Execution Management for Legacy Code Applications
IVOA	International Virtual Observatory Alliance
LOINC	Logical Observations Identifiers Names and Codes
MoSGrid	Molecular Simulation Grid
MSML	Molecular Simulation Markup Language
NIFSTD	Neuroscience Information Framework Standard Ontologies
OPM	Open Provenance Model
OWL	Web Ontology Language
PDB	Protein Data Bank format
PROV	W3C specification for PROVenance on the Web
RDF	Resource Description Framework
RDFS	RDF Vocabulary Description Language
RIF	Rules Interchange Format
SNOMED-CT	Systematized Nomenclature of MEDicine - Clinical Terms
SSP	SHIWA Simulation Platform
SPARQL	SPARQL Protocol and RDF Query Language
SZTAKI	Magyar Tudomanyos Akademia Szamitastechnikai Kutato Intezete
UCD	Unified Content Descriptor
UoW	University of Westminster
URI	Uniform Resource Identifier
URL	Uniform Resource Link
VERCE	Virtual Earthquake and seismology Research Community in Europe e-science environment
VObs	Virtual Observatory
W3C	World Wide Web Consortium

WeNMR	Nuclear Magnetic Resonance and structural biology
WP	Work Package
XML	Extensible Markup Language

Table 3. Glossary

5 Introduction

Semantics is the study of meaning. It applies to a wide-range of non-technical and technical concepts such as linguistics, programming languages, raw numbers, etc. In computer sciences, it focuses on the relation between digital artefacts such as data entities or data entity transformation processes and their signification.

Properly understanding the semantics of scientific data and computational processes is critical to implement scientific experiments and analyze experimental results. Traditionally, semantics of data or other scientific resources is not described directly in machine-readable formats. Scientists acquire the necessary knowledge to manipulate data and design meaningful data analysis pipelines through human-readable documentation and/or dedicated training. An increasing trend to formalize the semantics of data and computational processes is observed, especially in the context of large-scale consortiums where scientific resources are shared. They enable:

- Making semantics explicit and part of the scientific resources delivered.
- Disambiguating semantics and facilitating scientific resources reuse and repurposing.
- Improving computer legibility of scientific resources for automated manipulation.

Data or computational processes semantic description and manipulation techniques have been developed over the past years, especially in the context of the Semantic Web. Within scientific communities, these techniques led to the creation of domain-oriented semantic resources, such as:

- Specific vocabularies and taxonomies used to precisely define the terms used in a given scientific domain.
- Ontologies categorizing entities, defining the relations between them, and potential rules that systematically apply between different entities.
- Documented data repositories, which content is made explicit through semantic means.
- Documented computational processes, in which action on data is made explicit.
- Annotations associated to scientific resources to describe their context, provenance, validity, etc.

The aim of this document is to report on the ER-flow survey led among user communities to capture requirements for domain-specific data and workflows description. It describes the community consultation process adopted, the input collected and the analysis performed based on the collected data. The aim is to make an objective assessment of existing, planned and expected semantic resources within the user communities represented in the project to feed in the work of ER-flow Task 4.4 on “data semantics and workflows specification”.

6 Data and workflow semantics

Properly understanding the semantics of data, computational processes and workflows is critical to any scientific investigation. Semantics of scientific data, including description of the meaning of data / values, conditions of acquisition, validity, precision, encoding, etc, is needed to properly interpret data, (re)use it in different contexts, and analyze scientific results. Similarly, the semantics of computational processes is tightly coupled to the semantics of data, as any kind of computational process can be seen as a data transformation step, which alters the semantics of input data to produce new meaningful output data. Workflows, as a means to formally describe scientific processes, carry the computational semantics of these processes, complementarily to the semantics of data transformations involved. Describing semantics of data, computational processes and workflows explicitly leverages these scientific resources, facilitating their reuse and repurposing.

6.1 Explicit semantics

There are different types of data manipulated in scientific computing environments. In particular, scientific data is often manipulated as:

- Large amounts of raw data files.
- Structured data, accessible through databases with a specific query interface (especially relational databases).
- Metadata, annotating the raw data with content description, complementary information on the data acquisition context, and/or provenance information. It can be attached to raw data through various means: joint to the raw data files in structured file containers, as separate data files, into relational or RDF databases, attached to file catalogues, etc.
- Metadata provided as processing parameters for the computational method, specified either through configuration files or as program command line arguments.

Raw data files are often opaque, carrying no explicit semantics. A common mean to expose some of the data semantics to end-users is through the adoption of meaningful file names and file paths. File names thus commonly include production date, name of object described, etc, and file paths provide categorization information. This kind of semantic information is hardly explicit, although it can be used as a basis for semantic data generation through a transformation process. Some file formats also include semantic metadata describing the data enclosed. Metadata constitutes a basic source of semantics, more or less informative and explicit, depending on the amount of annotations integrated (which can be imposed or variable, depending on the care taken upon data generation), as well as the explicit definition of metadata categories. Similarly, metadata attached to raw files through file catalogues and databases may carry rich semantic information.

Structured data usually carries richer semantics, through the data model implemented. Relational databases, for instance, define data entities associated to precise categories (column types), and relations between these entities (entities belonging to a same row are correlated). However, the semantics carried is not explicit:

- Data categories are defined through data representation types associated to each column, e.g. *integer*, which gives limited information on the nature of data. The column names often carry complementary but non-documented information on the nature of data in that column.
- Relations between data entities are not explicit. Column names may help in identifying the relation (e.g. a relation between columns “name” and “date of birth” probably is “age”), but it is not explicit, nor documented in the data model.

Other database models may carry explicit semantics though, in particular RDF graphs described in the next sub-section.

Processing parameters are often defined in documentation about the tools (if at all). The semantics of processing parameters is explicit in case of semantic description of processing tools, such as Semantic Web Services.

Closely related to the semantics of data, the semantics of processing tools used to transform data is also highly relevant to scientists. Minimal semantics on the type of data consumed and produced is usually defined through processing tools invocation interface, but the semantics of the data transformation process is rarely explicit (except through the processing tool executable name), as well as the nature of data manipulated (a “string” of “filename” type of input / output parameter gives little information on the precise nature of this data). Defining processing tools semantics is achieved informally through tools documentation and can be achieved formally through structured tool specification and interface parameters annotation.

Workflow languages also formalize the interaction between different data processing tools. Workflows describe how multiple tools are invoked to achieve a particular computation, and as such they provide fine-grained information on a complex computation process. At a coarser grain, a complete workflow can be seen as a data processing tool of its own and described similarly to any processing tool.

The explicit description of data and processing tools semantics is relevant both for users of the resources thus described, who can find unambiguous documentation on these resources (attached to the resources themselves), and for scientific support platforms, which can make use of these descriptions to provide different levels of assistance such as validation of the coherence of user actions or design guidance. The rest of this section describes common means of semantic information specification and possible use in the context of the ER-flow project.

6.2 Possible use of machine-readable semantic data

Semantic information is used for annotating data with complementary information, which may describe data acquisition or production context, precise nature of data, data format, connection of this data with other data entities, etc. Depending on the kind of metadata available, these annotations may have different use, in particular:

- Information on data encoding and data format may be used to validate the compatibility of data with processing tools (given that similar information on tools inputs and outputs is available), or transform data to a different format.
- Information on data nature and role may be used to validate the proper use of data in a computational process. For instance, not all images can be processed through a given medical image analysis software; specific image modalities have to be considered. Furthermore, when several data type description taxonomies are in use, alignment techniques make it possible to reuse data described through one taxonomy in a context involving another, aligned taxonomy.
- Information on data precision and validity range may be used to check the validity or the coherence of data processing actions ordered.
- Information relating data to complementary data or similar data may be used to enrich data search.
- Information on data provenance, especially data produced by computational processes such as workflows, is relevant for scientists when analysing data. It describes the link between data and the computational processes used to generate it, clarifying the link between transformation tools function and the data, thus making data computation chain reanalysis possible (for debugging or scientific analysis purposes). Provenance data is also relevant in some data search contexts.
- Information on processing tools input and output data may be used to validate data processing chains created in workflows, at design time (designer assistance) and / or at run time (computation validity check).

- Information on processing tools function may be used to assist workflow designers by inferring existing tools from required function, defining alternatives to a given tool, or validating processing chains.

The various usages of semantic data presented above depend on capabilities for metadata search and to infer information from facts stated through metadata annotation. The automated exploitation of semantic data requires identifying the proper metadata (and indirectly the data bearing some annotation), or determining how a combination of annotations leads to some conclusion (by logical reasoning). Hence, semantic technologies are not limited to the description of annotations, but also to search and reasoning engines.

6.3 ER-flow motivations for investigating semantics

In the context of the ER-flow project, focusing on interoperability of multiple workflow systems, both semantics of data and semantics of computational processes are highly relevant:

- Data interoperability is a key aspect of workflow interoperability in a context of execution over multiple Distributed Computing Infrastructures (DCIs) making use of different data management systems. For instance, Korkhov and co-authors state that "While composing a meta-workflow, it is necessary to make sure that the data formats of the shared workflows are compatible which will enable them to interoperate. Data formats currently are not checked by the SHIWA platform and are in the full responsibility of users uploading and composing their workflows" [Korkh, 11]. Also refer to ER-flow deliverable 4.1 on data interoperability for details [D4.1].
- The reuse and repurposing of workflows that can be exchanged through the SHIWA Repository and integrated in meta-workflows through the SHIWA Simulation Platform require workflow designers to clearly understand the function and boundary conditions of workflows.

The forthcoming ER-flow deliverable D4.3 (Study of domain semantic data and workflow description) in particular will aim at describing the impact of data and computational semantics on the SHIWA Simulation Platform design and functionality.

6.4 Semantic Web technologies

Many kinds of meta-data (or annotation) mechanisms that provide complementary information on main data entities can be used to describe the semantics of this data. The most advanced and the most widely adopted set of specifications to describe and manipulate semantics are undoubtedly the technologies developed in the context of the *Semantic Web*[1]. Semantic Web standards cover all aspects related to semantic data representation and manipulation, including semantic data format, data model description, reference taxonomies description, data retrieval, reasoning over information sets, etc.

The *World Wide Web Consortium*[2] (W3C) defined multiple standards and recommendations related to explicit semantics description, in addition to the standards pertaining to the Web of data. The Semantic Web makes a particular focus on the relations between different data items (a.k.a. *Linked Data*[3]). The basis for the Semantic Web is the *Resources Description Framework*[4] (RDF), which makes it possible to uniquely identify any data resource available over the Web and relate it to other data resources. RDF entities are typically composed by triples defining a source data entity, a relation, and a target data entity. The resources involved (source, relation and entity) are described by unique identifiers, which unambiguously refer to physical data artefacts or concepts described in a well-defined vocabulary (or ontology).

[1] Semantic Web, http://www.w3.org/standards/semanticweb/

[2] World Wide Web Consortium (W3C), http://www.w3.org

[3] Linked Data, http://en.wikipedia.org/wiki/Linked_data

[4] Resource Description Framework (RDF), http://www.w3.org/TR/rdf-mt/

Several vocabulary definition languages, such as the *RDF vocabulary description language*[5] (RDFS) and the *Web Ontology Language*[6] (OWL) are specified to organize data. Vocabularies are both used to define terms within an area of concerns and classify them, characterizing possible relationships and defining possible constraints on using those terms. Vocabularies are therefore related to the type of reasoning that can be made upon entities described through these vocabularies. Different vocabulary description languages imply different reasoning abilities and different reasoning complexity.

RDF data sets can represent large databases of annotations. A set of RDF annotations can be seen as a graph composed of one or more connected components; nodes are the RDF triple sources and targets, and edges are the RDF triple relations. To retrieve relevant information in RDF graphs, the W3C defined the RDF-specific *SPARQL Protocol and RDF Query Language*[7] (SPARQL). SPARQL can be used to search for specific sub-graphs that match some search criterion (graph search pattern) in an RDF data set. Beyond its advanced pattern selection capabilities, SPARQL can also modify existing RDF triples or insert new data in RDF graphs through specific clauses. SPARQL is therefore a powerful and multipurpose query language.

