GOSC Event 'Introduction and Launch of Case Studies and Working Groups': Objectives and Agenda

On 28 June, we invite you to a Global Open Science Cloud (GOSC) event: 'Introduction and Launch of Case Studies and Working Groups'.

The objective of this event is to introduce the GOSC initiative and to invite participation in the initial Case Studies and Working Groups. We hope to come out of the meeting with agreements to participate (and expressions of interest) from both invited participants and from other experts.

Contents of this Document

Programme for the 28 June Meeting

The draft programme for the meeting on 28 June is as follows:

Meeting Co-Chairs: John Broome, Officer and Treasurer, CODATA; and Mark Leggott, Chair, CODATA International Data Policy Committee and RDA Council

Time	Activity
10:00-10:10 UTC	Opening remarks (Prof. Barend Mons, President CODATA and Director GO FAIR; WANG Zhenyu, Deputy Director General, Bureau of International Cooperation, CAS (TBC).)
10:10-10:25 UTC	Introduction to the GOSC Initiative (Prof. LI Jianhui, CNIC)
10:25-10:40 UTC	GOSC Thematic Working Groups (Dr. Simon Hodson, CODATA)
10:40-11:10 UTC	Discussion (Moderated by meeting co-chairs)
11:10-11:25 UTC	Case Studies Intro and Discussion (Paul Uhlir, Consultant, and Lili ZHANG, CNIC and CODATA International Data Policy Committee)
11:25-11:55 UTC	Discussion (Moderated by meeting co-chairs)
11:55-12:00 UTC	Closing remarks and adjourn

Objectives and Approach of GOSC

The Global Open Science Cloud (GOSC) initiative will encourage cooperation, alignment, and ultimately interoperability, between existing and emerging Open Science Clouds (OSCs). GOSC aims to do this through a combination of thematic Working Groups (building on existing experiences and addressing key areas of shared interest), and a set of detailed Case Studies that will demonstrate how international collaborative research communities and projects can be supported by Open Science Clouds. At the introduction event, we invite volunteers as co-chairs and participants for the Working Groups and Case Studies.

Working Groups

The thematic Working Groups (WGs) should address areas of potential alignment and interoperability. The following initial WG topics are proposed:

- Strategy, governance and sustainability.
- Policy and legal.
- Technical infrastructure.
- Data interoperability.

Initial descriptions of the proposed topics to be addressed by the WGs are provided below. We invite expert participants who are involved in OSCs as well as additional experts who might not be directly involved in an OSC, but who would contribute valuable insights and knowledge. Procedures and timescales for the WGs are described below. There will be a mechanism for suggesting additional WGs.

Case Studies

Topics of the initial Case Studies have been suggested by domain experts from communities close to CODATA in cooperation with the GOSC Steering Group. The following initial Case Studies are proposed:

- Incoherent scatter radar data fusion and computation.
- Biodiversity and ecology information platform.
- SDG-13 climate change and natural disasters.
- Sensitive data federation analysis model in population health.

Initial descriptions of the proposed topics to be addressed by the Case Studies are provided below. We invite expert participants who are involved in OSCs as well as additional experts who might not be directly involved in an OSC, but who would contribute valuable insights and knowledge. Procedures and timescales for the Case Studies are described below. There will be a mechanism for suggesting additional Case Studies.

Operation of the Working Groups and Case Studies

The WGs and Case Studies will be composed, on a voluntary basis, of experts from the various OSC initiatives, as well as other experts in the field. It is intended that each WG and Case Study should have 2-4 co-chairs. The initial lifespan for the WGs and Case Studies will be two years, after which they may be renewed, redirected or retired according to progress. It is felt that the matrix approach (thematic WGs, community-focussed Case Studies) is a useful way of breaking down the problem space and important to have focussed and usable outputs. Nevertheless, it is also clearly important to ensure that there are regular opportunities for communication and engagement among the WGs and Case Studies. This will be insured by open working practices and regular opportunities for joint activities (see the timeframes below).

Initiating the Working Groups

To initiate the WGs, we propose to:

- Invite at least one representative from each of the OSCs to each WG.
- Invite additional experts, stakeholders (as necessary and where additional input may be needed): e.g. members of CODATA Data Policy Committee; experts on e-infrastructure or on metadata, semantics etc.
- Allow interested and expert volunteers to join the WGs.
- Invite volunteers as co-chairs for the WGs.
- Assemble preliminary membership and co-chairs following the June launch meeting and start regular virtual meetings.
- Aim in early discussions to narrow down focus and identify an early set of tractable objectives and outputs, which will be presented and discussed at virtual SciDataCon in Oct/Nov 2021.

