RDA Case Statement

GORC International Benchmarking WG v.2

Last update 01April2021

1. Charter

The Global Open Research Commons (GORC) is an ambitious vision of a global set of interoperable resources necessary to enable researchers to address societal grand challenges including climate change, pandemics, and poverty. The realized vision of GORC will provide frictionless access to all research artifacts including, but not limited to: data, publications, software and compute resources; and metadata, vocabulary, and identification services to everyone everywhere, at all times.

The GORC is being built by a set of national, pan-national and domain specific organizations such as the European Open Science Cloud, the African Open Science Platform, and the International Virtual Observatory Alliance (see Appendix A for a fuller list). The GORC IG is working on a set of deliverables to support coordination amongst these organizations, including a roadmap for global alignment to help set priorities for Commons development and integration. In support of this roadmap, this WG will establish benchmarks to compare features across commons. We will not coordinate the use of specific benchmarks by research commons. Rather, we will review and identify features *currently* implemented by a target set of GORC organizations and determine how they measure their user engagement with these features.

In the first case we will collect and curate a set of benchmarks that will allow Commons developers to compare features across science clouds. For example, we would consider benchmarks such as evidence or the existence of :

1. A well defined decision making process
2. A consistent and openly available data privacy policy
3. Federated Authentication and Authorization infrastructure
4. Community supported and well documented metadata standard(s)
5. A workflow for adding and maintaining PIDs for managed assets
6. A mechanism for utilizing vocabulary services
7. A process to inventory research artefacts and services
8. An Open Catalogue of these artefacts and services
9. A proven workflow to connect multiple different research artefact types (e.g. data and publications; data and electronic laboratory notebooks; data and related datasets)
10. A mechanism to capture provenance for research artefacts
11. Mechanisms for community engagement and input; an element or scale for inclusion

These benchmarks will be an initial starting point of what we would expect to find in a mature research commons. We will then review each of the commons in the target list to see if they provide other features that should be included as benchmarks. As part of our review, we will document implementations of features in research commons.

We will collect information about each of the benchmarks we see "in the wild", but the benchmarks are not intended to be prescriptive regarding implementation. For example, the benchmark: "A mechanism for utilizing (or accessing) vocabulary services" is evidenced by the NERC Vocabulary Server (NVS) in the EOSC, and by Research Vocabularies Australia (RVA) in the Australian National Data Service (ANDS). NERC uses the Simple Knowledge Organization System (SKOS) to represent concepts in the vocabulary service and provides access via both SPARQL

and SOAP endpoints. ANDS RVA also serves SKOS-encoded vocabularies and provides a SPARQL endpoint, but also a RESTful API, and the option to bulk download complete vocabularies in a single file for local processing. The benchmark in this case is evidence of the ability to use a vocabulary service, satisfied by both RVA and NVS. Whenever possible we will collect information about the particular implementation of the benchmark or feature as we review the commons, but that is not the primary goal. The WG will collectively decide what constitutes a benchmark. For example, the ANDS RVA service also has the ability for users to self-register and create, edit or upload vocabularies, a function not available in the NVS. In this case, the WG will decide if the ability to create and edit, not just access, a vocabulary service should constitute a separate benchmark. Whenever possible we will utilize outputs from other RDA groups to identify benchmarks. In particular the [RDA 9 functional requirements for data discovery](#) will be very informative of the benchmarks associated with data repositories.

Secondly, the WG will collect information about how existing commons are measuring success, adoption or use of their services within their organization, such as data downloads, contributed software, and similar KPI and access statistics. The first set of benchmarks will be the existence of a feature or service and is comparable across organizations. The second set of benchmarks are quantitative measures used within an organization to measure the uptake or use of a feature or service.

2. Value Proposition

This WG is motivated by the broader goal of openly sharing data and related services across technologies, disciplines, and countries to address the grand challenges of society. The deliverables of the WG itself will inform roadmaps for development of the infrastructure necessary to meet that goal, while engagements and relationships formed during the work period will help forge strong partnerships across national, regional and domain focused members which are crucial to its success. Identifying observable and measurable benchmarks in

pursuit of the global open science commons will help create a tangible path for development and support strategic planning within and across science commons infrastructures. In the future, best practices for commons development will emerge based on the experience of what actions led to successful outcomes. This work will provide a forum for discussion that will allow members to identify the most important features and the minimal elements required to guide their own development and build a commons that is globally interoperable. Building interoperable commons will support many research efforts including work focused on societal grand challenges and UN Sustainable Development Goals (SDGs). Finally, it will support developers as they seek resources to build the global commons by helping them respond to funding agencies requirements for measurable deliverables.

