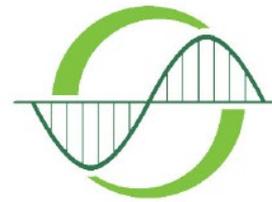




UNIVERSITÀ
degli STUDI
di CATANIA



DIPARTIMENTO di
FISICA e ASTRONOMIA
“Ettore Majorana”



The State of Play in the Open Science and Digital Repositories

Prof. Roberto Barbera (roberto.barbera@ct.infn.it)
University of Catania - Italy

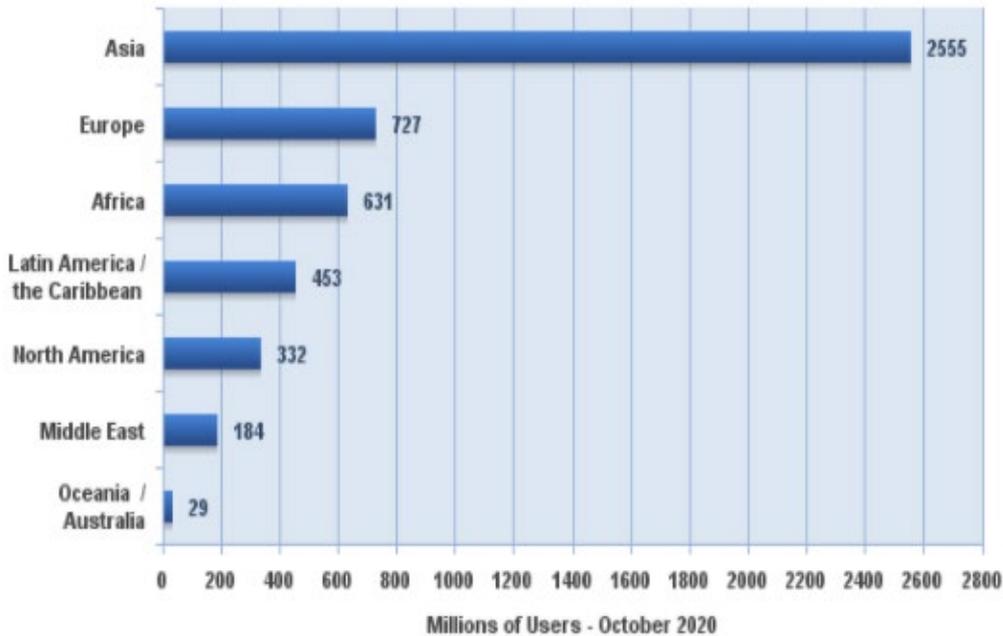
SORER Introduction Workshop for University Leadership - 4 January 2021



Outline

- ▶ Introductory concepts, definitions and considerations
- ▶ Open Science: definitions and tools
 - ▶ The role of FAIR principles compliant digital repositories and persistent identifiers
- ▶ A success story from Ethiopia
- ▶ The Knowledge Workflow and the Knowledge “Nexi”
- ▶ Summary and conclusion

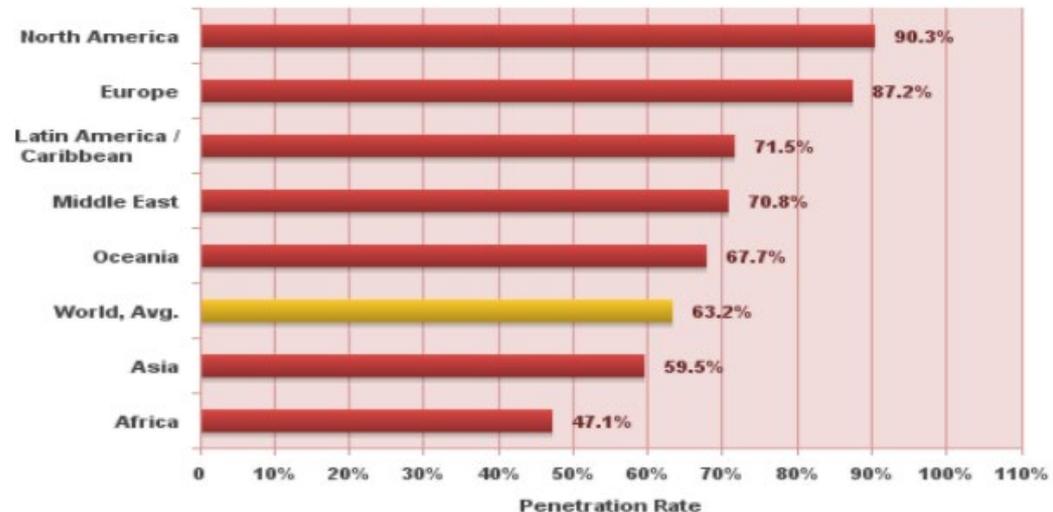
Internet Users in the World by Geographic Regions - 2020 Q3



Internet figures

- **2020 Q3: 63,2% of world population is on Internet (1)**
- **1,250% increase in the last 20 years**

