The Global Open Science Cloud Landscape

Date: October 2021

Editors

- Yin Chen, EGI Foundation, Netherlands
- Jianhui Li, CNIC, Chinese Academy of Sciences, China
- Simon Hodson, CODATA, EU
- Mark Dietrich, EGI Foundation, Netherlands
- Lili Zhang, CNIC, Chinese Academy of Sciences, China
- Tiziana Ferrari, EGI Foundation, Netherlands
- Eric Graventein, EGI Foundation – Graphic Designer, Netherlands

Authors

- Ane Persic, United nations Educational Scientific and Cultural Organisation, United Nations
- Geoffrey Boulton, International Science Council, EU
- Hussein Sherief, AASCTC, Africa
- Sarah Jones, EOSC Executive Board, EU
- Corina Pascu, European Commission, EU
- Ingemar Häggström, EISCAT Science Association, Swidden
- Xinna Yue, Institute of Geology and Geophysics, China
- Junyi Wang, Institute of Geology and Geophysics, China
- Eric Yen, Academia Sinica Grid Computing Centre, Taipei
- Chenzhou Cui, National Astronomical Observatories Chinese Academy of Science, China
- Guido Lemoine, European Commission, EU
- Kerstin Lehnert, Columbia University, US
- Happy Sithole, National Integrated Cyber-Infrastructure System, South Africa
- Rosie Hicks, Australian Research data Commons, Australia
- Natalia Manola, OpenAIRE, EU
- Erik Huizer, GÉANT, Netherlands
- Yan Zhuang, Chinese Academy of Science, China
- Kostas Glinos, European Commision, EU
- Manish Parashar, National Science Foundation, US

Date: October 2021

Abstract

We are at the threshold of an era with unprecedented opportunities for cross-disciplinary and cross-border research collaborations. Open Science that started years ago and is growing in momentum with the digitalization and many big investments such as those made by the European Commission. Open Science Cloud, Platform, Commons are developed at national and regional level. It is very timely for developing Global Open Science Cloud (GOSC) that aims to connect the different international, national and regional research infrastructures to create a global digital environment for borderless research and innovation. Fundamentally, it aims to make it easy for scientists to access cross-continent digital resources in order to resolve global challenges such as problems caused by the COVID19, and those identified in the UN sustainable Development Goals.

This report discusses the concept and landscape of GOSC. It is a major outcome from a workshop during EGI Conference 2020 that gathered global digital infrastructures, scientific communities, policy makers and funders to review the existing work, examine the available resources, and identify collaboration opportunities. This report is among the initial publications that document GOSC discussions with the aim to inform future developments.

Contents

Executive summary

The mission of GOSC is to connect different international, national and regional open science clouds and platforms to create a global digital environment for borderless research and innovation. It aims to provide better ways to harness digital resources from around the world, help bridge the division in infrastructure, technique and capacity building among different countries, support global science collaborations and foster truly international science.

This report collects the discussions and findings from the 1st Global Open Science Cloud (GOSC), organised during the EGI conference 2020. Each chapter focuses on one special aspect including, preliminary policy landscape, requirements of international science communities, the state-of-the-art of global digital infrastructures, and funding opportunities.

Chapter 1. GOSC, the Concept and the Preliminary Landscape.

It surveys the existing global policy landscape, including the UNESCO Recommendations on Open Science, the current International Science Council Action Plans, CODATA Decadal Program. It reviews the Open Science development in Africa, Latin American, Southeast Asia, and Europe (including European Open Science Cloud and RDA Global Open Research Common). It also introduces the GOSC initiative project funded by the Chinese Academy of Science.

There is common recognition that the **development of the Global Open Science Cloud is both timely and important.** Dramatic efforts have been made by various parties and communities, and key policies and technologies are in place. **It is seen as a long-term mission**, since it has been **pushing for a fundamental change in research culture**. The impacts and involvements it is making are much deeper and wider.

The GOSC will focus on digital infrastructure connectivity, aiming to join the existing regional platforms. Fundamentally, it pursues to ease international scientific collaborations **to deliver excellent research that may lead to resolve global challenges** such as problems caused by the COVID19, and those identified in the UN sustainable Development Goals.

Around that is a major opportunity for international cooperation and alignment, and actively inviting collaboration from a large number of stakeholders and organizations globally. The discussions will contribute to the developing countries, involve 'Global South' and initiatives in Africa and Latin America such as La Referencia. It will obviously do that in partnership and collaboration with global organizations like UNESCO, ISC, CODATA, RDA, GO FAIR, as well as others.

Chapter 2. Co-Design of GOSC with Research Communities.

It examines international science communities' needs from different scientific domains, including Photon and Neutron, Space Physics, Disaster Mitigation, Astronomy, Earth Observatory, and Geoinformatics. It invites world-class science communities to introduce the state of the art developments in the fields with the focus on global collaborations and expectations for GOSC.

The common understanding is that the growth of global research collaborations is speeding up in every domain field. Science has no borders. Scientists need data from other countries, regions and continents to develop more comprehensive models that lead to improved understanding and new discoveries. **Desire to access data and resources across regions and continents become more urgent.**

GOSC can contribute by providing e-Infrastructure service solutions, e.g. moving large data across continents, providing resources and distributed data processing solutions, etc.

facilitating best practices exchange, sharing of experiences and technology, avoiding duplicating efforts, and providing sustainable operation solutions. The development of GOSC infrastructure can start with interfacing into existing community applications, extending domain science platforms to be portable to GOSC, and further making them generic to support wider use.

Chapter 3. Globale e-Infrastructure: Challenges and Opportunities in Achieving the GOSC Vision.

It invites the advanced global digital infrastructures, including National Integrated Cyber Infrastructure System (NICIS) in South Africa, ARDC and Nectar in Australia, CSTCloud in China, OpenAIRE, EGI and GEANT in Europe, in order to share experiences and advice for GOSC development.

The observation is, **there is high technical readiness for digital infrastructures development at regional and continental level, and there is an incontestable interest in joining together to form a world-wide connected computing infrastructure.**

Technical challenges are commonly recognised. **Agreeable approach to enable easier access to compute at a global level is needed, that implies unified AAI, shared standards to enable interoperability, interoperation and cooperative service delivery and operation, etc. National production should be made internationally relevant and best practices and experiences should be shared,** i.e. how to move computing to data, how to handle sensitive data, how to provide integration access to HTC, HPC, and Cloud, etc. There are other aspects beyond technical issues, such as **governance, policies, organization of community support activities, Sustainable funding,** etc.

Chapter 4. Funding Agencies' Perspectives for GOSC.

It interviews international funding agencies, including Chinese Academy of Science, European Commission and US National Science Foundation, in order to gain a broader perspective on funding opportunities for GOSC.

The key messages are **there is a global effort around Open Science and Open Science Cloud or Open Science Commons. The policy makers and the funding agencies are confirming that there is a strong support of policies towards Open Science. The importance is widely recognized and major funding and initiatives are already happening.** For example, in the last 3 years, the European Commission has invested 250 M Euro in developing and prototyping the European Open Science Cloud, and will intensify the efforts in the new Horizon Europe Program. On the other hand, the Chinese Academy of Science has sustained the investments in the GOSC: in the past 2 years, seed money was provided for the initial setup of collaborative workshops, conferences and training. In the second stage, the funding is provided for pilot programs and the formulation of international projects that may lead to future international 'big science' projects such as SKA. From 2021 China will issue the 14th five-year plan on Science, Technology and Innovation. In the next five years, science, technology and innovation will be more emphasized and more resources and funding will flow into the communities in these areas.

The common vision is that science is global, and it is needed to enable cooperation in science and research. The principle on which to pursue international activities with partners is mutual interest, similar values (in terms of the research integrity), and ethics (to pursue collaborative activities with a Level Playing Field, so nobody has an unfair advantage on the basis of reciprocity).

In **Conclusion and Future Work**, it summarises the report and informs the activities that happened after. In particular, the launch of the CODATA GOSC Working Groups and Case Studies on 28th Jun 2021, that aim to address a number of challenges shared by Open Science Clouds, including:

- Governance and Sustainability: sharing experiences and good practice.
- Policy and Legal Dimensions: aligning principles, policies and exploring legal issues.
- Technical Infrastructure: including network connectivity and protocols, secure Authentication and Authorization Infrastructure (AAI), mechanisms for federation of computing , data and other services.
- Data Interoperability: including identifiers, semantic services, rigorous contextual and provenance metadata, analytical tools and virtual research environments.

In parallel, Case Studies of domain and cross-domain research will be developed to provide a concrete grounding and demonstrator of topics identified above. Four Case Studies are currently planned: Incoherent scatter radar data fusion and computation, Biodiversity and ecology information platform, SDG-13 climate change and natural disasters, and Sensitive data federation analysis model in population health.

As a final remark, **we would like to invite global Open Science infrastructures, international research communities, and research organisations to join together to build this new GOSC community.**

Introduction

The digital revolution has transformed the way in which data, information and knowledge are acquired, managed, analysed, repurposed, used and disseminated. We are at the threshold of an era with unprecedented opportunities for cross-disciplinary and cross-border collaboration for research and innovation. A new research paradigm is emerging which applies increasingly automated approaches and machine learning, and which harnesses the most advanced computing facilities and software, to handle huge and disparate cross-disciplinary data. The advanced infrastructure needed for this new paradigm and Open Science is rising up: it needs to be on demand, as a service, ubiquitous and seamless. In pursuit of this vision, digital infrastructures are developed at institutional, national and regional levels, such as the show cases in European Open Science Cloud (EOSC)[1], the Chinese Science and Technology Cloud (CSTCloud)[2], the Australia Research Data Commons (ARDC)[3], Malaysia Open Science Platform[4], the African Open Science Platform[5], etc.

Is it possible to share experiences and make a global framework to align and federate such Open Science clouds and platforms? Is there a way to better support research collaborations across continents to resolve global science challenges, such as the U.N. Sustainable Development Goals (SDGs), climate change, infectious diseases and pandemics, COVID19, coordinated and global disaster risk reduction, and so on? At the moment, a global fully connected digital infrastructure is not in place, making it difficult for scientists to access digital resources across regions and continents. The idea of a Global Open Science Cloud (GOSC) was agreed during the CODATA 2019 Beijing conference. **The mission of GOSC is to connect different international, national and regional open science clouds and platforms to create a global digital environment for borderless research and innovation.** It aims to provide better ways to harness digital resources from around the world, help bridge the division in infrastructure, technique and capacity building among different countries, support global science collaborations and foster truly international science.There are many challenges and difficulties, i.e., inconsistent access policies from country to country, lack of common standards for building a global-level data and e-infrastructure federation, differences in language and culture, varied funding models, etc. In order to investigate these issues and in order to develop consensus about initial concept and framework for GOSC, a workshop[6] was jointly organised by EGI, CNIC and CODATA on 4th-5th Nov 2020, that gathered representatives of international initiatives, policy makers, research communities, public digital infrastructure providers, and international funding agencies.

The workshop reviewed the existing work in global collaborations, discussed the needs and typical use cases from the research community, examined available resources and possible contributions from international e-infrastructure providers, identified the key barriers in policy, governance, standard and technique, and identified possible funding opportunities.This report presents the major outcomes from the workshop. It is a collection of the speeches from all presenters who are domain experts with a broad understanding of global collaborations in their special areas. Together they brought pieces from each field to form a unified picture -- the current landscape of Global Open Science Cloud, that reveals insights why it is important and timely development for GOSC and to inform future activities.

Chapter 1. GOSC, the Concept and the Preliminary Landscape

What could be a Global Open Science Cloud? Open Science is growing in momentum. It's becoming very much part of the way of doing science. Arguably it has been for some time in principle, but now with the digital age, we are seeing, in particular countries and regions' efforts and investments to support Open Science, in particular domains, and to some extent across domains. We're witnessing major investments in Open Science Clouds, platforms, and commons, and these initiatives have at least two motivations -- to achieve, from the funders and the institutional perspective, economies of scale, and to break down silos between domain infrastructures and to encourage collaboration across domains.The Global Open Science Cloud and the proposed collaboration aims to encourage dialogue and cooperation between such Initiatives, where alignment and interoperability are possible. The interoperability can take place on a number of levels, on the policy level, on the infrastructure and the technical level, and on the data levels (e.g., to do with the semantics and the metadata). This section explores some of these issues.

1.1 The UNESCO Recommendation on Open Science

UNESCO is the U.N. agency in charge of science together with education, culture and communication. The U.N. has been promoting science, technology and innovation for the benefits of society for peace and sustainable development. There is a huge interest in science technology innovation as it is one of the key means of implementation for achieving all the 17 Sustainable Development Goals (SDGs)[7]. It has realized the importance of having timely and free access to scientific data and publications, the importance of scientific collaborations and sharing of information at all different levels. This growing importance of science-policy-society dialogue is the key to deal with some of the issues that we have today e.g., the mistrust in science, fake news, etc. The COVID19 situation has brought up some of these issues even further. There is an acknowledgement, particularly on the policy side within member states and governments, a need for better access to science, and to open science more broadly.UNESCO's view on Open Science is, "a movement to transform and democratize the entire scientific process to ensure that science truly drives and enables the achievement of the United Nations Sustainable Development Goals for the benefits of all". Open Science is a game changer, an accelerator. It doesn't only concern the output of science, such as dataset and publications, but also the principles of openness.

