

# Introduction

- [Purpose and Scope](#)
- [Rationale](#)
- [Basis](#)
- [Approaches](#)
- [Conformance](#)
- [Related Work](#)
  - [Related Concepts](#)
  - [Related Reference Models](#)
    - [Committee Reference Models](#)
    - [Consensus Reference Models](#)
    - [Consultation Reference Models](#)
  - [Other Related Standards](#)

## Purpose and Scope

All research infrastructures for environmental sciences (the so-called 'ENVRI') although very diverse, have some common characteristics, enabling them potentially to achieve a greater level of interoperability through the use of common standards and approaches for various functions. The objective of the ENVRI Reference Model is to develop a common framework and specification for the description and characterisation of computational and storage infrastructures. This framework can support the ENVRI to achieve seamless interoperability between the heterogeneous resources of their different infrastructures.

The ENVRI Reference Model serves the following purposes [1]:

- to provide a way for structuring thinking that helps the community to reach a common vision;
- to provide a common language that can be used to communicate concepts concisely;
- to help discover existing solutions to common problems;
- to provide a framework into which different functional components of research infrastructures can be placed, in order to draw comparisons and identify missing functionality.

The present wiki / document describes the ENVRI Reference Model which:

- captures computational characteristics of data and operations that are common in ENVRI Research Infrastructures; and
- establishes a taxonomy of terms, concepts and definitions to be used by the ENVRI community.

The Reference Model provides an abstract logical conceptual model. It does not impose a specific architecture. Nor does it impose specific design decisions or constraints on the design of an infrastructure.

The *initial* model (versions 1.0 and 1.1) focused on the urgent and important issues prioritised for ENV research infrastructures including data preservation, data discovery and access, and data publication. It defines a minimal set of functionalities to support these requirements. The *initial* model does not cover engineering mechanisms or the applicability of existing standards or technologies.

Version 2.x of the model incrementally extends these core functionalities:

- Version 2.0 is a simplification of the way the Reference Model is presented, to make it easier to understand and become familiar with. Version 2.0 explicitly aligns the RM with a lifecycle oriented view of research data management.

## Rationale

Environmental issues will dominate the 21<sup>st</sup> century [2]. Research infrastructures that provide advanced capabilities for data sharing, processing and analysis enable excellent research and play an ever-increasing role in the environmental sciences as well as in solving societal challenges. The [ENVRIplus project](#) and its predecessor ENVRI project gathers many of the EU ESFRI and other environmental infrastructures ([ICOS](#), [EURO-Argo](#), [EISCA T-3D](#), [LifeWatch](#), [EPOS](#), [EMSO](#), etc.) to find common solutions to common problems, including use of common software solutions. The results, including the ENVRI Reference Model will accelerate the construction of these infrastructures and improve interoperability among them. The experiences gained will also benefit building of other advanced research infrastructures.

The primary objective of ENVRI is to agree on a reference model for joint operations. This will enable greater understanding and cooperation between infrastructures since fundamentally the model will serve to provide a universal reference framework for discussing many common technical challenges facing all of the ESFRI-ENV infrastructures. By drawing analogies between the reference components of the model and the actual elements of the infrastructures (or their proposed designs) as they exist now, various gaps and points of overlap can be identified [3].

The ENVRI Reference Model is based on the design experiences of the state-of-the-art environmental research infrastructures, with a view of informing future implementation. It tackles multiple challenging issues encountered by existing initiatives, such as data streaming and storage management; data discovery and access to distributed data archives; linked computational, network and storage infrastructure; data curation, data integration, harmonisation and publication; data mining and visualisation, and scientific workflow management and execution. It uses Open Distributed Processing (ODP), a standard framework for distributed system specification, to describe the model.

To our best knowledge there is no existing reference model for environmental science research infrastructures. This work intends to make a first attempt, which can serve as a basis to inspire future research explorations.

There is an urgent need to create such a model, as we are at the beginning of a new era. The advances in automation, communication, sensing and computation enable experimental scientific processes to generate data and digital objects at unprecedentedly great speeds and volumes. Many infrastructures are starting to be built to exploit the growing wealth of scientific data and enable multi-disciplinary knowledge sharing. In the case of ENVRI, most investigated RIs are in their planning / construction phase. The high cost attached to the construction of environmental infrastructures require cooperation on the sharing of experiences and technologies, solving crucial common e-science issues and challenges together. Only by adopting a good reference model can the community secure interoperability between infrastructures, enable reuse, share resources and experiences, and avoid unnecessary duplication of effort.

