CODATA: Global Collaboration to Advance Open Science and FAIR

Simon Hodson
Executive Director
CODATA
www.codata.org

Converging Science and Technology in a Digital Era
ISC Knowledge Sharing Event
15 and 16 February 2022

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.**
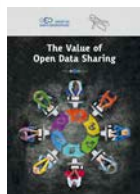
Making Data Work…

- **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

Data Policies

- 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

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

Data Skills

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

COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

International
Science Council

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

CODATA Task Group on Fundamental Constants

- CODATA's longstanding, foundational contribution to global science.

- TG sets the globally recognized values for fundamental constants.

- **Since 2019, these underpin the SI System of units.**

- TG is also supported by BIPM (including hosting the TG's portal and meetings).

- MoU signed 11 Oct 2021 formalizes this collaboration and makes BIPM a partner member of CODATA.

- Preparations underway for the 2022 adjustment (data to be used must be published or discussed in a preprint prior to 31 December 2022).

The Seven Base Units of the new SI are defined as in the table below*

Quantity	SI unit
time	The **second**, symbol s, is the SI unit of time. It is defined by taking the fixed numerical value of the caesium frequency $\Delta\nu_{Cs}$, the unperturbed ground-state hyperfine transition frequency of the caesium 133 atom, to be 9 192 631 770 when expressed in the unit Hz, which is equal to s^{-1}.
length	The **metre**, symbol m, is the SI unit of length. It is defined by taking the fixed numerical value of the speed of light in vacuum c to be 299 792 458 when expressed in the unit m s^{-1}, where the second is defined in terms of the caesium frequency $\Delta\nu_{Cs}$.
mass	The **kilogram**, symbol kg, is the SI unit of mass. It is defined by taking the fixed numerical value of the Planck constant h to be 6.626 070 15 $\times 10^{-34}$ when expressed in the unit J s, which is equal to kg m^2 s^{-1}, where the metre and the second are defined in terms of c and $\Delta\nu_{Cs}$.
electric current	The **ampere**, symbol A, is the SI unit of electric current. It is defined by taking the fixed numerical value of the elementary charge e to be 1.602 176 634 $\times 10^{-19}$ when expressed in the unit C, which is equal to A s, where the second is defined in terms of $\Delta\nu_{Cs}$.
thermodynamic temperature	The **kelvin**, symbol K, is the SI unit of thermodynamic temperature. It is defined by taking the fixed numerical value of the Boltzmann constant k to be 1.380 649 $\times 10^{-23}$ when expressed in the unit J K^{-1}, which is equal to kg m^2 s^{-2} K^{-1}, where the kilogram, metre and second are defined in terms of h, c and $\Delta\nu_{Cs}$.
amount of substance	The **mole**, symbol mol, is the SI unit of amount of substance. One mole contains exactly 6.022 140 76 $\times 10^{23}$ elementary entities. This number is the fixed numerical value of the Avogadro constant, N_A, when expressed in the unit mol^{-1} and is called the Avogadro number. The amount of substance, symbol n, of a system is a measure of the number of specified elementary entities. An elementary entity may be an atom, a molecule, an ion, an electron, any other particle or specified group of particles.
luminous intensity	The **candela**, symbol cd, is the SI unit of luminous intensity in a given direction. It is defined by taking the fixed numerical value of the luminous efficacy of monochromatic radiation of frequency 540 $\times 10^{12}$ Hz, K_{cd}, to be 683 when expressed in the unit lm W^{-1}, which is equal to cd sr W^{-1}, or cd sr kg^{-1} m^{-2} s^3, where the kilogram, metre and second are defined in terms of h, c and $\Delta\nu_{Cs}$.

DRUM (Digital Representation of Units of Measurement) TG

1. **Promote cooperation and coordination across initiatives**, and in particular mobilising the input of the various scientific domains, as represented by the ISUs/ISAs: https://codata.org/initiatives/task-groups/drum/

2. **Manifesto**, endorsed by the ISC and the ISUs/ISAs, calling for greater action and investment on the issue of units of measure (their definition, digital representation and conversion): https://doi.org/10.5281/zenodo.4081656

3. Preparing a **Call to Action** on Digital Representation of Units of Measurement, endorsed by CODATA DRUM and CIPM WG.

4. With CIPM WG, preparing **'Universal Metrology Data Model'** for Units of Measurement.

5. Preparing publication of the **Fundamental Constants** as machine-readable Linked Open Data.

6. **Sessions at IDW and a Units Summit and Hackathon, with BIPM, in planning for 2022.**

▪ **SciDataCon DRUM Session:** https://www.scidatacon.org/virtual-2021/sessions/329/

Digital SI
https://www.bipm.org/en/conference-centre/bipm-workshops/digital-si/

Bureau International des Poids et Mesures

FAIR Vocabularies in Domain and Cross-Domain Research Fields

- Group formed at Dagstuhl and led by Simon Cox produced **'10 Simple Rules for Making a Vocabulary FAIR'** **https://doi.org/10.1371/journal.pcbi.1009041**

- **IUSSP-CODATA Working Group on FAIR Vocabularies in Population Research** http://bit.ly/IUSSP-CODATA-FAIR-Vocabs

    - Working with owners of a set of key vocabularies in population science to apply 'ten simple rules for making a vocabulary FAIR' and other good practices.

