# GORC Interest Group: Typology and Definitions

Version: 1.01, August 11, 2023

## Context

The Global Open Research Commons is an Interest Group within the Research Data Alliance. The IG is working to reach a shared understanding of what a "Commons" is in the research data space; what functionality, coverage and characteristics does such an initiative require and how this can be coordinated at a global level.

**Commons:**

> A global trusted ecosystem that provides seamless access to high quality interoperable research outputs and services.

**Strapline:**

> Digital research resources for the common good

One of the outputs of the IG is a Typology of the essential elements in a Commons. In this context, a typology means a classification of elements into different types[1]. In developing this typology, the IG identified the need to also provide a set of definitions for each of the typology elements. This document is the formal statement of this typology with the associated definitions.

## Scope

This can be defined against a number of different dimensions:

- Discipline coverage: is the commons focussed on a specific discipline (e.g. astronomy) or is it a general commons (e.g. The Australian Research Data Commons)?
- Geographical coverage: is the commons national (e.g. KISTI), pan-national (e.g. EOSC) or international (IVOA)?
- Contents: what kinds of research objects (data, software, publications, models, etc) does it contain?
- Sources: what are the sources of input (instruments, surveys, manual data collection) into the commons?

[1] The detail of the elements and attributes associated with each type is part of the work plan for the GORC International Model WG.

## Contents

The set of possible research inputs and outputs vary widely by discipline, and thus the possible contents of different Commons may also be diverse. In some disciplines, physical collections are both an input to, and an output of, research (for instance herbaria, insect collections, map collections, geophysical samples, biobanks, 3d printed objects, etc.).

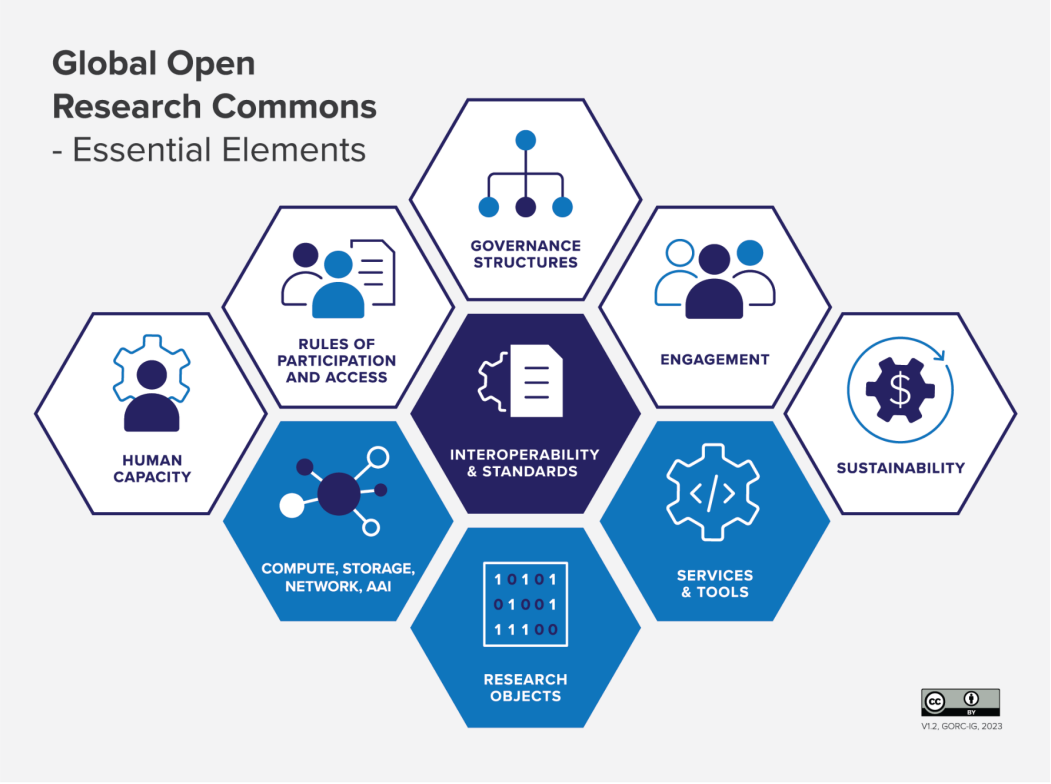
**The scope of the GORC is intentionally restricted to digital information.** This may involve surrogates for physical collections (e.g., 3d scans of insects, microscopy images of tissue samples, X-rays of rocks, etc.) but does not include the physical samples themselves.

