# OEILM: a semantic linking framework for environmental research infrastructures

Zhiming Zhao[a] Paola Grosso[a] Wouter Los[a] Cees de Laat[a]
Yin Chen[b] Alex Hardisty[b] Paul Martin[c] Schentz Herbert[d] Magagna Barbara[d]
[a] University of Amsterdam, the Netherlands {Z.Zhao|P.Grosso|W.Los|DeLaat}@uva.nl
[b] Cardiff University, UK {ChenY58|HardistyAR}@cardiff.ac.uk
[c] Edinburgh University, UK {Pmartin}@staffmail.ed.ac.uk
[d] Environmental Agency Austria, Austria {Barbara.Magagna|herbert.schentz}@umweltbundesamt.at

***Abstract*—The interoperability between research infrastructures, including not only cross invocations of services, but also the integration between data schemas, processing models and management policies and security controls, is essential to enable large scale data driven experiments. Analysing functional gaps between research infrastructures and decomposing the interoperability issues into separated viewpoints promote a generic information linking model between data and services in infrastructures. This paper proposes a multi viewpoint framework namely Open e-Science Information Linking Model (OEILM) in the context of environmental research infrastructures to semantically link data and services among infrastructures.**

**Key words:** Research infrastructure, Interoperability, semantic linking, Open distributed processing.

## I. Introduction

A research infrastructure (RI) provides advanced instruments and services for communities of users to conduct different levels of research in the field [1]; ICOS [2], EPOS [3], Lifewatch [4] and ESCAT [5] are examples in environmental sciences. These RIs all interface to large scale deployed sensors or observers, and deal with very large quantities of spatial aware data series. Collaborating data and services from different RIs enable scientific researches at a system level but also require integration and interoperability of operations at different service layers between infrastructures.

The EU FP7 project ENVRI[6], namely the "Common Operations of ENVironmental Research Infrastructures", emphasises on synergies between advanced developments, not only among the infrastructure facilities but also data driven experiments require multi-disciplinary sciences. The main mission is to define a common ontological framework and standards for the description and characterization of research infrastructures to achieve seamless interoperability between the heterogeneous data sources and resources. In this short paper, we briefly introduce the basic approach and discuss a semantic framework proposed in ENVRI called Open e-Science Information Linking Model (OEILM).

## II. Open e-Science Information Linking Model

Linking information and knowledge fragments that represent the semantics of services and data sources related to the data life cycle essentially enables further interoperability between RIs. A generic and globally operational ontology applicable for all possible situations is most probably not achievable; instead, interrelating different pairs of ontologies through semantic bridges promotes an evolutionary solution for many previously unrelated ontologies. A hybrid approach is taken in the ENVRI project. The Open e-Science Information Linking Model (OEILM) employs an ENVRI reference model to distinguish and formulate the commonality between the RIs, and uses a linking layer to bridge the semantic gaps between the reference model and other specific ontologies in RIs.

The Open Distributed Processing (ODP) modelling approach[7] captures the design and development issues in complex distributed systems from five corresponding viewpoints. The *enterprise viewpoint* models potential use cases, involved roles, behaviours and communities of the system; the *informational viewpoint* models the information objects and schemas related information; the *computational viewpoint* models operations of functional components and binding interfaces between them; the *engineering viewpoint* describes how the system should be constructed; and the *technological viewpoint* describes required technologies in the system development. Such multi viewpoint approach is suitable for modelling complex research infrastructures, and we follow the ODP philosophy to build the ENVRI Reference Model (ENVRI-RM) [8].

The OEILM has thus three layers 1) the *core ontology* of Open Distributed Processing (ODP) is used as the model basis, 2) the *ENVRI-Reference Model ontology* imports the core ontology and models concepts in ENVRI-RM, and 3) the *linking ontology* connects the reference model with the information models outside research infrastructures, such as description languages for underlying network infrastructures, and for domain specific data and services. Fig. 1 shows the basic idea of OEILM. The linking layer bridges the reference model ontology with five highlighted clusters of ontologies which are related to descriptions of abstract applications, services, data and metadata, middleware and physical infrastructure.

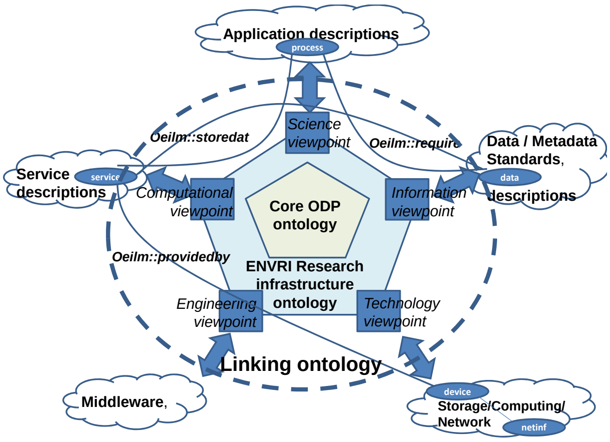


Fig. 1. The linking ontology connects ENVRI-RM ontology and the other ontologies.