Initiating the Case Studies

To initiate the Case Studies, we propose to:

- Invite the OSCs to nominate appropriate experts and researchers to each Case Study.
- Invite additional experts, stakeholders (as necessary and where additional input may be needed): i.e. experts on the particular topics that may not be associated with an OSC initiative.
- Allow interested and expert volunteers to join the Case Studies.
- Invite volunteers as co-chairs for the Case Studies.
- Assemble preliminary membership and co-chairs following the June launch meeting and start regular virtual meetings.
- Aim in early discussions to narrow down focus and identify tractable objectives and outputs, which will be presented and discussed at virtual SciDataCon in Oct/Nov 2021.

Timeframes for the WGs and Case Studies

WGs and Case Studies should hold regular virtual meetings and should proceed with their work and the preparation of outputs, with a view to the following milestones. These are based around events at which WGs and Case Studies will be asked to report on progress and will have the opportunity to hold workshops to advance their activities.

The following schedule of activity, reporting and public discussion of progress is proposed:

1. From 28 June 2021: WGs and Case Studies formed, co-chairs identified and approved by the GOSC SG, outputs scoped.
2. Oct-Nov 2021: GOSC Workshops as part of fully virtual SciDataCon 2021. WGs and Case Studies discuss and refine their mission, objectives, planned outputs and

progress.

3. 19-21 Oct 2021: EGI Conference (fully virtual), WGs and Case Studies (where on appropriate topics) report on and communicate activities to global eInfrastructure community.
4. June 2022: GOSC Workshops as part of International Data Week 2022, Seoul, Republic of Korea. WGs report on progress and outputs.
5. Sep 2020: International Symposium on Open Science Cloud, Beijing, China. WGs reports on progress and outputs.
6. Oct 2022: GOSC Workshops as part of 2nd FAIR Convergence Symposium, Leiden, Netherlands. Selected presentations on WGs progress.
7. Q4 2022: EGI Conference, report on progress and outputs to global eInfrastructure community
8. Q2 2023 (TBC): WG report on final outputs as part of first International Symposium on Open Science Clouds. WGs renewed, redirected or retired.

Additional Observations

- The WGs and Case Studies do not have to be symmetrical in size, activities and organisation. They should do what is useful for each topic and what works for the people involved.
- It will be better to identify small topics for cooperation and alignment, than aiming to address the whole range of issues suggested by each theme.
- In doing so, the WG or Case Study should identify tractable and useful outputs. The precise scope and nature of these is for the WG or Case Study to determine. Examples include recommendations for alignment between OSCs.
- Among the objectives of the WGs is simply to establish a forum for cooperation on these important issues.

Initial Working Groups

The topics for the WGs are described in more detail below. As appropriate the WGs will build on existing work, with a particular focus on issues of cooperation and alignment between OSCs in relation to each of these topics.

Strategy, Governance and Sustainability

As large, multidisciplinary infrastructures, OSCs around the world will face challenges of governance, funding, sustainability, rules of participation, inclusivity, making their value proposition, and engagement with researchers and existing research infrastructures and other e-infrastructures. The specificities of national and regional circumstances mean that the approaches of each OSC to these issues will often be relatively bespoke. Nevertheless,

there is much to be learnt from the exchange of experiences and strategies between initiatives. The WG will likely proceed by providing a forum to exchange information and identifying a discrete number of topics to explore in more detail. Possible initial questions include but are not limited to: What are the governance and accountability structures? What are the mechanisms for funding core OSC infrastructures and how does the OSC present its value proposition in order to sustain operations? What are the rules of engagement for public and private institutions?

