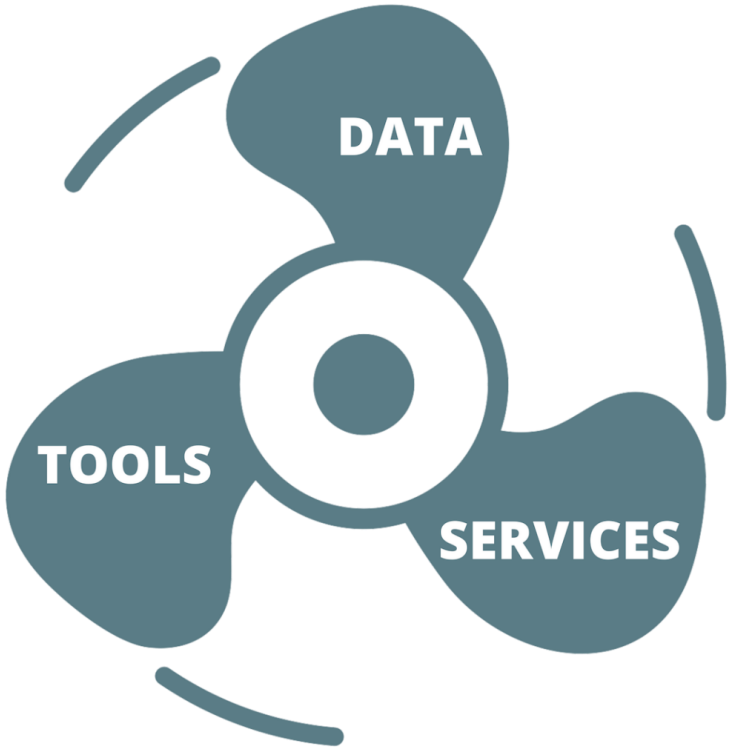


FAIR

DATA & SERVICES

GETTING READY FOR AI

Oya Beyan, 64. GMDS-Jahrestagung in Dortmund, 08. - 11. Sept. 2019

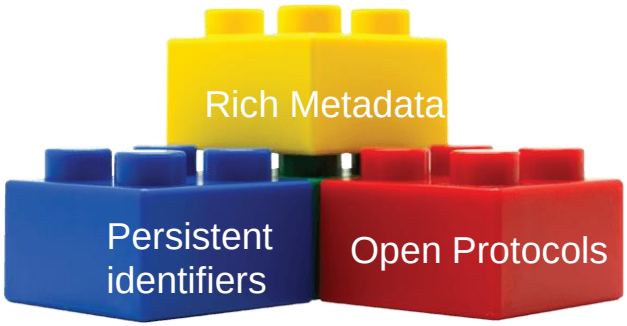
Fraunhofer FIT

Findable

Accessible

Interoperable

Reusable



Fraunhofer FIT

WHAT FAIR PRINCIPLES ARE NOT!

FAIR is **not** a standard

FAIR is **not** equal to ‘Open’ or ‘Free’

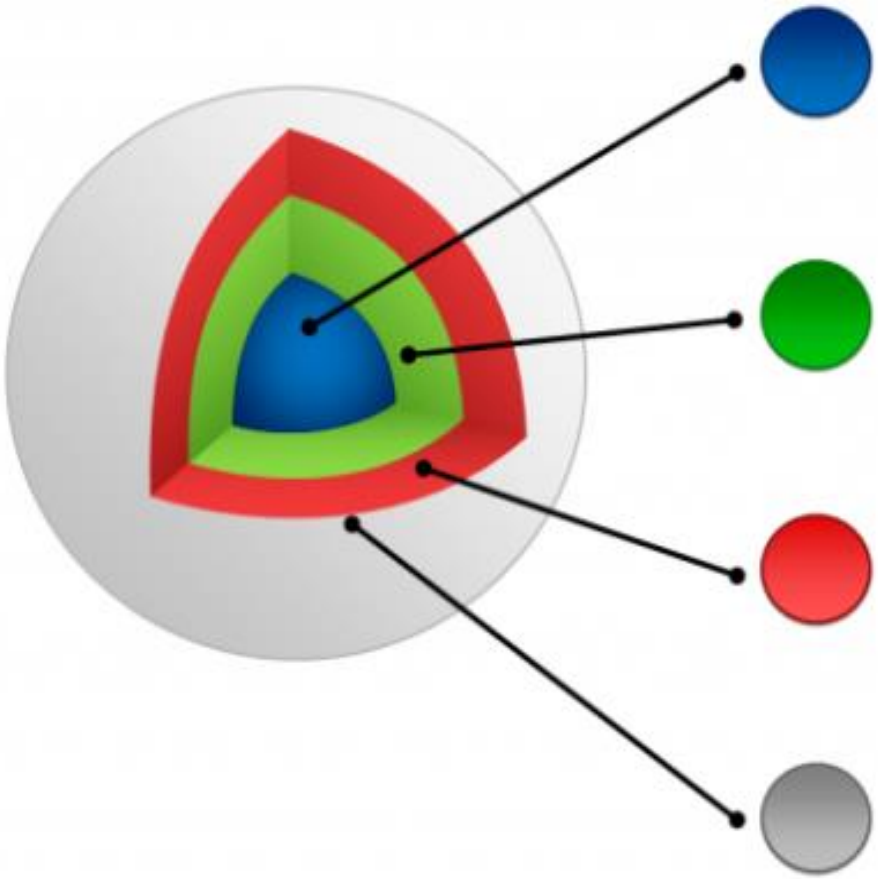
FAIR is **not** equal to RDF, Linked Data, or Semantic Web

FAIR is **not** assuming that only humans can find and re-use data

FAIR is **not** for humans only but for machines as well

Digital resources that are not FAIR are pretty ‘**Re-useless**’…..

FAIR Digital Object



DATA

The core bits

At its most basic level, data is a bitstream or binary sequence. For data to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and code. These layers of meaning enrich the data and enable reuse.

IDENTIFIERS

Persistent and unique (PIDs)

Data should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and supports citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders and associated research resources (RRIDs).

STANDARDS & CODE

Open, documented formats

Data should be represented in common and ideally open file formats. This enables others to reuse the data as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve . Data also need to be accompanied by the code use to process and analyse the data.

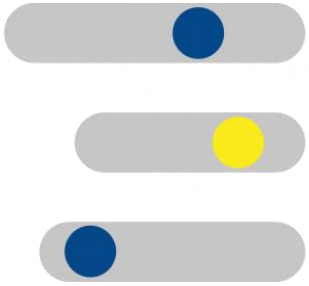
METADATA

Contextual documentation

In order for data to be assessable and reusable, it should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the data were created. To enable the broadest reuse, data should be accompanied by a 'plurality of relevant attributes' and a clear and accessible data usage license.

