Making Data Work for Cross-Domain Challenges

Simon Hodson
Executive Director
CODATA
www.codata.org

Open Science strategic approaches and the importance of open research data
Webinar
5-6 October 2020

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.**

Data Policies

GBIF Global Biodiversity Information Facility

- CODATA 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

ubiquity press open scholarship

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

Data Skills

GO FAIR

ICTP The Abdus Salam International Centre for Theoretical Physics

FAIRSFAIR

CODATA

- CODATA-RDA School of Research Data Science.
- CODATA China, PASTD and other training activities.
- #terms4FAIRskills and FAIRsFAIR Competence Centres.

Data to Improve our World

DDI DATA DOCUMENTATION INITIATIVE

- **Decadal Programme:** Making Data Work for Cross Domain Grand Challenges
- Promoting Good Data Practices
- Regional Open Science Platforms

COMMITTEE ON DATA CODATA INTERNATIONAL SCIENCE COUNCIL

International Science Council

Open Science for a Global Transformation

Open Science for a Global Transformation

CODATA coordinated submission to the UNESCO Open Science Review: https://bit.ly/UNESCO-CODATA-Submission and https://doi.org/10.5281/zenodo.3935461

Led by the CODATA Executive Director, members of the Data Policy Committee and representatives of GO FAIR, WDS and ICSTI.

Will publish Editorial in DSJ calling for further input to the consultation.

UNESCO Recommendation

Simon Hodson is Vice Chair of the International Advisory Group. Draft recommendation completed. Will be published late-Sept/early-Oct after internal UNESCO Review.

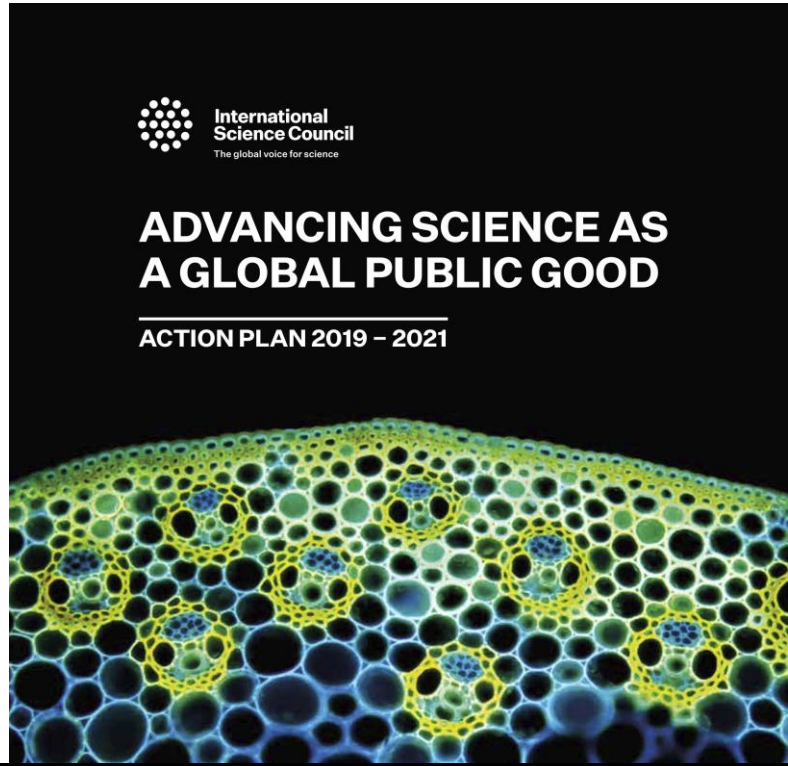
Further consultation Oct 2020-Jan 2021.

Open Science for a Global Transformation

- A movement for the transformation of science and global transformation.
- Science is a global public good. **Open Science aims to maximise the benefit of science for society and the engagement of society with science.**
- **Co-creation of knowledge and responsible scientific communication. Evidence and science-based decision-making.**
- **Open Science aims to maintain and promote good practice and scientific reproducibility by maximising access to robustly described data, code and methods underpinning scientific conclusions.**
- Open Science, through responsible governance, allows and requires necessary and proportionate protection of data, its sources, and derived information. *It categorically does not mean unrestricted openness.*
- Open by default. As open as possible, and only as closed as necessary.