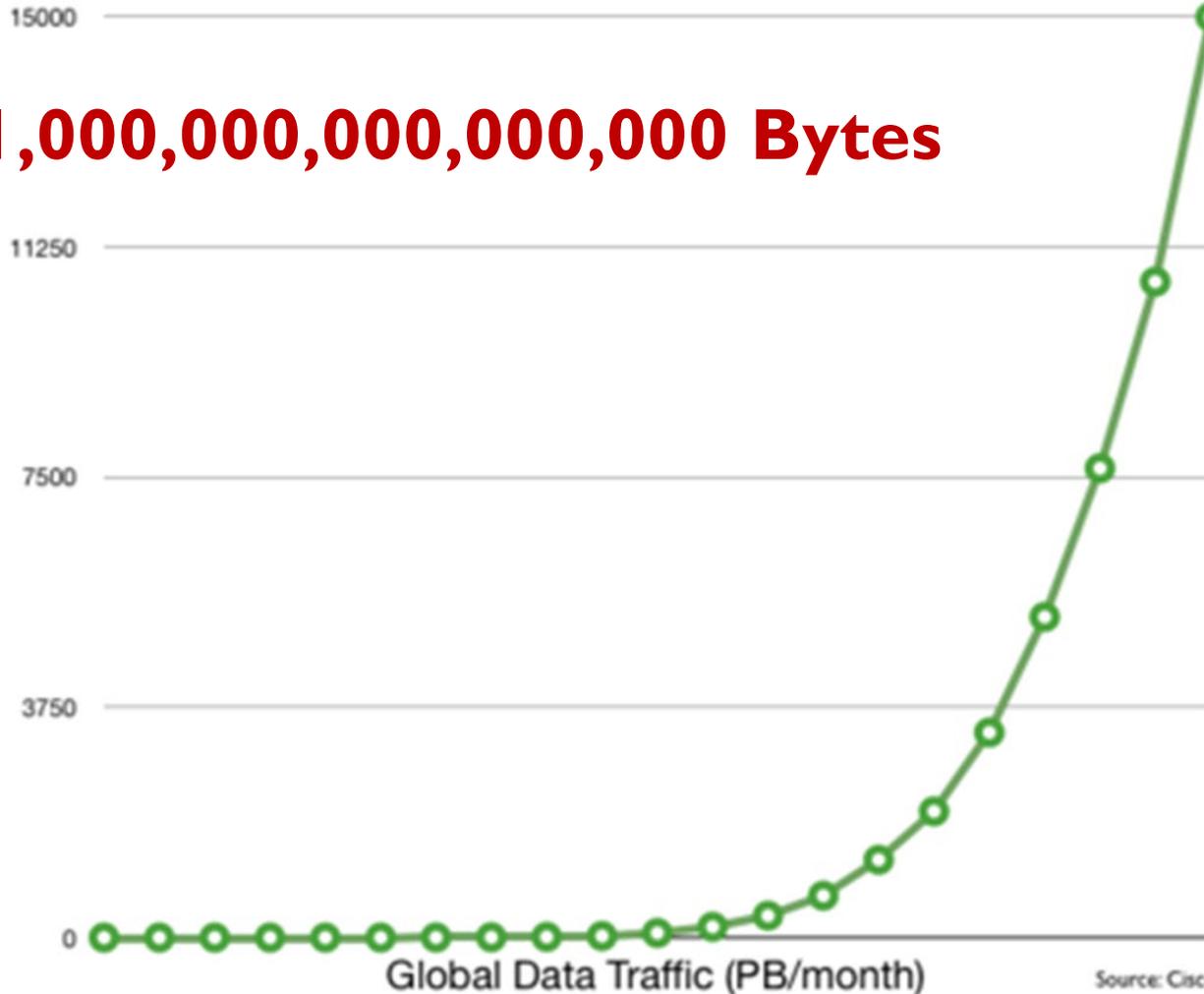
Internet World Penetration Rates by Geographic Regions - 2020 Q3



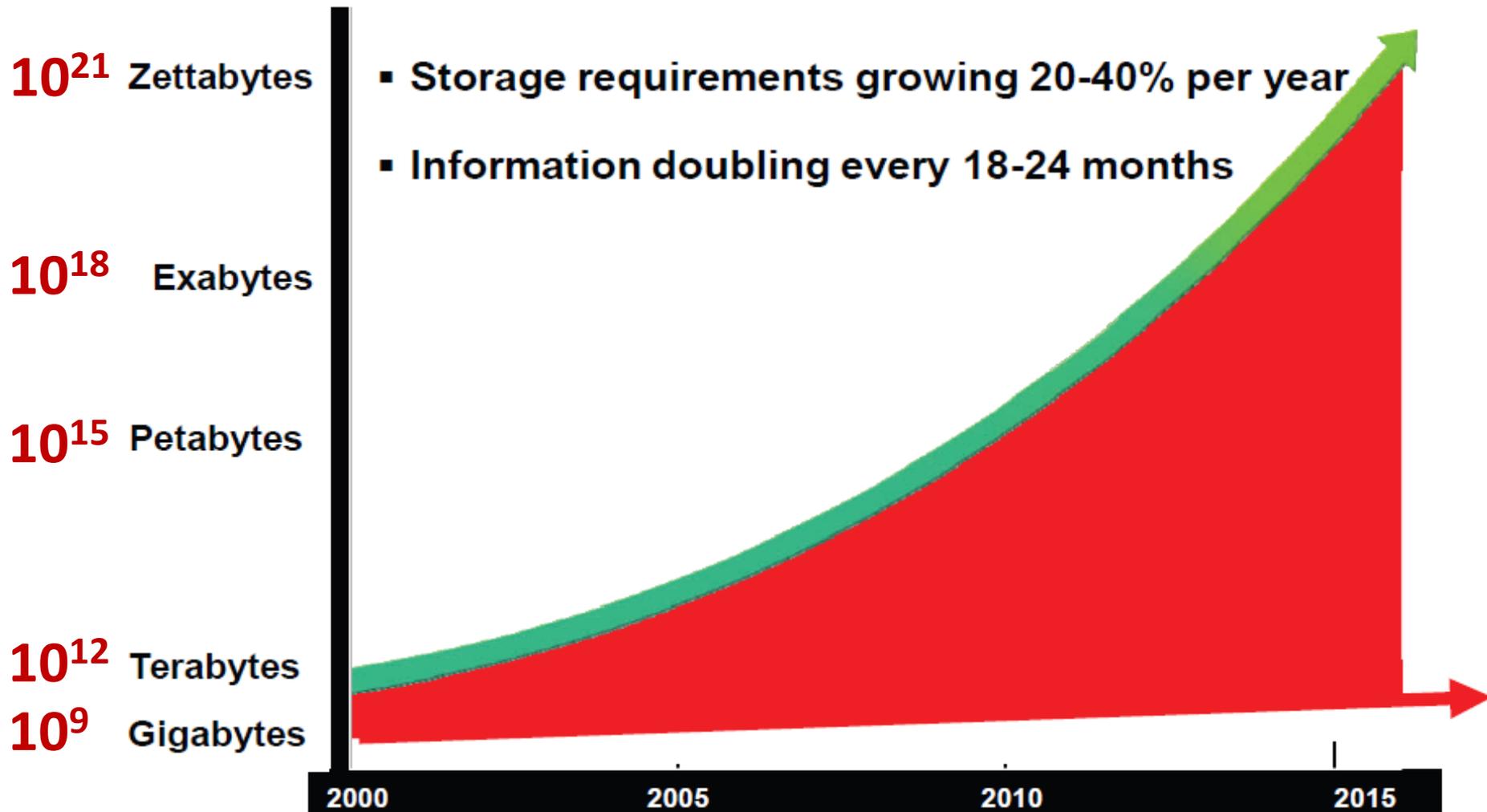
Internet Exponential Data Growth

Internet Data Growth (1990-2010)

1 PB = 1,000,000,000,000,000 Bytes



Estimated Data Growth - Worldwide



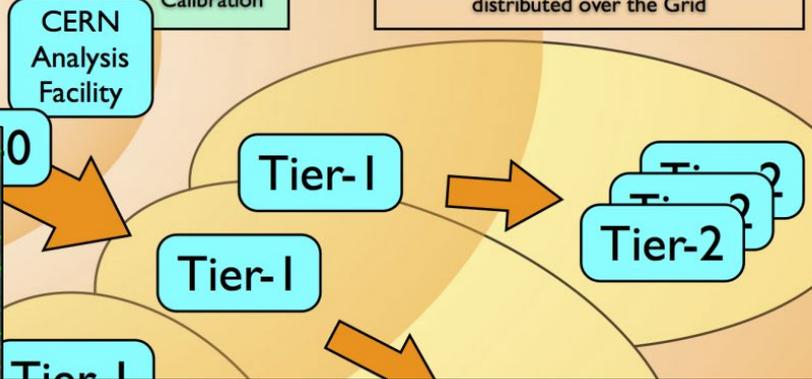
Scientific Data Growth (High-energy Physics)



