Knowledge-as-a-Service: A community knowledge base for research infrastructures in environmental and earth sciences

Zhiming Zhao*, Xiaofeng Liao,
Paul Martin
University of Amsterdam, the Netherlands
*Z.Zhao@uva.nl

Jordan Maduro, Peter Thijsse,
Dick Schaap
MARIS B.V.
Voorburg, the Netherlands

Markus Stocker,
TIB Leibniz Information Centre for Science and Technology, Germany

Doron Goldfarb, Barbara Magagna
Environmental Agency Austria

***Abstract*— The ENVRI Reference Model (ENVRI RM) and its ontological representation Open Information Linking for Environmental RIs (OIL-E) allow architects and engineers to describe the architecture and operational behavior of environmental and earth science research infrastructures (RIs) in a standardized way using community-agreed terminology. RI descriptions can be published as linked data, allowing discovery, querying, and comparison using established Semantic Web technologies. The ENVRI Knowledge Base is a community knowledge base which uses OIL-E to capture information about environmental and earth science RIs in the ENVRI community for query and comparison. Such Knowledge-as-a-Service supports identifying the technologies and standards used for particular activities and services and evaluating research infrastructure subsystems and behaviors against certain criteria, such as compliance with the FAIR data principles.**

Keywords*— *Research infrastructure, knowledge base, FAIRness, big data management

I. Introduction

Data-centric approaches play an increasing role in many scientific domains such as in the environmental and earth sciences. Such research activities also often require advanced computing and storage infrastructure in order to analyze, process, model, and simulate data. Advanced *research support environments* are needed to enable researchers to access data, software tools and services from different sources, and to integrate them into cohesive experimental investigations with well-defined, replicable workflows for processing data and tracking the provenance of results. Based on the types of functionality needed, we identified three key research support environments [1]: 1) *e-Infrastructures* for providing computing, storage and network resources; 2) *research infrastructures* (RI) for research assets and services within different scientific domains; and 3) *virtual research environments* for providing user-centered support for discovering and selecting data and software services from different sources.

Among research support environments, Research Infrastructures play a key role in the *lifecycle of research data, services, and other assets*, providing security and access policies for e.g., the acquisition, curation, publication, processing and other usages of research data. RIs in environmental and earth sciences support user communities with 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] (biodiversity), EPOS [2] (solid earth science), EURO-ARGO[3] (ocean monitoring), eLTER[4] (Ecosystem) , as well as ICOS[5] and EISCAT_3D[6] (atmosphere).

These infrastructures are developing into important pillars for their respective user communities, but are also intended to support interdisciplinary research as well as Copernicus[7] as a contribution to GEOSS[8]. As such, it is very important that data-related activities are well integrated in order to enable data-driven system-level science. This requires standard policies, models, and e-Infrastructures to ensure coordination, harmonization, integration, and interoperability of data, applications, and other services. However, the complex nature of environmental science seems to result in the development of environmental RIs that meet only the requirements and needs of their own domains, with very limited interoperability of data and isolated tools and operational policies. The root cause for the heterogeneity is that the RI development communities often lack effective sharing of technical practices about architectural design, service interfaces, selections of metadata standards, controlled vocabularies, and ontologies.

In this paper, we present a community knowledge base proposed in the EU H2020 ENVRI^{PLUS} [9] project that will be further developed in its follow-up project ENVRI-FAIR[10]. The knowledge base provides Knowledge-as-a-Service for the RI development communities to document the development and operation of RI services and to address engineering problems.

The paper is organized as follows. First, we introduce the research background and related work of the proposed solution. Next, we discuss the requirements for the ENVRI knowledge base and its architecture. Finally, we discuss the current prototype and demonstrate its use with two use cases.

[1] LifeWatch: www.lifewatch.eu
[2] EPOS: www.epos-ip.org
[3] EuroArgo: www.euro-argo.eu
[4] eLTER:www.lter-europe.net/elter
[5] ICOS: www.icos-ri.eu
[6] EISCAT-3D: www.eiscat.se
[7] COPERNICUS: www.copernicus.eu
[8] GEOSS: www.geoportal.org
[9] EU H2020 ENVRI^{PLUS}: envriplus.eu
[10] EU H2020 ENVRI-FAIR: envri.eu/envri-fair/

978-1-7281-3851-0/19/$31.00 ©2019 IEEE
DOI 10.1109/SERVICES.2019.00041

II. BACKGROUND

The ENVRI cluster projects, namely ENVRI (2011-2014), ENVRI^{PLUS} (2015-2019) and ENVRI-FAIR (2019-2022), are concerned with: 1) the identification of technical and organizational commonalities between environmental RIs; 2) prototyping reusable solutions to common challenges; and 3) improving the FAIRness of assets in different RIs and making them interoperable. One of the key challenges we have to tackle when sharing engineering practices and building interoperable data services is how to effectively communicate with different stakeholders, users, developers involved in RIs of different environmental sub-domains.

