GOSC, DataIO and WorldFAIR:
What are we trying to achieve?

Simon Hodson
Executive Director
CODATA
www.codata.org

CODATA's mission and operation

- **The mission of CODATA is to "Connect data and people to advance science and improve our world".**

- As the 'Committee on Data of the International Science Council (ISC)', CODATA supports the ISC's mission of 'advancing science as a global public good' by promoting Open Science and FAIR data. CODATA convenes a global expert community and provides a forum for international consensus building and agreements around a range of data science and data policy issues, from the fundamental physical constants to cross-domain data specifications.

- **CODATA's membership includes national data committees, scientific academies, International Scientific Unions and other organisations.**

Making Data Work…

Data Policies

Data Science

Data Skills

- **Decadal Programme: Making Data Work for Cross Domain Grand Challenges**
- Recommendations on core interoperability and FAIR
- FAIR Vocabularies with ISUs
- Cross-Domain Case Studies
- Global Open Science Cloud initiative
- Regional Open Science Platforms

- International Data Policy Committee http://bit.ly/data-policy-committee;
- One major policy report per year.
- 20-Year Review of GBIF published in May 2020
- Preparing Independent Review of CAS Earth data policy and practices

- Data Science Journal: **https://datascience.codata.org/**
- International Data Week and CODATA Conference series.
- Task Groups and Working Groups.

- CODATA-RDA School of Research Data Science.
- Beijing and other training workshops.
- #terms4FAIRskills and FAIRsFAIR

Global Open Science

- **UNESCO Recommendation :** Calls on Member States to promote 'North-South, North-South-South and South-South collaborations to optimize infrastructure use and joint strategies for shared, multinational, regional and national open science platforms, including through the promotion of research collaborations, sharing of open science infrastructures, technical assistance, transfer and coproduction of technology related to open science and exchange of good practices under mutually agreed terms' (iii.g)

- **International Science Council Action Plan, 4.2:** encourages the creation of national or regional Open Science platforms in the Global South, in order 'to position scientists and science systems in the Global South at the cutting edge of data-intensive open science':
https://council.science/actionplan/open-science/

- Includes ISC and CODATA engagement with the African Open Science Platform, Malaysian Open Science Platform; ISC and CODATA support for the UNESCO Recommendation.

- **AOSP Vision and Strategy:** Vision and Strategy Document:
https://doi.org/10.5281/zenodo.2222418 ; Executive Director and Deputy Director appointed
https://www.nrf.ac.za/african-open-science-platform-aosp-new-nrf-appointments/

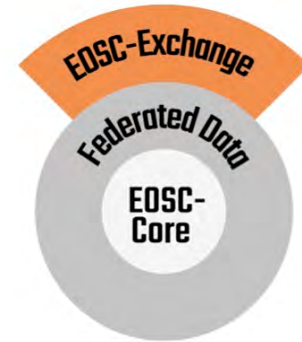- CODATA contributing through **OSCER** and **GOSC**.
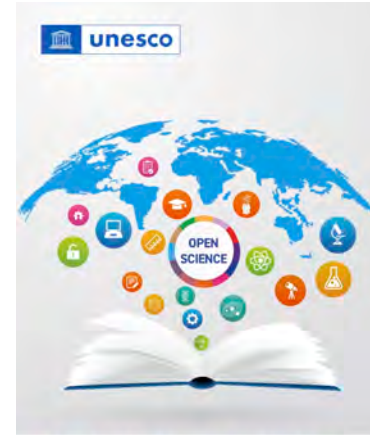
UNESCO Recommendation on Open Science
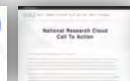
Global Open Science Cloud Initiative

- A number of Open Science/Research Clouds/Platforms/Commons emerging globally.

  - **EOSC** (European Open Science Cloud), **CSTCloud** (China Science and Technology Cloud), **ARDC** (Australian Research Data Commons), **Digital Research Alliance of Canada** (formerly NDRIO), **MOSP** (Malaysian Open Science Platform), **LA Referencia/Red Clara** (Latin America), **AOSP** (African Open Science Platform)…

- Advancing and supporting Open Science and FAIR, economies of scale, greater impact and RoI, more effective e-Infrastructures, greater realization of FAIR for established research domains and new cross-domain research areas.

- **Vertical alignment:** Bringing Open Science Infrastructures (NRENs, HPC, storage and other e-Infrastructures) closer to Research Infrastructures, research groups.

- **Horizontal interoperability:** Domain research infrastructures (WDCs, international and national domain data services and RIs, ESFRI process in Europe, Chinese data centres etc, Australian NCRIS… etc)

- **OSCER:** CODATA is providing secretariat support for the **Open Science Clouds Executive's Roundtable**. Periodic meetings between the leadership of these initiatives. **Summits in Seoul, June 2022; Leiden, October 2022.**

Schematic representation of the Minimal Viable EOSC, SRIA, p.73.

