

Implementation of FAIR Principles to African Marine and Coastal Data- Open Science Approach

23rd -27th Sept. 2024, KEFRI Mombasa

Prof Muliaro Wafula PhD, FCCS, FCSK

Chair CODATA ,Kenya

Lead Scientist, AOSP EA Node

Associate Professor JKUAT, Kenya

Why Open Science?

- significant amounts of research still locked behind paywalls and only accessible at a high cost, this creates a research and innovation divide. Open Science and Open Access can democratize higher education, by providing equitable access to literature resources to academic communities.



Good Research Practice

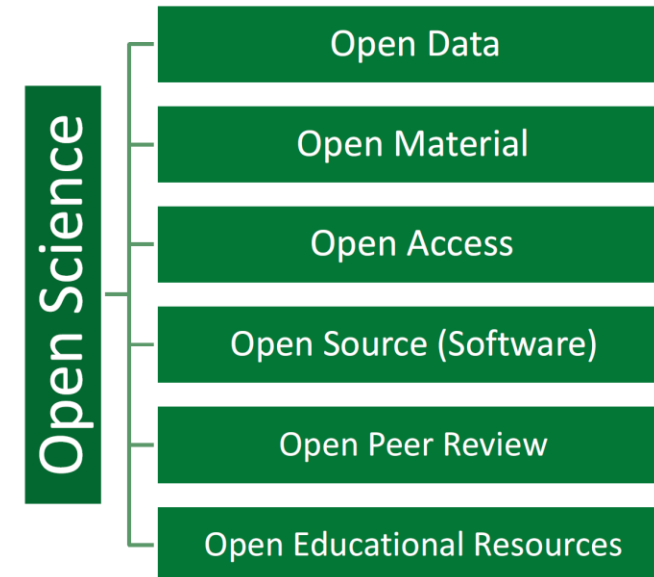
Good researchers

- Strive for excellence and take responsibility
- Respect the law, research ethics, and professional standards
- **Support a culture of transparency, openness, and honesty towards other researchers and the public**
- **Maximize public benefit and avoid resource waste**
- Continue learning and mentor others