III. RELATED WORK

Because a common ontological framework is essential, the ENVRI Reference model (RM) was proposed and developed in the ENVRI community (cluster projects) since 2011 [2].
The methodology for developing ENVRI-RM was to decompose system descriptions based on viewpoints. Open Distributed Processing (ODP) [3] provides five viewpoints from which to describe systems: *enterprise,* i.e., system scenarios, involved communities and roles; *computation,* i.e., system interfaces and bindings between system components; *information*, i.e., data objects and schemas of the system; *engineering*, i.e., system middleware, engineering principles; and *technology*, i.e., technology standards and decisions.

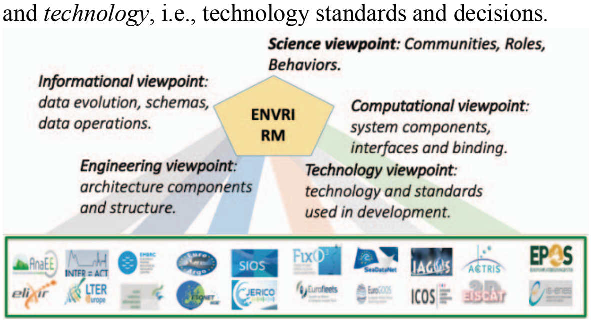


Fig. 1. The basic idea of the ENVRI Reference Model.

This decomposition of complex systems by viewpoints is a useful technique for managing the complexity and provide information tailored to different kinds of stakeholders [4,5,6]. ENVRI RM employs these viewpoints to model the characteristics of environmental research infrastructures but it replaces the *enterprise* viewpoint with the *science* viewpoint. This is to align the ODP with the RI view of the world. The current version is available online[11]. Fig. 1 depicts the basic viewpoints of the ENVRI RM.

Applications of ODP [3] have been studied extensively and ODP has been applied to the design of various kinds of infrastructure, including in the Internet of Things (IoT) [7] and Smart Cities [8]. The applicability of ODP, a standard that was developed in the 1990s, to modern concepts of service-oriented architecture and Cloud have been discussed before in research literature [9]. Indeed, the advancement and wide-scale adoption of virtualization and programmable infrastructure mean that the separation of concerns between the computational and engineering viewpoints (for example) are less clear than they perhaps were original. For instance, modelling a system deployed on virtual infrastructure and modelling the virtual infrastructure service itself would each result in a very different assignment of concepts between the two views. On the other hand, ODP supports the notion of transparencies, the selection of aspects of system design (such as authentication and migration of components) to not be explicitly modelled in specifications to reduce confusion, clutter or repetition in design documents. In this light, the explicit acknowledgment that the resources and channels described in the engineering view of an RI specification happen to be virtualized becomes simply another transparency option. Regardless of whether ODP can be considered to be a sufficiently contemporary specification for the modelling of modern distributed systems, the notion of specifying systems across multiple views is still well-regarded in software engineering research literature.
The ENVRI semantic linking framework was developed based on ENVRI RM. Open Information Linking for Environmental RIs (OIL-E) [10] was designed to provide an upper ontology for RI descriptions based on ENVRI RM that can be used to contextualize different kinds of RI assets from architectural or operational perspectives. This is in contrast to general-purpose ontologies for describing scientific phenomena such as ENVO [11]. OIL-E has more in common with conceptual models that focus on the products and tools of research rather than on scientific classification itself and is more concerned with providing a controlled vocabulary for environmental science RIs, in particular.
The foundation of OIL-E is the **oil-base** ontology, which provides a set of abstract concept classes derived from the most common elements observed in the ENVRI RM and distributed across the five ODP views. Defined from the ENVRI RM specification, the **envri-rm** ontology is the primary extension of **oil-base**. As ENVRI RM is an on-going development, with each release of the model, the **envri-rm** ontology must be updated accordingly. Currently, this is done via consultation within the relevant working group in the ENVRI community, based on demand for new stereotypes for RI entities or activities or discussion regarding the correctness of specific properties or other relationships.

