Reference Model Guided System Design and Implementation for Interoperable Environmental Research Infrastructures

Zhao, Z.; Martin, P.; Grosso, P.; Los, W.; de Laat, C.; Jeffrey, K.; Hardisty, A.; Vermeulen, A.; Castelli, D.; Legré, Y.; Kutsch, W.

DOI
10.1109/eScience.2015.41

Publication date
2015

Document Version
Final published version

Published in
Proceedings, 11th IEEE International Conference on eScience

Reference Model Guided System Design and Implementation for Interoperable Environmental Research Infrastructures

Zhiming Zhao, Paul Martin, Paola Grosso,
Wouter Los, Cees de Laat
Informatics Institute
University of Amsterdam, the Netherlands
{Z.Zhao|P.W.Martin|P.Grosso|W.Los|Delaat}@uva.nl

Keith Jeffrey
Natural Environment Research Council
United Kingdom
Keith.Jeffery@keithgjefferyconsultants.co.uk

Alex Hardisty,
Cardiff University, United Kingdom
HardistyAR@cardiff.ac.uk

Alex Vermeulen
Lund University, Sweden
alex.vermeulen@nateko.lu.se

Donatella Castelli
Consiglio Nazionale delle Ricerche, Italy
donatella.castelli@isti.cnr.it

Yannick Legré
EGI.eu, Netherlands
yannick.legre@egi.eu

Werner Kutsch
University of Helsinki, Finland
Werner.kutsch@helsinki.fi

***Abstract*—Environmental research infrastructures (RIs) support their respective research communities by integrating large-scale sensor/observation networks with data curation services, analytical tools and common operational policies. These RIs are developed as service pillars of intra- and interdisciplinary research, however comprehension of the complex, interconnected aspects of the Earth's ecosystem increasingly requires that researchers conduct their experiments across infrastructure boundaries. Consequently, almost all data-related activities within these infrastructures, from data capture to data usage, needs to be designed to be broadly interoperable in order to enable real interdisciplinary innovation. The *Data for Science* theme in the EU Horizon 2020 project ENVRI^PLUS intends to address this interoperability challenge as it relates to the design, implementation and operation of environmental science RIs; the theme focuses on key issues of data identification and citation, curation, cataloguing, processing, optimization, and provenance, supported by a generic cross-infrastructure reference model.**

Keywords—research infrastructure; system-level science; reference model; e-research; e-science; environmental research

I. INTRODUCTION

Many key problems in environmental science are intrinsically interdisciplinary; the study of climate change for example involves the study of atmosphere, but also earth processes, the oceans and the biosphere. Modelling these processes individually is difficult enough, but modelling their interactions is another order of complexity entirely. Scientists are challenged to collaborate across conventional disciplinary boundaries, but must first discover and extract data dispersed across many different sources and formats.

Data-driven research differs from classical approaches for analytical modelling or computer simulation insofar as new theories are measured first and foremost against huge quantities of observations, measurements, documents and other data sources culled from a range of possible sources. To enable such science, the underlying research infrastructure must provide not only the necessary tools for data discovery, access and manipulation, but also facilities to enhance collaboration between scientists of different backgrounds.

Environmental research infrastructures (RIs) support user communities by providing federated data curation, discovery and access services, analytical tools and common operational policies integrated around large-scale sensor/observer networks, often deployed on a continental scale. Examples in Europe include LifeWatch [1] (concerned with biodiversity), EPOS [2] (solid earth science), EURO-ARGO [3] and EMSO [4] (ocean monitoring), and ICOS [5] (atmosphere). These infrastructures are developing into important pillars of their respective user communities, but are also intended to support interdisciplinary research as well as more specifically Copernicus [6] as a contribution to GEOSS [7]. As such, it is very important that data-related activities are well integrated in order to enable data-driven system-level science [8]. This requires standard policies, models and e-infrastructure to improve technology reuse and ensure coordination, harmonization, integration and interoperability of data, applications and other services. We argue that key to this are interoperable metadata, and more specifically a rich metadata model acting as the 'switchboard' for interoperation—one with formal syntax and well-defined semantics. However, the complex nature of environmental science seems to result in the development of environmental

978-1-4673-9325-6/15 $31.00 © 2015 IEEE
DOI 10.1109/eScience.2015.41

RIs that meet only the requirements and needs of their own specific domains, with very limited interoperability of data, and isolated tools and operational policies.