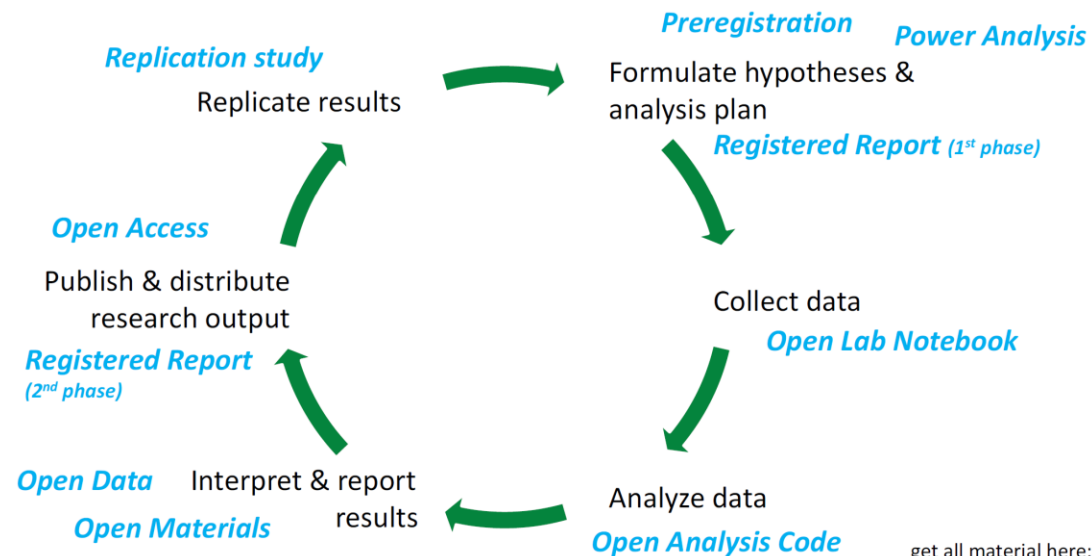


Open Science

Pillars of Open Science

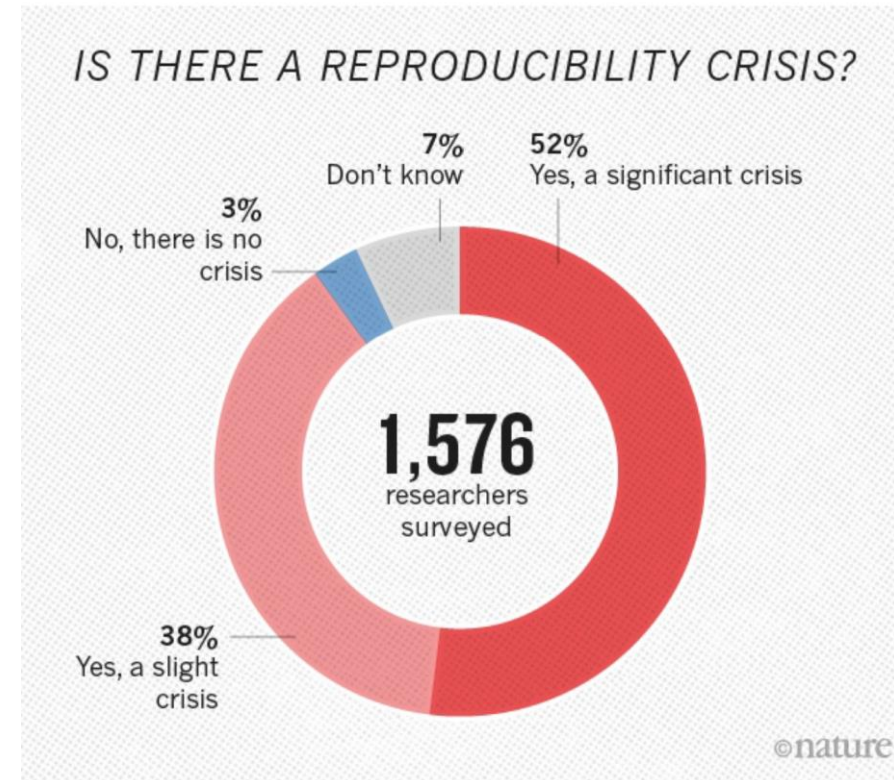


Open Science in the research process

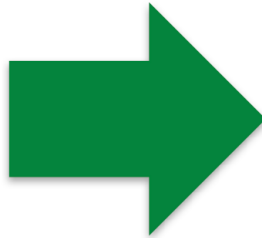


get all material here:
<https://osf.io/zjrhhu/>

Is there a Replication Crisis?



Current problems in science



Scientific progress is slowed down



Published results cannot be trusted



Resources are wasted

Preregistration

“The specification of a research design, hypotheses, and analysis plan prior to observing the outcomes of a study”



Nosek & Lindsay (2018)

Why?

- Prevent HARKing
- Reduce analytical flexibility
- Make selective reporting visible
- Get early feedback
- Take credit for your ideas
- Regulatory agencies require it

Apply for Registered Report

“Registered Reports are a form of empirical journal article in which methods and proposed analyses are pre-registered and peer-reviewed prior to research being conducted. High-quality protocols are then provisionally accepted for publication before data collection commences.”

COS(2018): www.cos.io/rr/

Why?

- Advantages of preregistration
- Guaranteed publication independent of results
- Peer review for your design

Open Lab Notebook

“Researchers use a lab notebook to document their hypotheses, experiments, and initial analysis or interpretation of these experiments. The lab notebook serves as an organizational tool, a memory aid, and can have a role in protecting intellectual property that comes from the research.”

Goyal, Malviya, & Kapoor (2012)

Why?

- Gain & share procedural knowledge
- Increase authenticity
- Protect your intellectual property

Open Data

“Open data should be available to everyone to access, use, and share.”

GO FAIR (2018)

**Why?**

- Make your analyses reproducible
- Enable re-use of data for answering other research questions
- Never lose valuable data in a file drawer
- Funding agencies require it

Open Materials

“Making components of the research methodology needed to reproduce the reported procedure and analysis publicly available.”



OSF (2016)

Why?

- Make your study reproducible
- Enable re-use of materials for other experiments

Open Access

Literature which is “digital, online, free of charge, and free of most copyright and licensing restrictions”.

Suber (2015)



Why?

- Enable faster progress in research by opening the access to knowledge
- Give back value to the community that funded you and not only to publishers

Open Analysis Code

“Clean, repeatable, script-based workflow [...] that links raw data through to clean data and to final analysis outputs.”

British Ecological Society (2017)

Why?

- Enable others to reproduce your analyses
- Understand your own code (after some time)
- Recreate your results with one click

Replication

“replication is a scientific method to verify research findings and [...] refers to a repetition of a research procedure to check the accuracy or truth of the findings reported.”

Schmidt (2009)

Why?

