Sustaining Heritage Access through Multivalent ArchiviNg

Large Scale Integrating Project

Grant Agreement no.: 216736

D13.1 – EXPERIMENTAL E-SCIENCE PROTOTYPE: PROOF-OF CONCEPT IMPLEMENTATION OF SHAMAN DATA GRID FOR SCIENTIFIC DATA

SHAMAN –WP13- D13.1

Project Number	FP7-ICT-216736
Due Date	NA
Actual Date	15/11/2011
Document Author/s:	INESC-ID
Version:	1.0
Dissemination level	Restricted
Status	Final
Contributing Sub-project and Workpackage	SHAMAN WP13
Document approved by	

SEVENTH FRAMEWORK PROGRAMME

Co-funded by the European Union

Document Version Control

Version	Date	Change Made (and if appropriate reason for change)	Initials of Commentator(s) or Author(s)
0.1	03/09/2011	First version of the deliverable	GA
0.2	01/10/2011	Contents added	GA, JB
0.3	11/11/2011	Additional Contents added	GA, JE, HP
0.5	14/11/2011	Additional Revisions	GA, JLB, JB
1.0	15/11/2011	Final Version	GA

Document Change Commentator or Author

Author Initials	Name of Author	Institution
GA	Gonçalo Antunes	INESC-ID
JB	José Barateiro	INESC-ID
JLB	José Borbinha	INESC-ID
HP	Helder Pina	INESC-ID
JE	João Edmundo	INESC-ID

Document Quality Control

Version QA	Date	Comments (and if appropriate reason for change)	Initials of QA Person
1.0	15/01/2011	Minor changes	IK

Catalogue Entry

Title	***Experimental e-science prototype: Proof-of-concept implementation of SHAMAN data grid for scientific data***
Creator	Gonçalo Antunes
Creator	José Barateiro
Creator	José Borbinha
Creator	Helder Pina
Creator	João Edmundo
Subject	Digital preservation prototypes, e-science
Description	This document describes the scenarios and the experimental demonstrators for the Integration and Demonstrator Subproject 3, which focus in the e-science domain. The demonstrator follows a service oriented approach combined with data grid technology, for greater redundancy and adaptability. In that sense, it adopts a similar approach as the other two integrated subprojects, but with significant adaptations, at the level of the services and of the data grid infrastructure, in order to accommodate the specific characteristics of this type of scenarios.
Publisher	SHAMAN Project
Contributor	
Date	15th November 2011
ISBN	
Type	Deliverable restricted to other programme participants (including the Commission Services)
Format	Text
Language	English
Rights	SHAMAN Project

Citation Guidelines

SHAMAN Consortium, Experimental e-science prototype: Proof-of-concept implementation of SHAMAN data grid for scientific data, (15/11/2011), http://www.shaman-ip.eu/

EXECUTIVE SUMMARY

SHAMAN Deliverable 13.1 describes the implementation of an experimental test bed for the acquisition and preservation of data in e-science scenarios. The test bed follows a Service Oriented Architecture for increased adaptability and flexibility, and makes use of federated data grids for storage. Three different e-Science scenarios are described, addressing a range of e-Science data: sensor data, scientific workflows, and experimental data. The application of a prototype to two of the scenarios is described, while the door is left open for future research in the preservation of e-science data in scenarios with highly distributed and highly heterogeneous data where ad-hoc data manipulation processes are used.

TABLE OF CONTENTS

1. INTRODUCTION

Nowadays, scientific communities engage in international collaborations, sharing huge amounts of knowledge and data in real time, with efficiency and reliability using a digital scientific infrastructure that enables this interchange. This collaborative environment comprising the infrastructures and services is usually known as e-Science (or enhanced science), and involves the requirement of interoperability and data sharing.

In a broad sense, e-Science concerns the set of techniques, services, personnel and organizations involved in collaborative and networked science. It includes technology but also human social structures and new large scale processes of making science. It also means, on the same time, a need and an opportunity for a better integration between science and engineering processes. Thus, long-term preservation can be thought as a required property for future science and engineering, to assure communication over time, so that information that is understood today is transmitted to an unknown system in the future.

The Integration & Demonstrator Subproject 3 (ISP3) addresses digital preservation in the e-Science domain. It has an experimental nature and aims to develop a proof-of-concept test bed for the acquisition and preservation of e-science data which: (*i*) make use of the SHAMAN Core Infrastructure and federated grids for storage; (*ii*) is based on reusable components, adopting a flexible and adaptable Service Oriented Architecture; (*iii*) performs workflow-based processing of data through the orchestration of data-processing services; and (*iv*) captures a large spectrum of e-Science data and context, such as sensor data, scientific workflows, experimental data, and simulation data.

For this purpose, three e-Science scenarios were addressed: (*i*) Acquisition and Preservation of Sensor Data, which addresses observational data captured by sensors used to predict the behaviour of engineering structures; (*ii*) Acquisition and Preservation of Scientific Workflows, which addresses the processes of making science, their description and additional data required for its understanding; and (*iii*) Acquisition and Preservation of Data and Simulations in Particle Physics, which addresses the data resulting from experiments and simulations in the high energy physics field.

This document is structured as follows. Sections 2, 3, and 4 describe each scenario in the same order taken above. Section 5 describes the prototype developed in general terms, while section 6 describes the specific application of the prototype to each of the cases. Finally, section 7 summarizes this deliverable.

Related Publications

J. Barateiro, G. Antunes, H. Manguinhas and J. Borbinha. *Archiving Sensor Data: Applied to Dam Safety Information*. Environmental Information Management Conference (EIM 2011). Santa Barbara, CA, USA. September 28 - 29, 2011.

G. Antunes and H. Pina. *Using Grids Federations for Digital Preservation*. In the 8th International Conference on Preservation of Digital Objects (iPRES 2011). Singapore. November 1 - 4, 2011.

2. SCENARIO 1 – ACQUISITION AND PRESERVATION OF SENSOR DATA

2.1 Scenario Description

The safety of large civil engineering structures like dams requires a comprehensive set of efforts, which must consider the structural safety, the structural monitoring, the operational safety and maintenance, and the emergency planning [1]. The consequences of failure of one of these structures may be catastrophic in many areas, such as: loss of life (minimizing human casualties is the top priority of emergency planning), environmental damage, property damage (e.g., dam flood plain), damage of other infrastructures, energy power loss, socio-economic impact, among others.

The risks associated with these scenarios can be mitigated by a number of structural and non-structural preventive measures, essentially to try to detect in advance any signs of abnormal behaviour, allowing the execution of corrective actions in time. The structural measures are mainly related to the physical safety of the structures, while the non-structural measures can comprise a broad set of concerns, such as operation guidelines, emergency action plans, alarm systems, insurance coverage, etc.

In order to improve the structural safety of large civil engineering structures, a substantial technical effort has been made to implement or improve automatic data acquisition systems able to perform real-time monitoring and trigger automatic alarms. This paradigm creates an imminent deluge of data captured by automatic monitoring systems (sensors), along with data generated by large mathematical simulations (theoretical models). Besides the fact that these monitoring systems can save lives and protect goods, they can also prevent costly repairs and help to save money in maintenance.

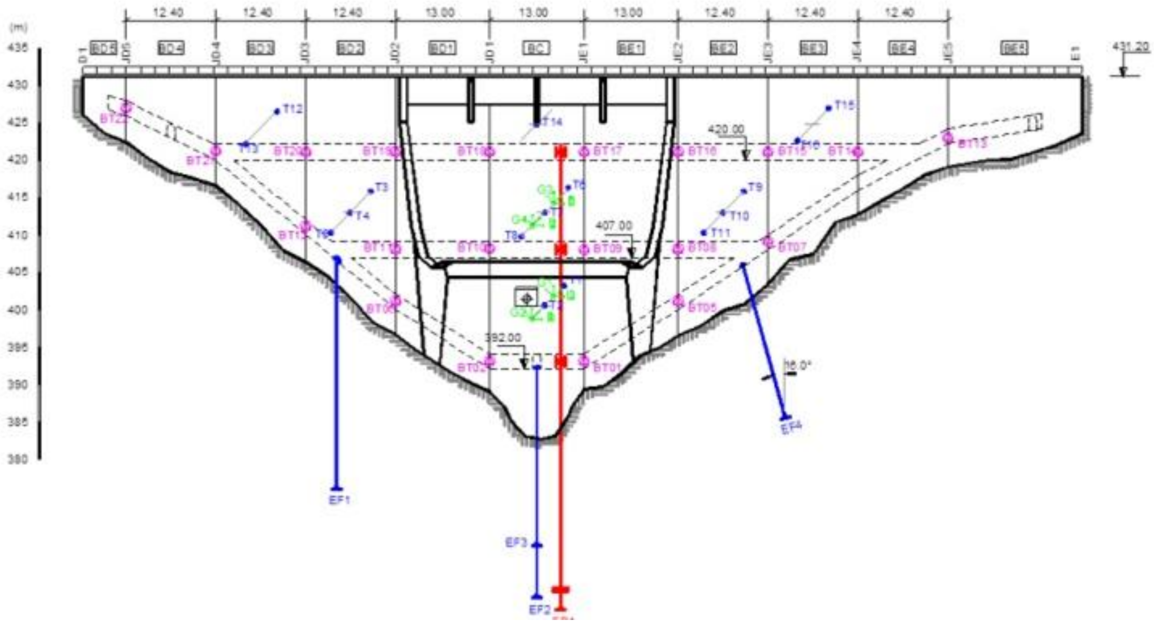

Figure 1 – Schematic representation of the instruments' location

The interpretation of the correlation of several parameters measured, in different physical locations of a structure, can be used to validate the current state of that structure and predict its future behaviour under specific and controlled conditions [1]. This is a key factor to detect potential anomalies and to be able to make decisions on time, reducing the risk of failures with catastrophic consequences. In the case of concrete dams, for example, their behaviour is continuously monitored by instruments (e.g., plumb lines, piezometers) installed in strategic points of the dam [2] [3], which can typically range from hundreds to few thousands of instruments or sensors. Figure 1 depicts a schematic representation of instruments installed in the dam structure.

The related raw data, usually known as “readings”, is collected manually by human operators or collected automatically by sensors. These readings are transformed, by specific algorithms, into engineering quantities (physical actions that can be used to assess the behaviour of the structure as, for example, a tension or a relative displacement). Actually, the term “reading” does not clearly correspond to raw data, since a reading is already a transformation from the raw data. For instance, an electrical instrument like an extensometer might provide raw data as a voltage (mV), which is then converted by a reading instrument (or by the sensor) into a resistance and a resistance relation, which are finally converted into an extension (engineering quantity). This monitoring information includes, essentially, instrument properties, readings and engineering quantities. Figure 2 depicts on the left a manual instrument used for measuring the displacement between concrete blocks, and on the right its automatic counterpart already installed in the structure.

Figure 2 - Measurement instruments

As a first assumption, one can consider that the main reason to preserve data is to preserve its value, as an asset. Consequently, it does not make sense to preserve valueless data. However, to determine and assess the value of data is a difficult and error-prone task. On the other hand, it could be an error to consider that data that cannot be used today will have no value in the future. For instance, today’s technology allows the simulation of mathematical models with a much higher resolution and volume of simulated data that was not possible a decade ago. From this perspective, we assume that the preservation of data concerning the safety of large civil engineering structures is crucial, since:

- observational data is unique and impossible to recreate
- complies with legal requirements or contracts established with third-parties
- allows the re-use of data for new research
- reduces costs (e.g., the retention of expensively generated data is cheaper to maintain than to re-generate) [4].

2.2 Scenario Host

The Portuguese regulations [5] state that the National Laboratory for Civil Engineering is responsible for keeping an electronic archive of data concerning the dam safety. Thus, the preservation of this data is a legal obligation. Moreover, that obligation defines the duties of the different parties involved in dam safety, namely the dam owners, the dam safety authority and the dam engineers and builders.

As a consequence, several entities are compelled to share data, and thus must face interoperability and preservation issues when dealing with heterogeneous sources of information [6].

Currently, LNEC uses a modular information system (*GestBarragens*) that provides components to manage dam observations, visual inspections, physical models and mathematical models. It also supports the management of technical documents and provides a set of exploitation tools, in the form of tabular and chart reports, graphical visualization of geo-referenced information, among others. However, the GestBarragens system was not designed for preservation purposes. Indeed, it supports the operational procedures to manage information concerning the dam safety, but does not assure the preservation of this information. It is a web-based system developed on the top of the *.NET framework*, where the underlying data is stored and managed in an *Oracle 10g* database. It uses a SOAP interface to provide and expose exploitation services as well as multiple ingest services. Figure 3 depicts the user interface of the *GestBarragens* system.

Figure 3 – The GestBarragens user interface

Table 1 summarizes an example of the data concerning the dam safety of a concrete dam. Currently, LNEC supports 32 different types of instruments with manual data acquisition and 25 different types of automatic monitoring instruments (implemented with sensors). Both the number and type of instruments installed in a specific structure depend on the stage of the structure's life and on a few hundred to thousand specific parameters that affect its behaviour. Currently LNEC monitors about 80 concrete dams, generating an average of 264,000 records per day that have to be processed and preserved.

Table 1 - Data registered for a representative concrete dam

Data Stage	***Description***	***# per day***	***Format***	***Notes***
Raw	Depend on the instrument type (e.g. voltage)	Currently discarded	Proprietary to the sensor	This information is currently discarded by sensors and not registered during manual acquisition
Processed readings	Transformed from raw data	Aprox. 3300 rows	.xls, .mdb, PDT, ascii	Sensors register data in .xls or .mdb and access a web service to send this information to LNEC. Manual acquisition can be registered into a PDT and automatically sent to LNEC or inserted via web interface or text file
Calculated engineering quantities	Calculated from readings	Aprox. 3250 rows	Oracle database	Algorithms to filter, clean and calculate engineering quantities are implemented as Oracle stored procedures (PL/SQL)
Analyzed	Tables, graphs, gis, mathematical simulations	Varies	.html, .xls, .pdf, .dxf (CAD), .xml	Uses several tools, including reporting tools and a geographic information system

3. SCENARIO 2 – ACQUISITION AND PRESERVATION OF SCIENTIFIC WORKFLOWS

3.1 Scenario Description

Scientific workflows are valuable knowledge assets which capture pieces of the scientific process and know-how that matter to preserve. Workflows are normally a result of collaborative design made by multidisciplinary teams, requiring skill to design, and its creation is often difficult and expensive [7].

The preservation of workflow is valuable mainly due to its potential reuse in the production of new knowledge: by scientists working on the same research group and on the same research project, involving sometimes the modification of some parameters or input data; by scientists working in other research groups belonging to other institutions, but performing similar research; or by scientists working in other areas of research which can benefit from research done in a different area.

Due to the value given to the re-usage of scientific workflows, provenance and authenticity are two dimensions that have a reinforced importance: provenance allows the tracking of the history and ownership of the experiment, and authenticity is a major factor in the validation of scientific experiments. These two dimensions are also important requirements of digital preservation, which helps to explain the potential of application of digital preservation knowledge and tools to help to dissolve existing barriers to the re-usage and sharing of scientific workflows. Those barriers are even bigger due to the fact that different scientific workflow systems are often employed by different research groups. Figure 4 depicts the interface of a system used for creating workflows (i.e., *Taverna*[1]).