Initial Contacts

Yu CHEN, CNIC, cy@cnic.cn
Hana Pergl, CODATA, hana@codata.org
Additional participants are invited: contact info@codata.org

Policy and Legal

Statements of principle and policy instruments at national and transnational levels have sought over the last 20 or so years to encourage and support emerging Open Science practices and to create a 'level playing field' in which scientific outputs are Open by default. Regional policy, as in the European Union, and international instruments, such as the recent OECD Declaration on Access to Research Data from Public Funding and the forthcoming UNESCO Recommendation on Open Science, are particularly significant. Different national concerns affect data policy priorities: nevertheless, to avoid the development of large silos and to achieve the potential of global Open Science, there is interest among OSCs to seek alignment of their policies, and where possible their legal frameworks. The WG will likely proceed by exchanging information about Open Science policies before exploring any areas of divergence. Possible questions include but are not limited to: What are the key policy instruments adopted by the OSCs? To what extent do these align and do they map well to major international statements of principle or policy? Are there any areas of policy misalignment that could hamper cooperation among OSCs and projects supported by OSCs? What are the most significant issues of legal interoperability to be addressed?

Initial Contacts

Mark Leggott, CODATA Data Policy Committee, RDA Council, Research Data Canada
mark.leggott@rdc-drc.ca
Lili ZHANG, CNIC, zhll@cnic.cn
Additional participants are invited: contact info@codata.org

Technical Infrastructure

Open Science Clouds necessarily build on existing global e-infrastructures, including provision

of high-speed academic networks, high performance computing, cloud computing, and so on. Among the objectives is to facilitate connectivity among international e-Infrastructures at national, regional, and global level, provide lightweight federation solutions and improve interoperability between different systems, and support international research collaborations. There are numerous important topics for e-infrastructure cooperation, alignment and interoperability, including: network connectivity and protocols, secure Authentication and Authorization Infrastructure (AAI), and mechanisms for federation of computing and other services.

Initial Contacts

LI Jianhui, CNIC, lijh@cnic.cn
Yin CHEN, EGI, yin.chen@egi.eu
Mark Dietrich, EGI, mark.dietrich@egi.eu
Additional participants are invited: contact info@codata.org and yin.chen@egi.eu

Data Interoperability

One important objective for OSCs is to enhance support for research that addresses the fundamental challenges of our age (including global sustainability, disaster risk reduction and so on). Such research topics often require an interdisciplinary approach and the ability to combine data from across traditional domain boundaries. Many OSCs explicitly aim to support and enhance the services provided by established Research Infrastructures, while also seeking to break down the silos that may inhibit data sharing and interoperability. The FAIR principles provide a framework for convergence and a number of topics can usefully be addressed to pursue alignment and interoperability among OSCs. These include but are not limited to: the emerging FAIR Digital Object Framework; the use of structural and provenance metadata to facilitate machine-actionability across data; and the alignment and development of good practice for semantic artefacts (including scientific vocabularies). The EOSC Interoperability Framework may provide a good starting point for discussions around how to align and how OSCs can contribute and engage with global efforts to address the I (interoperability) and the R (reusability) of FAIR.

Initial Contacts

Simon Hodson, CODATA, simon@codata.org
Additional participants are invited: contact info@codata.org

Initial Case Studies

The following initial Case Studies have been proposed and work has started to instigate them. We invite participation in each of these case studies. There will also be a mechanism for suggesting additional Case Studies.

Incoherent scatter radar data fusion and computation

Introduction

The EISCAT-3D radar, https://eiscat3d.se/, is a next-generation incoherent scatter radar system operated by the EISCAT association. It is a world-leading infrastructure using the incoherent scatter technique to study the atmosphere in the Fenno-Scandinavian Arctic and to investigate how the Earth's high latitude atmosphere is coupled to space.

The Sanya Incoherent Scatter radar (SYISR) is also a next-generation incoherent scatter radar system under design and construction by the Institute of Geology and Geophysics, Chinese Academy of Sciences (IGGCAS). It will use the incoherent scatter technique to study the atmosphere at the low latitude in the East Asia region and to investigate how the Earth's low latitude atmosphere is coupled to space.

Both radars aim to establish a system of distributed phased array radars that enable comprehensive three-dimensional vector observations of the atmosphere and ionosphere. The use of new radar technology, combined with the latest digital signal processing, will achieve ten times higher temporal and spatial resolution than obtained by present radars while simultaneously offering, for the first time, continuous measurement capabilities. Both systems will allow the study of atmospheric phenomena at both large and small scales unreachable by the present systems. The new system will be implemented for a wide range of users and applications. The continuous data coverage will facilitate the inclusion of detailed incoherent scatter radar data into climate and Earth system modelling.