- Enhance credibility of your research
- Gain confidence in your findings & solidify the basis of your research

Best Practices

- ✓ Promote open science culture through training and awareness raising
- ✓ Develop cloud computing facilities that provide networked computation, data access and analysis tools
- ✓ Promote science diplomacy
- ✓ Encourage citizen science
- ✓ Support open collaboration through ICTs
- ✓ Support Open Innovation
- ✓ Promote open software code
- ✓ Adopt Findable, Accessible, Interoperable and Reusable (FAIR) Data Principle

Implementers: Government, Policy Makers, Universities, Scholarly Societies, Publishers, Research Institutes, Research Funding Agencies, Repositories, Libraries, Archives, and Researchers

CODATA and WorldFAIR project

Making Data Work – WorldFAIR – WorldFAIR+

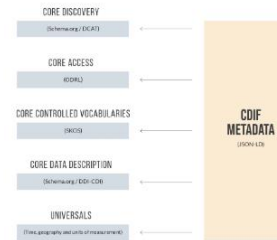
Making Data Work
(2018-2022)



WorldFAIR
(2022-2024)

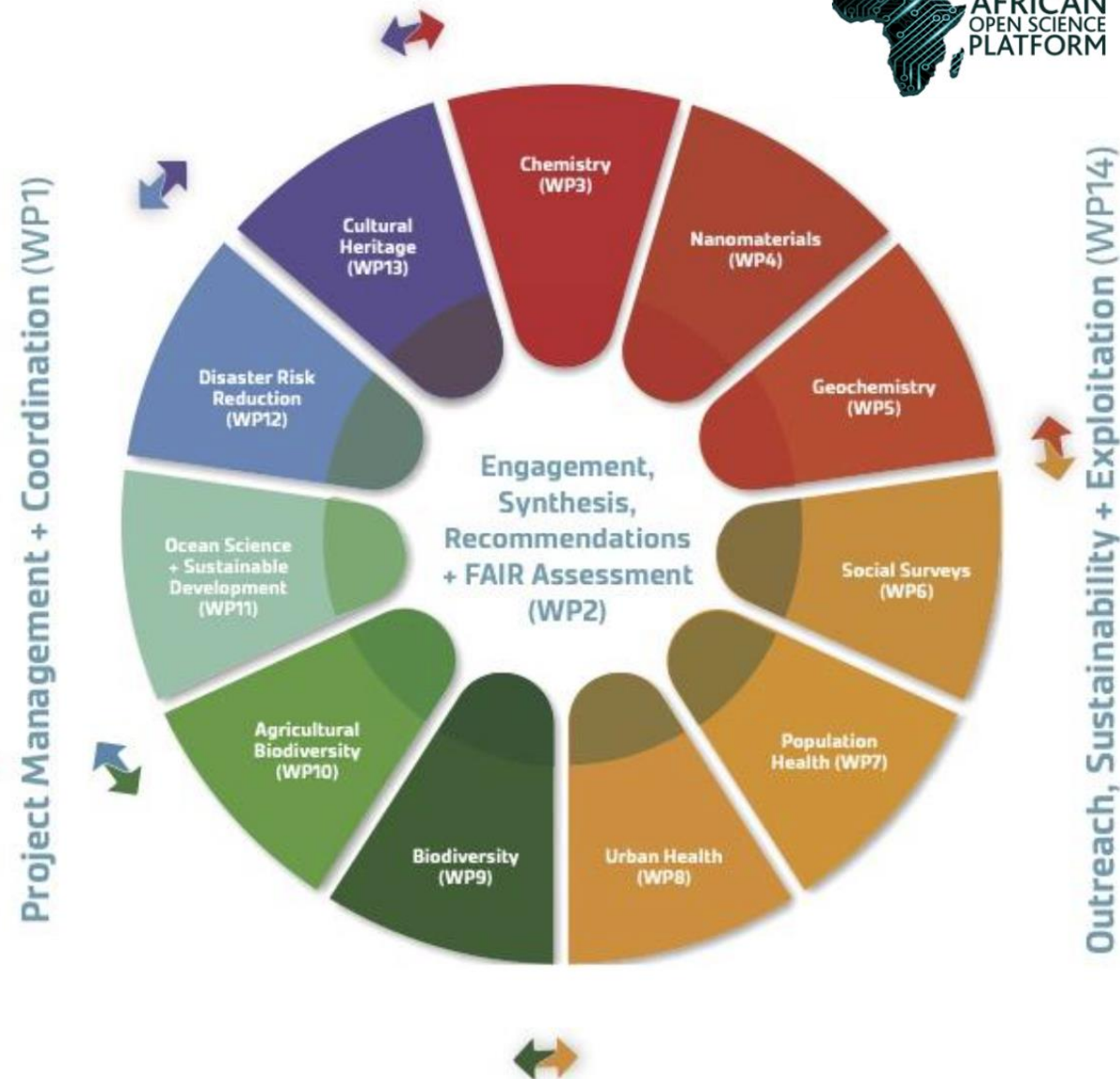


WorldFAIR+
(2024+)

