Initiatives for data and service federation that can connect local and global practices

session at SciDataCon; IDW2023 Salzburg, 23 October

An initiative from: EU project **EOSC Focus**
Organisers: Paolo Budroni (TU Wien)
   Franciska de Jong (CLARIN ERIC)
   Sara Garavelli (CSC)

Panel Composition

Moderator:

- Stefan Hanslik, Austrian Federal Ministry for Education, Science and Research;

  Open Science Austria;   Chair of e-IRG

Panelists

- Francesca Frontini, CNR/ILC (Italy); also involved in the EOSC Task Force Research careers and curricula
- Sarah Jones, Member of the board of directors of the EOSC Association
- Tshiamo Motshegwa, Director African Open Science Platform (ASOP)
- Wouter Schallier, Chief Hernán Santa Cruz Library and Web Services Unit, Latin America (UN/ECLAC) *remotely*
- Lee Wilson, Director of RDM at Digital Research Alliance of Canada

Digital Research Alliance of Canada

Initiatives for Data and Service Federation that can Connect Local and Global Practices

Lee Wilson, Director of RDM

Digital Research Alliance of Canada | **Alliance de recherche** numérique du Canada

Alliance plays a critical role as:

1. Leading federated **operations** of critical infrastructure

2. A **funder** for DRI initiatives

3. A leader of DRI **strategy** for Canada

…All with a focus on **serving** **Canada's** **researchers**

National RDM Services & Programs

Canada's Advanced Research Computing Platform

Digital Research Alliance of Canada

- 5 national host sites, federated user support
- Mix of general purpose and specialist systems
- Providing Community Cloud & Traditional HPC
- Serving a large user community with diverse needs

◎ **National Host Sites**

• **Support Sites**

Prairies DRI

Compute Ontario

Calcul Québec

ACENET

Recommendations (1/2)

Interoperability in national and global contexts

- Increased interoperability is needed to form a cohesive national data platform that provides services holistically and seamlessly across the research data lifecycle, and can connect to global infrastructure.

- Where appropriate, make use of open source, internationally supported software, standards, or infrastructure to ensure global interoperability.

- Investments in national PID infrastructure and strategy development is a means of advancing interoperability among disparate systems.

The complexity of multilingualism

- It is essential to approach development for new (and existing) services with multilingualism in mind from the outset and not rely on developing a system in one language and then attempting to translate it at the end.

- Multilingual support requires additional resourcing that must be planned for at the outset.

Recommendations (2/2)

Skills development

- Important to consider the different educational and experiential pathways for those supporting and using research data infrastructure services to inform professional development strategies.

- Continuous learning and upskilling opportunities are needed to contend with the rapidly changing landscape and technological developments.

- We should take a global perspective to skills development and consider approaches taken internationally around Data Stewardship training and education.

Development Of The African Open Science Platform (AOSP)
Critical Pieces – Governance, Policy, Networks, Diversity

Policy | Infrastructure | Data |Access| Skills | Incentives| Collaborations| Partnerships|

Dr Tshiamo Motshegwa
Director AOSP
Strategy, Planning and Partnerships (SPP)
NRF
South Africa
T.Motshegwa@nrf.ac.za

Website: https://aosp.org.za/

Data Policy Framework

- **Cooperatively enable data to flow on the continent;**
- **Create the necessary data capabilities**
- **Promote transversal data policy and agile regulation**
- **Create co-jurisdictional frameworks for the coordination**
- Develop **national legislations on personal data protection** and adequate regulations, particularly around data governance
- Establish or maintain independent, well-resourced and effective **Data Protection Authorities**
- **Promote interoperability, data sharing, and responsiveness** to data demand through the setting of open data standards
- **Promote data portability -** to avoid vendor lock in
- Improve **unevenly developed infrastructure** across the continent
- Establish an **integrated national data system** to enable data-driven public and private value creation
- Promote **research, development and innovation** in various areas
- Develop an **implementation plan** that takes into consideration the digital sovereignty of states