ALICE

200Hz - 400Hz
RAW: ~1.7-1.1 MB/evt

Event Summary Data (ESD): ~1 MB/evt
Analysis Object Data (AOD): ~100 kB/evt
derived data (dESD, dAOD, NTUP,...)
distributed over the Grid



TRENDING: Special Report: Platforms Play Big in the Cloud · Special Report: Hacking the election · Newsletters · Resources/White

COMPUTERWORLD
FROM IDG

INSIDER Sign In | Register

0.5 EB = 500 PB

Home > Cloud Computing > Cloud Storage

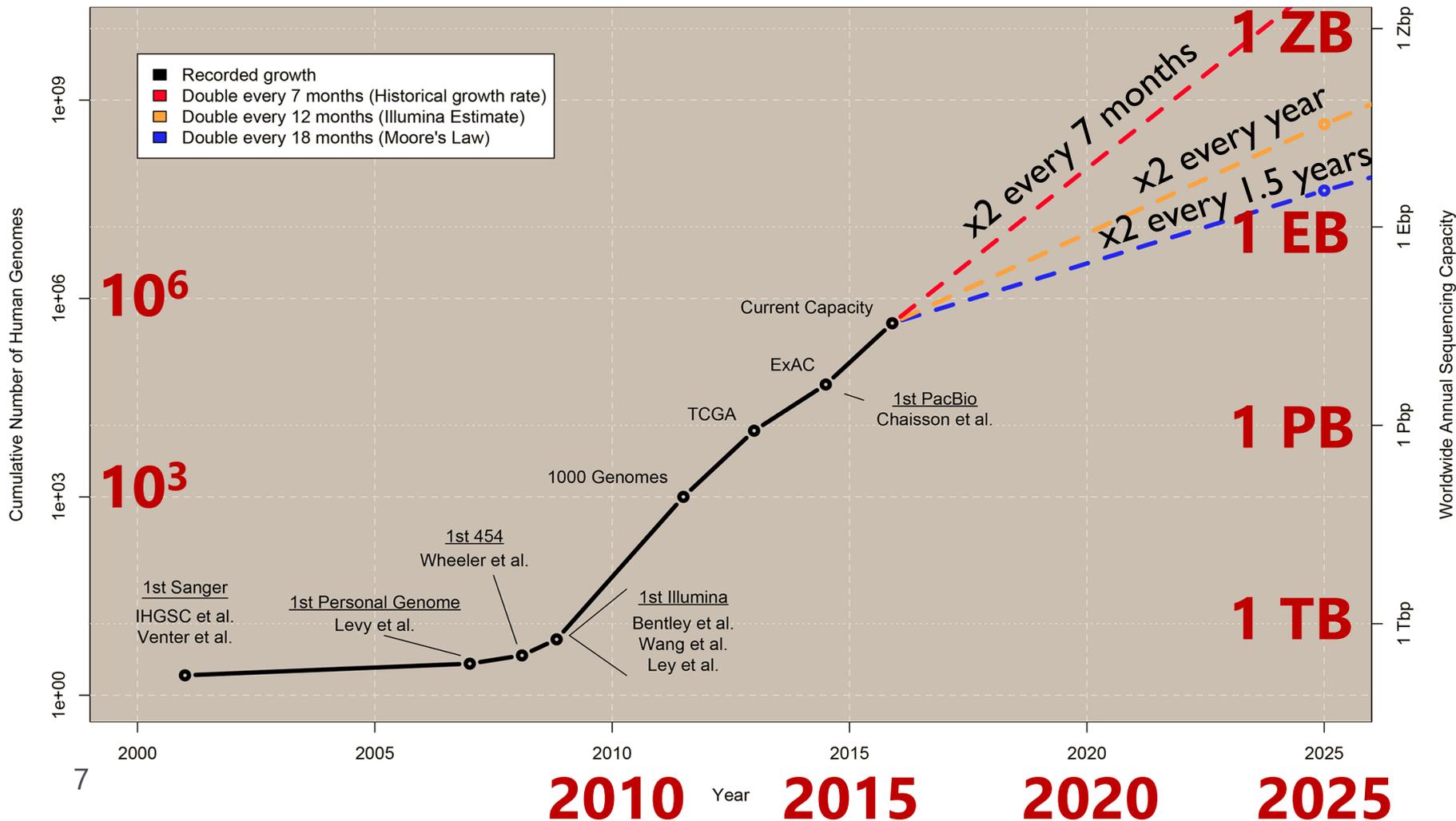
NEWS

CERN's data stores soar to 530M gigabytes

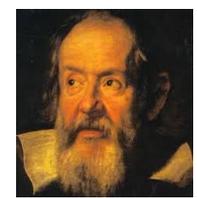
The Large Hadron Collider's detectors record 14 million photos a second

Scientific Data Growth (Bioinformatics)

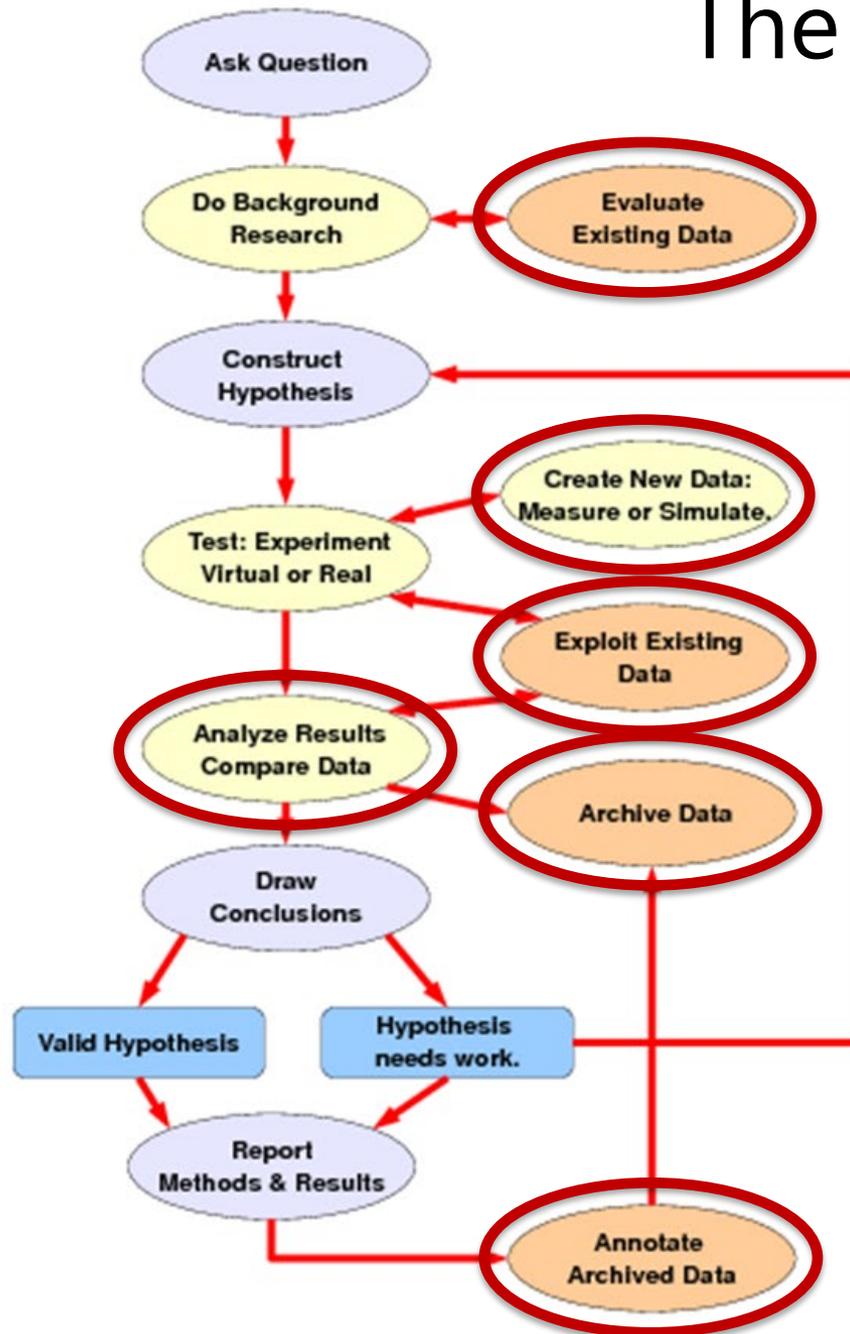
Global growth of DNA sequencing (per year)