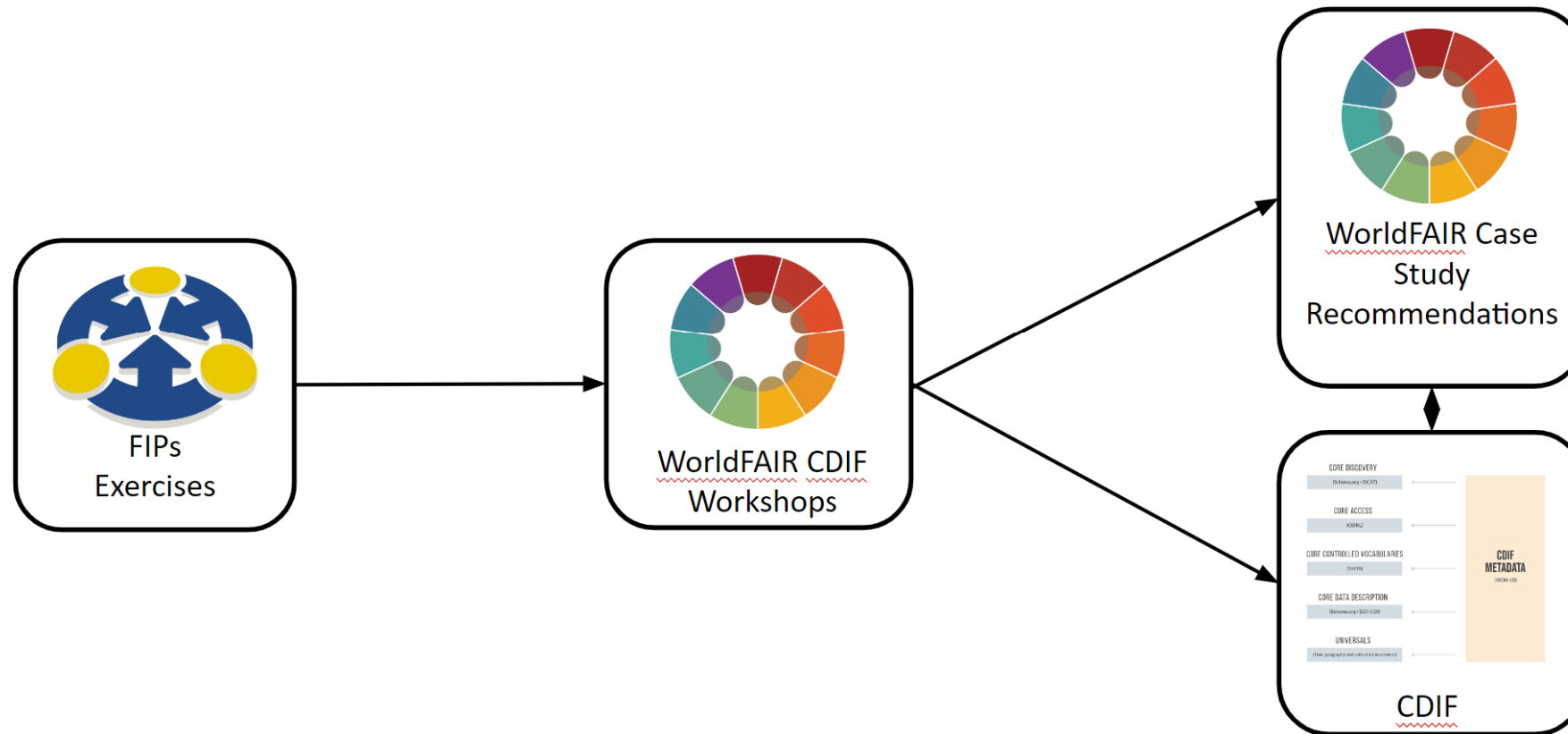


WorldFAIR Case Studies

- **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)