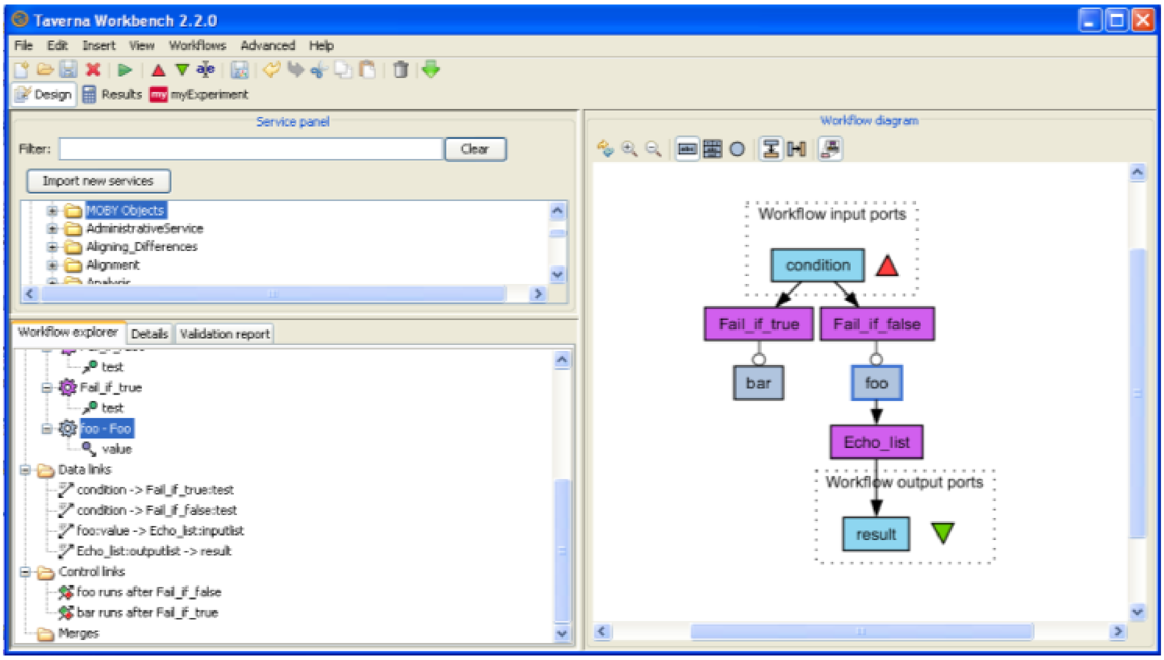

Figure 4 – Workflow creation using *Taverna*

[1] http://www.taverna.org.uk/

Workflows need additional information that allow and promote its sharing and reuse by scientists working inside or outside the organization. Examples of the kinds of additional data required include descriptions of function and purpose, example of input and output data, annotations, ownership, provenance, dependencies, among others.

Through the preservation of that information, knowledge can be preserved for future generations of scientists, which promotes one of the pillars of science making: the sharing of knowledge.

3.2 Scenario Host

The *myExperiment*[2] project is a virtual research environment to support scientists using scientific workflows. It was developed under the scope of the *myGrid*[3] research group, a multi-institutional and multi-disciplinary group involving different institutions such as the University of Manchester, the University of Southampton, the University of Oxford, among others. Figure 5 depicts a screenshot of the myExperiment interface.

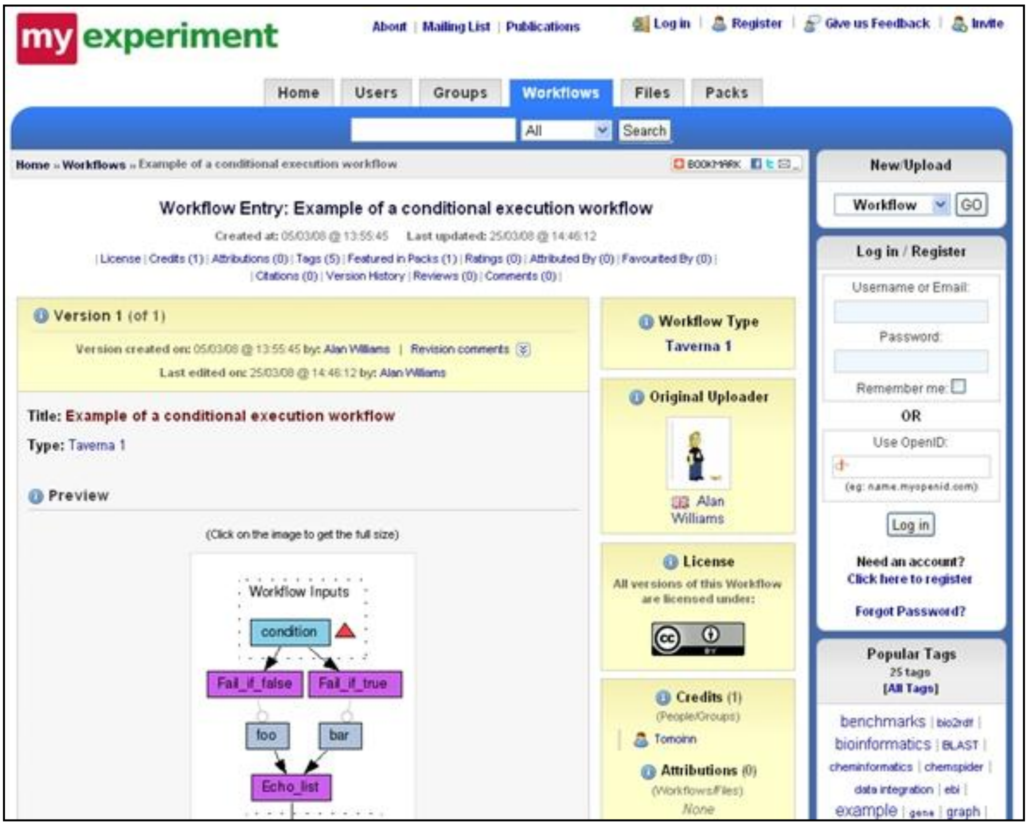

Figure 5 – *myExperiment* user interface

The *myExperiment* platform allows the publication and sharing, reuse, and repurposing of workflows, and carries several *web 2.0* features that make possible discussion around the workflows and associated objects, in a way independent of any workflow system. However, it does not provide preservation guarantees.

---

2 http://www.myexperiment.org/

3 http://www.mygrid.org.uk/

Several workflow formats are supported by the myExperiment environment, although the most usual format is *XSCUFL*, an *XML*-based format developed by the *myGRID* team. Figure 6 depicts an example of a workflow specification in the *XSCUFL* format.


```xml
<?xml version="1.0" encoding="UTF-8"?>
<s:scufl xmlns:s="http://org.embl.ebi.escience/xscufl/0.1alpha" version="0.2" log="0">
  <s:workflowdescription lsid="urn:lsid:www.mygrid.org.uk:operation:W0TIAJ5S3K0" author="Tom Oinn" title="Example
  <s:processor name="Fail_if_false">
    <s:local>org.embl.ebi.escience.scuflworkers.java.FailIfFalse</s:local>
  </s:processor>
  <s:processor name="Fail_if_true">
    <s:local>org.embl.ebi.escience.scuflworkers.java.FailIfTrue</s:local>
  </s:processor>
  <s:processor name="foo">
    <s:stringconstant>Foo</s:stringconstant>
  </s:processor>
  <s:processor name="bar">
    <s:stringconstant>Bar</s:stringconstant>
  </s:processor>
  <s:processor name="Echo_list">
    <s:local>org.embl.ebi.escience.scuflworkers.java.EchoList</s:local>
  </s:processor>
  <s:link source="condition" sink="Fail_if_true:test" />
  <s:link source="condition" sink="Fail_if_false:test" />
  <s:link source="foo:value" sink="Echo_list:inputlist" />
  <s:link source="bar:value" sink="Echo_list:inputlist" />
  <s:link source="Echo_list:outputlist" sink="result" />
  <s:source name="condition">
    <s:metadata>
      <s:description>Enter the string 'true' or 'false' here to show the conditional branching</s:description>
    </s:metadata>
  </s:source>
  <s:sink name="result" />
  <s:coordination name="bar_BLOCKON_Fail_if_true">
    <s:condition>
      <s:state>Completed</s:state>
      <s:target>Fail_if_true</s:target>
    </s:condition>
    <s:action>
      <s:target>bar</s:target>
```


Figure 6 – Workflow specification in XSCUFL

4. SCENARIO 3 – ACQUISITION AND PRESERVATION OF DATA AND SIMULATIONS IN PARTICLE PHYSICS

4.1 Scenario Description

Particle physics studies the basic constituents of matter. Research in these matters normally requires a technical infrastructure comprising particle colliders and detectors where particles are accelerated and made collide with each other or against fixed targets, or by studying particles of astronomical origin. All experiments require very complex and expensive infrastructure and for these reasons most particle physics experiments take place in the context of international collaborations.

The creation of a new experiment is usually a long process spanning many years, including the development of the particle detection apparatus, data acquisition systems, data processing and data analysis. Due to the specific characteristics of each experiment and the associated costs it is very unlikely that data obtained by a past particle physics experiment can be fully reproduced again by a new one. Therefore the data and the results from particle physics experiments are unique and must be saved for the future.

In the past it was wrongly thought that the potential of the data and results produced by an experiment was exhausted within the lifetime of the collaboration. However there are many cases where old data can be useful:

- New theories can lead to new predictions of physics effects that were not probed in the data when the experiment was running.
- Sometimes there is a need to cross check results from new experiments against results obtained by other previous experiments.
- The discovery of new phenomena in future experiments may demand that data from older experiments be analysed in search for things not yet known.
- New analysis techniques and Monte Carlo simulation models may create the opportunity to reprocess data and obtain higher precision results.
- New ideas for studies may appear in ranges of energy only available in old experimental data.
- Combined analysis by joining data from several experiments at once offers the possibility to reduce statistical and/or systematic uncertainties, or even to perform completely new analysis. This may require access to old data.

Moreover in particle physics it is estimated that the scientific production that can be attained from the data beyond the end of the experimental programme by continuing further analysis taking advantage of the whole data and better statistics represents 5 to 10 percent of the total scientific outcome.

However the complexity of the particle physics experiments is also reflected in the data, in the software and in the analysis process itself making the recovery of previous scientific work non-trivial.

Particle physics experiments are built to search for evidence of unknown physical phenomena predicted in theoretical models, or just to study more deeply certain known processes. Therefore there are two types of data, the data that is acquired from the experimental apparatus, also called raw data, and the data that is produced by Monte Carlo simulation according to the expected theoretical models.

The raw data has to be reconstructed to produce data with physics meaning (particle tracks). From the reconstructed data several types of summarized data are produced for analysis and from the summarized data each group or individual produces its own sets of data tailored to its own research and analysis tools.

Along the pipeline many types of other information are required, for instance calibration data, detector conditions, geometrical alignment of the detector, particle beam characteristics and many others. All these types of additional data add to the complexity.

The data is stored differently according to its nature and the way it is accessed resulting in many different data formats ranging from plain ASCII and binary files to databases of different types. The programs developed by the collaborations for the several steps of the pipeline produce data in different formats (raw, DST etc) that are frequently specific to each experiment. Both specific developed programs and generic tools are used for analysis. Again different data formats can be used depending on the tool. The final analysis including the generation of histograms is usually performed with tools such as ROOT[4] and/or in certain cases commercial software.

4.2 Scenario Host

The Portuguese Laboratory of Instrumentation and Experimental Particles (LIP) is a research laboratory in the fields of Experimental High Energy Physics, Astroparticles, Medical Physics, radiation detection instrumentation, data acquisition and data processing. Its research activities are developed in the framework of large collaborations at CERN and at other international organizations.

A typical scenario that can be observed in LIP is that of a PhD student or senior researcher performing small or medium collider or astroparticles experiments and analyzing data resulting from these experiments. Such an experiment will typically involve small data collections and might make use of several tools: (*i*) Experiment specific data processing software, which might be totally or partially developed by the researcher; (*ii*) Data analysis tools, which are used for analyzing the data produced in the context of the experiment; (*iii*) Simulation software for simulating the same experiments for comparing results; and (*iv*) other varied scripts.

The data that is used in such experiments is heterogeneous and might originate from diverse sources, and might include raw or reconstructed data, databases, and data resulting from simulations. Data resulting from such experiments might as well use several formats, some of them exclusive for the experiment. Such experiments result in several publications describing the results obtained.

[4] http://root.cern.ch/drupal/

5. PROTOTYPE

This prototype follows the same approach as the other two ISPs, being aligned with the SHAMAN Core Infrastructure ideal, and the SHAMAN Information Lifecycle (*Creation*, *Assembly*, *Archival*, *Adoption* and *Reuse*) [8].

The SHAMAN Core Infrastructure provides a top-down architecture approach which supports the delivery of those capabilities in a way that it can effectively address different preservation scenarios, being adaptable, reusable, and extensible. Governance capabilities are supported through the enforcement of policies; Business capabilities are supported through the ability to address preservation use cases through the deployment of business services; and Support capabilities are supported through the provision of services supporting the concretization of business services.

Figure 7 represents the approach made to the SHAMAN Core Infrastructure. The process specification includes the transformation of the preservation policy descriptions into process specification diagrams. This is supported by generic and standard process modelling tools like *Enterprise Architect* or *Eclipse BPMN* [cf. D5.1 section 3.3.3]. Using this approach, it is possible to generate standard machine readable process definitions, which can be interpreted by a service orchestration engine.

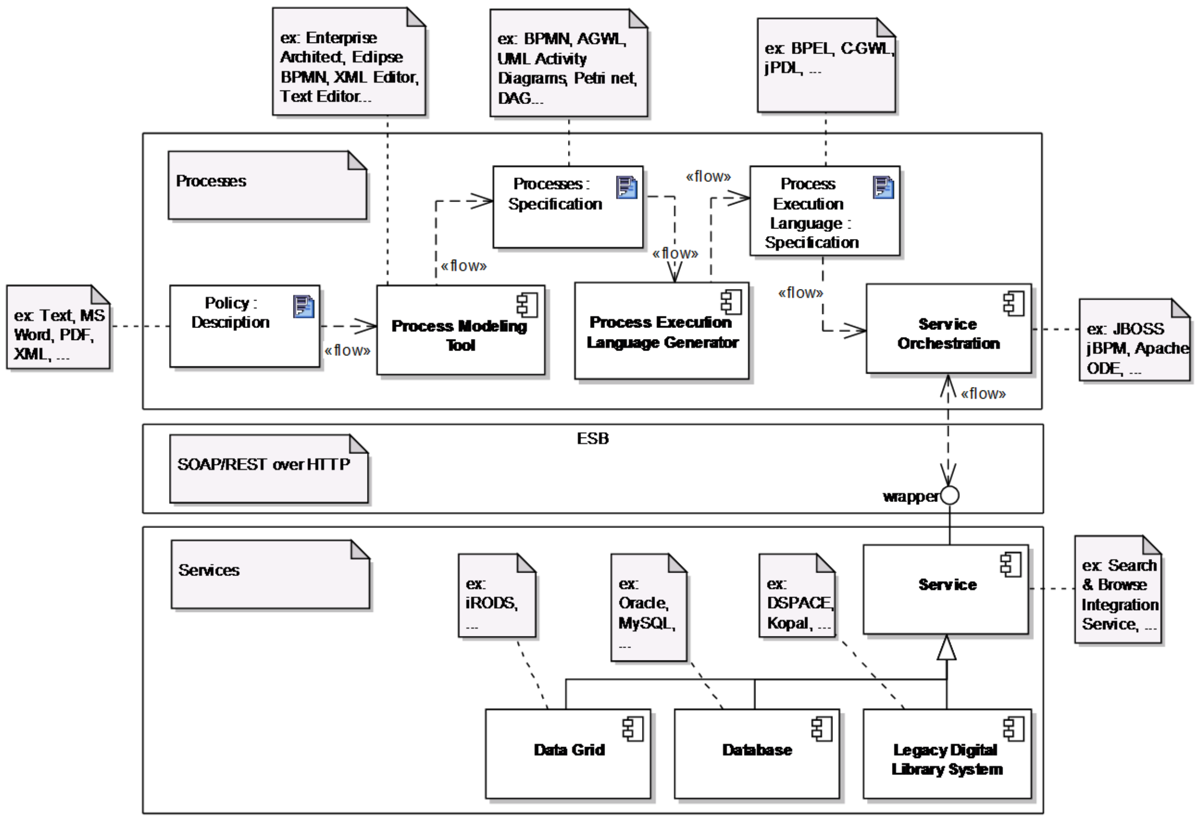

Figure 7 - Conceptual Architecture of the SHAMAN Core Infrastructure

The rationale for the proposed approach is based on the separation of definition of policies from the specification of services to execute those policies (service and technology independence).

The connection between policies (process level) and services is intermediated by an *Enterprise Service Bus* (ESB) middleware, to abstract the complexity of the implementation of deployed services. As a result, the implementation and deployment of services is independent from the process layer. It is just required to meet the specification of interfaces defined by the ESB layer.

The separation of concerns supported by this conceptual architecture follows the guidelines of the SHAMAN Reference Architecture (SHAMAN-RA). Moreover, the SHAMAN-RA was conceived to be able to effectively address changes in the context (e.g., Stakeholder’s concerns, drivers, requirements). This type of changes might impact the preservation policies and produce changes on the service specifications. The alignment of the services to the preservation policies is supported by the adaptability of the proposed architecture.