IV. ENVRI KNOWLEDGE BASE

The ENVRI knowledge base aims to provide a repository for RI architectural information and 'design wisdom' encoded using ENVRI RM that can be programmatically queried and analyzed. It serves as a database of information about technologies and standards used by RIs.

A. Target users and roles

The knowledge base serves different types of users in the ENVRI community based on their specific needs:

1. **RI architects,** designers of technical services for RIs, want to know what the best practices are for dealing with various problems common in the construction and management of RIs as well as the prevailing technologies and standards used by others that could be adapted to meet their own needs.
2. **Scientific application developers** of data science applications need to know what resources are available

[11] http://envri.eu/rm

to them, which RIs publishes them, any limitations on their use and the differences between them.

3. Similar to developers of data science applications in general, **Virtual Research Environment developers** need to know about the resources that their VREs might interact with, especially catalogue services that allow users of their VREs to search RI data collections but are particularly concerned about providing direct access to resources within their VREs, to facilitate on-demand programmatic discovery and integration of new resources.
4. **e-Infrastructure service developers** want to be able to tailor their offerings to the needs of researchers. To do this they need to know how their services are already used by RIs and where there is potential for further delegation of RI architecture to e-infrastructure.
5. **Teachers and students** in the domains of ICT, data science or environmental science, who are interested in reviewing how RIs are constructed, the variety of resources that they currently make available, and the potential of science conducted using RIs.
6. **Proposal writers and funding agencies** need to be able to understand the RI landscape and the resources that are currently available as well as gaps that new development may fill. Furthermore, funding agencies need to understand the RI landscape and profile RIs in order to inform their strategic planning.

B. Functional requirements

Based on the types of users and their specific needs, we can identify at least the following functional requirements:

1. **Flexible knowledge base query**, allowing for the development of arbitrarily complex queries, either directly by expert users or via other knowledge base interfaces (such as faceted search or model browsing, see below).
2. **Browsing** RI design models, including the ability to display all data associated with a model or its components in a human-readable manner on a client (e.g., a Web browser) and the ability to directly retrieve information by identifier.
3. **(Semi-)automated composition** of requests to an online RI resource using information about the resource, its API, and the type of request as provided by the user.
4. **Context-specific visualization** of RI design models and other common classes of knowledge base content, typically for embedding in a model browser but also for export (e.g., as an image to download).
5. **Model profiling** of RI design models against specific criteria, such as compliance with **FAIR (Findability, Accessibility, Interoperability and Reusability)** data principles [12], compliance with ENVRI RM templates, or satisfaction of user-provided constraints.
6. **Automated verification of updates** of new or revised data into the knowledge base, limited to restricted APIs subject to model validation and access permissions.
7. **Design space support** for alternative RI designs, including of the same RI, allowing RI architects to experiment, to differentiate between 'current' and 'planned' RI states, and to maintain different versions of the same RI design that can be retrieved separately by search and browsing facilities.
8. **Automated harmonization** of models within the same RI design space with overlapping concepts/concerns. Specifically, the generation of requirements based on correspondence rules between model concepts that are then checked against existing models in the knowledge base.
9. **Comparison** of two or more models in a manner that facilitates easy comprehension of similarities and differences.
10. **Gap Analysis** in an RI design based on a set of models, reference models (e.g., derived from ENVRI RM) and equivalent models of other RI designs, resulting in a set of suggestions as to what might be missing from the RI.
11. **Recommendation:** Generation of 'recommendations' by the knowledge base in response to a set of requirements for RI design, using information from reference models and existing RI designs. Requires the ability to evaluate models and components.

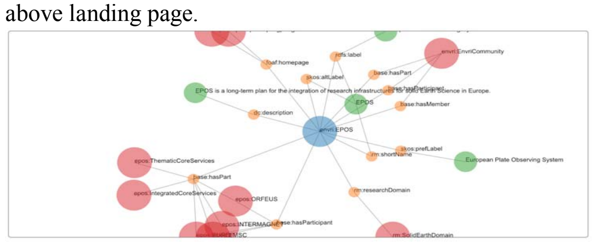
These requirements provide valuable input for knowledge base development.