The newly funded EU Horizon 2020 project ENVRI^{PLUS} follows up on the ENVRI project [9], but involves a much broader cluster of research infrastructures and has a clearer focus on the implementation of common operations. ENVRI was concerned with the identification of technical and organizational commonalities between environmental RIs and the prototyping of unified data discovery and access services; ENVRI^{PLUS} expands upon those objectives with a greater remit to implement common services applicable to the whole RI activity cycle. More specifically, the *Data for Science* theme within ENVRI^{PLUS} encompasses work on the development of common data services within a common semantic framework. In this theme, we will develop an engineering model for constructing domain-specific environmental RIs that are nevertheless broadly interoperable. Based on the results and lessons from ENVRI, this engineering model will: 1) combine both domain-specific characteristics and common abstractions; 2) harmonise RI-specific requirements with common operations; and 3) account for both existing generic e-infrastructures and solutions already adopted by existing RIs.

This paper first discusses the key challenges in the *Data for Science* theme of ENVRI^{PLUS}, and then presents the basic approach and development agenda adopted by this theme.

II. PROBLEM DESCRIPTION AND APPROACH

To perform system-level environmental science, scientists face challenges with respect to data accessing, processing and publication:

1. *Obtaining and harmonizing data from different sources:* Data are often in different formats, annotated using different metadata, and retrieved via catalogues with different interfaces.
2. *Selecting and combining data processing models from different domains:* Data processing models are often represented as workflows in different languages, and require different execution engines to realize.
3. *Selecting optimal infrastructure upon which to execute applications:* Infrastructures often provide different scheduling and monitoring tools.
4. *Publishing data objects in different research infrastructures:* Data objects should be both identifiable and citable regardless of their provenance.

Environmental RIs provide the tools to help with this, but only if their services are sufficiently interoperable.

A. Interoperability requirements for system-level environmental science

To enable interdisciplinary research across RIs from different sub-domains of environmental science, there are a number of principles that any interoperable services and their supporting infrastructure should adhere to:

- *Simple but effective:* Scientists should be able to use, analyze, compose and store data from distributed sources in an easy but effective way, with appropriate metadata generated at all stages in order to trace data provenance.
- *Bridgeable semantics:* A certain degree of semantic mapping is required to bridge the diverse complex knowledge organizing systems needed by different scientific and technical domains, but tools and resources need to be documented in a principled, formal way first.
- *Extensible and robust:* Available resources change and user demands fluctuate; core RI services must be elastic and fault-tolerant, and provide programmatic interfaces for service composition.
- *Open yet secure:* Although most research data is open, there is a need to protect the privacy of researchers, attribute credit to individuals and organizations, embargo new research prior to publication and preserve authority and accountability constraints when transferring data between different technical and political domains.

In order to meet these rather wide-ranging principles, the ENVRI^{PLUS} solutions will build upon the results of earlier projects, the expertise of individual RIs, and the services of e-infrastructure initiatives such as EGI [10] and HELIX-Nebula [11]. Filling in the gaps, we aim to:

1. Optimize data processing and develop common models, rules and guidance for research data workflow documentation.
2. Facilitate data discovery and use, and provide integrated end-user information technology to access heterogeneous data sources.
3. Make data citable by building upon existing approaches with practical examples, exchanges of expertise, and agreements with publishers.
4. Facilitate discovery of software services and their possible compositions.
5. Characterize users and build a community on top of existing RI communities.
6. Characterize ICT resources (including sensors and detectors) to allow virtualization of the environment (for instance onto grid- or cloud-based platforms) such that data and information management and analysis is optimized in terms of resource and energy expenditure.
7. Facilitate the connection of users, composed software services, appropriate data and necessary resources in order to meet end-user requirements.

B. State of the art

Table 1 shows the list of research infrastructures that participate the ENVRI^{PLUS} project, together with their development status. These RIs collectively play an important role in environmental research in Europe, with more than half of them prioritized in the roadmap of the European Strategy Forum on Research Infrastructures (ESFRI) [12], and several established as distinct legal entities or in the process of doing so.