The contribution of this work is threefold:

- The model captures the computational requirements and the state-of-the-art design experiences of a collection of representative research infrastructures for environmental sciences. It is the first reference model of this kind which can be used as a basis to inspire future research.
- It provides a common language for communication to unify understanding. It serves as a community standard to secure interoperability.
- It can be used as a base to drive design and implementation. Common services can be provided which can be widely applicable to various environmental research infrastructures and beyond.

## Basis

The ENVRI Reference Model is built on top of the Open Distributed Processing (ODP) framework [4, 5, 6, 7]. ODP is an international standard for architecting open, distributed processing systems. It provides an overall conceptual framework for building distributed systems in an incremental manner.

The reasons for adopting the ODP framework in the ENVRI project come from three aspects:

- It enables large collaborative design activities;
- It provides a framework for specifying and building large or complex system that consists of a set of guiding concepts and terminology. This provides a way of thinking about architectural issues in terms of fundamental patterns or organising principles; and
- Being an international standard, ODP offers authority and stability.

ODP adopts the **object modelling** approach to system specification. ISO/IEC 10746-2 [5] includes the formal definitions of the concepts and terminology adopted from object models, which serves as the foundation for expressing the architecture of ODP systems. The modelling concepts fall into three categories [4, 5]:

- Basic modelling concepts for a general object-based model;
- Specification concepts to allow designers to describe and reason about ODP system specifications;
- Structuring concepts, including organisation, the properties of systems and objects, management, that correspond to notions and structures that are generally applicable in the design and description of distributed systems.

ODP is best known for its use of viewpoints. A *viewpoint* (on a system) is an abstraction that yields a specification of the whole system related to a particular set of concerns. The ODP reference model defines five specific viewpoints as follows [4, 6]:

- The **Enterprise Viewpoint**, which concerns the organisational situation in which business (research activity in the current case) is to take place; For better communication with ENVRI community, in this document, we rename it as **Science Viewpoint**.
- The **Information Viewpoint**, which concerns modelling of the shared information manipulated within the system of interest;
- The **Computational Viewpoint**, which concerns the design of the analytical, modelling and simulation processes and applications provided by the system;
- The **Engineering Viewpoint**, which tackles the problems of diversity in infrastructure provision; it gives the prescriptions for supporting the necessary abstract computational interactions in a range of different concrete situations;
- The **Technology Viewpoint**, which concerns real-world constraints (such as restrictions on the facilities and technologies available to implement the system) applied to the existing computing platforms on which the computational processes must execute.

This version of the ENVRI Reference Model covers 3 ODP viewpoints: the science, information, and computational viewpoints.

## Approaches

The approach leading to the creation of the ENVRI Reference Model is based on the analysis of the requirements of a collection of representative environmental research infrastructures, which are reported in two ENVRI deliverables:

- D3.2: Assessment of the State of the Art
- [D3.3](#)

The ODP standard is used as the modelling and specification framework, which enables the designers from different organisations to work independently and collaboratively. The development starts from a core model and will be incrementally extended based on the community common requirements and interests. The reference model will be evaluated by examining the feasibilities in implementations, and the refinement of the model will be based on community feedback.

## Conformance

A conforming environmental research infrastructure should support the common functionalities described in [Model Overview](#) and the functional and information model described in [The ENVRI Reference Model](#).

The ENVRI Reference Model does not define or require any particular method of implementation of these concepts. It is assumed that implementers will use this reference model as a guide while developing a specific implementation to provide identified services and content. A conforming environmental research infrastructure may provide additional services to users beyond those minimally required functions defined in this document.

Any descriptive (or prescriptive) documents that claim to be conformant to the ENVRI Reference Model should use the terms and concepts defined herein in a similar way.

## Related Work

### Related Concepts

A **reference model** is an abstract framework for understanding significant relationships among the entities of some environment. It consists of a minimal set of unifying concepts, axioms and relationships within a particular problem domain [8].

A reference model is not a reference architecture. A **reference architecture** is an architectural design pattern indicating an abstract solution that implements the concepts and relationships identified in the reference model [8]. Different from a reference architecture, a reference model is independent from specific standards, technologies, implementations or other concrete details. A reference model can drive the development of a reference architecture or more than one of them [9].

It could be argued that a reference model is, at its core, an **ontology**. Conventional reference models, e.g., OSI [10], RM-ODP [4], OAIS[11], are built upon modelling disciplines. Many recent works, such as the DL.org Digital Library Reference Model [9], are more ontology-like.