In the ENVRI^{PLUS} and ENVRI-FAIR projects, those requirements are prioritized based on needs from the RI communities and their RI development and operation lifecycles. For instance, when an RI just starts its development, querying knowledge base, discovering existing solutions from other RIs, and analyzing the gaps will be the typical needs for the knowledge base.

C. Current prototype

The current knowledge base is prototyped using Apache Jena Fuseki, which provides a triple store for aggregated RDF data along with a service API and internal reasoning capabilities based on the OWL [13] standard. The knowledge base contains the complete set of OIL-E ontologies along with a representative sample of RI-specific data for the purposes of demonstration and experimentation. Access to the knowledge base is achieved via a SPARQL [14] endpoint. The main landing page for the knowledge base, which also provides a means to try and modify various sample queries via Web browser without needing a HTTP/SPARQL request client, can be accessed via the ENVRI community site[12].

Figure 3 shows a visualization of information in the current ENVRI knowledge base as can be viewed by visiting the above landing page.



property	data
http://www.w3.org/2000/01/rdf-schema#label	ORFEUS
http://purl.org/dc/elements/1.1/description	ORFEUS is the non-profit foundation to coordinate and promote digital, broadband seismology in the European-Mediterranean area.
http://xmlns.com/foaf/0.1/homepage	http://www.orfeus-eu.org/

Fig. 3. Visualizing information in the ENVRI knowledge base, showing entities associated with the EPOS RI.

[12] https://www.envri.eu/knowledgebase

Nodes are color-coded to distinguish concept classes from instance data and data properties, with additional information accessible by directly selecting individual nodes.

V. USE CASES

When resolving queries, the knowledge base is able to apply the relations and classifications defined by OIL-E in order to infer results beyond those explicitly asserted in the triple store. This allows the full set of ENVRI RM archetypes to be used to guide the discovery and search over all the RI data provided. For the ENVRI knowledge base, we identified four key knowledge capabilities:

1. **A survey of the technical landscape**. The web of knowledge created by semantic linking should help us understand what technologies (including software, standards and vocabularies) are being used by environmental science RIs.
2. **Comparative solution analysis**. It should be possible to compare solutions developed by environmental science RIs. Specifically, given the knowledge of how technologies are used in their proper context, we should be able to compare developments in equivalent contexts.
3. **Gap analysis and component recommendation**. Given a reference model for environmental science RIs (e.g., ENVRI RM), it should be possible to identify what is missing in the current development state of a given RI, and based on both the model and the solutions developed by other RIs, it should be possible to then make certain recommendations.
4. **Linked open research infrastructure**. The web of knowledge created by semantic linking should itself be publicly accessible, machine-navigable, and provide a gateway to the services and data held by the RIs. It should include (where available) data provenance and resource catalogues, and it should (where appropriate) make use of other ENVRI services such as the catalogue service for cross-RI search.

Given a sufficiently-detailed corpus of information regarding environmental and earth science RIs backed by a 'standard model' for how such RIs are constructed (i.e. ENVRI RM), it is possible to evaluate individual RIs or RI subsystems in terms of how they compare with similar RIs or against some kind of base criteria. In this section, we shall describe two use cases to demonstrate how the knowledge base can be used in the real ENVRI scenarios.

A. *Case 1: Sharing data quality control technologies among RIs*

Data quality control is an important activity when curating data. In the ENVRI community, raw data collected from sensors or observation stations has to be quality controlled before being published via catalogues. Using the knowledge base, it is possible to compare data quality control processes applied by RIs in the same domain (e.g. marine science) against another or against a specifically prescribed methodology. In this case, we use ENVRI RM to model the quality control procedure of different RI in a shared knowledge base.

First, we develop a detailed taxonomy for describing the Quality Control related methodologies, roles and processes, as shown in Fig. 4. The taxonomy is developed based on the data management lifecycle defined by ENVRI RM.

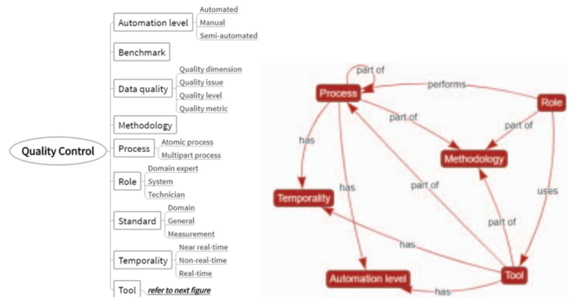


Fig. 4. Quality control concepts based on the information viewpoint in the ENVRI RM.

