OEILM: a semantic linking framework for environmental research infrastructures

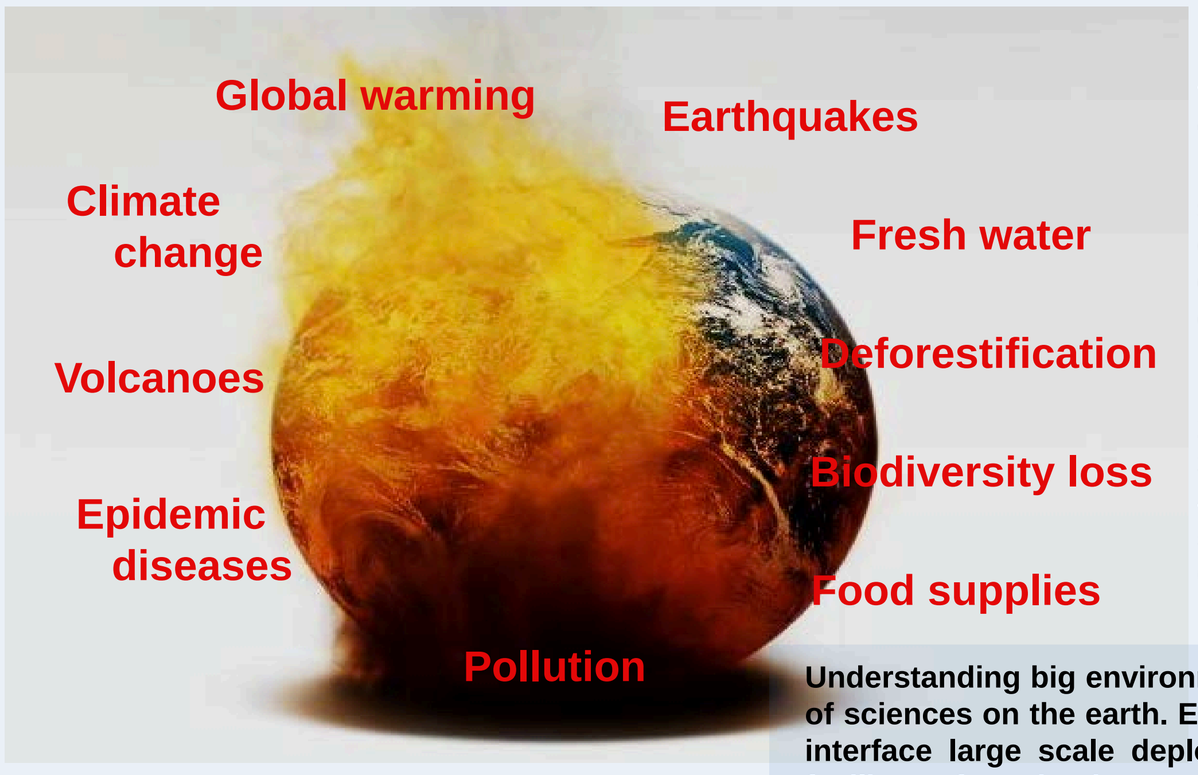
Zhao Z*., Wibisono A., Grosso P., Los W., de Laat C., Chen Y., Hardisty A., Martin P., Atkinson M., Schentz H., Magagna B.

