

SV Community Roles

A **role** in a community is a prescribing behaviour that can be performed any number of times concurrently or successively. A role can be either *active* (typically associated with a human actor) or *passive* (typically associated with a non-human actor, e.g. software or hardware components).

Active roles are identified in relation to people associated with a research infrastructure:

- those who use the research infrastructure to do science;
- those who work on resources to build, maintain and operate the research infrastructure; and
- those who govern, manage and administer the research infrastructure

Note

An individual may be a member of more than one community by undertaking different roles.

Passive roles are identified with subsystems, subsystem components, and hardware facilities. Active roles interact with passive roles to achieve their objectives.

- [Research Infrastructure](#)
- [Roles in the Data Acquisition Community](#)
- [Roles in the Data Curation Community](#)
- [Roles in the Data Publishing Community](#)
- [Roles in the Data Processing Community](#)
- [Roles in the Data Use Community](#)

Research Infrastructure

Research Infrastructure is the main entity being modeled in any ENVRI RM specification, additionally RI is a special role which can be part of the communities. This role is defined as follows.

Research Infrastructure: An active or passive role, which is the conglomeration of research resources providing some subset of data acquisition, data curation, data publishing, data processing and data use functionality to a research community.

Roles in the Data Acquisition Community

The main objective of the data acquisition community is to bring measurements into the system. Consider a typical data acquisition scenario: A measurement and monitoring model is designed by *designers* based on the requirements of *environmental scientists*. Such a design decides what data is to be collected and what metadata is to be associated with it, e.g., experimental information and instrument conditions. *Technicians* configure and calibrate a *sensor* or a *sensor network* to satisfy the experiment specifications. In the case where human sensors are to be used, *observers* or *measurers* input the measures to the system, e.g., by using mobile devices. *Data collectors* interact with a data acquisition subsystem to prepare the data or control the flow of data in order to automatically collect and transmit the data.

The following roles are identified in a data acquisition community:

- **Environmental Scientist:** An active role, which is a person who conducts research or performs scientific investigations. Using knowledge of various scientific disciplines, they may collect, process, analyse, synthesize, study, report, and/or recommend action based on data derived from measurements or observations of (for example) air, rock, soil, water, nature, and other sources.
- **Sensor:** A passive role, which is a converter that measures a physical quantity and converts it into a signal which can be read by an observer or by an (electronic) instrument.
- **Sensor network:** A passive role, which is a network consisting of distributed autonomous sensors to monitor physical or environmental conditions.
- **Measurement Model Designer:** An active role, which is a person who designs the measurements and monitoring models based on the requirements of environmental scientists.
- **Technician:** An active role, which is a person who develops and deploys sensor instruments, establishing and testing the sensor network, operating, maintaining, monitoring and repairing the observatory hardware.
- **Measurer:** An active role, which is a person who determines the ratio of a physical quantity (such as a length, time, temperature etc.), to a unit of measurement (such as the meter, second or degree Celsius).
- **Observer:** An active role, which is a person who receives knowledge of the outside world through his senses, or records data using scientific instruments.
- **Data collector:** Active or passive role, adopted by a person or an instrument collecting data.
- **Data Acquisition Subsystem:** In the Science Viewpoint, the data acquisition subsystem is passive role of the data acquisition community. It is the part of the research infrastructure providing functionalities to automate the process of data acquisition.

Acquisition Community Roles

Notation

The behaviours of the data acquisition community is described at [Acquisition Behaviours](#).

Roles in the Data Curation Community

The data curation community responds to provide quality data products and maintain the data resources. Consider a typical data curation scenario: when data is being imported into a curation subsystem, a *curator* will perform the quality checking of the scientific data. Unique identifiers will be assigned to the qualified data, which will then be properly catalogued by associating necessary metadata, and stored or archived. The main human roles interacting with or maintaining a data curation subsystem are *data curators* who manage the data and *storage administrators* who manage the storage facilities. Upon registering a digital object in a repository, its *persistent identifier (PID)* and the *repository* name or IP address is registered with a globally available system of *identification services (PID service)*. Users may subsequently present the *PID* to an *PID service* to learn the network names or addresses of repositories in which the corresponding digital object is stored. Here, we use a more general term "PID" instead of "*handle*", and identify the key roles involved in the data curation process.

We identified the following roles in this community:

- **Data Curator:** An active role, which is a person who verifies the quality of the data; annotates the data; catalogues, preserves and maintains the data as a resource; and prepares various required data products.
- **Semantic Curator:** An active role, which is a person who designs and maintains local and global conceptual models and uses those models to annotate the data and metadata.
- **Storage Administrator:** An active role, which is a person who has the responsibilities to design data storage, tune queries, perform backup and recovery operations, set up RAID mirrored arrays, and make sure drive space is available for the network.
- **PID Manager:** A passive role, a system or service that assigns persistent global unique identifiers to data and metadata products. The Manager invokes an external entity, the PID Service, to obtain the PIDs. The manager maintains a local catalogue of PIDs that are being used to reference data and metadata. If the data or metadata in the RI change location or are removed, the PID manager updates this information locally and informs the PID Service.
- **PID Service:** A passive role, a public system or service which can generate and assign persistent global unique identifiers (PIDs). The PID Service also maintains a public registry of PIDs for digital objects.
- **Storage System:** A passive role, which includes memory, components, devices and media that retain data and metadata for an interval of time.
- **Catalogue System:** A passive role, a catalogue system is a special type of storage system designed to support building logical structures for classifying data and metadata.
- **Data Curation Subsystem:** the data curation subsystem is a passive role of the data curation community. It is the part of the research infrastructure which stores, manages and ensures access to all persistent data and metadata produced within the infrastructure.