The openness shall happen at every step of the process in a way that should be just and equitable. On the other hand, Open Science helps achieve the human rights to science which is included in the declaration of human rights. This hasn't been talked about a lot in the past, but is coming more and more strongly, particularly, in the context of COVID19. The human rights to science and to make the society benefit from science is what Open Science can promote.Open Science has been happening for a while, and moved forward quickly with digitalization and a few big initiatives including those from the European Commission. For UNESCO, the question is no longer whether it's happening, rather how it's going to happen, and how everybody can contribute and benefit from the interpretation. The real concern is that Open Science would make even bigger gaps between those who have and those who have not -- while the idea originally behind Open Science is to reduce these gaps between developed and developing countries, and between those who have and have not within countries themselves. UNESCO's mission is to make sure the transition to Open Science is as equitable as possible, and contribute to reducing the gaps in science technology and innovation across nations.In November 2019, an Advisory

Committee was established consisting of global consultations, regional consultations, stakeholder consultations, and started a process of developing a global standard for Open Science. The standard is in the form of a recommendation[8], and was developed in a conservative manner. The effort has been made to gather existing inputs[9] (i.e., open data, open infrastructures, open access, open to society, citizen science etc.) and put them under the same umbrella of a global recommendation. With the support of a global community from 193 member states, the first draft of the recommendation[10] was produced and sent to the member states in September 2020. The aim is to get feedback from individuals and communities as much as possible, and to finalize and be approved by member states by November 2021.The draft recommendation defines the concept of Open Science, the different elements, guiding principles, etc. Specially, there are recommendations on Open Science infrastructures, and the key points are as follows:

- There should be community-led, interoperable, inclusive, interconnected and sustainable digital infrastructures and computing facilities to ensure long-term preservation, stewardship, and community control of research products.
- Community agreements should define community practices for data sharing, data formats, metadata standards, ontologies and terminologies, tools and infrastructure.
- Technical requirements for every digital object of significance for science, whether a datum, a dataset, metadata, code, or a publication should also be addressed.
- Certain core specifications, such as for example the FAIR and CARE principles for data stewardship.
- Joint strategies for shared, multinational, regional Open Science platforms to provide coordinated support for interoperability from the perspective of policy, practices and technical specifications.
- Platforms for exchanges and co-creation of knowledge between scientists and society.
- Monitoring and information systems to complement national, regional and global data and information systems.
- Capacity building at all possible levels to allow any projects on Open Science to include this feature.

1.2 International Science Council Action Plan and Open Science

The International Science Council (ISC) has been in existence (in one form or another) for over a hundred years. It tends to act as a global voice for science and has been an originator of many of the major international collaborative enterprises in science. It has around 200 members comprising national academies, regional science bodies, research councils, and the international scientific unions and associations. It has regional offices in Africa, Asia, Latin America and Caribbean (and regional social science councils in Africa, Latin America, Asia and Arab world).ISC has a very large global scientific network consisting of the ISC Science Support Bodies[11] and the ISC International Scientific Committees[12]. The ISC International Research Programs covers large variety of domain research fields (e.g., CROP (Comparative Research on Poverty), IRDR (Integrated Research on Disaster Risk), UHWB (Urban Health and Wellbeing), Future Earth, and WCRP (Climate Research Programme); the ISC Global Observing Systems (e.g., Global Climate Observing System (GCOS), and the Global Ocean Observing System (GOOS)).

ISC also has its own funding programs which are largely funded by external bodies but funneled through ISC in order to appropriately and effectively allocate those funds.The current (2019-2021) action plan of ISC[13] includes four major domains: 1) the 2030 Agenda for Sustainable Development; 2) the Digital Revolution; 3) Science in Policy and Public Action; and 4) the Evolution of Science and Science Systems. These action plans are regarded as relatively robust in a sense that they're likely to be important for the next few years. Open Science is a pervasive perspective underlies these domains of areas. A number of projects which are very explicitly to do with Open Science including, the digital planet for sustainability, data-driven interdisciplinarity (led by CODATA on behalf of ISC), global data resource and governance, Open Science Systems with a particular focus on Open Science in the Global South, etc. It's been recognized that the biggest and most urgent challenge for contemporary science is to identify pathways to global sustainability, and to assist in the creation and promotion of policies and public action that can advance societies along those pathways.ISC's viewpoints for effective delivery of global collaboration on Open Science are as follows:

- As an open collaborative enterprise, Open Science must be operationally connected and globally inclusive.
- It should be socially engaged, and it should be discovery and action-oriented.
- Open access to the record of science should be made available (not only the recent publication, but also those which have been published since the beginning of the scientific enterprise), also the access to infrastructure.
- There is a need to change funding policies and practices, redesigning assessment and reward systems to point our scientists towards Open Science rather than away from.
- Challenges also include how to bridge persistent global knowledge divides, and evolve into an enabling global commons.

1.3 The CODATA Decadal Program: Making Data Work for Cross Domain Grand Challenges

CODATA was invited to prepare a program that fits under the ISC action plan and addresses the challenges and opportunities emerging from the digital age. CODATA's vision for that -- "the major, pressing global scientific and human issues of the 21st century (i.e., climate change, mitigation and adaptation, the growth of cities and the resilience of cities, disaster risk reduction, meeting the 2030 agenda in various domains, etc.) can only be addressed through research that works across disciplines to understand complex systems, using a transdisciplinary approach to turn data into knowledge, and then into action". It requires researchers' ability to create data, relating them to those challenges and to combine data from many different sources, in order to understand, model and interpret those complex systems.

CODATA has been working for two years on the design of a decadal program named 'Making Data Work for Cross Domain Challenges', seeking to assist and support ISC' global coordinating research programs by addressing the data issues which are fundamental to those challenges. The program that can be divided into three areas of activity:

- Consensus and technical solutions for data interoperability (terminologies, ontologies, metadata, machine learning);
- Mobilising domains and breaking down silos (working with Unions, Associations and other domain organizations);
- Advancing solutions through cross-domain case studies.

The strategy is to work with existing domains on the data specifications and standards, at the same time, encourage particular domains through collaboration with international scientific unions to address issues of interoperability and data combination. Then to relate those advances to case studies in cross-domain research areas, e.g., resilient cities on infectious diseases, and an overarching activity of technical and consensus building for interoperability.The decadal program will be officially launched at the ISC General Assembly and the associated events, "Global Knowledge Forum", in Oman, 10-14 Oct 2021[14]. CODATA is playing a coordinating role, and they would want to get all organizations involved, to participate in and feel ownership of it, in particular, the Research Data Alliance (RDA), the World Data System, and other organizations.

During CODATA Beijing conference 2019[15], the idea of Global Open Science Cloud has been proposed and a list of priority tasks were identified as follows:

- FAIR object ecosystem (FAIR Digital Objects, PIDs/GUPRIs, Types)
- Core Community Resources: units, semantic resources, metadata schema
- Core community Resources: repositories/data resources, data stewardship services, analytical tools
- Community data sharing agreements (scientific responsibility, as open as possible, proportionate protections, embargoes)

- Open Science and FAIR Policies (institutions, funders)
- Infrastructure (network, compute, storage)
- Skills and Training: data science
- Skills and Training: data stewardship
- Business Models and Sustainability
- Governance and Rules of Engagement

The Beijing declaration is an indication of areas that need cooperation, alignment and interoperability. CODATA has set up an internal working group to prepare a draft project plan for a funded project which will be led by CNIC in China but with international cooperation and coordination.

1.4 Development in Africa, Latin American and Southeast Asia

The African Open Science Platform[16] is one of the major adventures which has been developing for several years. It's based on having effective hard and soft infrastructures federated across African countries and coordinated in an effective fashion, and issues of outreach and engagement, transdisciplinary action, education, and capacity building are crucial parts of it.The platform was designed as a hub and node structure. The hub will be located in Pretoria, South Africa. The nodes are essentially located in five regions of Africa: Southern, Eastern, Central, Western, and Northern. Crucial principle is that it's Africa-led. However, as their funding has dried up, many initiatives and internationally funded initiatives have drained into the sand.

This is largely because there's been no infrastructural or institutional base on which to ensure sustainability of the objective. Africa has major structural problems, e.g., low investment in science, poor electronic and transport connectivity, low intra-African collaboration such that African scientists collaborate more with European and North American scientists, but they cannot collaborate with each other because of few critical masses.In Sudan and Ethiopia , Africa Arab Science and Technology Cloud (AAcTCloud) funding has been applied for. Implementation is planned for end of 2021 or early 2022. The implementation will be done by AASCTC in collaboration with CNIC CAS.Spontaneously, there have been approaches from the very well-developed Latin American systems, particularly in relation to publication based on La Referencia[17] and RedCLARA[18] which have been developing now for 20 years. That discussion has been focused on collaboration in delivering good shared practice, a network for collaboration, equitable access publishing systems, etc.There are incipient moves in Southeast Asia to develop a Voice for the Global South in a way that science internationally has not heard before. Those South discussions are enormously important. One of the difficulties of a modern world is that although we would like there to be an international science community, it is largely a community dominated by the Global North. The major publications are all located in Europe and North America, and

they disadvantage the South quite dramatically. Developing the Voice for the Global South is crucial -- fundamentally, we cannot solve global problems unless we have global engagement. We should ensure that one part of the crucial global engagement takes place.

1.5 Development in Europe: the European Open Science Cloud

As an example of an Open Research Commons, the European Open Science Cloud is trying to develop a platform for European research which has been described most recently in the strategic research and innovation agenda, as the web of FAIR data and related services. The aim is to federate the different e-infrastructures and research infrastructures, and provide an environment in which the data and the relevant services can be brought together to allow researchers from different backgrounds to conduct research that address the societal challenges.In 2018, a governance structure for EOSC was established. There is a Governance Board consisting of representatives from the Member States and the European Commission; An Executive Board consisting of the EU stakeholder organisations and individual experts; A Stakeholder Forum to involve users, service providers, Public sector, Industry, SMEs, etc.; A series of working groups that gather a broad coalition of the doers to steer the discussion and implementation. At the moment there are working groups focusing on the EOSC landscape, the sustainability models, rules of participation, technical architecture, FAIR data, and the skills & training.Many work has been delivered by the working groups including[19], for example,

- The analysis of the existing national infrastructures, policies, the landscape and readiness;
- Agreed and tested Rules of Participation;
- Study of the financing model and the establishment of a legal entity, the EOSC association;
- Exploration of the post 2020 governance and the funding structures;
- Development of the EOSC architecture, the interoperability framework, and the development of Persistent Identifier (PID) architecture and policy;
- The metrics for FAIR data and certified services; and
- The skills curricula and training resource catalogue.

EOSC is part of a wider European Data Space Agenda which addresses different kinds of thematic areas including, industrial and manufacturing, the green deal, mobility, health, financial, energy, agricultural, and public administrations. In this context, EOSC is seen as horizontal cross cutting all of those aspects, playing a role to enable interoperability, creating community standards to make sure the data are understandable, to promote discovery across different research domains, and where possible to allow coordination and combination of different data sets.

The goal is to bring together disconnected silos in a federated infrastructure so that to make it easier for researchers to discover and combine data, and deliver their research.EOSC has actively used the Research Data Alliance (RDA) as a global forum, which is a rich forum to see what's happening elsewhere, to copy what works in other countries, and to learn from other people's mistakes. It's a good way to initiate a sounding board for discussing new ideas, and to co-develop pilots internationally that may lead to much more robust and internationally applicable solutions. There's a lot of groups that are looking at relevant aspects such as, PID policy, metadata standards for discovery, etc. In particular, the RDA Global Open Research Commons Interest Group was set up to provide a neutral space where people can have conversations about Open Science Commons and to define a shared understanding of the comments -- what are the building blocks, what are the common elements, and where are the opportunities to collaborate.

1.6 Coordination of global activities on the development of Open Science platforms – the RDA Global Open Research Commons Interest Group

The RDA Global Open Research Commons (GORC) interest group[20] was set up in 2019. It has a number of co-chairs and members coming from many continents, regions and countries, and many initiatives from various domains or disciplines.What the interest group does is to get an understanding of what the common is in the research data space. Initially, the discussion will focus on the data commons, then will be extended to the Open Science Commons.

As the work would progress in the coming years, the group would also proactively look outside of the RDA community to connect with other initiatives either national or regional or international initiatives such as, the G7 Open Science working group that was set up in 2016, co-chaired by the European Commission and by the Cabinet Office of Japan[21], and has members from all the G7 members. It's very actively working on an international policy alignment with Open Science, particularly into two areas[22]: 1) public access including incentives and rewards, and 2) infrastructures for Open Science.The group aims to coordinate a delivery of Global Open Research Commons, and monitors the progress made within/across other RDA working groups that are related to their work. Essentially they need to understand how to connect the national based data commons with regional or domain or discipline based initiatives. This is a very huge endeavor.A number of events have been organized since 2018, i.e. a workshop on towards a Global Open Science Commons[23] in the 11th RDA Plenary, March 2018; a session on delivering a Global Open Science Commons[24] in the SciDataCon conference, November 2018; Coordinating Global Open Science Commons workshop[25] in the 13th RDA Plenary, March 2019; Launch of the interest group[26] in the RDA 14th Plenary, October 2019; a session in the CODATA conference in Beijing, September 2019.The group planned activities include

- To reach a shared understanding of what a 'Common' is in the research data space, and propose a general description of an open research commons;
- Identify common elements that can be used to propose a typology of 'Commons' to provide a framework to make infrastructures more interoperable. Initially, the typology considers several dimensions, i.e., governance rules, engagement data services, tools and infrastructures, business models and sustainability, incentive skills, etc.;
- Validation of the definition and typology; and
- Identify concrete use cases for implementation.