Finally, new relations may be inferred from the known relations stated in RDF triples (facts database) and some additional information on the data stated in a vocabulary as a set of logical rules. For instance, if an RDF fact states that “universeSimulator” is a workflow and there is a rule specifying that any workflow is a data processing tool. Then, it can be inferred that “universeSimulator” is a data processing tool. New facts deduction is often referred to as *reasoning* over the data set. There are several languages to specify logical rules, either as parts of the vocabulary definition languages (RDFS, OWL…) or through the dedicated *Rule Interchange Format*[8] (RIF). Inference engines, also known as *reasoners*, are used to infer all possible facts from a set of known facts and a set of rules, or to validate the coherence of a set of known facts. SPARQL is often complemented with a reasoner when applying queries, so that not explicitly stated but deducible facts (*semantic entailment relations*) can be taken into account in the query process. Such a reasoning capability is known as a SPARQL *entailment regime*[9], defining which entailment relations are used and which queries are well formed for the regime.

Semantic technologies developed in the context of the extension of the Web of data towards richer and better-documented information are based on a rich set of widely adopted standards published by the W3C that are general enough to serve many purposes. In addition to this specification work, semantic Web technologies also benefit from a large tooling set implementation that reflects the level of adoption of these standards. Any work on semantic data creation and manipulation should carefully consider the techniques and tools already developed in this area.

[5] RDF Vocabulary Description Language (RDFS), http://www.w3.org/TR/rdf-schema/

[6] RDF Vocabulary Description Language (RDFS), http://www.w3.org/TR/rdf-schema/

[7] SPARQL Protocol and RDF Query Language (SPARQL), http://www.w3.org/standards/techs/sparql

[8] Rule Interchange Format (RIF), http://www.w3.org/standards/techs/rif

[9] SPARQL entailment regimes, http://www.w3.org/TR/sparql11-entailment

7 Communities consultation process

7.1 Method

Several communities were consulted through the same questionnaire (detailed in section 7.2) covering end-users awareness of semantic technologies as well as the use or the planned use of semantic technologies in domain-specific activities. This questionnaire was meant to manually analyze answers and compile them in this document. It was designed by WP4 and WP5 members to ensure a good coverage of both technical aspects and user-oriented concerns. The questionnaire was distributed to the four user communities involved in the ER-flow project, to two European projects cooperating with ER-flow (VERCE and DRIHM) and it was forwarded to the European Grid Infrastructure (EGI.eu) user community through its User Community Board. One month was left between the distribution of the questionnaire (at the beginning of June 2013) and the collection of results. Replies were received from all four ER-flow communities (Heliophysics, Computational Chemistry, Astronomy & Astrophysics, and Life Sciences), DRIHM (Hydrometeorology), VERCE (Seismology) and the WeNMR project (Structural Biology) reached through EGI.eu. In the next sections, WeNMR was grouped with Life Sciences to which it represents a sub-domain. Six main domains are thus presented.

It was expected that end-users receiving this questionnaire would have very different background regarding semantic technologies, and very different practice in their scientific activity. To alleviate the semantic background heterogeneity problem, the questionnaire was distributed together with the text from the Section 6 of this document, clarifying to the queried persons what is precisely meant by “semantic data and workflow description” and what can be considered as semantic description models or semantic data manipulation tools. Communities of different sizes were targeted. Feedback was thus collected both on small-scale and highly technical initiatives, as well as much larger consortia, which use of semantic technologies may be diversified. For that purpose, the questionnaire includes first parts to identify the profile of the replier and his/her acquaintance with semantic technologies.

The expected outputs of this survey are:

- Assessing the use of semantic technologies in different communities.
- Collecting user requirements for semantic technologies.
- Analyzing and summarizing the semantic-related requirements within each community and across communities.
- Identifying future expectations from end users.

7.2 Questionnaire

The questionnaire is summarized below. The complete version sent out to user community contacts is shown in Appendix A of this document.

Questions are grouped in several sections identifying different aspects for the questionnaire analysis. The two first sections deal with the replier profile (for determining his/her background and the kind of community he/she is representative of) and his/her familiarity with semantic technologies. Other sections relate to the use or the planned use of semantic technologies within the community considered. The third section identifies the need for semantic technologies known from the replier. Two major categories are distinguished: semantic description and manipulation of (i) domain data and (ii) domain processing tools used to analyze domain data. The former category is often the most well known and the best-developed aspect of semantic technology use. The second category is needed to link semantic of data with the data processing tools embedded in workflow execution systems. The fourth section identifies the semantic resources (data models, processing tools repositories, etc) already in use within the community, and the fifth section focuses on planned usage of semantic resources. The sixth and seventh sections address the used and planned semantic-aware tools within the community. The eighth and ninth sections open the

discussion to user expectations regarding semantic technologies and any further comments on semantic-related aspects relevant to the community.
The complete questionnaire is shown below:

1. Profile.

- Would you qualify yourself as?
  Scientist, *Workflow developer*, *Middleware developer*, *Other (explain)*.
- What is your community or scientific domain?
- Are you replying in quality of?
  Individual, *Research group (size?)*, *Community (Which one? Size?)*.

2. Familiarity with semantic technologies.

- How familiar are you with semantic technologies?
  Not at all, *Only heard of*, *Acquainted*, *Expert*.
- How familiar are you with the Semantic Web?
  Not at all, *Only heard of*, *Acquainted*, *Expert*.
- Are you aware of specific semantic resources in use in your community?
  No, *Only heard of*, *Yes*.
- Within your community, would you say that semantic technologies are…
  Not useful, *Not used but probably useful*, *Marginally used*, *Increasingly used*, *Commonly used*. (Pick one).

3. Identified need for semantic resources.

- Are you aware of semantic-related needs within your community?
  Yes, *No*. (If yes, give details).
- What are the data semantic-related needs?
  Content description, *Complementary information on acquisition context*, *Provenance information*, *Data Traceability*, *Other…*
- What are the computational semantic-related needs?
  Data processing tools description, *Data processing tools cataloguing*, *Computation coherency checking*, *Workflow design assistance*, *Other…*

4. Semantic resources in use.

- Are you aware of the use of some of the following semantic resources? (If yes, briefly explain which ones and what they are used for).
  - Domain-specific vocabularies and/or taxonomies?
  - Domain-specific ontologies?
  - Well-documented data models?
  - Well-documented data processing tools?
  - Well-documented data repositories?
  - Well-documented data processing tool repositories?
- Are you aware of other annotations or metadata associated to domain resources? Which ones?
- Other semantic resources in use? Which ones?

5. Semantic resources planned.

- Are you aware of future plans for using similar semantic resources? Briefly explain the resources and goals.

6. Semantic-aware tools in use.

- Are you aware of the use of some tools manipulating semantics of data and data processing tools?
  - Semantic-based data transformation frameworks?
  - Semantic Web-Services search/invocation frameworks?
  - Semantic-aware workflow specification languages?
  - Semantic-aware workflow execution environment?
  - Others?

7. Semantic-aware tools planned.

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals.

8. Community expectations related to semantic technologies.

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals.

9. Other.

- Please enter any other comment related to semantic technologies needs and uses within your community.

8 Community answers

Answers to the questionnaire detailed in Section 7 where received from all four communities represented in the ER-flow project (Heliophysics, Computatitonal Chemistry, Astronomy & Astrophysics and Life Sciences), from the DRIHM (HydroMeteorology) and the VERCE (Seismology) projects, and from the WeNMR (NMR and structural biology) project. All detailed answers are presented in Appendix B of this document.

	N answered questionnaires	Community size	Profile repliers
Heliophysics	6	6 individual answers	Scientists / Middleware developers
Computational Chemistry	1	100's users	Scientist / Workflow developer / Middleware developer
Astronomy & Astrophysics	1	7 repliers in name of a large community	Virtual Observatory services development
Life Sciences: Neuroscience Bioinformatics	 1[10] 1[10]	 550 ~100's	All profiles
DRIHM	1	1 project	Scientist
VERCE	1	10 organisations	Middleware developer
WeNMR	1	1 project	Scientist

Table 4. Overview of received questionnaires

As expected, the answers collected are very heterogeneous in the level of details and acquaintance with semantic technologies. In particular, the Life Science community replied to two questionnaires addressing two large sub-domains in this area: Bioinformatics and Computational neurosciences. Each of them results from questionnaires collected from and interviews conducted within the sub-community followed by a post-analysis of the group answer. In addition, the WeNMR activity (Structural Biology) falls in the boundaries of Life Sciences and it was considered as a sub-domain in the subsequent study. The Heliophysics community reported 6 individual answers to the questionnaire proposed. The other community/projects compiled a single answer based either on the input of one specialist within the community, or the analysis of multiple answers.

The answers received show that, although the level of awareness on semantic technologies and their usage among the different communities queried is different, most of the repliers usually consider themselves a reasonably acquainted with semantic technologies and they assess some regular use of semantic technologies among their community. This confirms the interest for the analysis of semantic-related data and processing tools methodologies being developed within diverse scientific communities that is proposed in this document. This also confirms that the user community consultation process set up could reach relevant actors in this area in most scientific communities queried.

The major differences observed per-community in the familiarity and usage of semantic technologies are discussed in the following sub-sections. Next, section 9 presents an overall analysis of the replies collected and outlines the main findings of this study.

8.1 Heliophysics community

The Heliophysics community returned 6 questionnaires from six individuals (scientists and middleware developers) showing limited acquaintance with semantic technologies. All but one replier give concrete examples of semantic resources already in use in the community.

[10] Summary of several questionnaires and inteveriews

[10]

All consider that semantic technologies are “probably useful” or “increasingly used” in the community. Among the survey conducted, this community shows one of the lowest levels of acquaintance concerning semantic technologies, but it still displays a clear interest and some existing semantic resources.

8.2 Computational Chemistry community

The Computational Chemistry community provided a compiled answer from scientists, workflow and middleware developers. The experience is collected from the MoSGrid project, which represents three sub-groups in Computational Chemistry: quantum chemistry, molecular dynamics and docking. The community has more than 100 users. Semantic technologies are commonly used and the level of awareness is rather high. This is confirmed by the highly detailed answers to the questions.

8.3 Astronomy and Astrophysics community

The Astronomy and Astrophysics community has a very good experience with semantic technologies, which was already identified in deliverable [D4.1], especially with the set up of the international “Virtual Observatory” and the design of a meta-catalogue to access all kinds of astronomical data. Their reply came from a specialized group at INAF Trieste on behalf of the community. The use of semantic technologies is considered as common and the repliers qualify themselves as experts in this field.

8.4 Life-Science community

Life Sciences constitute a broad area with many different sub-communities and interactions with other research fields (such as Chemistry and Nuclear Magnetic Resonance for instance). Answers in Life Sciences are collected from 3 sub-communities dealing with Bioinformatics, Computational neurosciences and Structural Biology. Bioinformatics and Computational neurosciences provided detailed answers compiled after a rather extensive query process and representative of hundreds of end users. The repliers’ familiarity with semantic technologies was high and semantic technologies are considered increasingly used within these areas. The Structural Biology community returned an individual answer self-considered as less acquainted with semantic technologies, but still mentioning and increasing interest for these and several semantic resources in use within this community.

8.5 Hydrometeorology community

The Hydrometeorology community returned an answer on behalf of the DRIHM consortium. The replier self-qualifies as familiar with semantic technologies and considers that there is a marginal use of these in the community. Still, some needs are well identified and existing semantic resources are mentioned.

8.6 Seismology community

The Seismology community provided a single answer from the VERCE consortium. It shows the lowest level of familiarity with semantic technologies. It seems this community (which size is not evaluated) has former experience in applying semantic technologies that turned out to be too costly to follow upon, although several needs are well identified.

9 User requirements analysis

As discussed above, the interest for semantic technologies is clear in all communities queried. The motivations for using semantic technologies are numerous and diverse, but some common trends can be identified:

- The Heliophysics community outlines the need for searching data and creating links between different data items (raw data, events, instruments). The need to link scientific publications with data and to reuse workflows is also explicitly mentioned.
- The Computational Chemistry community explains how the adoption of the Molecular Simulation Markup Language (MSML), providing application-independent description of chemicals and computational aspects of chemical simulations, facilitates resources sharing (data and processing tools), reuse, and results comparison.
- The Astronomy & Astrophysics community created a meta-catalogue of data repository (Virtual Observatory) and aims for better description of data, repositories, and data processing tools to improve resources search and alignment.
- The Life-Sciences community proposes a long list of potential uses of semantic technologies based on domain-specific ontologies, including data sharing, data transformation, heterogeneous data federation, long term hosting, links with literature, reuse and reanalysis of data, data analysis tools interoperability, workflows design assistance, tools and workflows repurposing.
- The HydroMeteorology community outlines the need for clear parameters and units definition. Beyond this, it expresses interest for all kinds of semantic technologies use (see Table 4).
- The Seismology community expresses some interest for semantically described services and advanced search engines.