AOSP OS Policy Tracker

- **Country**
- **Type of Policy Instrument** (Policy, Roadmap etc)
- **Lead Ministry or Authority**
- **Policy Development process**
- **Policy Pillars**
  - IP, Open Access, Open/FAIR Data, Education & Awareness, Rewards & Incentives, Infrastructure, Citizen Science, Advocacy, Resourcing
- **Associated Instruments** (Strategy, Implementation plan etc.)
- **Notable Initiatives**
- **Related Policies** – STI and Sector

National Open Science Policy Development

■ Draft ■ Adopted ■ inProgress

1. Ethiopia*
2. South Africa**
3. Cote D'Ivoire
4. Nigeria
5. Uganda
6. Tanzania
7. Botswana
8. Ghana
9. Somalia
10. Lesotho

ECMWF data coverage (all observations) - SYNOP-SHIP-METAR
2022101815 to 2022101821
Total number of obs = 250802
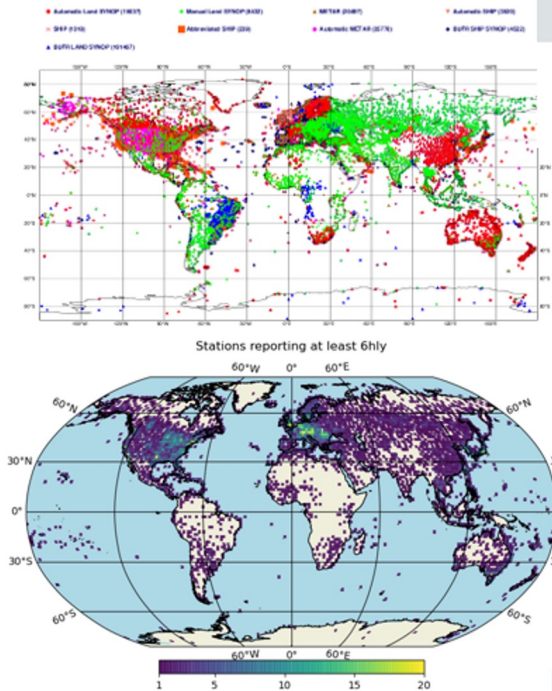
Stations reporting at least 6hly
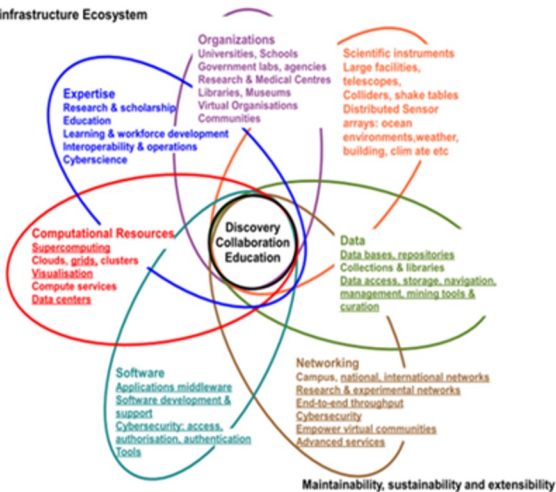
Number of Stations

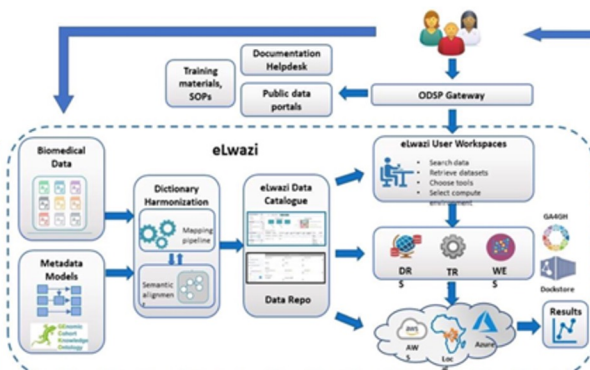**Strengthening Health and Disease Modeling for Public Health Decision Making in Africa**

Cyberinfrastructure Ecosystem

Organizations
Universities, Schools
Government labs, agencies
Research & Medical Centres
Libraries, Museums
Virtual Organisations
Communities