Making Data Work for Cross-Domain Challenges: the Premise

- 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 a **transdisciplinary** approach to turn data into knowledge and then into action.
- The digital and data revolution presents us with huge opportunities and significant challenges.
- Major challenges for many scientific domains – requires work on data specifications, semantics, infrastructures, etc.
    - **80% of effort used on data wrangling; conservative estimate of 10.2 Bn Euro opportunity cost from sub-optimal data stewardship.**
- Open Science and FAIR data provide solutions.
- Considerable global interest in data platforms (EOSC etc).



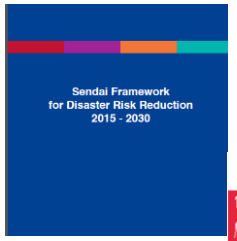
Data for Global Grand Challenges

- Addressing global grand challenges requires cross-domain collaboration.
- Needs the ability to gather data from many sources, to combine them and extract information from complex and heterogeneous data.
  - Combining data for SDG indicators is challenging.
  - Combining data for the scientific contribution to understanding of SDGs is very challenging!
- **ISC and ISC members (particularly Unions and Associations), and ISC programmes have a role to play.**
- Addressing how to access and combine data (issues of data interoperability) need input from domain experts and definitions agreed by communities.
- Major challenge of fundamental importance to science – **the work of a global decadal programme**.

futureearth
research for global sustainability

IRDR
Integrated Research on Disaster Risk

URBAN HEALTH AND WELLBEING
A SYSTEMS APPROACH



Findable Accessible Interoperable Reusable

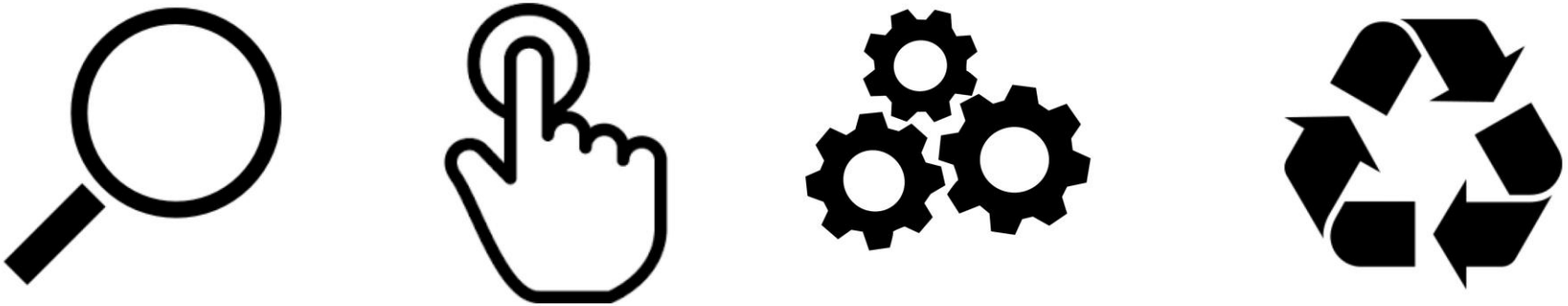


Image CC-BY-SA by SangyaPundir

(Mons, B., et al., The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, http://dx.doi.org/10.1038/sdata.2016.18)

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1 the protocol is open, free, and universally implementable
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

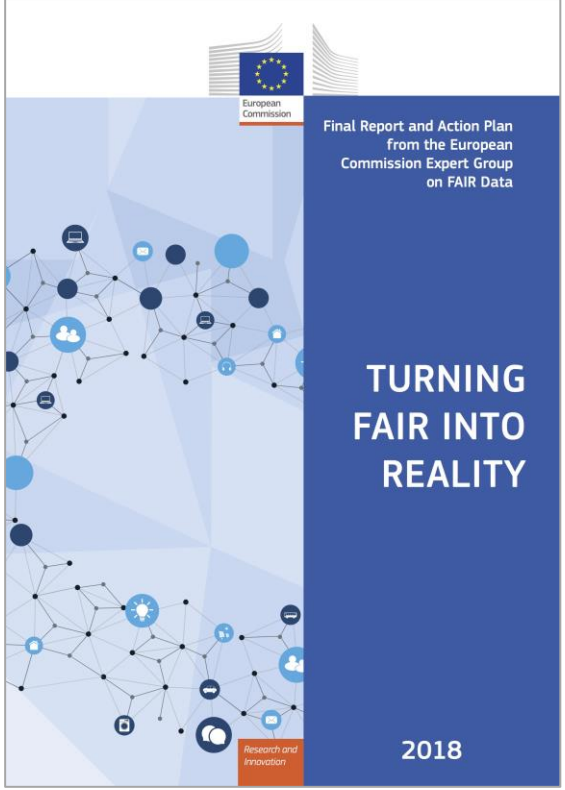
To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards

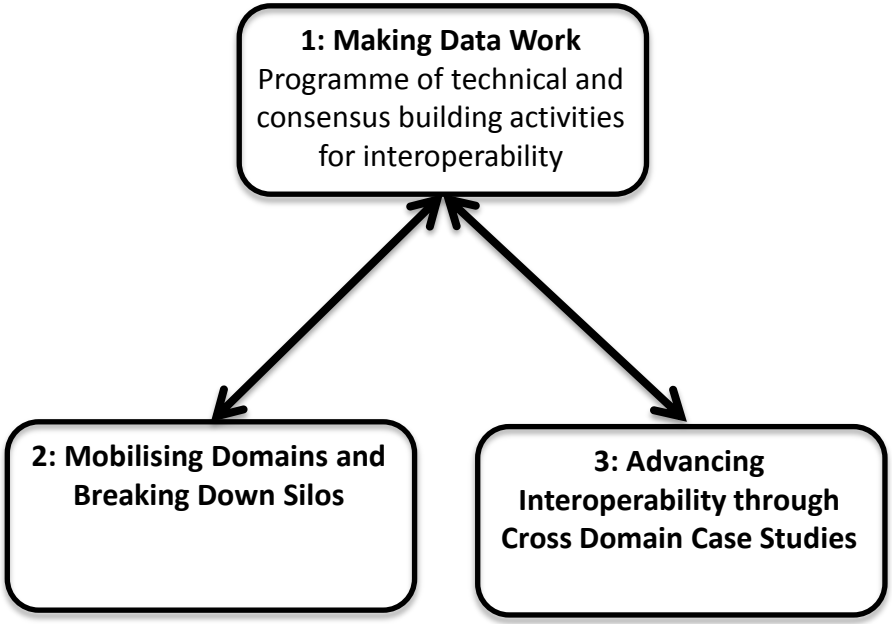
(Mons, B., et al., The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, http://dx.doi.org/10.1038/sdata.2016.18)



European Commission Expert Group, Chaired by Simon Hodson, Turning FAIR into Reality (2018)
https://doi.org/10.2777/1524

- **Findable:** have sufficiently rich metadata and a unique and persistent identifier, to enable discovery.
- **Accessible:** retrievable by humans and machines through a standard protocol; authentication and authorization where necessary.
  - Allows programmatic access for analysis.
- **Interoperable:** metadata use a 'formal, accessible, shared, and broadly applicable language for knowledge representation'.
  - The descriptions of variables etc follow a shared specification and are commensurable.
- **Reusable:** metadata provide rich and accurate information; clear usage license; detailed provenance.
  - Both humans and their analytical tools know what can be done with the data (license) and can assess its provenance.

Making Data Work: programme design



- Programme comprises three work areas.
  - 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 organisations);
  - Advancing solutions through cross-domain case studies.
- Current case studies in: **resilient cities, disaster risk reduction and infectious diseases**. **More planned and invited!**
- Working with domain and cross-domain areas, semantic solutions and machine learning.

Initial Pilot Activities

Initial Working Groups / Activities

1. Digital Representation of Units of Measure (TG is a key contribution to the decadal programme)
2. Semantic Interoperability and Conceptual Framework (good practice for semantic resources)
3. Supporting further refinement of the DDI-Cross Domain Integration specification
4. Policy Monitoring Indicators (SDGs, Sendai etc)
5. Infectious Diseases: projects looking at data integration in HIV and COVID
6. Resilient and Healthy Cities: large group with a number of cities and projects, identifying shared themes.