It can be seen that there is a lot of emphasis on data annotation, search and linking with other data items or related resources (such as scientific publications referring the data). The Semantic Web technologies offer a large variety of languages, methods and tools to achieve these goals (vocabularies description, data annotation, advanced querying, etc). Data format transformation and data model alignment are also frequently mentioned. Part of this requirement is domain-specific, as the vocabularies / ontologies used to categorize data need to be designed by community experts. Beyond the manipulation of data, and more related to the direct objectives of the ER-flow project, there is also a clear interest for sharing, reuse and possibly repurposing of data processing tools. In many cases, data processing tools cataloguing and workflow design assistance are also desirable.

9.1 Needs for semantic technologies identified and expectations

The questionnaire included a list of data-related and computation-related needs for semantic resources - Table 5 summarizes the answers of the various communities.

This table clearly shows that the classical use of semantic technologies anticipated in the questionnaire (content description, capturing information on the acquisition context, data provenance and data traceability for data-related needs; tools description, tools cataloguing, tools composition, coherency checking, and workflow design assistance for computation-related needs) are very well covered by a majority of communities represented in this survey. The only “other” needs mentioned are quite community-specific.

There were very few “expectations” reported in the questionnaire responses received, beyond what was mentioned as community needs. It seems that all communities face globalization challenges (data and processing tools broadly and often openly available, need for sharing and interoperability at a large scale, etc.) for which the requirements are already well identified and semantic technologies are perceived as a promising path to explore.

		Heliophysics	Computational Chemistry	Astronomy & Astrophysics	Life Sciences	HyrdoMeteorology	Seismology
Data-related	Content description	X	X	X	X	X	X
	Acquisition context	X		X	X	X	X
	Provenance	X	X	X	X	X	
	Traceability	X[11]		X	X[10]	X	
	Other					X[12]	
Computation-related	Tools description	X	X	X	X	X	
	Tools cataloguing	X	X	X	X		X
	Coherency checking			X	X	X	
	Workflow design assistance	X[13]			X	X	
	Other					X[14]	

Table 5. Per-community identified needs for semantic technologies

9.2 Data-related semantic resources

9.2.1 Survey summary

Astronomical data archives are the most important resources for the astronomical community. Given that astronomy is an observational science and that observed events and phenomena cannot be replicated, data collected during observations have to be preserved with great care. For this reason a considerable amount of resources is spent to create and maintain archives. The Astronomy and Astrophysics community has thus invested significant effort in the setup of an international-scale meta-repository to access distributed, cross-institution and cross-instruments data repositories in the context of the International Virtual Observatory Alliance[15]. The indexed repositories are heterogeneous, and semantic technologies are used to improve data search. To ensure non-ambiguous designation of astronomical objects, a community-wide Digital Object Identifier (DOI) scheme was set up. Most common astronomical quantities are defined in the IVOA Unified Content Descriptor models (UCDs). Other kinds of data are described through narrower use vocabularies or even much more informal "folksonomies". The IVOA progresses towards an ontological definition of astronomical objects, and the Simbad ontology for astronomical objects types already includes about 150 terms. New vocabularies are being developed for different sub-domains such as High Energy Astrophysics, Radio-Astronomy and Planetology. Prior to these efforts, different data models were in use. Among them is the standard file format FITS, which is too open to constitute a data model in itself (all metadata is optional), but provides a strong basis to define well-accepted data models.

The **Heliophysics community** inherits from the investment on semantic technologies conducted in the Astronomical community. It has numerous semantic resources in use, some of which are inherited from international collaborations such as IVOA and others are more specifically related to the particular heliophysics sub-domain. Many vocabularies (IVOA

[11] including data traceability from papers

[12] All use and retrieval metadata

[13] including workflow templates to reuse and query existing workflows

[14] Map visualisation

[15] IVOA: http://www.ivoa.net

UCD1+, IVOA Thesaurus, VO-Theory), ontologies (HELIO ontology[16], Space Physics Archive Search and Extract[17], HEK) and other data models (IVOA Characterization and Observation Data Models, ESA-FOREST data model for heliophysics, etc.) have been developed. There is a clear push towards open data publication[18]. There are many structured data repositories open to the community, among which NASA CDAS, HELIO/DPAS, the VSO and the JSOC have been cited. The FITS standard file format plays an important role for data interoperability. There are tools to link scientific publications with data (ADS). Further semantic resources are being developed, especially in the context of the FOREST and the SOLARIS projects.

The **Computational Chemistry community** uniformly uses the MSML formal language to describe both data and computational processes. There are plans to extend this language to cover more sub-domains. Beyond the adoption of this pivot format to foster interoperability, the MoSGrid repository was developed to enrich data files (stored in XtreemFS) with additional metadata. It is backed-up by a portlet-based data search engine.

There are many international-scale organizations that maintain websites and Web services for finding data, methods and vocabularies in the area of **Life Sciences**. As a consequence, many vocabularies, taxonomies and ontologies exist and are massively used in this domain. The topics cover medical, physio-pathological, radiological, biological and experiment setup concepts in particular. Some of the most commonly used taxonomies and ontologies are ConceptWiki, SNOMED Clinical Terms (SNOMED-CT), Convergent Medical Terminology (CMT), NCBI taxonomy, RADLex ontology of radiology terms, Foundational Model of Anatomy (FMA), NeuroLex (formerly BIRNLex), Logical Observations Identifiers Names and Codes (LOINC), Neuroscience Information Framework Standard Ontologies (NIFSTD), OntoNeuroLOG ontology, etc. A large number of data models and file formats in use are also mentioned. Structured and documented data-repositories are common in all sub-domains interviewed (Bioinformatics: GenBank, KEGG, Pathway databases, UniProt, etc; Neurosciences: MRI Atlases, PhysioNet, ADNI, OASIS, etc. Structural Biology: wwPDB), although they usually provide semantic annotations in their database, but not in a machine-readable format. Creating links between entities stored in different databases, and even links between data items and scientific publication is also mentioned as highly relevant in this area.

The **HydroMeteorology community** reports the use of numerous data models and repositories at a local scale. It seems no clear standard has emerged. The Climate and Forecast (CF) standard names vocabulary is in use. Semantic is seen as a way of solving the format transformations problem, but no implementation exists yet.

9.2.2 Discussion

Semantic data description, and in some cases semantic-aware search of data, have been widely adopted in many scientific areas. Some communities such as Astronomy & Astrophysics or Life Sciences are heavily relying on semantically enriched data. The globalization of scientific data and the trend towards on-line publication of open source data strongly pushes international-scale consortia to agree on standards and data models to archive and search data sets. The primary concerns raised are the sharing of data across sub-communities (understanding data content and converting data formats), the indexing of multiple and heterogeneous data sources, and advanced data search capabilities. Others, secondary use of semantic information, e.g., for data quality checking or long-term preservation, are sometimes mentioned but they are not considered as priorities yet.

Some of the data models developed are inspired by, or based on, semantic Web standards and technologies. The complete set of W3C standards, including the SPARQL semantic

[16] HELIO project: http://www.helio-vo.eu/
[17] SPACE: http://www.spase-group.org
[18] Helophysics Data Environment: http://hpde.gsfc.nasa.gov

query and inference language, is rarely used though. Data search capabilities are therefore often ad-hoc and database-specific.
There is also little use of machine-readable semantic annotations through semantic-aware processing tools, except for data format conversion. Semantic data models are often designed for human operators to non-ambiguously annotate data and reinterpret data produced by others.

9.3 Tools-related semantic resources

9.3.1 Survey summary

In the **Astronomy & Astrophysics community**, data semantics is used in data transformation tools such as VizieR and Simbad. Data mining is also becoming very popular in astronomy, and data processing tools dedicated to data mining are growing and becoming more sophisticated. Data clustering (spatial clustering) and data classification (classification of objects found in the astronomical zoo) are two of the typical problems that are now approached by means of data mining tools. Such tools are typically used to annotate data items with semantic information. There is no central repository for data processing tools, but several initiatives provide repositories such as the US National Virtual Observatory[19] or the IVOA applications[20]. There are future plans to develop semantic-aware processing tools in the context of astronomical science gateways federation (STARnet[21]).

Within Astronomy & Astrophysics, the **Heliophysics community** has documented processing tools (IVOA tools developed through the Euro-VO program, HELIO/HFC, HEK…) and a data tool repository was mentioned (solarsoft). The only semantic-aware workflow environment mentioned is Taverna.

The MSML language in use in the **Computational Chemistry community** describes both data and computational processes. The MoSGrid environment has parsers to extract metadata from input files. In addition, file format conversion tools are available for all major data formats in use in the community (PDB, SDF, MOL(2), GROMACS, etc.).

In the area of **Life Sciences**, many data processing toolboxes are available, either as executable software (as it is usually the case in image analysis: Freesurfer, FieldTrip, FSL, SPM, ITK, EEGLab, BrainVISA, MedInria, etc.) or as Web services (often the case in Bioinformatics). Some expose these processing tools through catalogues, in particular the BioCatalogue, the myExperiment and the SHIWA workflow repositories, and the LONI repository. Workflow environments with some level of customization for applications in Life Sciences exist (e.g.,LONI pipeline, Galaxy, Taverna), among which only Taverna was referred to as a semantic-aware workflow environment. The awareness is higher for tools concerning data transformation and Web services than for semantic-enabled workflow design and execution tools.

9.3.2 Discussion

Data processing tools developed in the context of various scientific disciplines are often packaged as coherent software suites gathering complete toolboxes. These software suites have progressively gained in modularity and reusability over time, to address the needs for data analysis tools sharing and experiment reproducibility arising in large-scale communities. Yet it is common that several, non-interoperable toolboxes are in use within different sub-community, or concurrent toolboxes are exploited within the same community. Data processing tools are rarely semantically described though, and the use of processing tools is typically documented in human-readable documentation.

[19] National Virtual Observatory: http://nvo.stsci.edu/vor10/index.aspx
[20] IVOA applications: http://wiki.ivoa.net/twiki/bin/view/IVOA/IvoaApplications
[21] STARnet federation: http://www.oact.inaf.it/STARnet

In some rare cases, tools are exposed in browsable catalogues. The only tool catalogues explicitly mentioned in the responses received are the LONI catalogue (an ad-hoc system developed by UCLA LONI in the context of neurosciences) and Web services for which several cataloguing options exist (e.g. BioCatalogue). It should be noted that in all cases these catalogues are only based on the syntax of the services hosted (e.g. Web services only describe a technical interface for service invocation, but they do not define any semantics in themselves). There are very few initiatives reported that deal with data processing services annotation using semantic information and the exploitation of this information by semantic-aware tools. Although the need for data processing tools sharing, reuse and possibly repurposing, as well as data processing tools cataloguing and workflow design assistance were clearly identified in many communities, there is a clear gap between the existing frameworks and the user expectations in this area.

It should be noted that several initiatives to extend Web services with semantic annotations were introduced over the past years, such as, for extended frameworks, OWL-S [Martin, 07], WSMO [Roman, 06], FLOWS [Grunin, 08], or for lighter approaches, SAWSDL [Farrell, 07] and WSMO-Lite [Vitvar, 08]. Although SAWSDL has been proposed by the W3C as a recommendation in 2007, no consensus clearly emerged, and OWL-S and SAWSDL provide good compromises for semantically annotating e-Science workflow components. Earlier work on semantic service integration within workflow emphasize on the need to properly annotate data processing tools with domain-specific information on the role of tool parameters to allow for workflow consistency checking [Gaigna, 11].

In the scientific domain, many toolboxes do not adopt the Web service standard for technical interface description and tool invocation, as most scientific codes are developed as simple command-line tools. Command-line tools interface are often very ad-hoc and their documentation accessible only through human-readable documents. In this case, wrapper tools with a formal interface representation such as GEMLCA [Delai, 05] or JGASW [Javier, 10] should be considered prior to inclusion in a semantic-aware catalogue.

10 Conclusions

This deliverable describes the process adopted in ER-flow to conduct a user survey on the usage and awareness about semantic technologies in multiple scientific domains. It summarizes the replies received from 6 communities and it discusses the most relevant points according to user feedback received.

