RESEARCH DATA ALLIANCE

EUROPEAN OPEN SCIENCE CLOUD

A Commons Compendium

Supporting coordination amongst national, pan-national and domain specific research commons.

Summary

The Research Data Alliance (RDA), a global organisation building the social and technical infrastructure to enable open sharing and re-use of data offers an excellent window to the world. It is contributing to the European Open Science Cloud (EOSC) through the effort of its 100+ Working and Interest Groups and is used as a key vehicle in implementing EOSC and supporting international activity. In particular, RDA4EOSC, a tripartite initiative with RDA, the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through an EOSC Secretariat open call, contributes to the internationalisation and implementation of EOSC.

The RDA4EOSC initiative paves the way for future alignment and collaboration between RDA activities and EOSC and lays the foundations for the RDA activities, within the framework of the EU-funded EOSC Future project, aimed at coordinating international research and science commons initiatives.

RDA4EOSC covers various EOSC engagement and international stakeholder and policy maker engagement, standards for EOSC implementation and more. One activity focuses on providing support to the RDA Global Open Research Commons (GORC) Interest Group[1], aiding their work in aligning the EOSC vision and implementation with similar activities worldwide.

Disclaimer

The information shown in the templates was provided to us in April 2021 by ARDC, CESSDA, DataONE, EOSC, GA4GH, IVOA, KISTI, LA Referencia, NeIC, NII.

[1]. https://www.rd-alliance.org/groups/global-open-research-commons-ig

RDA4EOSC

What is a Commons in the research data space?

The Global Open Research Commons Interest Group provides a neutral place for exchanges on Open Science | Open Research Commons and helps to both minimise data silos and to promote the adoption of standards and protocols in facilitating cross-country and cross-discipline global open science commons. The Group focuses on illustrating the common components of a range of national commons with the aim of harmonising on common elements to propose a shared definition for structuring collaborations.

The main goals of the GORC Interest Group are to:

1. Reach a shared understanding of a "Research Commons" in the research data space;
2. Connect with and coordinate international Research Commons and related initiatives;
3. Monitor progress of those Commons in the context of other RDA groups and external agencies.

A Deliberated Definition

The GORC Interest Group has worked carefully on formulating a definition for a Research Commons:

> ***A global trusted ecosystem that provides seamless access to high quality interoperable research outputs and services; digital research resources for the common good.***

The Collaborating Research Commons - an Analysis

This initial compendium includes insights on a series of commons from across the globe outlining their mission statement, funding sources, governance model, service and infrastructure components, standards and future services in planning[2]. As the work of the GORC IG progresses in furthering coordination of global research commons, additional contributions will be added. Representatives are invited to complete and submit the template (https://www.rd-alliance.org/gorc-ig-commons-overview-template) for inclusion in the Compendium.

The Global Research Commons

ARDC (Australian Research Data Commons)
CESSDA ERIC (Consortium of European Social Science Data Archives, European Research Infrastructure Consortium)
DataONE
EOSC (European Open Science Cloud)
GA4GH (Global Alliance for Genomics and Health)
IVOA (International Virtual Observatory Alliance)
KISTI (Korea Institute of Science and Technology Information)
LA Referencia
NeIC (Nordic e-Infrastructure Collaboration)
NII (National Institute of Informatics), Japan

Legend:
National Commons
Pan National Commons
Domain Commons

[2] A total of 10 Commons accepted the GORC IG's invitation and were asked to complete a template outlining their mission statement, funding source, governance model, service and infrastructure components, standards and future services in planning. These templates have been collated to provide this international global commons digest.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

Mission statements / objectives

Given the diversity of the status, contexts, cultures and conditions within which each of these research commons operate, the purpose of this compendium is to showcase the different commons and not to make comparisons or benchmarks. Some highlights of missions and objectives from the commons included in this draft are:

Research Commons in their early stages of development:

"Building a national research platform (DataON) to respond to the global trend." (KISTI)	*"Facilitate data-centric science and open science in Japan."* (NII)

More established commons:

"Providing Australian researchers with a competitive advantage through data. Accelerate research and innovation by driving excellence in the creation, analysis and retention of high-quality data." (ARDC) *collaborations."* (NeIC)	*"Provide a full-scale sustainable research infrastructure enabling the research community to conduct high-quality research in the social sciences contributing to the production of effective solutions to the major challenges facing society today and to facilitate teaching and learning in the social sciences."* (CESSDA)
"NeIC is a global role model for cross-border distributed and sustainable e-infrastructure	

Services planned for next 3 years

The more recently established Research Commons, KISTI are focusing their plans on stabilising a system/service, populating users, building a data market and extending Infra (storage, HW, GPUs) while NII will be facilitating a data governance service, data curation and analysis and RDM training. The more established commons aim to continue the provision of underpinning infrastructure (e.g. cloud compute and data retention) and build a portfolio of thematic Research Data Commons aligned to the needs of specific research domains (ARDC). Other strategic roadmaps for 2021/22 consist of community imperatives concerning cross collaboration with external standards organisations, internal alignment across standards and greater strategic engagement with the community (GA4GH). DataONE will be launching their branded portals and hosted repositories, new services that have been available on a trial basis since 2020 and will be completed with additional features and capabilities.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

How are the Commons funded?

Many of the Commons reported here receive government funding with others benefitting from national funding agency support and university in-kind support. DataONE is currently transitioning to a mixed sustainability model that includes grant support and fee-based services in addition to institutional support while the International Virtual Observatory Alliance (IVOA) does not receive funding with its member activities being funded by their own Virtual Observatory projects and countries.

Standards

Supporting and promoting the FAIR principles is common among the standards highlighted by the various Commons. DataONE identifies its support for a wide variety of open technology standards that are used across the diverse community of repositories including metadata, data representation and ontological standards. Such standards evolve over time as the community of data providers and users adopt new technological approaches. IVOA have 46 standards on interoperability defined in the last 20 years and NeIC complies with the EOSC Rules of Participation - accessible cross-border (in Nordics at least); service provider agrees to have it listed; service delivery is either automated or could be automated.