## III. USE CASE

Several use cases are defined in ENVRI[6]. One typical example is to investigate behaviours of the Eyjafjallajkull volcano in Iceland using data of different environmental research data infrastructures. The use case has the following conceptual processes: 1) discover required information from distributed locations, 2) obtain and gather data from different sources, 3) compose data processing or simulation workflows, 4) execute the workflow and analyse results, and 5) provide user support for sharing results about research and predications. The semantic information provided by OEILM will enable the data and resource discovery, and the planning for executing applications on distributed infrastructure.

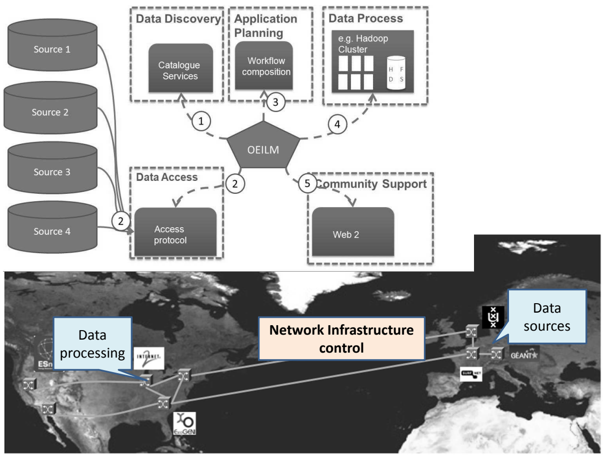


Fig. 2. The conceptual diagram of the use case: OEILM is used for data and resource discovery. The bottom part shows the test bed of the use case.

Fig. 2 shows the basic scenario. We prototype the data access and delivery part of the use case using a test bed provided by the OpenLab facilitate in University of Amsterdam [9], and will demonstrate it in Supercomputing 2013 [10].

## IV. SUMMARY

In this paper, we introduced our ongoing work in ENVRI. We propose OEILM to enhance the integration between RI and information of data and the underlying storing, computing and network elements. Compared to the other semantic linking frameworks, and data Grid technologies such as iRDOS [11] the work presented in this paper shows several differences. Firstly, the ODP approach used in the paper decomposes the context of functional components in a research infrastructure into different viewpoints. These viewpoints complement with each other, but can also deliver the information independently. Secondly, the OEILM links different levels of abstraction of resources that applications need. It can enable several use cases that traditional data intensive applications do not cover, for instance bringing network level QoS optimisation to the data movement that applications require. Finally, OEILM is meant for the communication between evolving research infrastructures, rather than building a infrastructure and promoting it to specific user communities.

We can summarize that interoperability is an important requirement for supporting semantic integration between data and services between different research infrastructures. An effective reference model synchronizes terminologies defined in different environmental RIs, and guides the further development of the common operations and functional components in the infrastructure. Semantic web technologies provide an open world view on modelling the linking among elements in research infrastructures. A semantic linking framework is important for realizing interoperability among research infrastructures, and Open e-Science Information Linking Model is evolving in this direction.

## ACKNOWLEDGMENT

We would like to thank the FP7 EU funded Integrated project ENVRI (project number 283465) and the Dutch national and education network SURFnet, through the GigaPort Research on Network (RoN) project for sponsoring this research.

## REFERENCES

[1] European commission. Research and innovation infrastructure, http://ec.europa.eu/research/infrastructures.

[2] ICOS. Icos (integrated carbon observation system), http://www.icos-infrastructure.eu. Research infrastructure.

[3] EPOS. Epos(the european plate observing system) epos; http://www.epos-eu.org. Research infrastructure.

[4] LifeWatch. Lifewatch infrastructure for biodiversity and ecosystem research, http://www.lifewatch.eu/. Research infrastructure.

[5] EISCAT. European incoherent scatter radar system (three-dimensional imaging), http://www.eiscat.com/. Research infrastructure.

[6] ENVRI. Common operations in environmental research infrastructures, http://envri.eu/. Research infrastructure, 2013.

[7] Peter F. Linington, Zoran Milosevic, Akira tanaka, and Antonio Vallecillo. *Building enterprise systems with ODP*. CRC press, Taylor & Francis group, 2012.

[8] Y. Chen, P. Martin, H. Schentz, B. Magagna, Z. Zhao, A. Hardisty, A. Preece, M. Atkinson, R. Huber, and Y. Legre. Building a common reference model for environmental science research infrastructures. In *Proceedings of EnviroInfo2013*, 2013.

[9] Openlab at university of amsterdam, http://sne.science.uva.nl/openlab/.

[10] Sc13, the international conference on high performance computing, networking, storage and analysis, http://sc13.supercomputing.org/.

[11] Mark Hedges, Adil Hasan, and Tobias Blanke. Curation and preservation of research data in an irods data grid. *2012 IEEE 8th International Conference on E-Science*, 0:457–464, 2007.