At the moment, the group is working on International Benchmarking[27] for global data commons, that will evaluate and recommend a minimal set of international benchmarks for global commons that allow researchers and developers to coordinate services and create roadmaps for international interoperability.

1.7 The GOSC Initiative

In recent years, many national regional open science infrastructure platform or cloud have been developed including, European Open Science Cloud (EOSC), African Open Science Platform (AOSP), Canadian New Digital Research Infrastructure Organisation (NDRIO)[28], U.S. National Research Cloud[29], China Science & Technology Cloud (CSTCloud), National Research Infrastructure for Australia (NCIRS)[30], etc.During the CODATA 2019 conference in Beijing, the idea of GOSC was born to coordinate the different national regional platforms and to make a globalized platform for open science and open data. Endorsed by CODATA, the Globe Open Science Cloud initiative, successfully received funding from the Chinese Academy of Science (CAS). Led by the Computer Network Information Center (CNIC)[31]of CAS, the project will run from 2021-2025.There are several key actions for the project, for example,

- Establish the international communication platform, and interconnect national and regional level infrastructures
- Harmonize policies for cross-border sharing of data, resources and services
- Explore governance rules and sustainable operation mechanism
- Implement key protocols & tool kits for interoperability, and to provide secure and transparent services
- Establish testbeds, involving international scientific communities i.e. SKA, SDG, EISCAT, WeNMR, and deliver demonstrations, for example, show cases of cross-continental large-scale data transmission, distributed simulation and process, collaborative machine-learning and federated learning, cooperational Cloud provision, access to e-Infrastructure platforms across nations and regions, etc.

The anticipated outputs include:

- Consensus and global collaboration network on Open Science Cloud
- The framework for GOSC policy and techniques

- Cross-border testbed to demonstrate for global science collaboration and open science.
- Towards the launch of a Big Science Program for GOSC in 2025

The GOSC initiative project promotes a long-term process driven by selected use cases from international scientific communities. Starting from a lightweight testbed and working towards a large-scale converged platform. It follows an iterative development and continuous evolution cycle, using co-design, co-develop and co-operate approaches, that would need to involve different stakeholders (digital infrastructures, technical experts, data professionals, policy makers and researchers, research communities, etc.). The project will actively involve international partners and communities.

1.8 Summary

This section surveys the existing landscape, looks at current developments of Open Science, and introduces the GOSC initiative. The emergence of the Global Open Science Cloud is not a surprise. It rides the wave of Open Science that started years ago and is growing in momentum with the digitalization and many big investments such as those made by the European Commission. The importance of Open Science is endorsed by the UN, governments, policy makers, and funding agencies. It is seen as a long-term mission, since it has been pushing for a fundamental change in research culture. The impacts and involvements it is making are much deeper and wider -- reaching down to individual citizen scientists and up to cross-region and continent collaborations. Such phenomena rarely happened before.It is very timely for developing the Global Open Science Cloud. With dramatic efforts from various parties and communities, many important pieces begin to appear. For example, UNESCO is developing consensus globally and developing engagement from the member states, and the recommendation for Open Science is about ready by the end of 2021; ISC fits the Open Science in its current Action Plan, and invites CODATA to design the data decadal program which will be launched in October 2021; The European Open Science Cloud (EOSC) contributes its successful experience in building the European urban science cloud. The RDA as a platform and community, particularly the GORC Interest Group (IG) is actively engaging global communities and the Global Open Research Commons is on the way. The GOSC initiative just started and sustains its funding for 5 years more of development.Revisiting the question what the Global Open Science Cloud is? In the existing landscape, the GOSC can position itself under the umbrellas of Open Science, be seen as an implementation of recommendations and global commons that are being developed, i.e., by UNESCO and RDA GORC IG. In turn, it will contribute to the validation of these standards with practical use cases and pilot implementations. The GOSC is a response to the calls for global science collaboration such as those made by ISC and CODATA, and serves as a doer to push the realisation of those strategies and plans. The GOSC will obviously leverage the successful experience of Europe with the European Open Science Cloud and can be included as a part of EOSC global extension to share the experiences with other countries and regions.

The GOSC community will largely benefit from the RDA platform and use it as one of the major venues for conversations and alignment. The GOSC will focus on digital infrastructure connectivity, aiming to join the existing regional platforms, to form a global highway that supports data transmission across-continent and regions. It will co-federate the digital resources, co-operate the provision, and co-deliver services, so that to support the needs of international scientific research for accessing digital resources at the global-level, collaborative data analysing and federated data processing. Fundamentally, it pursues to ease international scientific collaborations to deliver excellent research that may lead to resolve global challenges such as problems caused by the COVID19, and those identified in the UN sustainable Development Goals. Around that is a major opportunity for international cooperation and alignment, and actively inviting collaboration from a large number of stakeholders and organizations globally. The discussions will contribute to the developing countries and involve 'Global South', initiatives in Africa and

Latin America where there's initiatives such as La Referencia with which we will need to engage. The GOSC will obviously do that in partnership and collaboration with global organizations like UNESCO, ISC, CODATA, RDA, GO FAIR, as well as others.

Chapter 2. Co-Design of GOSC with Research Communities

What are the research communities' perspectives for GOSC? Community representatives from different scientific domains are invited including, photon and neutron, space physics, disaster mitigation, astronomy, earth observatory, and geoinformatics. They introduced the existing developments in the fields, specially focused on their global collaborations and expectations for GOSC.

2.1 Global Science Collaboration in the Photon and Neutron Community

Photon and Neutron (PaN) Research Infrastructures (RIs) are analytical facilities used for understanding the structure and functioning of matter.

PaN RIs include low energy storage rings such as, ALBA[32] in Spain, SOLEIL[33] in France and Diamond[34] in the UK and operate with electron beams with an energy between 3~4 GeV. The energy range of the electrons determines the energy of the X-ray beams used for probing matter. The High Energy Storage Rings are APS[35] in Chicago USA., ESRF[36] in Grenoble France, DESY[37] in Hamburg Germany, and SPRING8[38] in Sayo, Japan. These facilities have electron beams in the energy range of 6~8 GeV and can produce hard X-rays that can penetrate matter much more.

There are also Free Electron Lasers, large linear accelerators where the beam is pulsed and generates extremely intense light flashes. Very often, with such extremely intense beams the sample itself will be destroyed during the experiment; data is taken just before the sample disintegrates. .Photons and Neutrons are complementary and frequently used together to study samples. . There are four major Neutron facilities in the world: the SNS[39] in Oak Ridge USA, the ILL[40] in Grenoble France, the ESS[41] under construction in Sweden, and CARR[42] at CIAE in China.

PaN RIs are user facilities. Every year thousands of scientists come to the RIs with their samples. Typically, these researchers stay for a short time, e.g. three to seven days, to conduct experiments. Thereafter, they usually go home with terabytes of data, performing analysis with additional data, then writing publications. The user community of PaN RIs consists of ~40,000 scientists in Europe alone with an annual turnover of 20-25%. Thousands of publications annually originate from experiments with PaN facilities. The research spans all fields of matter and structure of matter and addresses many social challenges in health, energy, and environment.

PaN RIs create large and complex datasets which are not easy for scientists to analyse. Each year, PaN RIs generate many petabytes of data and the data volume is increasing rapidly causing a Big Data challenge. Data needs to be compressed and reduced in real-time using sophisticated algorithms. Large datasets (up to 100s of TBs) are difficult to transport and sometimes the best solution is to analyse data on-site. These lead to new challenging requirements, e.g., how to support complex data analysis software; optimization on multi CPU/GPU systems; batch processing; HPC/HTC environment; machine learning which is increasingly important for dealing with large datasets, etc.

Two European projects are aiming at connecting the PaN RIs to the European Open Science Cloud (EOSC), namely PaNOSC[43] and ExPaNDS[44]. These two projects are among the ESFRI science cluster projects[45] who have overall the same ambition to connect to the EOSC and bring data and user communities to make EOSC a reality and foster Open Science. A consensus has been reached in the PaN RI community to apply FAIR principles, to manage data in an organized manner, and profit from best practices and experience.

Through these joint activities, the priorities for the PaN RI community are:

- PaN RIs must implement data compression algorithms (specific to each scientific domain or method); data analysis software; and Machine Learning for automation and data analysis.
- PaN Users need to get easy access to open e-infrastructures (network, storage, CPU/GPU) that enable the re-use of open data.

2.2 Global EISCAT and Potential Collaboration with SYISR

Incoherent Scatter Radar (ISR) is the most powerful ground based instrument for ionosphere monitoring, with large aperture and high transmit power (1MW-10MW) it's capable to measure the parameters of the ionosphere (60~2000km) such as, density composition, velocity, temperature etc. EISCAT[46] runs 3 of the world's total 10 ISRs, and is located at the North pole region, supporting research for ionosphere, space debris, aurora, etc. EISCAT_3D is the next generation of ISR using phased array radar technology that will enable new capabilities such as, volumetric imaging and tracking, interferometric imaging, multistatic configuration, improved sensitivity, and transmitter flexibility. Each full array of EISCAT_3D will consist of 10,000 crossed dipole antennas and each of these antennas will produce high volume data that need to be handled at the local site and also between sites. The 5 sites, the phased-array antenna fields, are located in the northernmost areas of Norway, Finland and Sweden, connected with high speed network, terabyte links obtained by collaboration between the Nordic universities. Once in operation, EISCAT-3D will need very large compute and storage resources, e.g. during the High Power mode (10% of the time) with >100 Tflop/s and >20PB will be required. This will be upscaled every ~5 years as funding permits. During the Low Power mode or Off periods, the facility will be open for user processing. The EISCAT-3D ICT system will provide an application platform for analysis software, a co-location platform for sophisticated data discovery, and a Notebook application for algorithm development. The system will be hosted at a EISCAT high performance computer center, and it also plans to integrate with the EGI federation as processing mirror sites for users with extra needs and their own federated resources. Storage and network are needed for continuous data transfer between the sites and also for EISCAT Asian users (e.g., China, Japan, South Korea).

A number of applications have been developed:

- A data portal[47] is developed using DIRAC technology which provides a data catalogue for lower level observation data. EGI Check-In is integrated for authentication and authorization of accessing embargoed data. (There are embargo rules for data access).
- A metadata catalogue[48] is based on B2SHARE that enables searching for the High Level data which is the data products with physical parameters.
- A Madrigal-based distributed database[49] was built together with U.S. partners about 15 years ago. It's a database for higher level data supporting searches for physical parameters and combined parameters. It allows data downloading or plotting. It also contains Chinese radar data.

On the other side of the continent, SYISR radar is under construction at Sanya, China. Funded by National Natural Science Foundation of China (NSFC), SYISR[50] is a modular active electronically scanned phased-array with all-solid state transmitter and all-digital receiver. There will be two other receivers over the high islands to be built between 2020 and 2024. In the end there will be three receivers and one transmitter incoherent scatter radar over the low digital channel, that is called SYISR Tristatic System (TS).

SYISR radar and EISCAT_3D share many similarities. Both measure the same atmosphere layer (the ionosphere), use the similar Phased Array Radar technology, and the same algorithm in digital signal processing and data retrieval. The EISCAT_3D is the first multiple station Incoherent Scatter Radar (ISR) in the world, and the SYISR TS will be the only multiple station ISR in the equatorial region. On the other hand, the two radar systems are complementary to each other: their geographic locations are different, as the result their scientific focuses are very different. For example, the EISCAT_3D mainly focus on the high latitude ionosphere such as the precipitation, the aurora etc.; and the SYISR TS will fix on the low latitude ionosphere -- EISCAT_3D measures at the origin of ionospheric weather, and the SYISR TS will measure the result of space weather.

There are potential collaborations between EISCAT_3D and SYISR TS:

- Metadata catalogue: allows researchers to search both EISCAT-3D and SYISR TS data, which requests to federate metadata and access the repositories from both systems;
- Data processing: allows researchers to process data, which requests to enable the data processing on the federated resources, i.e., one can launch jobs in EGI FedCloud or CSTCloud depending on the location of data;
- Data transfer: allows users to move data to one location and process the integration data.

The main challenge is the big data issue, for example, for the Phased-Array radar, the original data will be gigabytes in minutes that needs to be processed to be higher level data that are only gigabytes in a year. How to process and store the big data, how to transfer and share the big data between China and Europe, and how to tackle data interoperability generated by different sites? It needs technical support for online data curation, online radar control, online data processing and visualization, etc.

2.3 Regional Collaboration on Disaster Mitigation

The original collaboration on e-science for Asian disaster mitigation started from the APAN[51] project in 2008. Because Asia is the most disaster-prone region in the world, and together with 10 Asia countries, the Asian Mitigation Community Center has been created and focusing on capacity development for disaster mitigation based on a Deeper Understanding approach which aims to discover the root cause and physical processes of targeted events, develop accurate simulations of the whole life cycle of target events, and to quantify the risks.

Since the EGI-Engage project in 2015, a disputed cloud infrastructure has been built based on EGI and EOSC-Hub technologies. A Case Study Simulation Portal and a Knowledge Base constitutes the framework for disaster mitigation capacity development. Case studies drive the advancement of science technology and the Simulation Portal applications. Collaboration partners and users are able to reproduce the hazard events through numerical simulations, and carry out risk analysis. All the observation data (i.e., the parameters, simulation algorithm, event facts, publications, etc.) are compiled into the Knowledge Base. Based on lessons learned from case studies, a user is able to make use of the numerical simulations for better risk analysis of other events and conducting their local case studies.