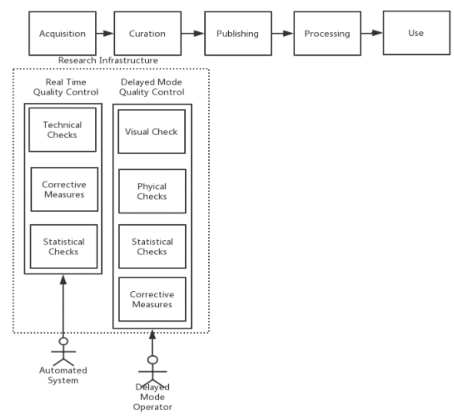


Fig. 5. Quality control processes survey of the EuroArgo research infrastructure.

Second, we use the taxonomy and ENVRI RM to model the quality control procedures for each RI (Fig. 5 displays an example) and to describe them in the ENVRI knowledge base.

Finally, the quality control tools are reviewed based on the taxonomy (as shown in Table 1). Such information is valuable for the community to share.

		Temporal			License		Interface			Architecture				Source		Task		
Tool	Real-time	Near real-time	Manual	Free	Limited	Commercial	Graphical user interface	Command line	Embedded	Stand alone	Web Service	Library	Implementation / Algorithm	Open Source	Proprietary	Testing	Visualization	Manipulation
Ocean Data View			x		x		x			x					x		x	
ODM Tools			x	x			x			x				x			x	
AutoQC		x		x				x		x				x		x		
CoTeDe		x		x				x		x				x		x		
Java Ocean Atlas			x		x		x			x					x		x	
SeaBird			x			x	x			x					x	x		
QC Tool Hydroffice			x	x			x			x				x			x	
Spike		x		x					x				x	x		x		
Global Range	x			x					x				x	x		x		
PyNGL			x	x				x				x		x			x	
HDFView			x	x			x			x					x		x	
Pan2Applic			x	x				x		x				x				x
Panoply			x	x			x			x					x		x	
Integrated Data Viewer			x	x			x			x					x		x	
Saga			x	x			x			x				x			x	
ToolsUI			x	x			x			x					x			x
ncBrowse			x	x			x			x					x		x	
GRASS			x	x			x			x				x			x	
Single Calculus Chain	x			x				x			x				x	x		
QuinCe QC system			x	x			x				x				x		x	

Table 1. A review of QC tools in ocean observation.

In this way, the quality control procedure and tools of a specific RI can be discovered and shared with other RI communities, as suggested in Figure 6.

```
CONSTRUCT {
  iagos:PIQualityControl ?property ?object .
} WHERE {
  iagos:PIQualityControl ?property ?object .
  FILTER( !isBlank(?object) ) .
}
```

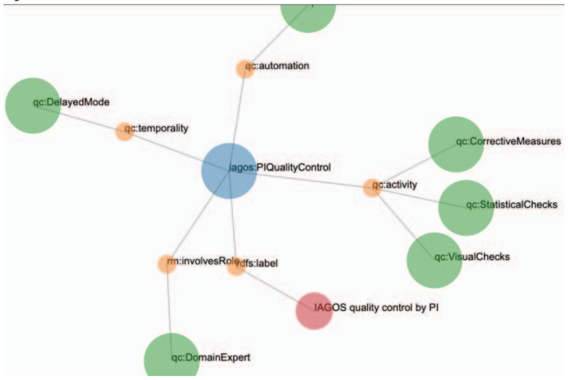


Fig. 6. The ENVRI knowledge base for QC related query and output.

B. Case 2: analyze the FAIRness of an RI

Another possibility is to provide tools for RI designers to evaluate their own RIs in terms of compliance to the FAIR data principles[13]. Wilkinson et al. [12] proposed guidelines for Findable, Accessible, Interoperable and Reproducible data that would better support open data science.

There are a number of possible approaches that could be taken to evaluate some of these guidelines based on the content of an ENVRI knowledge base:

1. **Findability**. Which published data products include globally unique persistent identifiers? Are those identifiers included in the product metadata? What other core metadata does each published product include (or not include)? Does the RI provide an index or registry for search and discovery of data products? Does it contribute to any external registries?
2. **Accessibility**. Can data product metadata be retrieved by a standard, open and free communication protocol, and if so, which one? Does the RI define an authentication and authorization process for accessing data, and does it use standard, open mechanisms? Are metadata accessible via some means even if the data product described is no longer available?
3. **Interoperability**. What data formats, metadata schemes and controlled vocabularies are used to describe/represent (meta)data in the RI? Do those terminological resources comply themselves with the FAIR principles?
4. **Reusability**. How rich are the metadata provided for data products? Under what licenses can data be used? Is detailed provenance included in the metadata, and does the RI include provenance tracking in its internal processes? Do RI (meta)data meet domain-specific community standards?