TABLE 1—THE CURRENT STATE OF ENVIRONMENTAL RIS IN THE ENVRI PLUS PROJECT.

	Domain	Current status	On ESFRI roadmap	Legal entity?
ACTRIS	ATM	I3	No	No
ANAEE	BIO/ECO	PPP	Yes	No
EISCAT_3D	ATM	PPP	Yes	Yes
ELIXIR	BIO/ECO	Ope	Yes	Yes
EMBRC	MARINE, BIO/ECO	Con/Ope	Yes	No
EMSO	MARINE	Ope	Yes	In process
EPOS	SOLID	PPP	Yes	No
ESONET	MARINE	Con/Ope	No	Yes
EURO-ARGO	MARINE	Ope	Yes	Yes
EUROFLEETS	MARINE	I3	No	No
EUROGOOS	MARINE	Ope	No	Yes
FIXO3	MARINE	I3	No	No
IAGOS	ATM	Ope	Yes	Yes
ICOS	ATM, MARINE BIO/ECO	Con/Ope	Yes	In process
INTERACT	BIO/ECO	I3	No	No
IS-ENES	ATM	I3	No	No
JERICO	MARINE	I3	No	No
LifeWatch	BIO/ECO	Con/Ope	Yes	In process
LTER	BIO/ECO	I3	No	No
SeaDataNet2	MARINE	Ope	No	No
SIOS	All domains	PPP	Yes	No

I3 = Integrated Infrastructures Initiative
PPP = Preparatory Phase Project
Con = Under construction
Ope = Operational

From Table 1, we can see that the RIs are all based around four environmental domain: atmosphere (abbreviated ATM); bio- or ecological (BIO/ECO); marine (MARINE); and solid earth (SOLID). There is considerable variation in their states of development—only approximately a third of RIs are operational, mostly in the marine domain.

From current literature, we learn that existing interoperability solutions focus on specific layers: infrastructure [13, 14]; middleware [15]; and workflow [16]. Typically these solutions are realized iteratively, building adapters or connectors between two components and then deriving new service layer models for standardization via community effort. Such a process of iteration can gradually promote the evolution of new standards for both infrastructures and the service layers above them, but will not completely solve all interoperability problems while the diversity between infrastructures and the gaps between standards remain significant [17]. Providing interoperability solutions only for a specific service layer without a deep view of the complete technology stack hampers the convergence of service layer design. White et al. [18] argued that an interoperability reference model is needed to complement models of application and infrastructure. In ENVRI we argued that such a reference model should be applied early at the design phase of new systems [19].

For those environmental RIs that are currently under construction or in preparation, it therefore becomes urgent to guide their development so that they can be immediately interoperable once operational.

C. *The reference model guided approach*

The reference model guided approach in the ENVRI PLUS *Data for Science* theme expands upon developments from the preceding ENVRI project. It includes three parts, as shown in Fig. 1.

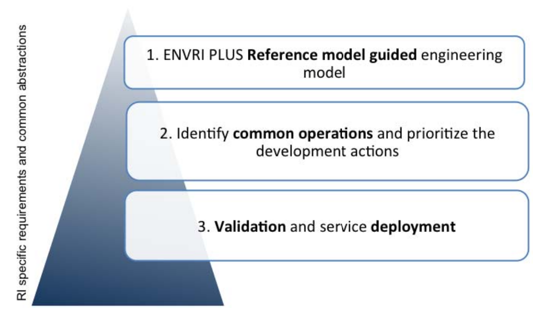

Figure 1—The basic idea of the reference model guided approach.

The **ENVRI PLUS reference model guided engineering model** builds upon abstracted concepts derived from analyzing common operations of a selected set of RIs and subsequently defines an ontological reference model for all environmental research infrastructures. To benefit from existing technologies, we will review early results from specific RIs and interact with computational e-infrastructures (such as EGI), data infrastructures (such as EUDAT [20]), and other initiatives (such as D4Science [21]) that work on related issues. We will review other interoperation technologies including CERIF [22] from EPOS for describing datasets, users, software, facilities, services and resources, and DCAT [23] for high-level exposure of basic dataset information.