The answers received show that the repliers usually consider themselves a reasonably acquainted with semantic technologies, and that it approves representative usage of semantic technologies among their community. This confirms the interest for the analysis of semantic-related data and processing tools methodologies being developed among diverse scientific communities, which motivated the study proposed in this document. This also confirms that the user community consultation process set up could reach knowledgeable actors in this area in most scientific communities queried.

10.1 Towards automatic semantic data manipulation

The survey shows that much emphasis is put on data annotation, search and linking with other data items or related resources. The primary concerns raised are the sharing of data across sub-communities (understanding data content and converting data formats), the indexing of multiple and heterogeneous data sources, and advanced data search capabilities. Other secondary use of semantic information, e.g. for data quality checking or long-term preservation, are sometimes mentioned but they are not considered as priorities yet. Beyond the manipulation of data, and more related to the direct objectives of the ER-flow project, there is also a clear interest for sharing, reuse and possibly repurposing of data processing tools. In many cases, data processing tools cataloguing and workflow design assistance are also desirable.

The globalization of scientific data and the trend towards on-line publication of open source data strongly pushes international-scale consortia to agree on standards and data models to archive and search data sets. Although some of the data models developed are inspired by, or based on semantic Web standards, the complete set of W3C standards, including the SPARQL semantic query and inference language, is rarely used. There is also little use of machine-readable semantic annotations through semantic-aware processing tools, except for data format conversion. Semantic data models are often designed for human operators to non-ambiguously annotate data and reinterpret data produced by others.

Data processing tools developed in the context of various scientific disciplines are often packaged as coherent software suites gathering complete toolboxes. Data processing tools are rarely semantically described though, and the use of processing tools is documented in human-readable documentation. There are very few initiatives dealing with data processing tools annotation using semantic information and the exploitation of this information by semantic-aware tools. Although the need for data processing tools sharing, reuse and possibly repurposing, as well as data processing tools cataloguing and workflow design assistance, were clearly identified in many communities, there is a clear gap between the existing frameworks, and the user expectations in this area.

10.2 Future work in ER-flow

The findings of this work will feed the study on data semantic and workflows specification from Task 4.4 in the ER-flow project. In particular, the gap between user expectations for linking semantics of data with that of processing tools, data processing tools reusability and workflow composition will be studied.

References

[D4.1] Virtual Data Objects specification, ER-flow deliverable 4.1, April 2013. https://documents.egi.eu/public/ShowDocument?docid=1740

[Delai, 05] T. Delaitre, T. Kiss, A. Goyeneche, G. Terstyanszky, S. Winter, P. Kacsuk. "GEMLCA: Running Legacy Code Applications as Grid Services", Journal of Grid Computing (JoGC), 3(1-2):75-90, 2005.

[Farrell, 07] J. Farrell, H. Lausen. “Semantic Annotations for WSDL and XML Schema”. http://www.w3.org/tr/sawsdl. Online. August 2007.

[Gaigna, 11] A. Gaignard, J. Montagnat, B. Wali and B. Gibaud, “Characterizing semantic service parameters with Role concepts to infer domain-specific knowledge at runtime”, International Conference on Knowledge Engineering and Ontology Development, KEOD 2011, Paris, France, 2011.

[Grunin, 08] M. Gruninger, R. Hull and S. McIlraith, “A short overview of flows: A first-order logic ontology of web services”, IEEE Data Engineering Bulletin. 31(3):3–7, 2008.

[Javier, 10] J. Rojas Balderrama, J. Montagnat, D. Lingrand. "jGASW: A Service-Oriented Framework Supporting High Throughput Computing and Non-functional Concerns", IEEE International Conference on Web Services (ICWS 2010), Miami, FL, USA, July 2010. doi:10.1109/ICWS.2010.59

[Korkh, 11] V. Korkohv, D. Krefting, T. Kukla, G. Terstyanszky, M. Caan and S. Olabarriaga, “Exploring workflow interoperability tools for neuroimaging data analysis”, Proceedings of the 6th workshop on Workflows in support of large-scale science (WORKS’11), Seattle, USA, November 2011, pp 87-96, ACM. http://dx.doi.org/10.1145/2110497.2110508.

[Martin, 07] D. Martin, M. Burstein, D. Mcdermott, S. Mcilraith, M. Paolucci, K. Sycara, D. L. Mcguinness, E. Sirin and N. Srini- vasan, “Bringing semantics to web services with OWL-S”, World Wide Web 10(3):243–277, 2007. doi:10.1007/s11280-007- 0033- x.

[Roman, 06] D. Roman, J. de Bruijn, A. Mocan, H. Lausen, J. Domingue, C. Bussler and D. Fensel, “WWW: WSMO, WSML, and WSMX in a nut- shell”, 2006, pp. 516–522. doi:10.1007/11836025_49.

[Vitvar, 08] T. Vitvar, J. Kopecky, J. Viskova and D. Fensel, “WSMO-Lite Annotations for Web Services”, in: 5th European Semantic Web Conference (ESWC2008), 2008, pp. 674–689.

Appendix A: questionnaire sent to communities

As part of the ER-Flow European Research Project, we are conducting a study of the current uses and further requirements for **domain semantics of data and data processing tools** in the context of **scientific workflows**.

By **domain semantics**, we mean meta-data specifying either the nature or content of scientific data or the domain goal or method used in a scientific protocol. Those meta-data are often themselves specified via ontologies, but more informal or more confidential meta-data schemes used in specific communities also count as domain semantics.

By **scientific workflows**, we mean systems meant to automate and perform scientific experiments (a.k.a. simulations) on distributed computing infrastructures (e.g. web services, grids, clouds).
Examples: ASKALON, Kepler, MOTEUR, P-GRADE, Pegasus, Taverna, Triana.

Ideally, we would like to obtain combined answers for each user community involved in the ER-Flow project. It would be greatly helpful if community representatives could circulate this survey inside their respective communities and compile a summary that would best represent the needs of their respective communities.
However, if needs and/or current usages differ too wildly inside a given community, please do not hesitate to forward to us as many distinct completed surveys as necessary to fully capture the variety of requirements.

Please complete the survey and return it to your community representative. If there is no such representative for your community or you cannot find out who that person is, please forward your answer to:
Nadia Cerezo cerezo@i3s.unice.fr and Johan Montagnat johan.montagnat@cnrs.fr

Thank you very much for your time.

Questionnaire

Profile

- Would you qualify yourself as? (Pick one)
  - ○ *Scientist*
  - ○ *Workflow developer*
  - ○ *Middleware developer*
  - ○ *Other*
    Explain: ____
- What is your community or scientific domain?
  ____
- Are you replying in quality of? (Pick one)
  - ○ *Individual*
  - ○ *Research group*
    Size: ____
  - ○ *Community*
    Which one: ____
    Size: ____

Familiarity with semantic technologies

- How familiar are you with semantic technologies? (Pick one)
  - ○ *Not at all*
  - ○ *Only heard of*

  - ○ *Acquainted*
  - ○ *Expert*
- How familiar are you with the Semantic Web? (Pick one)
  - ○ *Not at all*
  - ○ *Only heard of*
  - ○ *Acquainted*
  - ○ *Expert*
- Are you aware of specific semantic resources in use in your community? (Pick one)
  - ○ *No*
  - ○ *Only heard of*
  - ○ *Yes*
- Within your community, would you say that semantic technologies are… (Pick one)
  - ○ *Not useful*
  - ○ *Not used but probably useful*
  - ○ *Marginally used*
  - ○ *Increasingly used*
  - ○ *Commonly used*

Identified need for semantic resources

- Are you aware of semantic-related needs within your community? (Pick one)
  - ○ *Yes*
    Please give details: ____
  - ○ *No*
- What are the data semantic-related needs?
  - ○ *Content description*
  - ○ *Complementary information on acquisition context*
  - ○ *Provenance information*
  - ○ *Data Traceability*
  - ○ *Other:* ____
- What are the computational semantic-related needs?
  - ○ *Data processing tools description*
  - ○ *Data processing tools cataloguing*
  - ○ *Computation coherency checking*
  - ○ *Workflow design assistance*
  - ○ *Other:* ____

Semantic resources in use

- Are you aware of the use of some of the following semantic resources? (If yes, briefly explain which ones and what they are used for)
  - ○ Domain-specific vocabularies and/or taxonomies?
    ____
    ____
  - ○ Domain-specific ontologies?
    ____
    ____
  - ○ Well-documented data models?
    ____
    ____
  - ○ Well-documented data processing tools?
    ____
    ____
  - ○ Well-documented data repositories?
    ____
    ____
  - ○ Well-documented data processing tool repositories?

- Are you aware of other annotations or metadata associated to domain resources? Which ones?
- Other semantic resources in use? Which ones?

Semantic resources planned

- Are you aware of future plans for using similar semantic resources? Briefly explain the resources and goals.

Semantic-aware tools in use

- Are you aware of the use of some tools manipulating semantics of data and data processing tools?
  - Semantic-based data transformation frameworks? Yes/No
  - Semantic Web-Services search/invocation frameworks? Yes/No
  - Semantic-aware workflow specification languages? Yes/No
  - Semantic-aware workflow execution environment? Yes/No
  - Others: ____________________

Semantic-aware tools planned

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals.

Community expectations related to semantic technologies

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals.

Other

- Please enter any other comment related to semantic technologies needs and uses within your community.

Thank you very much for your participation!

Appendix B: questionnaires returned

This appendix collects all answers received to the questionnaire sent to various user communities:

- The four communities represented in the ER-flow project: Heliophysics (B.1), Computational chemistry (B2), Astronomy & Astrophysics (B3) and Life Sciences (B4). The Heliophysics community returned 6 questionnaires (B.1.1 to B1.6) and the Life-Science community returned two questionnaires covering two large sub-domains in this area: Bioinformatics (B4.1) and Computational neurosciences (B4.2).
- DRIHM – HydroMeteorology (B5) and VERCE – Seismology (B6) projects who signed a MoU with ER-flow.
- The WeNMR – structural biology project, active on the EGI.eu infrastructure, which answer was gathered with other Life-Science answers (B4.3).

Note that empty answer fields and unchecked options were omitted to improve legibility.

B.1 Answers from the Heliophysics community

B.1.1 Middleware developer

Profile

- Would you qualify yourself as? ***Middleware developer***
- What is your community or scientific domain? ***Heliophysics***
- Are you replying in quality of? ***Individual***

Familiarity with semantic technologies

- How familiar are you with semantic technologies? ***Acquainted***
- How familiar are you with the Semantic Web? ***Acquainted***
- Are you aware of specific semantic resources in use in your community? ***Yes***
- Within your community, would you say that semantic technologies are… ***Increasingly used***

Identified need for semantic resources

- Are you aware of semantic-related needs within your community? (Pick one) ***Yes. Some understanding of needs in heliophysics with regard to semantic linking between instruments, etc.***
- What are the data semantic-related needs? ***Content description, Complementary information on acquisition context, and Provenance information***
- What are the computational semantic-related needs? ***Data processing tools description and Data processing tools cataloguing***

Semantic resources in use

- Are you aware of the use of some of the following semantic resources? (If yes, briefly explain which ones and what they are used for)
    - Domain-specific vocabularies and/or taxonomies? ***IVOA UCD1+, IVOA Thesaurus, etc.***
    - Domain-specific ontologies? ***HELIO Ontology, other heliosphere/heliophysics ontologies***
    - Well-documented data models? ***IVOA Characterization and Observation Data Models and related IVOA DMs***
    - Well-documented data repositories? ***NASA CDAS and similar***
- Are you aware of other annotations or metadata associated to domain resources? Which ones? ***FITS file header keyword metadata***

Semantic resources planned

- Are you aware of future plans for using similar semantic resources? Briefly explain the resources and goals. ***"FOREST" ESA project developing semantic search for quicklook data for heliophysics, with reference to IVOA standards, HELIO services.***

Semantic-aware tools in use

- Are you aware of the use of some tools manipulating semantics of data and data processing tools?
  - Semantic-based data transformation frameworks? ***No***
  - Semantic Web-Services search/invocation frameworks? ***Yes***
  - Semantic-aware workflow specification languages? ***No***
  - Semantic-aware workflow execution environment? ***No***