## Sources

In some disciplines, instruments of all kinds (e.g., telescopes, NMR scanners, camera traps, seismic sensors, remote sensing, including satellites, photogrammetry, drones, microscopes, etc.) are significant generators of data.

While these are essential to the wider conduct of research, the **GORC scope intentionally does not include the instruments that generate the data** that is made available through the commons. This may be revisited in a later version.

# Typology of Elements



The three elements in blue are the underpinning digital elements that constitute the parts of the commons with which people interact. The five elements in white are the social/human elements that are needed to make the commons succeed. The central element in dark blue represents the way in which standards are at the core of a commons.

# Definitions

NOTE: These definitions are current as of mid 2023. As this field evolves, the definitions and typology may need to be revisited.

## Governance & Leadership

Governance is focused on defining theorganization's purpose and the development of the strategies, objectives, values, and policies that frame how that purpose will be pursued by management and internal personnel of the commons. It includes the development of such things as strategic plans, including mission statements, values, organizational performance metrics, risk management frameworks; policies and guidelines for financial and operational matters such as commitments to community-endorsed principles and frameworks like FAIR and data ethics; and the creation and maintenance of governance structures, their interactions with stakeholders, and the ways of working with management. Typically the governance processes will be operated via a series of steering groups or boards, involving key stakeholders for the commons such as funders, national services and community representatives. Data governance is classified under the Rules of Participation and Access element.

## Rules of Participation & Access

The Rules of Participation and Access element is the set of policies defining a minimal set of rights, obligations and accountability governing the activities of those participating in the Commons (this definition and the concept of Rules of Participation are drawn from the European Open Science Cloud). Participant rights vary by Commons and are not enumerated here, but could include rights "that enable and foster reuse and reproducibility." These should fulfill the following criteria, inspired by the EOSC Rules of Participation as listed in this report from the EOSC Executive Board (but noting that this is under active consideration within the EOSC):

- The RoP apply to all users of the Commons
- They apply not only to data and services, but also, for example, to publications, software, tools, workflows, training, and consultancy
- They apply to all resources provided or accessed via the Commons
- They apply to users of the Commons, including users from research, industry and the public
- They apply to publicly-funded and commercial data and service providers as well as consumers

These overlap with, and reinforce, Elinor Ostrom's Eight Principles for Managing Commons (written for natural resource commons, but also applicable here).

## Engagement

Methods used to interact with the broad stakeholder community to involve them in activities. These could include requirements gathering exercises, community needs assessments, consultations, usability testing, communications, events, and training amongst others.

## Human Capacity

The ability of the Commons to create a human-friendly environment for all stakeholders and community members in all aspects, specifically for users, providers, and intermediaries, so that the Commons can set and achieve objectives, perform functions, solve problems, and continue to develop the means and conditions required to enable this process (adapted from https://www.fao.org/3/y5613e/y5613e08.htm).

The human capacity of a Commons should be viewed in the context of a community which includes all human individuals and entities that could be considered stakeholders, users, providers, members of the Commons, and intermediaries (i.e., those who do not interact directly with the Commons, but use information about or provided by the commons, e.g., policy makers, journals, funders) in the past, present, and future as well as all research communities that the research Commons is a part of in a regional, national, and global context. Depending on the structure of the Commons, stakeholders may include funding and government bodies as well as related commercial entities; these may be in scope through the provision of related services or an association with related research communities.