EC Expert Group on Turning FAIR Data into Reality

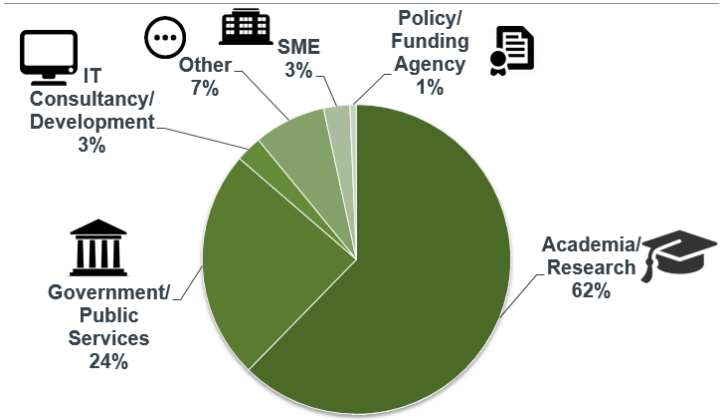
Enabling FAIR via Community Efforts



EOSCFAIR
Executive Board Working Group

Research Data Alliance created to Accelerate Development of Research Data Sharing Infrastructure Worldwide

- RDA community focuses on building **social, organizational and technical infrastructure** to
  - reduce barriers to data sharing and exchange
  - accelerate the development of coordinated global data infrastructure



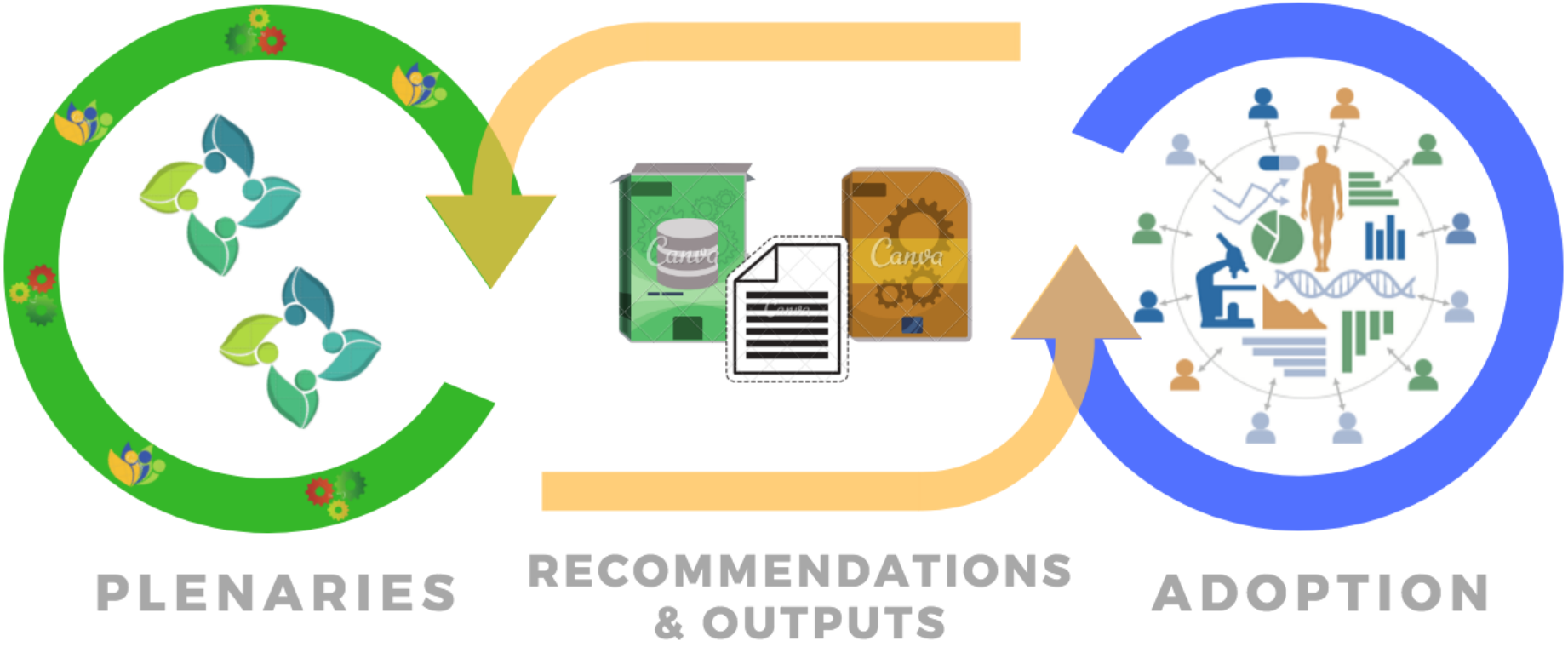
CREATE → ADOPT → USE

RDA Working Group Infrastructure Deliverables are:

- **Focused pieces of adopted code, policy, infrastructure, standards, or best practices** that enable data to be shared and exchanged
- **"Harvestable" efforts** for which 12-18 months of work can eliminate a roadblock for a substantial community
- **Efforts that have substantive applicability** to "chunks" of the data community, but may not apply to everyone
- **Efforts for which working scientists and researchers can start today** while more long-term or far-reaching solutions are appropriately discussed in other venues