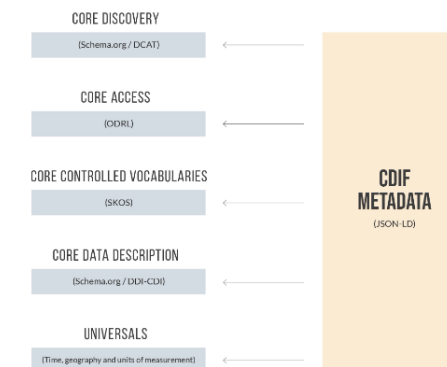


WorldFAIR Methodology



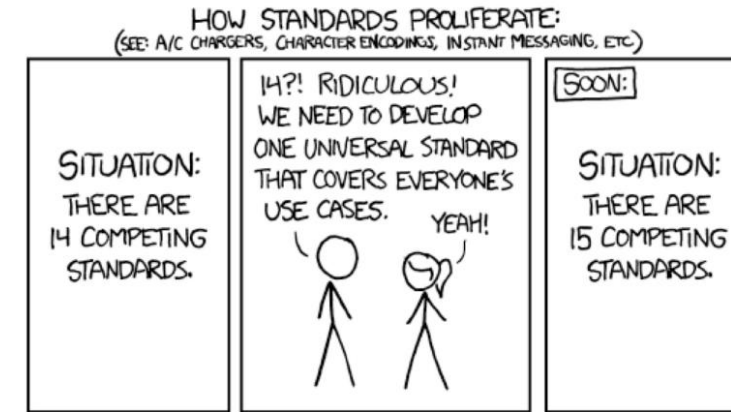
WorldFAIR Outputs

- See WorldFAIR Final Message from Coordinator
<https://bit.ly/WorldFAIR-Final-Coordinator-Message> and
<https://worldfair-project.eu/worldfair-plus/>
 - **Summary of project, links to key outputs and next steps.**
- Reports, recommendations, guidelines, implementation examples and training materials from 11 Case Studies: <https://zenodo.org/communities/worldfair-project/records> and
<https://worldfair-project.eu/project-outcomes-and-documents/>
 - **So many useful materials for all the subjects covered by WorldFAIR!**
- Policy Recommendations: <https://doi.org/10.5281/zenodo.11242702> :
 - **We urgently need to shift from a bibliographic to an engineering approach to data stewardship.**
 - **We need metadata uplift to support interdisciplinary research and the responsible use of AI.**
- Cross-Domain Interoperability Framework (CDIF):
<https://doi.org/10.5281/zenodo.11236871> (**approaching 2000 downloads!**)
 - **A practical guide to FAIR implementation! Adopt widely used web standards and use them in line with good practice.**

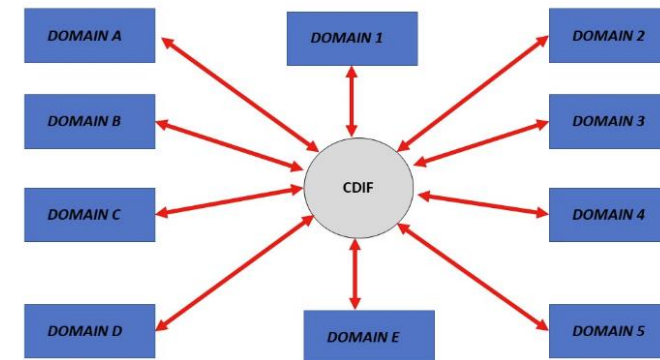


What is the CDIF (Cross-Domain Interoperability Framework)?

- Pushes beyond the F and A, to the I and R (and harder parts of A).
- **Identifies a set of functional requirements for interoperability, particular for steps in data combination, and identifies good practices for each of these requirements.**
- Use cases: domain or cross-domain projects or data services that need to combine data for analysis, modelling etc.
- **Categorically not a new standard. Rather it is a framework of existing and emerging standards.**
- A framework of standards/specifications to provide a *lingua franca*.
- Significant proportion of CDIF rests on good web practice, domain neutral standards and good practice: disciplines can adopt or map.
- Directed at implementers: use cases, standards, how to implement them.



Source: xkcd.com



CDIF Functional Requirements and Standards/Specifications

- Aims to discuss and recommend good practice and standards for a set of functions related to the FAIR principles
 - Discover (**DCAT, schema.org as JSON-LD**)
 - Access, negotiate access to non-public data (**ODRL, DPV, DUO**)
 - Assess and Integrate/Combine:
 - Understand and transform data structure (**DDI-CDI**)
 - Understand and align semantics (**SKOS/XKOS, OWL, SSSOM**)
 - Determine origin/context (**PROV-O, I-ADOPT/O&M**)
 - Understand processing (**DDI-CDI, SDTL**)
 - 'Universals' (**time, place, units of measurement, authoritative domain recommendations**);
 - Organising concepts and connectors (**e.g. 'unit types' and 'conceptual variables'**) *NB: different terms for these in different domains...*
 - Metadata for ML; Packaging...
- CDIF WG subgroups on Design Principles, Discoverability, Data Integration, API/Architecture, Semantics: <https://worldfair-project.eu/cross-domain-interoperability-framework/>
- **V01 CDIF published on 31 May:** <https://doi.org/10.5281/zenodo.11236871>



Get involved with CDIF!