The Scientific Method



G. Galilei



The «pillars» of the Scientific Method

- **Repeatability**

- The closeness of agreement between independent results obtained with the same method on identical test material, under the same conditions (same operator, same apparatus, same laboratory and after short intervals of time)
- Affected by *random errors*

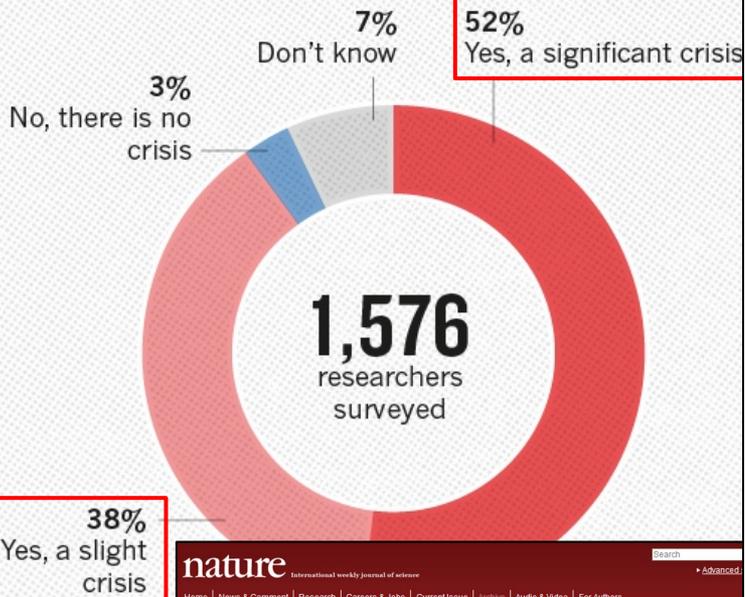
- **Reproducibility**

- The closeness of agreement between independent results obtained with the same method on identical test material but under different conditions (different operators, different apparatus, different laboratories and/or after different intervals of time)
- Affected by *systematic errors*

Is science reproducible?

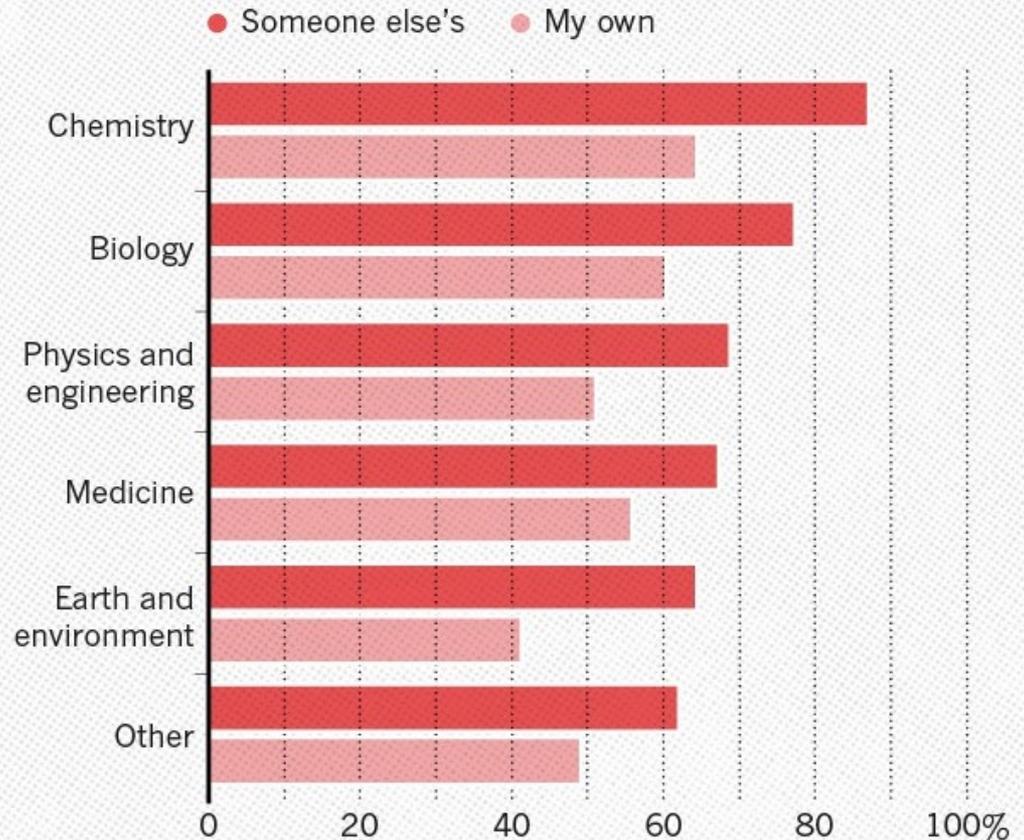
Reproducibility «crisis» survey

IS THERE A REPRODUCIBILITY CRISIS?



HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Specials & supplements archive > Challenges in irreproducible research

Take part in Nature Publishing Group's annual reader survey here for the chance to win a MacBook Air. Find out more

SPECIAL See all specials

CHALLENGES IN IRREPRODUCIBLE RESEARCH

No research paper can ever be considered to be the final word, and the replication and corroboration of research results is key to the scientific process. In studying complex entities, especially animals and human beings, the complexity of the system and of the techniques can all too easily lead to results that seem robust in the lab, and valid to editors and referees of journals, but which do not stand the test of further studies. Nature has published a series of articles about the worrying extent to which research results have been found wanting in this respect. The editors of Nature and the Nature life sciences research journals have also taken substantive steps to put our own houses in order in improving the transparency and robustness of what we publish. Journals, research laboratories and institutions and funders all have an interest in tackling issues of irreproducibility. We hope that the articles contained in this collection will help.

Free full access

- Editorial
- Features
- News and analysis
- Comment
- Perspectives and reviews

EDITORIAL

Journals unite for reproducibility
Consensus on reporting principles aims to improve the quality of the biomedical literature
Nature 515, 7 (6 November 2014)

Code share
Papers in Nature journals should make code available
Nature 514, 559 (29 October 2014)

Reducing our irreproducibility
Nature 496, 398 (25 April 2013)

Further confirmation needed
A new mechanism for independently repeating experiments must realize the pitfalls of work
Nature Biotechnology 30, 905 (10 September 2012)

Error prone
Biologists must realize the pitfalls of work
Nature 487, 400 (20 July 2012)

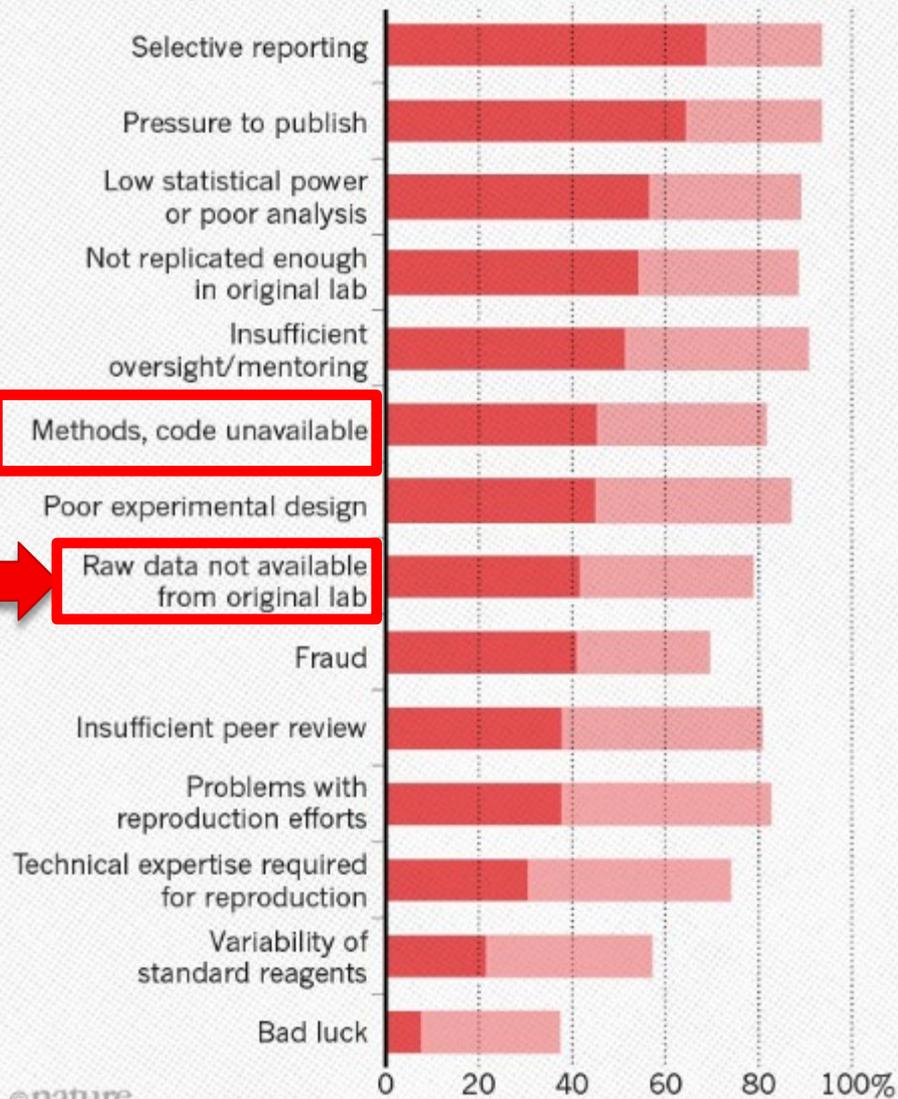
Must try harder
Too many sloppy mistakes are creeping into the data — and at themselves.
Nature 483, 909 (29 March 2012)

Reproducibility «crisis» survey

WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

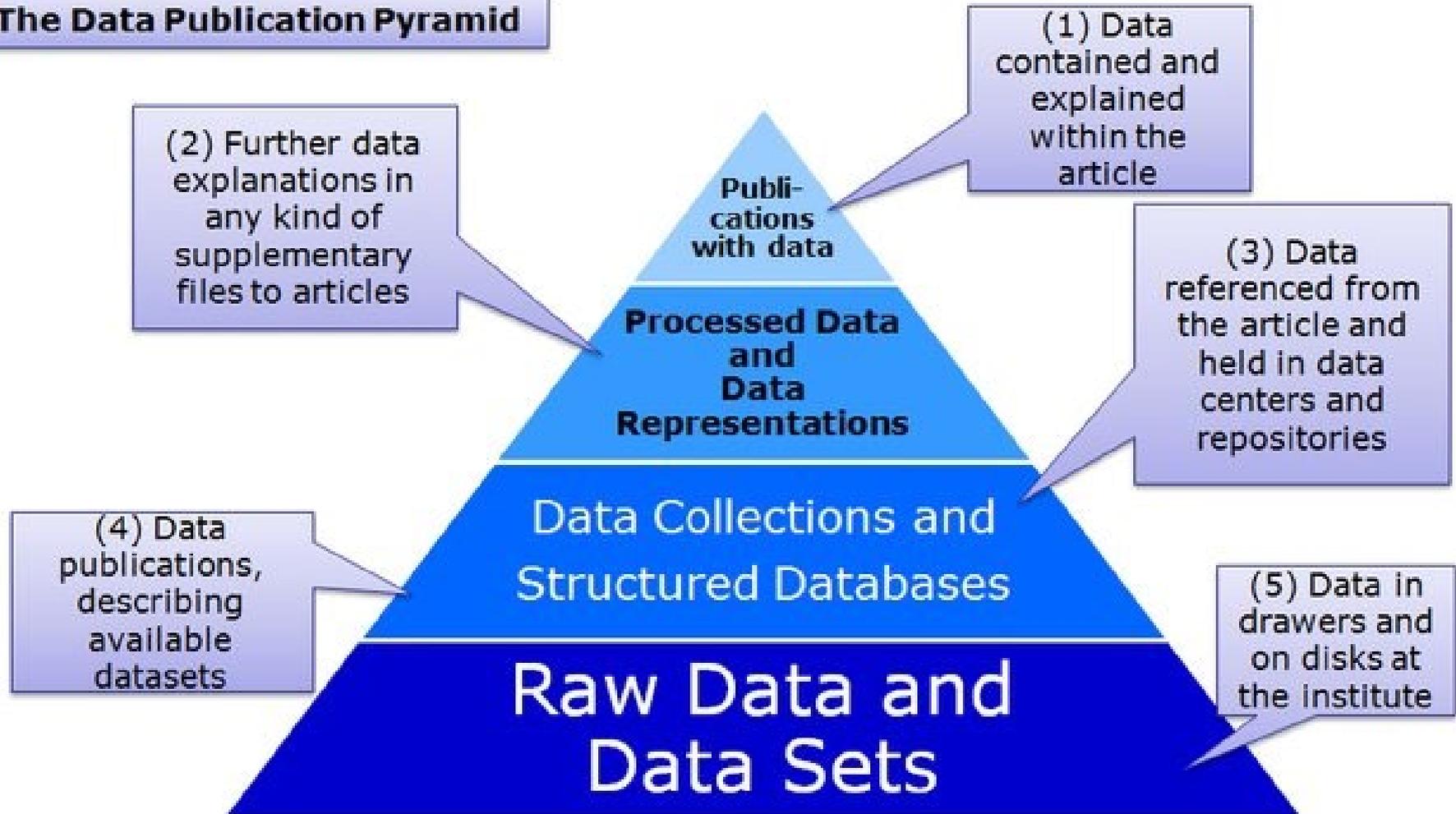
Many top-rated factors relate to intense competition and time pressure.

