Open Information Linking for Environmental Research Infrastructures

Martin, P.; Grosso, P.; Magagna, B.; Schentz, H.; Chen, Y.; Hardisty, A.; Los, W.; Jeffery, K.; de Laat, C.; Zhao, Z.

DOI
10.1109/eScience.2015.66

Publication date
2015

Document Version
Final published version

Published in
Proceedings, 11th IEEE International Conference on eScience

2015 IEEE 11th International Conference on eScience

Open Information Linking for Environmental Research Infrastructures

Paul Martin[1], Paola Grosso[1], Barbara Magagna[2], Herbert Schentz[2], Yin Chen[3],
Alex Hardisty[4], Wouter Los[1], Keith Jeffery[5], Cees de Laat[1], Zhiming Zhao[1]
[1]*University of Amsterdam, Netherlands, {p.w.martin|p.grosso|w.los|delaat|z.zhao}@uva.nl*
[2]*Environmental Agency Austria, Austria, {barbara.magagna|herbert.schentz}@umweltbundesamt.at*
[3]*EGI.eu, Netherlands, yin.chen@egi.eu*
[4]*Cardiff University, United Kingdom, HardistyAR@cardiff.ac.uk*
[5]*Natural Environment Research Council, United Kingdom, Keith.Jeffery@keithgjefferyconsultants.co.uk*

***Abstract*—Environmental research infrastructures (RIs) support data-intensive research by integrating large-scale sensor/observer networks with dedicated data curation services and analytical tools. However the diversity of scientific disciplines coupled with the lack of an accepted methodology for constructing new RIs inevitably leads to incompatibilities between the data models, metadata standards and service descriptions used by different RIs, inhibiting their usefulness for interdisciplinary research. In the absence of a common global ontology of science and infrastructure, these inconsistencies may best be counteracted by selectively bridging the semantics of the various vocabularies, standards and models used by the RIs at present. Open Information Linking for Environmental RIs (OIL-E) was developed within the FP7 project ENVRI to provide a framework for semantic linking of knowledge resources used by different environmental RIs. Built around a multi-viewpoint reference model ENVRI-RM, OIL-E is intended to act as a central exchange for linking information fragments and identifying gaps in the conceptual models of RIs.**

***Keywords*-research infrastructures; system-level science; reference model; e-research; e-science; semantic linking**

I. Introduction

The study of complex environmental problems is increasingly dominated by data-driven approaches whereby defining assumptions, extracting evidence and validating theories is based on the bulk analysis of large quantities of data in the form of observations, measurements, documents and other derived products [1]. For example understanding climate change requires analysis and integration of measurement data acquired not only from atmospheric readings, but also observation of the oceans, earth processes and the biosphere. It is difficult enough to model these environmental systems separately using classical simulation approaches, but it is especially challenging to model the interactions between them, requiring a certain degree of interdisciplinary collaboration. To enable good data-driven system-level science [2], researchers need not only good tools for searching, accessing and integrating data and software from a variety of sources, but also facilities to ease collaboration with other researchers of different disciplines.

Environmental science research infrastructures (RIs) integrate large deployments of sensors or observers (often on a continental scale) with dedicated facilities for data curation, typically providing a unified interface for discovering, accessing, and sometimes even processing that data, which is often distributed across multiple data centers. Most RIs specialize in a specific domain, e.g. atmospheric science (ICOS [3] and EISCAT_3D [4]), ocean/marine science (EURO-ARGO [5] and EMSO [6]), biodiversity (LifeWatch [7]), or solid earth science (EPOS [8]). These RIs are intended to become vital pillars supporting their respective research communities in the conduct of data-intensive research, but are also intended to contribute to wider-scale technical and political collaborations in Europe and beyond [9], [10], [11].

Despite these cooperation efforts, there exist characteristic differences between different environmental domains that have resulted in a diversity of standards for annotating, cataloguing and publishing data—such diversity makes the discovery, access and integration of data and services from different RIs difficult, especially for individual researchers and research teams that wish to conduct innovative science founded on global collaboration [12]. In the absence of global standards, it is necessary to consider efficient mechanisms for bridging the gulf between the different existing standards and models. Semantically linked metadata along with comprehensive descriptions of services and data sources essentially enable researchers to adopt new discovery, integration and processing tools to utilize data and services from different sources. Typical scenarios include:

1) *Accessing and harmonizing data from the catalogues and repositories of different RIs:* semantically linked metadata standards allow data discovery and integration tools to correctly query otherwise heterogeneous catalogue services.
2) *Selecting and combining data processing methods and tools from different sources:* semantically linked description models of service interfaces and of the languages that describe compositions of services allow the composition of complex workflows calling upon services from different RIs.
3) *Selecting optimal data and computing infrastructures*

978-1-4673-9325-6/15 $31.00 © 2015 IEEE
DOI 10.1109/eScience.2015.66

for executing applications: semantically mapped resource descriptions, from data catalogues to software, storage, computing and network requirements, enable execution scheduling tools to better select resources.