B.1.2 Scientist

Profile

- Would you qualify yourself as? ***Scientist (mainly) but I like to think I'm also a workflow dev and somehow a amateur developer***
- What is your community or scientific domain? ***Solar physics, heliophysics***
- Are you replying in quality of? ***Individual***

Familiarity with semantic technologies

- How familiar are you with semantic technologies? ***Only heard of***
- How familiar are you with the Semantic Web? ***A bit more than "Only heard of"***
- Are you aware of specific semantic resources in use in your community? ***SPASE http://www.spase-group.org/***
- Within your community, would you say that semantic technologies are… ***Not used but probably useful / Marginally used***

Identified need for semantic resources

- Are you aware of semantic-related needs within your community? (Pick one) ***Yes (don't know to what degree)***
- What are the data semantic-related needs? ***Content description (like features detected from observational data), Data traceability from papers (that would be awesome!)***
- What are the computational semantic-related needs? ***Don't know what this is about...***

Semantic resources in use

- Are you aware of the use of some of the following semantic resources? (If yes, briefly explain which ones and what they are used for)
  - Domain-specific ontologies? ***SPASE is that... and in HELIO we had something. I should ask Anja.***
  - Well-documented data processing tools? ***Well-documented in solar physics?? Ha! That's a joke! But yes, we try.***
  - Well-documented data repositories? ***Data repositories are normally well documented... but not really well...***
  - Well-documented data processing tool repositories? ***well... we have solarsoft with some instructions on how to install it... does that answer the question?***
- Are you aware of other annotations or metadata associated to domain resources? Which ones? ***I think HELIO and HEK (lmsal) use metadata.***
- Are you aware of other annotations or metadata associated to domain resources? Which ones? ***I would say ADS, the place we look for papers in astrophysics has a lot of this stuff behind, extract information from papers about the data observed/used and it tries to gather it from the sources.***

Other

- Please enter any other comment related to semantic technologies needs and uses within your community. ***I think I need a better understanding of all these semantic technologies, tools... All names sounds familiar, but since I've never used them directly I'm lost with these detailed questions.***

B.1.3 Workflow developer / Middleware developer

Profile

- Would you qualify yourself as? ***Workflow developer and Middleware developer***
- What is your community or scientific domain? ***Heliophysics***
- Are you replying in quality of? ***Individual***

Familiarity with semantic technologies

- How familiar are you with semantic technologies? ***Only heard of***
- How familiar are you with the Semantic Web? ***Only heard of***
- Are you aware of specific semantic resources in use in your community? ***Yes***
- Within your community, would you say that semantic technologies are… ***Increasingly used***

Identified need for semantic resources

- Are you aware of semantic-related needs within your community? ***Yes***
- What are the data semantic-related needs? ***Content description, Complementary information on acquisition context, Provenance information. Go to*** http://www.mygrid.org.uk/files/presentations/HELIOposterA0.pdf ***for a nice overview of Semantic and workflows in Heliophysics.***
- What are the computational semantic-related needs? ***Data processing tools description, Data processing tools cataloguing, Workflow design assistance***

Semantic resources in use

- Are you aware of the use of some of the following semantic resources?
  - Domain-specific ontologies? ***The Semantic Mapping Services (SMS) of the HELIO (***http://www.helio-vo.eu/***) project has developed a web service that offers an Ontology for Heliophysics.***
  - Well-documented data models? ***The ESA Forest project (***http://figshare.com/articles/FOREST_a_new_heliophyics_data_system/95815***) is developing a Data model for Heliophysics. In*** http://hpde.gsfc.nasa.gov/ ***there is a short description of data models and Heliophysics.*** http://science.nasa.gov/media/medialibrary/2011/02/10/Heliophysics_Data_Policy_2009Apr12.pdf ***is a good source of information for Data in Heliophysics.***

Semantic resources planned

- Are you aware of future plans for using similar semantic resources? Briefly explain the resources and goals. ***I do not have any specific plan but I will follow this research with attention as I reckon it could prove very useful to my research.***

Semantic-aware tools in use

- Are you aware of the use of some tools manipulating semantics of data and data processing tools?
  - Semantic-based data transformation frameworks? ***No***
  - Semantic Web-Services search/invocation frameworks? ***Yes***
  - Semantic-aware workflow specification languages? ***No***
  - Semantic-aware workflow execution environment? ***No***

Other

- Please enter any other comment related to semantic technologies needs and uses within your community. ***I think that semantic tools will become increasingly necessary with the introduction of workflows in the community and the already spreading use of web-service based distributed architecture. As these architectural approaches have among their strong advantages sharing and exchange of knowledge and computational resources, semantic technologies and widely accepted data models will become fundamental in the community.***

B.1.4 Scientist

Profile

- Would you qualify yourself as? ***Scientist***
- What is your community or scientific domain? ***Solar Physics***
- Are you replying in quality of? ***Individual***

Familiarity with semantic technologies

- How familiar are you with semantic technologies? ***Only heard of***
- How familiar are you with the Semantic Web? ***Only heard of***
- Are you aware of specific semantic resources in use in your community? ***Only heard of***
- Within your community, would you say that semantic technologies are… ***Not used but probably useful***

Identified need for semantic resources

- Are you aware of semantic-related needs within your community? ***Yes. No uniform description of our data exists but we need one***
- What are the data semantic-related needs? ***Content description, Complementary information on acquisition context, Provenance information***
- What are the computational semantic-related needs? ***Data processing tools description and Workflow design assistance***

Semantic resources in use

- Are you aware of the use of some of the following semantic resources? (If yes, briefly explain which ones and what they are used for)
  - Domain-specific vocabularies and/or taxonomies? ***In VO-Theory, a specific vocabulary has been developed.***
  - Domain-specific ontologies? ***HELIO ontology for Heliophysics, but only fit a part of data. I heard also about SKOS (?) ontology used in the frame of VO-Theory (from IVOA).***
  - Well-documented data models? ***SPASE DM for plasma physics data.***
  - Well-documented data processing tools? ***ALADIN, VIZIER, TOPCAT, VO-SPEC (all IVOA tools developed through the Euro-VO program).***

Other

- Please enter any other comment related to semantic technologies needs and uses within your community. ***The above questions are too much specific for me. What I know is that we need to define a unique description of all data set we use in solar and plasma physics and then, from that build a data model that could be proposed to our community so we can develop generic tools for data processing, the same way as IVOA for stars, galaxies, …***

B.1.5 Scientist

Profile

- Would you qualify yourself as? ***Scientist***
- What is your community or scientific domain? ***Solar physics***

- Are you replying in quality of? ***Individual***

Familiarity with semantic technologies

- How familiar are you with semantic technologies? ***Only heard of***
- How familiar are you with the Semantic Web? ***Not at all***
- Are you aware of specific semantic resources in use in your community? ***Only heard of***
- Within your community, would you say that semantic technologies are… ***Not used but probably useful***

Identified need for semantic resources

- Are you aware of semantic-related needs within your community? ***No***

B.1.6 Scientist

Profile

- Would you qualify yourself as? ***Scientist***
- What is your community or scientific domain? ***I study solar flare activity, sunspot group evolution, and solar wind propagation***
- Are you replying in quality of? ***Individual***

Familiarity with semantic technologies

- How familiar are you with semantic technologies? ***Only heard of***
- How familiar are you with the Semantic Web? ***Only heard of***
- Are you aware of specific semantic resources in use in your community? ***No***
- Within your community, would you say that semantic technologies are… ***Not used but probably useful***

Identified need for semantic resources

- Are you aware of semantic-related needs within your community? ***Yes. Being able to search for data or events that have occurred on or near the Sun and be able to link them together***
- What are the data semantic-related needs?
  - *Content description* ***Yes***
  - *Complementary information on acquisition context* ***Probably***
  - *Data Traceability* ***Yes***
  - *Other:* ***Also, what events appear within the data***
- What are the computational semantic-related needs?
  - *Data processing tools description* ***Yes***
  - *Data processing tools cataloguing* ***Yes***
  - *Workflow design assistance* ***Yes***
  - *Other:* ***Also include lots of templates for scientists to use when doing queries/WFs, etc***

Semantic resources in use

- Are you aware of the use of some of the following semantic resources? (If yes, briefly explain which ones and what they are used for)
  - Domain-specific vocabularies and/or taxonomies? ***Cassifying data and events?***
  - Domain-specific ontologies? ***HELIO and HEK, I believe- both include data and events***
  - Well-documented data models? ***I think HELIO/DPAS does***

  - Well-documented data processing tools? ***HELIO/HFC, EGSO, and HEK feature finding might***
  - Well-documented data repositories? ***HELIO/DPAS, the VSO, the JSOC***
  - Well-documented data processing tool repositories? ***Maybe HELIO/HFC?***
- Are you aware of other annotations or metadata associated to domain resources? Which ones? ***Solar data FITS headers?***

Semantic resources planned

- Are you aware of future plans for using similar semantic resources? Briefly explain the resources and goals. ***I have heard of SOLARIS, and FOREST but don't think they are in action. I think the idea is to make a google interface for solar data.***

Semantic-aware tools in use

- Are you aware of the use of some tools manipulating semantics of data and data processing tools?
  - Semantic-based data transformation frameworks? ***No***
  - Semantic Web-Services search/invocation frameworks? ***No***
  - Semantic-aware workflow specification languages? ***No***
  - Semantic-aware workflow execution environment? ***Maybe Taverna?***

Semantic-aware tools planned

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals. ***FOREST – I don't know the details***

Community expectations related to semantic technologies

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals. ***I heard that Solar Orbiter is planning to take advantage of that sort of thing.***

Other

- Please enter any other comment related to semantic technologies needs and uses within your community. ***The only comment I have is that most researchers in solar physics have an adverse reaction to new web interfaces for getting data, etc. So, I think people will need a lot of convincing that what you are building is A. Useful, B. Easy to use, and C. Will be around for years (i.e. Won't disappear after it is built and the development funding ends).***

B.2 Answer from the MoSGrid community (Computational Chemistry)

Profile

- Would you qualify yourself as? ***Scientist, Workflow developer, Middleware developer***
- What is your community or scientific domain? ***Molecular Simulation Grid (MoSGrid) comprising Quantum Chemistry, Molecular Dynamics and Docking***
- Are you replying in quality of? ***Community, Molecular Simulation Grid (MoSGrid), 100 users***

Familiarity with semantic technologies

- How familiar are you with semantic technologies? ***Acquainted***
- How familiar are you with the Semantic Web? ***Only heard of***
- Are you aware of specific semantic resources in use in your community? ***Only heard of***

- Within your community, would you say that semantic technologies are… ***Commonly used***

Identified need for semantic resources

- Are you aware of semantic-related needs within your community? ***Yes. Meta-data annotation for chemical simulations by MSML (Molecular Simulation Markup Language), an advancement of the Chemical Markup Language (CML). MSML abstracts all chemical as well as computational aspects of simulations. An application and its results can be described with a common semantic. Utilizing such application independent descriptions users can easily switch between different applications or compare them.***
  Especially within the frame of the core functionality, most simulations can be performed by more than one program. However, each program demands a specific syntax with regard to its input data, i.e. all the input files for the fundamentally same calculation are different. Furthermore, the results of such simulations describe the same or similar scientific content, but again the syntax and format of the result files is different.
  A solution to these problems is provided by the Molecular Simulation Markup Language (MSML). It offers the ability to describe a simulation in an abstracted manner, i.e. the semantics of a simulation is expressed in this language with the same syntax for different programs. At the beginning of a calculation, the MSML document is translated into the input format needed by the respective program. The same holds for the simulation results: important results are extracted from the program specific output and stored in the quasi-standardized format MSML.
  This unified and standardized way of describing simulations and associated results in MSML makes these metadata descriptions independent of a program. Often the usage of certain programs in a working group is traditionally grown or given due to technical or organizational requirements. Nonetheless, similar scientific topics are investigated in different groups with different programs. Hence, MSML offers a possibility to easily compare and share molecular simulations without the need of learning different input and output syntaxes.
  MSML is capable of storing molecular information such as atom coordinates, bond information, and molecule properties. This information can then be used to perform data analysis not on basis of the, sometimes quite different, output files of the simulation runs, but on the well defined MSML files created by simulations. MSML can be processed by all portlets in the MoSGrid science gateway including the portlets for visualisation of protein structures and the graph portlet. These can therefore be used to evaluate the data in respect to their quality.
- What are the data semantic-related needs? ***Content description, Provenance information***
- What are the computational semantic-related needs? ***Data processing tools description, Data processing tools cataloguing***

Semantic resources in use

- Are you aware of the use of some of the following semantic resources? (If yes, briefly explain which ones and what they are used for)
  - Domain-specific vocabularies and/or taxonomies? ***Yes, every chemical domain has its own vocabulary which can be used in MSML.***
  - *Well-documented data processing tools?* ***Yes, MoSGrid has parsers which extract metadata from the raw data (e.g. functional, basis set, calculation type). The input is extracted and the output as well in order to facilitate the search in the repository. Moreover, there are parsers***

which extract detailed results from the output (e.g. energies, frequencies) which facilitates data analysis to the scientist.