Note

PID Service was called PID Generator, however, the analysis of the Identification and Citation practices made evident that the generation can be done inside the RI (by PID Manager), shared between the RI and the PID Service, or completely delegated to a PID Service. Consequently the names were changed after version 2.1 of the ENVRI RM.

The PID generator does not disappear completely, it is a refinement (specialisation/subclass) which can be implemented by PID service or the PID Manager

Curation Community Roles

Notation

The behaviours of the data curation community are described at [Curation Behaviours](#).

Roles in the Data Publishing Community

The objectives of the data publishing community are to publish data and assist discovery and access. We consider the scenarios described by Kahn's data publication model [34]: an originator, i.e., a user with digital material to be made available for public access, makes the material into a digital object. A digital object is a data structure whose principal components are digital material, or data, plus a unique identifier for this material (and, perhaps, other material). To get a unique identifier, the user requests one from an authorised PID service. A user may then deposit the digital object in one or more repositories, from which it may be made available to others (subject, to the particular item's terms and conditions, etc.).

The published data are to be discovered and accessed by data consumers. A semantic mediator is used to facilitate the heterogeneous data discovery.

In summary, the following roles are involved in the data publication community:

- **Data Originator:** Either an active or a passive role, which provides the digital material to be made available for public access.
- **Data Repository:** A passive role, which is a facility for the deposition of published data.
- **Semantic Mediator:** A passive role, which is a system or middleware facilitating semantic mapping (i.e., executing mapping and translation rules), discovery and integration of heterogeneous data.
- **Data Publisher:** An active role, is a person in charge of supervising the data publishing processes.

- **Data Publishing Subsystem:** In the Science Viewpoint, the data publishing subsystem represents a passive role of the data publication community. It is the part of the research infrastructure enabling the discovery and retrieval of scientific data. The access to this subsystem could require authorisation at different levels for different roles.
- **Data Consumer:** Either an active or a passive role, which is an entity who receives and uses the data.
- **Metadata Harvester:** A passive role, which is a system or service collecting metadata which supports the construction/selection of a global conceptual model and the production of mapping rules

Publishing Community Roles
Notation

The behaviours of the data publishing community are described at [Publishing Behaviours](#).

Roles in the Data Processing Community

The data processing community provides various application services such as data analysis, mining, simulation and modelling, visualisation, and experimental software tools, in order to facilitate the use of the data. We consider scenarios of service oriented computing paradigm which is adopted by the ENVRI implementation model, and identify the key roles as below. These concepts are along the lines of the existing standards such as OASIS Reference Model for Service Oriented Architecture.

- **Data Provider:** Either an active or a passive role, which is an entity providing the data to be used.
- **Service:** A passive role, in which a functionality for processing data is made available for general use.
- **Service Consumer:** Either an active or a passive role, which is an entity using the services provided.
- **Service Provider:** Either an active or a passive role, which is an entity providing the services to be used.
- **Service Registry:** A passive role, which is an information system for registering services.
- **Capacity Manager:** An active role, which is a person who manages and ensures that the IT capacity meets current and future business requirements in a cost-effective manner.
- **Data Processing Subsystem:** In the Science Viewpoint, the data processing subsystem represents a passive role of the data processing community. It is the part of the research infrastructure providing services for data processing. These services could require authorisation at different levels for different roles.
- **Processing Environment Planner:** An active agent that plans how to optimally manage and execute a data processing activity using RI services and the underlying e-infrastructure resources (handling sub-activities such as data staging, data analysis/mining and result retrieval).

Processing Community Roles
Notation

The behaviours of the data processing community are described at [Processing Behaviours](#).

Roles in the Data Use Community

The main role in the data use community is a *user* who is the ultimate consumer of data, applications and services. Depending on the purposes of use, a user can be one of the following active roles:

- **Scientist (synonym: Researcher):** An active role, which is a person who makes use of the data and application services to conduct scientific research.
- **Engineer (synonym: Technologist):** An active role, which is a person who develops and maintains the research infrastructure.
- **Educator (synonym: Trainer):** An active role, which is a person who makes use of the data and application services for education and training purposes.
- **Policy Maker (synonym: Decision Maker):** An active role, which is a person who makes decisions based on the data evidence.
- **Stakeholder (synonyms: Private Investor, Private Consultant):** An active role, which is undertaken by a person who makes use of the data and application service for predicting markets so as to make business decisions on producing related commercial products.
- **Citizen (synonyms: General Public, Media):** An active role, which is a person or organisation interested in understanding the knowledge delivered by an environmental science research infrastructure, or discovering and exploring the **knowledge base** enabled by the research infrastructure.
- **Citizen Scientist:** An active role, member of the general public who engages in scientific work, often in collaboration with or under the direction of professional scientists and scientific institutions (also known as amateur scientist).
- **Data Use Subsystem:** In the Science Viewpoint, the data use subsystem represents a passive role of the data use community. It is the part of the research infrastructure supporting the access of users to an infrastructure. The data use subsystem manages, and tracks user activities and supports users to conduct their roles in different communities.

Use Community Roles
Notation

The behaviours of the data use community are described at [Use Behaviours](#).