For more information or to get involved in GORC visit:
https://www.rd-alliance.org/groups/global-open-research-commons-ig.

Membership of the RDA and all its community driven groups is free of charge. Register to become a Research Data Alliance community member at https://rd-alliance.org/user/register.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

ARDC (Australian Research Data Commons)

https://ardc.edu.au/

Rosie Hicks, CEO

1 Overview

1.1 General description

In July 2018, the Australian Research Data Commons (ARDC) was formed under the Federal Government's National Collaborative Research Infrastructure Strategy (NCRIS) through the merger of three existing eResearch infrastructure capabilities: the Australian National Data Service (ANDS), National eResearch Collaboration Tools and Resources (Nectar) and Research Data Services (RDS).

The ARDC was formed to drive the development of the Commons, in collaboration with partners, by focussing on specific strategic themes. These themes were identified during a consultation process undertaken in 2018/19 with various partners, collaborators, thought leaders and the research community:

- Theme 1 – Coordination & Coherence: Facilitating an Australian research data commons
- Theme 2 – People & Policy: Connecting ARDC
- Theme 3 – Data & Services: Maximising the value of Australia's data assets
- Theme 4 – Platforms & Software: Accelerating research insights and supporting collaboration
- Theme 5 – Storage & Compute: Providing foundation infrastructure

The ARDC brings together people, skills, data, related resources (storage, compute, software, models) and partnerships to enable researchers to conduct world class data-intensive research.

1.2 Mission statement, objectives and/or rationale for the services

The purpose of the ARDC is to provide Australian researchers with competitive advantage through data, providing access to leading edge eResearch collections, tools, infrastructure and services. Its mission is to accelerate research and innovation by driving excellence in the creation, analysis and retention of high-quality data assets.

The ARDC is defined by the following principles:

- Transforming research through better tools, by providing better software, platforms and data across the research lifecycle.
- Focusing on national scale opportunities to help develop a nationally coherent eResearch infrastructure environment in a global context.
- People are essential and we will continue to raise awareness of this and support communities in order to build skills and culture in the sector.
- Building strong partnerships and collaborations are at the heart of everything that we do. As one part of a national and international system, we work with others to inform, magnify and sustain common work objectives.
- Be a catalyst for and complement the sector by accelerating innovation through projects, infrastructure, services, consultancy and outreach.
- Commitment to sustainable expertise and services, digital infrastructure for data and tools.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

2 Organisation

2.1 Participating stakeholders and roles

The ARDC is part of a group of five 'Digital Data and eResearch Platforms' (DDeRP) organisations that contribute to the Australian eResearch ecosystem. They include the Australian Access Federation (AAF) providing identity, access and authentication services; Australia's Academic and Research Network (AARNet) providing network infrastructure; and National Computational Infrastructure (NCI) and Pawsey Supercomputing Centre providing high performance and cloud computing resources.

2.2 Funding/income

The ARDC is funded by the Australian Government through the National Collaborative Research Infrastructure Scheme (NCRIS). For the 2019-2023 funding cycle, the ARDC has a budget of approximately AUD $200m.

The ARDC model for co-investing in eResearch infrastructure projects includes both financial and in-kind contributions. This is typically matched at a 1:1 ratio, but this requirement has been relaxed for some programs to accommodate partners' COVID-associated income pressures.

2.3 Organisational and governance model

In May 2019, the ARDC became a company limited by guarantee and is registered with the Australian Charities and Not-for-profits Commission. Currently, the ARDC has 20 members made up of Australian universities and a government research agency (CSIRO).

The ARDC is governed by its constitution. The ARDC governance structure is:
The Board of Directors
The Chief Executive Officer
The Executive Management Team

The Board has two sub-committees (Nominations & Remuneration and Audit & Risk). The CEO is also advised by an independent Research & Technology Advisory Committee.

2.4 Organisational context: related/associated/partner/complementary initiatives

In addition to the DDeRP organisations, the ARDC partners on projects directly with research organisations. Key stakeholders include: peak bodies and funders; research organisations (universities, publicly funded research agencies, Cooperative Research Centres, ARC Centres of Excellence, medical research institutes); commercial and government infrastructure providers; federal and state government agencies.

3 Services

3.1 Services and infrastructure components

Cloud-based compute- Nectar Research Cloud
Cloud-based tools- Platforms hosted on Nectar Research Cloud
Data analysis- Virtual Laboratories and tools
Data discovery- Research Data Australia
Metadata / Vocabulary support- Research Vocabularies Australia

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

Citation support- Identifier Services
Information services (FAIR Principles, Data Management, Sensitive Data, Metadata, Governance)
Policy submissions
eResearch skills support
Data systems architecture and analysis support
Communities of practice

3.2 Future services planned for next 3 years

The ARDC has run a number of open calls to select relevant data infrastructure projects for collaboration. .
The ARDC will build on its deep understanding of data infrastructures and the needs of the Australian research community to drive convergence. It will do so by strategically selecting thematic areas of activity to meet the needs of the greatest number of researchers and ensure that Australian researchers are best positioned to collaborate internationally.

Over the next three years, the ARDC will continue the provision of underpinning infrastructure (e.g. cloud compute and data retention), as well as build a portfolio of thematic Research Data Commons (RDCs) aligned to the needs of specific research domains. This strategy will drive aggregation of research infrastructure needs and ensure greater integration with and between NCRIS facilities. RDC activities will be at sufficient scale to be transformational.

3.3 Standards

The ARDC promotes the FAIR principles, supporting and driving a number of international and national initiatives, as well as finding the right resources for users to ensure they're always using best practice methods within their research.

The ARDC has a number of resources to support the uptake of FAIR principles, including a FAIR data self-assessment tool to assess the 'FAIRness' of a dataset, tips on how to enhance FAIRness and an in depth series of webinars exploring the four FAIR principles.