Fraunhofer FIT

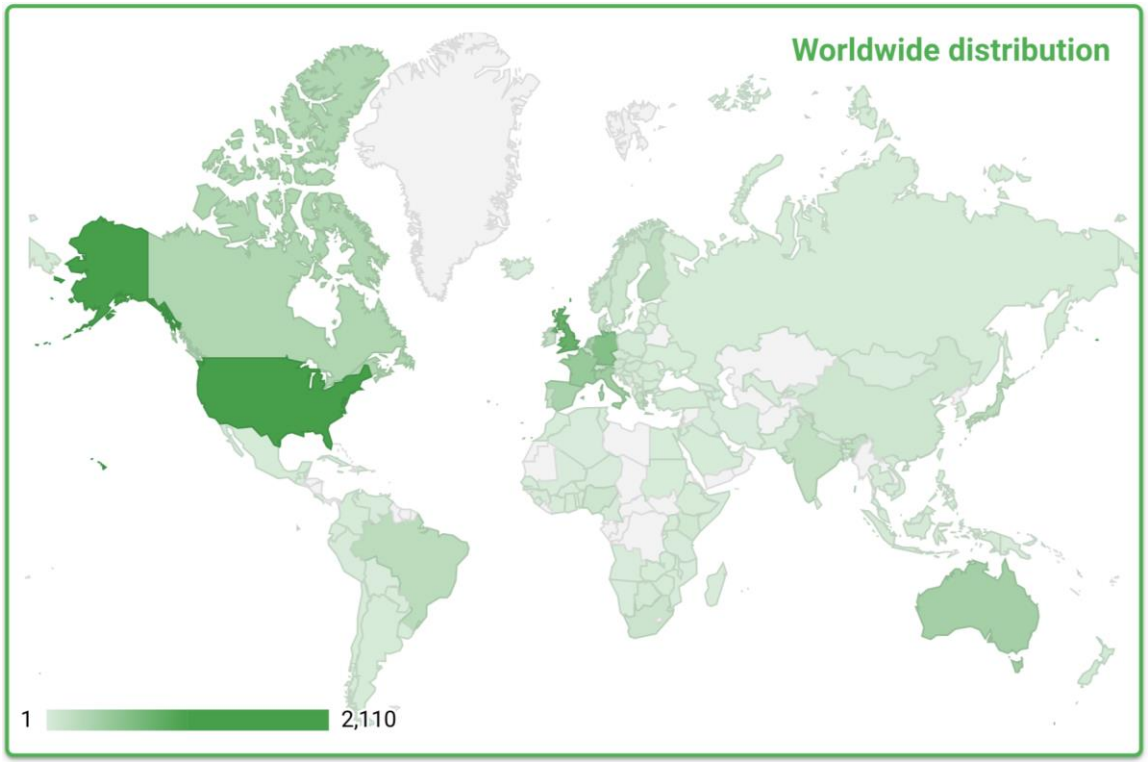
RESEARCH DATA ALLIANCE

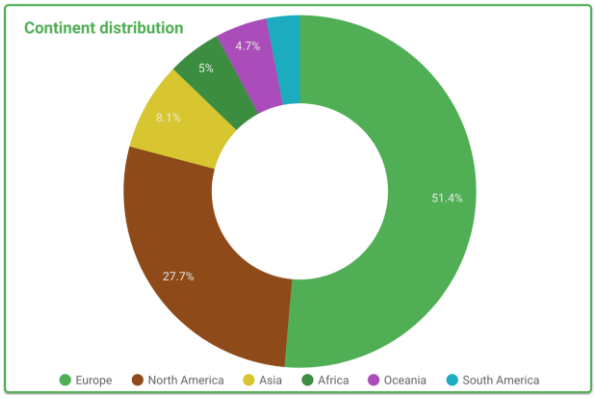


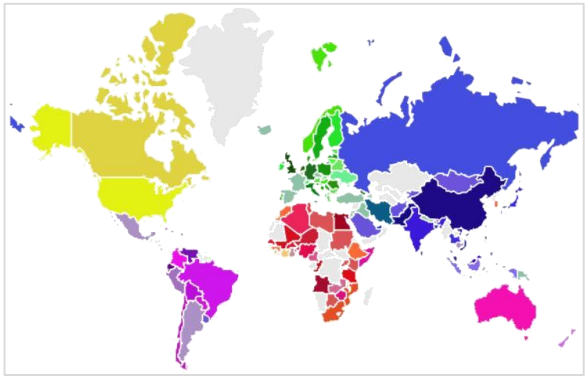
Vision: Researchers and innovators openly share data across technologies, disciplines, and countries to address the grand challenges of society.

RDA Members

8,917 Individual & 58 Organisational Members







Fraunhofer

RDA Working and Interest Groups

Total 87 groups:
32 Workind Groups & 55 Interest Groups

Domain Science - focused

- **Agrisemantics WG**
- **FAIRSharing Registry WG**
- **Capacity Development for Agriculture Data WG**
- **On-Farm Data Sharing (OFDS) WG**
- **Rice Data Interoperability WG**
- **Wheat Data Interoperability WG**
- **Reproducible Health Data Services WG**
- **Preserving Scientific Annotation WG**
- Agricultural Data IG (IGAD)
- Biodiversity Data Integration IG
- Chemistry Research Data IG
- Digital Practices in History and Ethnography IG
- ESIP/RDA Earth, Space, and Environmental Sciences IG
- Geospatial IG
- Global Water Information IG
- **Health Data IG**
- Linguistics Data IG
- RDA/CODATA Materials Data, Infrastructure & Interoperability IG
- Research data needs of the Photon and Neutron Science community IG
- Small Unmanned Aircraft Systems' Data IG
- From Observational Data to Information IG
- Social Sciences & Humanities Research Data IG
- Research Data Management in Engineering

Partnership Groups

- **RDA / TDWG Metadata Standards for attribution of physical and digital collections stewardship WG**
- **RDA/WDS Scholarly Link Exchange WG**
- **RDA/WDS Publishing Data Workflows WG**
- ELIXIR Bridging Force IG
- RDA/NISO Privacy Implications of Research Data Sets IG

RDA Working and Interest Groups

Reference and Sharing – focused

- **Data Citation WG**
- **Data Description Registry Interoperability WG**
- **Empirical Humanities Metadata WG**
- **International Materials Resource Registries WG**
- **Research Data Collections WG**
- **Research Data Repository Interoperability WG**
- **Data Usage Metrics WG**
- Data Discovery Paradigms IG
- National Data Services IG
- RDA/CODATA Legal Interoperability IG
- Reproducibility IG
- Sharing Rewards and Credit (SHARC) IG

Community Needs – focused

- **RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World WG**
- CODATA/RDA Research Data Science Schools for Low and Middle Income Countries
- Archives & Records Professionals for Research Data IG Data for Development IG
- Early Career and Engagement IG
- Education and Training on handling of research data IG
- Ethics and Social Aspects of Data IG
- International Indigenous Data Sovereignty IG
- Open Questionnaire for Research Data Sharing Survey IG
- Data for Development IG
- Research Funders and Stakeholders on Open Research and Data Management Policies and Practices IG

Data Stewardship and Services – focused

- **Brokering Framework WG**
- **DMP Common Standards WG**
- **Exposing Data Management Plans WG**
- **WDS/RDA Assessment of Data Fitness for Use WG**
- **Data Versioning WG**
- **FAIR Data Maturity Model WG**
- Active Data Management Plans IG
- Data in Context IG
- Data Rescue IG
- Domain Repositories IG
- Virtual Research Environments IG
- Libraries for Research Data IG
- Physical Samples and Collections in the Research Data Ecosystem IG
- Preservation Tools, Techniques, and Policies IG
- RDA/WDS Certification of Digital Repositories IG
- Repository Platforms for Research Data IG
- Research Data Architectures in Research Institutions IG
- Data policy standardisation and implementation IG
- GO FAIR IG
- Open Science Graphs IG

Base Infrastructure – focused

- **Array Database Assessment WG**
- **Data Type Registries WG**
- **Metadata Standards Catalog WG**
- **PID Kernel Information WG**
- **Persistent Identification of Instruments WG**
- **Software Source Code ID WG**
- Data Fabric IG
- Data Foundations and Terminology IG
- Disciplinary Interoperability Framework IG
- Big Data IG
- Metadata IG
- PID IG
- Software Source Code IG
- Vocabulary Services IG
- Federated Identity Management IG
- Data Economics IG