DRUM

- DRUM TG sent a position paper to the International Scientific Unions, to do the following:
  - Make the case for the importance of digital units of measure;
  - Invite an ‘ambassador’ from each Union / Association to be the point of contact for DRUM and engage with the TG;
  - Where appropriate, the ‘ambassador’ will be proposed also to be the liaison with BIPM and nominated for an important workshop on the Digital SI;
  - **Invite the Union to present use cases that demonstrate the utility and importance of digital representation of units of measure, or illustrate pain points.**
  - **Address conversions of non-SI units, digitally referenceable units and conversions.**

Digital SI

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

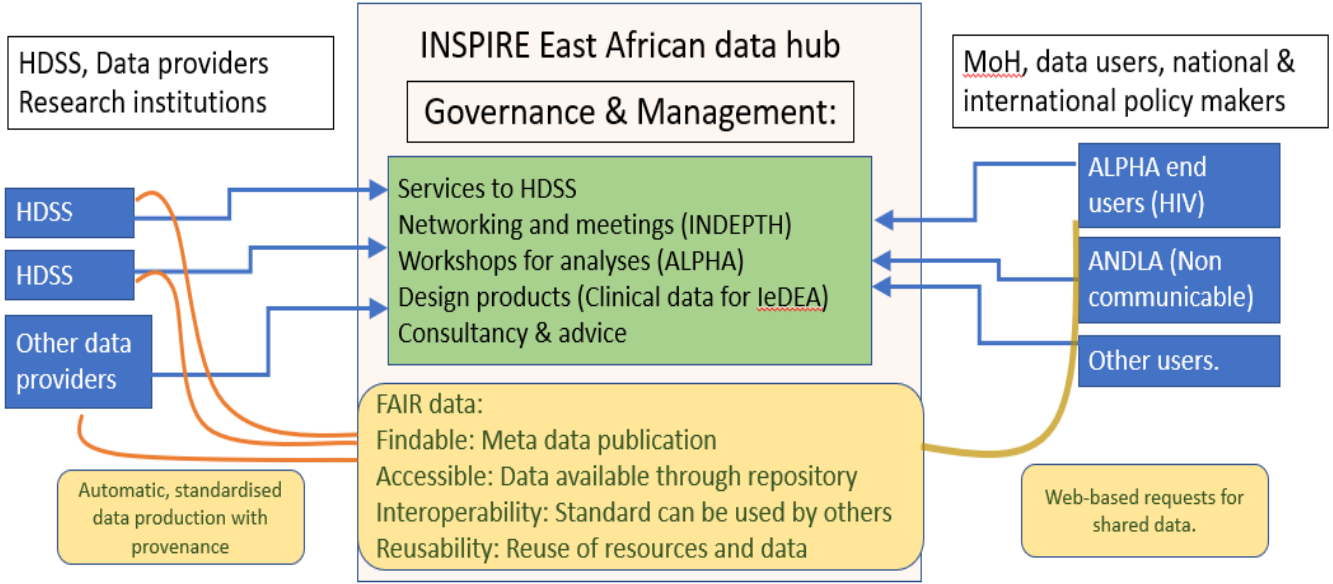
Bureau International des Poids et Mesures