University of Amsterdam
the Netherlands
*** Email Z.Zhao@uva.nl**

Cardiff University
UK

Edinburgh University
UK

Environmental Agency Austria
Austria



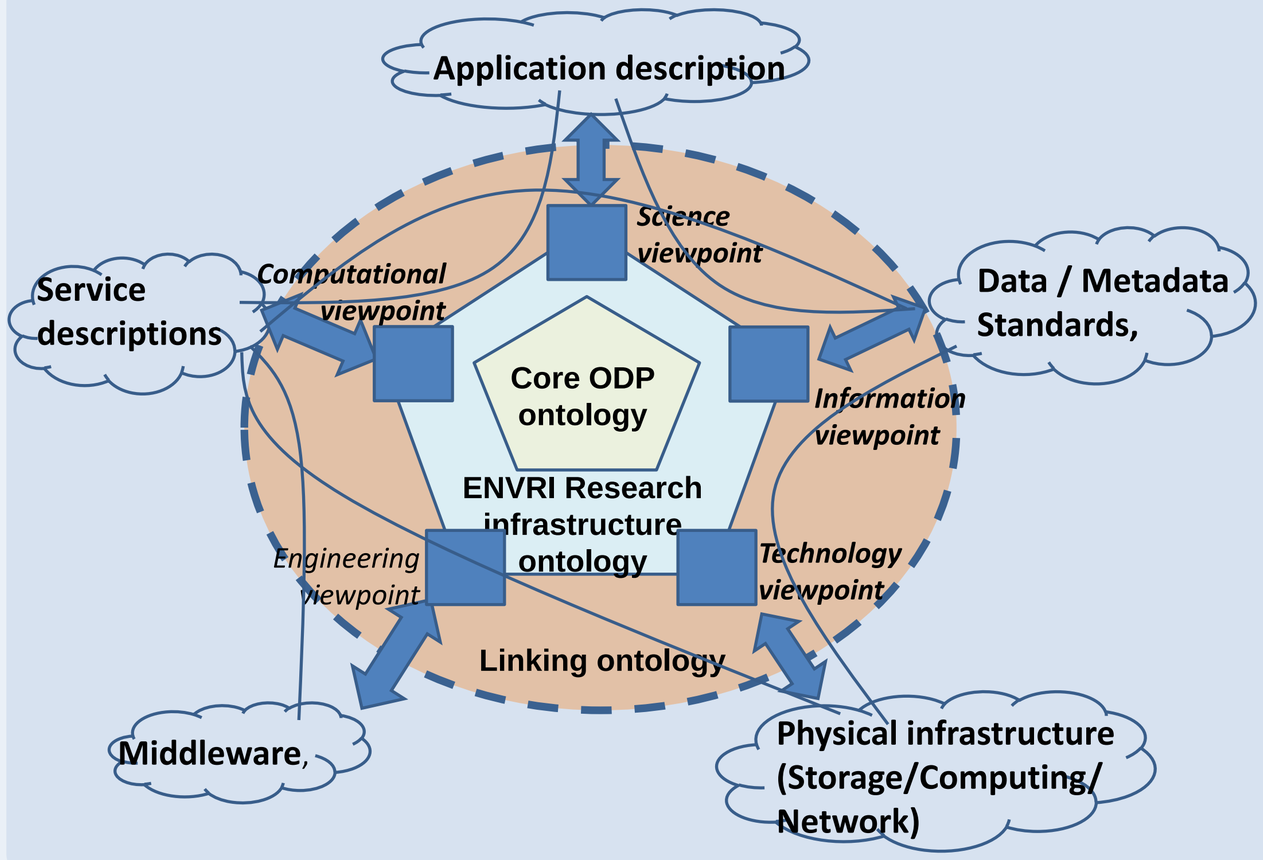
Environmental Research Infrastructure

The EU FP7 project **ENVRI**, namely the "**Common Operations of ENVironmental Research Infrastructures**", emphasizes on synergies between advanced developments, not only among the infrastructure facilities but also data driven experiments require multi-disciplinary sciences.

1. ICOS: http://www.icos-infrastructure.eu/
2. EURO-Argo: http://www.euro-argo.eu/
3. EISCAT-3D: http://www.eiscat3d.se/.
4. LifeWatch: http://www.lifewatch.eu/
5. EPOS: http://www.epos-eu.org/.
6. EMSO: http://www.emso-eu.org/

Understanding big environmental problems requires system level of sciences on the earth. Environmental Research infrastructures interface large scale deployed sensors, and provide advanced facilitates for scientists to perform advanced research.