So far, 18 case studies of six types of hazards in almost every partner country have been conducted in the Simulation Portal, and 13 of them have been completed. The research covers tsunami, typhoon & storm surge, dust transportation, forest fire/smoke/haze monitoring and lightning. For each type of hazard, the leading scientist or science group had been assigned to conduct the Deeper Understanding of events and develop and validate the simulation processes. Now, partners such as Philippines, Thailand and Vietnam, could conduct their own case studies and contribute to the regional collaboration framework. During the annual conferences and meetings, all partners are encouraged to bring their case studies and users to work together, and to share their requirements, case study, experiences etc. The original collaboration on disaster mitigation keeps moving forward, with more case studies and partners, new types of hazards and application fields, and IT technologies. There are more regional collaborations from different institutes and organizations such as, UNESCO, SA, EGI, ISGC[52], Remote Sensing community, Asian Center for Water Research, ASEAN community[53], etc. In the future, more case studies will be supported by collaboration projects with EGI and capacitive building is needed for dealing with more complex multi-hazard scenarios. The collaboration is extending to agricultural areas, and the space-based resource federations. Systematically, efficiency, Authentication Authorization Infrastructure, advanced data analysis pipeline, GPU/Container/Jupiter over cloud are the key requirements.

2.4 Virtual Observatory and Science Platforms in Astronomy

Astronomy is a Big Data science research field. Nowadays, there are many global telescopes that generate large scale observation data. Driven by these facilities, astronomy remains at the leading place in Big Data science.

About 20 years ago, the community initiated the idea of Virtual Observatory (VO) to satisfy the Big Data challenges. As the web is 'transparent', and the goal of the Virtual Observatory is to achieve the same feeling for astronomical data that it is all available to explore in a single transparent system. The VO allows astronomers to interrogate multiple data centers in a seamless and transparent way, provide new powerful analysis and visualization tools within that system, and gives data centers a standard framework for publishing and delivering services using their data. Like the World Wide Web, the VO is not a fixed system but rather a way of doing things.

Created in 2002, the International Virtual Observatory Alliance (IVOA)[54] was initiated which is an organization that debates and agrees on the technical standards that are needed to make the VO possible, a focal point for VO aspirations, and a framework for discussing and sharing VO ideas and technology. Currently, the IVOA consists of 21 national and international members.

During the last two decades, the VO community designed and developed dozens of interoperable applications and services, and now the VO framework has been a daily tool for astronomers around the world. VO tools are also embedded in all different kinds of astronomy services such as the European Southern Observatory (ESO) Science Portal, WWT, EAS Sky, Grav waves --the telescope of microsoft research, the CDS reference data service, the Firefly Caltech-IPAC, SVO Filter Profile service, etc.

The amount and complexity of astronomy data will increase greatly in the near future, and many VO science platforms are being developed for researchers to efficiently analyze big data sets. These science platforms will enable analysis close to the data, support online data mining and machine learning. Most science platforms in astronomy employ a similar architecture and technologies to provide interactive data analyzed environments, for example, a cloud-based platform, and JupyterHub and JupyterLab are used as an interface for exploratory data mining and analysis. Such an interactive environment is normally deployed using container technology (e.g., docker).

As an example, the Astro Data Lab[55] is operated by the National Optical Astronomy Observatory (NOAO), the national center for ground-based nighttime astronomy in the United States. Astro Data Lab provides a large catalog (a TB-scale database), pixel data (images & spectro in NOAO Science Archive), a virtual storage (1TB per user to minimise data transfer), visualization (for data exploration), compute processing (workflows run close to the data); and other services for access to published datasets, data publication, exportable workflows, distributed software, etc. CANFAR[56], a Canadian Advanced Network for Astronomical Research, is operated by the Canadian Astronomy Data

Center, which provides users storage, user management, date publication, data access, batch processing, openstack-based Cloud services, Cloud immigration, etc. SciServer[57] is a revolutionary new approach to do science by bringing the analysis to the data. It consists of data hosting services coupled with integrated tools to create a full-featured system. SciServer enables a major upgrade of SkyServer[58], one of the largest and most successful projects in the history of astronomy since 2001. SkyServer provides Internet access to the public Sloan Digital Sky Survey (SDSS) data, which is a 5-year survey of the Northern sky (characterised about 200M objects, measures the spectra of a million objects). The China-VO is a Chinese virtual repository, and the official name is National Astronomical Data Center (NADC). NADC[59] was launched by the Chinese government in June 2019 as one of the first 20 national certificate data centers. The science platform supports a whole astronomical research lifecycle -- from the proposal submission to a telescope operation, and services are provided for data processing, data publishing, data archiving and scientific paper publication, etc.

Regarding the Global Open Science Cloud, it will share the common challenges and requirements as the VO community. Currently all these VO science platforms have been run for several years and the technical solutions have been or at least partly implemented and tested, that can be taken as prototypes for GOSC. On the other hand, as a global science infrastructure, GOSC will need strong financial and community support, and it will be very useful for the whole science community.

2.5 Big Data Analytics needs for the Earth Observation Science Community

Earth Observation (EO) science is undergoing a paradigm shift. Geoscientists were used to having access to isolated images for specific studies with limited spatial and time impact. Nowadays, researchers around the world can easily access a wide range of high quality EO data from, e.g., EU Copernicus Sentinels[60], US Landsat[61], Japanese ALOS[62], Chinese Gaofen, etc. These systems produce data at rates of multiple TBs/day, and annual archives are at petabyte-scale. The EO research community is facing the challenge of responding to such rapid growth of data and information. Big Data Analytics (BDA) emerges as the combination of data access and best possible processing techniques especially based on modern machine and deep learning techniques. Deployment of BDA on cloud compute solutions that are logically close to the large data stores eliminates the need for massive data downloads.

Complex physical models at higher spatial, temporal and hybrid radiometric resolutions are behind all this information that are further integrated with richer reference and valuable validation data. EO data are also made available in a much wider variety in natural and social sciences, economics and biology domains that are far beyond the traditional remote sensing and spatial analysis disciplines. There are many challenges. EO science use patterns are highly diverse, but increasingly towards global applicability (e.g. U.N. Sustainable Development Goals). The global monitoring systems derive environmental information that can be used to assess the overall processes of important issues like global climate change, which needs consistent monitoring to produce statistical indicators on change, starting from a consistent baseline. The key requirement is permanent open access to full archives of historical data.

This is a huge data curation task which has not been addressed to the right extent. Sensor data needs to be made available in analysis ready formats so that researchers can directly use them and do not need to go through the complex pre-processing steps. In the global context, the full archives are often distributed over different open science clouds that need high connectivity, efficient data movement, and advanced technology such as smart caching, use pattern prediction, etc. Global upscaling implies the capability to scale computing with the massively parallel cloud compute resources. The integration of specialized hardware solutions such as GPUs is still very limited in the Earth Observation domain.. Many processing tasks are no longer compute constraint but very much constrained by the way to get the data from the archive into the computing and then back to the user or to the storage. It becomes storage limited rather than compute limited. Open source APIs are essential in providing parallel compute solutions that combine open data with open clouds.

As an example for EO applications, Sentinel-1 is used to monitor agricultural land at a continental scale and in a timely manner. Figure 2.1 shows the detailed European benchmark[63] for crop type mapping that is produced based on application ready geocoded backscattering data, exclusively derived from Sentinel-1, combined with the systematic collection of ground samples.

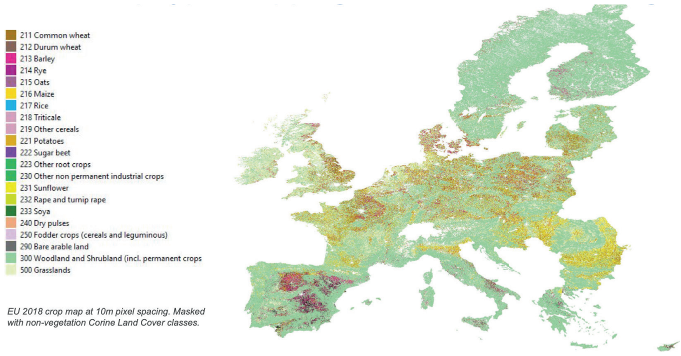

Fig 2.1. A European benchmark for continental scale crop type mapping with machine learning

Another example. Fig 2.2 shows a fragment of a global scope map product (covering West Africa) that combines human settlement detection on Sentinel-2 with surface water detection based on Landsat[64]. It combines the specific thematic outputs of different research groups to support contextual analysis (e.g. flood risk). These products can be generated or updated in weeks rather than years as it used to be done.

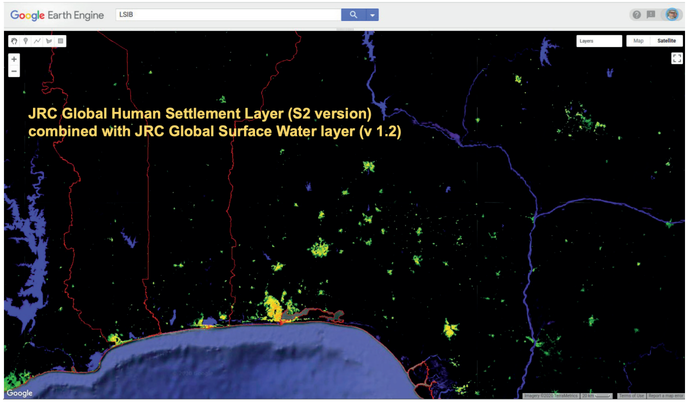

Fig 2.2. JRC Global Human Settlement Layer (S2 version)[65] combined with JRC Global Surface Water Layer (v1.2)[66]

In the field of EO, Google Earth Engine is the de facto standard in BDA. Data collections are made available on Google Cloud infrastructure. Although it is not an open cloud infrastructure, it is free for science use. Google Earth Engine also provides a massively parallel API for rich data analytics that is unmatched by other cloud infrastructures. The software can be applied to the huge data stack and the programming interface is intuitive and easy to use. EC Copernicus Data and Information Access Service (DIAS)[67] offers IaaS on commercial cloud infrastructure that is coupled to a multiple Petabytes S3 archive . Some DIAS instances and smaller equivalents are federated in the European Open Science Cloud, but overall in Europe, the familiarity and technical expertise to scale EO science use on open cloud solutions is still scarce. There are two important trends. In Europe, the Digital Europe agenda[68] is shaping up as the key initiative for a federated European Open Science Cloud. There is a specific topic, Destination Earth[69], which deals with modeling and monitoring and is especially relevant to the EO science domain which extends the "European Data Space"[70] to unlock thematic potential. The current open cloud landscape is very fragmented, e.g., per application domain, national initiatives, etc. It needs consolidation through coordination of funding streams, making sure that the funding streams are not only directed to particular infrastructure resources, but also the use of these infrastructures. Europe has world-class Open Source geospatial expertise that increasingly moves into applied machine/deep learning with rich reference data. But this is still limited to specific groups and needs to be promoted wider in the whole EO community. In the Global Open Science Cloud context, all these initiatives can be automatically scaled or ported to other continental Open Science Cloud initiatives that accommodate additional open data streams, globally distributed archives, unlock validation and reference data, and that will largely stimulate the Open Science development. It is an urgent requirement to be ready for upcoming open data streams. EU Sentinel continuation missions start launching in 2022, new US missions providing public access are scheduled for 2022 and beyond. Furthermore, open science clouds provide novel mechanisms to integrate increasingly sophisticated commercial data streams.

2.6 Supporting Global Open Science with Collaboration in GeoInformatics

Different from the EO community, GeoInformatics has a lot of small data communities. In recent years, a lot of efforts have been made to collaborate internationally that pertain to data integration, infrastructure integration, and capacity building integration. The drivers in the geoinformatics community to work together is the need to converge on essential elements of data infrastructure on development and operation. Stabilizing solutions through common adoption (technical standards, best practices, policies, and value) make these capabilities usable in minimizing confusions for users, better understanding and appreciating each other's experience, requirements, and constraints, so that not to make the same mistakes.

This can largely maximize the use of limited resources, avoid duplication of efforts, leverage expertise, efficient scale and share infrastructure where possible. As an example, OneGeochemistry[71] is an international effort that brought together over a hundred countries' geological surveys to build a shared and integrated geological map of the world. It also aims to address the need for the FAIR data, with respect to the interoperability and reusability. OneGeochemistry partners with relevant professional societies and science unions (e.g., IUGS-CGI, IUPAC), and brings together large research infrastructure providers across the world (e.g., AuScope/Australian GeoChemistry Network, EPOS/Multi-scale Laboratories, EarthCube, EarthChem, etc) to create momentum in the adoption of these standards.

These infrastructure providers are willingly participating because they've been struggling to establish and use the standards that can make them interoperable on a global scale. Sustainability remains as a key issue, because many data providers of these small communities lack funding for continuous operation. Another example. Founded in 2014, COPDESS[72], the Coalition for Publishing Data in the Earth and Space Sciences, aimed to foster international dialogue and coordination between publishers and data facilities, and to implement shared best practices for open and FAIR data. A statement of commitment that was signed by publishers and data repositories that lead to the creation of a new project called Enabling FAIR Data led by the American Geophysical Union (AGU), that developed and promoted recommendations and guidelines for implementing the FAIR data principles at data repositories and publishers to realize the full research data ecosystem.