Reproducible Health Data Services WG

Home » Working And Interest Groups » Working Group » Reproducible Health Data Services WG

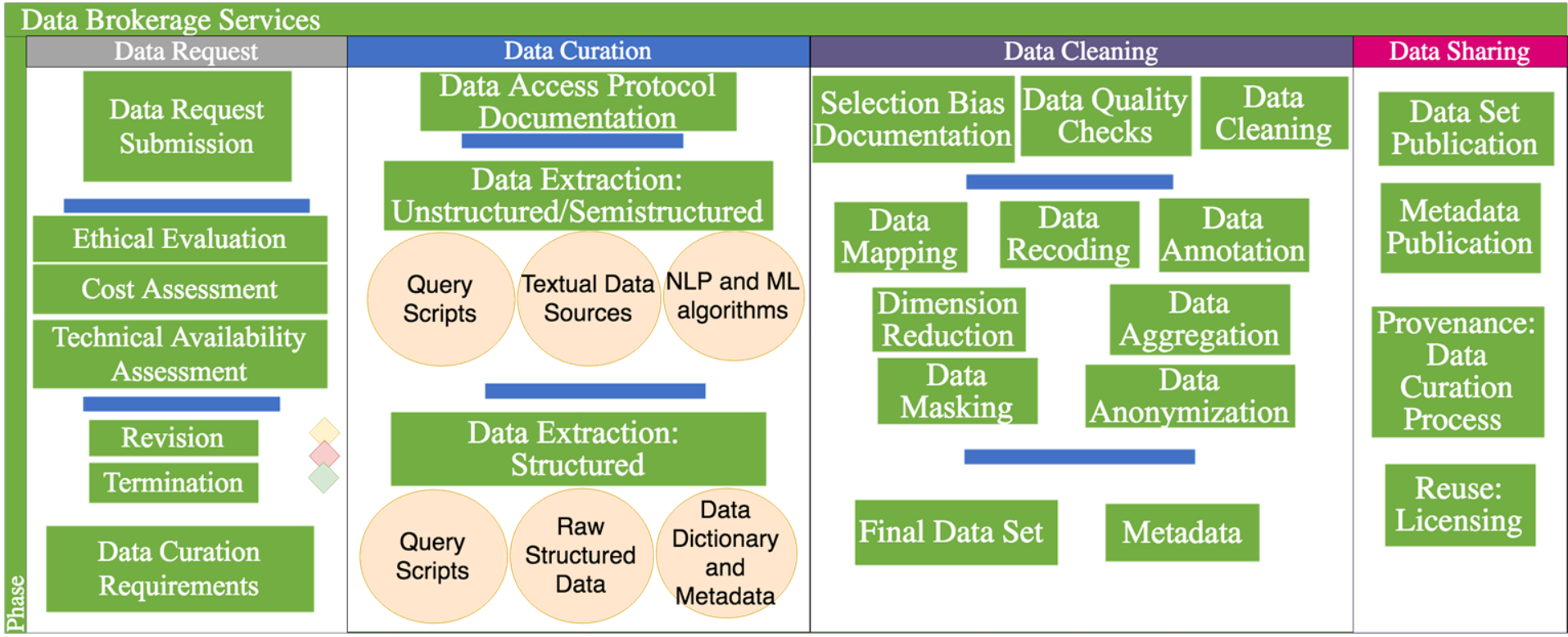
WG

Group details

Status: In Group Revisions
Chair (s): Ludovica Durst, Oya Beyan, Anthony Juehne
TAB Liaison: Frankie Stevens

"to improve the reuse of health data by providing recommendations for reproducible data curation and brokerage workflow services"

Data Curation Processes



Stanford Adoption Case Study

Title	
Cost Assessment	
Data Access	1/10/18 7:00 AM
Data Extraction - Structured	1/10/18 6:13 PM
Data Extraction_Unstructured Data	1/10/18 6:16 PM
Ethical Assessment Review	1/9/18 1:50 AM
Participant Exclusion Criteria	2/9/18 8:55 PM
Participant Inclusion Criteria	1/11/18 2:32 AM
Project Metadata	1/11/18 6:59 AM
RDA_Health Data Service Center_Data Brokerage Workflow	1/11/18 7:21 AM

"The objective of the EOSC is to give the Union a global lead in ***research data management***

and ensure that ***European scientists reap the full benefits of data-driven science,***

by offering in science and technology ***a virtual environment***

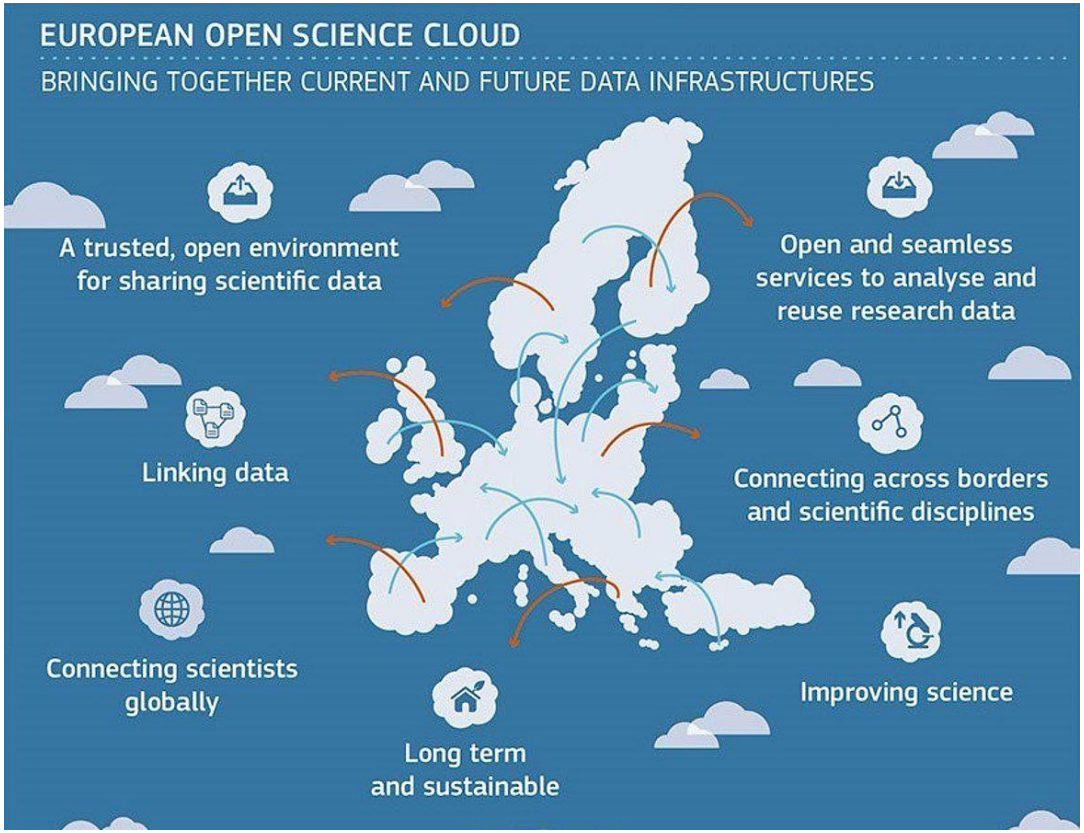
with ***free at the point of use, open and seamless services***

for ***storage, management, analysis and re-use of research data,***

across borders and scientific disciplines."