In addition, the ARDC's current Data Retention Project seeks to increase the impact of investment in underpinning infrastructure that stores important data collections. Partnerships in the Data Retention project will leverage contemporary research data management practices to enrich data collections with controlled and consistent structural metadata to drive the FAIR data principles, particularly the findability, accessibility and reusability of data collections.

The ARDC Identifier Services use the DOI, Handles, IGSN and RaID standards as their basis. The Nectar Research Cloud uses the Openstack open-source codebase. The Research Data Australia data discovery service uses the RIF-CS metadata standard (https://documentation.ands.org.au/display/DOC/About+RIF-CS), based on the ISO2164:2010 information model standard. It is also investigating adding support for schema.org.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

CESSDA (Consortium of European Social Science Data Archives) / European Research Infrastructure Consortium (ERIC)

https://www.cessda.eu

Ivana Ilijasic Versic, Chief Operations Officer

1 Overview

1.1 General description

CESSDA provides large-scale, integrated and sustainable data services to the social sciences. It brings together social science data archives across Europe, with the aim of promoting the results of social science research and supporting national and international research and cooperation.

The CESSDA Strategy 2018-2022 has four pillars:
Building on TRUST
Renown for TRAINING
Proficient in TECHNOLOGY
User-friendly TOOLS & SERVICES.

CESSDA builds trust in social science research by ensuring its quality and that it is available for future research. By acquiring the status of a trust repository, CESSDA Service Providers demonstrate their reliability to researchers as well as national and international research funders.

CESSDA supports continuous learning and training of its Service Provider staff and the social science user community. The areas covered include research data management, data discovery and reuse, digital preservation and data archiving, as well as CESSDA tools and services.

Technology refers to the technical infrastructure behind CESSDA that ensures a stable and up-to-date backbone for our products and services, making it simple and easy to both deposit and access data.

1.2 Mission statement, objectives and/or rationale for the services

The mission of CESSDA is to provide a full scale sustainable research infrastructure enabling the research community to conduct high-quality research in the social sciences contributing to the production of effective solutions to the major challenges facing society today and to facilitate teaching and learning in the social sciences.

Having evolved from a network of European Service Providers into a legal entity and large-scale infrastructure under the auspices of the European Strategy Forum on Research Infrastructures (ESFRI) it became an ERIC (European Research Infrastructure Consortium) on 14 June 2017.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

CESSDA was previously organised as a limited company under Norwegian law. Furthermore, CESSDA was recognised as an ESFRI Landmark in the ESFRI 2016 Roadmap in the field of social and cultural innovation.

2 Organisation

2.1 Participating stakeholders and roles

The CESSDA Consortium is currently composed of 21 member countries and one observer. Several European countries are in the process of becoming a CESSDA member or observer. CESSDA also has partners in a number of countries outside of the consortium.

Within CESSDA, each member country must assign a national Service Provider (or a consortium of service providers). These Service Providers are at the core of CESSDA and cooperate on providing services to researchers.
Being part of CESSDA implies that Service Providers:
Can share their expertise
Be more efficient and provide better services by

Setting up a common technology
Developing and using standards
Eligibility and better access to European funding.

2.2 Funding/income

As the European Research Infrastructure Consortium (ERIC), CESSDA is owned and financed by the individual member states' ministries of research or a delegated institution. Norway is hosting CESSDA, and the main office is located in Bergen. At the moment, there is a very limited amount of IKC ruled by a special agreement with a country involved offering IKC.

2.3 Organisational and governance model

The CESSDA ERIC Consortium is comprised of the General Assembly, the Director, the Scientific Advisory Board and the Service Providers Forum. Its main office is in Bergen, Norway.

2.4 Organisational context: related/associated/partner/complementary initiatives

- other SSH ERICs (ESS, SHARE, DARIAH and ClARIN) together in EOSC related cluster project- SSHOC (https://www.sshopencloud.eu)
- EOSC Association
- ESFRI SWG on SSH
- RDA
- ERIC Forum
-....etc.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

3 Services

3.1 Services and infrastructure components

CESSDA aims to offer a seamless social science data archive service for the whole of the European Research Area (ERA), which is capable of supporting the research needs of the next generation of social scientists wherever in Europe they may be, or beyond.
Find out more:
https://www.cessda.eu/Tools-Services
https://datacatalogue.cessda.eu
https://www.cessda.eu/Training

3.2 Future services planned for next 3 years

- upgrade of CDC
- Metadata Office project (on metadata standards)
- upgrade of training services, international training associations engagement
- global outreach
- etc.

3.3 Standards

CESSDA Data Access Policy
https://doi.org/10.5281/zenodo.4054793

CESSDA ERIC Persistent Identifier Policy
https://doi.org/10.5281/zenodo.3611327

CESSDA Publication Policy & Procedures
https://doi.org/10.5281/zenodo.3904264

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

DataONE (Data Observation Network for Earth)

https://www.dataone.org/

Amber Budden, Director for Community Engagement and Outreach

1 Overview

1.1 General description

DataONE was founded in 2009 and the infrastructure formally launched in 2012, and it extended and was informed by earlier projects focused on federating data across the environmental sciences such as the Knowledge Network for Biocomplexity and the Science Environment for Ecological Knowledge. DataONE simplifies discovery, access and reuse of data by providing a federated network of data repositories that serve specific research communities. DataONE develops software to support interoperability and collaborates on initiatives in service of open data science. Additional services include the ability to host a data repository through DataONE and to create custom portals featuring curated collections of data from across the full DataONE network. These portals include comprehensive project information, data usage metrics and metadata assessments in addition to other information. DataONE also provides data management training and learning resources, and supports community collaboration and education through community calls and webinars.

1.2 Mission statement, objectives and/or rationale for the services

DataONE is designed to enable new science and knowledge creation through universal access to data about life on earth and the environment that sustains it. DataONE is a community driven program providing access to data across multiple member repositories, supporting enhanced search and discovery of Earth and environmental data. DataONE promotes best practices in data management through responsive educational resources and materials. DataONE envisions researchers, educators, and the public using DataONE to better understand and conserve life on earth and the environment that sustains it.