Significance of the Case Study

Free flow for massive research data is one of the typical use scenarios within different subjects. The EISCAT-3d and SYISR radar data use case focuses more on the technical interoperability of the data. The large scale of data that will be transferred and managed online provides the GOSC Initiative with an excellent scenario in which many cloud federation techniques can test their deployments and iterate further developments.

Research challenges and requirements for GOSC

The challenge of EISCAT-3D and SYISR data management system is how to handle large-scale experimental data that will be massively generated at great speeds and volumes. During its first operation stage in 2018, EISCAT-3D will produce 5PB data per year, and the total data volume will rise up to 40PB per year in its full operations stage in 2023. While for SYISR, during its first operation stage in 2021, SYISR will produce 1PB data per year, and the total data volume will rise up to 3PB per year in its full operations stage in 2023.

In this case, the EISCAT and SYISR radar data fusion and computing may require further

technical supports from the GOSC Initiative within the following aspects:

1. Secure check-in services for accessing cross-border cloud services
2. DIRAC for job submission
3. Radar data storage and federated processing
4. On-demand data movement

Engagement with the GOSC Initiative

Platform and semantic interoperability are two key issues to be addressed in EISCAT-3D & SYISR radar data cooperation. Under the umbrella of the GOSC initiative, joint activities for EISCAT and SYISR may include:

1. (Meta)data federation: Metadata searchable via an integrated portal, with data stored in the EISCAT repository and the SYISR repository. Such a search portal can be based on the EISCAT Portal (by either configuring the EISCAT Portal to be able to access the SYISR metadata, or by hosting a customized instance of the EISCAT Portal in the CNIC cloud) and configuring it to be able to access the SYISR and EISCAT metadata.
2. Federated processing: After discovering relevant EISCAT and SYISR data, the user could launch analysis jobs in the cloud to process those data. These jobs would run in the European or the Chinese clouds depending on the user's affiliation (e.g., Chinese users would access the CNIC portal and run jobs in the CNIC cloud, others would use the EISCAT portal and run jobs in the EGI/EISCAT cloud). Other clouds could also be included when necessary.
3. On-demand data movement: If the data must be moved, the user would be able to move data from EISCAT and SYISR into a single cloud site and perform online analysis with both sets.
4. Collaborations on necessary training and activities for SYISR and EISCAT community outreach.

Possible deliverables

After the implementation of the five-year GOSC program, the anticipated outputs from the EISCAT-3D and SYISR Case Study may focus on the following aspects:

1. Exploration of technical solutions for EISCAT and SYISR (Meta)data federation.
2. Deployment of GOSC technical solutions supporting the cloud federated data processing and on-demand data movement.
3. Lessons and good practices for analogous research infrastructures in OSCs.

Initial Contacts

Ingemar Haagstrom, EISCAT Science Association, Sweden, ingemar.haggstrom@eiscat.se
Xinan Yue, Institute of Geoscience and Geophysics, CAS, China, yuexinan@mail.iggcas.ac.cn
Kaichao WU, CNIC, CAS, kaichao@cnic.cn

Yin Chen, EGI, yin.chen@egi.eu
Additional participants are invited: contact info@codata.org

Biodiversity and Ecology: an open cloud service for camera trap data management and intelligent analysis

Introduction

In the past 10 years, infrared trigger camera and camera trap technology have been widely used in nature reserves because of its obvious advantages over traditional survey methods, especially for the monitoring and research on the terrestrial large and medium-sized animals. According to incomplete statistics, camera trap technology has been used for wildlife monitoring and research in 80% of China's national nature reserves, and images and videos of more than 200 rare and endangered species have been obtained.

Followed by a large number of camera deployments and massive amounts of image and video data are two challenges to a nature reserve. One is how to effectively store, organize and manage massive amounts of camera trap data. The other one is how to intelligently analyze massive amounts of camera trap data. To help nature reserve solve these challenges, the Chinese Biodiversity Observation Network – Mammals (Sino BON-Mammals) and Computer Network Information Center (CNIC), CAS developed an open cloud service for camera trap data management and intelligent analysis. Based on the cloud service, nature reserves will be able to create sample areas, add sample points, and upload camera trap data in a self-service manner. On the cloud side, the uploaded camera trap data is stored and managed, and through object detection and species identification, image and video data are intelligently analyzed based on deep learning technology.
We seek other biodiversity and ecology related initiatives to participate in this Case Study, particularly if focussed on the technological challenges listed here and the integration and enrichment of biodiversity data.