Based on the requirements collected from each of the four main environmental science domains and their respective RIs, we will **identify and develop common operations**, by characterizing RI's individual current solutions with consideration given to underlying common technologies and engineering challenges. These individual operations will be characterized in terms of the engineering model, which will then be used in the design and implementation of common operations.

This approach is used to (a) reduce risk; (b) maximize utilization of e-infrastructures in individual RIs developed with EC or other public funding; (c) provide an opportunity for convergence of ideas among the RIs without discarding work already done; and (d) maximize the chances of successful interoperation between environmental RIs both technically and socially.

The third step is **validation and service deployment**, deploying the implemented common operations within generic e-infrastructures (such as EGI or EUDAT), and operating them in the service of specific RIs. This approach aligns with ongoing work and trends in the provision of e-Infrastructure—especially grid-based (e.g. EGI), cloud-based (e.g. HELIX-Nebula) and data-centric projects (e.g.

EUDAT), as well as the developments being proposed (and implemented) under the umbrella of RDA [24].

III. REFERENCE MODEL FOR ENVIRONMENTAL RESEARCH INFRASTRUCTURES

A reference model for a computational system provides an ontological framework for involved parties to clearly communicate. In both the ENVRI and ENVRI^PLUS projects, a reference model has been recognized as a promising contribution for realizing interoperability for diverse environmental research infrastructures. In this section, we will first review the work of the ENVRI Reference Model, and then summarize the lessons learned. Afterwards, we will discuss the approach for the ENVRI^PLUS Reference Model.

A. ENVRI Reference Model and lessons learned

In the ENVRI project, the development of the Reference Model (ENVRI-RM) was based on analysis of six RIs involved in the project: ICOS, EURO-Argo, EISCAT_3D [25], LifeWatch, EPOS, and EMSO. By interviewing specialists from each of these RIs, and examining the requirements, design documents, and use-cases collected, we abstracted some common operations and design patterns. This analysis had to cope with different viewpoints and varying vocabularies between (and even within) RIs.

The methodology for developing ENVRI-RM was to decompose system descriptions based on viewpoints. Open Distributed Processing (ODP) [26] provides five viewpoints from which to describe systems: *enterprise* (about system scenarios, involved communities and roles), *computation* (about system interfaces and bindings between system components), *information* (about data objects and schemas of the system), *engineering* (about system middleware, engineering principles) and *technology* (technology standards and decisions). This decomposition of complex systems by viewpoint is a useful technique for managing that complexity and providing information tailored to different kinds of stakeholder. ENVRI-RM employs these viewpoints to model the characteristics of environmental research infrastructures. The current version is available online [27].

ENVRI also looked at semantic linking. Linking the semantics of services and data sources related to the data lifecycle potentially enables further interoperability between RIs. A generic and globally operational ontology applicable for all possible situations is probably not achievable, not to mention cumbersome to develop. Instead, interrelating pairs of ontologies via semantic bridges may promote a pragmatic solution for previously unrelated semantic descriptions. Open Information Linking for Environment systems (OIL-E) uses ENVRI-RM to identify and formulate the commonality between the RIs, and has a linking layer that can be used to bridge the semantic gaps between ENVRI-RM and other specific ontologies in RIs.

OIL-E has three layers that compartmentalize the different key components of the framework (viewpoint decomposition, environmental RI modeling, and semantic linking). These layers are the ODP ontology, the ENVRI-RM ontology that extends the ODP ontology and defines RI-specific vocabularies, and the linking ontology used to connect ENVRI-RM with RI-specific semantic components. Fig. 2 shows the basic idea of ENVRI-RM and OIL-E.

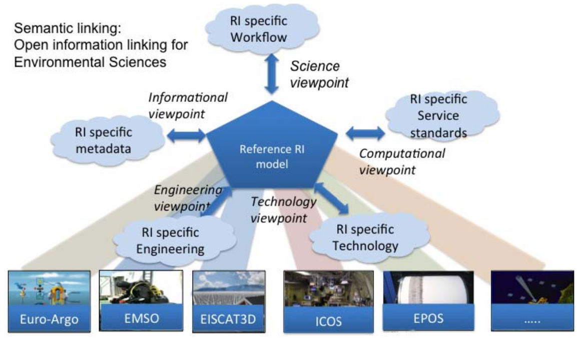

Figure 2--The basic idea of ENVRI-RM and semantic linking.