As a third example, IGSN, the International Collaboration to Make Samples FAIR, aims to create a global unique and persistent identifier for physical samples. This small community builds heavily on observations made on physical specimens collected in nature that need to be integrated into a research data ecosystem that provides FAIR open transparent and reproducible scientific results. IGSN has received funding from Alfred P. Sloan Foundation for further development to create a sustainable business model and to design a more scalable and flexible architecture that can scale to new communities as it moves beyond the geosciences into other sciences domains like biology and archaeology. It is in the process of moving the IGSN Central Services to the cloud and transitioning from XML-based metadata to a semantic web architecture based on sitemaps The last example is very parallel to developments on a Global scale for Open Science Cloud. The U.S. Council of Data Facilities (CDF) started in 2014 to provide a collective voice on behalf of the member data facilities to the National Science Foundation (NSF) and other foundations and associations. CDF endorses and promotes standards and best or exemplary practices in the organization and operation of a data facility Identity, and supports the development and utilization of shared infrastructure services, including computing services, professional staff development and training services, and related activities.

Efforts are going on to create an organizational and social structure to work among these data facilities together with cyberinfrastructure providers both commercial and academic funded by NSF. In parallel, a new project is planned to create the prototype of a Next Generation Joint Data Facility with shared infrastructure in the cloud. It is evident that mechanisms are needed to internationally link a lot of different ongoing efforts to build global infrastructure, and for that an environment is needed where all these research efforts can create globally interoperable networks of solid Earth science data, information systems, software and researchers that could lead us into the global collaboration.

2.7 Summary

Over recent years, the growth of global research collaborations is speeding up in every domain field. Science has no borders. Scientists need data from other countries, regions and continents to develop more comprehensive models that lead to improved understanding and new discoveries. Desire to access data and resources across regions and continents become more urgent. Different communities are in their different developing stages. For example, VO and EO communities have built many advanced science platforms or virtual research environments that provide interactive graphical interfaces for easily searching, observing, and visualizing data generated by remote sensors. Advanced approaches, i.e., based on machine learning or deep learning using GPUs, are experimented as new solutions for extracting knowledge from Big Data. There are normally big developer communities behind to support the creation of smart tools, data mining algorithms that largely accelerate the evolution of technologies.

GOSC can benefit from the developments of such communities, e.g., interfacing into existing community applications, extending domain science platforms to be portable to GOSC and further making them generic to support wider use. On the other hand, GOSC can contribute by providing e-Infrastructure service solutions, e.g. moving large data across continents in timely manners, providing resources and distributed data processing solutions, etc. There are smaller and medium-sized communities having not (yet) gained momentum and sometimes do not speak fluent digital. They have limited development resources (developers, fundings, computing, etc.). GOSC can facilitate best practices exchange, transmit experiences and technology from other communities to support new needs, avoid duplicating efforts, and provide sustainable operation solutions, etc.

Chapter 3. Globale e-Infrastructure: Challenges and Opportunities in Achieving the GOSC Vision

What is the development status of global digital Infrastructures? This section examines the state-of-the-art e-Infrastructure technologies, identifying gaps in supporting communities' common requirements.

3.1 African Infrastructure for GOSC

The African National Integrated Cyber Infrastructure System (NICIS)[73] combines all important aspects of cyberinfrastructure including, the networking (offered by the South African National Research Network (SANReN)[74]), the data intensive research (offered by the Data Intensive Research Initiative of South Africa (DIRISA)[75]), and the high performance computing (offered by the Center for High Performance Computing (CHPC)[76]). The motivations of building such asset in South African are driven by the needs of many large science projects that South Africa is participating in, e.g., the Square Kilometer Array (SKA) flagship project, CERN facility, projects in Bioinformatics and Climate Change, the SADC Cyber-Infrastructure Framework[77], etc. All these large-scale research collaboration projects require continental-level e-Infrastructure resources. NICIS is also looking at how to work together with the other African countries.

The DIRISA75 infrastructure provides Cloud resources which is a data infrastructure that allows users to link the CHPC HPC systems (1.6 petaflops, 6PB Fast Parallel File System), and interact with the data storage through iRODS. Currently, 8PB storage is made available for active data (near real-time interactive access), and 40PB for long-term archival data & staging (accessible via Virtual Machines (VMs)).

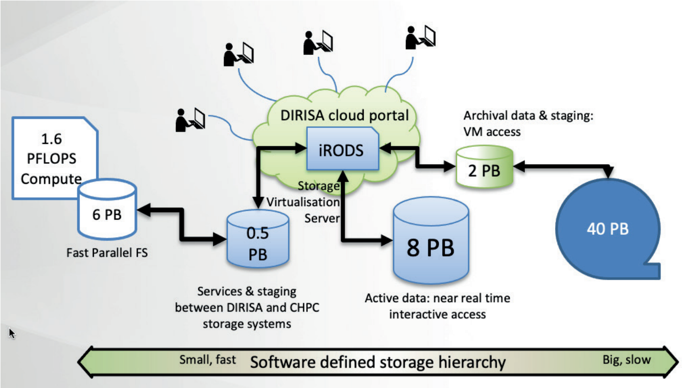

Fig 3.1 DIRISA Architecture

The SANReN74 serves as a backbone connecting all institutions across the country. There are also investments on the international cable, on the west through the WACS[78] and on the east the SEACOM[79], to be able to connect to Europe and the rest of the world. The WACS cables pass through West Africa, so the West African countries can be able to tap into that network, and the SEACOM passes through the Eastern part of Africa and Eastern countries can be able to connect there. This forms the basis of connecting networks across the countries.

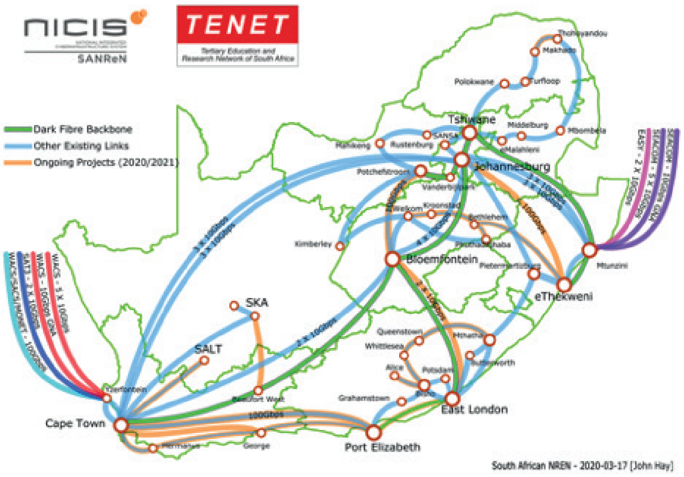

Fig 3.2 SANReN backbone

The CHPC76 owns the fastest supercomputer in the continent, the Lengau HPC system, offering 1,000 teraflops speed, and has been included in the TOP500 list for the past 4 years. Lengau comprises 1,039 Dell PowerEdge servers, based on Inter Xeon processors totalling 19 racks of computer nodes and storage. It has a total Dell Storage capacity of 5 PB and uses Dell Networking ethernet switches and Mellanox EDR InfiniBand with a maximum interconnect speed of 56 GB/s[80].

The infrastructure has been deployed across various countries in Africa, including Mauritius, Namibia, Botswana, Zambia, Madagascar, Ghana, Mozambique, and Kenya. Training has been delivered to people in those countries and there are projects already taking place in collaborations.

The collaborative plans in the context of GOSC include but not limited to:

- Build the OpenStack based cloud federation across South Africa, and to combine all existing resources in the country, e.g., from the Astronomy and Bioinformatics communities.
- Extend it to develop a continent-wide federated cloud, providing the HPC systems to the South African Development Community (SADC)79 and the rest of the continent.
- Create an operational African Open Science Platform (AOSP).
- Create synergy with European Open Science Clouds and connect AOSP with the Global Open Science Cloud.

The key message here is that South Africa has user communities, the infrastructure is already built, and the technologies are ready. Not just South Africa but the rest of Africa is ready to be able to connect. Although some of the African countries still have network problems, through the GOSC activities, we can expect good progress to connect them.

3.2 The ARDC's Nectar Research Cloud: Challenges and Opportunities for the GOSC

The Australian Research Data Commons (ARDC)[81] was formed by the Australian Federal Government in July 2018 through the merger of three existing National Collaborative Research Infrastructure Strategy (NCRIS) e-Infrastructure capabilities, including the Australian National Data Service (ANDS), National eResearch Collaboration Tools and Resources (Nectar), and the Research Data Services (RDS). The vision of ARDC is to transform digital infrastructure to support leading edge research and innovation, aiming to provide Australian researchers with competitive advantage through data, accelerating research and innovation by driving excellence in the creation, analysis and retention of high quality data assets.

ARDC plays a key role in transforming Australia's research culture through its strategic vision, leadership, programs and resource provision, and through its mandate to provide national coherence capacity via the Digital Data and eResearch Platform (DDeRP)[82].

The portfolio of ARDC services is very broad and drives the FAIR data agenda through a range of data and services activities. ARDC provides high capacity digital data storage, cloud compute, HPC/HTC, Data Retention infrastructure, Network and Identity & Access service. It also provides high level services and tools for data management, analysis, visualisation, virtual labs, platforms and gateway. Community support is active in the areas of outreach, training, advice, consultation, and policy.

ARDC runs the Nectar cloud[83] which is a leading edge federated model across 10 NCRIS sites, 11 ARDC Centres of Excellence, 14 Cooperative Research Centres, supporting over 3,500 projects, and 17,500 users. This distributed model enables partnership and co-investment that allows sites to have their own business model, in alignment with the block funding provided to researchers, together allowing strategic research planning and infrastructure procurement.

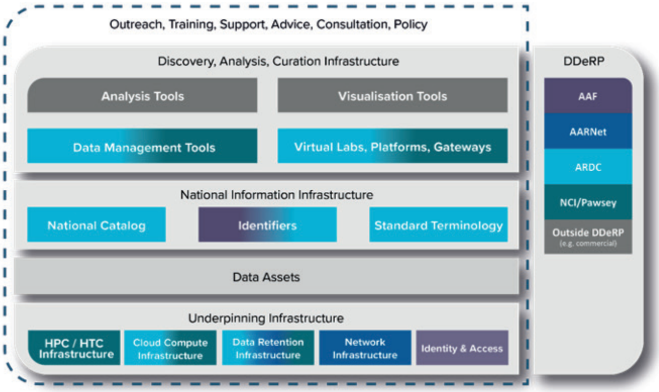

Fig 3.3 ARDC's Areas of Activities

There are many challenges, e.g., it's difficult to balance a federation agreement on a standard approach. This is because local customization is desired in order to encourage innovation. Other challenging issues include, e.g., access to compute on a national scale; capacity to support all types of research disciplines and domains; access and sharing of sensitive data; easy navigation for researchers between resources and different types of compute platforms (Cloud, HTC, HPC); accessing the growing need for multi-skilled resources across the research ecosystem. Many of these challenges are not unique to the Australian e-Infrastructure. The Nectar is moving with new activities such as the kubernetes, the container, and orchestration, that ARDC can partner and complete with European and global collaborations .

ARDC brings the important voice from the southern hemisphere and the rich experiences of sharing skills, best practice and approaches that can inject into the discussion of the GOSC. The areas that ARDC seeks for collaborating with the GOSC include but are not limited:

- Move the compute and analysis to the data (Australian researchers need access to large international datasets).
- Make research collaboration easier, and to use digital infrastructure in a coherent way across disciplines and institutions.
- Skills for research infrastructure, transition of specialist skills to shared commodity skills.
- Share best practice and approaches.
- Standardised authentication and authorization mechanisms across different resources.

ARDC also provides valuable suggestion for action priorities of GOSC include:

- Alignment in developing standards, open-source tools, best practices that allowing the different infrastructures to be used by researchers and developers/providers in a common way;
- Federation at a global level: although this would be very challenging, it can start with domain specific pilots such as the BioCommons[84] which relies heavily on international access to genomics data sets;
- Opportunities to collaborate: Common eResearch infrastructure challenges are globally. ARDC can share the successful experiences for making research. For example, during the COVID19 Melbourne lockdown, ARDC made research datasets readily available and converted data insight into digital tools to tackle contact-tracing.

3.3 Research e-Infrastructure Federation in China

Founded by the Chinese Academy of Sciences (CAS), the China Science and Technology Cloud (CSTCloud)[85] is a national infrastructure for CAS scientific communities and China's top research. It federates the national computing resources including, the China Science & Technology Network, China National Grid, and National Scientific Data Centers. As one of the key national infrastructures, it is included in the 'China National 13th Five-year Plan for Informationization', obtaining sustainable support for long-term development.