Significance of the Case Study

The capabilities of on-demand provision of cloud service can help nature reserve store and manage massive amounts of camera trap data. Intelligent analysis and visualization can significantly improve the data analysis capabilities of the nature reserve. Through object detection and species identification based on deep learning technology, the service can help reserve filter out invalid data, implement automatic identification of animal species, and greatly reduce the workload of researcher and reserve staff. It also provides effective support for subsequent analysis like wildlife distribution and active rhythms, etc. This open cloud service will also be a collaborative platform for global camera trap data sharing and analysis service and contributes to global biodiversity research and protection.

Research challenges and requirements for GOSC

The cloud service involves cloud computing, big data, artificial intelligence, data visualization, and other technologies, and requires interdisciplinary cooperation and multi-party coordination. Challenges include on the following issues:

1. Stewardship of massive amounts of camera trap data storage, movement, management methods, and mechanisms under cloud environment.
2. Deep learning algorithms and methods to improve intelligent analysis performance.
3. Workflow or pipeline to implement camera trap data management, intelligent analysis, and visualization process automation.
4. Mechanisms for integrating cloud computing, big data, artificial intelligence, and visualization technologies into a cloud service.
5. Data standard specification, service integration standard, and multi-party collaboration mechanism.

Engagement with the GOSC Initiative

This case deals with distributed big data management, intelligent analysis, and cloud computing for high-quality integration and optimization for camera trap data uploaded by a nature reserve. The cloud service will be a collaborative platform for global camera trap data sharing and analysis service and contributes to global biodiversity research and protection.

Joint work under the GOSC Initiative umbrella may include:

1. Exploring alignment of the infrastructure supporting camera trap data storage and AI computing, for example, cloud storage space and GPU resources for AI models training and prediction, etc.
2. A toolkit offering online cloud services of various modular functions, which may include but not be limited to real-time image processing, intelligent recognition, and other functions yet to be determined.
3. Collaborations on necessary standards and specifications, training of data scientists and data managers, and other activities promoting the deployment of interoperable services in nature reserves around the world.

Possible deliverables

Anticipated outputs from this Case Study may focus on the following aspects:

1. A software system for camera trap data management and intelligent analysis.
2. Development of an online computing and processing toolkit for big data acquired from biosphere reserves.
3. A specification for biosphere reserves data systems, comprising standards for data and metadata, for monitoring equipment, and for applications.
4. Lessons and good practices for analogous cloud services.

Initial Contacts

Ze LUO, Scientific Research Monitoring Information Platform for Nature Reserves, CNIC, CAS, luoze@cnic.cn
Zhishu XIAO, Institute of Zoology, CAS, China
Additional participants are invited: contact info@codata.org

SDG-13 climate change and natural disasters

Introduction

CASEarth for Sustainable Development Goals (CASEarth4SDGs), www.sdgs.casearth.cn, is a platform system of data sharing and online computing for monitoring, measuring, and evaluating SDG indicators under the Big Earth Data Science Engineering Program (CASEarth). Once assembled, it will provide a tool to generate knowledge from numerous and complex data sources, with the objective of supporting and understanding a sustainable human society that is essential to the protection of the planet. This Case Study concentrates on SDG-13, which focuses on climate change and natural disasters.
We seek other SDG-13, climate change and disaster risk reduction related initiatives to participate in this Case Study.

Significance of the Case Study

The sharing of research data across numerous disciplines is vitally important for decision making in the assessment of and response to the effects of climate change and natural disasters. Climate change and its effects are increasingly an existential crisis for all humanity and requires concerted international cooperation to help resolve. This Case Study focuses on SDG-13 and aims to:

1. establish an integrated professional database of climate change, extreme climate and natural disasters;
2. develop a dynamic monitoring and risk assessment model system;
3. study the large-scale patterns and heterogeneity of climate change and natural disasters;
4. achieve high-precision comprehensive assessment and seasonal prediction on the temporal changes and spatial patterns of extreme climate events and natural disasters; and
5. explore the causes and frequency of regional extreme climate and related natural disasters, providing science-based support for disaster mitigation.