Scientific instruments
Large facilities,
telescopes,
Colliders, shake tables
Distributed Sensor
arrays: ocean
environments,weather,
building, clim ate etc

Expertise
Research & scholarship
Education
Learning & workforce development
Interoperability & operations
Cyberscience

Discovery
Collaboration
Education

Data
Data bases, repositories
Collections & libraries
Data access, storage, navigation,
management, mining tools &
curation

Computational Resources
Supercomputing
Clouds, grids, clusters
Visualisation
Compute services
Data centers

Software
Applications middleware
Software development &
support
Cybersecurity: access,
authorisation, authentication
Tools

Networking
Campus, national, international networks
Research & experimental networks
End-to-end throughput
Cybersecurity
Empower virtual communities
Advanced services

Maintainability, sustainability and extensibility

EU2022.CZ

Czech Presidency of the Council
of the European Union

Brno Declaration
on Fostering a Global Ecosystem
of Research Infrastructures

SOUTH AFRICAN | RESEARCH INFRASTRUCTURE ROADMAP
FIRST EDITION

Training
materials,
SOPs

Documentation
Helpdesk

Public data
portals

ODSP Gateway

Biomedical
Data

eLwazi

Dictionary
Harmonization

eLwazi Data
Catalogue

eLwazi User Workspaces
• Search data
• Retrieve datasets
• Choose tools
• Select compute
  environment

Mapping
pipeline

Metadata
Models

Semantic
alignmen

Data Repo

DR
S

TR

WE
S

GA4GH

Dockstore

Results

AW
S

Loc

Azure

eLwazi
OPEN DATA SCIENCE PLATFORM

Enabling data science applications for health in Africa

science & technology
Department:
Science and Technology
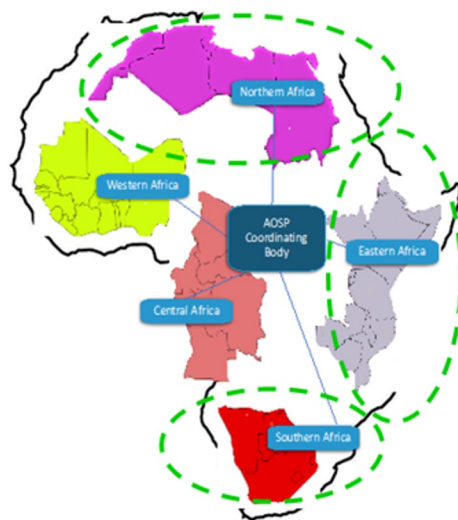REPUBLIC OF SOUTH AFRICA

AOSP – The What

AOSP is an Open Science Diplomacy platform **stimulating interactivity** and creating opportunity through the development of **efficiencies of scale**, **building critical mass** through **shared capacities**, **amplifying impac**t through a **commonality of purpose and voice**, and to engage in Global Commons to address continental and global challenges through **joint action**

AOSP – The Why?

- ✓ **Operationalization of the UNESCO Open Science Recommendations**,
- ✓ **Enhancing and entrenching dialogue** on open science, skills development, infrastructure , sustainability
- ✓ **Promotion of development of interoperable** and aligned open science policy frameworks,
- ✓ **Showcasing African Research** addressing developmental agenda and societal impact
- ✓ **Coordination and creating linkages** between open data and open science programmes across disciplines across continents
- ✓ **Contribution to the development** of global standards
- ✓ Plugging in to **global architecture for international open science framework**,
- ✓ **Pragmatic incremental approach in implementation.**

AOSP – The How?

Northern Africa

Western Africa

AOSP Coordinating Body

Eastern Africa

Central Africa

Southern Africa

Concrete action

How to galvanise around concrete projects to build trust, collaborations and advance open science globally: action the 2022 Brno declaration

During the closing day of ICRI 2022, the Brno Declaration on Fostering a Global Ecosystem of Research Infrastructures which highlights the essential role of research infrastructures was introduced. This call for action is supported by the European Union Member States and the European Commission, which are open to using their own processes, and the European Strategy Forum on Research Infrastructures (ESFRI) to underpin the development of a thriving global Research Infrastructures ecosystem.