Initial Pilot Activities: DDI-CDI

DDI-Cross Domain Integration Collaboration

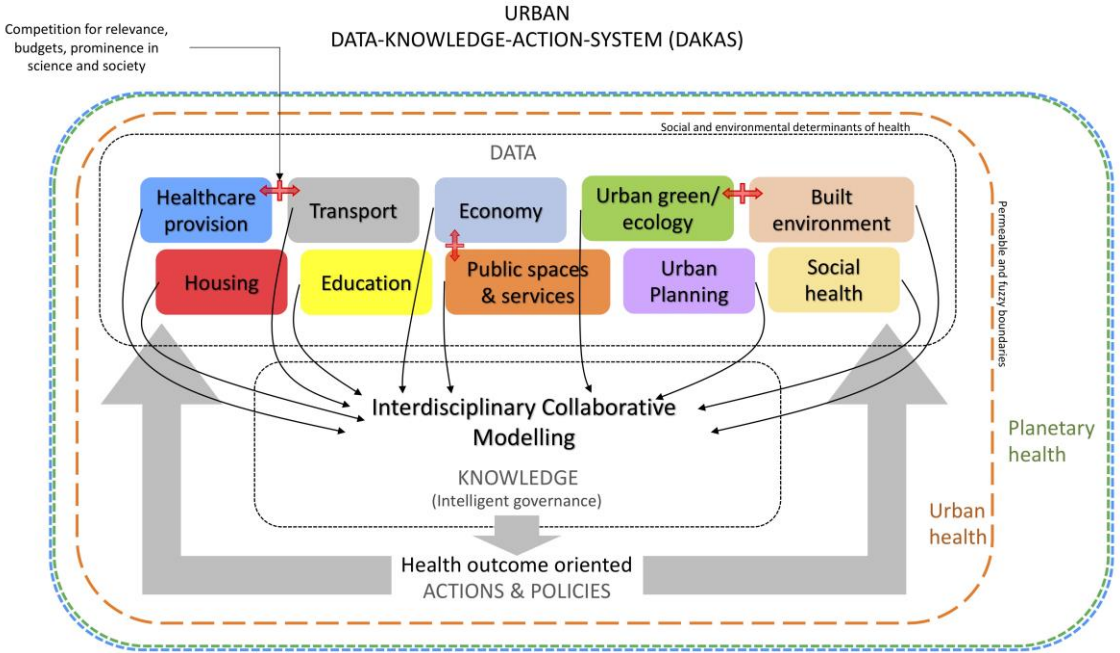
- DDI-CDI (Cross Domain Integration) is designed to interface with other standards and to help interoperability between different data types, standards, formats.
- Series of webinars to assist review of specification: https://bit.ly/DDI-CDI-Webinars
- Invite participation of International Scientific Unions and domain experts in a series of virtual workshops to identify use cases and further refine the specification.
- **EOSC funding to consult on DDI-CDI with European Research Infrastructures and EOSC and further refine the specification.**
- Upcoming workshops on representation of units, on provenance, on environmental use cases and health/medical use cases.

Initial Pilot Activities: Combining Social Science and Health Data



- Project with LSHTM combining HDSS (health and demographic surveillance system) data with clinical data.
- Secure data system.
- Consistent semantics and metadata.
- DDI-CDI also playing a role.
- Aim to apply the same approach to COVID Data.
- Diagram courtesy Chufundo Kanjala, INSPIRE Project.