As mentioned above, MSML holds a central role in MoSGrid. Several tools have been designed and implemented to support this role. The so-called structure parsers convert biochemical structure formats to MSML. Before being able to start a simulation, information is translated from MSML to application specific input files, which are performed by so called adapters. A generic parser extracts data from more or less obscure output files using a combination of regular expressions. Finally, so-called extractors have been implemented to convert the metadata from MSML to the JSON format to be indexed and made searchable by UNICORE. Due to the plethora of available structure formats in biochemistry the development of parsers between the molecular data and MSML is required. The most common formats in the simulations offered by MoSGrid are Gromos 87 format (.gro), Protein Data Bank file (.pdb), MDL Molfile (.mol), structure-data file (.sdf), and Tripos MOL2 file (.mol2) format. For all these file formats suitable converters were developed to allow the translation of structural information into MSML.

The structure parsers are written in Java with the contribution of BioJava (REF: Prlić, Yates et al. 2012) and the Chemistry Development Kit (REF: Steinbeck, Hoppe et al. 2006) for PDB, SDF, MOL, and MOL2 files, as well as GROMACS (Hess, Kutzner et al. 2008) for molecular dynamics simulation trajectory files.

To enable the use of MSML for later data mining and data retrieval it was necessary that the data extraction is loss-less. All of the users input and all data in structure files are retained. This includes coordinates, calculated scores, and simulation parameters used.

- Well-documented data repositories? ***Yes, the MoSGrid repository was developed based on standard technology. It is formed by joining three metadata storage parts of MoSGrid. XtreemFS serves as the underlying storage system. UNICORE provides access to XtreemFS and interfaces an Apache Lucene service, which provides the metadata index. The Portlet-API provides a convenient interface for searching this index to find raw data stored in the repository. The underlying storage system XtreemFS stores preliminary as well as raw data and results from chemical simulations. This allows re-using computationally expensive results. Therefore, XtreemFS was integrated with three layers of the MoSGrid architecture, namely the grid-middleware UNICORE, the high level middleware gUSE, and its user-interface WS-PGRADE.***

  Before a simulation is started by a user, all input data and the job description are written to XtreemFS in the form of an MSML file. The data can be uploaded directly. The user can access data stored in XtreemFS via the domain-specific portlets and the dedicated XtreemFS portlet. The portlets offer features to upload, move, and download files to or from XtreemFS. In MoSGrid each user has his/her own home directory to which by default all file transfers are staged. Within the home directory, all files related to a specific workflow are grouped in a separate folder.

  When the user starts a simulation, an MSML file is assembled and transferred to the UNICORE job directory via the gUSE UNICORE submitter. The submitter is responsible for matching job and resource requirements and installed simulation codes, starting and monitoring a job, uploading the data from XtreemFS to the job directory and back. To

allow for XtreemFS access within UNICORE job descriptions, a newly developed URL schema was integrated into WS-PGRADE and the Job Submission Description Language (JSDL) for UNICORE. As the MSML file is aggregated with results in the course of a simulation, at the end of a simulation it is written back to XtreemFS by UNICORE.
All these steps are developed without the interaction of the user by XtreemFS, UNICORE, the gUSE UNICORE submitter, and the application portlets

Semantic resources planned

- Are you aware of future plans for using similar semantic resources? Briefly explain the resources and goals. ***MoSGrid plans to expand the use of MSML for data annotation and data exchange between different domains. This shall facilitate the combination of different domains within inter-domain workflows.***

B.3 Answer from the Astronomy and Astrophysics community

Profile

- Would you qualify yourself as? ***Development of software for astronomical data archives and VObs (Virtual Observatory) services.***
- What is your community or scientific domain? ***The whole astronomy and astrophysics community.***
- Are you replying in quality of? ***Research group: IA2[22] at INAF Trieste, composed of 7 people. IA2 is an ambitious Italian Astrophysical research infrastructure project that aims at coordinating different national initiatives to improve the quality of astrophysical data services. It aims at coordinating these developments and facilitating access to this data for research purposes.***
The questionnaire has been compiled by Cristina Knapic[23] on behalf of the whole IA2/VObs.it Group at INAF Trieste.

Familiarity with semantic technologies

- How familiar are you with semantic technologies? ***Expert***
- How familiar are you with the Semantic Web? ***Expert***
- Are you aware of specific semantic resources in use in your community? ***Yes***
- Within your community, would you say that semantic technologies are… ***Commonly used***

Identified need for semantic resources

- Are you aware of semantic-related needs within your community? ***Yes. In the framework of the VObs (Virtual Observatory) it is necessary to produce new instances of the VObs Registries for a better characterization of VObs services. New data semantics are currently under definition for High-energy astrophysics, Radio-astronomy and Planetology.***
- What are the data semantic-related needs? ***Content description, Complementary information on acquisition context, Provenance information, Data Traceability.***
- What are the computational semantic-related needs? ***Data processing tools description, Data processing tools cataloguing, Computation coherency checking.***

Semantic resources in use

- Are you aware of the use of some of the following semantic resources? (If yes, briefly explain which ones and what they are used for). ***All the semantic resources listed***

[22] http://ia2.oats.inaf.it/
[23] http://www.oats.inaf.it/component/qcontacts/66-people/43-knapic-cristina

below are used within the VObs; better, they can be considered as constituting elements of the VObs.

- Domain-specific vocabularies and/or taxonomies? ***Astronomical information of relevance to the VObs is not confined to quantities easily expressed in a catalogue or a table. Fairly simple things such as position on the sky, brightness in some units, times measured in some frame, redshifts, classifications or other similar quantities are easily manipulated and stored in VOTables and can currently be identified using IVOA Unified Content Descriptors (UCDs) [std:ucd]. However, astrophysical concepts and quantities use a wide variety of names, identifications, classifications and associations, most of which cannot be described or labelled via UCDs.***

  There are several basic forms of organised semantic knowledge of potential use to the VObs. Informal "folksonomies" are at one extreme, and are a very lightly coordinated collection of labels chosen by users. A slightly more formal structure is a "vocabulary", where the label is drawn from a predefined set of definitions which can include relationships to other labels; vocabularies are primarily associated with searching and browsing tasks.
- Domain-specific ontologies? ***Ontologies allow to capture the domain in a set of logical classes, typically related in a subclass hierarchy.***

  An astronomical ontology is necessary if we want to have a computer (appear to) "understand" something of the domain. There has been some progress towards creating an ontology of astronomical object types [std:ivoa-astro-onto] to meet this need. However there are distinct use cases for letting human users find resources of interest through search and navigation of the information space.

  An example of ontology in astronomy is the ontology of object types: e.g. Ontology of Simbad astronomical object types which relies on Simbad object types and uses about 150 terms to classify objects.
- Well-documented data models? ***An astronomical data model is the result of a conceptual analysis of the characteristics of astronomical data and the relationships that obtain between those kinds of data. This analysis is mapped onto a set of graphical or linguistic conventions, able to faithfully represent the characteristics and complexity of the data. Each component of the mapping must have a physical interpretation. The entire model designates a state of affairs that exists, has existed, or might possibly exist in reality.***

  The physical interpretation is an essential part of the concept of an astronomical data model. It is in virtue of this aspect that a data model can be said to be true or false. In other words, data models have meaning; they make assertions about the nature of reality.

  FITS, for instance, is not an astronomical data model. In its current form as a standard transport mechanism, FITS does not require a physical interpretation. All astronomical keywords and even units are optional. FITS is merely a convention for exchanging bits in a manner that is independent of hardware. A FITS image might have nothing to do with astronomy; it might be a bit mapped image of Greek text. However, the FITS standardization process can become a vehicle for defining standard models of the basic astronomical data concepts. To be a data model of an astronomical image, a FITS image must require sufficient astronomical keywords to provide a meaningful interpretation, including units and a world coordinate system.

  - Well-documented data processing tools? ***Data mining is becoming very popular in astronomy and data processing tools dedicated to data mining are growing and becoming more sophisticated; data clustering and data classification are two of the typical problems that are now approached by means of data mining tools.***
    Clustering usually has a very specific meaning to an astronomer – that is "spatial clustering" (more specifically, angular clustering on the sky). In other words, we see groupings of stars close together in the sky, which we call star clusters.
    The other major dimension of astronomical research is the assignment of objects to classes. This was historically carried out one-at-a-time, as the data were collected one object at a time. ML (Machine Learning) and data mining classification algorithms were not explicitly necessary. However, in fact, the process is the same in astronomy as in data mining: (1) class discovery (clustering); (2) discover rules for the different classes (e.g. regions of parameter space); (3) build training samples to refine the rules; (4) assign new objects to known classes using new measured science data for those objects. Hence, it is accurate to say that astronomers have been data mining for centuries. Classification is a primary feature of astronomical research. We are essentially zoologists – we classify objects in the astronomical zoo.
  - Well-documented data repositories? ***Astronomical data archives are the most important resources for the astronomical community. Given that astronomy is an observational science and that observed events and phenomena cannot be replicated, data collected during observations have to be preserved with great care. For this reason a considerable amount of resources is spent to create and maintain archives. An increasing number of then is now federated in the Virtual Observatory and this greatly enhance their discovery, retrieval and exploitation.***
    Semantic data, i.e. data that says how to interpret astronomical data to extract as much scientific information as possible and produced by semantics data processing tools are store in such repositories as well.
  - Well-documented data processing tool repositories? ***At present, a unique centralized repository for astronomical data processing tools does not exist. Within the Virtual Observatory as well, each research group sets up a repository for its developed and maintained tools. Such tools include also data semantics tools.***
  - ***However, some links are available in order to allow end users to discover and access the software tools. We report here two of these links; the first one is maintained by the National Virtual Observatory (US); the second one is accessible through the IVOA wiki page: http://nvo.stsci.edu/vor10/index.aspx and http://wiki.ivoa.net/twiki/bin/view/IVOA/IvoaApplications.***
- Are you aware of other annotations or metadata associated to domain resources? Which ones? ***Besides metadata which usage is very popular in astronomy for data and software discovery and reuse, we mention here DOI, the Digital Object Identifier. A DOI is a character string used to uniquely identify an astronomical data object. Metadata about the object is stored in association with the DOI name; such metadata may include a location, such as a URL, where the object can be found. The DOI associated to an object is permanent, whereas its location and other metadata may change. Referring to an online object by its DOI provides more stable linking than simply referring to it by its URL, because if its URL changes, the publisher needs only update the metadata for the DOI to link to the new URL.***

- Other semantic resources in use? Which ones? ***We mention here RDA (Research Data Alliance) and WDS (World Data System).***
  The Research Data Alliance implements the technology, practice, and connections that make Data Work across barriers.
  The Research Data Alliance aims to accelerate and facilitate research data sharing and exchange.
  The ICSU World Data System (WDS) was created by the 29th General Assembly of the International Council for Science (ICSU) and builds on the 50-year legacy of the former ICSU World Data Centres (WDCs) and former Federation of Astronomical and Geophysical data-analysis Services (FAGS).
  WDS strives to form a worldwide 'community of excellence' for multidisciplinary scientific data, which ensures the long-term stewardship and provision of quality-assessed data and data services to the international science community and other stakeholders. Its concept aims at a transition from existing stand-alone components and services to a common globally interoperable distributed data system, with searchable common data directories and catalogues that incorporates emerging technologies and new scientific data activities. Disciplinary and multidisciplinary data networks within WDS will play a key role in moving this concept forward.