2 Organisation

2.1 Participating stakeholders and roles

DataONE is a distributed network with many partner organizations and individuals. Administration of DataONE occurs via the National Center for Ecological Analysis and Synthesis at the University of California Santa Barbara. Two of the current management team are based there. Other management team members are based at the University of Tennessee Knoxville, University of Kansas, University of New Mexico and North Carolina State University. Participating repositories are located across the United States and internationally.

2.2 Funding/income

DataONE was previously funded via two NSF DataNet awards, totalling $35M across 10 years, as well as in-kind support from universities (UC Santa Barbara, University of New Mexico, University of Kansas, and University of Tennessee) and agencies (e.g., US Geological Survey). DataONE is currently transitioning to a mixed sustainability model that includes grant support and fee-based services in addition to institutional support. DataONE is built using open source software and multiple other projects utilize the same software, leveraging investments made and contributing back to the infrastructure.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

2.3 Organisational and governance model

During the period of NSF funding DataONE comprised a leadership team, external advisory board and user's group. As part of our current transition we are also undertaking a change in governance structure. We now comprise a smaller, agile management team and are supporting a DataONE Community that informs decision making and represents the larger, engaged DataONE community.

2.4 Organisational context: related/associated/partner/complementary initiatives

DataONE, as a federated network, engages with many dozens of data repositories (see each of the member repositories in the network (https://www.dataone.org/network/#list-of-member-repositories) and related curatorial groups. These include academic institutions, local, state, and federal government agencies, and other non-governmental organizations. In addition, we interact extensively on interoperability and community engagement with data community groups, including ESIP, RDA, EarthCube, ROpenSci, Make Data Count, COUNTER, Whole Tale, and others.

3 Services

3.1 Services and infrastructure components

DataONE provides three primary services. 1) Easy search, discovery and access to data held within member repositories through a single, integrated discovery platform that harmonizes metadata across the network. 2) Branded data portals enabling a custom view of your research topic or lab data that spans repositories in the DataONE network. 3) Hosted repositories, where DataONE will manage and preserve your data, software and derived products within a branded repository built with our hardware and software. Each of these services comprise features developed by DataONE in collaboration with partners, including data usage metrics and associated reporting, semantic annotation and metadata quality checks among others.

3.2 Future services planned for next 3 years

The branded portals and hosted repositories are new services that are available since 2020 on a trial basis and will be officially launched within the next year. Additional features and capabilities are planned for these services.

3.3 Standards

DataONE supports a wide variety of open technology standards that are used across the diverse community of repositories. These include metadata standards (such as the ISO 19115 family, EML, and schema.org), data representation standards (such as HDF5 and netCDF), ontological standards for provenance and semantics (such as PROV-O, ProvONE, ENVO, and ECSO), and a variety of data services (such as the DataONE web services, OGC Catalog Services for the Web, and the ORCID user identity services). The dozens of standards evolve over time as the community of data providers and users adopt new technological approaches. The DataONE infrastructure is designed and implemented to support and promote the FAIR data principles.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

EOSC (European Open Science Cloud)

https://eosc.eu/
https://eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf

Karel Luyben, President EOSC Association

1 Overview

1.1 General description

The scope is the complete spectrum of Open Science.

1.2 Mission statement, objectives and/or rationale for the services

Please refer to page 58 of the SRIA for the Objectives tree of the EOSC ecosystem.
The European Open Science Cloud was launched to service around 2 million researchers in Europe.

2 Organisation

2.1 Participating stakeholders and roles

Please see the website for the members: https://eosc.eu/members
The Organisation will be in the form of a co-programmed partnership between the European Union and the EOSC Association.

2.2 Funding/income

Funding for the coming 7 years (Horizon Europe programme) will be roughly EURO 500 M for the European Union in cash and some EURO 500 M from the participants in kind.

2.3 Organisational and governance model

Please refer to page 53 of the SRIA.