This demonstration of data uses for SDG-13 and the related technical and policy study will cast light on the extreme climate induced disasters in various geographies and countries.

Research challenges and requirements for GOSC

During its first stage of operation in 2021, CASEarth4SDGs programme will produce 3 Petabytes (PB) of data. The total data volume will increase to 10PB during its full operational stage in 2023. As one of its main components, the research work of SDG-13 will produce 1.2 PB data focusing on:

1. temporal and spatial patterns of climate change;
2. collection and sharing of research data on extreme climates and disasters; and,
3. short-term forecasting and seasonal prediction of climate disasters, as well as the monitoring and assessment of natural disasters.

Challenges include how to handle 'in real-time' high-resolution and multi-source Big Earth data at global, regional, and national scales that will be generated at great speeds and volumes.

Engagement with the GOSC Initiative

This case study will address the interoperability (technical, semantic, and policy) and other data management issues using high-performance computing facilities by exploring models for cross-border data exchanges for decision-making and multi-source data analytics. Because the CASEarth program and this case study intersect closely with the Digital Belt and Road (DBAR) Initiative, some partners have already been identified from established DBAR International Centers of Excellence (ICOEs) that are doing work in SDG-13. **Other partners on the Case Study are invited.**

The three ICOEs initially identified include ICOE Bangkok, ICOE Zambia, and ICOE Morocco. While the ICOE Bangkok is focused on climate change, the other two ICOEs are concerned with agriculture and water, respectively, which are two well-known disaster areas affiliated with climate change. Joint work under the GOSC Initiative umbrella may include:

1. Secure check-in data services for accessing cross-border services for improving modelling, simulation, and prediction.
2. Data sharing model(s) and federated data activities.
3. Collaborations on necessary activities for community engagement, involving CASEarth4SDGs and other initiatives, and with a focus on SDG-13.

Possible deliverables

Anticipated outputs from this Case Study may focus on the following aspects:

1. Metadata and database federation, between CASEarth4SDGs and other initiatives, especially focusing on SDG-13.
2. Cooperative development of the online Computing and Processing Toolkit for SDG-13 indicators.

3. Exploration of cloud federation techniques supporting on-demand data processing and analysis for SDG-13.

Contacts

Gensuo Jia, Institute of Atmospheric Physics, CAS, jiong@tea.ac.cn
Lili ZHANG, CNIC, CAS, zhll@cnic.cn
Paul Uhlir, Consultant (US National Academy of Sciences, retired), pfuhlir@gmail.com
ICOE Bangkok: Monthip Sriratana, National Research Council of Thailand, Thailand, monthip2007@gmail.com
ICOE Zambia: Luke E. Mumba, lemumba2004@gmail.com
ICOE Morocco: Kamal Labbassi, labbassi@ucd.ac.ma
Additional participants are invited: contact info@codata.org

Sensitive data federation analysis model in population health

Introduction

Reusing Real World Observations (RWO) and health data for research, health innovation and policy is key to better health in general, pandemic preparedness and imminent cost savings. However, the generally accepted notion that 'citizens should be in control of the reuse of their personal data' remains a paper mantra unless we design and implement a user friendly, trusted and sustainable environment that allows the realisation of that ambition. Performing GDPR compliant research will be entirely dependent on solving the trusted data federation challenge.
We seek projects and initiatives working on population health data, clinical data, and genetic data to participate in this Case Study.

Significance of the Case Study

This Case Study will explore how innovative technology can change the current methods in care and improve the crucial influence of healthy citizens and patients in an advanced globally interoperable health data system. The resulting technology will be relevant for future initiatives that need to reuse sensitive data of individuals. The current limited or non-existent level of reuse of critical data on infection, viral spread and post-vaccination as well as long term effects related to COVID-19 will severely hamper preparedness for future SARS-CoV-2 related problems, including the emergence of new variants. Reuse of sensitive RWO is of much wider use than just for COVID-19. Hence this Case Study aims to greatly contribute to the generic abilities of the global society to tackle future health issues.