- **Apply to join the CDIF Working Group or Advisory Group! Deadline 30 September.**
- WG generally meets every two weeks. Focus is on reviewing the existing profiles and developing the new profiles.
- AG to meet 2-4 times a year as required. Receives profile documentation to review.
- Application form: <https://forms.gle/DhQ4rnnTWW8DzpGUA>
- First review of applications will occur in early October.

- Join the CDIF Community list: cdif-community@lists.codata.org
- Register at: <https://bit.ly/cdif-community-list>

- CDIF Book:
<https://cross-domain-interoperability-framework.github.io/cdifbook/introduction.html>
- CDIF Github: <https://github.com/Cross-Domain-Interoperability-Framework>



New 'WorldFAIR+', CDIF Implementation Projects

1. **"Data Science Without Borders"**: CODATA is a partner in this Wellcome-funded project which builds on the work of WorldFAIR WP07. The project includes CDIF implementation (particularly data interoperability/integration and privacy management) to enable federated analysis across four African **health research** centres (Kenya, Ethiopia, Senegal, Cameroon). Three years. Underway. Africa.
2. **"FAIR Data and Emergencies"**: CODATA will coordinate a project funded by ISC applying the WorldFAIR methodology and implementing CDIF components in **emergencies and cascading hazards** case studies on earthquake data (Turkey) and flooding / cholera (Malawi). 18 months. Started 1 September 2024. Africa and Turkey.
3. **"CDIF-4-XAS"**: an OSCARS project cascading grant, to implement, test and refine the CDIF, to prepare **X-ray absorption spectroscopy data for interdisciplinary reuse**. Two years. Starts 1 October 2024, once OSCARS processes are completed. Europe (Germany and UK, but with global relevance).
4. **"CLIMATE-ADAPT4EOSC"**: CODATA is a member of the consortium for a major four-year project on FAIR data and innovative services for climate adaptation. A central part of the project will be the implementation and further development of the CDIF and related tooling. The project comprises three case studies: **urban heat** (Greece); **oceans / coastal management data** (Portugal); **clay soils / hydrology / built environment / insurance** (France). Four years. Starts 1 Jan 2025 once EC processes are completed. Europe (Greece, Portugal, France).
5. **"FAIR Principles implementation for DDE"**: Implementation of FAIR principles, alignment of IUGS CGI standards with CDIF, for cross-domain research topics and data reuse in **geology**. Three years. From Jan 2025. Global. Funding from IUGS.



WorldFAIR+

Vision:

- federation of case studies (existing and new), with parallel funding and supported by a coordinating mechanism with technical expertise.

Approach:

- Use and refine the Case Studies and FIPs workshop approach.
- Further develop the work on describing 'context' and provenance that was explored at Dagstuhl.
- CDIF implementation and refinement.
- Case study-led interoperability frameworks.
- Continue and expand the CDIF Working Group and Advisory Group (underway).



Implementation, mechanisms and governance:

- Approach discussed at WorldFAIR GA, with AC and other stakeholders.
- ISC has approved WorldFAIR+ as part of its portfolio of activities: <https://bit.ly/ISC-WorldFAIR-PLUS>
- Levels of participation: A, case study; B, case study, plus contribution in kind to coordination; C, case study, plus financial contribution to coordination. **Will shortly reissue a public version of this document.**
- Interest from existing and new case studies, particularly from case studies with new projects to be part of the initiative.
- Challenge remains how to resource the coordinating function.