Both models and ontologies are technologies for information representation, but have been developed separately in different domains [13]. Modelling approaches have risen to prominence in the software engineering domain over the last ten to fifteen years [12]. Traditionally, software engineers have taken very pragmatic approaches to data representation, encoding only the information needed to solve the problem in hand, usually in the form of language, data structures, or database tables. Modelling approaches are meant to increase the productivity by maximising compatibility between systems (by reuse of standardised models), simplifying the process of design (by models of recurring design patterns in the application domain), and promoting communication between individuals and teams working on the system (by a standardisation of the terminology and the best practices used in the application domain) [13]. On the other hand, ontologies have been developed by the Artificial Intelligence community since the 1980s. An ontology is a structuring framework for organising information. It renders shared vocabulary and taxonomies which models a domain with the definition of objects and concepts and their properties and relations. These ideas have been heavily drawn upon in the notion of the Semantic Web [13].

Traditional views tend to distinguish the two technologies. The main points of argument include but are not limited to:

1. Models usually focus on realisation issues (e.g., the Object-Oriented Modelling approach), while ontologies usually focus on capturing abstract domain concepts and their relationship [14].
2. Ontologies are normally used for run-time knowledge exploitation (e.g., for knowledge discovery in a **knowledge base**), but models normally do not [15].
3. Ontologies can support reasoning while models cannot (or do not) [13].
4. Finally, models are often based on the Closed World Assumption while ontologies are based on the Open World Assumption [13].

However, these separations between the two technologies are rapidly disappearing in recent developments. Study [13] shows that 'all ontologies are models', and 'almost all models used in modern software engineering qualify as ontologies.' As evidenced by the growing number of research workshops dealing with the overlap of the two disciplines (e.g., SEKE [16], VORTE [17], MDSW [18], SWESE [19], ONTOSE [20], WoMM [21]), there has been considerable interest in the integration of software engineering and artificial intelligence technologies in both research and practical software engineering projects [13].

We tend to take this point of view and regard the ENVRI Reference Model as both a model and an ontology. The important consequence is that we can explore further in both directions, e.g., the reference model can be expressed using a modelling language, such as UML (UML4ODP). It can then be built into a tool chain, e.g., to plugin to an integrated development environment such as Eclipse, which makes it possible to reuse many existing UML code and software. On the other hand, the reference model can also be expressed using an ontology language such as RDF or OWL which can then be used in a **knowledge base**. In this document we explore principally from model aspects. In another ENVRI task, T3.4, the ontological aspect of the reference model will be exploited.

Finally, a reference model is a **standard**. Created by ISO in 1970, OSI is probably among the earliest reference models, which defines the well-known 7-layered network communication. As one of the ISO standard types, the reference model normally describes the overall requirements for standardisation and the fundamental principles that apply in implementation. It often serves as a framework for more specific standards [22]. This type of standard has been rapidly adopted, and many reference models exist today, which can be grouped into 3 categories, based on the type of agreement and the number of people, organisations or countries who were involved in making the agreement:

- **Committee reference model** – a widely-based group of experts nominated by organizations who have an interest in the content and application of the standard build the standard.
- **Consensus reference model** – the principle that the content of the standard is decided by general agreement of as many as possible of the committee members, rather than by majority voting. The ENVRI Reference Model falls into this group.
- **Consultation reference model** – making a draft available for scrutiny and comment to anyone who might be interested in it.

Some examples from each of the categories are discussed below, with emphasis on approaches of building the model and technologies the model captures.

## Related Reference Models

### Committee Reference Models

In this category, we look at those defined by international organisations, such as the Advancing Open Standards for the Information Society (OASIS), the Consultative Committee for Space Data Systems (CCSDS), and the Open Geospatial Consortium (OGC).

The Open Archival Information System (OAIS) Reference Model [11] is an international standard created by CCSDS and ISO which provides a framework, including terminology and concepts for archival concept needed for Long-Term digital information preservation and access.

The OASIS Reference Model for Service Oriented Architecture (SOA-RM) [8] defines the essence of service oriented architecture emerging with a vocabulary and a common understanding of SOA. It provides a normative reference that remains relevant to SOA as an abstract model, irrespective of the various and inevitable technology evolutions that will influence SOA deployment.

The OGC Reference Model (ORM) [23], describes the OGC Standards Baseline, and the current state of the work of the OGC. It provides an overview of the results of extensive development by OGC Member Organisations and individuals. Based on RM-ODP's 5 viewpoints, ORM captures business requirements and processes, geospatial information and services, reusable patterns for deployment, and provides a guide for implementations.

The Reference Model for the ORCHESTRA Architecture (RM-OA) [24] is another OGC standard. The goal of the integrated project ORCHESTRA (Open Architecture and Spatial Data Infrastructure for Risk Management) is the design and implementation of an open, service-oriented software architecture to overcome the interoperability problems in the domain of multi-risk management. The development approach of RM-OA is standard-based which is built on the integration of various international standards. Also using RM-ODP standard as the specification framework, RM-OA describes a platform neutral (abstract) model consisting of the informational and functional aspects of service networks combining architectural and service specification defined by ISO, OGC, W3C, and OASIS [24].