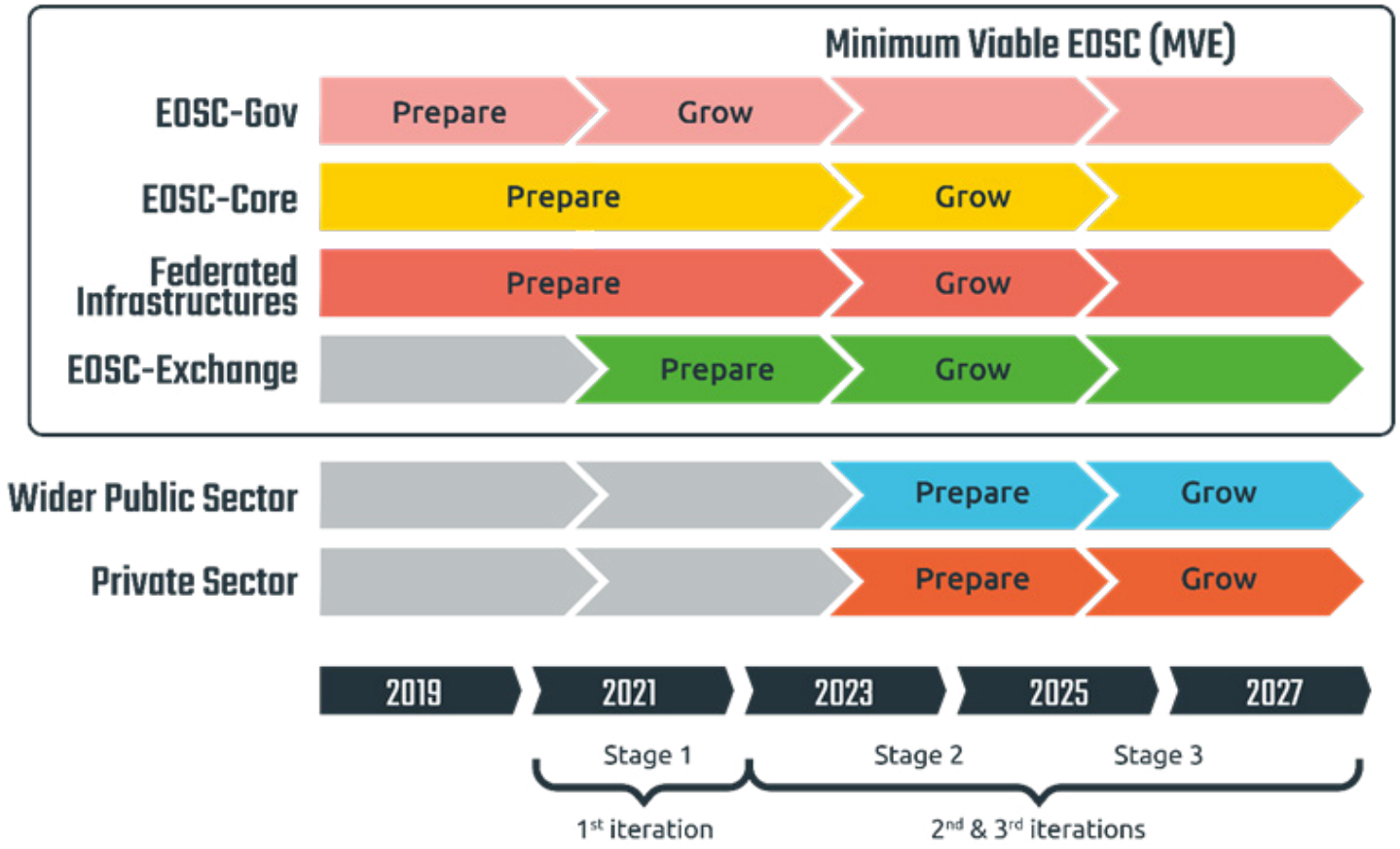
2.4 Organisational context: related/associated/partner/complementary initiatives

In principle all Research Performing, Research Funding and Research Service Providing Organisations in Europe. Thus also all other partnerships active in Horizon Europe.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

3 Services

3.1 Services and infrastructure components



3.2 Future services planned for next 3 years

See above figure.

3.3 Standards

A set of five guiding principles will help position EOSC within the Horizon Europe programme during the next seven years. These are:

- Multi-stakeholderism – EOSC will succeed if and only if it follows a multi-stakeholder approach;
- Openness – EOSC will ensure that research artefacts are 'as open as possible, as closed as necessary';
- FAIR principles – EOSC will assemble research artefacts that are findable, accessible, interoperable and reusable;
- Federation of infrastructures – EOSC will federate existing and upcoming research infrastructures (data- and e-infrastructures);
- Machine-actionable – EOSC will strike the right balance between machines and people in delivering the services that will serve the needs of European scientists.

See Section 4 of the SRIA

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

GA4GH (Global Alliance for Genomics and Health)

https://www.ga4gh.org/
Angela Page, Communications Manager

1 Overview

1.1 General description

GA4GH is an international, nonprofit alliance formed in 2013 to accelerate the potential of research and medicine to advance human health. Bringing together 600+ leading organizations working in healthcare, research, patient advocacy, life science, and information technology, the GA4GH community is working together to create frameworks and standards to enable the responsible, voluntary, and secure sharing of genomic and health-related data.

1.2 Mission statement, objectives and/or rationale for the services

GA4GH aims to accelerate progress in genomic research and human health by cultivating a common framework of standards and harmonized approaches for effective and responsible genomic and health-related data sharing.

2 Organisation

2.1 Participating stakeholders and roles

GA4GH consists of 24 Driver Projects (www.ga4gh.org/how-we-work/driver-projects/) — real-world genomic data initiatives that help guide our development efforts and pilot our tools. Stakeholders around the globe, including but not limited to our 650+ member organizations (www.ga4gh.org/community/organizational-members), advocate, mandate, implement, and use our frameworks and standards in their local contexts. Four GA4GH Host Institutions (EMBL-EBI, Broad Institute, Ontario Institute for Cancer Research, and Wellcome Sanger Institute) provide in-kind support by provisioning staff to the GA4GH Secretariat team.

2.2 Funding/income

GA4GH receives a total of $3M USD per year in direct funding from the US (NIH), UK (MRC, NIHR), Canada (CIHR, Genome Canada), and Wellcome.

2.3 Organisational and governance model

GA4GH is a collaborative, global alliance led by a four person Executive Leadership Team and Standards Steering Committee which consists of 24 Driver Project Champions and 16 Work Stream Leads. We receive guidance from a Strategic Advisory Board and operational support from a distributed staff team (currently 12 members). Details at www.ga4gh.org/about-us/governance-and-leadership-2/.

2.4 Organisational context: related/associated/partner/complementary initiatives

GA4GH interfaces with several additional large consortia, including the International Hundred-thousand-plus Cohorts Consortium (IHCC), the International Common Disease Alliance (ICDA), Health Data Research UK / International COVID-19 Data Alliance (ICODA), the Human Pangenome Research Consortium, and the Public

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

Health Alliance for Genomic Epidemiology (PHA4GE).

3 Services

3.1 Services and infrastructure components

GA4GH technical standards and policy frameworks are developed by a collaborative, cross-sector community of more than 200 active contributors. These individuals populate Eight Work Streams, which are the primary GA4GH production teams. Each Work Stream tackles a specific area of the genomic data lifecycle: Clinically & Phenotypic Data Capture & Exchange, Cloud, Data Security, Discovery, Genomic Knowledge Standards, Large Scale Genomics, and Regulatory & Ethics. Collectively, these groups have produced nearly 20 technical standards and policy frameworks to help support responsible, international genomic data sharing. Driver Projects are required to dedicate at least two full time equivalents across at least two GA4GH Work Streams. For more details on the GA4GH matrix organizational structure, as well as details on each of the Driver Projects and Work Streams. Please visit www.ga4gh.org/how-we-work/.