4) *Sharing and reusing architectures and products:* besides the actual execution of system-level science, semantically linked design documents allow developers to share architecture design and reuse results, which leads to more wide-spread implementation interoperability.

A traditional approach of working towards a single global ontology may work for a small number of RIs catering to specific scientific community, but is unrealistic for RIs in general—there are more than thirty large-scale developments underway in Europe at the time of writing [13], not including RIs funded only at a national level. Alternatively, a bottom-up approach based on bridging specific information gaps between two or more RIs in a limited context based on community demands can potentially promote the emergence of a connected semantic ecosystem for all RIs. However *ad hoc* semantic connections can only have limited influence guiding the future convergence of such an ecosystem given the lack of a framework by which to establish a global context within which to consider individual connections. To overcome this problem, the EU FP7 project ENVRI [14] adopted a hybrid approach based on a mix of a single core ontology and bottom-up semantic bridging, referred to as Open Information Linking for Environmental RIs (OIL-E). A reference model that abstracted a common taxonomy and common patterns from a cluster of environmental RIs [15], looking not only at the architectural design of those RIs but also the data lifecycle [16], served as the basis for OIL-E.

In this paper, we first give an introduction to the background of the ENVRI project and then describe the basic idea and current development state of OIL-E, finishing with a discussion of how future work might proceed.

II. Problem description and approach

Combining all environmental domains into one single RI is neither feasible in development nor manageable in operation. During the past several years, interoperability between infrastructures has been extensively studied, with different interoperability solutions proposed for different levels of interoperation: between computing infrastructures [17], [18], between middleware [19], and between computational workflows [20]. These solutions iteratively build adapters or connectors between two infrastructures and then derive new service standards via focusing community efforts. Such iteration promotes the evolution of services in infrastructures, but cannot fully realize infrastructure interoperability while these solutions only focus on specific layers of the global problem without considering the overall e-science context [21]. Meanwhile, White et al. [22] argued the importance of an ontological reference model in the development of interoperable services in infrastructure.

Table I
The current state of environmental RIs in the ENVRI^{PLUS} project.

RI	Domain	Status
ACTRIS	ATM	I3
ANAEE	BIO/ECO	PPP
EISCAT_3D	*ATM*	*PPP*
ELIXIR	BIO/ECO	Ope
EMBRC	MARINE, BIO/ECO	Con/Ope
EMSO	*MARINE*	*Ope*
EPOS	*SOLID*	*PPP*
ESONET	MARINE	Con/Ope
Euro-Argo	*MARINE*	*Ope*
EUROFLEETS	MARINE	I3
EUROGOOS	MARINE	Ope
FIXO3	MARINE	I3
IAGOS	ATM	Ope
ICOS	*ATM, MARINE, BIO/ECO*	*Con/Ope*
INTERACT	BIO/ECO	I3
IS-ENES	ATM	I3
JERICO	MARINE	I3
LifeWatch	*BIO/ECO*	*Con/Ope*
LTER	BIO/ECO	I3
SeaDataNet2	MARINE	Ope
SIOS	All domains	PPP

I3 = Integrated Infrastructures Initiative
PPP = Preparatory Phase Project
Con = under construction
Ope = operational

We now briefly look at the status of current environmental RIs in Europe, the requirements for semantic linking, and the approach taken by ENVRI.

A. Current status of environmental RIs

One of the objectives of ENVRI was to improve the interoperability of different environmental RIs by first abstracting common functions and patterns typical to all RIs, and then build a reference model for guiding system design along with a semantic linking framework for sharing information. The successor project to ENVRI, namely ENVRI^{PLUS}, has recently been granted funding under Horizon 2020 to further develop this ontological framework and utilize it within a larger cluster of environmental RIs. Table I shows the RIs currently involved with ENVRI^{PLUS}, with the original six RIs involved in ENVRI in italics.