There are no reference model standards yet for environmental science research infrastructures.

### Consensus Reference Models

In this category, we discuss those created by non-formal standard organisations.

The LifeWatch Reference Model [25], developed by the EU LifeWatch consortium, is a specialisation of the RM-OA standard which provides the guidelines for the specification and implementation of a biodiversity research infrastructure. Inherited from RM-OA, the reference model uses the ODP standard as the specification framework.

The Digital Library Reference Model [9] developed by DL.org consortium introduces the main notations characterising the whole digital library domain, in particular, it defines 3 different types of systems: (1) Digital Library, (2) Digital Library System, and (3) Digital Library Management System; 7 core concepts characterising the digital library universe: (1) Organisation, (2) Content, (3) Functionality, (4) User, (5) Policy, (6) Quality, and (7) Architecture; and 3 categories of actors: (1) DL End-Users (including, Content Creators, Content Consumers, and Digital Librarians), (2) DL Managers (including, DL Designer, and DL System Administrators), and (3) DL Software Developers.

The Workflow Reference Model [26] provides a common framework for workflow management systems, identifying their characteristics, terminology and components. The development of the model is based on the analysis of various workflow products in the market. The workflow Reference Model firstly introduces a top level architecture and various interfaces it has which may be used to support interoperability between different system components and integration with other major IT infrastructure components. This maps to the ODP Computational Viewpoint. In the second part, it provides an overview of the workflow application program interface, comments on the necessary protocol support for open interworking and discusses the principles of conformance to the specifications. This maps to the ODP Technology Viewpoint.

The Agent System Reference Model [27] provides a technical recommendation for developing agent systems, which captures the features, functions and data elements in the set of existing agent frameworks. Different from conventional methods, a reverse engineering method has been used to develop the reference model, which starts by identifying or creating an implementation-specific design of the abstracted system; secondly, identifying software modules and grouping them into the concepts and components; and finally, capturing the essence of the abstracted system via concepts and components.

### **Consultation Reference Models**

The Data State Reference Model [28] provides an operator interaction framework for visualisation systems. It breaks the visualisation pipeline (from data to view) into 4 data stages (Value, Analytical Abstraction, Visualisation Abstraction, and View), and 3 types of transforming operations (Data Transformation, Visualisation Transformation and Visual Mapping Transformation). Using the data state model, the study [29] analyses 10 existing visualisation techniques including, 1) scientific visualisations, 2) GIS, 3) 2D, 4) multi-dimensional plots, 5) trees, 6) network, 7) web visualisation, 8) text, 9) information landscapes and spaces, and 10) visualisation spread sheets. The analysis results in a taxonomy of existing information visualisation techniques which help to improve the understanding of the design space of visualisation techniques.

The Munich Reference Model [30] is created for adaptive hypermedia applications which is a set of nodes and links that allows one to navigate through the hypermedia structure and that dynamically “adapts” (personalise) various visible aspects of the system to individual user’s needs. The Munich Reference Model uses an object-oriented formalisation and a graphical representation. It is built on top of the Dexter Model layered structure, and extends the functionality of each layer to include the user modelling and adaptation aspects. The model is visually represented using in UML notation and is formally specified in Object Constraint Language (which is part of the UML).

While these works use a similar approach to the development of the reference model as the ENVRI-RM, which is based on the analysis of existing systems and abstracts to obtain the ‘essence’ of those systems, a major difference is that these works have not normally met with significant feedback or been formally approved by an existing community, with the consequence that they express less authority as a standard.

### **Other Related Standards**

Data Distribution Service for Real-Time Systems (DDS) [31], an Object Management Group (OMG) standard, is created to enable scalable, real-time, dependable, high performance, interoperable data exchanges between publishers and subscribers. DDS defines a high-level conceptual model as well as a platform-specific model. UML notations are used for specification. While DDS and the ENVRI share many similar views in design and modelling, DDS focuses on only one specific issue, i.e., to model the communication patterns for real-time applications; while ENVRI aims to capture a overall picture of requirements for environmental research infrastructures.

Published by the web standards consortium OASIS in 2010, the Content Management Interoperability Services (CMIS) [32] is an open standard that allows different content management systems to inter-operate over the Internet. Specially, CMIS defines an abstraction layer for controlling diverse document management systems and repositories using web protocols. It defines a domain model plus web services and Restful AtomPub bindings that can be used by applications to work with one or more Content Management repositories/systems. However as many other OASIS standards, CMIS is not a conceptual model and is highly technology dependent [32].