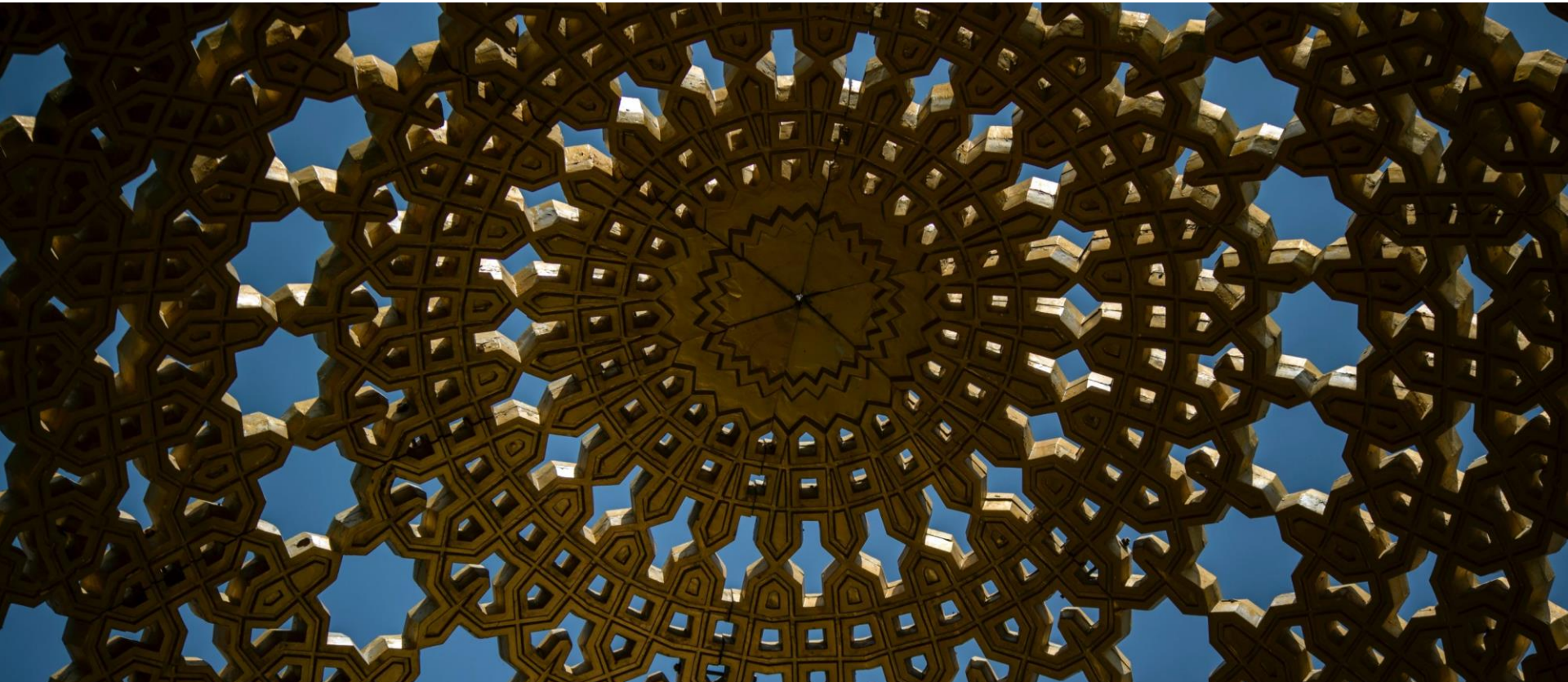
Initial Pilot Activities: Resilient Cities (Data-Knowledge-Action System)



- Decadal Programme Pilot Working Group on Resilient and Healthy Cities.
- Partnership with the ISC Programme on Urban Health and Wellbeing.
- Developing a conceptual model for Data Knowledge Action System.
- Data case studies: application of data audit, FAIR data.
  - E.g. mobility and contact tracing.
- Feeds into Interdisciplinary Collaborative Modelling with both data, community and expert inputs.
- Diagram courtesy of Franz Gatzweiler, UHWB.

Making Data Work for Cross-Domain Challenges

- Aim to launch the Decadal Programme at the ISC GA and associated events ‘Global Knowledge Forum’ in Oman, 10-14 October 2021: https://council.science/about-us/governance/general-assembly/muscatassembly



FAIR Convergence Symposium

- Entirely virtual event.
- 30 November-4 December.
- Keynotes, interactive sessions, posters.
- Preparatory workshops on key themes: Sept-Nov.
- **Call for Sessions, Posters and Lightning Talks: https://conference.codata.org/FAIRconvergence2020/**
- **Deadline for Session Proposals 30 September.**
- **Strongly encourage session proposals from CODATA China.**

INTERNATIONAL DATA WEEK 2021

Data to Improve our World

8 -11
NOVEMBER
2021

SEOUL,
REPUBLIC OF
KOREA

Convened by

CODATA
COMMITTEE ON DATA
INTERNATIONAL SCIENCE COUNCIL

WORLD DATA SYSTEM

RDA
RESEARCH DATA ALLIANCE

KISTI
www.kisti.re.kr
Korea Institute of
Science and Technology Information

KRISS
Korea Research Institute of Standards and Science

SUNGKYUNKWAN UNIVERSITY
1398

Follow CODATA!

- CODATA Website: http://www.codata.org/
- CODATA Blog: http://codata.org/blog/
- CODATA International News and Discussion List: http://bit.ly/CODATA-International-List
- CODATA Data Science and Data Stewardship Careers List: http://bit.ly/CODATA_Careers_List
- CODATA on Twitter: @CODATANews and @simonhodson99
- Facebook: https://www.facebook.com/codata.org/
- Insta: https://www.instagram.com/codatainternational/

COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

International
Science Council

Thank you for your attention

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