We see that all RIs belong to at least one of four high-level domains: atmosphere (abbreviated ATM), bio- or ecological (BIO/ECO), ocean/marine (MARINE), and solid earth (SOLID). There is considerable variation in their states of development, with fewer than half operational, mostly in the marine domain. Different metadata standards have been observed from those RIs that are in operation, including NASA DIF [23] and SensorML [24] in EMSO, ISO 19115 [25] geospatial metadata in SeaDataNet and ISO 19139 [26] geospatial XML in EUROGOOS, and a combination of ISO 19115, INSPIRE [27] and NetCDF-CF [28] based standards in IAGOS [29]. One of the goals of ENVRI^PLUS is to link the different information models of these RIs and share resource descriptions, in order to implement interoperable services for accessing, processing and citing data.

B. Semantic linking and requirements

Semantic linking is often investigated in the context of ontology matching, mapping or alignment [30]. The key task is to compare similarity between entities from different semantic models and measure the similarity distances at different layers: the *data* layer, comparing data values and objects; the *ontology* layer, comparing the labels and concepts of entities; and the *context* layer, comparing semantic entities with inclusion of application contexts. We posit that there are five main application contexts applicable to environmental RIs. To semantically link conceptual models across different RIs, it is first necessary to determine which contexts are applicable to the particular information fragments being linked:

1) The *user and application* context includes description models for scientific methods (workflows), users (roles), and operational policies.
2) The *data* context includes data standards, metadata, data quality attributes, and other information models used to manage the data lifecycle.
3) The *computing* context includes models for describing interfaces, quality-of-service attributes, and the logical functionality of software tools, services and processes.
4) The *engineering* context includes models for describing physical infrastructures (e.g. of storage, computing and network) and other engineering issues (e.g. fault tolerance).
5) The *technology* context includes semantic models for describing software, hardware and standards.

Not only is it necessary to identify the specific context (or viewpoint) that a given information model applies to (the concerns of operational policies being very different from those of standards for service interfaces or schemas for metadata), but it is also important to identify the correspondences between entities and concepts defined in different contexts for the same infrastructure.

C. Challenges and the ENVRI approach

The typical process for semantic linking involves several iterations of the following steps: 1) preprocessing of features by a small set of excerpts of the overall ontology definition to describe a specific entity; 2) definition of the search space in the ontology for candidate alignment; 3) computation of the similarity between two entities from different ontologies; 4) aggregation of the different similarity results of each entity pair, depending on the algorithms used; and 5) derivation of the final linking between entities using different interpretation mechanisms, including the analysis of human experts.

Semantically linking information models from different environmental RIs is difficult however. The information resources (e.g., datasets, documents and descriptions) from different RIs often do not share common vocabularies given their individual idiosyncracies coupled with the different contexts these information sources address. Moreover, diverse metadata standards from different RIs (in particular, their potential evolutions or adjustments made in those RIs due to specific needs) make it difficult for any semantic linking model to be sustainable and usable by other RIs.

To handle these difficulties an effective linking model should identify and analyse information gaps between RIs by: 1) capturing the domain characteristics and specific viewpoints that information is based on; 2) structuring semantic links in the context of the entire lifecycle of data; and 3) designing new RIs with consideration of future support for system-level environmental research.

In ENVRI, Open Distributed Processing (ODP) [31] was used as the basis for modelling RIs. The ODP model captures the design and development issues in complex distributed systems from five viewpoints:

1) The *enterprise* viewpoint includes the concepts related to the core business and usage of the system, potential use cases, involved roles, behaviours and interactions. In ENVRI, this was referred to as the *science* viewpoint in deference to the domain of discourse.
2) The *information* viewpoint models the schemas of data objects in the systems and their permitted transitions.
3) The *computational* viewpoint models the operation and the binding interfaces used by the systems logical components.
4) The *engineering* viewpoint describes the distribution and connectivity between physical components of the system.
5) The *technological* viewpoint describes technology, standards, hardware and software deployed by the system.

Each of the different ODP viewpoints independently facilitates the design and understanding of a system for a different core purpose and agent perspective, but nevertheless there still exist correspondence points between the viewpoints that ensure they still effectively describe the same system. This multi-viewpoint modelling approach is very suitable for modelling complex RIs and can be supported in systems/software engineering tool chains (regardless of whether one personally favors the particular decomposition of views espoused by ODP); the viewpoints here match the RI application contexts described in the previous section.