5.1 Process Layer

The proposed solution elaborates on the creation and description of information packages to control the media obsolescence vulnerabilities that occur when the representation format becomes obsolete and unable to be rendered, even if the "bit stream" survives over time.

Since the information package is composed by the data to be preserved, along with its contextual and representation information, a network of object have to be aggregated to create a meaningful object in whatever e-science context. The *Metadata Encoding and Transmission Standard* (METS[5]) is a widespread metadata representation to encode structural metadata in *XML*. The use of *METS* provides and extensible way to represent the aggregations required by the scenario described.

On the other hand, to address the management of schema representations and their dynamic nature (new or updated schemas to represent the same information), it is critical to manage metadata that describes the information representation. This is not a new requirement in the community, where, for instance, previous work developed the Metacat framework [9], which is able to store, retrieve and transform *XML* documents managed stored in a relational database. In this work, we use the concept of *Metadata Registry* (*MDR*), which was conceived to represent a system that allows the management of multiple schemas (not limited to *XML*) and the export of information about the schema. It also supports the creation and management of mappings between different schemas. This concept is formalized by the ISO 11179[6] series of standards.

For demonstrating the preservation of data in these scenarios, we developed a *Service Oriented Architecture* (*SOA*) solution, as shown in Figure 8. Our proposal comprises services for: harvesting the data to be preserved from a configured data provider, acquiring a description of the schema representation, packaging the data together, and ingesting the data package into the archival system. Such a solution is controlled by a service orchestrator (**Service Orchestration** component), which orchestrates the following services, controlling their flow and execution:

- **Data Extractor**: Extracts data from the source system, according to the parameters defined by the Assembly Orchestration. To support the dynamic nature of e-science data, it has the option to define the time window for data extraction, full extraction, incremental extraction and the type of items to extract. The recursive use of full data sets uses more space, while incremental data sets require the recomposition of data sets on access.

[5] http://www.loc.gov/standards/mets/

[6] http://metadata-stds.org/11179/

- **Metadata Registry (MDR)**: Supports the registration and management of multiple data schemas, addressing “the semantics of data”, “the representation of data”, and “the registration of the descriptions of that data”. It also supports the creation and management of the mappings between data schemas, as well as the export of both schema and mapping information.

- **Data Aggregator**: the METS schema is used to “wrap” all the information, acting as structural metadata. It is configured by a XML file that defined the nodes composing the object to aggregate. The proposed solution, defines objects as a hierarchical tree of nodes, where each node can have several children, and be defined by a set of files. This structure is represented in Figure 9.

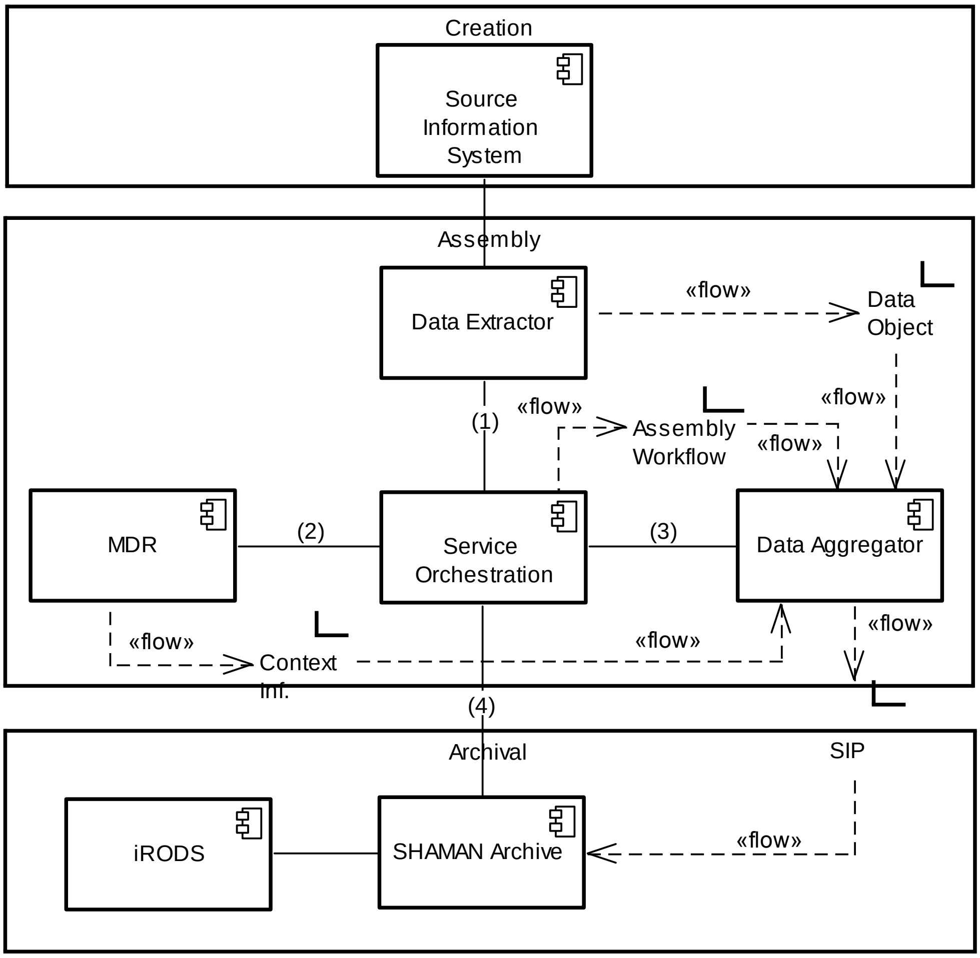

Figure 8 – Overview of the solution

The Assembly Orchestration component starts by (1) harvesting data from the source system through the Data Extractor component, specifying both the type of export (time window, full, or incremental from last export) and the item or list of items, and (2) acquiring the related schema information by requesting it from the MDR, depending on the type of data exported in the previous step. It continues, by (3) requesting the generation of a METS file to package the information and (4) submitting the package into the archive (using the ingest Web Service of the SHAMAN archive). When the submitted package enters the SHAMAN archive, it is then managed as common information packages, as those constructed for typical data objects like images or text. In the case of the SHAMAN archive, an information package is encoded in plain zip and includes an OAI-ORE manifest[7] to aggregate resources contained in the information packages (e.g., information content, preservation metadata).

When the data is accessed from the archive for future use (adoption), the information package is self-contained, in the sense that it includes, not only the preserved data, but also all the information required to render this data (structural information provided by the schema representation extracted from the MDR), in addition to the context information required to understand the data itself.

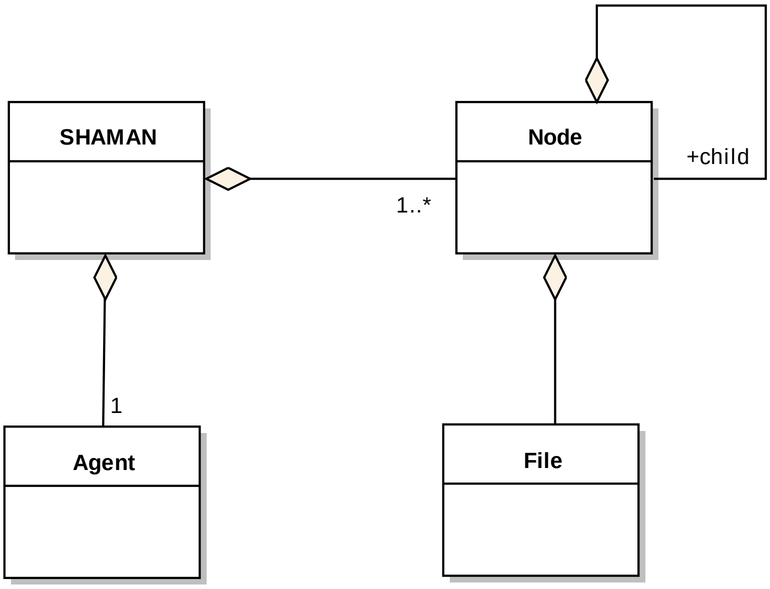

Figure 9 – Data Aggregator configuration

Finally, the integration of a MDR and the decoupling between data, its schema representation and the context information in the information packages, support the use of migration techniques inside the archive. In fact, migration is one of the most effective techniques used in digital preservation to avoid the obsolescence of data representations/formats. For observational data, mappings between schemas supported by the MDR provide an effective tool to migrate from an obsolete schema representation to an updated representation. Note that migrations can be often lossy, making it critical to plan when, how and what to migrate [10].

5.2 Implementation of the Process Layer

To demonstrate this prototype, we developed a Java 2 Platform Enterprise Edition (J2EE)[8] application, composed by a set of services that create complex Information Packages from the source information systems, and submit these Information Packages into the SHAMAN archive (Assembly phase of the digital preservation life-cycle). The proposed solution is intended to be generic, and was applied to the sensor data, scientific workflow and data and simulation in particle physics.

[7] http://www.openarchives.org/ore

[8] http://java.sun.com/j2ee

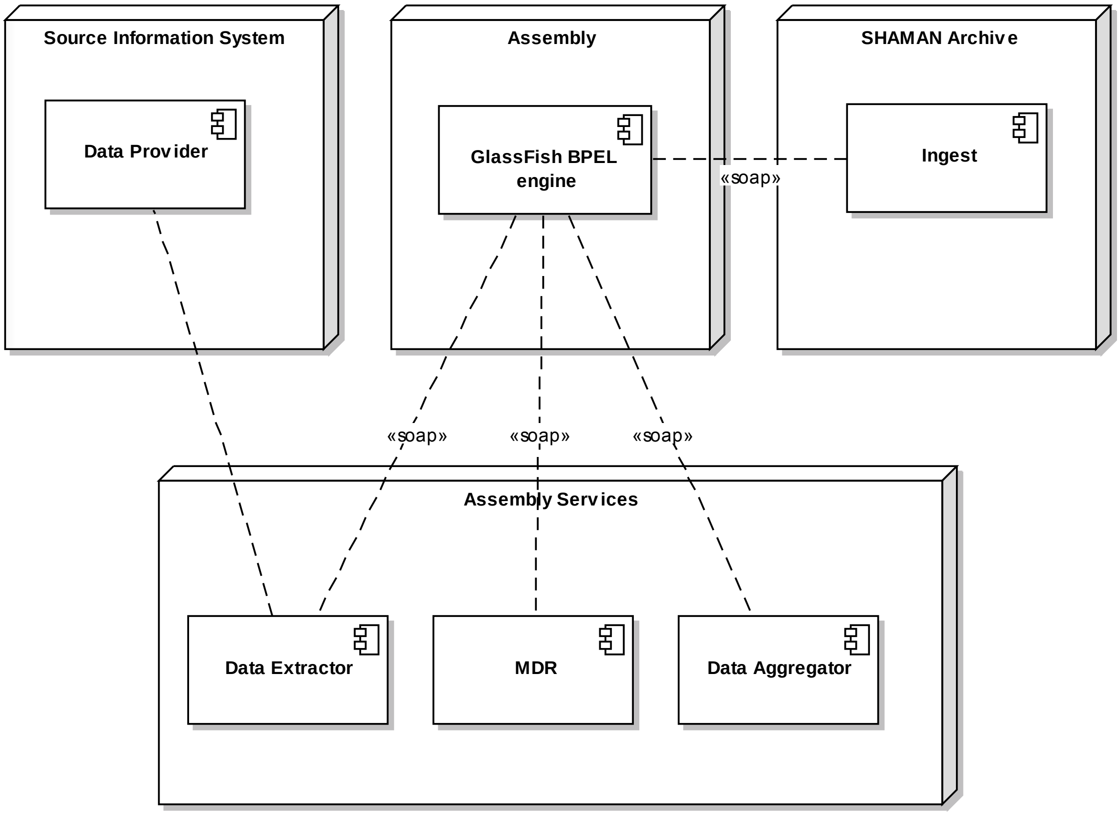

Figure 10 – Prototype deployment diagram

Those services were implemented as *Java Web Services* and are controlled through the *GlassFish Open ESB BPEL*[9] engine. This provides a flexible tool to schedule and parameterize the creation and submission of digital objects.

Figure 10 details a generic deployment of the developed components. The Assembly Orchestration component orchestrates the involved services, controlling their flow and execution. The Data Extractor component is specific to each of the analyzed scenarios, while MDR requires previous registration of the involved schemas and the Data Aggregator is controlled through the *data aggregator configuration* (see Figure 9).

The BPEL service orchestration can be defined with any tool that supports BPEL edition. In the scope of this prototype, the *NetBeans IDE*[10] was used. Figure 11 depicts the creation of the *Assembly Orchestration* using *NetBeans.* Finally, in order to control and monitor the execution of both orchestration and SHAMAN services, we instantiate a *BPEL Monitor* component as shown in Figure 12.

---

[9] http://openesb-dev.org

[10] http://netbeans.org/

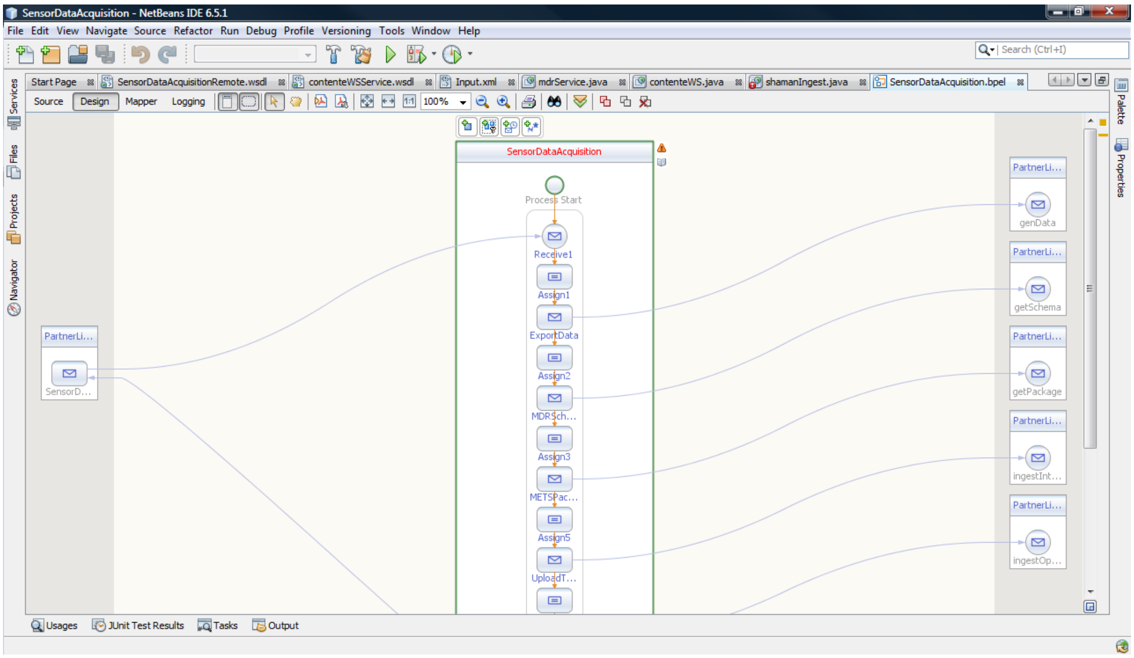

Figure 11 – NetBeans BPEL designer

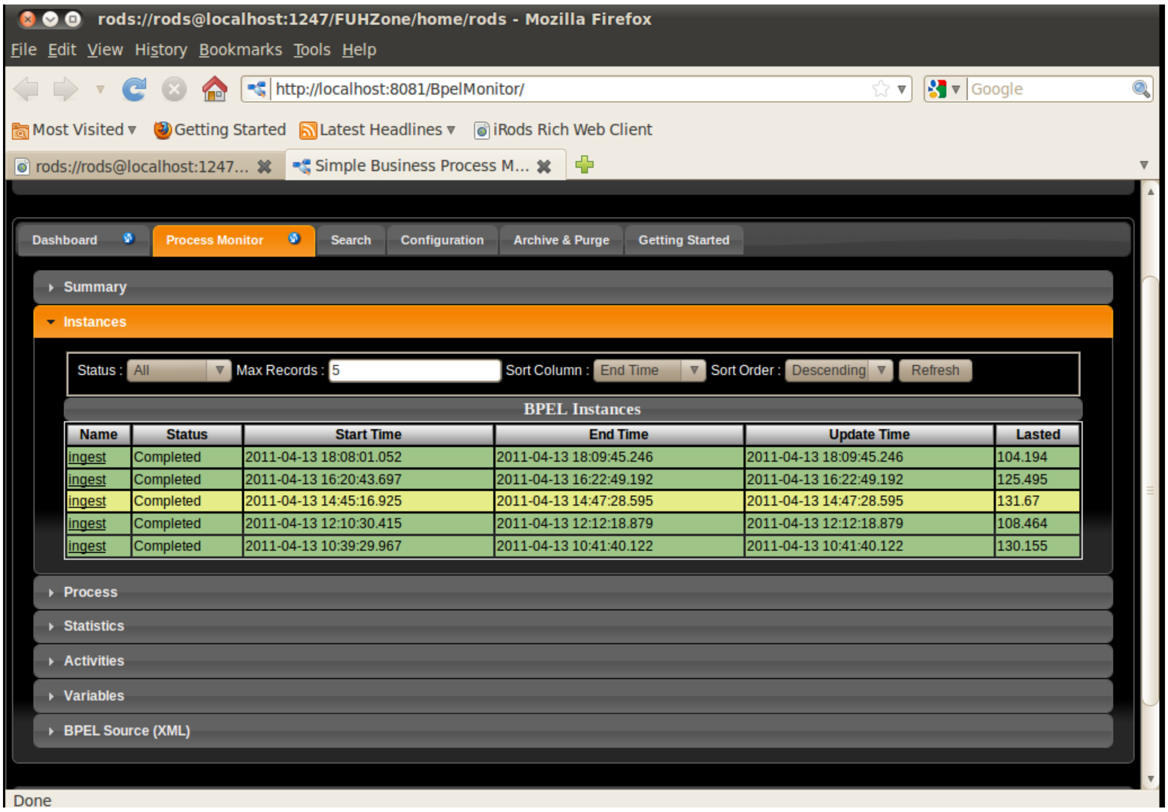

Figure 12 – BPEL monitor

5.3 Infrastructure Layer

The SHAMAN project developed an archival infrastructure that follows the OAIS reference model and uses the iRODS[11] data grid as storage substrate. The work presented in previous subsections relies on this infrastructure to address the digital preservation risks related to media faults, process and infrastructure vulnerabilities. However, the domain addressed by this ISP has special characteristics concerning the issues related to the volume of data and its imminent deluge.