● Always/often contribute ● Sometimes contribute

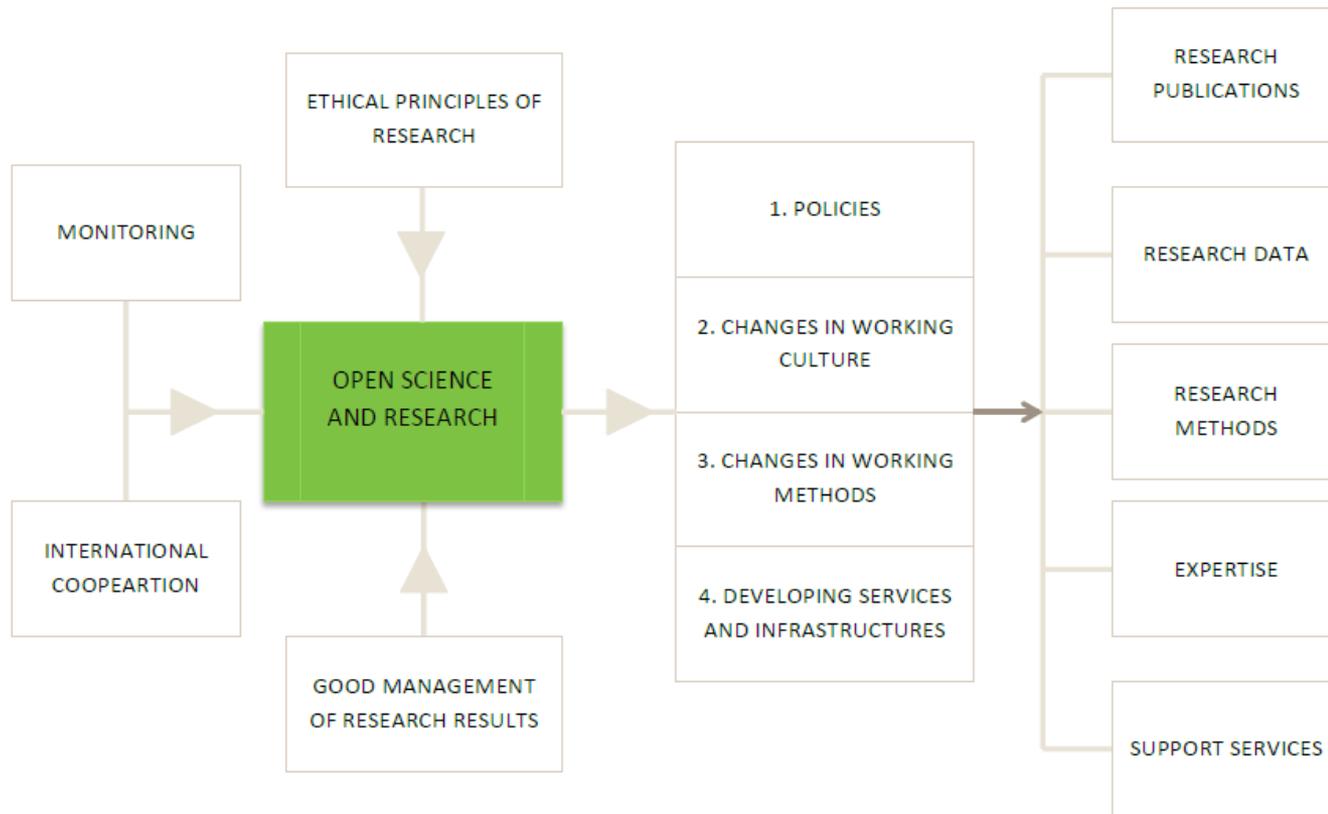


The Data pyramid

The Data Publication Pyramid



Open Science enablers and beneficiaries



Researchers

- Visibility
- Credits
- Funding
- Networking

Research teams

- Visibility
- Funding
- Cost-effectiveness
- Networking

Organisations

- Funding
- Quality
- Cost-effectiveness

Decision-makers and financiers

- Decisions based on better information
- Increased impact of funding
- Quality

The general public

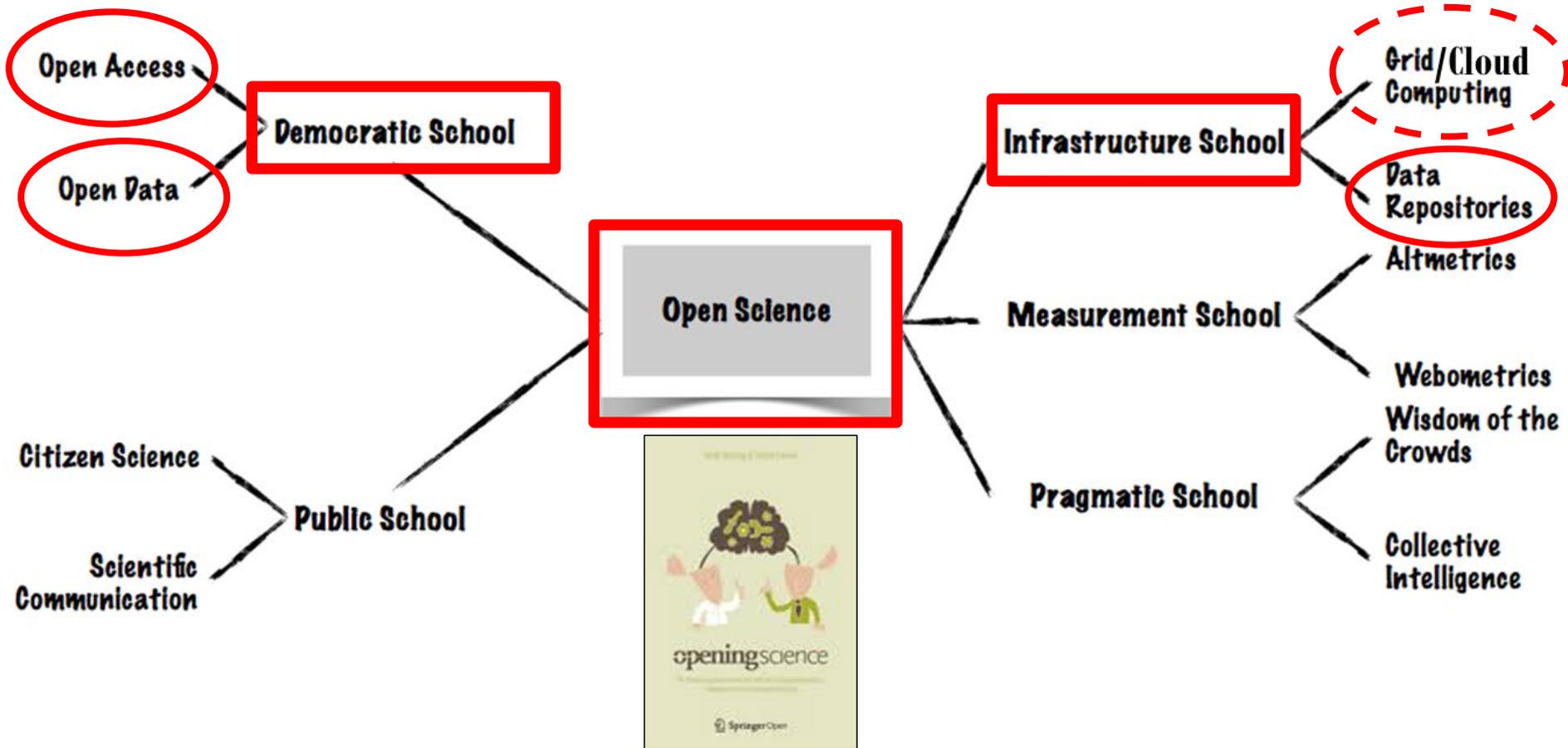