OEILM: Open E-Science Information Linking Model



The ODP core ontology

The Open Distributed Processing (ODP) modeling approach captures the design and development issues in complex distributed systems from five corresponding viewpoints:

1. The ***enterprise viewpoint*** models organizational issues of the system;
2. The ***informational viewpoint*** models the information objects and schemas;
3. The ***computational viewpoint*** models functional components and binding interfaces;
4. The ***engineering viewpoint*** describes the construction of the system;
5. the ***technological viewpoint*** describes required technologies in the development.

The ENVRI Reference model ontology

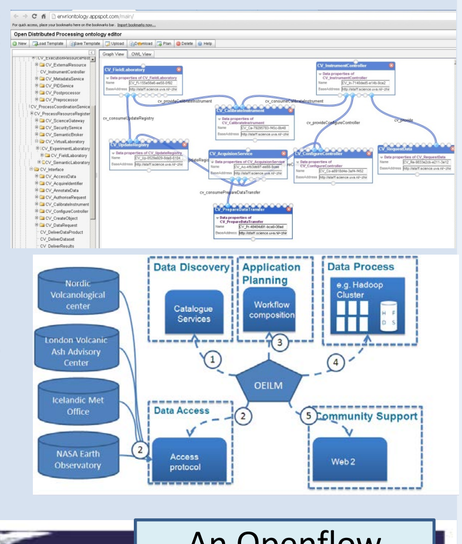
The ENVRI RM ontology imports the ODP ontology and models the functional components of a research infrastructure based broadly on a notion of data life-cycle evident in all existing ESFRI. Five sub systems are distinguished.

1. The ***data acquisition*** subsystem collects raw data from sensor arrays and other instruments, as well as from human observers, and brings those data into the system.
2. The ***data curation*** subsystem stores, manages and ensures access to all persistent data-sets produced within the infrastructure.
3. The ***data access*** subsystem enables discovery and retrieval of data housed in data resources managed by the data curation subsystem.
4. The ***data processing*** subsystem is able to aggregate data from various sources and conduct a range of experiments and analyses upon that data.
5. The ***community support*** subsystem manages, controls and tracks users' activities and supports users to fulfill their roles in their communities.

Use case

One typical example is to investigate behaviours of the Eyjafjallajkull volcano in Iceland using data of different environmental research data infrastructures. The semantic information provided by OEILM will enable the data and resource discovery, and the planning for executing applications on distributed infrastructure. We prototype the data access and delivery part of the use case using a test bed provided by the OpenLab facilitate in University of Amsterdam

The infrastructure descriptions are based on Network Modeling Language, the network control services are described as using CineGrid Description Language, and the application workflow is described using Abstract QoS Workflow schema. Using the linking framework, the data, service, and infrastructure descriptions are aggregated by a workflow planning component called NEWQoSPlanner. The network controllability is achieved using Network Service Interface (NSI) and Openflow.



The Linking ontology

The linking ontology models the relations between the ENVRI RM and the external specific information models such as for data and physical infrastructure.

1. linking to **application schemas** enables developers to discover data and services from research infrastructures to solve the constraints in application descriptions.
2. linking to domain specific **data and metadata standards** enables data discovery agents to do cross infrastructure data discovery.
3. linking to **service description schemas** enables service discovery agents to select software component from research infrastructures.
4. linking to **middleware descriptions** enables runtime resource selection components to select and schedule the execution of an application.
5. linking to **infrastructure descriptions** enable the agents to select low level physical or virutlized services such as network for executing application.

Summary

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

Acknowledgement. We would like to thank the EU FP7 ENVRI project, National e-Science Research Center and the Dutch national and education network SURFnet, through the GigaPort Research on Network (RoN) project, for sponsoring this research.

References.
[1] Chen, Y. et al., Building a common reference model for environmental science research infrastructures. In *Proceedings of EnviroInfo2013*, 2013.
[2] Zhao, Z., et al, (2012). Quality guaranteed media delivery over advanced network, chapter in book Next Generation Content Delivery Infrastructure ISBN 978-1-4666-1794-0
[3] Openlab at university of amsterdam, http://sne.science.uva.nl/openlab/.
[4] OEILM Ontology: staff.science.uva.nl/~zhiming/Ontology
[5] Use case GUI: newqosplanner-demo.appspot.com