Data grids in general offer the particularity of having, among other desired characteristics, a replication feature. Moreover, the fact that grids belonging to different institutions can engage in federations with other grids, while retaining full administrative control of their domains can be a useful feature in preservation scenarios, since additional storage space is obtained this way. This would allow a preservation system to take advantage of "borrowed" resources belonging to another administrative domain.

The iRODS data grid system supports both replication features and the creation of federations. However, it requires customization in order to take advantage of these features, namely through the micro-service mechanism. In addition to this, the federation feature of iRODS only allows a limited control of the resources and data of remote federated grids, due to each grid having its own administrative domain. For instance, access to data in a remote grid is possible for an authenticated user, but writing new data or updating existing data is a limited feature, requiring some tweaking. See [12] for more details on the composition of rules and micro-services.

This allows taking advantage of two types of scenarios: (*i*) grids *exclusive* for preservation, which comprise machines dedicated to running the data grid exclusively for digital preservation, which are likely to be under administration of the data owner; and (*ii*) grids *extended* for preservation, in which existing grid clusters, initially created for data processing, can be federated through the installation of an iRODS instance and extended for preservation. Their spare disc space, CPU, and bandwidth can be used to store data according to the preservation requirements.

We can consider that we have two kinds of abstract actors: a Business Administrator, which is responsible for the creation and enforcement of replication rules and might not have strong technical skills, and a System Administrator, which is responsible for the administration and maintenance of the technical aspects of the system, and thus of replication.

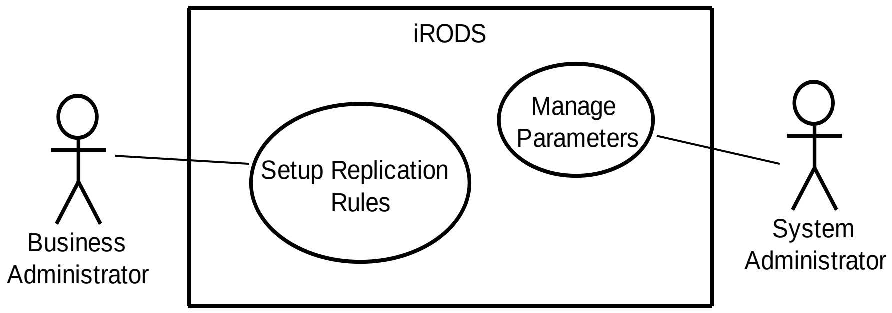

Figure 13 - Use case diagram of the proposal

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

We propose an easy to use interface for managing replication rules so that business administrators without any specific technical skills can manage replication rules. That interface would support two use cases: *Setup Replication Rules* and *Manage Parameters*, which are represented in Figure 13.

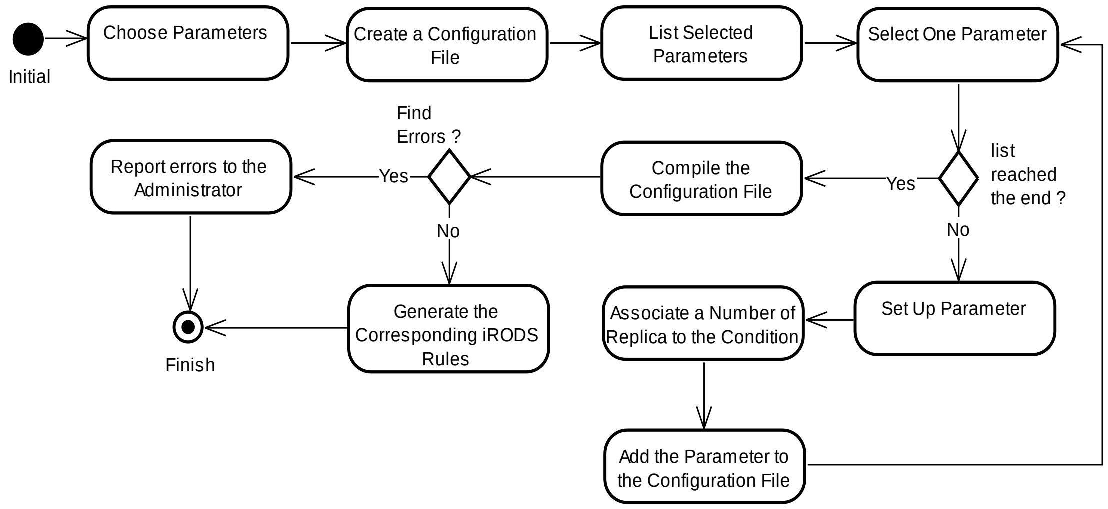

Figure 14 – Sequence diagram of the *Setup Replication Rules*

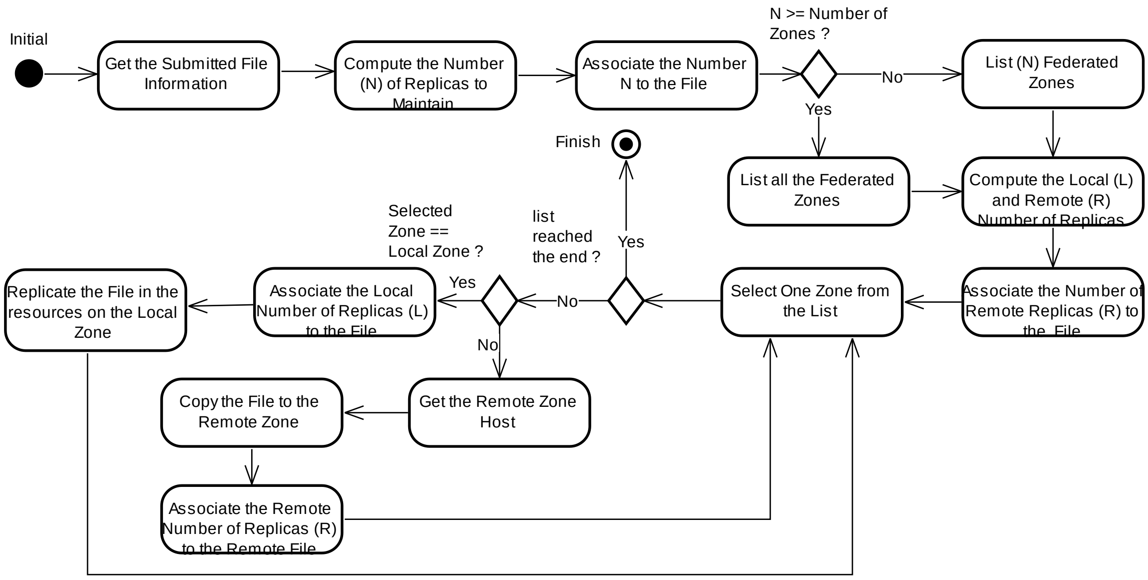

Figure 15 - Activity diagram of the *Replication Service*

In the *Setup Replication Rules* use case, the business administrator is able to create, in an easy way, specific rules based on determined parameters (e.g., file format, file size, user identity, etc.), in order to determine the minimum number of replicas to maintain of a file. The Parameters are listed as check boxes. The business administrator can select a list of parameters to be setup and then proceed to the configuration. When doing the configuration, for each parameter the business administrator can specify one condition (e.g., equal, different from, etc.) and enter a minimum number of replicas to be associated in case the condition is verified. The respectively configuration will be added to a configuration file. When finished, the configuration file is compiled and searched for any syntax errors. If no error is found, iRODS rules are generated and added to the rule base. Otherwise, the system will show the resultant errors. Figure 14 depicts the workflow sequence of this use case.

In the *Manage Parameters* use case, the system administrator can manage which parameters the business administrator can select and customize. The system can interpret certain parameters. The system administrator can choose from a previous list which parameters will be available to the business administrator. The system administrator can add or remove parameters to the actual list (the one that the business administrator can see), and apply the changes.

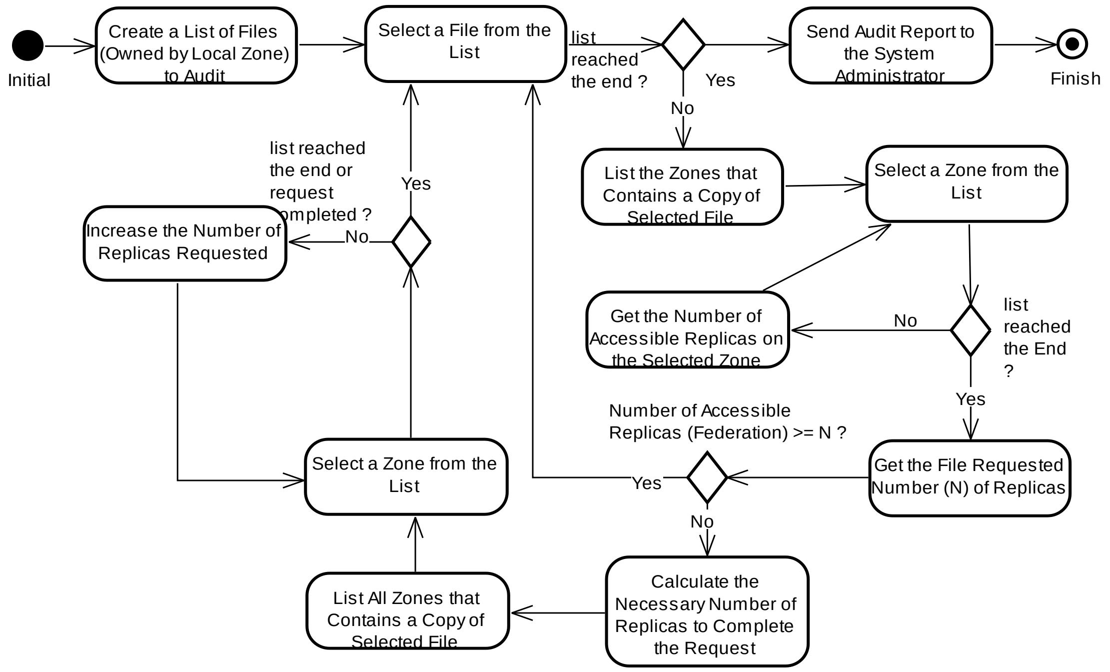

Figure 16 - Activity diagram of the *Audit of Locally Owned Files*

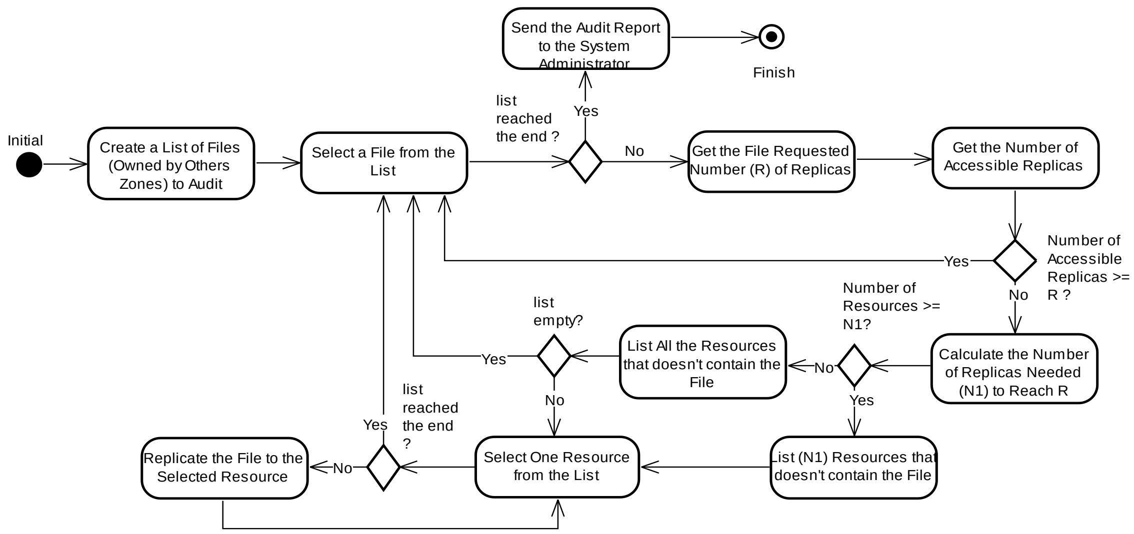

Figure 17 - Activity diagram of the *Audit of Remotely Owned Files*

Upon ingestion of a new file into the local iRODS deployment, the replication service checks if any of the rules configured by the business administrator apply to the file and, according to the rules, computes the number N of replicas to maintain of that file. That number is associated to the file through its metadata.

After that, the service checks the number of different federated grids. If the number of replicas is bigger or equal than the number of federated grids, a list of all federated grids is compiled. Otherwise, the first N federated grids listed are compiled into the list. Based on the list of federated grids available and on the number N of replicas, the number of replicas to be stored on the local iRODS deployment and on the remote federated grids is computed. The number of replicas to be stored in each zone, local or remote, is the integer division of N by the number of Zones. The remaining number of replicas until filling N is stored in the local Zone. The number of remote replicas R to be created is then registered and associated to the file metadata.

Then, one by one, each Zone contained in the federation list, will be used to create and store the replicas. If the local grid deployment is selected, the number of effectively created replicas L is registered and associated to the file metadata. If it is the case of a remote Zone, the file is copied and the number of remote replicas R that file should have is associated to that copied file. The file is not directly replicated, since the federation configuration does not allow the direct replication of files to a remote Zone. The file has to be copied and the replication has to be executed by the audit service running in the remote grid, which we will explain in the next section. Also, when copying a file to a remote zone, the associated metadata is not copied, hence the association of the desired number of replicas R after the copy. A reference to the file copied to the remote zone is also maintained. Figure 15 depicts the activity diagram of the replication service.

The replica audit service functions at two different levels. The first level audits the number of replicas stored in the local zone and the replicas of copies of local files stored in remote zones. The second level audits the number of replicas in the local zone which are owned by other zones (in other words, files which have been copied to the local zone from a remote zone).