Notably, such evaluation does not rely solely on the specification of data products (information view), but also on information about the services provided or delegated by an RI (computational view), the technologies used (technology view) and the general processes defined (science view). Thus the holistic multi-view specifications permitted by OIL-E using ENVRI RM stereotypes potentially allows for a much more sophisticated analysis of RI status that would be provided by (for example) a catalog of metadata schemes used by RIs for their primary data products.

The detailed FAIRness assessment and analysis will be presented in a separated paper.

VI. Discussion

The knowledge base and OIL-E are both the basis for more tools with which to support several useful functions.

A. Semantic web technologies and knowledge base

The Semantic Web relies on a number of foundational technologies for representing and associating semantics to information, including RDF [15], OWL [13] and SKOS [16], along with standards for interacting with semantic information (e.g., for search and discovery) such as SPARQL. Considerable attention has been given to the openness, extensibility, and computability of such standards, with different options for controlled vocabulary specification. While RI designs could be specified using something other than Semantic Web technologies (for example based on traditional relational database models), the openness and extensibility of the Semantic Web fit well with the heterogeneity of RI designs and the varying levels of detail specific aspects of RI design may or may not be modelled. It should also be noted that RI models are not themselves particularly large in terms of data volume. Indeed, they consist of relatively high-level and highly structured semantic information. This aspect also fits the Semantic Web knowledge graph meta-model.

B. Advantages of a community knowledge base

We can envisage a number of avenues of further development (or in most cases, alignment with existing developments for mutual benefit). These include:

Cross-RI search and discovery. OIL-E provides a standard taxonomy for various entities and activities related to RIs, which can be used to classify different kinds of resources as part of a faceted search pipeline. An OIL-E knowledge base can act directly as a catalogue service for multiple RIs. However, this is not necessarily the best possible approach as OIL-E is optimized for describing RI design and contextualizing RIs' component parts, rather than providing a more traditional metadata scheme for describing RI resources.

Faster RI specification using ENVRI RM. Detailed descriptions of RIs in terms of their architecture, core data products and processes allows for more in-depth investigations and comparisons of RI solutions to various technical problems. ENVRI RM provides the basis for such descriptions but requires a specialist's expertise to be used effectively. It has previously been used manually, which resulted in the creation of a body of documentation for each modelled RI.

Requirements recommendation. Using tools such as OIL-E and the ENVRI knowledge base, it is possible to do a comparative analysis of the solutions provided by RIs in terms of technology and processes to address various

[13] FORCE 11: https://www.force11.org/

common problems regarding the handling of research data (among other aspects). Tools that can interact with the knowledge base on behalf of users, constructing and interpreting queries behind a user-friendly interface, could be very useful for taking full advantage of the corpus of knowledge resulting from RI modelling.

Provenance exploration. There are two notable ways in which OIL-E data can interact with provenance data, especially data encoded to the W3C PROV standard [17]: 1) as linking data to various provenance repositories, contextualizing the role of the repositories and providing a reference to where the provenance is and how it can be extracted; and 2) as a validation framework. Given descriptions of RI processes encoded in OIL-E, provenance can be checked against those descriptions by mapping agents, entities, and activities to the correct OIL-E concepts and then checking whether the relationships described in provenance match those prescribed by the process model.

Natural language based document analysis and annotation. A significant corpus of existing information about RIs exists in the form of written documentation produced by RI architects and developers. The ability to apply a framework such as OIL-E to annotate uploaded documents, identifying possible references to concepts defined in ENVRI RM in text, would be useful both to contextualize documents automatically and provide initial descriptions for the RIs and RI components described by the documents. Machine learning tools would thus provide a valuable additional source of data for the knowledge base, or to validate existing models of RIs.

VII. Conclusions and Future Work

In this paper, we described how the ENVRI Reference Model (ENVRI RM) and its ontological representation (OIL-E) are used as the basis for building a community knowledge base for RIs in environmental and earth sciences.