1) Lessons learned

ENVRI-RM focused on the design of a small set of RIs and was produced at a time when most of them were in their preparatory phase of development. Since ENVRI began, many them have made significant progress in their development, to some extent exceeding the expressiveness of ENVRI-RM. As such, a number of lessons can be drawn:

1. The rapid evolution of environmental RIs was not sufficiently taken into account when building ENVRI-RM; there are mismatches between new insights in the RIs and what was encoded in the model.
2. Supporting use cases are required for validation and demonstration; the lack of this in ENVRI led to drifting requirements and difficulty explaining the model to potential users.
3. The development of the model did not involve enough domain-aware ICT specialists from the RIs themselves. This was partly due to the early development state of the RIs, but meant that the model was not really applied to that development.

B. ENVRI^PLUS Reference Model and semantic linking

The extended ENVRI^PLUS Reference Model takes the above lessons into account. It takes input not only from ENVRI-RM but also from a wider range of RIs, as well as from developments in generic data and computing infrastructures postdating the original ENVRI project. Use-cases, for instance integrating European marine biological data (i.e. data curated by EMSO, SeaDataNet [28], JERICO [29] and EMBRC [30]) as a joint contribution to EMODNET Biology [31] the COPERNICUS provider, are being collected during the project and will be used for validating and improving the new model. This time, the development team explicitly includes participants from the actual RIs.

Semantic linking between architecture descriptions from different RIs, and services descriptions from other e-infrastructures, will be included in a new version of OIL-E, which will: (a) establish a common semantic framework for ENVRI^PLUS RIs, abstracting and harmonizing the models and knowledge organization systems already available in different domains; (b) provide tools for semantically describing data, services and technologies; (c) provide flexible mechanisms to keep descriptions adaptable when technical details change; and (d) provide tools to map between high level data and services from different RIs to bridge any gaps.

IV. COMMON SERVICES AND INTEROPERABILITY

The Reference Model assists in defining commonalities in the operations of environmental RIs, e.g. common services that support a particular sub-domain of environment research, or set of such sub-domains. ENVRI^{PLUS} is not concerned with unique services inside a specific RI. The focus on is on common services that are useful for significant subsets of environmental RIs.

A. Highlighted interoperable services

We have identified six common concerns based on the demands of the RIs involved in ENVRI^{PLUS}, which we will work to provide solutions for.

Data identification and citation requires the implementation of a common policy model for handling persistent identifiers for publishing and citing data. Moreover, services for assigning and handling identifiers and for retrieving data based on identifiers should also be provided. The service provided by ENVRI^{PLUS} will be built on existing approaches (e.g. DataCite [32]) and further developments by upcoming activities. It will be operated in close cooperation with existing initiatives (e.g. RDA and ICSU WDS [33]) and will elaborate a common data citation solution for all the involved RIs.

Interoperable data processing, monitoring and diagnosis services make it significantly easier for scientists to aggregate data from multiple sources and to conduct a range of experiments and analyses upon those data. Expanding upon the data processing workflow modeled in ENVRI, this service focuses on the engineering aspects of managing the entire lifecycle of computing tasks and application workflows for efficient utilization of underlying e-infrastructure. In particular, the service enables scientists to enrich the data processing environment by injecting new algorithms to be reused by others.

Performance optimization for big data science is increasingly required in environmental science. ENVRI^{PLUS} will focus on high-level, generically-applicable optimization mechanisms for making decisions on resources, services, data sources and potential execution infrastructures [34], and on scheduling the execution of big data applications.

Data quality control and annotation were modeled as basic curation services in ENVRI-RM, although they have different (but related) requirements. **Self-adaptable data curation for system-level science** covers different levels of data. The service provided by ENVRI^{PLUS} will comply with data and metadata standards such as OASIS [35] and INSPIRE [36] and will provide rich, interoperable metadata for geospatial semantic annotation. The quality of user experience, when checking the quality of data and when annotating different data using the aforementioned metadata standards, will be explicitly modeled and taken into account in the development of curation services.