In the ENVRI project, a semantic linking model (OIL-E) was proposed based on the multi-viewpoint model of ODP used by the ENVRI Reference Model (ENVRI-RM). ENVRI-RM abstracted the data and computing processes in the data lifecycle of a cluster of RIs and provided the basis for OIL-E, which was designed as a linking framework

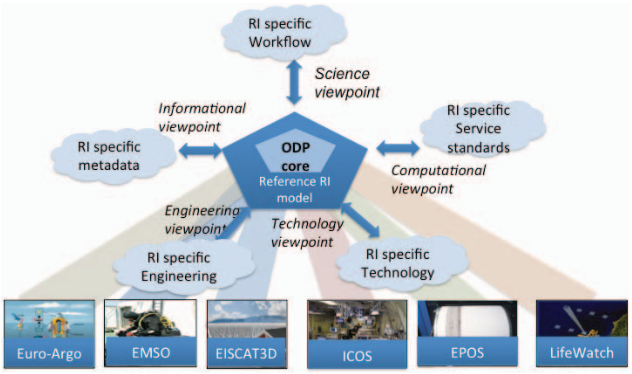

Figure 1. The abstract structure of OIL-E.

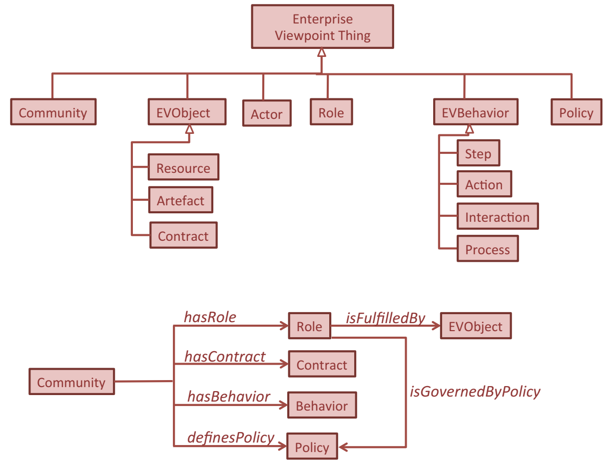

Figure 2. Sample of *enterprise* concepts described within the ODP ontology.

by which to build linking ontologies between the specific metadata standards, service description schemas and other semantic models used by ENVRI RIs.

III. Open Information Linking for Environmental Research Infrastructures

Fig. 1 shows the abstract structure of OIL-E in the ENVRI context. OIL-E has three parts:

1) The *core ontology* of ODP provides basic classes and properties to describe a system.
2) The *ENVRI-RM ontology* models the common functional components in environmental RIs from each of the five viewpoints defined by ODP.
3) The *linking ontology* connects the reference model with the information models outside research infrastructures, such as schemas for underlying physical infrastructures, and for domain specific data and service.

OIL-E has been developed by progression through a number of stages:

1) The development of an ODP core ontology to describe the semantics of components in complex distributed systems.
2) The development of an ENVRI-RM ontology to describe the semantics of data and resources in the six ENVRI research infrastructures.
3) A review of information models for data, metadata, resources and infrastructures in current ESFRI [32] projects, identifying a representative subset.
4) A definition of links between the ENVRI-RM ontology and the representative subset of models identified in the previous stage.
5) Some initial validation of OIL-E via use-cases.

We now go into that development in more detail.

A. Core ODP ontology

The ODP ontology defines the basic vocabulary for describing a distributed system. Alain et al. [33] discussed early work on an ODP ontology. Based on existing work and the ODP standard [34], we modelled the basic ODP vocabularies using OWL [35]. The ODP ontology provides a vocabulary for distributed systems from each of the five ODP viewpoints. Fig. 2 shows part of the ODP ontology for enterprise viewpoint concepts, which focus on user community, interaction behavior, system scope and purpose, and system policies.

B. ENVRI reference model ontology

The ENVRI reference model describes common concepts and components identified in the six ENVRI RIs by analysing their functionality and design documents from three of the five viewpoints (missing the engineering and technology viewpoints due to resource constraints and the lack of established standard implementations of key systems at the time of analysis[1]). The ENVRI-RM ontology serves to provide a standard structured vocabulary for these commonalities. Chen et al. [15] describes some of the common elements identified in the RIs and further deconstructs the archetypical RI into five subsystems for *data acquisition*, *data curation*, *data access*, *data processing* and *community support*, which can be used to deconstruct RIs othogonally to the viewpoints in ENVRI-RM.