Francesca Frontini (CNR/ILT)

- a distributed network of 70 centres

- 23 members: AT, BE, BG, CY, CZ, DE, DK, EE, ES, FI, GR, HR, HU, IS, IT, LT, LV, NL, NO, PL, PT, SE, SI

- 3 observers: CH,UK,ZA

- 1 third party

eosc

EOSC-A Task Forces

Research careers and curricula

The most important stakeholders for EOSC are the researchers.

- ∨ **Data Stewardship, Curricula and Career Paths Task Force**
- ∨ **Research Careers, Recognition and Credit Task Force**
- ∨ **Upskilling Countries to Engage in EOSC Task Force**
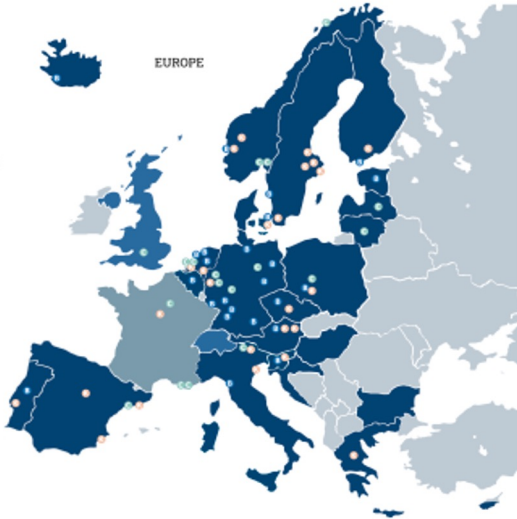- ∨ **Researcher Engagement & Adoption Task Force (REA TF)**

Interoperability across domains, regions

is an objective, but can also represent a **cost** and an additional burden (resources, efforts, skills) for infrastructures, which are focused on catering for the needs of their communities; the benefits may not be immediately visible to all interested parties. To be achieved, we need both a system of incentives, and some **convincing "success stories"**.

Skills Development

Data Stewardship is certainly well developed in some regions, and an asset for our data infrastructures, but a lot has to be done.

Ensure a system of *incentives for* **Researchers** and their careers.

Ensure *adequate training and career paths* for **Data professionals** also in terms of harmonisation across countries.

Global mobility in the Open Science professions, can also be a mean to facilitate interoperability.

Multilingualism

Multilingualism, especially in data and service infrastructures, should **not be an afterthought**; multilinguality for metadata is easier to achieve with the progress of language technologies, but provisions should be made for platforms and interfaces to be able to easily accommodate it.

Multilingual terminologies for the Open Science domain can also help to build bridges across communities.

Wouter Schallier
Hernán Santa Cruz Library at UN ECLAC

- ECLAC (www.cepal.org)
  - one of the 5 Regional Commissions of the United Nations
  - HQ is based in Santiago de Chile
  - unique regional coverage in Latin America and the Caribbean
  - Library has played a crucial role in offering a regional platform to Open Science community
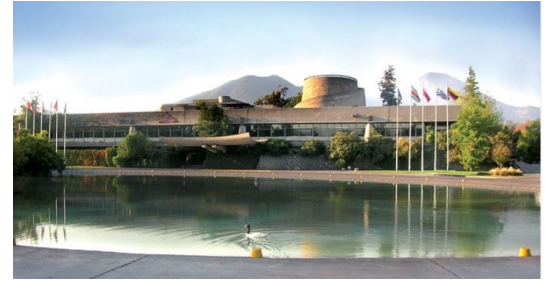
EOSC and Open Science will be international, or not be…

Before we can talk about interoperability, we first need to talk about availability and usability of data.

Access to quality data from regions outside Europe is vital for the further growth of EOSC.

More concrete steps should be made to equip research institutions and researchers in Latin America, the Caribbean and Africa with data skills in order to guarantee the availability of FAIR and CARE data.

Open Science an instrument for equity and inclusion?