The development of the ENVRI knowledge base is ongoing. The current prototype demonstrates how an information corpus for RIs might be used to analyze and compare RI designs, as well as to document the technologies, software, and standards used by RIs in their operational contexts.

The next major objective of the ENVRI community is to facilitate the adoption of the FAIR principles for research data gathered in the atmospheric, marine, solid earth and biodiversity domains, and to develop sustainable FAIR data services for research communities as part of the broader push towards better open data science and more seamless interoperability between different data providers.

Acknowledgment

This work was supported by the European Union's Horizon 2020 research and innovation program under grant agreements No. 654182 (ENVRI^{PLUS} project), No. 824068 (ENVRI-FAIR project) and No. 825134 (ARTICONF project).

References

[1] Koulouzis, S., Martin, P., Zhou, H., Hu, Y., Wang, J., Carval, T., Grenier, B., Heikkinen, J. de Laat, C., and Zhao, Z.(2019), Time-critical data management in clouds: challenges and a Dynamic Real-time Infrastructure Planner (DRIP) solution, Concurrency Computat Pract Exper [doi:10.1002/cpe.5269]

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

[3] P. F. Linington, Ed., Building enterprise systems with ODP: an introduction to Open Distributed Processing. Boca Raton, Fla.: CRC Press, 2012.

[4] Ken Peffers, Tuure Tuunanen, Marcus Rothenberger, and Samir Chatterjee. 2007. A Design Science Research Methodology for Information Systems Research. J. Manage. Inf. Syst. 24, 3 (December 2007), 45-77. DOI=http://dx.doi.org/10.2753/MIS0742-1222240302

[5] Shirley Gregor and Alan R. Hevner. 2013. Positioning and presenting design science research for maximum impact. MIS Q. 37, 2 (June 2013), 337-356. DOI: https://doi.org/10.25300/MISQ/2013/37.2.01

[6] Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. 2004. Design science in information systems research. MIS Q. 28, 1 (March 2004), 75-105.

[7] Laassiri, Jalal, and Salah-ddine Krit. "Internet of Things—Architecture and concepts in ODP information language." 2016 International Conference on Engineering & MIS (ICEMIS). IEEE, 2016.

[8] Moshrefzadeh, Mandana, and Thomas H. Kolbe. "Smart Data Infrastructure for Smart and Sustainable Cities." 13th International Conference on Design & Decision Support Systems in Architecture and Urban Planning. 2016

[9] Zhao Z., Grosso P., de Laat C. (2012) OEIRM: An Open Distributed Processing Based Interoperability Reference Model for e-Science. In: Park J.J., Zomaya A., Yeo SS., Sahni S. (eds) Network and Parallel Computing. NPC 2012. Lecture Notes in Computer Science, vol 7513. Springer, Berlin, Heidelberg

[10] Martin, P., Grosso, P., Magagna, B., Schentz, H., Chen,Y., Hardisty, A., Los, W., Jeffrey, K., de Laat, C., Zhao, Z. (2015) Open Information Linking for Environmental Research Infrastructures, 1st International workshop on Interoperable infrastructures for interdisciplinary big data research (IT4RIs) In the proceedings of IEEE eScience Munich Germany, p546-553 [doi:10.1109/eScience.2015.66]

[11] Buttigieg, P. L., Morrison, N., Smith, B., Mungall, C. J., & Lewis, S. E. (2013). The environment ontology: contextualising biological and biomedical entities. Journal of Biomedical Semantics, 4(1), 43. doi:10.1186/2041-1480-4-43

[12] Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." Scientific data 3 (2016).

[13] "OWL 2 Web Ontology Language Document Overview (Second Edition)". W3C. 11 December 2012.

[14] Prud'hommeaux, E., Seaborne, A.: SPARQL Query Language for RDF. W3C Recommendation (January 2008), http://www.w3.org/TR/rdf-sparql-query/

[15] http://www.w3.org/TR/PR-rdf-syntax/ "Resource Description Framework (RDF) Model and Syntax Specification"

[16] Baker, T., Bechhofer, S., Isaac, A., Miles, A., Schreiber, G., Summers, E.: Key choices in the design of simple knowledge organization system (SKOS). Web Semantics: Science, Services and Agents on the World Wide Web 20, 35–49 (2013)

[17] PROV-Overview,http://www.w3.org/TR/2013/NOTE-prov-overview-20130430/