EOSC Implementation Roadmap (March 2018)

EOSC



Fraunhofer FIT

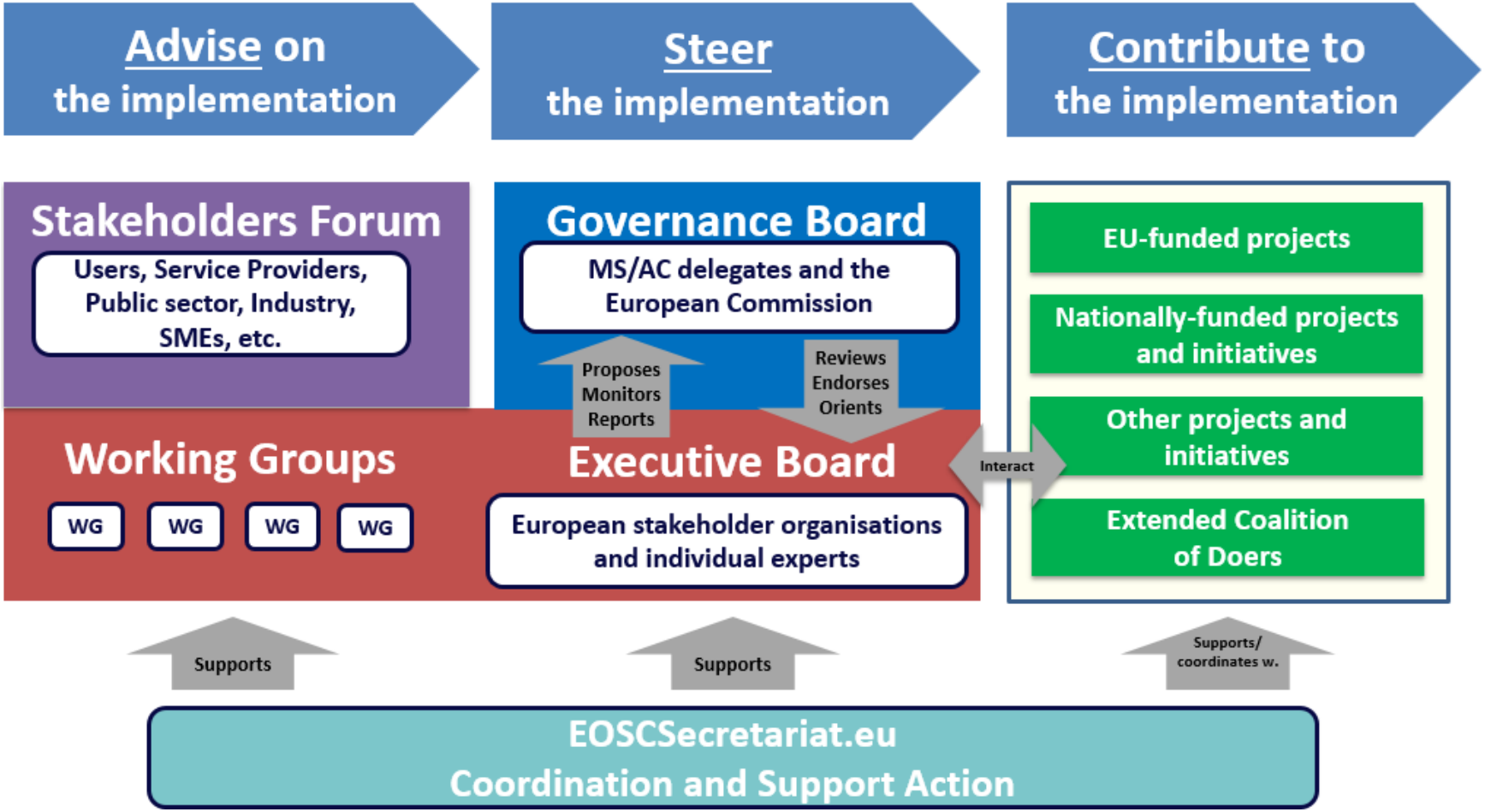
Why the EOSC?

- ✓ Enable Open Science and the digital transformation of science
- ✓ Empower 1.7 million European researchers to reap the benefits of data-driven science
- ✓ Build an open research data commons in Europe widely inclusive of all disciplines and Member States, sustainable in the long-term
- ✓ Give Europe a global lead in research data management; help developing an internet of FAIR digital objects
- ✓ Reduce fragmentation by federating existing Research Infrastructures (including e-Infrastructures)
- ✓ Enable interdisciplinary research to address societal challenges
- ✓ Enable the Digital Single Market and stimulate the emergence of a competitive EU cloud sector