3.2 Future services planned for next 3 years

The 2021/22 GA4GH strategic roadmap consists of three community imperatives (cross collaboration with external standards organizations, internal alignment across GA4GH standards, and greater strategic engagement with the clinical community), as well as 30+ new or expanded technical standards and policy frameworks. More details can be found at www.ga4gh.org/roadmap.

3.3 Standards

All GA4GH work builds on the foundational principles and core elements laid out in the Responsible Data Sharing of the Framework for Responsible Sharing of Genomic and Health-Related Data (www.ga4gh.org/framework), published by the Regulatory & Ethics Work Stream in 2014. Each technical standard is subject to rigorous product approval process, which includes detailed review by a Product Review Committee, external to the development team, as well as review by both the Regulatory & Ethics and Data Security Work Streams. Wherever appropriate, GA4GH aims to align with existing standards used by the community and has developed strategic relationships with ISO, HL7, CDISC, and others.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

IVOA (International Virtual Observatory Alliance)

http://www.ivoa.net
Chenzhou Cui, Executive Commitee Chair

1 Overview

1.1 General description

The Virtual Observatory (VO) is a collection of interoperating data archives and software tools that facilitate astronomical research. The overall goal is to support innovative research in astronomy by exploiting the full power of growing and emerging datasets and interoperable services.

The International Virtual Observatory Alliance1 (IVOA) is an organisation that debates and agrees on the technical standards that are needed to make the VO possible. The IVOA was formed in June 2002 with a mission to "facilitate the international coordination and collaboration necessary for the development and deployment of the tools, systems and organizational structures necessary to enable the international utilization of astronomical archives as an integrated and interoperating virtual observatory."

1.2 Mission statement, objectives and/or rationale for the services

The work of the IVOA focuses on the development of standards. Working Groups are constituted with cross-project membership in those areas where key interoperability standards and technologies have to be defined and agreed upon. The Working Groups develop standards using a process modeled after the World Wide Web Consortium, in which Working Drafts progress to Proposed Recommendations and finally to Recommendations.

2 Organisation

2.1 Participating stakeholders and roles

The IVOA now comprises 20 VO programs from Argentina, Armenia, Australia, Brazil, Canada, Chile, China, Europe, France, Germany, Hungary, India, Italy, Japan, Russia, South Africa, Spain, Ukraine, the United Kingdom, and the United States and an inter-governmental organization (ESA). Membership is open to other national and international projects according to the IVOA Guidelines for Participation.

2.2 Funding/income

There is no funding for the Alliance. Activities of the IVOA members are funded by their own VO projects and countries.

2.3 Organisational and governance model

The IVOA work is pursued by Working Groups (WG) and Interest Groups (IG), coordinated by the Technical Coordination Group (TCG), guided by a scientific priorities committee (CSP), with the overall direction provided by the IVOA Executive Committee.
Senior representatives from each national VO project form the IVOA Executive Committee. A chair is chosen from among the representatives and serves a one-and-half year term, preceded by a one-and-half term as

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

deputy chair. The Executive Committee meets 3-4 times a year to discuss goals, priorities, and strategies. The IVOA holds two Interoperability Workshops each year: a week-long meeting in spring, typically May, and a shorter meeting in fall that is either coordinated with the annual ADASS conference or with a regional VO project meeting. These meetings are opportunities for the Working Groups and Interest Groups to have face-to-face discussions and for the more difficult technical questions to be resolved.

2.4 Organisational context: related/associated/partner/complementary initiatives

IVOA is evolving as an organisation in a rapidly changing landscape, with the emergence of new large initiatives for research data sharing such as RDA with strong support of the FAIR principles. Astronomy is also changing as we enter an era of very large data, and multi-wavelength and multi-messenger astrophysics where there is an essential need for high level interoperability of data, simulations, tools and services.

3 Services

3.1 Services and infrastructure components

Major IVOA accomplishments include standards for data and metadata (Data Models), data exchange methods (Data Access Layer; Query Language), and a registry that lists available services and identifies what can be done with them.

3.2 Future services planned for next 3 years

Recent IVOA activities have focused on the engagement with future large data producing projects, and priorities have most recently focused on multi-dimensional data, and time domain astronomy.

The roadmap for development activities by the various IVOA working and interest groups in 2020/2021 are outlined at: https://wiki.ivoa.net/twiki/bin/view/IVOA/2020BRoadmap

3.3 Standards

About 46 standards on interoperability have been defined during the last 2 decades. Detailed list is available at: https://ivoa.net/documents/index.html

The work of the IVOA is all aimed at enabling new and innovative science. The IVOA standards, defined by the community, for the community, form an important part of the astronomy data infrastructure. Astronomy is rapidly entering into a new era of Big Data, and interoperability is increasingly important for multi-wavelength, multi-messenger and time domain astronomy. One of the common themes that is being driven by the needs of these big data projects is the concept of science analysis platforms that will enable analysis of the data with capabilities for providing computational resources close to the data.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

KISTI (Korea Institute of Science and Technology Information)

https://dataon.kisti.re.kr
Seongchan Kim, Senior Researcher

1 Overview

1.1 General description

In 2018, a pilot service was designed and implemented, and a research data platform (DataON) was built in 2019. Last year (2020) DataON had a major update. DataON services include the research data registration, search, link(OAI-PMH, OpenAPI), community, personal cloud storage, and analysis workspace.

1.2 Mission statement, objectives and/or rationale for the services

Recently, recognizing the importance of research data, countries around the world have been developing strategies to manage and utilize research data. Korea is also building a national research data platform(DataON) to respond to the global trend. DataOn is expected to be used for users to improve the result of their research.

2 Organisation

2.1 Participating stakeholders and roles

KISTI is managing and developing DataON

2.2 Funding/income

About one million dollars (USD) per year.