To perform complex data-driven experiments, scientists want simple but effective mechanisms to discover data recorded in catalogues, and to integrate data into computing processes. An **interoperable data cataloguing** service provides interoperable solutions for accessing, retrieving and integrating data from different categories. The service will extend the open search tools developed in the ENVRI project by reusing the latest technologies. It will investigate key issues in interoperable cataloguing and metadata harmonization with consideration of other ongoing initiatives.

Higher-level data products provided by RIs have to be clearly reproducible. Therefore, provenance services that record the evolution of data by tracking each operation processed have to be further developed, and integrated within existing RIs. A **cross-RI data provenance** service provides tracing services for data manipulation between different infrastructures. Standardized interfaces for querying, accessing and integrating provenance data will be realized, building on current standardization efforts such as W3C-PROV [37].

B. Interoperability-aware design and development

The development of common services will be guided by the ENVRI^{PLUS} Reference Model, and be validated with use-cases from various RIs. They will benefit from services provided by underlying e-infrastructures or other similar RI projects. The architecture design for RIs and their common operations will depend critically on a metadata catalogue describing the characteristics (data, users, software services, computing resources, etc.) of existing and planned RIs.

C. Service validation and operation

To enable the final usage of developed common services, the results will be tested and deployed in RIs, possibly via computing and data infrastructures such as EGI and EUDAT.

A complete use-case will be defined and its full implementation will be analyzed and validated. The use-case will be selected to preferably involve several research infrastructures, have the interest and participation of active scientific communities, and have a clear impact. An example of such a use-case could be the study of the mechanisms of carbon sequestration in the biosphere, but will finally be selected in the context of agreement among RIs. This use-case analysis will be described in accordance with the ENVRI^{PLUS} Reference Model. Well-defined success criteria for this use-case will be assigned and tracked alongside validation. The resources available or required on e-infrastructures will be identified, from network connectivity to data storage and processing capabilities.

It must be remarked that among the set of specific services that will be developed and integrated, many of them pose a challenge due to the potential resources needed, the level of integration or the complexity of deployment and operation, and will require closer tracking. For example, the setup of earth observation systems requires data from the RIs as well as data obtained from Copernicus satellite services.

V. SUMMARY

Conducting system-level environmental science research requires advanced e-infrastructures for collecting, curating and providing access to scientific data products. Various environmental research infrastructures (RIs) are being constructed to address this requirement, however there is no coherent standard approach to constructing interoperable RIs that would permit the kind of interdisciplinary research needed to fully exploit the data now being made available.

In this paper, we briefly introduced the basic approach of the *Data for Science* theme in the newly funded EU Horizon 2020 project ENVRI^{PLUS}. ENVRI^{PLUS} started officially in May 2015, and will last four years; the project will provide for environmental RIs standard implementations for various commonly required services based on a generic reference model. During the project, we will have to continuously monitor developments and specifically draw upon the lessons we learned from the previous ENVRI project.

ACKNOWLEDGMENT

This work was supported by the European Union's FP7 funded Integrated project under grant agreement No. 283465 (ENVRI project), and Horizon 2020 research and innovation program under grant agreements No. 654182 (ENVRI^{PLUS} project) and 643963 (SWITCH project). We thank all ENVRI^{PLUS} consortium members for their contributions.

REFERENCES

[1] LifeWatch, an e-science Infrastructure for biodiversity and ecosystem research, http://www.lifewatch.eu/.

[2] EPOS, a European Research Infrastructure on earthquakes, volcanoes, surface dynamics and tectonics, http://www.epos-eu.org/.

[3] EURO-Argo, the European contribution to Argo, which is a global ocean observing system, http://www.euro-argo.eu/.

[4] EMSO, a European network of seafloor observatories for the long-term monitoring of environmental processes related to ecosystems, climate change and geo-hazards, http://www.emso-eu.org/.

[5] ICOS, a European distributed infrastructure dedicated to the monitoring of greenhouse gases (GHG) through its atmospheric, ecosystem and ocean, http://www.icos-infrastructure.eu/.