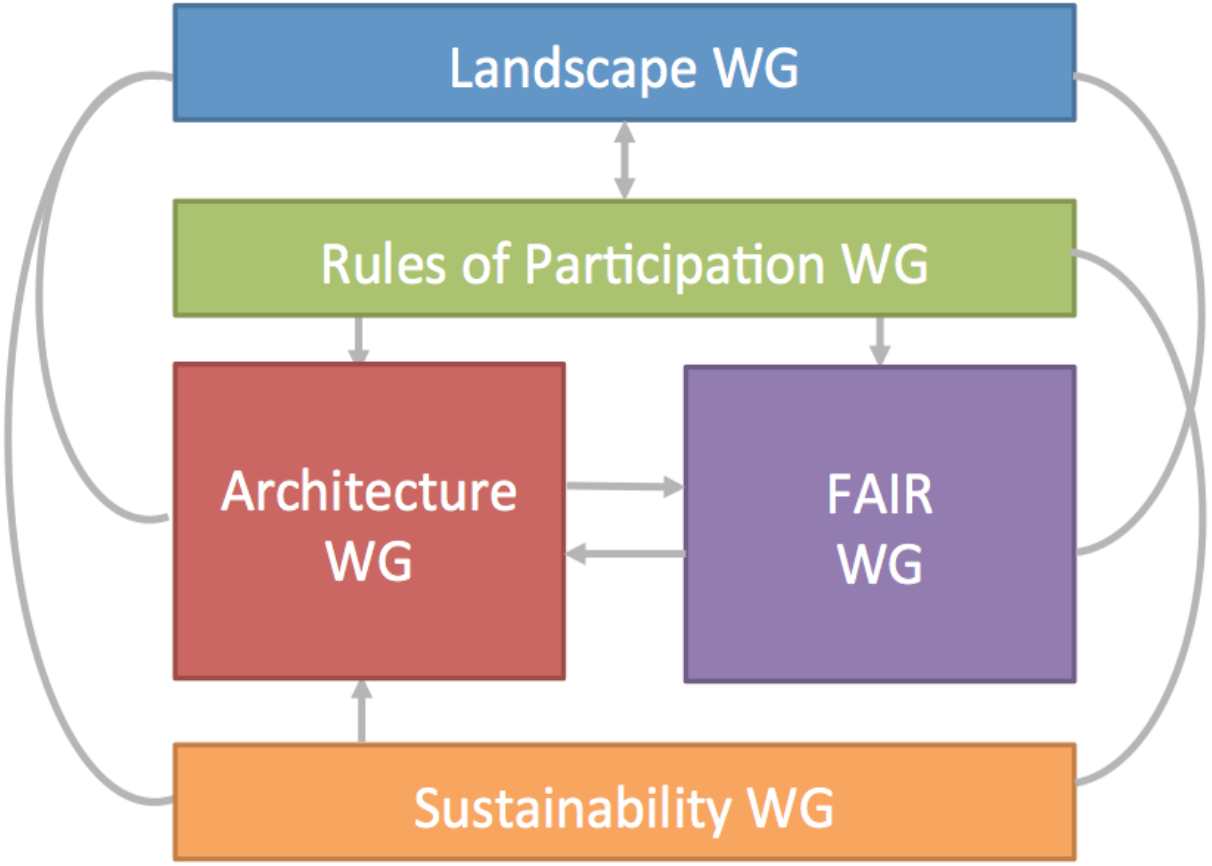
What is the EOSC?

- ✓ A trusted and open virtual environment for the scientific community with seamless access to services (with highest TRLs) addressing the whole research data life cycle:
  - ✓ Federating core services to federate and connect existing or planned RIs
  - ✓ Services to make data FAIR, store them, ensure their long-term preservation
  - ✓ Services to find, access, combine, analyse and process data
  - ✓ Protected, personalized work environment
- ✓ FAIR competition, no lock-in by individual service providers
- ✓ Free at the point of use for the researchers
- ✓ User-oriented and inclusive (across borders and disciplines)
- ✓ Accessible through a non-exclusive, simple, universal gateway
- ✓ Governed by a minimal set of Rules of Participation
- ✓ Steered by a governance based on increasing mutuality

EOSC governance structure



EOSC Executive Board Working Groups



https://www.eoscsecretariat.eu/eosc-working-groups

EOSCFAIR
Executive Board Working Group

Chair: Sarah Jones
Independent Expert

WHAT

Provide recommendations on the implementation of FAIR, including corresponding requirements for services, in order to foster cross-disciplinary interoperability

WHY

Connect people, data and service via standards

Be the glue

HOW

1. Data standards & sharing agreements
2. Upscale best-practice solutions
3. EOSC Interoperability Framework
4. Identify service requirements for FAIR
5. Persistent Identifier Policy for EOSC
6. Frameworks to assess FAIR data and certify services that enable FAIR
7. Converge towards globally-accepted frameworks

Q2 2019
2019 Annual FAIR work plan

Q4 2019
PID policy defined

Outline metrics for FAIR data & service certification

Q1 2020
2020 Annual FAIR work plan

testing & iterating

Q3 2020
EOSC Interoperability Framework

Q4 2020
Updated PID policy

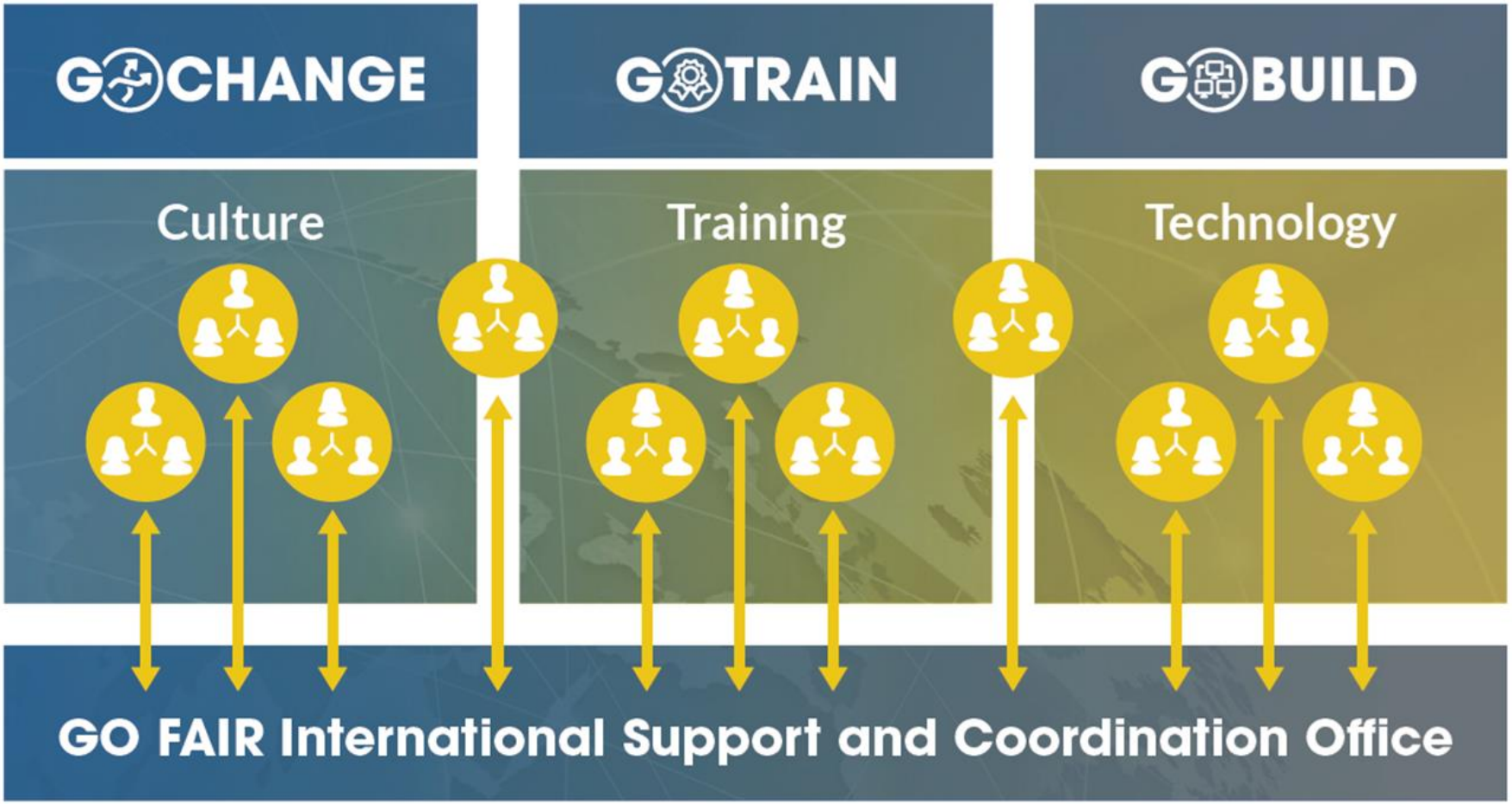
Updated FAIR metrics & service certification