# Global Open Science Cloud (GOSC) Working Groups and Case Studies

- 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.

Four initial WGs, exploring key topics of shared concern:

1. Strategy, governance and sustainability.
2. Policy and legal.
3. Technical infrastructure.
4. Data interoperability.

Five initial Case Studies, exploring practical areas for data access across clouds:

1. Incoherent scatter radar data fusion and computation
2. Open reproducible raw diffraction data for access in pandemics
3. Biodiversity and ecology information platform
4. SDG-13 climate change and natural disasters
5. Sensitive data federation analysis model in population health
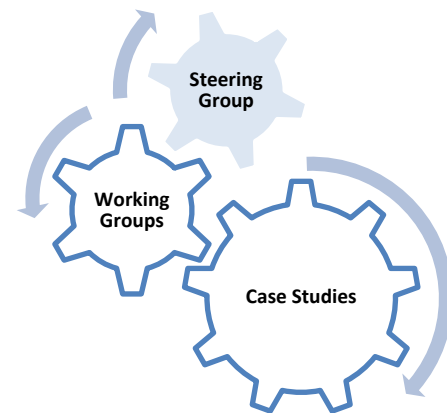
- GOSC Overview: https://bit.ly/GOSC-Overview
- Join GOSC WGs, Case Studies: https://bit.ly/3jwZHNg
- Propose New Case Studies: https://bit.ly/GOSC-Propose-New-Case-Study

GOSC Data Interoperability WG

- The GOSC DataIO WG seeks to explore, document and encourage good practice for the interoperability of data and metadata.

- The WG will work with Case Studies from GOSC and elsewhere to understand challenges, successes and emerging solutions.

  - Stakeholder analysis.

  - Survey of initiatives and solutions.

  - Detailed Case Studies.

- Recommendations on good practice and alignment across OSCs to address (data and metadata) interoperability.
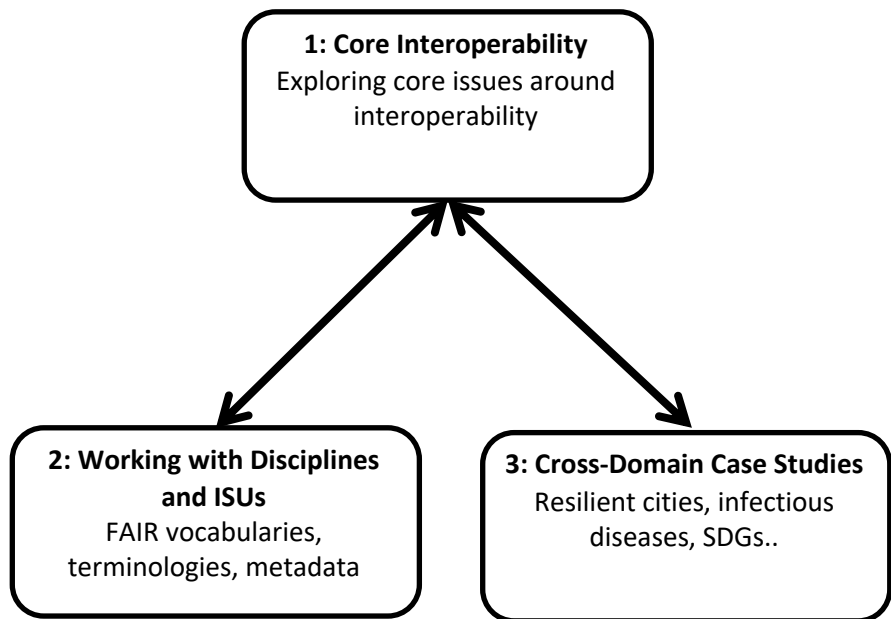
Suggested GOSC Working Group Timeline

1. **From 28 June 2021: WGs and Case Studies formed**, co-chairs identified
2. **From 2 Sept 2021, Co-Chairs Meeting:** prepare WG and Case Study work plans.
3. **27 Oct 2021: GOSC Workshops as part of fully virtual SciDataCon 2021.** WGs and Case Studies discuss their workplans: mission, objectives, planned outputs…
4. **1-2 June 2022: First GOSC AHM.** WGs and CSs report on progress and outputs, discuss alignment.
5. **20-23 June 2022: GOSC Sessions as part of SciDataCon / International Data Week 2022**, Seoul, Republic of Korea and online.
6. **24-26 Oct 2022 Second GOSC AHM / GOSC Workshops as part of 2nd FAIR Convergence Symposium**, Leiden, Netherlands. WGs and CSs report on progress and outputs.
7. **Oct 2023: WG report on final outputs as part of International Data Week, Salzburg, Austria. WGs and CSs renewed, redirected or retired.**