Concerning the first level, a list of files contained in the local zone and owned by the local zone is compiled. Then, one by one, each file is selected and list of zones containing a replica of that file is compiled and, for each zone of the list, the number of replicas effectively stored in the selected zone is determined. When the list of zones is fully processed, if the total number of existing replicas is smaller than the replica number N contained in the file metadata, the number of necessary replicas in order to get N replicas is calculated. Then, a list containing all the zones where a copy of the file exists is compiled, and the number of replicas is increased to be as close as possible to L, in case of a local zone, or to R in case of a remote zone. After all files have been verified, an audit report of the status of each file is compiled and sent to the system administrator, so that, in case of need, he can take informed decisions. Figure 16 depicts the activity diagram of the first level of the auditing service.

The second level of audit begins with the creation of a list of files contained in the local zone which are owned by remote zones. Then, one by one, each file is selected and the number of replicas that the file should have (R) is retrieved. The number of accessible replicas is retrieved and is compared with the desired number of replicas. If the number of accessible replicas is smaller than R, than the number of replicas needed in order to reach R is calculated. If the number of existing resources is smaller than the number of replicas needed to reach R, a list of all the resources that do not contain a replica is retrieved. If the list is empty, then the issue is registered for the audit report and another file is selected for verification. Otherwise, files are replicated throughout available resources. After all files have been verified, an audit report is compiled and sent to the system administrator, so that, in case of need (e.g., add more storage resources) he can take informed measures. Figure 17 depicts the activity diagram of the second level of auditing.

5.4 Implementation of the Infrastructure Layer

The interface for the composition of rules was implemented using HTML, PHP, and Javascript. For the communication with the iRODS server for loading the rules into the system, the phyton API for iRODS – PRODS - was used. The currently supported replication parameters are presented in Table 2.

Table 2 – Supported replication parameters

Parameter	Configuration entry
File format	File format
File size	Condition (=, >, < , >= , <=) File Size
User	Username
Submission date	[Start_date, End_date]
Resource	Resource name
File name	File name

After authentication, the user is guided through the configuration, being able to select for each parameter the entry values that will be added to the condition. For instance, when configuring a rule for the file format parameter, the format can be selected from a predefined list. If the format does not exist the user can add it in the moment. Figure 18 depicts the selection of a file format when configuring a rule.

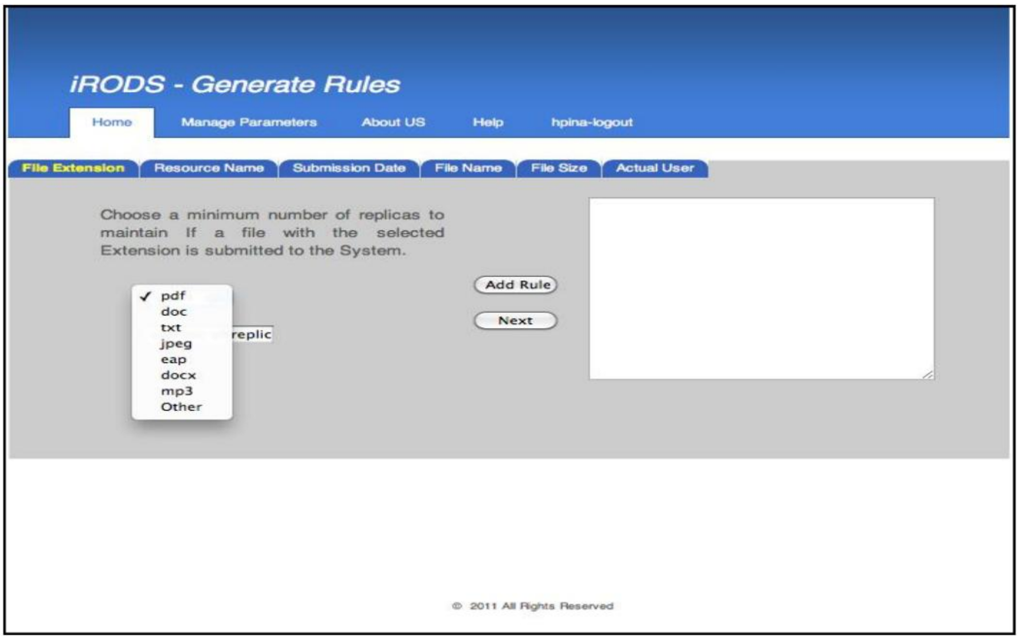

Figure 18 – Composing a rule based on the file format parameter

When the composition of rules is finished, a XML configuration file is generated and is directly processed by a compiler. The compiler is written in C. For each replication parameter it should verify the syntax and generate the corresponding iRODS rules. The compiler output is a set of iRODS rules as an iRODS rule base file so it can be included in the iRODS installation.

Both *Replication* and *Audit* Services are implemented using the rule mechanism and workflow capabilities provided by iRODS. The services are defined as a set of actions within a rule. The actions are composed by a set of micro-services. For the development of those micro-services we used the C language API provided by iRODS.

The *Replication* Service is triggered by a file submission. To access the file information we use available iRODS session variables. We send this information as input to the rules previously generated by the compiler. The execution of the rule results in a minimum number of replicas. This number is then associated to the file as a metadata attribute, using a micro-service already packed with iRODS.

When using resources located in remote zones for replication, the file has to be copied to the remote zone and the number of copies to maintain is associated with the file metadata. The remote zone then takes care of the replication. While we had to develop a set of micro-services to perform some of the operations involved, we also used micro-services already included in the installation.

The Audit Service is composed by two iRODS rules, one for auditing files owned by the local zone, and the other for auditing the files owned by remote zones. Again, some micro-services were developed to specifically for this purpose. Other micro-services were already packed with iRODS, such as the case of *msiSendMail*, which is used to send the audit report to the system administrator.

6. APPLYING THE PROTOTYPE TO THE SCENARIOS

6.1 Scenario 1 - Acquisition and Preservation of Sensor Data

The preservation of dam safety sensor data raises several challenges because of the data and process characteristics. First, data is not static (a data set is continuously increasing). Second, since new sensors (with different characteristics and results) have to be accommodated in the future, new data representations must be handled. Third, the representation of a dataset can evolve in the future (new devices can use different representations to store the same data), limiting the ability to understand the same type of data, as well as relating the same type of data when it was captured by devices using different data representations. Finally, the nature if complex and interlinked objects composed by datasets and their representation (an isolated dataset is useless to interpret the structural behaviour).

In order to control the complexity of data representations, some communities developed their own metadata initiatives as, for instance, the Ecological Metadata Language (EML[12]), or the Federal Geographical Data Committee (FGDC[13]). Yet, there will never be a unified metadata schema for all possible data. Thus, in a scenario that is not covered by current metadata initiatives, or when the information can be represented in heterogeneous schemas that can continuously change (like the sensors used in the civil engineering domain), the use of standard languages to describe data representations [11] is an expected solution. The use of a Metadata Registry (*MDR*) is a way of implementing such a solution in an extensible fashion.

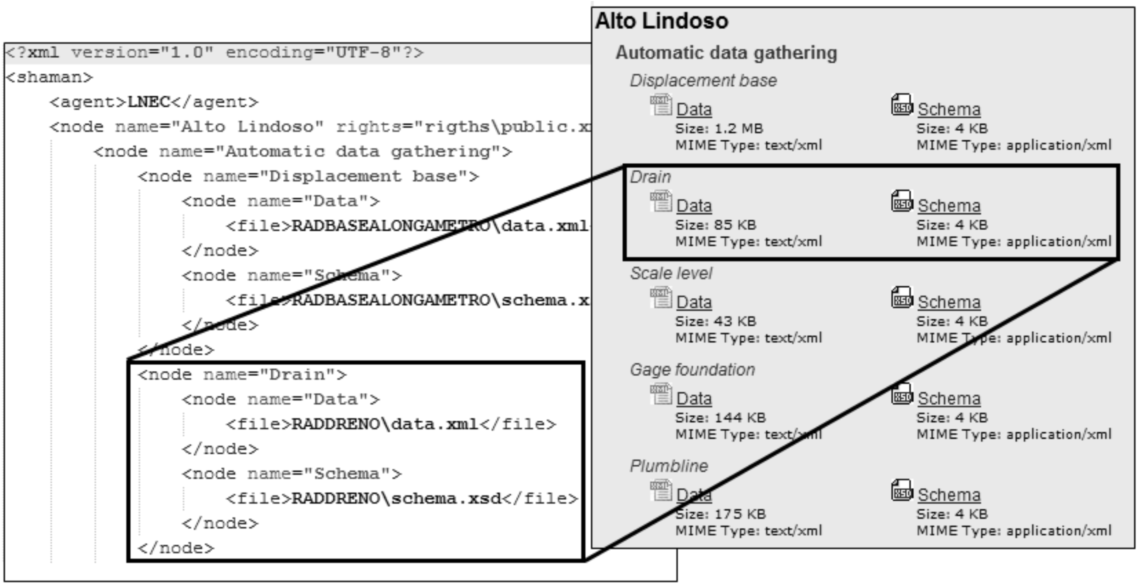

Figure 19 – LNEC's Information package HTML visualization

In the dam safety context, an information package is an interlinked object that must aggregate the sensor data, the information on the sensors that produced the data, as well as the description of the schemas used to encode them (it can include syntactic and semantic representation). In order to construct such a package, the components of the solution described in the previous section should be

[12] http://knb.ecoinformatics.org/software/eml

[13] http://www.fgdc.gov

configured in order to: (*i*) harvest the data objects from the GestBarragens information system; (*ii*) create a logical structure to organize the objects; (iii) register the used schemas in the *MDR* (Appendix A1.1 presents an example of schema representation for the subset of sensor data shown in Appendix A1.2).

A specific *Data Extractor* component was developed to interface with the *GestBarragens* system. It can be configured to extract a full set of data, a time window or an incremental extraction of data for a specific dam. At this step, the data sets and the sensor properties were generated, and their representation information is requested to the *MDR* component that produces a set of files to specify the schema representation. Finally, the *Data Aggregator* component is used to "wrap" the information generated by the *Data Extractor* (data about the characteristics of the sensor which produced the readings; and the data captured by these sensors); and by the *MDR* (schema information for the data generated by the *Data Extractor*) components and create a *METS* file to aggregate them (an example *METS* file can be seen in the Appendix A1.3). Additionally, this component also creates a set of *HTML* files to facilitate a human navigation/browse under the *METS* file. Figure 19 shows an example of a portion of this html for a digital object of the Alto Lindoso dam.

When dam safety data is accessed from the archive for future use (*Adoption*), the information package is self-contained in the sense that it includes not only the preserved data, but also all the information required to render this data. This includes structural information provided by the schema representation extracted from the *MDR* and context information required to understand the data itself, such as the type and characteristics of sensors, location, and data units.

6.2 Scenario 2 - Acquisition and Preservation of Scientific Workflows

For the second scenario, a similar approach to that of the previous scenario was taken, providing a unified way for retrieving the data from the *myExperiment* project and preserve it in the SHAMAN archive.

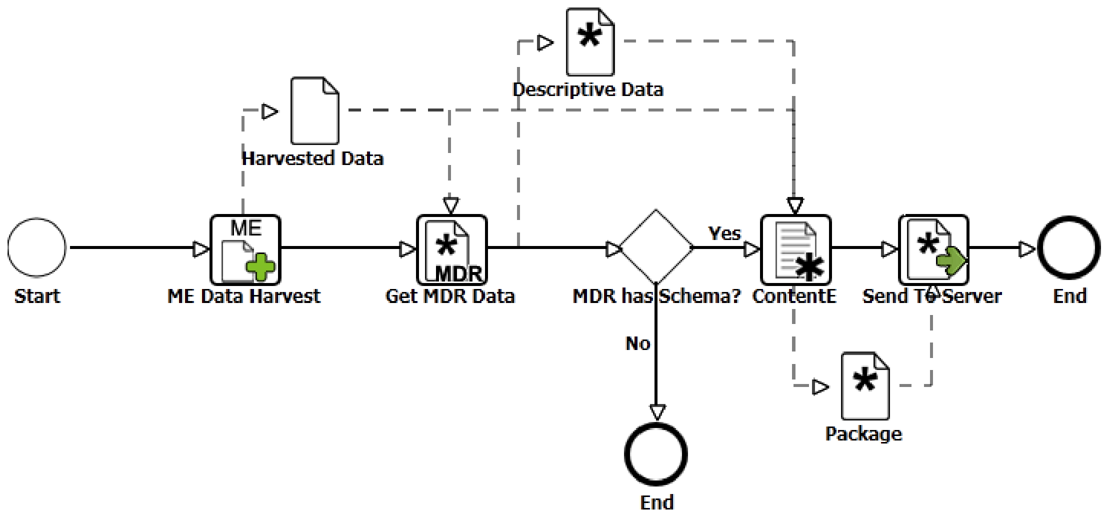

Figure 20 – Visual *myExperiment* harvest process definition in our solution's process modeler

As shown in Figure 20, we defined a process through our visual process modeler which represents the whole harvest process from the *myExperiment* project. In this process we start by harvesting the data using the REST services available in the *myExperiment* platform, allowing us to collect all the information related to a workflow. To achieve this, a specific *Data Extractor* was developed to gather information from the *myExperiment* system. This information contains the main workflow description and its visual representation (Appendix A2.1 contains an example of a workflow description in XML). Trough this description we were able to acquire the workflow's ratings, license, type, credits, tags, and user data responsible for its creation.

After collecting this data through its XML representation, we ask the *MDR* for the schema definition correspondent to the harvested data (Appendix A2.2 depicts an example of a schema of a workflow object). According to its existence we continue the process or end it. We end it because in order to manage all the schema definitions in an organized fashion they must be stored in the *MDR*. However, if the MDR has the schema, then an input file (Figure 21 – left side) is produced for the creation of the *METS* file within the *Data Extractor* component. Finally, and using the data generated by the *Data Extractor*, the *Data Aggregator* component is able to create a *METS* file representing the harvested data and their matching schemas (an example of a generated *METS* file is shown in A2.3). Additionally, this component also creates a set of *HTML* files to facilitate a human navigation/browse under the *METS* file. Figure 21 shows an example of a portion of this html for a digital object of the *scientific workflow 173* in the *myExperiment* system.

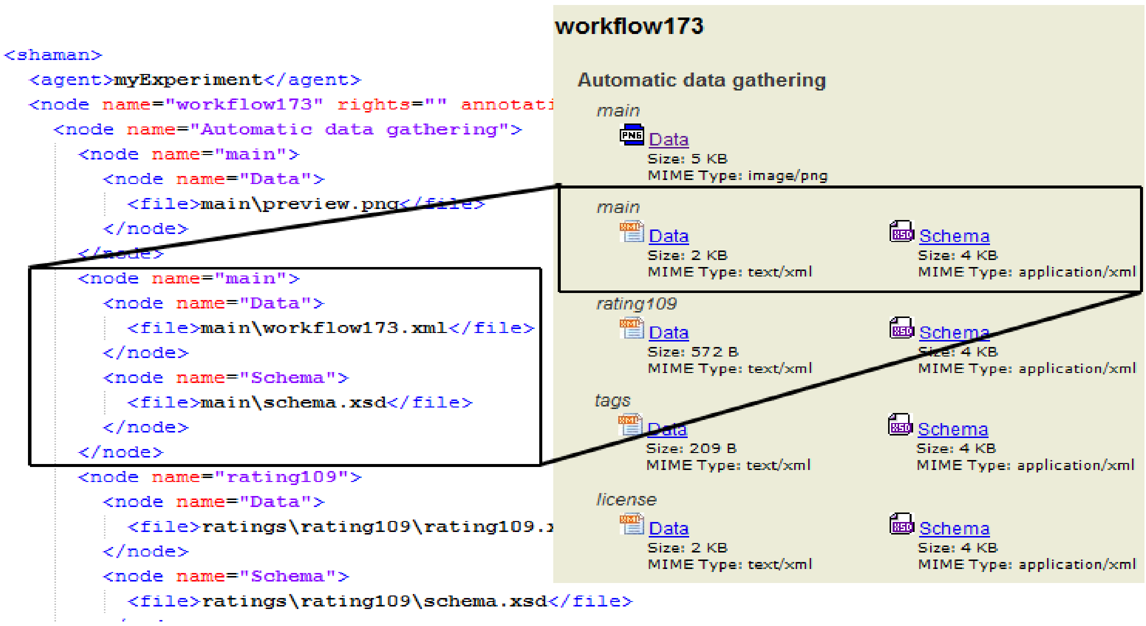

Figure 21 – *myExperiment's* information package HTML visualization

After the *Data Aggregator* component finishes the creation of the *METS* file, this information is sent to the SHAMAN archives in order to be preserved.