EOSCsecretariat webinar, 1st July 2019

Fraunhofer FIT

- supports existing and emerging community practices for establishing the EOSC as part of the **Global Internet of FAIR Data and Services** (IFDS)
- aims to **kick-start the development of the EOSC** through communities of excellence ‘Implementation Networks’ voluntarily committed to collectively engage in the IFDS

GO FAIR Pillars



GO FAIR Implementation Networks

Fraunhofer FIT

Distributed Analytics on Sensitive Medical Data

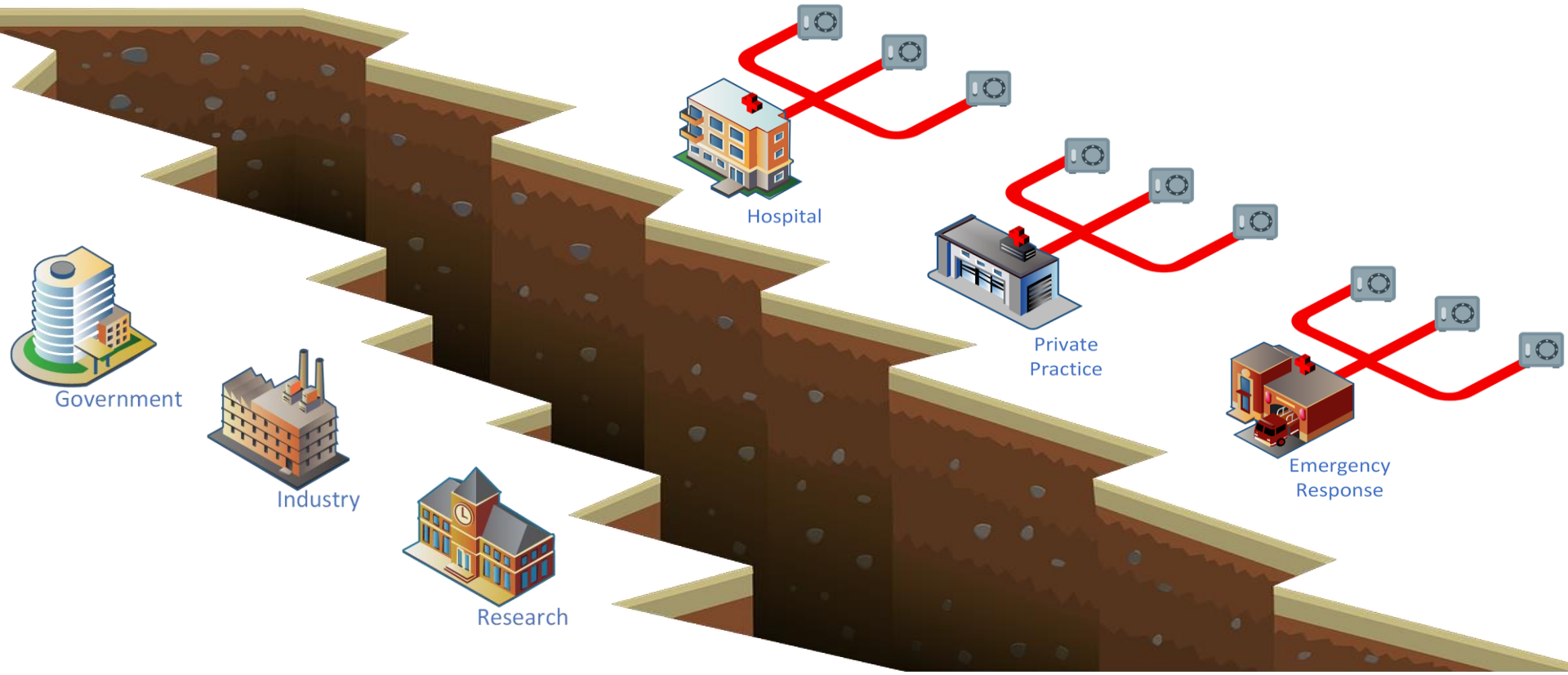
Learning from Data

Fraunhofer FIT

Data Sharing

Can we really share sensitive data ?