The ENVRI-RM science (enterprise) viewpoint defines concepts specific to environmental RIs based on the ODP enterprise view, presenting a schema for how RIs interact with a research community. The information view models the various states and schemas of data at various points in the RI data lifecycle, capturing the evolution of research data from raw input to published results. The computational viewpoint captures the internal dependencies between common

[1]ENVRI^{PLUS} is intended to redress this omission based on some of the new developments that have occurred in recent years.

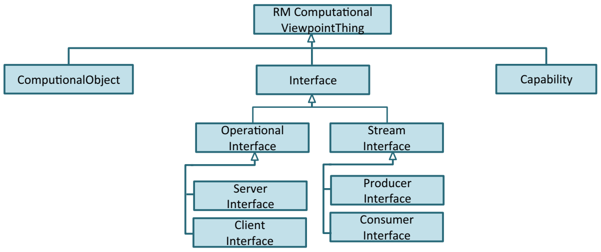

Figure 3. Sample of computational viewpoint concepts described within the ENVRI-RM ontology.

logical operations that ESFRI infrastructures share (such as data transfer, cataloguing, instrument configuration, etc.) by identifying essential components and their interfaces with one another (see for example Fig. 3, as well as [36]).

C. Linking ENVRI-RM with other semantic fragments

The linking component of OIL-E glues concepts both *inside* ENVRI-RM and *between* ENVRI-RM and external concepts belonging to outside vocabularies. The ENVRI-RM ontology only contains a limited set of vocabularies derived from common functionality and patterns, so linking ENVRI-RM with external RI-specific concepts will enable RI-specific extensions to the ENVRI-RM vocabulary. Similarly, linking ENVRI-RM with external vocabularies provides bridge between those vocabularies and ENVRI-RM, and indirectly between the vocabularies themselves. Notably, the internal correspondences between different ENVRI-RM viewpoints (enterprise, information, etc.) can potentially be used to indirectly link external vocabularies of quite different foci (data, services, infrastructure, etc.).

1) Correspondences between internal concepts: In the ODP model, concepts from different viewpoints can correspond to one another. For example convolutions of data (information viewpoint) correspond to specific abstract interactions (science view) and to interface bindings between functional components (computational view). In ENVRI, we examined correspondences between ENVRI-RM concepts based on not only definitions in the ODP model, but also for specific interactions that arise in the operation of analyzed RIs. Fig. 4 shows how concepts from different viewpoints can be linked when considering the configuration of instruments used for data acquisition in most RIs.

2) Linking between ENVRI-RM and external concepts: ENVRI-RM was developed based on six RIs, which were all in their preparation phase during ENVRI. It was thus not intended for ENVRI-RM to include all intended standards for data and service metadata, not least due to the predictably rapid evolution of environmental RIs in general in the intervening years. The information that is not included in ENVRI-RM must be connected via the different viewpoints of ENVRI-RM, as was illustrated in Fig 1.

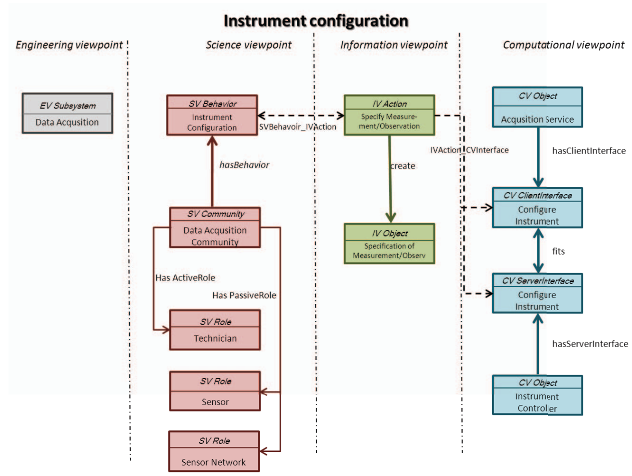

Figure 4. Correspondence between concepts from different viewpoints on instrument configuration.

Table II
METADATA STANDARDS MAPPED TO KEY RI SUBSYSTEMS IDENTIFIED BY ENVRI-RM.

Standard	Acq.	Cur.	Acc.	Pro.	Com.
SensorML	Y	Y		Y	
NetCDF			Y	Y	
ISO19115	Y	Y		Y	Y
ISO19156	Y	Y			
CSR	Y	Y			
Dublin Core				Y	

Acq. = Data acquisition, Cur. = Data curation, Acc. = Data access, Pro. = Data processing, Com. = Community support.