[6] Copernicus (formerly GMES), http://copernicus.eu/

[7] Global Earth Observation System of Systems (GEOSS), http://www.earthobservations.org/geoss.php

[8] I. Foster, and C. Kesselman. Scaling system-level science: Scientific exploration and IT implications. *Computer*, (11), 31-39, 2006.

[9] EU FP7 ENVRI project, www.envri.eu

[10] European Grid Initiative, egi.eu, www.egi.eu

[11] Helix Nebula, http://www.helix-nebula.eu/

[12] European Strategic Forum of Research Infrastructure (ESFRI) Roadmap, http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri.

[13] Y. Charalabidis, M. Janssen, O. Glassey: Introduction to cloud infrastructures and interoperability minitrack. Hawaii International Conference on System Sciences, 0:2177 (2012)

[14] L. Ngan, Y. Feng, S. Rho, R. Kanagasabai: Enabling interoperability across heterogeneous semantic web services with owl-s based mediation. Asia-Pacific Conference on Services Computing. IEEE (2011)

[15] G. Blair, P. Grace: Emergent middleware: Tackling the interoperability problem. IEEE Internet Computing, 16:78–82 (2012)

[16] Z. Zhao, S. Booms, A. Belloum, C. de Laat, B. Hertzberger, VLE-WFBus: a scientific workflow bus for multi e-Science domains Proceedings of the 2nd IEEE International conference on e-Science and Grid computing, IEEE Computer Society Press, 2006, 11-19

[17] M. Riedel, E. Laure, T. Soddemann, et al.: Interoperation of world-wide production e-science infrastructures. Concurr. Comput. : Pract. Exper., 21(8):961–990 (2009)

[18] L. White, N. Wilde, T. Reichherzer, M. El-Sheikh, G. Goehring, A. Baskin, B. Hartmann, M. Manea: Understanding interoperable systems: Challenges for the maintenance of soa applications. Hawaii International Conference on System Sciences, 0:2199–2206 (2012)

[19] Z. Zhao, P. Grosso, & C. de Laat, OEIRM: An Open Distributed Processing based Interoperability Reference Model for e-Science Cloud&Grid interoperability workshop, Gwangju, Korean., 2012

[20] European Data Infrastructure (EUDAT) http://www.eudat.eu/

[21] L. Candela, D. Castelli, A. Manzi, P. Pagano, Realising Virtual Research Environments by Hybrid Data Infrastructures: the D4Science Experience. International Symposium on Grids and Clouds (ISGC) 2014, Proceedings of Science PoS(ISGC2014)

[22] K. Jeffery, N. Houssos, B. Jörg, A. Asserson, Research information management: the CERIF approach, International Journal of Metadata, Semantics and Ontologies , Volume 9 Issue 1, February 2014

[23] Data Catalog Vocabulary (DCAT) http://www.w3.org/TR/vocab-dcat/

[24] Research Data Alliance (RDA) https://rd-alliance.org/

[25] EISCAT-3D, a European new-generation incoherent-scatter research radar for upper atmospheric science, https://www.eiscat3d.se/.

[26] P. F. Linington, Z. Milosevic, A. & Vallecillo, A. Building enterprise systems with ODP CRC press, Taylor & Francis group, 2012

[27] ENVRI Reference Model: www.envri.eu/rm

[28] SeaDataNet, Pan European Infrastructure for Ocean and Marine Data Management: http://www.seadatanet.org/

[29] JERICO, European research infrastructure network for coastal observations: http://www.jerico-fp7.eu/

[30] European Marine Biology Resource Centre (EMBRC): http://www.embrc.eu/

[31] European Marine Observation and Data network, Biology data discovery and access service, http://www.emodnet-biology.eu/

[32] DataCite: https://www.datacite.org/

[33] Trusted data services for global science: https://www.icsu-wds.org/

[34] Z. Zhao, P. Grosso, J. van der Ham, R. Koning, and C. de Laat. An agent based network resource planner for workflow applications. Multiagent Grid Syst., 7(6):187–202, November 2011

[35] Advancing Open Standards for the Information Society, https://www.oasis-open.org/

[36] INSPIRE infrastructures for spatial information, http://inspire.ec.europa.eu

[37] P. Groth, and Luc Moreau. PROV-Overview, W3C Working Group Note. http://www.w3.org/TR/prov-overview/, 2013.