Semantic resources planned

- Are you aware of future plans for using similar semantic resources? Briefly explain the resources and goals. ***New standards related to high-energy astrophysics, radio-astronomy and planetology are currently under definition. New similar standards for other branches of astrophysics could be defined very soon.***

Semantic-aware tools in use

- Are you aware of the use of some tools manipulating semantics of data and data processing tools?
  - Semantic-based data transformation frameworks? ***VizieR[24] and Simbad.***
  - Semantic Web-Services search/invocation frameworks? ***A typical example in this context is represented by astronomical registries.***
  - Semantic-aware workflow specification languages? ***The ADQL[25] (Astronomical Data Query Language) has been developed based on SQL92. A subset of the SQL grammar are supported by ADQL. Special restrictions and extensions to SQL92 have been defined in order to support generic and astronomy specific operations.***

Semantic-aware tools planned

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals. ***According to the experience acquired by the astronomical community through the SHIWA, SCI-BUS and ER-flow projects, it is possible to state that Science Gateways are of utmost importance to expand the community of end users who make use of DCIs and their correlated resources and services. In light of this experience we started the production of Science Gateways specialized for the needs of research groups who contributed applications for the first year of ER-flow. Given this successful experience we are planning now to propose specialized science gateways also to research groups that are going to contribute applications for the second year of the project. Once we have a significant number of specialized Gateways geographically distributed, we plan to create a network of Science Gateways named STARnet[26]. Part of the Gateways federated in STARnet will act as entry***

[24] http://vizier.u-strasbg.fr/
[25] http://www.ivoa.net/documents/latest/ADQL.html
[26] http://www.oact.inaf.it/STARnet/

points toward relevant astronomical archives, including those federated in the Virtual Observatory and to its resources. Semantic-aware data processing tools are in the pool of resources that we plan to offer through STARnet.

Community expectations related to semantic technologies

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals. ***Given the relevance of semantic technologies for astronomy, we expect further investments within our community to: 1) increase the number of semantic-aware tools; 2) refine them to enhance their power and efficiency. Just to mention a couple of examples:***
    - ***Aladin[27] is an interactive software sky atlas allowing the user to visualize digitized astronomical images, superimpose entries from astronomical catalogues or databases, and interactively access related data and information from the Simbad database, the VizieR service and other archives for all known sources in the field. VObs tools dedicated to educational aspects could probably be integrated in Aladin in a short time.***
    - ***TAP[28] (Table Access Protocol) defines a service protocol for accessing general table data, including astronomical catalogues as well as general database tables. Access is provided for both database and table metadata as well as for actual table data. The current version of the protocol includes support for multiple query languages, including queries specified using ADQL, the Astronomical Data Query Language and the Parameterised Query Language (PQL, under development) within an integrated interface. It also includes support for both synchronous and asynchronous queries. Special support is provided for spatially indexed queries using the spatial extensions in ADQL. A multi-position query capability permits queries against an arbitrarily large list of astronomical targets, providing a simple spatial cross-matching capability. More sophisticated distributed cross-matching capabilities are possible by orchestrating a distributed query across multiple TAP services. TAP could be revised in order to make it suitable for the export of astronomical data of different nature.***

B.4 Answers from the Life-Science community

B.4.1 Answer from the Bioinformatics community

This questionnaire was filled by Shayan Shahand and Silvia Olabarriaga from the AMC based on input provided by four bioinformaticians of the AMC and one external researcher. A mix of self-filled questionnaires and interviews was used. The answers of all were summarized and interpreted to compile the answers below.

For privacy reasons the identity of the persons has been removed from that table.

Profile

- Would you qualify yourself as? ***The answers covered all the profiles.***
- What is your community or scientific domain? ***Mostly bioinformatics and one biomedical/semantic web researcher.***
- Are you replying in quality of? ***Most answers were provided by individuals representing their own opinions and perspectives. In fact, from the answers we can see that the persons also took into account the general trends in their areas and not so much their personal experience.***

[27] http://aladin.u-strasbg.fr/

[28] http://www.ivoa.net/documents/TAP/

Familiarity with semantic technologies

- How familiar are you with semantic technologies? ***We observe that the level of familiarity is high, including one "Expert".***
- How familiar are you with the Semantic Web? ***Same as above. We feel that most people could not really distinguish between these two questions.***
- Are you aware of specific semantic resources in use in your community? ***Most people are aware and could name a few in the next questions.***
- Within your community, would you say that semantic technologies are… ***The trend here was to consider "increasingly" or "commonly" used (if ontologies in specific domains are included)***

Identified need for semantic resources

- Are you aware of semantic-related needs within your community? ***All persons are aware. For details they mentioned the following (some of them were mentioned by several people):***
  - ***Long term hosting (data and resources)***
  - ***Standards***
  - ***Annotation tools***
  - ***Change mindset***
  - ***Develop domain-specific ontologies***
  - ***Languages for data exchange***
  - ***Nano-publications***
  - ***Integration of heterogeneous data***
- What are the data semantic-related needs? ***The persons indicated roughly equal needs for all listed capabilities.***
- What are the computational semantic-related needs? ***The persons indicate roughly the same need for tools description, coherence checking and cataloguing, and less for design assistance. Possibly this is explained by the profile of the persons, all of them are well skilled in the design of pipelines for data analysis.***

Semantic resources in use

- Are you aware of the use of some of the following semantic resources? (If yes, briefly explain which ones and what they are used for) ***The answers provided here varied a lot. All persons listed some resource. However, we noticed a difficulty to characterise them under the categories that were asked. For example, taxonomies are mixed with ontologies, data models are mixed with repositories, and methods are mixed with generic tools such as Galaxy. Perhaps the terminology adopted in the questionnaire caused confusion.***

  Below we list selected resources. A complete list with links and a brief explanation can be found in the "resources bioinformatics" tab of https://docs.google.com/spreadsheet/ccc?key=0Aum1EYSLrLE0dHBRQTZTV21COTRya3hyRUlrZXZDWkE&usp=sharing
  - Domain-specific vocabularies and/or taxonomies?
  - Domain-specific ontologies? ***Several vocabularies, taxonomies, and ontologies exist and are massively used in bioinformatics. The topics cover both medical and biological concepts. For example, for human anatomy, physiopathology, medical images, other biological samplings and clinical information.***
  - ***Some of the most commonly used taxonomies and ontologies are listed below. Because usually taxonomies are mixed with ontologies, here we list them together:***
    - ***ConceptWiki***
    - ***SNOMED CT: SNOMED Clinical Terms***
    - ***CMT: Convergent Medical Terminology***
    - ***NCBI taxonomy***

  - Well-documented data models? ***Existing data models are expressed as languages to describe the data, with metadata. Various languages and file formats exist and are associated to completely different families of processing tools.***
    - ***UMLS: Unified Medical Language System***
    - ***Systems Biology Markup Language (SBML)***
    - ***BioPAX: Biological Pathways Exchange***
    - ***RxNorm: drugs***
    - ***GO: Gene Ontology***
    - ***Peroxisome Knowledge Database (no longer maintained)***
    - ***MIAME Minimum Information About a Microarray Experiment***
    - ***MINSEQE: Minimum Information about a high-throughput SEQuencing Experiment"***
  - Well-documented data processing tools? ***This is a tricky classification. The examples here refer to systems used as development and execution environment, and that allow the publication of pipelines or workflows. One could also see as "repository"***
    - ***Galaxy***
    - ***R Bioconductor***
    - ***Entrez Programming Utilities (E-utilities)***
    - ***A large amount of bioinformatics web services***
    - ***A large amount of packages***
  - Well-documented data repositories? ***This is very developed in the field of bioinformatics. In genomics, which covers the fields of most interviewed persons, there are many repositories. They are called "databases" in the community jargon, although many are actually flat files with a collection of data, for example, DNA sequences. The annotation is normally a community process. In many cases the annotation exists only at the database level (e.g., describing the data source), and not explicitly stored in machine-readable format. Examples of repositories mentioned by the interviewees are:***
    - ***GenBank***
    - ***KEGG***
    - ***Pathway databases***
    - ***UniProt***
    - ***ArrayExpress***
    - ***Nucleotide***
    - ***EST: Expressed Sequence Tag***
    - ***GSS: Genome Survey Sequence***
  - Well-documented data processing tool repositories?
    - ***BioCatalogue***
    - ***MyExperiment***
    - ***(the SHIWA repository also have some bioinformatics tools, but it was not mentioned by the persons interviewed)***
- Are you aware of other annotations or metadata associated to domain resources? Which ones?
  - ***SKOS (Simple Knowledge Organization System)***
  - ***Nano-publications: this is new concept about publishing findings as small facts that can be processed programmatically. A nano-publication is the smallest unit of publishable information: an assertion about anything that can be uniquely identified and attributed to its author.***
- Other semantic resources in use? Which ones? ***One remarkable point is that there are many organizations that maintain websites and web services for finding***

data, methods and vocabularies for bioinformatics. Many of these resources can be accessed both by humans and programs. Examples are:

- *OBO: Open Biological and Biomedical Ontologies*
- *EBI: European Bioinformatics Institute*
- *IMI consortium: Innovative Medicines Initiative*
- *BioPortal*
- *Open PHACTS: Open Pharmacological Space (platform)*

NOTE: from here on too few answers were given, and they were ambiguous. It seems that the awareness about the available semantic web technologies is high, the potential of its application is recognized, but that this is all taking place far from practice. Therefore the tools actually in use and plans are few and they are far from the current users of workflow management systems on DCIs.

Semantic resources planned

- Are you aware of future plans for using similar semantic resources? Briefly explain the resources and goals. *Nano-publications, research objects (My Experiment) and the IMI consortium were cited here, but no explanations were given. These sound more like intensions than concrete plans.*

Semantic-aware tools in use

- Are you aware of the use of some tools manipulating semantics of data and data processing tools? *Awareness is higher for tools concerning data transformation and web services (3 and 2 persons answered positively). Only one person indicated awareness for tools for workflow specification or execution.*

Semantic-aware tools planned

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals. *Only one person answered and indicated nano-publications.*

Community expectations related to semantic technologies

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals. *Only one person answered and mentioned SPARQL endpoints.*

Other

- Please enter any other comment related to semantic technologies needs and uses within your community. *Only one person answered and indicated that semantic web tools should leave the research arena and start being deployed in practice.*

B.4.2 Answer from the Computational Neurosciences community

This questionnaire was filled by Shayan Shahand and Silvia Olabarriaga from the AMC based on input provided by representatives of medical imaging communities in Europe. A mix of self-filled questionnaires and interviews was used.
The answers of all were summarized and interpreted to compile the answers below.
For privacy reasons the identity of the persons has been removed from that table.

Profile

- Would you qualify yourself as? *The answers covered all the profiles.*
- What is your community or scientific domain? *Medical imaging, mostly with focus on neuroimaging.*

- Are you replying in quality of? ***Most answers were provided for a community of size from 10 to 400. In total, around 550 researchers are represented in the answers.***

Familiarity with semantic technologies

- How familiar are you with semantic technologies? ***We observe that the level of familiarity varies a lot. The medical imaging scientists "only heard of", the middleware developers were "acquainted". The only "Expert" represents a project that heavily exploited and used semantic web technologies.***
- How familiar are you with the Semantic Web? ***Same as above. We feel that most people could not really distinguish between these two questions.***
- Are you aware of specific semantic resources in use in your community? ***Most people are aware and could name a few in the next questions.***
- Within your community, would you say that semantic technologies are… ***Again the answers vary (probably, marginally, increasingly). One person answered "commonly" for the usage of ontologies in specific domains.***

Identified need for semantic resources

- Are you aware of semantic-related needs within your community? ***All interviewed persons are aware. For details they mentioned the following (some of them were mentioned by several people):***
  - ***Annotation tools***
  - ***Sharing of data and models***
  - ***Literature search and discovery***
  - ***Integration and federation of heterogeneous data***
  - ***Repurposing and reanalysis of data***
  - ***Assistance for experiment design and implementation***
- What are the data semantic-related needs? ***The persons indicated equal needs for all listed capabilities. PS: we could not distinguish well between data provenance and data traceability.***
- What are the computational semantic-related needs? ***The persons indicate equal needs for tools description and design assistance, and less for cataloguing and coherence checking.***