The proposed WG was discussed at the [RDA 16 virtual plenary](#).[1] Participants discussed the initial work packages and agreed during the meeting this was a worthy goal and an appropriate approach.

3. Engagement with Existing Work

This WG will review all appropriate IG and WG outputs to determine intersection with this work, and engage with the WG/IGs as appropriate. Some of the efforts are reasonably well known now: the GORC IG builds on, and incorporates the previous [National Data Services IG](#), which was embarking on a similar exercise when the GORC started; the [Domain Repositories IG](#), specifically the repository-specific discovery metrics/benchmarks. The Commons that will be investigated in this WG are likely either to have considered or implemented outputs from other RDA groups, such as the [Data Fabric IG](#), and the [Virtual Research Environment IG](#), just to name a few. These groups and many others outside of RDA will have recommendations that speak to

---

[1] P16 [session notes](#) and [presentation](#)

functionality and features of various components of Commons; for example the EOSC [FAIR WG](#) and [Sustainability WG](#) that seek to define the EOSC as a Minimum Viable Product (MVP). We will review these and other related outputs to see if they have identified benchmarks that we feel will support our goals. This review period will ensure that we do not duplicate existing efforts. [Appendix B](#) of this case statement identifies a few of these existing efforts, both within and without RDA; this list will be expanded and reviewed by the WG members.

4. Work Plan

To create these deliverables, members of the group will:

1. Create a target list of Commons ([Appendix A](#))
2. Create a database structure to capture benchmarks
3. Create an initial list of benchmarks
4. Create an online form to capture benchmarks
5. Create task groups within the Benchmarking WG, each responsible for reviewing a subset of the target list and ancillary documents
6. Each task group reviews public facing documentation of their assigned Commons to extract benchmarking information (both KPIs and feature lists) and reports back to the larger WG.
7. A separate task group reviews public facing documentation of recommendations and roadmaps from related communities to extract benchmarking information ([Appendix B](#)) and reports back to the larger WG. This evaluation phase will include an examination of the outputs from other RDA WGs and position papers available in the wider science infrastructure community, along with experiences gathered by the WG's members.

8. Because benchmarking information may not be easily found in public documents we will conduct outreach to Commons representatives and related organizations to ask for additional feedback and information about benchmarks used by their community. This may include benchmarks already in use, as well as benchmarks that organizations feel would be useful but which are not yet implemented.
9. Begin drafting adoption plan
10. Synthesize and document the benchmarks into 3 deliverables, described below.

There are multiple ways for the WG to create task groups. The WG will decide if they would rather define the task group according to the deliverables, creating a Commons Internal Benchmarking TG and a Commons External Benchmarking TG, or if they would rather subdivide according to a typology of the commons, for example with some members looking at pan-national, national, or domain specific commons, or by some other subdivision of labor.

The WG will proceed according to the following schedule:

Month	Activity
Jan-Mar 2020	Group formation 1. Agreement on the scope of work and deliverables (broad scope) 2. Case statement community review
Apr-Jun 2021	RDA17 Refine scope: Agree to target list of commons and organizational approach Begin to define methodology, especially the form for data collection and initial set of commons

	Begin literature review of public facing documents from Science Commons and related organizations Report on progress to International Symposium on Global Open Science Cloud (June 2021)
Jul-Sep 2021	Recruit additional members to WG, continue lit review
Oct-Dec 2021	Begin outreach to Science Commons and related organizations Update at RDA18 Report on progress to https://internationaldataweek.org/ (Nov 2021)
Jan-Mar 2022	First draft: External Benchmarks distributed for community review
Apr-Jun 2022	First draft: Internal Benchmarks distributed for community review Update at RDA19
Jul-Sep 2022	Develop adoption plan
Oct-Dec 2022	Final deliverables Update at RDA20

5. Deliverables

This group will create Supporting Outputs in furtherance of the goals of the GORC IG.

Specifically, 3 documents:

D1: a list of observable international benchmarks of features, structures and functionality that can help define a Commons and that will feed into a roadmap of Commons interoperability. The benchmark criteria needs to remain simple, understandable and not skewed towards the particular reality of some of the commons so as not to appear as irrelevant or unattainable to commons developers. It will include a description of implementations observed or planned in Commons examined in this work.