We use the information viewpoint as an example to demonstrate how linking is performed. During ENVRI, we reviewed a list of metadata standards currently used by ENVRI RIs, including Dublin Core (ISO 15836) [37], SensorML, ISO 19156 (geographic observations and measurements) [38], ISO 19115, SeaDataNet Cruise Summary Reports metadata [39], CERIF [40], CSMD [41] and INSPIRE. These standards can be linked via the information viewpoint of ENVRI-RM and mapped to functional subsystems of RIs, as demonstrated by Table II. Information viewpoint concepts in ENVRI-RM were then mapped to concepts found in those standards as further elaborated in [36].

D. Use-cases and validation

OIL-E was developed and given some initial evaluation within the ENVRI project. RI use-cases were collected both to guide the modelling of OIL-E and to validate its component ontologies. OIL-E can be validated via three phases:

1) Using OIL-E to annotate the natural language description of the collected scenarios.
2) Validating the common operations or services derived

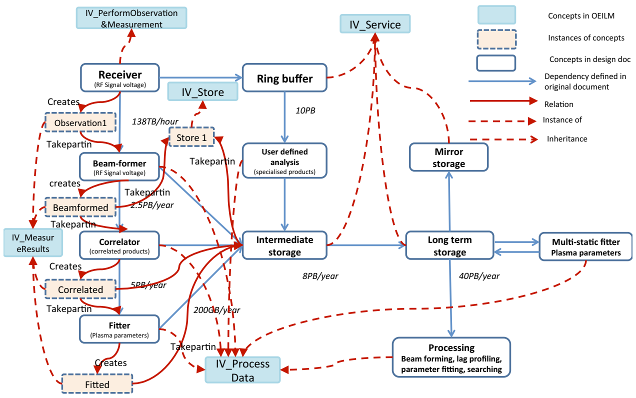

Figure 5. Example of scenario annotation using OIL-E.

from the annotated description by interviewing technical experts from the RIs.

3) Applying OIL-E based descriptions in the applications to realize the optimal resource selection from different RIs, for instance for complex workflow execution.

Fig. 5 shows an example of using OIL-E concepts to annotate a dataflow, in this case for EISCAT_3D. By annotating dataflows, design documents from different RIs can then be compared and common operations (and gaps in functionality) inferred. Future work in ENVRI^{PLUS} will provide opportunity to significantly expand on the validation and evaluation of OIL-E performed thus far.

IV. Summary

In this paper, we introduced a semantic linking framework called OIL-E, which was developed in the ENVRI project to integrate information fragments in environmental RIs in order to enhance interoperability among RIs and to provide further basis for inter-RI resource allocation, scheduling and optimization. While the approach seems promising, further development is needed to prove its practical value: the ENVRI^{PLUS}project provides an opportunity to test it further by linking the information fragments of a much broader set of RIs in order to realize a core set of common services.

A. Discussion

As a rapidly growing field, the development of services in data-oriented RIs is driven by research activities both inside and outside the intended scientific domain. For instance extending metadata models to a broader range of RIs is important for promoting interoperable semantic frameworks for RIs—Jeffery et al. [40] highlights discovery, context and details as three basic levels in the CERIF system to organize the metadata models between RIs. It highlights formal information systems in the model, but does not focus on same level of architecture design details in the way that OIL-E does. Linking CERIF with ENVRI RM (or its successor model) will one of the tasks investigated within the ENVRI^{PLUS} project.

How to model the as-yet-unknown facets of future RI projects and keep ENVRI-RM and OIL-E open and extensible (in particular how to bring the models into practical use by RI developers) will remain important questions. Moreover, the data-driven e-science experiments that the RIs intend to support often require customized processing services for special research purposes. How to balance the constraints of developing new services and adapting existing ones from other research infrastructures will be important issue when promoting the ENVRI RM and OIL-E to the architect/developer community targeted by ENVRI and now ENVRI^{PLUS}.

B. Conclusions and future work

From the current work and reading of associated literature, we can argue that:

- Semantic linking is a pragmatic means to support interoperability between data and services from different research infrastructures. An effective reference model synchronizes the vocabularies used in different environmental RIs, and can potentially guide the further development of common operations and functional components in RIs provided that there is sufficient incentive for developers to actually use the model.
- Modelling a distributed system like a research infrastructure requires decomposition of modelling issues based on different stakeholder viewpoints. The Open Distributed Processing model provides a suitable mechanism to do this.
- Semantic web technologies provide an open framework for modelling linking between different elements in research infrastructures. If a semantic linking framework is important for realizing interoperability between research infrastructures, then Open Information Linking for Environmental Sciences (OIL-E) should be a step in the right direction if properly developed.