Semantic resources in use

- Are you aware of the use of some of the following semantic resources? (If yes, briefly explain which ones and what they are used for) ***The answers provided here varied a lot. All persons listed some resource. However, we noticed a difficulty to characterise them under the categories that were asked. For example, taxonomies are mixed with ontologies, and data models are mixed with repositories. With the exception of answers by two projects, most listed resources are consulted manually, and not programmatically. This indicates they are not using semantic web technologies in practice. Below we list selected resources. A complete list with links and a brief explanation can be found in the "resources neuroimaging" tab of https://docs.google.com/spreadsheet/ccc?key=0Aum1EYSLrLE0dHBRQTZTV21COTRya3hyRUlrZXZDWkE&usp=sharing***
  - Domain-specific vocabularies and/or taxonomies?
  - Domain-specific ontologies? ***Several vocabularies, taxonomies, and ontologies exist, for example for human anatomy, physiopathology, medical images, other biological samplings and clinical information. Some of the most commonly used taxonomies and ontologies are listed below. Because usually taxonomies are mixed with ontologies, here we list them together:***
    - ***RADLex: taxonomy and ontology of radiology terms***

    - ***FMA: Foundational Model of Anatomy***
    - ***NeuroLex (formerly BIRNLex)***
    - ***LOINC: Logical Observations Identifiers Names and Codes***
    - ***NIFSTD: Neuroscience Information Framework Standard Ontologies***
    - ***OntoNeuroLOG: NeuroLOG ontology***
    - ***OntoVIP: Ontology for Virtual Imaging Platform***
  - Well-documented data models?
    - ***DICOM: Digital Imaging and Communications in Medicine***
    - ***XNAT: archival with extendable data model***
    - ***EDF: biological and physical signals***
    - ***CDISC: Standardization effort for clincal data in medical research***
    - ***SHANOIR: Sharing NeurOImaging Resource***
  - Well-documented data processing tools? ***Tools are normally packaged and well documented (mostly for humans), for example Freesurfer, FieldTrip, FSL, SPM, ITK, EEGLab, BrainVISA, MedInria, etc. Some data processing tools are also available as "workflows" or pipelines for existing management systems (e.g. LONI)***
  - Well-documented data repositories?
    - ***MRI Atlases (several online websites)***
    - ***** Also see below***
  - Well-documented data processing tool repositories?
    - **SHIWA** repository
    - **LONI**: Laboratory of Neuro Imaging

  ** Some repositories include both data and data processing tools, examples are:
    - **PhysioNet**: Physiologic signals and related software
    - **ADNI**: Alzheimer's Disease Neuroimaging Initiative
    - **OASIS**: Open Access Series of Imaging Studies
    - **Siesta DB**: sleep research
- Are you aware of other annotations or metadata associated to domain resources? Which ones? ***Standards such as DICOM and HL7, and literature databases (with citations).***
- Other semantic resources in use? Which ones? ***Many projects have websites with links to several resources that are usually for humans and not for programmatic consumption. For example, NCBO, BIRN, NeuroLog, I2B2, IDASH, EU Bioimaging Project, etc.***

Semantic resources planned

- Are you aware of future plans for using similar semantic resources? Briefly explain the resources and goals. ***People agreed that the interest for using semantic resources is increasing, but there were few answers to this question. We suspect the question was ambiguous. One person replied that processing tools semantic is increasingly captured and exploited in e-Science platforms that support neurosciences.***

Semantic-aware tools in use

- Are you aware of the use of some tools manipulating semantics of data and data processing tools? ***Generally people were aware of all mentioned frameworks, languages, and environments. However they were slightly more aware of data transformation and Web-Service search/invocation frameworks. One person indicated that nowadays semantic-aware workflow languages and execution environments only use semantics for provenance information.***

Only few people answered the rest of the questionnaire. We summarize their answers here.

Semantic-aware tools planned

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals. ***Semantic-aware tools are planned for quality assessment and federation of heterogeneous data.***

Community expectations related to semantic technologies

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals. ***Ontologies definition is a complex problem as a trade-off has to be found between the ontology quality / level of details and its usability / coverage. Standardisation among many different initiatives is clearly needed. There will hardly be a one-match-all-needs ontology in the end.***

Other

- Please enter any other comment related to semantic technologies needs and uses within your community. ***There is a clear need for semantics to support / enable:***
  - ***Interdisciplinary and cross-domain studies.***
  - ***Provenance of publications (data and methods)***
  - ***Experiment reproducibility***
  - ***Data sharing***
  - ***Enrich the e-Science platforms with more intelligence base on semantics at various levels (data, method, provenance, user).***

  Semantic annotation is doubtlessly useful, however, the development effort required to build/integrate ontologies and to design/implement useful semantic-aware tools is quite high.

B.4.3 Answer from the WeNMR (structural biology) community

Profile

- Would you qualify yourself as? ***Scientist***
- What is your community or scientific domain? ***Macromolecular NMR spectroscopy and structural biology.***
- Are you replying in quality of? ***Individual***

Familiarity with semantic technologies

- How familiar are you with semantic technologies? ***Only heard of.***
- How familiar are you with the Semantic Web? ***Not at all.***
- Are you aware of specific semantic resources in use in your community? ***Yes.***
- Within your community, would you say that semantic technologies are… ***Increasingly used.***

Identified need for semantic resources

- Are you aware of semantic-related needs within your community? ***Yes. Data transfer, program interoperability, (semi)automatic pipeline/workflow set-up, data and result deposition.***
- What are the data semantic-related needs? ***The actual data (not just a description of them) are semantically complex and need a precise data model for transfer, interoperability, deposition and storage.***
- What are the computational semantic-related needs? ***Computation coherency checking and precise, standardized interface descriptions for programs and services, to use in pipeline building.***

Semantic resources in use

- Are you aware of the use of some of the following semantic resources? (If yes, briefly explain which ones and what they are used for)
  - Well-documented data models? ***CCPN data model: (maintained by my project) with Data I/O libraries, designed for complete- application-***

independent, consistent storage. Used as basis for applications (CcpNmrAnalysis suite, CcpNmr FormatConverter) for passing data between NMR and structural biology programs (ad-hoc via FormatConverter or as part of collaborative integration efforts and pipelines (e.g. WeNMR, CASD-NMR, CCPN project), and for data cleaning, curation and deposition (internally in BioMagResBank, PDBe deposition tool ECi and others, RECOORD/NRG). mmCif: structural data deposition and storage model for macromolecular structures, maintained by RCSB. Used for wwPDB deposition, data extraction, and recently set up as data communication standard between macromolecular crystallography programs. NMR-STAR: NMR data deposition and storage model, maintained by BioMagResBank. Used for NMR deposition and data extraction, as underpinning for a number of NMR analysis programs, and for data exchange, e.g. of chemical shift assignments.

- Well-documented data processing tools? ***Our own CcpNmr Analysis suite has particularly precise data I/O documentation, seeing that it uses the CCPN data model for all data. Many structure validation and generation programs in the field are well documented for users (CING, ARIA, CYANA., UNIO, ASDP, CS-Rosetta), but not necessarily as building blocks for a data flow. Also precise format of accepted data files is often a problem.***
- Well-documented data repositories? ***wwPDB (protein and macromolecular structures deposition database). BioMagResBank (NMR data deposition database)***

- Are you aware of other annotations or metadata associated to domain resources? Which ones? ***In neighbouring domains: CML (Chemical markup language, for chemistry); MIAME and related models, for microarray data and related fields.***

Semantic resources planned

- Are you aware of future plans for using similar semantic resources? Briefly explain the resources and goals. ***mmCif is currently being adopted as a standard data transfer format between structural biology programs, in addition to being a deposition and storage format, in a collaboration between the wwDB and the major crystallographic software developers. The NMR Validation Task Force is studying the possibility of setting up a light-weight common working format for structural NMR data, in collaboration with the major software developers in the area.***

Semantic-aware tools in use

- Are you aware of the use of some tools manipulating semantics of data and data processing tools?
    - Semantic-based data transformation frameworks? ***No***
    - Semantic Web-Services search/invocation frameworks? ***No***
    - Semantic-aware workflow specification languages? ***Yes***
    - Semantic-aware workflow execution environment? ***Yes***

Other

- Please enter any other comment related to semantic technologies needs and uses within your community. ***The experience and awareness of 'semantic tools' in my research domain presented here specifically relates to explicit data models for complex research data. The models themselves are metadata and make up a semantic tool, but this is likely a special case, compared to the more generic and more 'meta' nature of e.g. ontologies or more general markup languages.***

B.5 Answer from the DRIHM (HydroMeteorology) community

Profile

- Would you qualify yourself as? ***Scientist***
- What is your community or scientific domain? ***HydroInformatics***
- Are you replying in quality of? ***Community, DRIHM Distributed Research Infrastructure for HydroMeteorology, 10 Organisations performing project, many more to adopt when complete***

Familiarity with semantic technologies

- How familiar are you with semantic technologies? ***Acquainted***
- How familiar are you with the Semantic Web? ***Only heard of***
- Are you aware of specific semantic resources in use in your community? ***Yes***
- Within your community, would you say that semantic technologies are… ***Marginally used***

Identified need for semantic resources

- Are you aware of semantic-related needs within your community? ***Yes***
  Please give details: ***Parameter Definitions, unit definitions***
- What are the data semantic-related needs? ***Content description, Complementary information on acquisition context, Provenance information, Data Traceability, Other: All use and retrieval metadata***
- What are the computational semantic-related needs? ***Data processing tools description, Computation coherency checking, Workflow design assistance, Other: Map visualisation***

Semantic resources in use

- Are you aware of the use of some of the following semantic resources? (If yes, briefly explain which ones and what they are used for)
  - Domain-specific vocabularies and/or taxonomies? ***yes, usually CF standard names, various unit definitions***
  - Well-documented data processing tools? ***Loads – each model will have accompanying tools. Some will be well documented!***
  - Well-documented data repositories?
    Loads, each country will have its own sources of environmental data e.g. NERC data centres
  - Well-documented data processing tool repositories? ***FluidEarth, Many also on SourceForge***
- Are you aware of other annotations or metadata associated to domain resources? Which ones? ***ISO19115, 139, 136, 156, Certain OpenMI Constructs***

Semantic resources planned

- Are you aware of future plans for using similar semantic resources? Briefly explain the resources and goals. ***Yes. DRIHM incorporates a chain of models and data sources which need to pass results between them. Each of these comes from one of three communities: meteorological, hydrological, hydraulic. It is imperative that these communities understand one another. As such, data and metadata standards are being sought or created to ease this communication. Semantics is one key aspect.***

Semantic-aware tools in use

- Are you aware of the use of some tools manipulating semantics of data and data processing tools?
  - Semantic-based data transformation frameworks? ***Yes***
  - Semantic Web-Services search/invocation frameworks? ***Yes***
  - Semantic-aware workflow specification languages? ***Yes***

- Semantic-aware workflow execution environment? ***No***

Semantic-aware tools planned

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals. ***There are currently no plans to invoke such semantic aware technologies unless incorporated into tools required for other reasons. The issue is being addressed on a case-by-case basis at present.***

Community expectations related to semantic technologies

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals. ***See answer above.***

B.6 Answer from the VERCE (Seismology) community

Profile

- Would you qualify yourself as? ***Middleware developer***
- What is your community or scientific domain? ***HPC and Computer Science***
- Are you replying in quality of? ***Individual***

Familiarity with semantic technologies

- How familiar are you with semantic technologies? ***Not at all***
- How familiar are you with the Semantic Web? ***Not at all***
- Are you aware of specific semantic resources in use in your community? ***Only heard of***
- Within your community, would you say that semantic technologies are… ***Marginally used***

Identified need for semantic resources

- Are you aware of semantic-related needs within your community? Yes. ***Conceptual Search Engines, Ontology-Based Services.***
- What are the data semantic-related needs? ***Content description, Complementary information on acquisition context***
- What are the computational semantic-related needs? ***Data processing tools cataloguing***

Semantic resources in use

- Are you aware of the use of some of the following semantic resources? (If yes, briefly explain which ones and what they are used for)
  - Domain-specific ontologies? ***Creating ontologies for web portals include gathering data representative information and knowledge about concepts definition from the domain expert. The domain expert then creates the ontology which is integrated in the application.***

Community expectations related to semantic technologies

- Are you aware of future plans for using similar semantic-aware tools? Briefly explain the resources and goals. ***Ontology-based services have been tested in my community in the past in order to analyse how to automatically configure and (eventually) deploy infrastructure resources and services on the basis of users requirements.***

Other

- Please enter any other comment related to semantic technologies needs and uses within your community. ***Actually these kind of tests and experiments are not progressing too much, due to effort limitations and because of the difficulties found adopting such kind of technologies.***