D2: a non-redundant set of KPIs and success metrics currently utilized, planned or desired for existing science commons, and classified by functional layers defined by the GORC IG; how do we define a minimal interoperability

D3: Adoption plan, described below

6. Mode and Frequency of Operation

The WG will meet monthly over Zoom, at a time to be determined by the membership. The WG will also communicate asynchronously online using the mailing list functionality provided by RDA and via shared online documents. If and when post-Covid international travel is restored during the 18 month work period of this WG then we will propose and schedule meetings during RDA plenaries and at other conferences where a sufficient number of group members are in attendance.

7. Addressing Consensus and Conflicts

The WG will adhere to the stated [RDA Code of Conduct](#) and will work towards consensus, which will be achieved primarily through mailing list discussions and online meetings, where opposing views will be openly discussed and debated amongst members of the group. If consensus

cannot be achieved in this manner, the group co-chairs will make the final decision on how to proceed.

The co-chairs will keep the working group on track by reviewing progress relative to the deliverables. Any new ideas about deliverables or work that the co-chairs deem to be outside the scope of the WG defined here will be referred back to the GORC IG to determine if a new WG should be formed.

8. Community Engagement

The working group case statement will be disseminated to RDA mailing lists and communities of practice related to Commons development that are identified by the GORC IG in an effort to cast a wide net and attract a diverse, multi-disciplinary membership. The GORC Benchmarking effort is also being facilitated by the RDA Secretariat, providing a strong intersection with the EOSC community - this will provide an additional level of community engagement. Similarly, the CODATA GOSC work, and the associated coordination of both efforts by the Data Together group, will provide additional engagement and outreach to the WDS and GO FAIR communities. When appropriate, draft outputs will also be published to relevant stakeholders and mailing lists to encourage broad community feedback, this will include both the GORC WG and GORC IG membership. When appropriate we will ask members of the WG to reach out to their own networks.

9. Adoption Plan

○ The Adoption Plan will be detailed in an additional document that will provide additional information for the 2 primary outputs, and will include the following.
  i. Integration of the benchmarks into the Typology and larger GORC roadmap being created by the parent IG.

ii.   Integration/intersections with the CODATA GOSC work, including use cases.

iii. Promoted/tested in additional infrastructures not part of the CODATA GOSC or GORC IG work (some of which are listed below in Appendix A).

10. Initial Membership

Co-chairs:

1. Karen Payne <ito-director@oceannetworks.ca>
2. Mark Leggott <mark.leggott@rdc-drc.ca>
3. Andrew Treloar <andrew.treloar@ardc.edu.au>

[Current members](#) represent Europe, the U.S., Canada, Australia, and the UK. It is anticipated that additional membership will include colleagues from organizations that were part of the pre-P17 outreach, as well as members of the GORC IG and CODATA GOSC WG. The CODATA-led GOSC Symposium being planners for September 2021, will also generate additional memberships.

Appendix A: List of Commons

Pan National Commons

1. [European Open Science Cloud](#)
2. [African Open Science Platform](#)
   a. including [H3Africa](#)?
3. [Nordic e-Infrastructure Collaboration](#)
4. [the Arab States Research and Education Network, ASREN](#)

National Commons

European Roadmaps - The European Commission and European Strategy Forum on Research Infrastructures (ESFRI) encourage Member States and Associated Countries to develop national roadmaps for research infrastructures.

1. German National Research Data Infrastructure (NFDI)

2. DANS

3. GAIA-X (non- member state?; see also) (focused on data sharing in the commercial sectors - without excluding research)

4. UK JISC Open Research Framework

Non-European

5. China Science and Technology Cloud (CSTCloud); see also

6. Australian Research Data Commons

7. NDRIO (Canada)

8. NII Research Data Cloud (Japan)

9. KISTI (South Korea)

Domain Commons

10. International Virtual Observatory Alliance (IVOA) (including SKA?)
11. Earth Sciences[2]

    a. DataOne Federation

    b. Federation of Earth Science Information Partners (ESIP)
    c. EarthCube
    d. GEO / GEOSS (GEOSS Requirements lists functionality; GEOSS Common Infrastructure - GCI)
    e. Near-Earth Space Data Infrastructure for e-Science (ESPAS, prototype)
    f. Polar

---

[2] Both the Earth Sciences and Health and Life Sciences groups do not have an overarching governance structure and are not identifiable as a GORC per se. We anticipate creating sub groups that review the interoperability and development plans of these communities.