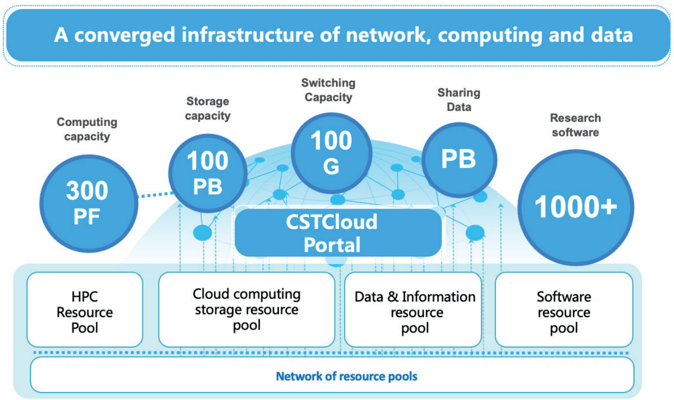

Fig 3.4 CSTCloud's capacities

The design of the CSTCloud is based on the principle of 'Openness and Sharing'. It aims to develop an open architecture that is capable of integrating various national and international computing resources in order to support multidisciplinary open science research. For example, it provides computing facilities for Chinese advanced research projects including CASEarch, CAS Space Science Missions, and research related with big facilities or observation stations such as the Five-hundred-meter Aperture Spherical Telescope (FAST)[86] and the Large-High-Altitude Air Shower Observatory(LHAASO) [87].

CSTCloud has established global partnerships and collaborations, i.e., with EGI Foundation and GEANT in Europe,, the African Open Science Platform, and other organisations which share common interests to provide cloud solutions for scientific research from a global perspective.

The CSTCloud is among the pioneer e-infrastructures that support GOSC vision. Resources have been allocated for the GOSC initiative to set up a testbed that joins the EGI federation in Europe and later other countries' facilities. The work will lead to a discussion of a GOSC federation model driven by the GOSC working groups and case studies.

3.4 OpenAIRE Experience with Global Open Science

OpenAIRE takes the Open Science infrastructure approach from the scholarly communication perspective, initially focused on Open Access to publications, and now extends its mission to assist the EC in implementing Open Science policies.

OpenAIRE witnessed the movement from Open Access to Open Science in Europe. What is evident in the past years is that Open Science can provide better access, reproducibility, and accountability, which lead to collaboration, innovations resulting in better research. It's not just about having access to knowledge, but about actionable and sustainable knowledge, data retention and connection. Open Science is becoming the new norm. However, the challenge is how we can achieve it -- that everyone has access to Open Science resources in a way that is equitably, transparently and trusted.

Towards a Global Science Commons, it is time to act at different levels -- local, national, regional, and global. The focus should be on interlinking policies, technologies, and social infrastructure. There should be decentralized common values without lock-in, and communities need to think about shared benefits, sustainable knowledge and sustainable infrastructures. A global science commons cannot exist without Open Science, and it is the only way to accelerate the implementation of the U.N. Sustainable Development Goal.

OpenAIRE is approaching to the Global Open Research Commons through three main pillars of actions:

1. Building synergies via regional networks. The starting point is metadata. What OpenAIRE has achieved is through a set of guidelines to make sure there is a minimal metadata set that can be used for data exchange at early discovery. There are OpenAIRE guidelines[88] for literature repositories, data repositories, software repositories, CRIS Managers (a heritage system), Usage Statistics, etc. Around metadata activities, OpenAIRE has built a network across Europe and linked to infrastructures around the world. The National Open Access Desks are established serving as ambassadors of Open Science and representatives of every European member. OpenAIRE's view is that if other regions or countries can build similar networks, we can depend on each other's layer of Open Science Cloud to work together. The lesson learned is that we need to realize that not everyone is on the same speed, for example, the Australian data services are much more advanced than some of the countries in Europe, and we need to exchange best practices and help each other through networking. Another lesson learned is to make sure that there is commitment from the governments because this is something to be sustained and requires commitment through funding and resources.
2. Producing tools for the global community. In scholarly communication, there are many resources nowaday are paid and locked-in. OpenAIRE aims to co-produce services with other regions and around the world in order to open the global resources to the community. There are a bunch of infrastructural services or services for end users available[89], and these services are put out into the global commons. The services are not just for researchers but for the whole range of stakeholders and actors involved in the research process (from discovery, publishing, to monitoring).
3. Co-producing global assets for scholarly communication. The focus is on the OpenAIRE research graph[90], that links repositories for journals, data, software, together with other authoritative global sources. Such an open scholarly communication resource could be used for discovery, monitoring, and for research communities to use on their own domain. This research graph is at a level of scholarly communication. If other communities can build their knowledge graph based on their domains, we can all hook them up in order to have a global interlinked Open Science System.

3.5 EGI Experience in Supporting International Scientific Collaborations

As an European infrastructure for data-driven, computing intensive and exabyte-scale processing, EGI's core user communities are international groups that need to process and analyze data across Europe and with other continents and international institutes. EGI's vision is that researchers should have a seamless access to all the services, resources and expertise to collaborate and conduct world-class research and innovation. During the past 15 years, EGI has been delivering open solutions to advanced computing and data analytics, supporting hundreds of international user communities reaching a scientific impact of about 2,000 open access publications per year, and about 71,000 registered users worldwide.

EGI has a very strong user community in physical sciences with high level maturity. This is due to the fact that there has been radical practices in international collaborations towards big experiments such as that happened in the LHC lab, and they have been successfully pulling and connecting all infrastructures they need for their science. But there is a new wave of other scientific disciplines that are more and more challenged by data, for example, Earth Science, which today largely benefit from EGI Federation Cloud and data access services. Social Sciences and Humanities are growing in readiness and maturity, and they're quickly moving from data integration into provisioning of services for data analysis. The level that these communities have reached is hundreds of international institutes from tens of countries, resulting in thousands of researchers worldwide working together focusing on the same scientific challenge. EGI has already successfully leveraged the support to large scale international collaborations, but definitely encountered even big challenges ahead. Many international scientific collaborations are sufficiently well organized to succeed in opening infrastructures for their scientific purposes. But in communities of practice, this is a way harder, because these communities typically gather the data around the scientific applications and they don't have long-term projects that provide international governance for their collaboration. For example, WeNMR is probably the largest EGI user group, which is a structural biology community that has reached a level which is even bigger than many other scientific collaborations. However, it is today not properly served, and the reason is that it lacks a governance to open up the infrastructure to serve them up to the requirements that they have in this community of practice. Thanks to bilateral agreements with the United States, Asia-Pacific and Latin America, EGI has pooled resources for WeNMR. With these bilateral agreements, WeNMR scientists from everywhere in the world can use their simulation tools which are supported by an infrastructure layer where bilaterally infrastructures in the world come together recognizing the value of this community and bringing capacity to support this open access. Without these bilateral agreements, it would be impossible for scientists like structural biology, in this case where only 25 percent of the users are European the rest are from other continents, to be supported in a coordinated manner, and open scientific applications would be otherwise only accessible in Europe or to specific countries in Europe.

How has EGI achieved these? EGI is creating governance in Europe through a membership model. EGI is a membership organisation with 26 members from Jan 2021. Where national organizations are leveraging national funding and federating it at European level, this is not sufficient to support collaborations internationally. This worldwide federation has been constructed by having bilateral agreements with as many partners as possible in the world, for example, EGI has created agreements with China, Australia, U.S., etc. What are the key elements in EGI experience? To have successful scientific collaborations at international level, it needs to have a community level governance without which it's very hard to have homogeneous access policies and agreed access policies to data and to the software to process the data. One of the biggest barriers we need to face is that our national infrastructures in Europe but also worldwide are typically funded to serve the national user communities, which means they are not open by default or at least a part of their capacity is not open by default. It is important to create a global infrastructure which can be accessible and made available to user communities that are not strictly related in terms of funding to specific national funding agencies. Collaborations are also enabled by seamless access policies and across all the federation members. These policies have to be agreed, for example, there must be pledge capacity which is agreed. This is particularly important in computing because computing is depletable, so we cannot support an infinite number of users, and there must be agreed capacity allocated by the providers to the user communities.

Interoperation is also important. When we use the word interoperation, it doesn't mean interoperability, which is not just about using the same standards (which is sometimes impossible across communities or not even achievable across infrastructure). Interoperation means having higher level services which create a glue that allows users to access infrastructures which are intrinsically heterogeneous. We need some form of coordinated service management, so users can be supported in a coordinated manner. When they are international and that means the providers have to find a way to do this. Usage has to be accounted for to create an impact and to make a case for funding agencies. Security has to be coordinated, for example, incident response. Where we have security as a concern in the assets which are being federated and interoperable, trust and identity frameworks are a necessary requirement for secure access to data to applications and infrastructures.

There are three but not least other gaps should be addressed internationally:

- Governance. We need a lightweight international-level federated e-infrastructure both at European level and worldwide level. Through partnership, we can leverage each other's networks and reach out to our peers in other continents more easily if we can coordinate at continent level at least in Europe. At international level, we don't have a governance where all these infrastructures can come together and discuss the strategies and cooperate. The current bilateral agreements EGI has with a larger portfolio, but this doesn't scale, for example, when collaborating with OpenAIRE with GEANT. We still have a lot of bilateral agreements which means there is a long lead time when we have user communities that have requirements to the point when there is a coordinated infrastructure available for them to be used. These lead times can be years, which is not acceptable for modern science.
- Organisation. It needs some form of coordinated human networks having experts that work together to support user communities, because infrastructures are necessary to reengineer the data repositories that need high skill. If there is no coordinated support, effective usage is very hard to achieve.
- Access Policies. There is still a major blocker when we have an international user community with interest in data in the United States, Europe, China or other regions. The first question for a researcher is how do I analyze this data? Typically this data is so large, in many cases it cannot be downloaded locally, and it has to be processed where the data is available and that means having access to additional infrastructure capacity which is coming together with the data which is very hard to organize at international level. This is why having open infrastructure policies next to open access policies to data and publications is one of the major hurdles that we need to tackle together. We hope that through European Open Science Cloud this is going to improve in Europe, and we see an opportunity in this global effort to be able to scale up our efforts at international level.

3.6 GEANT Experience with Global Open Science

GEANT is the largest and most advanced research and education (R&E) network in the world, through interconnections with 40 National Research and Education Network (NREN) organisations in Europe, connecting over 50 million users at 10,000 institutions[91]. The backbone network operates at speeds of up to 500Gbps and reaches over 100 national networks worldwide. GEANT international network capacity is shared with other continental and regional research and education networks in the world and together it forms the so-called REN, the global Research and Education Network. One of the notable projects, Bella[92], has recently launched a totally new subsea cable going between Portugal and Brazil which will increase the capacity for cooperation between Latin America research and European research and will allow European researchers to access, for example, the telescopes in Chile. In this way, GEANT contributes to creating a federated worldwide network that is purely based on sharing, in the same way that the internet is based on, sharing capacity and cost and allowing others to use your resources. This is done almost without any bilateral agreements, but on mutual recognition -- that 'if I can use your stuff and you can use my stuff', both parties can have a win-win situation.

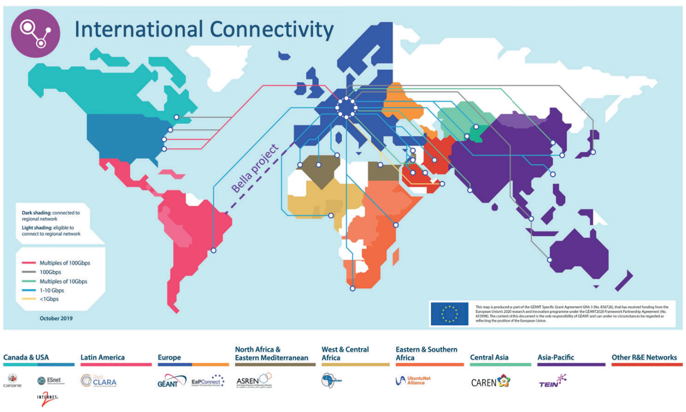

Fig 3.5 GEANT' International Connectivity

This Global Research and Education Network (GREN), is built upon the foundation of regional networks like GEANT. GEANT is currently building out connections that cover all countries, so this global Research and Education Network can bring all the way to the National Research and Education Networks (NREN) in Europe. As an example, NORDUnet[93] who are building their next generation network in the Scandinavian countries which seamlessly integrates into the GEANT network, and into the GREN. The Regional Research and Education Networks are built on the foundation of the National Research and Education Networks (NREN), one for every country. As an example, Swedish National Research and Education Network will connect all the research institutes and all the universities inside Sweden through SUNET[94]. This GREN where we share each other's capacity and resources makes sure that anybody from any university in Sweden can share resources with researchers in China connected to the Chinese CERNET[95] or CSTnet. Hereby, GEANT has helped to build an infrastructure that is fully capable of supporting a Global Open Science Cloud, and on top of this network infrastructure, it supports the global scale services such as, eduroam, eduGAIN, to make sure that access to a Global Open Science Cloud can be guaranteed. GEANT also supports the global scale security services to make sure that the access to the Global Open Science Cloud can be as secure as possible. Global cooperation is done on a basis of trust and recognition -- "by sharing, we together grow bigger rather than being on our own island", and a Global Open Science Cloud should be based on the same principle as much as possible.

Different from EGI, the nature of its organisation model needs more agreements and governance, GEANT's global Research and Education Network has a lightweight governance that is stable and solid and provides for audits at the highest level, but most of the cooperation is done at a community level by working in work groups and task forces. With the Global Research and Education Network, GEANT is ready to support this Global Open Science Cloud federation. Interoperability, shared standards, and a close collaboration will be the key to its success.

3.7 Summary

There is high technical readiness for digital infrastructures development at regional and continental level, and there is an incontestable interest in joining together to form a world-wide connected computing infrastructure. Thanks to GEANT, who has been among the pioneers paving the path for building up a global linked high speed network backbone.