In the ENVRI^{PLUS}project, OIL-E will be extended both horizontally to a broader cluster of RIs, and vertically to formal models of the services and data required by the environmental RIs. This further development will provide tools for semantically describing data, services and technologies, provide flexible mechanisms to keep descriptions adaptable when technical details change, and provide tools to interlink and map between high level data and services from different RIs to bridge any gaps at both conceptual representation level and at data processing levels.

Acknowledgment

This work was supported by the European Unions FP7 research and innovation programme under grant agreement No. 283465 (ENVRI project) and Horizon 2020 research and innovation programme under grant agreements No. 654182 (ENVRI^{PLUS} project) and 643963 (SWITCH project).

References

[1] T. Hey, S. Tansley, and K. Tolle, Eds., *The fourth paradigm: data-intensive scientific discovery*. Microsoft Research, 2009.

[2] I. Foster and C. Kesselman, "Scaling system-level science: Scientific exploration and it implications," *Computer*, no. 11, pp. 31–39, 2006.

[3] Integrated Carbon Observation System (ICOS). Accessed: 2015-07-21. [Online]. Available: https://www.icos-ri.eu/

[4] EISCAT Scientific Association. EISCAT_3D. Accessed: 2015-07-21. [Online]. Available: https://eiscat3d.se/

[5] Euro-Argo—European contribution to Argo program. Accessed: 2015-07-21. [Online]. Available: http://www.euro-argo.eu/

[6] European multidisciplinary seafloor & water column observatory (EMSO). Accessed: 2015-07-21. [Online]. Available: http://www.emso-eu.org/

[7] LifeWatch—e-Science European infrastructure for biodiversity and ecosystem research. Accessed: 2015-07-21. [Online]. Available: http://www.lifewatch.eu/

[8] European Plate Observing System (EPOS). Accessed: 2015-07-21. [Online]. Available: http://www.epos-eu.org/

[9] E. Christian, "Planning for the global earth observation system of systems (GEOSS)," *Space Policy*, vol. 21, no. 2, pp. 105–109, 2005.

[10] C. Jacobs, "A vision for, and progress towards EarthCube," in *EGU General Assembly Conference Abstracts*, vol. 14, 2012, p. 1227.

[11] C. Waldmann, J. Pearlman, and R. Huber, "COOPEUS — strengthening the transatlantic cooperation on common data policies and standards for research infrastructures in the environmental field," in *EGU General Assembly Conference Abstracts*, vol. 15, 2013, p. 10816.

[12] C. L. Smith *et al.*, *Knowledge, Networks and Nations: Global Scientific Collaboration in the 21st Century*, The Royal Society, 2011.

[13] A. Robin, A. M. Johansson, and J. Delahaut, *Global change: towards global research infrastructures (updated edition, 2014)*, European Commission, Directorate-General for Research Innovation, 2014.

[14] ENVRI—Common operations of environmental research infrastructures. Accessed: 2015-07-21. [Online]. Available: http://www.envri.eu/

[15] Y. Chen, P. Martin, B. Magagna, H. Schentz, Z. Zhao, A. Hardisty, A. D. Preece, M. P. Atkinson, R. Huber, and Y. Legré, "A common reference model for environmental science research infrastructures." in *EnviroInfo*, 2013, pp. 665–673.

[16] Z. Zhao, P. Grosso, and C. de Laat, "OEIRM: An open distributed processing based interoperability reference model for e-science," in *Network and Parallel Computing*. Springer, 2012, pp. 437–444.

[17] Y. Charalabidis, M. Janssen, and O. Glassey, "Introduction to cloud infrastructures and interoperability minitrack," in *2012 45th Hawaii International Conference on System Sciences*. IEEE, 2012, p. 2177.

[18] L. D. Ngan, Y. Feng, S. Rho, and R. Kanagasabai, "Enabling interoperability across heterogeneous semantic web services with OWL-S based mediation," in *Services Computing Conference (APSCC), 2011 IEEE Asia-Pacific*. IEEE, 2011, pp. 471–476.

[19] G. Blair and P. Grace, "Emergent middleware: Tackling the interoperability problem," *IEEE Internet Computing*, no. 1, pp. 78–82, 2012.

[20] Z. Zhao, S. Booms, A. Belloum, C. De Laat, and B. Hertzberger, "VLE-WFBus: a scientific workflow bus for multi e-science domains," in *e-Science and Grid Computing, 2006. e-Science'06. Second IEEE International Conference on*. IEEE, 2006, pp. 11–11.