i. The [Arctic Data Committee landscape map](#) of the Polar Community

ii. [Polar View](#) - The Canadian Polar Data Ecosystem (includes international initiatives, infrastructure and platforms)

iii. [Polar Commons](#) / Polar International Circle (PIC) [not sure if this is active]

iv. [PolarTEP](#)

g. [Infrastructure for the European Network for Earth System Modelling](#) (IS-ENES)

h. [Global Ocean Observing Systems](#) (composed of [Regional Alliances](#))

i. [Global Climate Observing System](#)

j. [CGIAR Platform for Big Data in Agriculture](#)

12. Health and Life Sciences

a. [ELIXIR Bridging Force IG](#) (in the process of being redefined as "Life Science Data Infrastructures IG")

b. [NIH Data Commons](#); [Office of Data Science Strategy](#) (USA)
c. [AIRR Data Commons](#)
d. Global Alliance for Genomics and Health (GA4GH)

13. [Social Sciences & Humanities Open Cloud](#) (SSHOC)

14. Dissco https://www.dissco.eu/ Research infrastructure for natural collections (a commons for specimens and their digital twins)

15. [Datacommons.org](#) - primarily statistics for humanitarian work

Appendix B: Draft List of WG/IG, documents, recommendations, frameworks and roadmaps from related and relevant communities to be reviewed during research phase

1. [RDA Outputs and Recommendations Catalogue](#)
2. [RDA Data publishing workflows](#) ([Zenodo](#))
3. [RDA FAIR Data Maturity Model](#)
4. [RDA 9 functional requirements for data discovery](#)
5. [Repository Platforms for Research Data IG](#)
6. [Metadata Standards Catalog WG](#)
7. [Metadata IG](#)
8. [Brokering IG](#)

9. [Data Fabric IG](#)

10. [Vocabulary Services IG](#)

11. [Repository Platform IG](#)

12. [International Materials Resource Registries WG](#)

13. [RDA Collection of Use Cases](#) [(see also)](#)

14. Existing service catalogues (for example the [eInfra service description template](#) used in the EOSC)

15. [the Open Science Framework](#)

16. [Matrix of use cases and functional requirements for research data repository platforms](#).

17. Activities and recommendations arising from the interdisciplinary [EOSC Enhance](#) program

18. [Scoping the Open Science Infrastructure Landscape in Europe](#)

19. Docs from https://investinopen.org/about/who-we-are/

20. [Monitoring Open Science Implementation in Federal Science-based Departments and Agencies: Metrics and Indicators](#)

21. [Next-generation metrics:Responsible metrics and evaluation for openscience](#). Report of the European Commission Expert Group on Altmetrics ([see also](#))

22. Guidance and recommendations arising from EOSC [FAIR WG](#) and [Sustainability WG](#)

23. Outputs from the [International FAIR Convergence Symposium](#) (Dec 2020) (particularly the session [Mobilizing the Global Open Science Cloud (GOSC) Initiative: Priority, Progress and Partnership](#)

24. The European Strategy Forum on Research Infrastructures ([ESFRI) Landscape Analysis](#) "provides the current context of the most relevant Research Infrastructures that are available to European scientists and to technology developers"

25. [NIH Workshop on Data Metrics](#) (Feb 2020)

26. WMO's Global Basic Observing Network ([GBON](#)) has [internationally agreed metrics to guide investments](#), "using data exchange as a measure of success, and creating local benefits while delivering on a global public good."

27. [Evolving the GEOSS Infrastructure: discussion paper on stakeholders, user scenarios and capabilities](#)

28. There is a [national open access policy in Ethiopia](#) that was released last year, one in the first in Africa to my knowledge. Part of AOSP?

29. [Briefing Note for CODATA Officers:](#) CAS GOSC (Global Open Science Cloud) Project

30. [UNESCO Open Science Recommendation](#)

31. [Open Science in the ISC Science Action Plan](#)

32. [CODATA: Coordinating Global Open Science Commons Initiatives](#)

33. [CODATA: Policies and Interoperability for Global Big Earth Data: a joint CASEarth and CODATA Workshop Session](#)

34. [CODATA: Building a global network infrastructure for international cooperation on data-intensive science](#)

35. Outputs from European Plate Observing System ([EPOS](#)) under ERI (European Research Infrastructure Consortium) upcoming work package: "Strategy for engagement across solid

Earth research infrastructures on a global scale" in the section Key initiative and infrastructure [architecture]
36. A Research Data Infrastructure for Materials Science
37. CeNAT (Costa Rica)
38. Canada's Roadmap for Open Science
39. Are there any ontologies for metrics and measurements we should be aware of?