Open Science can also become an instrument for inequality and exclusion just like the traditional scholarly communication system.

If researchers from Europe from the start have an important strategic advantage over their colleagues from the Global South,

If the researchers in the Global South don't have the skills and instruments to exploit data in the same way as their colleagues in Europe,

If commercial companies take control over research data,

Then researchers in the Global South will lose control over their research data. And Europe may also lose access to data from the Global South.

Conclusions

Global Open Science initiatives should support and work with regions such as Latin America where there is a long tradition of Diamond Open Access to publications, but not necessarily a data culture.

We need our global collaboration to act as a catalyst for open data, across European borders, with special attention to the Global South.

Sarah Jones (EOSC Association / GÉANT)

∞ eosc - Large EC initiative to build an OS solution

- EOSC = European Open Science Cloud

- Huge EC investment in infrastructure – €350 million in initial development phase and at least €1 billion co-investment foreseen for next 7 years

- Governance is a three-party collaboration

- Partnership MOU between EC and EOSC Association (legal entity) to govern/oversee the implementation

- Representation of Member States in Steering Board

- Attempt to blend top down & bottom up

EC public procurement
Managed Services for the European Open Science Cloud platform

Status update
- 4 May 2022: Prior Information Notice
- 4 May to 5 August 2022: Open Market Consultation (3 months)
- 20 December 2022: Tender Publication
- 2 May 2023: Submission of Initial Tender Response for Dialogue
- May to July 2023: Dialogue Stage (9 weeks)
- 28 July 2023: Final invitation to tender
- 7 September 2023: Final submission
- Ongoing: Evaluation (to be followed by contract signatures by the end of this year)

Procurement Lot Structure

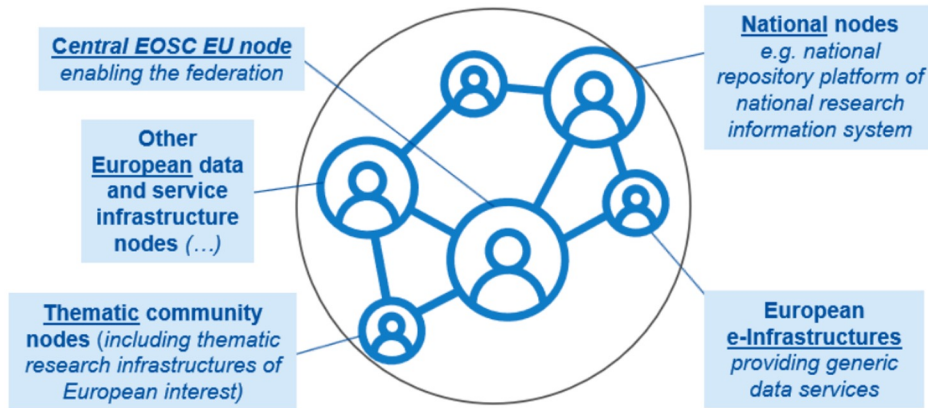|	
Lot 3	Managed Collaborative Data Platform, Interactive Data Analytics Platform and Visualization Services for the EOSC Exchange (Application Services)
Lot 2	Managed Container Platform and Virtual Machine Services for the EOSC Exchange (Infrastructure Services)
Lot 1	Managed Services for the Development, Integration, Deployment and Operations of the Federated EOSC Core

The **EOSC EU node** should provide from next year onwards core functionalities enabling the **EOSC federation**

Central EOSC EU node enabling the federation

National nodes e.g. national repository platform of national research information system

Other **European** data and service infrastructure nodes (…)

Thematic community nodes (including thematic research infrastructures of European interest)

European **e-Infrastructures** providing generic data services

Interoperability is biggest global challenge

- Lack of connection between infrastructure at different levels

- Some successes with federated EOSC AAI and resources catalogue

- Cross-domain interoperability still a way off → need to identify and prioritise concrete use cases based on global challenges e.g. SDGs

To the on site and online audience: Vielen Dank!

For a summary of the recommendations and follow-up, please check this page:

https://eosc.eu/events/international-data-week-2023/