6.3 Scenario 3 - Acquisition and Preservation of Data and Simulations in Particles Physics

The high energy physics scenario presents challenges that are distinct from the previous scenarios, mainly due to the high heterogeneity of the data and software used, and the fact that the processes used for making science are completely ad-hoc. Moreover, due to the fact that information about experiments is highly dispersed through different systems, the data harvesting mechanism used in the previous scenarios is not usable since information is not centralized.

A possible workaround would pass by making the acquisition of data manual. In other words, the researcher producing an experiment would be the ultimate responsible for flagging the data, software and other necessary information, in a tool specifically designed for that effect. That tool would then provide this input to the data aggregator component.

Due to all these issues, which need a more profound study, and due to time constraints, the practical addressing of this specific scenario will be explored in more detail in the project TIMBUS[14], where it will be the basis of a specific demonstration case, and where INESC-ID will be involved.

Anyway, the analysis of this case in the scope of the SHAMAN project was fundamental to give us a relevant insight of this class of challenges, and thus played a fundamental role for the requirements and design of this ISP.

[14] http://timbusproject.net/

7. SUMMARY

This deliverable described the design and implementation of an experimental test bed for the acquisition and preservation of e-Science, which makes use of the SHAMAN Core Infrastructure and federated data grids for storage. Three different e-Science scenarios were described which address different types of e-Science data: sensor data, scientific workflows, and experimental data. A prototype was developed with a Service Oriented Architecture for flexibility and adaptability, and changes were made at the level of the SHAMAN Core Infrastructure so that it takes advantage of the federation configuration of data grids, leading to improved availability of data. The prototype was applied to two of the scenarios (sensor data and scientific workflows). The third scenario presented significantly different research challenges which will be addressed in a future project, due to time constraints.

Bibliography and References

[1] M. Wieland and R. Mueller. *Dam safety, emergency action plans and water alarm systems*. International Water Power & Dam Construction. January 2009

[2] ASCE – American Society of Civil Engineers. *Guidelines for instrumentation and measurements for monitoring dam performance*. 2000. ISBN 0-7844-0531-X.

[3] ICOLD - International Commission on Large Dams. *Guidelines for automated dam monitoring systems*. 1999.

[4] P. Lord and A. Macdonald. *E-Science curation report – Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision*. 2003.

[5] RSB. *Dam safety regulation, DL n.344/2007*. October 15th. DR, Lisbon, 2007 (in Portuguese).

[6] M. Franklin, A. Halevy and D. Maier. *From databases to dataspaces: a new abstraction for information management*. SIGMOD Record, 34(4):27-33. 2005.

[7] D. De Roure, C. Goble, S. Aleksejevs, S. Bechhofer, J. Bhagat, D. Cruickshank, P. Fisher, D. Hull, D. Michaelides, D. Newman, R. Procter, Y. Lin and M. Poschen. *Towards Open Science: the myExperiment Approach*. ConcurrencyComputat.: Pract. Exper. 22:2335-2357, 2010.

[8] H. Brooks, A. Kranstedt, G. Jaschke and M. Hemmje. *Modeling context for digital preservation. Smart Information Knowledge Management*, 197-226. 2010.

[9] M. Jones, C. Berkley, J. Bojilova and M. Schildhauer. *Managing Scientific Metadata*. IEEE Internet Computing 5: 59-68. 2001.

[10] C, Becker, H. Kulovits, M. Guttenbrunner, S. Strodl, A. Rauber, and H. Hofman. *Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans*. International Journal on Digital Libraries (IJDL), 10(4):133-157. December 2009.

[11] M. Westhead, T. Wen and R. Carroll. *Describing data on the GRID*. GRID Computing, 134-140. 2003.

[12] SHAMAN Consortium. *D5.3 Implementation of the SHAMAN Core Infrastructure*. 2011.

APPENDIX A. EXAMPLE DATA FILES AND SCHEMAS

A1. Scenario 1 - Acquisition and Preservation of Sensor Data

A1.1. Example data schema

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified"
targetNamespace="http://www.lnec.pt/schemas/sensor"
xmlns:sensor="http://www.lnec.pt/schemas/sensor">
  <xs:element name="dataset">
    <xs:complexType>
      <xs:sequence>
        <xs:element maxOccurs="unbounded" ref="sensor:sensor-record"/>
        <xs:element maxOccurs="unbounded" ref="sensor:data-record"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="sensor-record">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="sensor:ID"/>
        <xs:element ref="sensor:X"/>
        <xs:element ref="sensor:Y"/>
        <xs:element ref="sensor:Z"/>
        <xs:element ref="sensor:IDENTIFICACAO"/>
        <xs:element ref="sensor:CODIGOSIOBE"/>
        <xs:element ref="sensor:MANUAL"/>
        <xs:element ref="sensor:MODULO1"/>
        <xs:element ref="sensor:CANAL1"/>
        <xs:element ref="sensor:MAX1"/>
        <xs:element ref="sensor:MIN1"/>
        <xs:element ref="sensor:MODULO2"/>
        <xs:element ref="sensor:CANAL2"/>
        <xs:element ref="sensor:MAX2"/>
        <xs:element ref="sensor:MIN2"/>
        <xs:element ref="sensor:DATAINSTALACAO"/>
        <xs:element ref="sensor:DATADESACTIVACAO"/>
        <xs:element ref="sensor:DATALEITURAINICIAL"/>
        <xs:element ref="sensor:FREQLEITURA"/>
        <xs:element ref="sensor:COMENTARIOS"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="ID" type="xs:integer"/>
  <xs:element name="X" type="xs:decimal"/>
  <xs:element name="Y" type="xs:decimal"/>
  <xs:element name="Z" type="xs:decimal"/>
  <xs:element name="IDENTIFICACAO" type="xs:NCName"/>
  <xs:element name="CODIGOSIOBE" type="xs:integer"/>
  <xs:element name="MANUAL" type="xs:integer"/>
  <xs:element name="MODULO1" type="xs:integer"/>
  <xs:element name="CANAL1" type="xs:decimal"/>
  <xs:element name="MODULO2" type="xs:integer"/>
  <xs:element name="CANAL2" type="xs:decimal"/>
  <xs:element name="DATAINSTALACAO">
    <xs:complexType/>
  </xs:element>
  <xs:element name="DATADESACTIVACAO">
    <xs:complexType/>
  </xs:element>
  <xs:element name="DATALEITURAINICIAL">
```

```
    <xs:complexType/>
  </xs:element>
  <xs:element name="FREQLEITURA" type="xs:decimal"/>
  <xs:element name="COMENTARIOS">
    <xs:complexType/>
  </xs:element>
  <xs:element name="data-record">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="sensor:INSTRFIXO"/>
        <xs:element ref="sensor:DATARES"/>
        <xs:element ref="sensor:RESULTADO1"/>
        <xs:element ref="sensor:ESTADO1"/>
        <xs:element ref="sensor:MIN1"/>
        <xs:element ref="sensor:MAX1"/>
        <xs:element ref="sensor:RESULTADO2"/>
        <xs:element ref="sensor:ESTADO2"/>
        <xs:element ref="sensor:MIN2"/>
        <xs:element ref="sensor:MAX2"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="INSTRFIXO" type="xs:integer"/>
  <xs:element name="DATARES" type="xs:string"/>
  <xs:element name="RESULTADO1" type="xs:double"/>
  <xs:element name="ESTADO1" type="xs:string"/>
  <xs:element name="RESULTADO2" type="xs:decimal"/>
  <xs:element name="ESTADO2" type="xs:string"/>
  <xs:element name="MAX1" type="xs:double"/>
  <xs:element name="MIN1" type="xs:double"/>
  <xs:element name="MAX2" type="xs:decimal"/>
  <xs:element name="MIN2" type="xs:double"/>
</xs:schema>
```

A1.2. Example data

```
<dataset xmlns="http://www.lnec.pt/schemas/sensor">
        <sensor-record>
                <ID>341</ID>
                <X>0.2825153</X>
                <Y>0.14528537</Y>
                <Z>0.18533128</Z>
                <IDENTIFICACAO>PZ4-5</IDENTIFICACAO>
                <CODIGOSIOBE>4</CODIGOSIOBE>
                <MANUAL>14440</MANUAL>
                <MODULO1>1</MODULO1>
                <CANAL1>63.320805</CANAL1>
                <MAX1>0.13195592</MAX1>
                <MIN1>0.060967743</MIN1>
                <MODULO2>1</MODULO2>
                <CANAL2>63.169632</CANAL2>
                <MAX2>0.73938346</MAX2>
                <MIN2>0.62638235</MIN2>
                <DATAINSTALACAO></DATAINSTALACAO>
                <DATADESACTIVACAO></DATADESACTIVACAO>
                <DATALEITURAINICIAL></DATALEITURAINICIAL>
                <FREQLEITURA>1.7133925</FREQLEITURA>
                <COMENTARIOS></COMENTARIOS>
        </sensor-record>
        <sensor-record>
                <ID>342</ID>
                <X>0.47543406</X>
                <Y>0.7410047</Y>
                <Z>0.76579344</Z>
                <IDENTIFICACAO>PZ7-8</IDENTIFICACAO>
```

```
<CODIGOSIOBE>7</CODIGOSIOBE>
                <MANUAL>14443</MANUAL>
                <MODULO1>1</MODULO1>
                <CANAL1>59.118736</CANAL1>
                <MAX1>0.8170759</MAX1>
                <MIN1>0.71550655</MIN1>
                <MODULO2>1</MODULO2>
                <CANAL2>59.481487</CANAL2>
                <MAX2>0.99022555</MAX2>
                <MIN2>0.5151579</MIN2>
                <DATAINSTALACAO></DATAINSTALACAO>
                <DATADESACTIVACAO></DATADESACTIVACAO>
                <DATALEITURAINICIAL></DATALEITURAINICIAL>
                <FREQLEITURA>1.2941413</FREQLEITURA>
                <COMENTARIOS></COMENTARIOS>
        </sensor-record>
 (…)
<data-record>
                <INSTRFIXO>341</INSTRFIXO>
                <DATARES>2009-11-25 16:00:00.0</DATARES>
                <RESULTADO1>3.3287005</RESULTADO1>
                <ESTADO1>Alto</ESTADO1>
                <MIN1>0.42179298</MIN1>
                <MAX1>2.4863694</MAX1>
                <RESULTADO2>3.0451293</RESULTADO2>
                <ESTADO2>Alto</ESTADO2>
                <MIN2>0.82892954</MIN2>
                <MAX2>2.4320564</MAX2>
        </data-record>
        <data-record>
                <INSTRFIXO>342</INSTRFIXO>
                <DATARES>2009-11-26 00:00:00.0</DATARES>
                <RESULTADO1>3.4385567</RESULTADO1>
                <ESTADO1>Alto</ESTADO1>
                <MIN1>0.46555614</MIN1>
                <MAX1>2.225185</MAX1>
                <RESULTADO2>2.7541509</RESULTADO2>
                <ESTADO2>Alto</ESTADO2>
                <MIN2>0.15687603</MIN2>
                <MAX2>2.509763</MAX2>
        </data-record>
</dataset>
```

A1.3. Example METS file

```
<?xml version="1.0" encoding="utf-8" standalone="no"?>
<mets OBJID="shaman1t10" LABEL="Alto Lindoso" TYPE="ContentE v.2.9"
xmlns="http://www.loc.gov/METS/" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:bndTechInfo="http://schemas.bn.pt/tech-info"
xsi:schemaLocation="http://www.loc.gov/METS/ http://schemas.bn.pt/mets/v1.5/metsv1.5.xsd">
 <metsHdr CREATEDATE="2011-04-13T17:07:01" LASTMODDATE="2011-04-13T17:07:10"
RECORDSTATUS="COMPLETED">
  <agent ID="agent.CONTENTE.METSAgent.ag16" ROLE="CREATOR" TYPE="ORGANIZATION">
   <name>LNEC</name>
  </agent>
 </metsHdr>
 <amdSec>
  <techMD ID="tech.CONTENTE.BNDTech.t0">
   <mdRef LOCTYPE="URL" xlink:type="simple" xlink:href="shaman1t10_metadata/tech/info.xml"
MDTYPE="OTHER" OTHERMDTYPE="BND-CONTENTE" MIMETYPE="application/xml"/>
  </techMD>
  <rightsMD ID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r1">
```

```
    <mdRef LOCTYPE="URL" xlink:type="simple"
xlink:href="shaman1t10_metadata/rights/private.xml" MDTYPE="OTHER" OTHERMDTYPE="RIGHTSMD"
MIMETYPE="application/xml"/>
   </rightsMD>
   <rightsMD ID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
    <mdRef LOCTYPE="URL" xlink:type="simple" xlink:href="shaman1t10_metadata/rights/public.xml"
MDTYPE="OTHER" OTHERMDTYPE="RIGHTSMD" MIMETYPE="application/xml"/>
   </rightsMD>
  </amdSec>
  <fileSec>
   <fileGrp ID="RADBASEALONGAMETRO">
    <file ID="xml.RADBASEALONGAMETRO.data" MIMETYPE="text/xml" SIZE="1246105"
CHECKSUM="c36e99b2349f9fa0eef9ec29c283d1d9" CHECKSUMTYPE="MD5" GROUPID="RADBASEALONGAMETRO">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="RADBASEALONGAMETRO/data.xml"/>
    </file>
    <file ID="xsd.RADBASEALONGAMETRO.schema" MIMETYPE="application/xml" SIZE="4003"
CHECKSUM="b5dba721902f04a8a4f1ae27ce64833d" CHECKSUMTYPE="MD5" GROUPID="RADBASEALONGAMETRO">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="RADBASEALONGAMETRO/schema.xsd"/>
    </file>
   </fileGrp>
   <fileGrp ID="RADDRENO">
    <file ID="xml.RADDRENO.data" MIMETYPE="text/xml" SIZE="86602"
CHECKSUM="5bc997904cb039433bc6a5ca3d621e94" CHECKSUMTYPE="MD5" GROUPID="RADDRENO">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="RADDRENO/data.xml"/>
    </file>
    <file ID="xsd.RADDRENO.schema" MIMETYPE="application/xml" SIZE="4003"
CHECKSUM="b5dba721902f04a8a4f1ae27ce64833d" CHECKSUMTYPE="MD5" GROUPID="RADDRENO">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="RADDRENO/schema.xsd"/>
    </file>
   </fileGrp>
   <fileGrp ID="RADESCALANIVEL">
    <file ID="xml.RADESCALANIVEL.data" MIMETYPE="text/xml" SIZE="43694"
CHECKSUM="30be038e22998f9823823c2ba4d27910" CHECKSUMTYPE="MD5" GROUPID="RADESCALANIVEL">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="RADESCALANIVEL/data.xml"/>
    </file>
    <file ID="xsd.RADESCALANIVEL.schema" MIMETYPE="application/xml" SIZE="4003"
CHECKSUM="b5dba721902f04a8a4f1ae27ce64833d" CHECKSUMTYPE="MD5" GROUPID="RADESCALANIVEL">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="RADESCALANIVEL/schema.xsd"/>
    </file>
   </fileGrp>
   <fileGrp ID="RADEXTFUNDACAO">
    <file ID="xml.RADEXTFUNDACAO.data" MIMETYPE="text/xml" SIZE="147635"
CHECKSUM="bcf30f2b6e15bcb3ccf98ba10da3ec37" CHECKSUMTYPE="MD5" GROUPID="RADEXTFUNDACAO">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="RADEXTFUNDACAO/data.xml"/>
    </file>
    <file ID="xsd.RADEXTFUNDACAO.schema" MIMETYPE="application/xml" SIZE="4003"
CHECKSUM="b5dba721902f04a8a4f1ae27ce64833d" CHECKSUMTYPE="MD5" GROUPID="RADEXTFUNDACAO">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="RADEXTFUNDACAO/schema.xsd"/>
    </file>
   </fileGrp>
   <fileGrp ID="RADFIOPRUMO">
    <file ID="xml.RADFIOPRUMO.data" MIMETYPE="text/xml" SIZE="178869"
CHECKSUM="3f03e6c55f6f81636e3d247528521167" CHECKSUMTYPE="MD5" GROUPID="RADFIOPRUMO">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="RADFIOPRUMO/data.xml"/>
    </file>
    <file ID="xsd.RADFIOPRUMO.schema" MIMETYPE="application/xml" SIZE="4003"
CHECKSUM="b5dba721902f04a8a4f1ae27ce64833d" CHECKSUMTYPE="MD5" GROUPID="RADFIOPRUMO">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="RADFIOPRUMO/schema.xsd"/>
    </file>
   </fileGrp>
   <fileGrp ID="RADPIEZOMETRO">
    <file ID="xml.RADPIEZOMETRO.data" MIMETYPE="text/xml" SIZE="219491"
CHECKSUM="761b5ba1c59a6789452f5fda18960ba1" CHECKSUMTYPE="MD5" GROUPID="RADPIEZOMETRO">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="RADPIEZOMETRO/data.xml"/>
    </file>
```