Technical challenges are commonly recognised. Agreeable approach to enable easier access to compute at a global level is needed, that implies unified AAI, shared standards to enable interoperability, interoperation and cooperative service delivery and operation, etc. National production should be made internationally relevant and best practices and experiences should be shared, i.e. how to move computing to data, how to handle sensitive data, how to provide integration access to HTC, HPC, and Cloud, etc.

There are other aspects beyond technical issues. What is an appropriate governance structure to coordinate global resources federation and service provision? How to organize community support activities across regions and continents? What are suitable access policies that accommodate differences from countries to countries, organisations to organisations? Sustainable funding sources are mentioned as one of the key aspects which we will investigate in the next section.

Chapter 4. Funding Agencies' Perspectives for GOSC

In order to realise the GOSC, funding for support to international collaboration is both necessary and important. How the funding agencies have been supporting the international collaborations in science, will support this priority in the future? What are the directions? This section explores these issues by interviewing with three funding agencies from different regions and countries.

4.1 Chinese Funding Agency's Strategy and Vision for GOSC

Chinese Academy of Sciences (CAS), is China's leading national scientific institution in natural sciences and high-tech development, and a supreme scientific adversary body. It has established a mixed funding mechanism both top-down and bottom-up for international cooperation and science collaborations (i.e., EISCAT-3D & SYRSI radar data fusion and computing, CASEarth4SDG). Its featured international competition programs cover different topics, disciplines, specific or even critical subjects. Through the invested collaboration projects, CAS has established a formal corporational partnership with over different countries at three levels -- the academy level, the institute's level and the individual level. In the near future, the focus will be still focusing on major global issues, for example, climate change, ocean, earth science, astronomy, and so on.

A new project fund recently approved under the CAS International Partnership Program (IPP), will sustain CNIC to develop the GOSC for 5 years (starts in 2021). The aim is to boost collaboration with international partners to harness the full potential of cooperation networks and digital infrastructure to advance science and address global challenges.

Other possible funding opportunities that open to GOSC or similar activities include the CAS PIFI (President's International Fellowship Initiative), and from other key Chinese funding agencies including, the Ministry of Science and Technology, China Association for Science and Technology, and Beijing Municipal S&T Commission. The discussion is very positive. The CAS's strategy for investing in the GOSC is to provide funding for different stages of the development: in the past 2 years, seed money was provided for initial setup of collaborative workshops, conferences and training. With the developments entering the second stage, the funding is provided for pilot programs and the formulation of international projects that may lead to future international 'big science' projects such as SKA.

From 2021 China will issue the 14th five-year plan on Science, Technology and Innovation. In the next five years, science, technology and innovation will be more emphasized and more resources and funding will flow into the communities in these areas.

4.2. European Commission's priorities and visions for EOSC and global collaborations

In the past few years, the European Commission (EC) has put significant efforts in developing the EOSC. The Europe Open Science Cloud is the federation of the resources (most of them are existing already), motivated by the vision of the web of digital objects (dataset, research outputs, publication, algorithm, models, software) and services. Most resources are from member states of the European Union and their associated countries. The European Commission doesn't own them, but makes the federation's efforts.

EOSC is funded under the policy of Open Science. The vision is that science is transitioning, and becoming more collaborative, more open, more digital and more data-driven. Eurpe wishes to remain at the front-end of this transformation. There are many challenges, for example, issues on open access to scientific outputs (publications, software, models -- Europe is leading this area and taking the same steps as what US and France do); making data FAIR; opening science up between researcher and society/citizens and open data across disciplines; research productivities; reproducibility of research results to conclude scientific publications, and so on. These areas are all covered in the EC program.

EC pursues Open Science quality. In order to pursue them effectively, it needs enabling digital infrastructures and frameworks, where a researcher is able to locate the resource, and be sure whether can use them (i.e., what are the terms, the licence, etc.). Another enabling factor is the so-called rewards incentive/research assessment system, by which researchers and research institutions are assessed, and to make it transition to a system that favour collaboration rather than rewarding major individual activity; rewards quality and impacts rather than where the research publish (e.g., high impact journals) or the amount of the publication is produced. This is the current focus and will be intensive in the future.

In summary, the EOSC comes from as an integral part of European Open Science policy, is an enabling technical infrastructure on which to go forwards openness in science, where openness means high quality, high impact, more productivity, and more innovation.

The European Commission is both a research funder and the policy maker. What it has made for policy, it has to implement them in the program and vice versa. As a research funder, EC mainstreams Open Science practice in its funded projects, for example, in the new Horizon Europe program framework[96], the data management and producing FAIR data will be mandatory for all research projects. Regarding FAIR data, although very often FAIR is open, it isn't necessary. International cooperation needs to work with data, knowledge, and resources across the world. If a scientist needs data, s/he shouldn't be aware where data comes from (whether it is in China, in Canada, or in Europe). Ideally, the dataset should be FAIR globally, (e.g., the National Institutes of Health (NIH) and National Science Foundation (NSF) in the US have advanced policies in that aspect.) EC aims to make it a contractual obligation for all of its beneficiaries from next year onwards to provide their data in FAIR form based on the principle that is as open as possible and as close as necessary.

EC mainstreams Open Science policies across its funding programs. For years, EC has invested a significant amount of money. For example, in the last 3 years, EC has invested 250M Euro in developing and prototyping EOSC, and will intensify the efforts from next year and on. It is going to intensify it both structurally as well as quantitatively. Structurally, EOSC will be in a new form, a form of partnership between EC, member states and research performing organisations. An association has been set up[97], a legal entity where the stakeholders (already close to 200 applications) are grouped to join as members or observers. This association will represent EC and the member states. The association will provide a strategic research agenda. A draft version was made ready and went into public consolidation during the summer 2020, to receive feedback from EU member states and other parts of the world.

In the international aspects, the vision is that science is global, and it is needed to enable cooperation in science and research. The principle on which to pursue international activities with partners is mutual interest, similar values (in terms of the research integrity), and ethics (to pursue collaborative activities with a Level Playing Field, so nobody has an unfair advantage on the basis of reciprocity). With cloud initiatives in many countries including Australia, Canada, China, Europe, South Africa, US etc. it certainly needs to work together at both the governance and technical level. The common aim should be to promote interoperability between the different components.

In Europe, there is a lot of data FAIR, and there will be even more in the near future. The approaches and standards that were developing for the Cloud and EOSC are also public. There are a significant number of policy papers produced that reflect the policies that EC are taking on different aspects that can be found on the web. This is practical reaching-out for a collaborative hand to all partners globally and help other countries making their own reports, data, publication and research outputs FAIR (findable, accessible if possible open). EC also collaborates with many initiatives in relation to data including GoFAIR, RDA (together with NIST, the National Science Foundation, the Cyberinfrastructure community, and Australian partners, etc.). This cooperation goes on and seems to be the right framework, for example, there is a RDA working group to develop interoperability solutions for technical standards; a working group working on FAIR maturity assessment in which EC has invested efforts together with American colleagues. The vision of the RDA interest group on Open Research Commons is also important in that respect.

In summary, EC has the corresponding policies and aims to establish a Global Commons for science resources. This is the desired end point and EC wants to engage in very concrete actions of openness and cooperation on the basis of reciprocity and solidarity in research.

4.3 US funding agency's perspectives on Global Open Science Collaborations

The US National Science Foundation (NSF) has a very broad mission including, promoting the progress of science, advancing national health prosperity and welfare, and securing national defence. The way it does this is by supporting foundational applied and translational research across all of science and engineering research and education. NSF supports all fields of science span areas such as, biological sciences, engineering, mathematical and physical sciences, compute and information science and engineering (where the Cyberinfrastructure is offered), geosciences, education and human resources, social behavioral and economic sciences, integrative activities, and international science and partnership. It has a very broad community that is very heterogeneous in the type of research, researchers, data, methods, and approaches that are used. The funding mechanisms include, supporting grants for fundamental research, education and training grants, community grants, and funds for research resources that range from large instruments and shared experimental facilities, telescopes, observatories, and the cyberinfrastructure is a big part of that. Grants for Cyberinfrastructure includes a large computing leadership class, computing resources, campus research resources, capacity resources, cloud infrastructure testbeds, edge resources, networking, data and computing services, and a broad spectrum of infrastructure both in size scale and capabilities that support research. The approach that NSF has both in policies and mechanisms are targeted to the specific needs of this broad community across many different disciplines as well as the integration in transdisciplinary practice. NSF's funding program is complementing the other sister agencies within the U.S. research ecosystem, but having unique angles, e.g., in the ownership of IP, ownership of data, and a lot of solutions targeted to the unique nature of NSF mission and of its community that it supports.

Open Science and openness is fundamental to NSF's vision, that is integral to its mission and built in all research that's supported by NSF. NSF emphasizes the importance of Open Science which fuels scientific discovery and economic gain by making the products of federally funded research more easily accessible and usable. Open Science can improve scientific rigour by directly linking the products of research (data and software) to their associated publications, making it easier for others to confirm the validity of a scientific result reported in a journal or juried conference proceedings.

Collaborations with reciprocity, FAIR, Level Playing Field, shared values and visions are also integral to NSF's mission. NSF addresses this by taking a long-term view and philosophy towards it. It needs to create the incentive mechanisms, the policy infrastructure, and the Cyberinfrastructure, that's needed to be able to have the grassroots change to have a culture of openness. And we must keep in mind the heterogeneous nature of our community.

NSF's been making efforts to set up the right policy frameworks, for example, the data sharing policy requires funded research projects to include the data management plan. The aim is to bring this culture change at the institution level, at the principal investigator or the researcher level to build the culture of openness and sharing. The launch of the NSF Public Access Initiative[98] aims to accelerate this and advance the frontiers of knowledge. Efforts also made on creating a Cyberinfrastructure, consisting of repositories, individual PIs (Personal Investors) with their datasets, large computing facilities, networking and software, that facilitates sharing and interoperability across these different pieces. There are also community building efforts that allow the community to come together around the principle of open science and sharing. For example, NSF has a public access repository which allows researchers to share publications. Researchers are required to think about data management as a fundamental aspect of any research they propose where the concept of sharing is very important.

The impacts already happen. An example is the creation of the COVID19 HPC consortium. Given the idea of interoperability and sharing as a fundamental concept or requirement in NSF's investment, it's been realised that it was necessary to bring researchers and resources together to address the pandemic that was/is affecting all of the world. NSF very quickly came across not only the NSF community, but also the other funding agencies, US government, industry, academia to create a consortium for resource sharing, and these literally happened in days. All the communities were able to bring and combine resources, and make them accessible to the research community in an open way, to be able to collectively address the pandemic. Because sharing, interoperability and collaboration was integral to the way the resources were built and operated, bringing them together was very simple, and NSF was able to set up this sharing mechanism on top of the existing platforms in about just a couple of days.

NSF has been increasingly investing in the so-called missing middles. There are many different disciplinary and transdisciplinary resources, repositories, dataset, and researchers. Science happens when they all come together to address the important problems of our day. How to ensure that scientists can access this data across different ecosystems, discover and use data in an effective way as part of end-to-end workflows? NSF's been focusing on building out the missing middles that bring these different pieces together to serve the need of science.

Engaging the community around the vision of Open Science and shared infrastructure is very important, and NSF has been fostering research community efforts such as RDA and CODATA. NSF's also been working with other US agencies to align related policies, for example, the National Science and Technology Council has a subcommittee on Open Science and NSF holds a leadership role so that to be able to bring all the communities in the different agencies to work on the Open Science policy and approaches and enable this broader vision of openness and sharing.

4.4 Summary

There is a global effort around Open Science and Open Science Cloud or Open Science Commons. The policy makers and the funding agencies are confirming that there is a strong support of policies towards Open Science. The importance is widely recognized and major funding and initiatives are already happening.

From the funding aspect, there are opportunities for targeted actions in a top-down approach or bottom-up approach. CAS provides seed funding initializing collaborations i.e., starting activities, training, and outreach. It also provides support that scales up to high collaboration and mega science or projects. NSF looks more from a bottom-up perspective, concerning how to make sure the funding provided across both small infrastructure projects and large facilities and instruments, and the principles of Open Science, sharing, and interoperability is built into them. The European Commission has moved away from the top-down approach, and in the past few years has become far more bottom-up. For instance, there is a part of the EC program that's devoted to oceans, clean water and sustainability including international partners in the program. On the other hand, there are still case-by-case basis top-down actions that are aimed at facilitating international collaboration.

There is still a need to evolve with existing policies and policies remain key and fundamental as an enabler of international collaboration. Collaboration will be efficient and productive if collinear policies are in place, for example, the openness of data is there for everybody more or less according to the same rules. These are policies that an organization can impose on their staff, for instance, the research organization or a university or a funding agency can impose on their researchers. It needs everybody to take a position on these policies and to implement them.

COVID19 provokes unprecedented challenges for us all. In NSF, COVID19 HPC consortium is started with a handful of industry government agencies in the U.S. and academic partners, and it has grown not only across industry in the U.S. but internationally and the UK, Sweden Switzerland, Australia, Japan, South Korea joined it. It was able to do this in a matter of weeks and months in a very agile way. What was the enabler for this success? The observation is that there was a collective desire to fundamentally address this pandemic that we were all facing. When there is a shared desire for a shared need, people can come together.