… *reality*



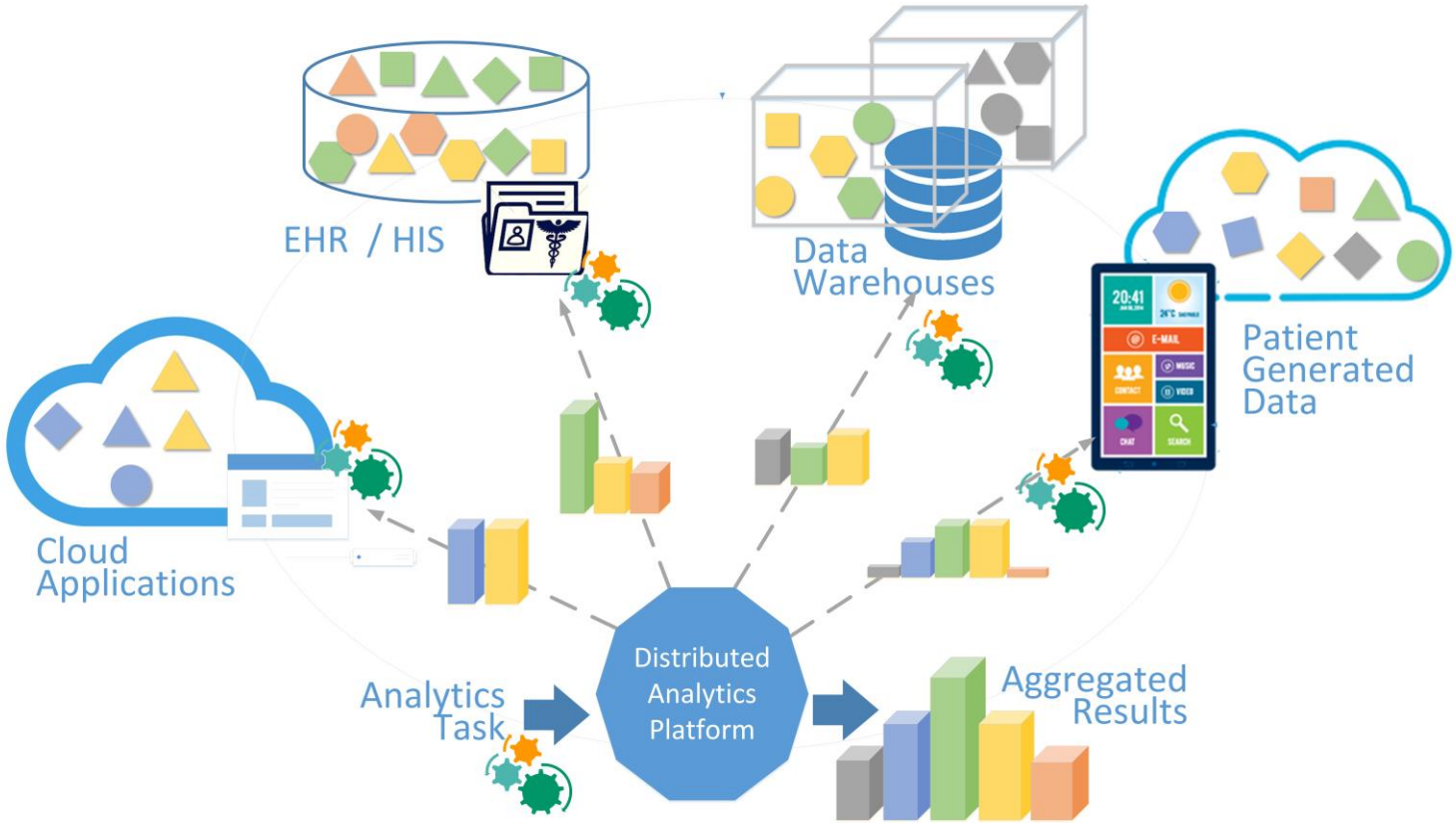
Fraunhofer FIT

Personal Health Train Approach

bring algorithms to the data

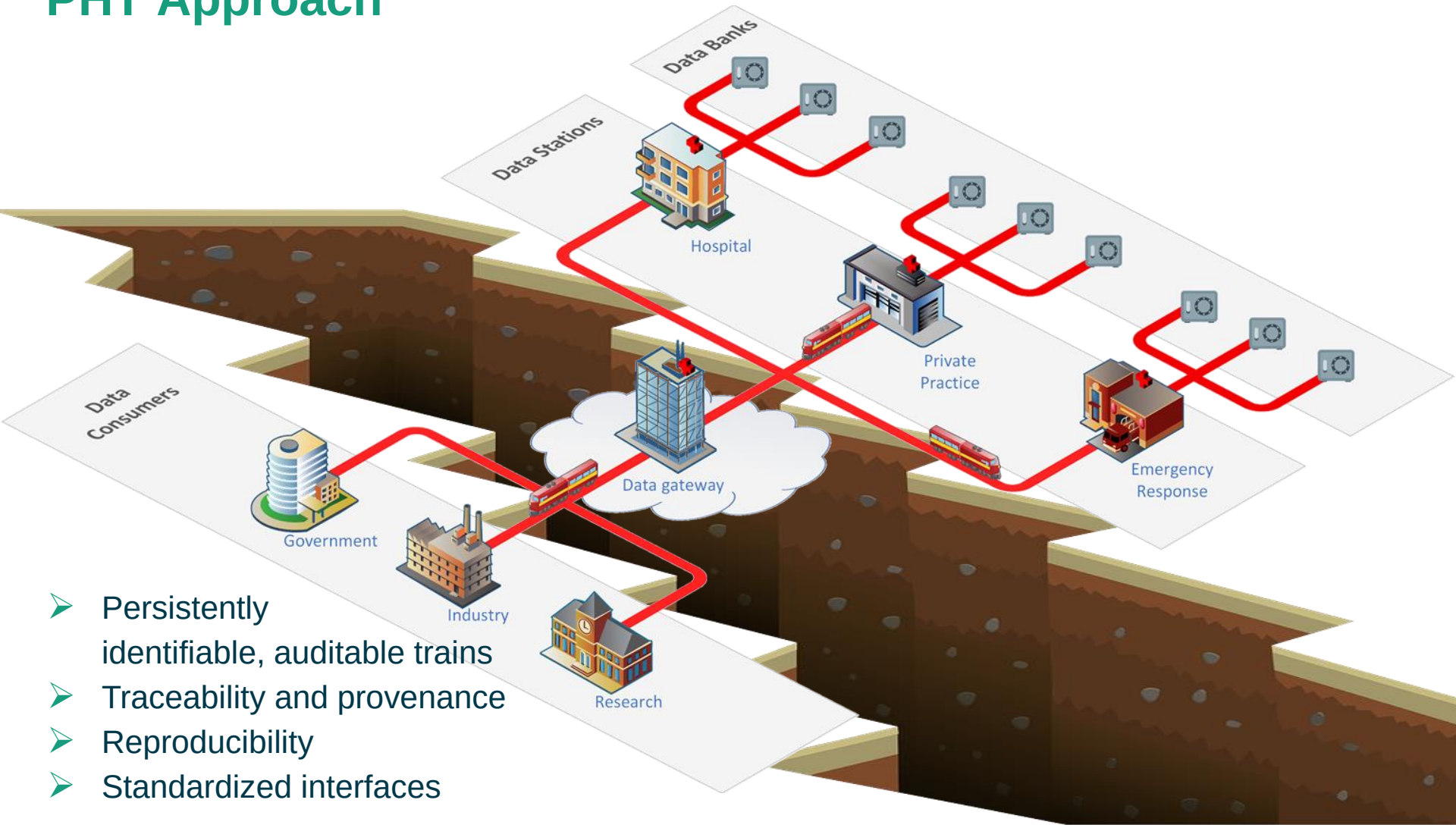
……rather than bringing all data to a central repository

Sharing Algorithms to Develop Analytical Models



Fraunhofer FIT

PHT Approach



- Persistently identifiable, auditable trains
- Traceability and provenance
- Reproducibility
- Standardized interfaces

Fraunhofer FIT

We are working together

…… **be part of it !**

contact us

Reproducible Health Data Services WG

https://www.rd-alliance.org/groups/reproducible-health-data-services-wg

Personal Health Train Implementation Network

https://www.go-fair.org/implementation-networks/overview/personal-health-train/

EOSCFAIR Executive Board Working Group

Oya Beyan beyan@fit.fraunhofer.de

https://www.eoscsecretariat.eu/working-groups/fair-working-group