Making Data Work for Cross-Domain Grand Challenges

- **ISC Action Plan entrusts CODATA with an initiative 'Making Data Work for Cross-Domain Grand Challenges': establish a global (decadal) programme to address these issues.**

- **Calls for initiatives to support Open Science platforms.**

- The major, pressing global scientific and human issues of the 21st century can **ONLY** be addressed through **research that works across disciplines to understand complex systems**, and which uses **interdisciplinary and transdisciplinary** approaches to turn data into knowledge and then into action.

- **Preparatory Phase:** exploring technical issues and case studies through Dagstuhl workshops, TGs and WGs, funded projects etc, to understand the challenges and prepare the programme.

- **Core Interoperability Framework:** units, vocabularies, data structure, data description…

- **Case Studies:** in a range of domain and cross-domain research areas.

- **Global Open Science Cloud Initiative:** initiative launched in 2021, with Working Groups and Case Studies to promote cooperation, alignment and ultimately interoperability across OS Platforms/Commons/Clouds

Making Data Work for Cross Domain Grand Challenges

1: Core Interoperability
Exploring core issues around interoperability

2: Working with Disciplines and ISUs
FAIR vocabularies, terminologies, metadata

3: Cross-Domain Case Studies
Resilient cities, infectious diseases, SDGs..

1. **Exploring core issues around interoperability:**
   - Digital Representation of Units of Measure TG and the Digital SI: https://codata.org/initiatives/task-groups/drum/
   - FAIR Vocabularies: https://doi.org/10.1371/journal.pcbi.1009041
   - Collaboration with DDI Alliance on Cross Domain Integration (CDI): https://doi.org/10.5281/zenodo.4707263
   - Participation in FAIR Digital Objects Forum: https://fairdo.org/about/
   - Preparation of FAIR Implementation Profiles: https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/

2. **Working with disciplines and International Scientific Unions:** promoting principles of FAIR vocabularies, explore with Unions and other discipline representatives

3. **Cross Domain Case Studies:** resilient cities, infectious diseases, DRR, SDGs and more… **World FAIR funding proposal under consideration to expand these cases studies.**

   ▪ **Overview: https://bit.ly/CODATA-Decadal-Programme**

Cross-Domain Interoperability Framework?

- Aims to add some detail to the FAIR Principles, particularly I and R. Proposal that we can define a useful set of guidelines for domain agnostic metadata to support Interoperablity and Reuse.

    - **Units of Measurement:** the Digital SI, units services and the emerging 'Universal Metrology Data Model'.

    - **FAIR Vocabularies:** guidelines for good practice in FAIR vocabularies, '10 Simple Rules' https://doi.org/10.1371/journal.pcbi.1009041; exploring implementation with a number of domains; available standards and services **(SKOS; RVA, NERC, OBO etc)**

    - **Structure:** The roles played by datums in various data structures (wide data, long data/event data, multi-dimensional/cube data, key-value/big data); toolkit for describing data structure. **(DDI-CDI)**

    - **Variable Description:** Approach to describing individual "datums" and associated foundational metadata (concepts, variables, classifications, coding, etc.). **(Variable Cascade in a number of standards)**

    - **Provenance:** Description of provenance and processing of data **(e.g. PROV, SDTL, other standards/descriptions to describe processing chain)**

- Webinar on 'the Emerging Core Interoperability Framework', 31 March: https://bit.ly/cross-domain-IF-webinar
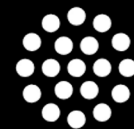
WorldFAIR Project

- Project proposal approved by the European Commission. Formal announcement on completion of Grant Agreement in April 2022.

- Project scheduled to start 1 June 2022.

- Will form the core of CODATA's contribution to ISC Action Plan Project 2.1: 'Making Data Work…'

- Exploring features of a Core Interoperability Framework with case studies from a range of research areas.

- FAIR Implementation Profiles (how do you implement FAIR?)

- Units, vocabularies, data description, data structure, provenance…

- Partnership with the Research Data Alliance.

- **Explicitly aim to add additional Case Studies through further funding and collaboration.**

- **Sessions at SciDataCon/International Data Week in June; FAIR Convergence Symposium in October.**

WorldFAIR Case Studies

- **Chemistry** - making IUPAC assets FAIR

- **Nanomaterials** - applying NanoInchi and FAIR recommendations in Nanosafety.

- **Geochemistry** - recommendations for FAIR in geochemistry, particularly vocabularies.

- **Social Surveys Data** – data harmonisation between ESS and AussiESS.

