COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

GOSC Case Studies Introduction and Discussion

Paul Uhlir (Consultant, GOSC Initiative) and
Lili ZHANG (CNIC, CAS and CODATA IDPC member)

28 June 2021

International
Science Council

COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

Global Open Science Cloud (GOSC)
Case Studies

I. Background
II. Approach
III. Timelines
IV. Substance
V. Discussion

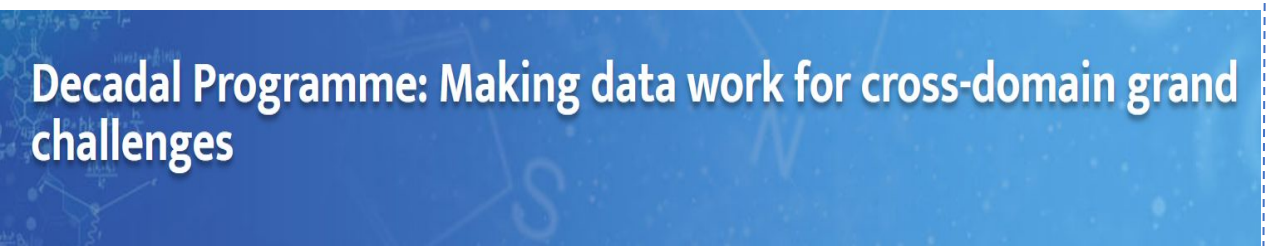
Background

Decadal Program: Making data work for cross-domain grand challenges

- Theme 1: Enabling Technologies and Good Practice for Data-Intensive Science.
- Theme 2: Mobilizing Domains and Breaking Down Silos
- Theme 3: Advancing Interoperability **Through Cross-Domain Case Studies**

GOSC Project Activities, Deliverables

- Policy alignment
- Governance and sustainability
- Technical Interoperability and Alignment
- Semantic Interoperability and Convergence
- **Demonstrators** and testbeds for convergence and interoperability
- **Case studies** and exemplar research activities facilitated by the GOSC



COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

Approach and procedures

Approach

Case Studies will be carried out to demonstrate how international collaborative research communities and projects can be supported by Open Science Clouds.

Objective

Demonstrations of typical research scenarios across subjects and regions supported by the GOSC testbed.

Procedures

Scenarios analysis - Case Studies – Draft plan for each Case Study – Implementation of Case Studies – [Additional Case studies]—Community Outreach

Timelines

2021.6- WGs and Case Studies formed, co-chairs identified and approved by the GOSC SG, outputs scoped.

2021.10-11 GOSC Workshops as part of fully virtual SciDataCon 2021. EGI Conference (fully virtual).

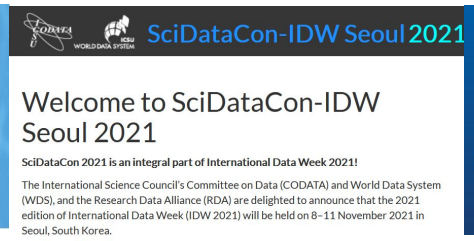
2022.6 GOSC Workshops as part of International Data Week 2022, Seoul, Republic of Korea.

2022.9 **International Symposium on Open Science Cloud, Beijing, China.**

2022.10 GOSC Workshops as part of 2nd FAIR Convergence Symposium, Leiden, Netherlands.

2022.12 EGI Annual Conference GOSC session (TBC)

2023.6 WG report on final outputs as part of International Symposium on Open Science Clouds. WGs renewed, redirected or retired.



Case Study Descriptions

Incoherent scatter radar data fusion and computation

Ingemar Haagstrom, EISCAT Science Association, Sweden
Xinan Yue, Institute of Geoscience and Geophysics, CAS, China
Kaichao Wu, CNIC, CAS, China
Yin Chen, EGI, Netherland

An open cloud service for camera trap data management and intelligent analysis

Ze Luo, CNIC, CAS, China
Zhishu Xiao, Institute of Zoology, CAS, China
And other international co-chair(s) (TBD)

SDG-13 climate change and natural disasters

Gensuo Jia, Institute of Atmospheric Physics, CAS, China
Li Wang, Aerospace Information Research Institute, CAS, China
Zhaohui Lin, Institute of Atmospheric Physics, CAS, China
And other international co-chair(s) (TBD)

Sensitive data federation analysis model in population health

Barend Mons, CODATA, GO FAIR and LUMC, Netherland
Lei Liu, University of Fudan, China
And other international co-chair(s) (TBD)

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

Case 1 Space Physics

Incoherent scatter radar data fusion and computation

EISCAT-3D radar, next generation incoherent scatter radar system, EISCAT association.