```
    <file ID="xsd.RADPIEZOMETRO.schema" MIMETYPE="application/xml" SIZE="4003"
CHECKSUM="b5dba721902f04a8a4f1ae27ce64833d" CHECKSUMTYPE="MD5" GROUPID="RADPIEZOMETRO">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="RADPIEZOMETRO/schema.xsd"/>
    </file>
   </fileGrp>
   <fileGrp ID="RADTERMOMETROAR">
    <file ID="xml.RADTERMOMETROAR.data" MIMETYPE="text/xml" SIZE="28822"
CHECKSUM="cda4f5621c09ce81257b1d070a4f1770" CHECKSUMTYPE="MD5" GROUPID="RADTERMOMETROAR">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="RADTERMOMETROAR/data.xml"/>
    </file>
    <file ID="xsd.RADTERMOMETROAR.schema" MIMETYPE="application/xml" SIZE="4003"
CHECKSUM="b5dba721902f04a8a4f1ae27ce64833d" CHECKSUMTYPE="MD5" GROUPID="RADTERMOMETROAR">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="RADTERMOMETROAR/schema.xsd"/>
    </file>
   </fileGrp>
   <fileGrp ID="shaman1t10_HTML">
    <file ID="b0" MIMETYPE="image/gif" SIZE="856" CHECKSUM="f10336b367bf91cd7e1bf198a7b8e1f1"
CHECKSUMTYPE="MD5" GROUPID="shaman1t10_HTML">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="icons/icon-master.gif"/>
    </file>
    <file ID="b1" MIMETYPE="image/gif" SIZE="432" CHECKSUM="951c7018d5e34445ceb2ccda89743c30"
CHECKSUMTYPE="MD5" GROUPID="shaman1t10_HTML">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="icons/icon-xml.gif"/>
    </file>
    <file ID="b2" MIMETYPE="image/gif" SIZE="898" CHECKSUM="7f6a18c5e38ea00e9179e875e5ee3c3a"
CHECKSUMTYPE="MD5" GROUPID="shaman1t10_HTML">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="icons/icon-xsd.gif"/>
    </file>
    <file ID="info" MIMETYPE="text/html" SIZE="4360"
CHECKSUM="3ff942b6c8b36d05dad00b253b95d013" CHECKSUMTYPE="MD5" GROUPID="shaman1t10_HTML">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="info.html"/>
    </file>
   </fileGrp>
  </fileSec>
  <structMap>
   <div ID="w" LABEL="Alto Lindoso" ADMID="tech.CONTENTE.BNDTech.t0
rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2" TYPE="BOOK">
    <div ID="w_i0" ORDERLABEL="[Master]" LABEL="[Master]"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r1" TYPE="INDEX">
     <div ID="w_i0_n0" ORDER="0" LABEL="RADBASEALONGAMETRO"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r1">
      <fptr FILEID="xml.RADBASEALONGAMETRO.data"/>
      <fptr FILEID="xsd.RADBASEALONGAMETRO.schema"/>
     </div>
     <div ID="w_i0_n1" ORDER="1" LABEL="RADDRENO"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r1">
      <fptr FILEID="xml.RADDRENO.data"/>
      <fptr FILEID="xsd.RADDRENO.schema"/>
     </div>
     <div ID="w_i0_n2" ORDER="2" LABEL="RADESCALANIVEL"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r1">
      <fptr FILEID="xml.RADESCALANIVEL.data"/>
      <fptr FILEID="xsd.RADESCALANIVEL.schema"/>
     </div>
     <div ID="w_i0_n3" ORDER="3" LABEL="RADEXTFUNDACAO"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r1">
      <fptr FILEID="xml.RADEXTFUNDACAO.data"/>
      <fptr FILEID="xsd.RADEXTFUNDACAO.schema"/>
     </div>
     <div ID="w_i0_n4" ORDER="4" LABEL="RADFIOPRUMO"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r1">
      <fptr FILEID="xml.RADFIOPRUMO.data"/>
      <fptr FILEID="xsd.RADFIOPRUMO.schema"/>
     </div>
```

```
    <div ID="w_i0_n5" ORDER="5" LABEL="RADPIEZOMETRO"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r1">
     <fptr FILEID="xml.RADPIEZOMETRO.data"/>
     <fptr FILEID="xsd.RADPIEZOMETRO.schema"/>
    </div>
    <div ID="w_i0_n6" ORDER="6" LABEL="RADTERMOMETROAR"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r1">
     <fptr FILEID="xml.RADTERMOMETROAR.data"/>
     <fptr FILEID="xsd.RADTERMOMETROAR.schema"/>
    </div>
   </div>
   <div ID="w_i1" ORDERLABEL="Alto Lindoso" LABEL="Alto Lindoso"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2" TYPE="INDEX">
    <div ID="w_i1_n0" LABEL="Automatic data gathering"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
     <div ID="w_i1_n0_n1" LABEL="Displacement base"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n1_n2" ORDER="0" LABEL="Data"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <fptr FILEID="xml.RADBASEALONGAMETRO.data"/>
      </div>
      <div ID="w_i1_n0_n1_n3" ORDER="1" LABEL="Schema"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <fptr FILEID="xsd.RADBASEALONGAMETRO.schema"/>
      </div>
     </div>
     <div ID="w_i1_n0_n4" LABEL="Drain"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n4_n5" ORDER="2" LABEL="Data"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <fptr FILEID="xml.RADDRENO.data"/>
      </div>
      <div ID="w_i1_n0_n4_n6" ORDER="3" LABEL="Schema"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <fptr FILEID="xsd.RADDRENO.schema"/>
      </div>
     </div>
     <div ID="w_i1_n0_n7" LABEL="Scale level"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n7_n8" ORDER="4" LABEL="Data"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <fptr FILEID="xml.RADESCALANIVEL.data"/>
      </div>
      <div ID="w_i1_n0_n7_n9" ORDER="5" LABEL="Schema"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <fptr FILEID="xsd.RADESCALANIVEL.schema"/>
      </div>
     </div>
     <div ID="w_i1_n0_n10" LABEL="Gage foundation"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n10_n11" ORDER="6" LABEL="Data"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <fptr FILEID="xml.RADEXTFUNDACAO.data"/>
      </div>
      <div ID="w_i1_n0_n10_n12" ORDER="7" LABEL="Schema"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <fptr FILEID="xsd.RADEXTFUNDACAO.schema"/>
      </div>
     </div>
     <div ID="w_i1_n0_n13" LABEL="Plumbline"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n13_n14" ORDER="8" LABEL="Data"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <fptr FILEID="xml.RADFIOPRUMO.data"/>
      </div>
```

```
      <div ID="w_i1_n0_n13_n15" ORDER="9" LABEL="Schema"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <fptr FILEID="xsd.RADFIOPRUMO.schema"/>
      </div>
     </div>
     <div ID="w_i1_n0_n16" LABEL="Piezometer"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n16_n17" ORDER="10" LABEL="Data"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <fptr FILEID="xml.RADPIEZOMETRO.data"/>
      </div>
      <div ID="w_i1_n0_n16_n18" ORDER="11" LABEL="Schema"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <fptr FILEID="xsd.RADPIEZOMETRO.schema"/>
      </div>
     </div>
     <div ID="w_i1_n0_n19" LABEL="Air thermometer"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n19_n20" ORDER="12" LABEL="Data"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <fptr FILEID="xml.RADTERMOMETROAR.data"/>
      </div>
      <div ID="w_i1_n0_n19_n21" ORDER="13" LABEL="Schema"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <fptr FILEID="xsd.RADTERMOMETROAR.schema"/>
      </div>
     </div>
    </div>
   </div>
  </div>
 </structMap>
</mets>
```

A2. Scenario 2 - Acquisition and Preservation of Scientific Workflows

A2.1. Example data schema

```
<?xml version="1.0"?>
<xs:schema id="NewDataSet" xmlns="" xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:msdata="urn:schemas-microsoft-com:xml-msdata">
  <xs:element name="workflow">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="id" type="xs:string" minOccurs="0" msdata:Ordinal="0" />
        <xs:element name="title" type="xs:string" minOccurs="0" msdata:Ordinal="1" />
        <xs:element name="description" type="xs:string" minOccurs="0" msdata:Ordinal="2" />
        <xs:element name="created-at" type="xs:string" minOccurs="0" msdata:Ordinal="4" />
        <xs:element name="preview" type="xs:string" minOccurs="0" msdata:Ordinal="5" />
        <xs:element name="svg" type="xs:string" minOccurs="0" msdata:Ordinal="6" />
        <xs:element name="content-uri" type="xs:string" minOccurs="0" msdata:Ordinal="7" />
        <xs:element name="content-type" type="xs:string" minOccurs="0" msdata:Ordinal="8" />
        <xs:element name="type" nillable="true" minOccurs="0" maxOccurs="unbounded">
          <xs:complexType>
            <xs:simpleContent msdata:ColumnName="type_Text" msdata:Ordinal="2">
              <xs:extension base="xs:string">
                <xs:attribute name="uri" type="xs:string" />
                <xs:attribute name="resource" type="xs:string" />
              </xs:extension>
            </xs:simpleContent>
          </xs:complexType>
        </xs:element>
        <xs:element name="uploader" nillable="true" minOccurs="0" maxOccurs="unbounded">
          <xs:complexType>
```

```
            <xs:simpleContent msdata:ColumnName="uploader_Text" msdata:Ordinal="2">
              <xs:extension base="xs:string">
                <xs:attribute name="resource" type="xs:string" />
                <xs:attribute name="uri" type="xs:string" />
              </xs:extension>
            </xs:simpleContent>
          </xs:complexType>
        </xs:element>
        <xs:element name="license-type" nillable="true" minOccurs="0" maxOccurs="unbounded">
          <xs:complexType>
            <xs:simpleContent msdata:ColumnName="license-type_Text" msdata:Ordinal="2">
              <xs:extension base="xs:string">
                <xs:attribute name="uri" type="xs:string" />
                <xs:attribute name="resource" type="xs:string" />
              </xs:extension>
            </xs:simpleContent>
          </xs:complexType>
        </xs:element>
        <xs:element name="tags" minOccurs="0" maxOccurs="unbounded">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="tag" nillable="true" minOccurs="0" maxOccurs="unbounded">
                <xs:complexType>
                  <xs:simpleContent msdata:ColumnName="tag_Text" msdata:Ordinal="2">
                    <xs:extension base="xs:string">
                      <xs:attribute name="uri" type="xs:string" />
                      <xs:attribute name="resource" type="xs:string" />
                    </xs:extension>
                  </xs:simpleContent>
                </xs:complexType>
              </xs:element>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:sequence>
      <xs:attribute name="uri" type="xs:string" />
      <xs:attribute name="resource" type="xs:string" />
      <xs:attribute name="version" type="xs:string" />
    </xs:complexType>
  </xs:element>
  <xs:element name="NewDataSet" msdata:IsDataSet="true" msdata:UseCurrentLocale="true">
    <xs:complexType>
      <xs:choice minOccurs="0" maxOccurs="unbounded">
        <xs:element ref="workflow" />
      </xs:choice>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

A2.2. Example data

```
<?xml version="1.0" encoding="UTF-8"?>
<workflow uri="http://www.myexperiment.org/workflow.xml?id=173"
resource="http://www.myexperiment.org/workflows/173" version="2">
  <id>173</id>
  <title>Unique tags</title>
  <description>This workflow takes a comma separated list of tags and removes duplicate
entries. Tags may have multiple words in them. An example string is
"carrots,handbags,carrots,cheese".</description>
  <type uri="http://www.myexperiment.org/type.xml?id=1"
resource="http://www.myexperiment.org/content_types/1">Taverna 1</type>
  <uploader resource="http://www.myexperiment.org/users/22"
uri="http://www.myexperiment.org/user.xml?id=22">Don Cruickshank</uploader>
  <created-at>Tue Mar 11 16:52:42 +0000 2008</created-at>
  <preview>http://www.myexperiment.org/workflows/173/previews/full</preview>
```

```
  <svg>http://www.myexperiment.org/workflows/173/previews/svg</svg>
  <license-type uri="http://www.myexperiment.org/license.xml?id=2"
resource="http://www.myexperiment.org/licenses/2">by-sa</license-type>
  <content-
uri>http://www.myexperiment.org/workflows/173/download/unique_tags_18054.xml</content-uri>
  <content-type>application/vnd.taverna.scufl+xml</content-type>
  <tags>
    <tag uri="http://www.myexperiment.org/tag.xml?id=555"
resource="http://www.myexperiment.org/tags/555">example</tag>
    <tag uri="http://www.myexperiment.org/tag.xml?id=450"
resource="http://www.myexperiment.org/tags/450">scampi</tag>
    <tag uri="http://www.myexperiment.org/tag.xml?id=760"
resource="http://www.myexperiment.org/tags/760">design pattern</tag>
  </tags>
</workflow>
```