- **IUPAC: how to make IUPAC assets FAIR?**

    - Discussions on collaboration on a similar joint initiative.

- **Invite similar collaborations with other International Scientific Unions or similar discipline representative bodies: e.g. a set of joint WGs exploring vocabularies and other semantic artifacts to make them FAIR.**

- **SciDataCon FAIR Vocabularies Session:** https://www.scidatacon.org/virtual-2021/sessions/351/
- **FAIR Vocabularies Workshop 'Downunder':** https://bit.ly/FAIR-vocabs-downunder

Global Collaboration on FAIR Vocabularies
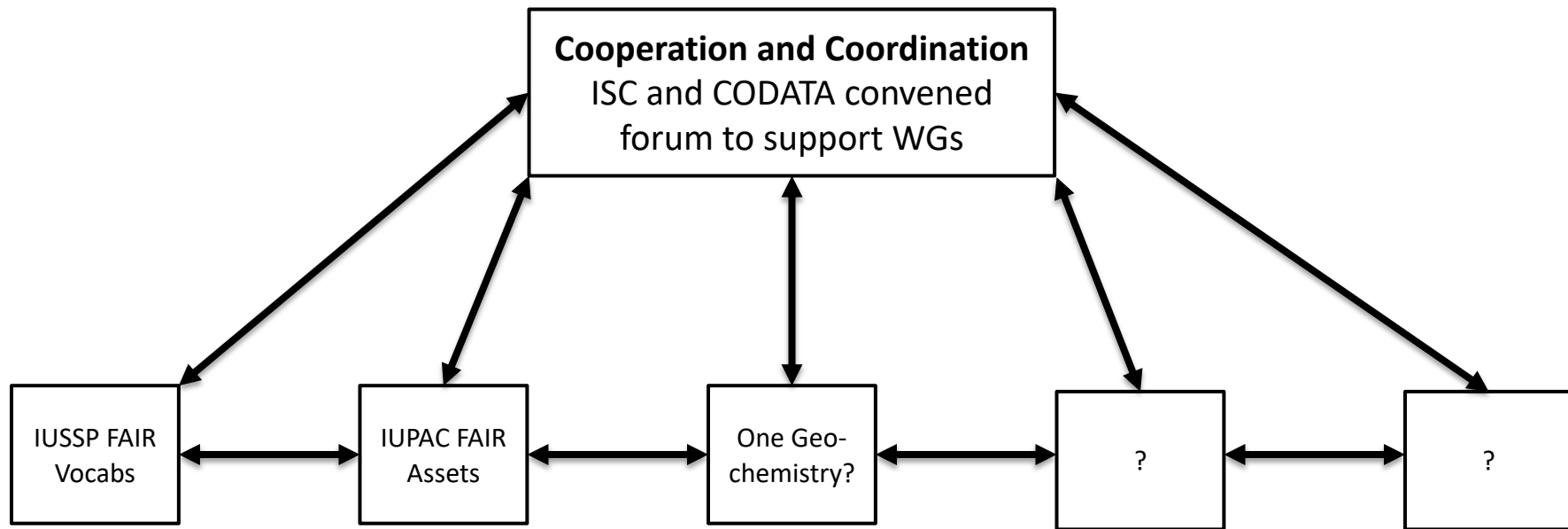
WorldFAIR Project

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

International Data Week — A FESTIVAL OF DATA

23-26 OCT 2023 SALZBURG

"The name WorldFAIR was chosen to evoke the great international technological exhibitions of more optimistic times. CODATA has its origins in the recognition of the need for international cooperation around data collection and data standards that was ignited by the great International Geophysical Year of 1958. … The outputs of the project … will be presented as a Great Exhibition, both virtually and at International Data Week 2023, Salzburg, Austria in order to inspire and guide other research areas."

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**.
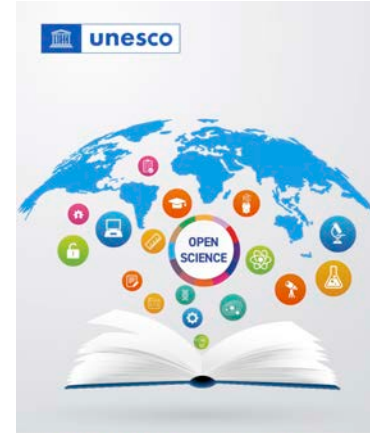
Global Open Science Cloud Initiative

- 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
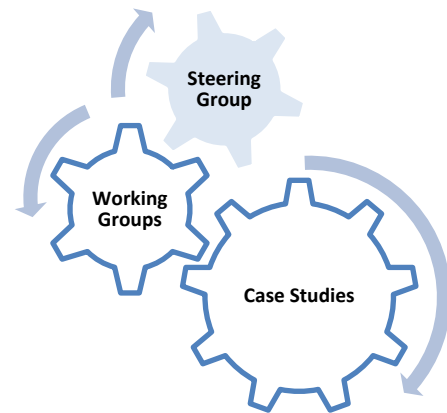
- GOSC Overview: https://bit.ly/GOSC-Overview
- Join GOSC WGs, Case Studies: https://bit.ly/3jwZHNg
- Propose New Case Studies: https://bit.ly/GOSC-Propose-New-Case-Study

Thank you for your attention

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