2.3 Organisational and governance model

There is an operation team with 6 members. The team has four sub-roles: portal management, data link management, analysis service management, and infra-management(SW/HW). Also, we have a development team of 5 members covering portal, data link, analysis service, and infra.

2.4 Organisational context: related/associated/partner/complementary initiatives

3 Services

3.1 Services and infrastructure components

not defined

3.2 Future services planned for next 3 years

Stabilizing System/Service
Populating Users
Building Data Market
Extending Infra(Storage, HW, GPUs etc.)

3.3 Standards

not defined

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

LA Referencia

https://www.lareferencia.info/en/
https://github.com/lareferencia
https://guidelines.openaire.eu/en/latest/
Bianca Amaro, President

1 Overview

1.1 General description

LA Referencia was born through the Cooperation Agreement, signed in Buenos Aires in 2012, which reflects the political will to offer in open access the scientific production of Latin America as a regional public good with emphasis on the results financed with public funds.

1.2 Mission statement, objectives and/or rationale for the services

The Federated Network of Institutional Repositories of Scientific Publications, or simply LA Referencia, is a Latin American network of open access repositories. Through its services, it supports national Open Science strategies in Latin America through a platform with interoperability standards, sharing and giving visibility to the scientific production generated in institutions of higher education and scientific research.

2 Organisation

2.1 Participating stakeholders and roles

This experience is based on technical and organizational agreements between public science and technology agencies (Ministries and S&T Organisms) of the member countries, together with RedCLARA.
From the national nodes, scientific articles, doctoral and master's theses are integrated, coming from more than a hundred universities and research institutions from the ten countries that now form LA Referencia. Argentina, Brazil, Chile, Costa Rica, Colombia, España, Ecuador, El Salvador, Panamá, Perú and Uruguay.

2.2 Funding/income

Each country pays an annual membership of USD 14,500

2.3 Organisational and governance model

LA Referencia is governed by a Council integrated by country representatives, each representative has a vote, and the decisions are taken by single majority. LA Referencia council elects a president as political representative in international forums. Bianca Amaro from Ibict Brasil is the president of LA Referencia (2020-2022)

2.4 Organisational context: related/associated/partner/complementary initiatives

RedCLARA, COAR, OpenAIRE, Zenodo

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

3 Services

3.1 Services and infrastructure components

- LA Referencia Harvester / Validator / Enricher / Indexing- Open Source Platform
- Regional Harvester (based on LA Referencia software)
- Regional Search Portal (based on LA Referencia software)
- National Aggregator / National Search Portal (based on LA Referencia software)
- Usage statistics pilot

3.2 Future services planned for next 3 years

- Common regional Usage Statistics
- CRIS Systems metadata integration
- Data Repositories harvesting
- Dashboard for repository managers

3.3 Standards

OpenAIRE Metadata Guidelines https://guidelines.openaire.eu/en/latest/
OAI-PMH Protocol

All services are based on open source components developed by LA Referencia technical team in collaboration with national nodes teams, or in third party open source projects (such as Vufind)

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

NeIC (Nordic e-Infrastructure Collaboration)

https://neic.no/

Monica Lassi, Executive Manager

1 Overview

1.1 General description

NeIC was established in the key area of e-infrastructure in 2012, and its sustainability was consolidated through a Memorandum of Understanding between the five national research councils and NordForsk. NeIC has two major roles. The first role is to assume operational responsibility for the Nordic distributed Tier-1 facility that is part of the Worldwide LHC Computing Grid (WLCG) that provides computing and storage for CERN and is used by high energy physicists worldwide. The second is to jointly or collaboratively explore, evaluate, develop and deploy innovative infrastructure services in response to the strategic priorities in the area of e-infrastructure and the needs of the national e-infrastructure providers, their users and selected ESFRI projects of joint Nordic interest. NeIC e-infrastructure projects are collaborations on the building blocks for the future capabilities and NeIC is NordForsk's main tool to implement the Nordic eScience Action Plan 2.0. NeIC is at an intersection point between national, Nordic and international strategies. NeIC has worked to increase collaborations through a set of projects that include the national providers (CSC, DeIC, RHnet, SNIC, UNINETT Sigma2 and ETAIS), partners in ESFRI projects, and other institutions in the Nordics and Europe. NeIC projects are in place in the areas of Physics and Engineering Sciences, Environmental Sciences, Humanities, Culture and Society, Life Science and e-Sciences.

1.2 Mission statement, objectives and/or rationale for the services

Our vision: NeIC is a global role model for cross-border distributed and sustainable e-infrastructure collaborations.

Strategic objectives:

- Beneficial collaborations are the principal way that NeIC brings together the needs, interests and resources to create e-infrastructure to support research excellence in the Nordic region.
- Nordic Influence is the effect that NeIC has in improving and advancing e-infrastructure for researchers and for society.
- Motivated People are essential in the collaborations and to bring the results into the research domains and society.
- Effective processes bind NeIC into an organization that is able to realize the benefits and influences that come from the collaborations.

Rationale for NeIC's launch: The first role is to assume operational responsibility for the Nordic distributed Tier-1 facility that is part of the Worldwide LHC Computing Grid (WLCG) that provides computing and storage for CERN and is used by high energy physicists worldwide. The second is to jointly or collaboratively explore, evaluate, develop and deploy innovative infrastructure services in response to the strategic priorities in the area of e-infrastructure and the needs of the national e-infrastructure providers, their users and selected ESFRI projects of joint Nordic interest.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

2 Organisation

2.1 Participating stakeholders and roles

NeIC is a joint initiative between the Nordic countries, hosted by NordForsk since 1. January 2012. NordForsk appoints the NeIC Board based on nominations by the national e-infrastructure provider organisations. These strategic partner organisation are CSC (Finland), SNIC (Sweden), UNINETT Sigma2 (Norway), DeIC (Denmark), RH Net (Iceland) and ETAIS (Estonia). The NeIC Board consists of one representative from each of these.