## Sustainability

Models and agreements made on how to ensure the viability and continued or transitioned operations of the Commons, including funding and resourcing activities, in a way that can be sustained over the long-term. This may include mixed streams of investment and cost recovery through subscriptions or service payment models to ensure operation of the Commons, as well as the input of effort/time by contributors to, and maintainers of, all elements of Commons infrastructure and interactions with stakeholders. Sustainability should also include defining and developing a strategy for long-term sustainability for all operations and holdings, as well as keeping issues of sustainability in mind when choosing or building Commons components. For instance, reuse of existing components is an effective strategy for more sustainable Commons infrastructures.

## Interoperability

The ability of data or tools from non-cooperating resources to integrate or work together with minimal effort. (Wilkinson, et al. 2016). This ability often supports, but is not limited to, discoverability, access, reuse and reproducibility. It can also enable a wider range of cross-commons use-cases. Types of interoperability include: Technical interoperability (how artefacts are exchanged), Syntactic interoperability (how to structure information), Semantic interoperability (data are interpreted the same way) and Pragmatic interoperability (agreements between organizations). (Janssen et al 2014)

## Standards & Conventions

A standard is a repeatable, harmonised, agreed and documented way of doing something (IRENA). Standards can be either *de jure* or *de facto. De jure* standards, or standards according to law, are endorsed by a formal standards organization, such as the ISO. The organization ratifies each standard through its official procedures and gives the standard its stamp of approval.

*De facto* standards, or conventions, are adopted widely by a community. These standards arise when they become part of the accepted way of doing things within a community. *De facto* standards can become *de jure* standards if they are submitted to a formal standards organization and approved.

Within the data commons context, standards and conventions may cover various things:

- Metadata standards used to describe data according to agreed schema
- Controlled vocabularies and ontologies used to label and assign keywords according to semantically agreed and standardised terms
- Encoding standards for multilingual content
- Data formats used to structure data in a common way that is accepted by the community
- Service endpoints that allow humans and machines to bind and consume resources.
- Authentication and authorization protocols

## ICT Infrastructure (Compute, Storage, Network, AAI)

By "ICT infrastructure" we mean the physical components that a computer system requires to function and are necessary to conduct research. This includes:

- Compute: access to processing cycles, delivered via on-premise hardware or off-premise cloud services
- Storage: access to data storage, delivered via on-premise hardware or off-premise cloud services
- Network: required to connect compute and storage, as well as to access Internet resources
- AAI: Authentication and authorization infrastructure refers to services and procedures that enable members of different institutions to access protected information that is distributed on different servers

## Services and Tools

The context for these definitions is the emerging and complex intersection of tools, disciplines, services, platforms, hardware, resources, and the people (users, researchers, developers, stakeholders, personnel and communities, etc.) who use and contribute to them.

Service (as defined by IVOA) A service is any Commons element that can be invoked by the user to perform some action on their behalf. Services are usually intended for use by machines.

Tools enable researchers to perform one or more operations, typically on data, and often with data as the output. Tools are usually intended for use by humans. In this context we are explicitly excluding physical instruments.

As research infrastructure, services and tools are often made available through research platforms (variously referred to as virtual science labs, virtual research environments (VREs), or Science Gateways,) that are deployed to support both the research workflows and the communities of practice engaged in collaborative research. Typically, a research platform's capabilities include data acquisition and management, processing and visualization, storage and preservation, sharing and discovery; platforms may provide the full spectrum or a subset of components. Science Gateways may be discipline-specific, and may support and enhance scientific collaboration and scholarly communication by facilitating citizen science engagement as well.

## Research Objects

These are the outputs of the research process, but can also be inputs to later processes. Here, we limit our scope to digital research objects. The research object approach is primarily motivated by a desire to improve reproducibility of scientific investigations. Central to the proposal is a need to share research artifacts commonly distributed across specialist repositories on the Web including publications, lab notebooks, blog entries, supporting data, software executables, source code, presentation slides, and presentation videos. See https://www.researchobject.org/ for more.

# References

The GORC IG would like to acknowledge a number of existing definition documents that were used as inspiration for these definitions:

- The CASRAI glossary
- The EOSC glossary
- A glossary of terms relating to open scholarship
- The draft wikidata open infrastructure elements and definitions