Concrete Progress

- Discussions with various partners and CODATA members.
- Exploring with ISC, UNEP and UNESCO to explore further case studies and implementation pilots.
- Five new projects for CDIF implementation.
- **Keen to discuss potential case studies with ISC members (national members, unions etc).**
- **Keen to discuss any other partnerships, mechanisms and funding!**

African Open Science Platform (AOSP)

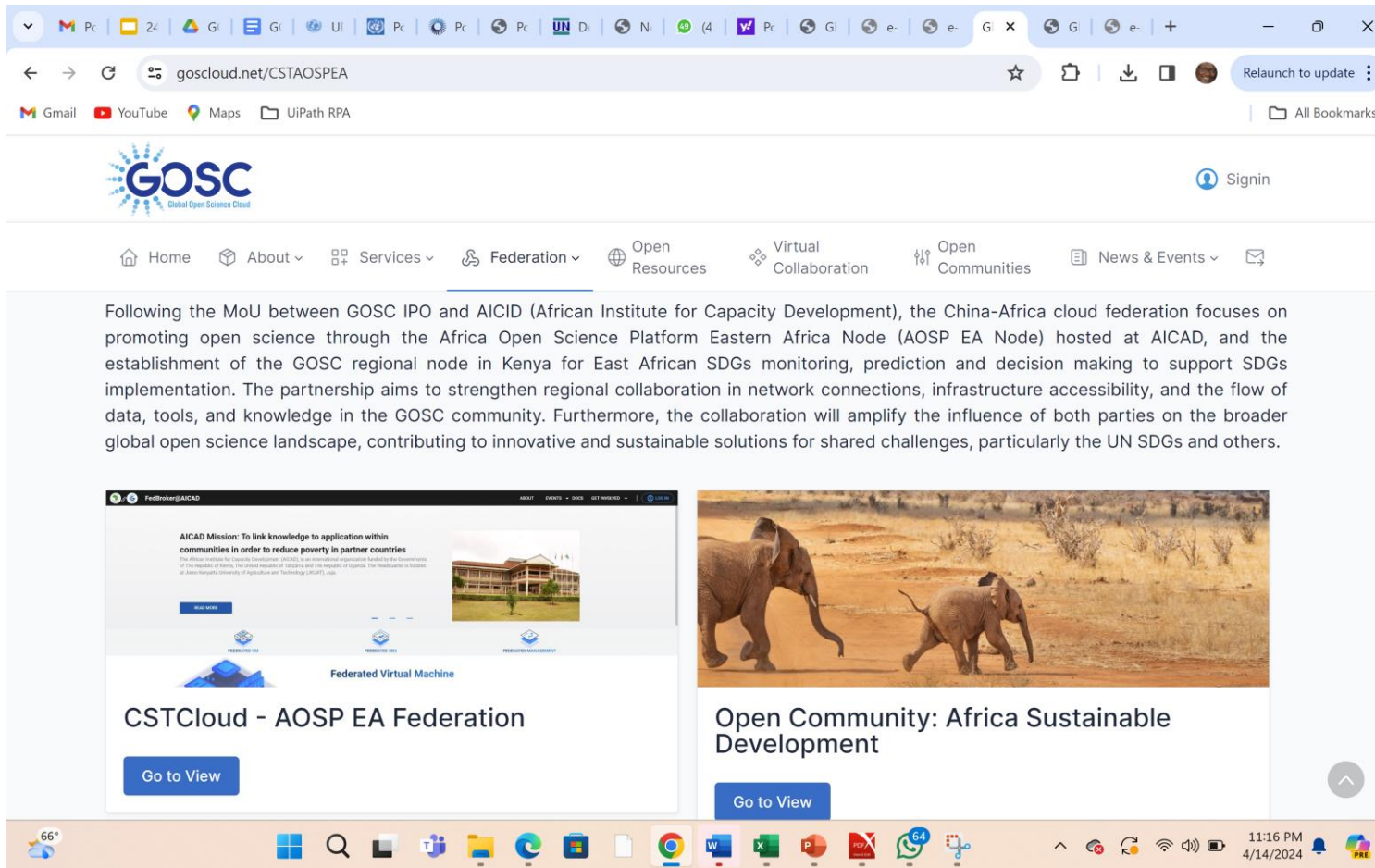
The African Open Science Platform
(AOSP) with HQ in SA
(<https://aosp.org.za/>)

- It has 3 regional nodes:
 1. Egyptian National Authority for Remote Sensing and Space Sciences (NARSS) is the Northern African Node based in Egypt;
 2. The African Institute for Capacity Development (AICAD) is the East African Node based in Kenya, and
 3. The UbuntuNet Alliance is the Southern Africa Node based in Malawi

Key Objectives

- Create a platform, with the potential to be a powerful lever of social, cultural and scientific vitality and of economic development.
- put African scientists at the cutting edge of contemporary, data-intensive science as a fundamental resource for a modern society
- Enable researchers to access a wide diversity of data streams, which, when integrated, reveal deep patterns in complex phenomena.
- Make science a more public enterprise that engages actively with business, policymakers, governments, communities and citizens as knowledge partners
- facilitate data-intensive, solutions-oriented research, bringing scientists and non-scientists together as knowledge partners in open networks of collaborative learning and problem solving

CSTCloud-AOSP EA Cloud federation



The screenshot shows a web browser window with the URL goscloud.net/CSTAOPEA. The page features the GOSC logo and a navigation menu with links to Home, About, Services, Federation, Open Resources, Virtual Collaboration, Open Communities, and News & Events. A "Signin" button is also present.