- **Population Health** - INSPIRE - Integration of population surveys with clinical and genomics data for COVID-19 research in eastern and southern Africa.

- **Urban Health** - terminologies and making urban health data FAIR

- **Biodiversity** – improving GBIF data model in collaboration with TDWG - GBIF (Global Biodiversity Information Facility)

- **Agricultural Biodiversity** - pollinator data (KALRO, Embrapa, Meise, HiveTracks)

- **Ocean Science** - Implementing FAIR in the ODIS (Ocean Data and Information System) for the UNESCO Oceans' decade.

- **Disaster Risk Reduction** - recommendations on making DRR data and terminologies FAIR, case studies in Africa and Pacific Islands

- **Cultural Heritage** - recommendations on making cultural heritage data FAIR (particularly digital representation of heritage artefacts)
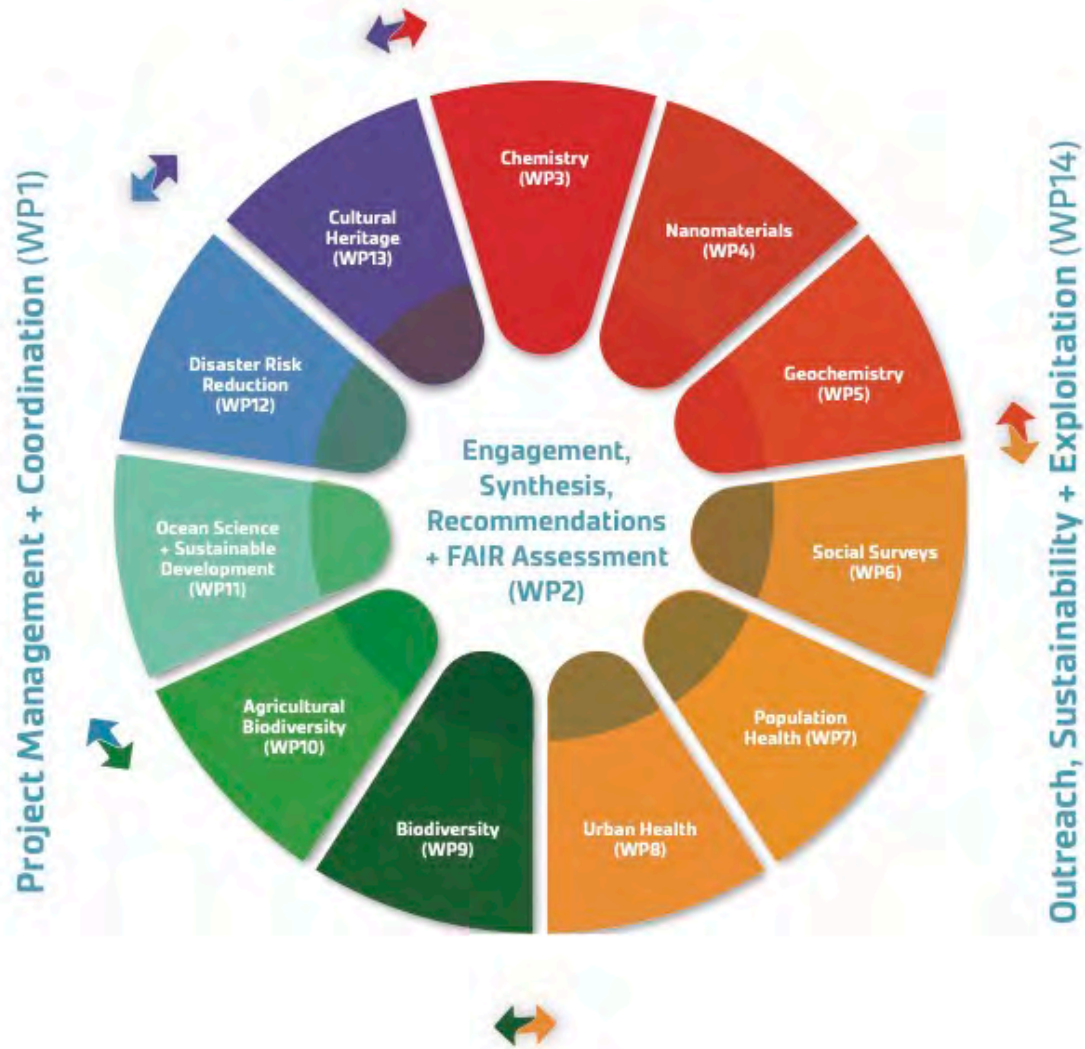
Thank you for your attention

Simon Hodson, CODATA
www.codata.org
simon@codata.org
@simonhodson99 ; @CODATAnews