Research challenges and requirements for GOSC

Better use of data and evidence based personalised medicine could potentially save society hundreds of billions of Euros each year. Successful and reusable components of a system enabling this societal health benefit and cost saving can be reused. Many public organisations and private companies have indicated keen interest in a GDPR compliant access to a real world, citizen and patient controlled system that can help them to learn from the data. The Case Study will adhere to joint principles of open protocols and standards, around which multiple vendor solutions can be operational to create a distributed, safe and scalable environment. Reuse can be for any purpose for which consent of the citizen is pre-recorded in the system, and can range from improved care outcome studies, to outbreak management, to vaccine follow up to drug rationalisation and virtual clinical trials or post marketing surveillance by the pharmaceutical industry.

Engagement with the GOSC Initiative

Controlled and globally distributed access to sensitive data is a very important issue for social and health sciences, particularly when topics with global dimensions. The Case Study will seek to demonstrate how to make better steps to make available those parts of the data which can be made shared. Commonly agreed FAIR implementation profiles will be created and FAIR data points established in various 'GOSC' regions. The feasibility for distributed analytics over datasets held in various regions will then be explored and demonstrated. Community accepted standards will be used throughout and only if needed new choices and challenges in terms of standards and technology will be addressed.

Possible deliverables

1. A FAIR-based system with optimal scaling potential and no vendor lock in, entirely based on FAIR Digital Objects.
2. Fully distributed and GDPR compliant analytics and learning with full respect for and actual involvement of the citizen.
3. FAIR Data Points in a number of locations, with synthetic (and if possible real world data) to demonstrate cross-regional-OSC re-use of sensitive data for analytical purposes.

Initial Contacts and Participants/Invitees

Simon Hodson, CODATA, simon@codata.org
Hana Pergl, CODATA, hana@codata.org
Barend Mons, CODATA, GO FAIR and LUMC
Erik Schultes, GO FAIR, erik.schultes@go-fair.org
LIU Lei, University of Fudan, liulei@fudan.edu.cn
Additional participants are invited: contact info@codata.org

Working Group Plans

Among their early activities, once co-chairs have been identified and members recruited, WGs should focus their activities and prepare a WG Plan which should be concise and practical. The primary purpose is to help the WG prioritise and organise its activities. They are also helpful in communicating the value and significance of the activities. We recommend that the WG Plans follow the following template.

Common Template for Working Group Plans

1. Mission Statement: presenting the case for the WG, why it is important and its overall goals.
2. Objectives: briefly stating the specific measurable objectives of the WG and how these contribute to the mission and goals.
3. Tasks: the specific activities that the WG will undertake to achieve its objectives.
4. Deliverables: concrete outputs and achievements.
5. Working Arrangements: The plan should state any agreements on way of working, including: regularity of virtual meetings, use of collaborative platforms, document management.
6. Resources: the plan should list the resources, including: Co-Chairs, Members, Secretariat Support.

Proposing Additional Case Studies

Suggestions of additional Case Studies are encouraged, in particular on topics of interest to communities from more than one Open Science Cloud/Platform. The main criteria for selection of new Case Studies will be the feasibility of the activity, its potential for demonstrating how OSCs may support particular research communities, and the likely value of the proposed outputs and recommendations.

The template below should be used to propose a new GOSC Case Study. The completed form should be sent by the proposer to info@codata.org. Following a review of scope and quality by the CODATA Secretariat it will be submitted to the GOSC Steering Group for consideration. Feedback will be shared with the Case Study proposer within ±30 days since the Case Study submission.

Common Template for the GOSC Case Studies

1. Name of the case study.
2. Description of the case study: who/where (which research projects), what (scope of research), why (do they need to share data / collaborate using an OSC?). Significance of the research issue. Why this Case is important for the a) immediate scientific

community, b) the broader research community, c) society, and d) as a GOSC pilot project.

3. Significance of the research issue. Why this case is important for the a) immediate scientific community, b) the broader research community, c) society, and d) as a GOSC pilot project.
4. Description of some example research questions that researchers are addressing and that have a clear societal benefit.
5. Data requirements for the case study.
    a. Description of the data area and dataset(s) that would be considered, including a) the name of the principal contact(s), b) the location of the project and the data cloud services, c) the relevant dates (when the project started, when the database was formed and placed in the cloud repository, and d) link(s) to more official information and the dataset(s).
6. Statement of the problem(s) that need to be addressed by GOSC, e.g.:
    a. Policy/legal interoperability (existing cloud data access and use policy (a summary and link to more info, if available).
    b. Platform interoperability.
    c. Semantic interoperability.
7. Key (likely/possible) deliverables.
8. Main contacts.