Following the MoU between GOSC IPO and AICAD (African Institute for Capacity Development), the China-Africa cloud federation focuses on promoting open science through the Africa Open Science Platform Eastern Africa Node (AOSP EA Node) hosted at AICAD, and the establishment of the GOSC regional node in Kenya for East African SDGs monitoring, prediction and decision making to support SDGs implementation. The partnership aims to strengthen regional collaboration in network connections, infrastructure accessibility, and the flow of data, tools, and knowledge in the GOSC community. Furthermore, the collaboration will amplify the influence of both parties on the broader global open science landscape, contributing to innovative and sustainable solutions for shared challenges, particularly the UN SDGs and others.

The page displays two featured articles:

- CSTCloud - AOSP EA Federation**: AICAD Mission: To link knowledge to application within communities in order to reduce poverty in partner countries. Includes a "Go to View" button.
- Open Community: Africa Sustainable Development**: Includes a "Go to View" button.

The bottom of the browser window shows the Windows taskbar with various application icons and the system clock indicating 11:16 PM on 4/14/2024.

TREs to facilitate secure and ethical data sharing.

Example: DASSA Project for Africa Research Communities –APHRC Funded

- TREs implement rigorous security measures, including data **anonymization**, **encryption**, and **controlled access**, to safeguard against unauthorized access or breaches.
- TREs enable researchers to access and use data in a controlled, transparent manner that adheres to **ethical guidelines**.
- TREs ensure that data is used responsibly, respecting participants' rights and mitigating the risk of misuse through **enforcing governance, and legal (National/International) frameworks** eg General Data Protection Regulation (GDPR) in the EU or the Health Insurance Portability and Accountability Act (HIPAA) in the U.S.
- TREs **foster collaboration** among researchers by providing secure access to shared datasets, while also ensuring that the integrity and confidentiality of the data are maintained,
- TREs **build trust among researchers, data providers, and the public**. This trust is crucial for encouraging data sharing and collaboration, which are essential for **addressing complex global challenges**
- TREs by protecting sensitive data and building trust among stakeholders, they enable collaborative research while **safeguarding individual privacy and data integrity**

FAIR Implementation Profile (FIP)

- A FAIR Implementation Profile (FIP) is a list of declared technologies made as a collective decision by the members of a particular community of practice.
- the FIP Wizard captures FIPs by means of a questionnaire that prompts a representative of the community (the Community Data Steward) to provide answers that explicitly profile the FAIR implementation approach of that community.
- FIPs are published by the FIP Wizard as FAIR (machine-readable) and Open data
- FIPs will need to be revised to reflect the evolving needs of the community and the ongoing development of FAIR technologies.
- Nanopublications: The FIP Wizard makes use of a data format called nanopublications to capture FIPs as FAIR data (in this case, as FAIR Digital Objects).

FIP Example

- WorldFAIR WP11 Oceanography
[https://drive.google.com/file/d/1k9Ofin5syNKNHpnn2WAZ6MwEctYaEpFx/view?usp=drive link](https://drive.google.com/file/d/1k9Ofin5syNKNHpnn2WAZ6MwEctYaEpFx/view?usp=drive_link)

5 Projects Exercise

FIP IOC Mombasa Training Exercise

[https://docs.google.com/spreadsheets/d/1kDk66FZUAYH20WltDJ89_ODvwg_L-LWCm/edit?usp=drive link&ouid=101924755130327600330&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1kDk66FZUAYH20WltDJ89_ODvwg_L-LWCm/edit?usp=drive_link&ouid=101924755130327600330&rtpof=true&sd=true)

FIP IOC Mombasa Training Template

[https://docs.google.com/spreadsheets/d/14T-7hCTiWOtoX1AxIsIwVRBTC5-zxNOptQvpJMuw2Ic/edit?usp=drive link](https://docs.google.com/spreadsheets/d/14T-7hCTiWOtoX1AxIsIwVRBTC5-zxNOptQvpJMuw2Ic/edit?usp=drive_link)

Thank you!

muliaro@icsit.jkuat.ac.ke