Sanya Incoherent Scatter radar (SYISR), next generation incoherent scatter radar, IGGCAS.

Use Case Scenarios

EISCAT & SYRSI (Meta)data federation

Federated processing

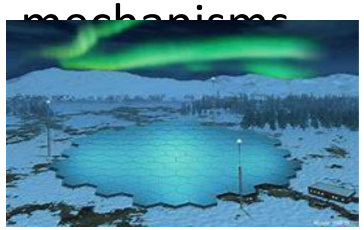
On-demand data movement

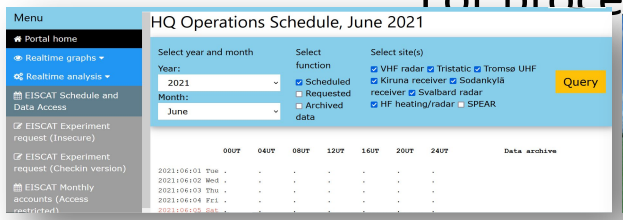
Key deliverables

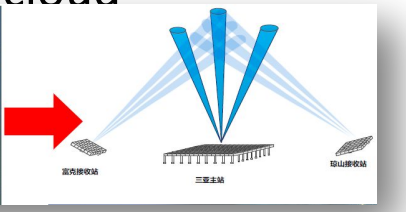
Cross-continental data migration and federated computing; high-speed and high-capacity experimental data processing; data sharing policies and mechanisms

EISCAT_3D/SYISR radar data (Häggström, 2020)

- Similar hardware
  Multistatic phased array radars
- Separate repositories
  Metadata federation
- Federation processing
  Launch jobs in EGI/CNIC clouds
- On demand data movement
  For processing in either cloud







This slide is prepared based on the "Potential collaboration and challenge between SYISR and EISCAT" by Ingemar Häggström and Xinan Yue (2020).

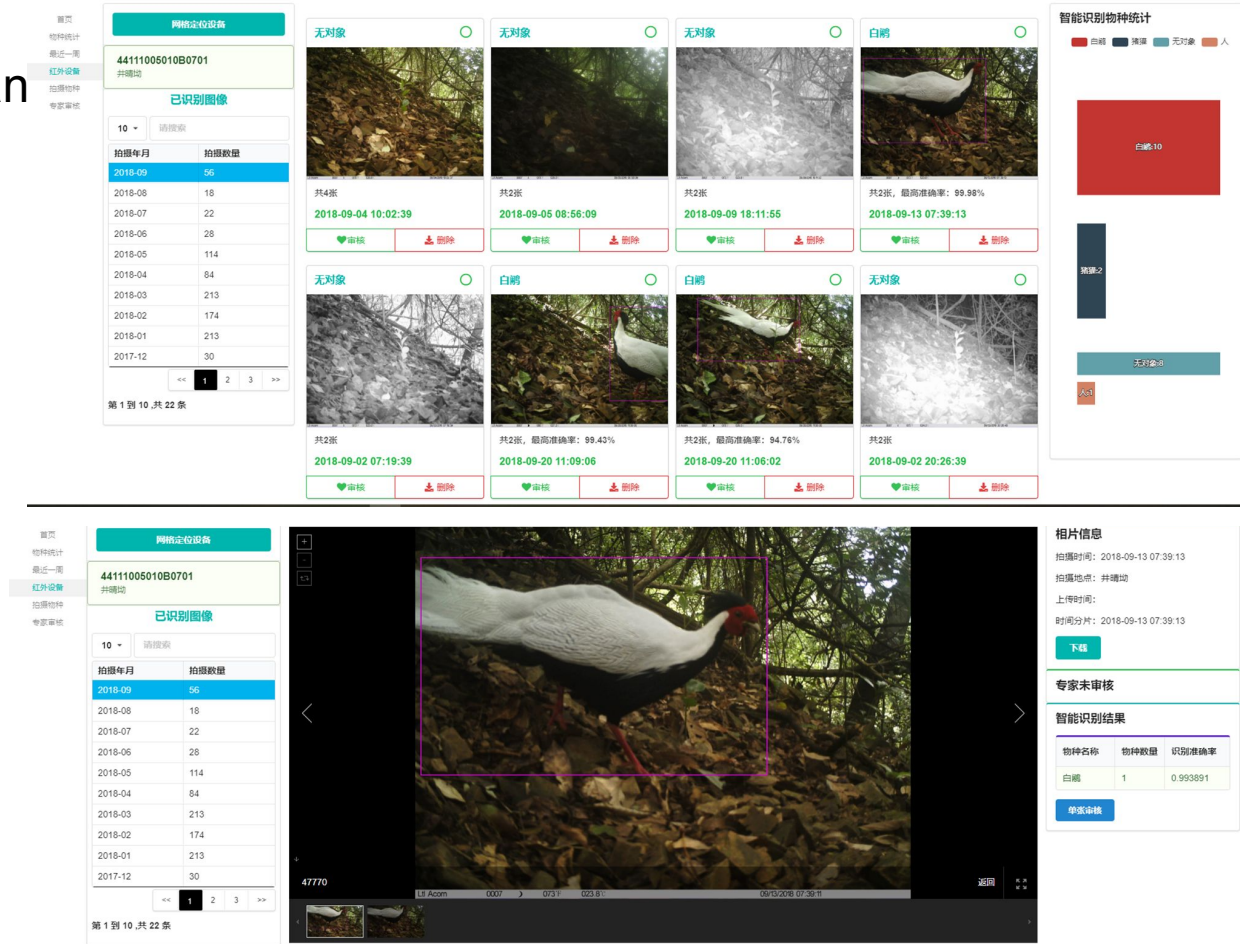
COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