A symmetric effort in Europe. With the motivation of the pandemic, things moved very quickly. For example, RDA within a record time of a month, produced a recommendation for COVID19 data that would have taken years to produce. On the other hand, a global data platform is launched, hosted by the Molecular Biology Laboratory EBI in Cambridge and the ELIXIR consortium. Although this is a European data resource, it is open internationally. There are many depositors of sequences coming from China, U.S., and many other countries. And the majority of users of these platforms are outside Europe. It is a platform to fight COVID19 that brings together genomic, molecular, epidemiological, and clinical data to be deposited and accessible. A mutual agreement with NCBI in the U.S. was established, not only for COVID19, but it is by default that any data (genomic sequences) uploaded in the U.S. can also be seen in Europe, and vice versa. There are 70 000 individual researchers that use the platform today. This year shows a sign that international collaboration is possible. For example, in energy physics there was a decision to devote a part of the infrastructure and pull it with the United States with Open Science Grid to support molecular docking simulations across the U.S. and Europe. And the impression is that it is more exceptional under the pressure of emergency than practice.

Conclusion and Future Work

In this report, we have reviewed the existing political landscape, the high level initiatives and the role of them. We have made dialogues with users communities who have been bringing up their requirements and sense of urgency in tackling the data delusion and the problems for data sharing to enable excellent science. We have learned the perspectives of international infrastructure providers in different technical domains from networking to international e-Infrastructures.

We have jointly recognised that international science needs coordinated policies, e-Infrastructures, fundings, and support in different ways. We also have jointly realized the GOSC action is timely in order to discuss global cooperation, leveraging the experience of Europe with the European Open Science Cloud and the existing international collaborations and policy initiatives.

These discussions triggered enthusiasm for many more questions. What is the missing bridge that connects the initiatives in different continents and countries that brings our efforts together in a more coordinated and aligned manner? What is the missing building block that we need to construct to move ahead from policy to implementation? What is the first priority to make this collaboration structurally part of our way of operating? A list of tasks for GOSC was identified from a number of workshops that is shown in Fig 5.1.

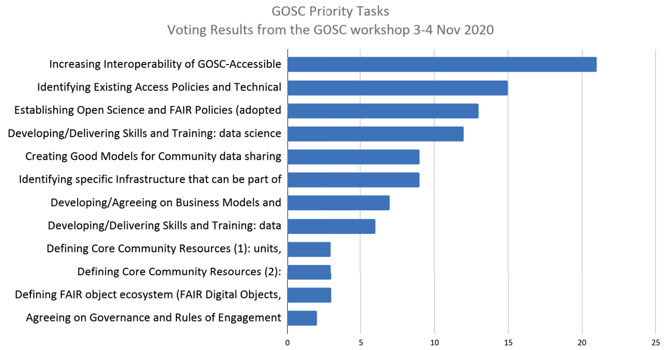

Fig 5.1 GOSC Priority Tasks

To pursue GOSC priority tasks, on 28th June 2021, CODATA launched the GOSC Working Groups and Case Studies[99]. CODATA GOSC thematic Working Groups aim to address a number of challenges shared by Open Science Clouds, including:

- Governance and Sustainability: sharing experiences and good practice.
- Policy and Legal Dimensions: aligning principles, policies and exploring legal issues.
- Technical Infrastructure: including network connectivity and protocols, secure Authentication and Authorization Infrastructure (AAI), mechanisms for federation of computing , data and other services.
- Data Interoperability: including identifiers, semantic services, rigorous contextual and provenance metadata, analytical tools and virtual research environments.

In parallel, Case Studies of domain and cross-domain research will be developed to provide a concrete grounding and demonstrator of topics identified above. Four Case Studies are currently planned: Incoherent scatter radar data fusion and computation, Biodiversity and ecology information platform, SDG-13 climate change and natural disasters, and Sensitive data federation analysis model in population health.

GOSC invites global Open Science infrastructures, international research communities, and research organisations to join together to build this new GOSC community.

Acknowledgement

We'd like to thank all speakers in the Global Open Science Cloud Workshop, Nov 2020, for their valuable contributions.

References

- [1] European Open Science Cloud, https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud
- [2] Chinese CSTCloud, http://www.cstcloud.cn/
- [3] ARDC, https://ardc.edu.au/
- [4] Malaysia Open Science Platform, https://www.akademisains.gov.my/mosp/about/
- [5] Africa Open Science Platform, http://africanopenscience.org.za/
- [6] GOSC workshop in EGI conference 2020, https://indico.egi.eu/event/5255/
- [7] UN Sustainable Development Goals: https://sdgs.un.org/goals
- [8] In general, there are two ways for an international organization like UNESCO to set standards. One is a convention which is illegal binding that needs a heavier process to handle; the other one is the recommendations which are not legally binding but having a monitoring mechanism allowing member states to report back and be accountable for the provisions that they have agreed upon in that particular recommendation.
- [9] Open Science is still very fragmented, there are many related concepts and communities, but they don't often speak to each other.
- [10] The first draft of UNESCO recommendation on Open Science: https://unesdoc.unesco.org/ark:/48223/pf0000374409.locale=en.page=10
- [11] ISC Science Support Bodies include CODATA (the Committee on Data for Science and Technology), WDS (the World Data System), INGSA (the International Network for Government Science Advice)
- [12] ISC International Scientific Committees are largely created for ISC research, e.g., SCAR for Antarctic research, IUCAF for radio astronomy and space science, SCOR for ocean research, SCOR for space science, and SCOSTEP for solar terrestrial physics.
- [13] ISC 2019-2021 action plan: https://council.science/actionplan/
- [14] ISC GA 2021: https://council.science/about-us/governance/general-assembly/muscatassembly
- [15] CODATA2019 conference: http://codata2019.csp.escience.cn/dct/page/65586
- [16] The African Open Science Platform: https://codata.org/initiatives/strategic-programme/african-open-science/
- [17] La Referencia: http://www.lareferencia.info/en/institutional/about-us
- [18] RedCLARA: https://www.redclara.net/index.php/en/
- [19] EOSC working groups and the outcomes: https://www.eoscsecretariat.eu/eosc-governance/eosc-executive-board-outputs
- [20] RDA Global Open Research Commons Interest Group: https://www.rd-alliance.org/groups/global-open-research-commons-ig
- [21] Japan Cabinet Office: https://www.cao.go.jp/en/about.html
- [22] G7 open science working group action point: https://www.zbw-mediatalk.eu/2016/11/the-g7-open-science-working-group-action-points-speeding-up-open-science/
- [23] Towards a Global Open Science Commons workshop: https://rd-alliance.org/towards-global-open-science-commons-rda-11th-plenary-bof-meeting
- [24] Delivering a Global Open Science Commons workshop: https://www.scidatacon.org/IDW2018/sessions/252/
- [25] Coordinating Global Open Science Commons workshop: https://www.rd-alliance.org/bof-coordinating-global-open-science-commons-initiatives-rda-13th-plenary-meeting
- [26] Launch of the GORC Interest Group: https://www.rd-alliance.org/open-science-commons-interest-group-working-group-0
- [27] BoF-Global Open Research Commons International

Benchmarking: https://www.rd-alliance.org/global-open-research-commons-international-benchmarking-bof
- [28] Canadian NDRIO: https://engagedri.ca/
- [29] U.S. National Research Cloud https://hai.stanford.edu/national-research-cloud-joint-letter
- [30] NCRIS: https://www.ncris-network.org.au/
- [31] CNIC: http://english.cnic.cas.cn/
- [32] ALBA, https://www.cells.es/en/
- [33] Synchrotron SOLEIL, https://www.synchrotron-soleil.fr/fr
- [34] Diamond Light Source, https://www.diamond.ac.uk/
- [35] Advanced Photon Source, https://www.aps.anl.gov/
- [36] European Synchrotron Radiation Facility (ESRF), https://www.esrf.fr/
- [37] German Electron Synchrotron DESY, https://www.desy.de/
- [38] Spring8, http://www.spring8.or.jp/
- [39] Spallation Neutron Source (SNS), https://neutrons.ornl.gov/sns
- [40] ILL, https://www.ill.eu/
- [41] European Spallation Source (ESS), https://europeanspallationsource.se/
- [42] China Advanced Research Reactor (CARR), http://www.ciae.ac.cn/eng/carr/index.htm
- [43] PaNOSC, https://www.panosc.eu/
- [44] ExPaNDS, https://expands.eu/
- [45] ESFRI Science Cluster Projects: ENVRI_FAIR, EOSC-Life, ESCAPE, PaNOSC, and SSHOC
- [46] EISCAT: https://eiscat.se/
- [47] EISCAT data portal: https://marketplace.eosc-portal.eu/services/eiscat-data-access-portal-3bac4832-dec5-46d8-8107-bdf7c1ec1924
- [48] EISCAT B2SHARE metadata catalogue: https://b2share-testing.fz-juelich.de/communities/EISCAT
- [49] EISCAT Madrigal Database, https://madrigal.eiscat.se/madrigal/
- [50] SYISR, 10.1109/TGRS.2020.3008427
- [51] APAN, the Asia Pacific Advanced Network, https://apan.net/
- [52] ISGC: International Symposium on Grids and Clouds
- [53] ASEAN, the Association of Southeast Asian Nations, https://asean.org/
- [54] IVOA, https://ivoa.net/
- [55] NOAO Data Lab, https://datalab.noao.edu/, operated by the Association of Universities for Research in Astronomy (AURA).
- [56] CANFAR, https://www.canfar.net/
- [57] SciServer, https://www.sciserver.org/
- [58] SkyServer, http://skyserver.sdss.org/dr16/en/home.aspx
- [59] NADC, https://nadc.china-vo.org
- [60] Copernicus, https://www.copernicus.eu/, Europe's eyes on Earth.
- [61] Landsat Missions, https://www.usgs.gov/core-science-systems/nli/landsat, soon will be NISAR.
- [62] ALOS, https://www.eorc.jaxa.jp/ALOS/a/en/index_e.htm, Advanced Land Observing Satellite
- [63] d'Andrimont et al, "From parcel to continental scale – A first European crop type map based on Sentinel-1 and LUCAS Copernicus in-situ observations", Rem. Sens. Env., 266, 2021, 112708, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2021.112708.
- [64] Landsat Missions, https://www.usgs.gov/core-science-systems/nli/landsat
- [65] https://ghsl.jrc.ec.europa.eu/
- [66] https://global-surface-water.appspot.com/
- [67] DIAS, https://www.copernicus.eu/en/access-data/dias
- [68] Digital Agenda for Europe, https://ec.europa.eu/digital-single-market/en/shaping-digital-single-market
- [69] Destination Earth, https://digital-strategy.ec.europa.eu/en/policies/destination-earth
- [70] Europe Data Space, https://digital-strategy.ec.europa.eu/en/policies/strategy-data
- [71] OnGeoChemistry, https://www.earthchem.org/communities/onegeochemistry/
- [72] COPDESS, https://copdess.org, is now supported by the Earth Science Information Partners Federation (ESIP).
- [73] The African National Integrated Cyber Infrastructure (NICIS), https://nicis.ac.za/
- [74] The South African National Research Network (SANReN):

http://sanren.ac.za
- [75] The Data Intensive Research Initiative of South Africa (DIRISA): https://www.dirisa.ac.za/
- [76] Center for High Performance Computing (CHPC): https://www.chpc.ac.za/
- [77] The SADC Cyber-Infrastructure Framework: https://datascience.codata.org/articles/10.5334/dsj-2019-034/
- [78] WACS (West Africa Cable System): https://en.wikipedia.org/wiki/West_Africa_Cable_System
- [79] SEACOM: https://www.submarinenetworks.com/systems/asia-europe-africa/seacom#:~:text=SEACOM%20is%20a%2017%2C000%20km,1.5Tbps%20 capacity%20in%202018.
- [80] The Lengau HPC System: https://www.hpcwire.com/off-the-wire/chpc-unveils-lengau-supercomputer/
- [81] ARDC, https://ardc.edu.au/
- [82] Identified in the 2016 National Research Infrastructure Roadmap, the DDeRP includes Australian Research Data Commons (ARDC), Australia's Academic and Research Network (AARNet), the Australian Access Federation (AAF), National Computational Infrastructure (NCI), Pawsey Supercomputing Centre (the tier one high performance computing facilities).
- [83] The Nectar Cloud: https://ardc.edu.au/services/nectar-research-cloud/
- [84] Australian BioCommons, https://www.biocommons.org.au/
- [85] CSTCloud, www.cstcloud.cn
- [86] FAST, https://fast.bao.ac.cn/
- [87] LHAASO, http://english.ihep.cas.cn/lhaaso/
- [88] OpenAIRE Guidelines: https://guidelines.openaire.eu/en/latest/
- [89] OpenAIRE Service Catalogue: http://catalogue.openaire.eu/
- [90] OpenAIRE Research Graph: https://graph.openaire.eu/
- [91] Numbers in 2020.
- [92] Bella, https://bella-programme.redclara.net/index.php/en/
- [93] NORDUnet, https://www.nordu.net/
- [94] SUNet, https://www.sunet.se/
- [95] CERNET, http://www.edu.cn/english/
- [96] Horizon Europe program, https://ec.europa.eu/info/research-and-innovation/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en
- [97] The EOSC Secretariat: https://www.eoscsecretariat.eu/
- [98] https://www.nsf.gov/about/budget/fy2014/pdf/45_fy2014.pdf
- [99] Launch of CODATA GOSC Working Groups and Case Studies, https://codata.org/initiatives/strategic-programme/global-open-science-cloud/

License

This pager is licensed under a Creative Attribution-NonCommercial-ShareAlike 4.0 international License (https://creativecommons.org/licenses/by-sa/4.0/)

The Global Open Science Cloud Landscape