[21] M. Riedel, E. Laure, T. Soddemann, L. Field, J.-P. Navarro, J. Casey, M. Litmaath, J. P. Baud, B. Koblitz, C. Catlett *et al.*, "Interoperation of world-wide production e-science infrastructures," *Concurrency and Computation: Practice and Experience*, vol. 21, no. 8, pp. 961–990, 2009.

[22] L. White, N. Wilde, T. Reichherzer, E. El-Sheikh, G. Goehring, A. Baskin, B. Hartmann, and M. Manea, "Understanding interoperable systems: Challenges for the maintenance of SOA applications," in *System Science (HICSS), 2012 45th Hawaii International Conference on*. IEEE, 2012, pp. 2199–2206.

[23] Z. B. Miled, S. Sikkupparbathyam, O. Bukhres, K. Nagendra, E. Lynch, M. Areal, L. Olsen, C. Gokey, D. Kendig, T. Northcutt *et al.*, "Global change master directory: object-oriented active asynchronous transaction management in a federated environment using data agents," in *Proceedings of the 2001 ACM symposium on Applied computing*. ACM, 2001, pp. 207–214.

[24] Open Geospatial Consortium. Sensor Model Language (SensorML). Accessed: 2015-07-21. [Online]. Available: http://www.opengeospatial.org/standards/sensorml

[25] *Geographic information—Metadata*, International Organization for Standardization Std. 19 115, 2014.

[26] *Geographic information—Metadata—XML schema implementation*, International Organization for Standardization Std. 19 139, 2007.

[27] Infrastructure for Spatial Information in the European Community (INSPIRE). Accessed: 2015-07-21. [Online]. Available: http://inspire.ec.europa.eu/

[28] Latest CF Conventions Documents (NETCDF-CF). Accessed: 2015-07-22. [Online]. Available: http://cfconventions.org/latest.html

[29] D. Boulanger, B. Gautron, V. Thouret, M. Schultz, P. van Velthoven, B. Broetz, A. Rauthe-Schöch, and G. Brissebrat, "Latest developments for the IAGOS database: Interoperability and metadata," in *EGU General Assembly Conference Abstracts*, vol. 16, 2014, p. 6510.

[30] M. Ehrig, *Ontology Alignment: Bridging the Semantic Gap*, ser. Semantic Web and Beyond. Springer, 2007, vol. 4.

[31] P. F. Linington, Z. Milosevic, A. Tanaka, and A. Vallecillo, *Building enterprise systems with ODP: an introduction to open distributed processing*. CRC Press, 2011.

[32] European Strategic Forum of Research Infrastructure (ESFRI). Accessed: 2015-07-21. [Online]. Available: http://ec.europa.eu/research/esfri/

[33] A. Wegmann and A. Naumenko, "Conceptual modeling of complex systems using an rm-odp based ontology," in *Enterprise Distributed Object Computing Conference, 2001. EDOC'01. Proceedings. Fifth IEEE International*. IEEE, 2001, pp. 200–211.

[34] *Information technology—Open Distributed Processing—Reference model*, International Organization for Standardization Std. 10 746, 1998.

[35] W3C OWL Working Group. OWL 2 Web Ontology Language Document Overview (Second Edition). Accessed: 2015-07-24. [Online]. Available: http://www.w3.org/TR/owl-overview

[36] Z. Zhao, P. Grosso, C. de Laat, B. Magagna, H. Schentz, Y. Chen, A. Hardisty, P. Martin, and M. Atkinson. (2014) Interoperability framework for linked computational, network and storage infrastructures, version 2. Accessed: 2015-07-21. [Online]. Available: http://envri.eu/

[37] *Information and documentation—The Dublin Core metadata element set*, International Organization for Standardization Std. 15 836, 2009.

[38] *Geographic information—Observations and measurements*, International Organization for Standardization Std. 19 156, 2011.

[39] SeaDataNet Cruise Summary Reports. Accessed: 2015-07-24. [Online]. Available: http://www.seadatanet.org/Standards-Software/Metadata-formats/CSR

[40] K. Jeffery, N. Houssos, B. Jörg, and A. Asserson, "Research information management: the cerif approach," *International Journal of Metadata, Semantics and Ontologies*, vol. 9, no. 1, pp. 5–14, 2014.

[41] The Core Scientific Metadata Model (CSMD). Accessed: 2015-07-21. [Online]. Available: http://purl.org/net/CSMD