The NordForsk Board has delegated to the NeIC Board the authority to make strategic decisions regarding computing and data-storage infrastructure and react on upcoming opportunities. This includes allocating budget, implementing the organisation structure as well as prioritising and coordinating Nordic collaboration projects. The NeIC Board develops and maintains a strategy for NeIC. The NeIC Board recommends the NeIC Director to be appointed by NordForsk.
NeIC is managed by an Executive Team chaired by the NeIC Director. Most of NeIC's activities are organized in projects. The Executive Team coordinates the activities and participate in project steering groups as project owners.
In collaboration with national eInfrastructure providers and user-community representatives, NeIC engages IT experts to participate in projects and operational activities. Projects are developed in accordance with the directions set out in the Nordic eScience Action Plan 2.0 (2015).

2.2 Funding/income

Funding of NeIC's activities is provided through national funding agencies, NordForsk and participating project partners.

2.3 Organisational and governance model

Day-to-day management of NeIC is handled by the Executive Team (XT), chaired by the Director. Functions of the XT include strategy implementation, project ownership, project portfolio management, committee facilitation, NeIC office management, facilitating the work of the NeIC board, NeIC commons. Project ownership includes steering group representation. Project portfolio management includes supporting project managers and stimulating mutual learning and improvement.

Functions of the Project Managers: Project management (team leadership), facilitation of project steering group meetings, engaging with project's stakeholders and beneficiaries, contribution to NeIC commons.
The 'NeIC commons' is the wide range of common functions and services that keep our organisation flourishing. Some of these are offered through dedicated roles but several others need to be contributed by team members at large. NeIC staff are prepared to volunteer a fraction (say, 10%) of their available time to the NeIC commons in agreement with their team leader.

2.4 Organisational context: related/associated/partner/complementary initiatives

Related/associated/partner/complementary initiatives.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

3 Services

3.1 Services and infrastructure components

NeIC has only one infrastructure component with a long-term mandate. The majority of NeIC's work are developing services during projects, of which results are maintained by other stakeholders after the project.

NORDIC WLCG TIER-1 FACILITY

The Nordic distributed tier-1 facility for the worldwide computing grid serving the large hadron collider at CERN. NDGF Tier1 is one of 14 regional computing centres of the Worldwide LHC Computing Grid – the huge international e-infrastructure built to provide computing and storage for the CERN. This Nordic solution is unique in being distributed across four countries: Denmark, Finland, Norway and Sweden.

3.2 Future services planned for next 3 years

Not defined.

3.3 Standards

A Nordic e-Infrastructure Service is a service fulfilling the following criteria:

1. Complies with the EOSC Rules of Participation;
2. Can be accessed cross-border (in the Nordics at least);
3. Service provider, e.g. National e-infrastructure, agrees to have it listed;
4. Service delivery is either automated or could be automated.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

NII (National Institute of Informatics), Japan

https://rcos.nii.ac.jp/en/

Kazu Yamaji, Center Director

1 Overview

1.1 General description

NII Research Data Cloud
Start development in 2017
Start operation in 2021

NII Research Data Cloud (NII RDC) aims to support the whole research data life cycle by the different components.
NII RDC, at this moment, is composed of a Research Data Management (RDM) platform, publication (repository) platform, and discovery platform.

1.2 Mission statement, objectives and/or rationale for the services

The mission of the NII RDC is to facilitate data-centric science and open science in Japan.
NII RDC first appeared in the document entitled “Recommendations Concerning an Approach to Open Science that Will Contributes to Open Innovation” (http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t230-en.pdf), which was published by Science Council of Japan.

2 Organisation

2.1 Participating stakeholders and roles

NII is developing and operating the services. JAIRO Cloud (Japanese Institutional Repositories Online: a shared repository service provided by NII since April 2012) has the joint operation structure with JPCOAR (Japan Consortium for Open Access Repositories), each of which has the following responsibilities:
NII: Software development, Service operation.
JPCOAR: Training course, Needs survey, Initial test of the new function.

2.2 Funding/income

The success model is to receive initial development and incubation budget from the government. Once we reach the critical mass, we start to discuss the business model with the user institution and decide the price model. Until that phase, we request government support for operational costs in addition to the R&D costs.

2.3 Organisational and governance model

NII has an operational committee for NII RDC. The committee is composed of external members from universities and research institutions. For instance, service operation policy is resolved by the committee.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.

2.4 Organisational context: related/associated/partner/complementary initiatives

AXIES (Academic eXchange for Information Environment and Strategy: https://en.axies.jp/) is the IT center, and JPCOAR is the library community in Japan. They are the main partners of our service development and operation. They have correspondence WG to our service and then support our activities.

3 Services

3.1 Services and infrastructure components

RDM Platform: GakuNin RDM
Publication Platform: JAIRO Cloud (WEKO3)
Discovery Platform: CiNii Research
are running based on
Identity Management Federation: GakuNin
Cloud Security Checklist Service: GakuNin Cloud
Network: SINET5

3.2 Future services planned for next 3 years

* Data Governance Service (DMP based service composer)
* Data Analysis Service
* Data Curation Service
* RDM Training Service

3.3 Standards

* SAML: Federated access management.
* OAuth2: Service integration at RDM service.
* Time Stamping Protocol: Proof of data existence at RDM service.
* DMP Common Standard: DMP based service composer (under development).
* BagIt: Contents packaging at curation service (under development).
* OAI-ORE: Contents packaging at curation service (under development).
* SWORD: Contents deposit from the RDM service to a repository (publication platform).
* Open Graph Protocol: Page summary generation at third party service.
* ResourceSync: Data provision from the repository to the discovery.
* OAI-PMH: Data aggregation from the repository to the discovery.

This document was produced under RDA4EOSC, a tripartite initiative between the Research Data Alliance (RDA), the Digital Repository of Ireland (DRI) and the Digital Curation Centre (DCC), funded through the EOSC Secretariat to contribute to the internationalisation and implementation of EOSC.