A2.3. Example METS file


```
<mets OBJID="workflow173" LABEL="workflow173" TYPE="ContentE v.3.5"
xmlns="http://www.loc.gov/METS/" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:bndTechInfo="http://schemas.bn.pt/tech-info"
xsi:schemaLocation="http://www.loc.gov/METS/ http://schemas.bn.pt/mets/v1.5/metsv1.5.xsd">
 <metsHdr CREATEDATE="2011-10-19T16:38:21" LASTMODDATE="2011-10-19T16:38:21"
RECORDSTATUS="COMPLETED">
  <agent ID="agent.CONTENTE.METSAgent.ag16" ROLE="CREATOR" TYPE="ORGANIZATION">
   <name>LNEC</name>
  </agent>
 </metsHdr>
 <amdSec>
  <techMD ID="tech.CONTENTE.BNDTech.t0">
   <mdRef LOCTYPE="URL" xlink:type="simple" xlink:href="workflow173_metadata/tech/info.xml"
MDTYPE="OTHER" OTHERMDTYPE="BND-CONTENTE" MIMETYPE="application/xml"/>
  </techMD>
  <rightsMD ID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
   <mdRef LOCTYPE="URL" xlink:type="simple"
xlink:href="workflow173_metadata/rights/rights.xml" MDTYPE="OTHER" OTHERMDTYPE="RIGHTSMD"
MIMETYPE="application/xml"/>
  </rightsMD>
 </amdSec>
 <fileSec>
  <fileGrp ID="attributions">
   <fileGrp ID="attributions.file5">
    <file ID="xml.attributions.file5.file5" MIMETYPE="text/xml" SIZE="1170"
CHECKSUM="57606650e9289ba1e02e4897b81a3e85" CHECKSUMTYPE="MD5" GROUPID="attributions.file5">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="attributions/file5/file5.xml"/>
    </file>
    <file ID="xml.attributions.file5.license" MIMETYPE="text/xml" SIZE="1861"
CHECKSUM="2025593e6317ac1d9445b2467fd22358" CHECKSUMTYPE="MD5" GROUPID="attributions.file5">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="attributions/file5/license.xml"/>
    </file>
    <file ID="xsd.attributions.file5.schema" MIMETYPE="application/xml" SIZE="3797"
CHECKSUM="ef00d4035e00373aa65c38351b877a80" CHECKSUMTYPE="MD5" GROUPID="attributions.file5">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="attributions/file5/schema.xsd"/>
    </file>
    <file ID="xml.attributions.file5.type" MIMETYPE="text/xml" SIZE="326"
CHECKSUM="c915b5eb30fdff194e5c4476db04a0a4" CHECKSUMTYPE="MD5" GROUPID="attributions.file5">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="attributions/file5/type.xml"/>
    </file>
    <file ID="xml.attributions.file5.uploaderUser" MIMETYPE="text/xml" SIZE="619"
CHECKSUM="7180ed66d6cbcd417a977da5182f0aa4" CHECKSUMTYPE="MD5" GROUPID="attributions.file5">
     <FLocat LOCTYPE="URL" xlink:type="simple"
xlink:href="attributions/file5/uploaderUser.xml"/>
    </file>
   </fileGrp>
```

```
  </fileGrp>
  <fileGrp ID="content">
   <file ID="xml.content.content" MIMETYPE="text/xml" SIZE="2297"
CHECKSUM="5e0ea286f6f476dbdb100d9e360849a8" CHECKSUMTYPE="MD5" GROUPID="content">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="content/content.xml"/>
   </file>
   <file ID="xsd.content.schema" MIMETYPE="application/xml" SIZE="3797"
CHECKSUM="ef00d4035e00373aa65c38351b877a80" CHECKSUMTYPE="MD5" GROUPID="content">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="content/schema.xsd"/>
   </file>
  </fileGrp>
  <fileGrp ID="credits">
   <file ID="xml.credits.creditUser22" MIMETYPE="text/xml" SIZE="619"
CHECKSUM="7180ed66d6cbcd417a977da5182f0aa4" CHECKSUMTYPE="MD5" GROUPID="credits">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="credits/creditUser22.xml"/>
   </file>
   <file ID="xsd.credits.schema" MIMETYPE="application/xml" SIZE="3797"
CHECKSUM="ef00d4035e00373aa65c38351b877a80" CHECKSUMTYPE="MD5" GROUPID="credits">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="credits/schema.xsd"/>
   </file>
  </fileGrp>
  <fileGrp ID="license">
   <file ID="xml.license.license" MIMETYPE="text/xml" SIZE="1992"
CHECKSUM="b067984bb725f8657eae473a4e8cba7a" CHECKSUMTYPE="MD5" GROUPID="license">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="license/license.xml"/>
   </file>
   <file ID="xsd.license.schema" MIMETYPE="application/xml" SIZE="3797"
CHECKSUM="ef00d4035e00373aa65c38351b877a80" CHECKSUMTYPE="MD5" GROUPID="license">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="license/schema.xsd"/>
   </file>
  </fileGrp>
  <fileGrp ID="main">
   <file ID="png.main.preview" MIMETYPE="image/png" SIZE="4789"
CHECKSUM="aa08658da0cb9e49169e2edda88bc834" CHECKSUMTYPE="MD5" GROUPID="main">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="main/preview.png"/>
   </file>
   <file ID="xsd.main.schema" MIMETYPE="application/xml" SIZE="3797"
CHECKSUM="ef00d4035e00373aa65c38351b877a80" CHECKSUMTYPE="MD5" GROUPID="main">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="main/schema.xsd"/>
   </file>
   <file ID="xml.main.workflow173" MIMETYPE="text/xml" SIZE="1588"
CHECKSUM="af13f33b34c05ef09ffb1ab5231e0021" CHECKSUMTYPE="MD5" GROUPID="main">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="main/workflow173.xml"/>
   </file>
  </fileGrp>
  <fileGrp ID="ratings">
   <fileGrp ID="ratings.rating109">
    <file ID="xml.ratings.rating109.rating109" MIMETYPE="text/xml" SIZE="572"
CHECKSUM="2f2936ed901fc7a56d2701fcb232dfb3" CHECKSUMTYPE="MD5" GROUPID="ratings.rating109">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="ratings/rating109/rating109.xml"/>
    </file>
    <file ID="xml.ratings.rating109.ratingOwner" MIMETYPE="text/xml" SIZE="438"
CHECKSUM="2d0d9d6e4f10c5dd711cd5f7652ab61c" CHECKSUMTYPE="MD5" GROUPID="ratings.rating109">
     <FLocat LOCTYPE="URL" xlink:type="simple"
xlink:href="ratings/rating109/ratingOwner.xml"/>
    </file>
    <file ID="xsd.ratings.rating109.schema" MIMETYPE="application/xml" SIZE="3797"
CHECKSUM="ef00d4035e00373aa65c38351b877a80" CHECKSUMTYPE="MD5" GROUPID="ratings.rating109">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="ratings/rating109/schema.xsd"/>
    </file>
   </fileGrp>
   <fileGrp ID="ratings.rating124">
    <file ID="xml.ratings.rating124.rating124" MIMETYPE="text/xml" SIZE="583"
CHECKSUM="20f221a2b1d125e31c63710ea1c87732" CHECKSUMTYPE="MD5" GROUPID="ratings.rating124">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="ratings/rating124/rating124.xml"/>
```

```
    </file>
    <file ID="xml.ratings.rating124.ratingOwner" MIMETYPE="text/xml" SIZE="391"
CHECKSUM="eb6af8db3cc3a67edb182532e4b805c3" CHECKSUMTYPE="MD5" GROUPID="ratings.rating124">
     <FLocat LOCTYPE="URL" xlink:type="simple"
xlink:href="ratings/rating124/ratingOwner.xml"/>
    </file>
    <file ID="xsd.ratings.rating124.schema" MIMETYPE="application/xml" SIZE="3797"
CHECKSUM="ef00d4035e00373aa65c38351b877a80" CHECKSUMTYPE="MD5" GROUPID="ratings.rating124">
     <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="ratings/rating124/schema.xsd"/>
    </file>
   </fileGrp>
  </fileGrp>
  <fileGrp ID="tags">
   <file ID="xsd.tags.schema" MIMETYPE="application/xml" SIZE="3797"
CHECKSUM="ef00d4035e00373aa65c38351b877a80" CHECKSUMTYPE="MD5" GROUPID="tags">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="tags/schema.xsd"/>
   </file>
   <file ID="xml.tags.tag450" MIMETYPE="text/xml" SIZE="206"
CHECKSUM="51ad1d07bd279967a1fa5313ebe86b85" CHECKSUMTYPE="MD5" GROUPID="tags">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="tags/tag450.xml"/>
   </file>
   <file ID="xml.tags.tag555" MIMETYPE="text/xml" SIZE="209"
CHECKSUM="e98397cbdb36b53d18b6706ff0e0eebb" CHECKSUMTYPE="MD5" GROUPID="tags">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="tags/tag555.xml"/>
   </file>
   <file ID="xml.tags.tag760" MIMETYPE="text/xml" SIZE="215"
CHECKSUM="c8c5d72cb78136abdf95af4e7e78c1bb" CHECKSUMTYPE="MD5" GROUPID="tags">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="tags/tag760.xml"/>
   </file>
  </fileGrp>
  <fileGrp ID="type">
   <file ID="xsd.type.schema" MIMETYPE="application/xml" SIZE="3797"
CHECKSUM="ef00d4035e00373aa65c38351b877a80" CHECKSUMTYPE="MD5" GROUPID="type">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="type/schema.xsd"/>
   </file>
   <file ID="xml.type.type" MIMETYPE="text/xml" SIZE="1553"
CHECKSUM="3c845218369e6a822b08edbb902e5c65" CHECKSUMTYPE="MD5" GROUPID="type">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="type/type.xml"/>
   </file>
  </fileGrp>
  <fileGrp ID="user">
   <file ID="xsd.user.schema" MIMETYPE="application/xml" SIZE="3797"
CHECKSUM="ef00d4035e00373aa65c38351b877a80" CHECKSUMTYPE="MD5" GROUPID="user">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="user/schema.xsd"/>
   </file>
   <file ID="xml.user.user" MIMETYPE="text/xml" SIZE="619"
CHECKSUM="7180ed66d6cbcd417a977da5182f0aa4" CHECKSUMTYPE="MD5" GROUPID="user">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="user/user.xml"/>
   </file>
  </fileGrp>
  <fileGrp ID="workflow173_HTML">
   <file ID="b0" MIMETYPE="image/gif" SIZE="878" CHECKSUM="7bc3b9013af0ffea8b3d7aa06edb7374"
CHECKSUMTYPE="MD5" GROUPID="workflow173_HTML">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="icons/icon-png.gif"/>
   </file>
   <file ID="b1" MIMETYPE="image/gif" SIZE="432" CHECKSUM="951c7018d5e34445ceb2ccda89743c30"
CHECKSUMTYPE="MD5" GROUPID="workflow173_HTML">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="icons/icon-xml.gif"/>
   </file>
   <file ID="b2" MIMETYPE="image/gif" SIZE="898" CHECKSUM="7f6a18c5e38ea00e9179e875e5ee3c3a"
CHECKSUMTYPE="MD5" GROUPID="workflow173_HTML">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="icons/icon-xsd.gif"/>
   </file>
   <file ID="index" MIMETYPE="text/html" SIZE="17262"
CHECKSUM="8d0d08ba4ba2474d08e1833b46cff456" CHECKSUMTYPE="MD5" GROUPID="workflow173_HTML">
```

```
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="index.html"/>
   </file>
   <file ID="css" MIMETYPE="text/css" CHECKSUM="9ab99debd6cf8c7bf87f2454ebbfcc77"
CHECKSUMTYPE="MD5" GROUPID="workflow173_HTML">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="cssStyle.css"/>
   </file>
   <file ID="info" MIMETYPE="text/html" SIZE="4347"
CHECKSUM="f597a8e3acc7e00991060404dcc8f10d" CHECKSUMTYPE="MD5" GROUPID="workflow173_HTML">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="info.html"/>
   </file>
  </fileGrp>
 </fileSec>
 <structMap>
  <div ID="w" LABEL="workflow173" ADMID="tech.CONTENTE.BNDTech.t0
rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2" TYPE="BOOK">
   <fptr FILEID="index"/>
   <fptr FILEID="css"/>
   <div ID="w_i1" ORDERLABEL="workflow173" LABEL="workflow173"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2" TYPE="INDEX">
    <div ID="w_i1_n0" LABEL="Automatic data gathering"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
     <div ID="w_i1_n0_n1" LABEL="file5"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n1_n2" ORDER="0" LABEL="Data">
       <fptr FILEID="xml.attributions.file5.file5"/>
      </div>
      <div ID="w_i1_n0_n1_n3" ORDER="1" LABEL="Schema">
       <fptr FILEID="xsd.attributions.file5.schema"/>
      </div>
     </div>
     <div ID="w_i1_n0_n4" LABEL="file5"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n4_n5" ORDER="2" LABEL="Data">
       <fptr FILEID="xml.attributions.file5.license"/>
      </div>
      <div ID="w_i1_n0_n4_n6" ORDER="3" LABEL="Schema">
       <fptr FILEID="xsd.attributions.file5.schema"/>
      </div>
     </div>
     <div ID="w_i1_n0_n7" LABEL="file5"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n7_n8" ORDER="4" LABEL="Data">
       <fptr FILEID="xml.attributions.file5.type"/>
      </div>
      <div ID="w_i1_n0_n7_n9" ORDER="5" LABEL="Schema">
       <fptr FILEID="xsd.attributions.file5.schema"/>
      </div>
     </div>
     <div ID="w_i1_n0_n10" LABEL="file5"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n10_n11" ORDER="6" LABEL="Data">
       <fptr FILEID="xml.attributions.file5.uploaderUser"/>
      </div>
      <div ID="w_i1_n0_n10_n12" ORDER="7" LABEL="Schema">
       <fptr FILEID="xsd.attributions.file5.schema"/>
      </div>
     </div>
     <div ID="w_i1_n0_n13" LABEL="content"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n13_n14" ORDER="8" LABEL="Data">
       <fptr FILEID="xml.content.content"/>
      </div>
      <div ID="w_i1_n0_n13_n15" ORDER="9" LABEL="Schema">
       <fptr FILEID="xsd.content.schema"/>
      </div>
```

```
      </div>
      <div ID="w_i1_n0_n16" LABEL="credits"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <div ID="w_i1_n0_n16_n17" ORDER="10" LABEL="Data">
        <fptr FILEID="xml.credits.creditUser22"/>
       </div>
       <div ID="w_i1_n0_n16_n18" ORDER="11" LABEL="Schema">
        <fptr FILEID="xsd.credits.schema"/>
       </div>
      </div>
      <div ID="w_i1_n0_n19" LABEL="license"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <div ID="w_i1_n0_n19_n20" ORDER="12" LABEL="Data">
        <fptr FILEID="xml.license.license"/>
       </div>
       <div ID="w_i1_n0_n19_n21" ORDER="13" LABEL="Schema">
        <fptr FILEID="xsd.license.schema"/>
       </div>
      </div>
      <div ID="w_i1_n0_n22" LABEL="main"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <div ID="w_i1_n0_n22_n23" ORDER="14" LABEL="Data">
        <fptr FILEID="png.main.preview"/>
       </div>
      </div>
      <div ID="w_i1_n0_n24" LABEL="main"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <div ID="w_i1_n0_n24_n25" ORDER="15" LABEL="Data">
        <fptr FILEID="xml.main.workflow173"/>
       </div>
       <div ID="w_i1_n0_n24_n26" ORDER="16" LABEL="Schema">
        <fptr FILEID="xsd.main.schema"/>
       </div>
      </div>
      <div ID="w_i1_n0_n27" LABEL="rating109"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <div ID="w_i1_n0_n27_n28" ORDER="17" LABEL="Data">
        <fptr FILEID="xml.ratings.rating109.rating109"/>
       </div>
       <div ID="w_i1_n0_n27_n29" ORDER="18" LABEL="Schema">
        <fptr FILEID="xsd.ratings.rating109.schema"/>
       </div>
      </div>
      <div ID="w_i1_n0_n30" LABEL="rating109"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <div ID="w_i1_n0_n30_n31" ORDER="19" LABEL="Data">
        <fptr FILEID="xml.ratings.rating109.ratingOwner"/>
       </div>
       <div ID="w_i1_n0_n30_n32" ORDER="20" LABEL="Schema">
        <fptr FILEID="xsd.ratings.rating109.schema"/>
       </div>
      </div>
      <div ID="w_i1_n0_n33" LABEL="rating124"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <div ID="w_i1_n0_n33_n34" ORDER="21" LABEL="Data">
        <fptr FILEID="xml.ratings.rating124.rating124"/>
       </div>
       <div ID="w_i1_n0_n33_n35" ORDER="22" LABEL="Schema">
        <fptr FILEID="xsd.ratings.rating124.schema"/>
       </div>
      </div>
      <div ID="w_i1_n0_n36" LABEL="rating124"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
       <div ID="w_i1_n0_n36_n37" ORDER="23" LABEL="Data">
        <fptr FILEID="xml.ratings.rating124.ratingOwner"/>
```

```
      </div>
      <div ID="w_i1_n0_n36_n38" ORDER="24" LABEL="Schema">
       &lt;fptr FILEID="xsd.ratings.rating124.schema"/&gt;
      </div>
     </div>
     <div ID="w_i1_n0_n39" LABEL="tags"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n39_n40" ORDER="25" LABEL="Data">
       &lt;fptr FILEID="xml.tags.tag450"/&gt;
      </div>
      <div ID="w_i1_n0_n39_n41" ORDER="26" LABEL="Schema">
       &lt;fptr FILEID="xsd.tags.schema"/&gt;
      </div>
     </div>
     <div ID="w_i1_n0_n42" LABEL="tags"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n42_n43" ORDER="27" LABEL="Data">
       &lt;fptr FILEID="xml.tags.tag555"/&gt;
      </div>
      <div ID="w_i1_n0_n42_n44" ORDER="28" LABEL="Schema">
       &lt;fptr FILEID="xsd.tags.schema"/&gt;
      </div>
     </div>
     <div ID="w_i1_n0_n45" LABEL="tags"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n45_n46" ORDER="29" LABEL="Data">
       &lt;fptr FILEID="xml.tags.tag760"/&gt;
      </div>
      <div ID="w_i1_n0_n45_n47" ORDER="30" LABEL="Schema">
       &lt;fptr FILEID="xsd.tags.schema"/&gt;
      </div>
     </div>
     <div ID="w_i1_n0_n48" LABEL="type"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n48_n49" ORDER="31" LABEL="Data">
       &lt;fptr FILEID="xml.type.type"/&gt;
      </div>
      <div ID="w_i1_n0_n48_n50" ORDER="32" LABEL="Schema">
       &lt;fptr FILEID="xsd.type.schema"/&gt;
      </div>
     </div>
     <div ID="w_i1_n0_n51" LABEL="user"
ADMID="rights.CONTENTE.METSRights.urn.bnd.bn.rights.r2">
      <div ID="w_i1_n0_n51_n52" ORDER="33" LABEL="Data">
       &lt;fptr FILEID="xml.user.user"/&gt;
      </div>
      <div ID="w_i1_n0_n51_n53" ORDER="34" LABEL="Schema">
       &lt;fptr FILEID="xsd.user.schema"/&gt;
      </div>
     </div>
    </div>
   </div>
  </div>
 &lt;/structMap&gt;
&lt;/mets&gt;
```