Case 2: Biodiversity and Ecology

An open cloud service for camera trap data management and intelligent analysis

Introduction

Through deep learning technology, the service can help reserve filter out invalid camera trap data, implement automatic identification of animal species, and provide cloud storage for massive camera trap data.



Possible deliverables

- A cloud service for camera trap data management and intelligent analysis
- Software toolkits
- Lessons and good practices for analogous cloud service in GOSC

Case 3 Earth Sciences and Applications

SDG-13 for climate change and natural disasters

This Case Study concentrates on the UN Sustainable Development Goal, SDG-13, with focuses on climate change and natural disasters, especially temporal and spatial patterns of climate change; collection and sharing of research data on extreme climates and disasters; short-term forecasting and seasonal prediction of climate disasters, and monitoring and assessment of natural disasters.

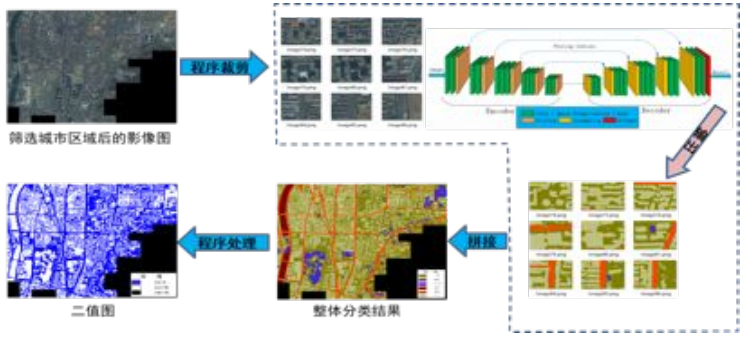
Possible deliverables

- Metadata and database federation.
- Cooperative development of the online Computing and Processing Toolkit for SDG-13 indicators.
- Exploration of cloud federation techniques supporting on-demand data processing and analysis for SDG-13.

We seek other SDG-13, climate change and disaster risk reduction related initiatives to join us.

www.sdgs.casearth.cn

Sharing of data and algorithm



Research & Decision making

COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

Case 4 Population Health

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.

Key deliverables

- A FAIR-based system with optimal scaling potential and no vendor lock in, entirely based on FAIR Digital Objects.
- Fully distributed and GDPR compliant analytics and learning with full respect for and actual involvement of the citizen.
- 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.

We seek projects and initiatives working on population health data, clinical data, and genetic data to participate in this Case Study.

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

How to engage

Proposal submission → **Case selection** → **Demo plan** → **Implement.** → **Community outreach**

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)** 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**.

COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

Thank you for your attention!