- Increased understanding and expertise
- Increased opportunities to have an influence

National level

- The promotion of human rights and democracy
- Decisions based on research data

Open Science Schools of Thought

(<http://book.openingscience.org>)



FAIR Data Principles

(<https://www.go-fair.org/fair-principles/>)

Findable:

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

Accessible:

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

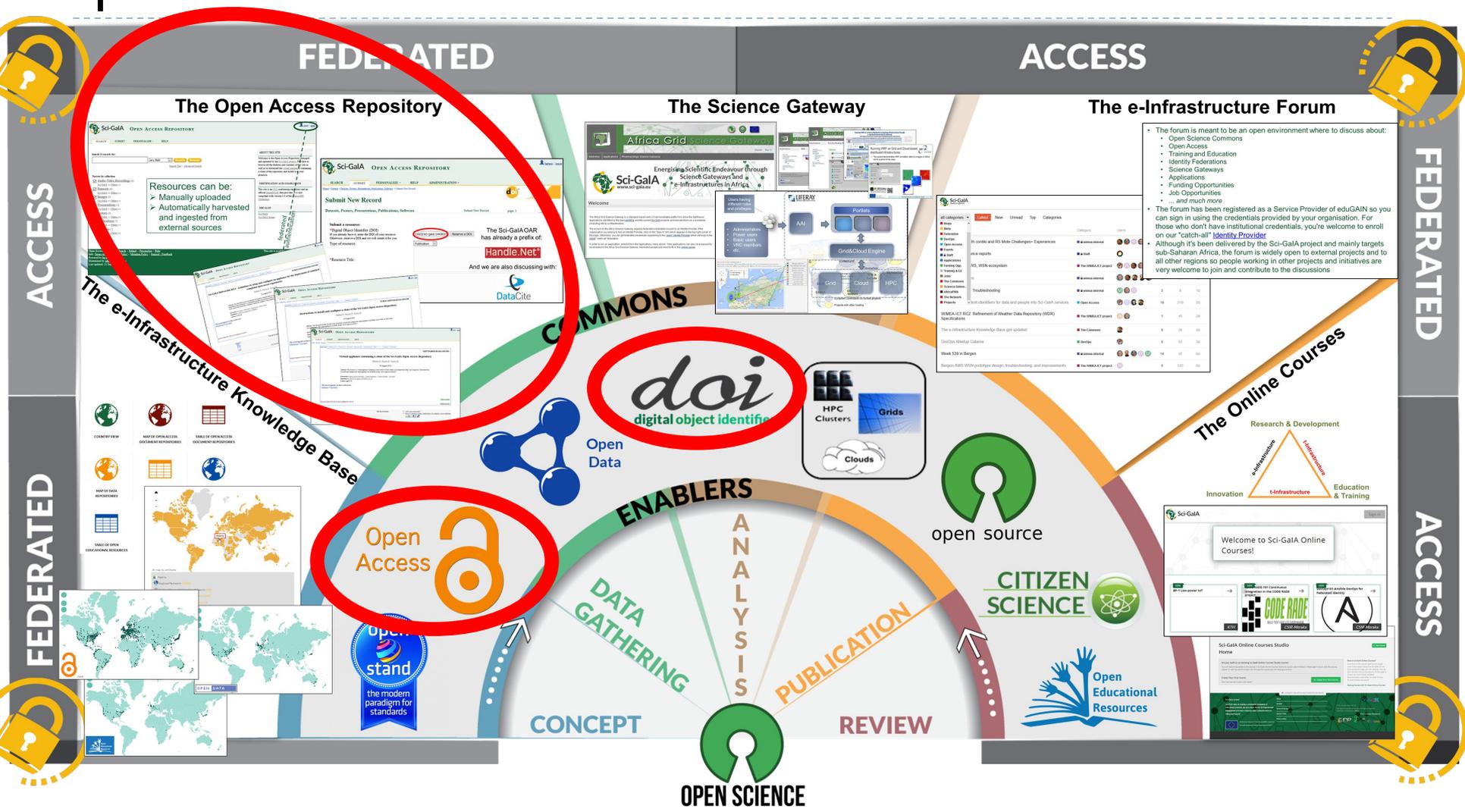
Interoperable:

- 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

Reusable:

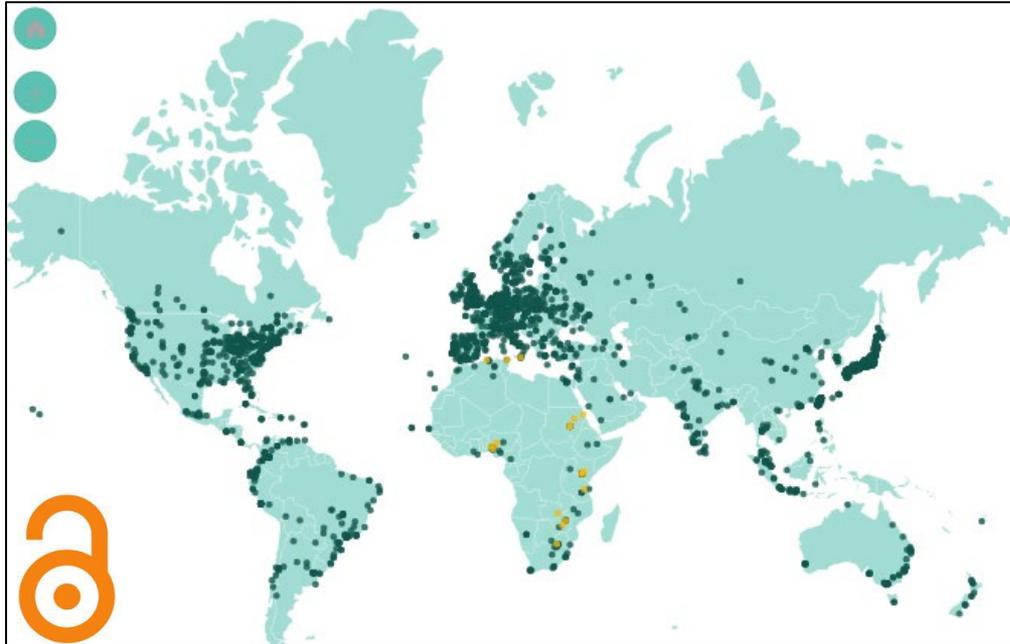
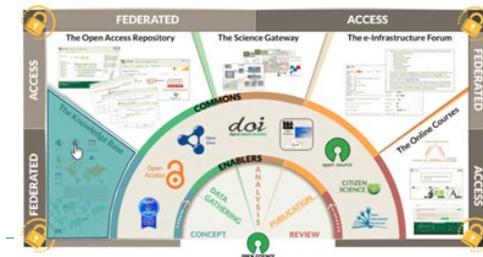
- R1. Meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (Meta)data are released with a clear and accessible data usage license
 - R1.2. (Meta)data are associated with detailed provenance
 - R1.3. (Meta)data meet domain-relevant community standards

The Sci-GaIA Federated Platform for an Open Science Commons in Africa



The e-Infrastructure Knowledge Base

(www.sci-gaia.eu/knowledge-base)

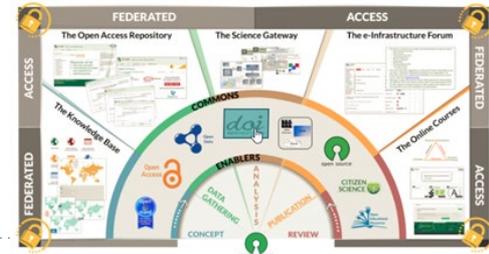


OPEN DATA

- > 4,000 repositories
- $3 \cdot 10^7$ docs/datasets/OERs
- **Few repositories in Africa, even less the FAIR-compliant ones**



Persistent identifiers for African research outputs

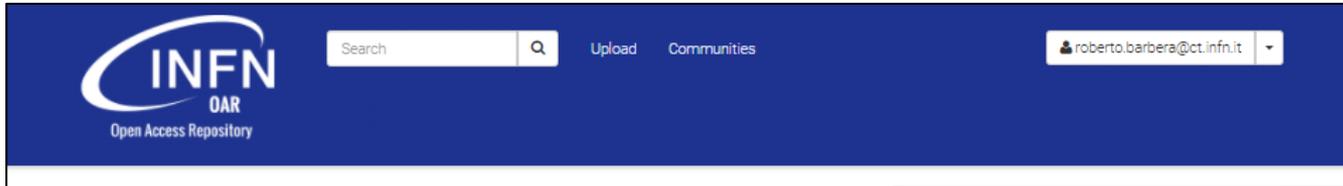


- In January 2017, an agreement was established between UNICT and the Conference of Italian University Rectors (CRUI) to extend the possibility to provide Digital Object Identifiers (DOI) prefixes to African organizations wishing to deploy an institutional Open Access Repository
- 4 DOI prefixes released so far in Africa:
 - The African Population and Health Research Centre (KE and Intl.)
 - The Eko-Konnect Research and Education Initiative (NG)
 - The Ministry of Education of Ethiopia, through EthERNet
 - **The UbuntuNet Alliance (MW and Intl.)**



The INFN Open Access Repository

(www.openaccessrepository.it)



Latest entries

January 2, 2021 (v270) [Plot](#) [Open Access](#) [View](#)

CovidStat project summary plots

Lista, Luca; Menasce, Dario; Mezzetto, Mauro; Pedrini, Daniele; Spighi, Roberto

This record contains the daily update summary plots of the data of the CovidStat project. CovidStat is a project carried out by the CovidStat Working Group at INFN, whose creation was promoted within the Italian National Institute of Nuclear Physics with the aim of making a statistical analysis...

Uploaded on January 3, 2021
269 more version(s) exist for this record

January 2, 2021 (v269) [Dataset](#) [Open Access](#) [View](#)

CovidStat project data

Lista, Luca; Menasce, Dario; Mezzetto, Mauro; Pedrini, Daniele; Spighi, Roberto

This record contains the daily updated data of the CovidStat project. CovidStat is a project carried out by the CovidStat Working Group at INFN, whose creation was promoted within the Italian National Institute of Nuclear Physics with the aim of making a statistical analysis of the data provided...

Uploaded on January 3, 2021
268 more version(s) exist for this record

December 28, 2020 (v8) [Software](#) [Open Access](#) [View](#)

CovidStat project analysis code

Lista, Luca; Menasce, Dario; Mezzetto, Mauro; Pedrini, Daniele; Spighi, Roberto; Varaschin, Antonella; Zoccoli, Antonio

This record contains the software to analyse the data of the CovidStat project. CovidStat is a project carried out by the CovidStat Working Group at INFN, whose creation was promoted within the Italian National Institute of Nuclear Physics with the aim of making a statistical analysis of the...

November 23, 2020 [Technical note](#) [Open Access](#) [Edit](#)

Synthetic procedure to replicate the INFN Open Access Repository (based on Invenio v3 and Zenodo)

Fargetta, Marco; Rotondo, Riccardo; Barbera, Roberto

This document contains a synthetic procedure to replicate the INFN Open Access Repository

Preview

(Version 4 - November 23, 2020)

Synthetic procedure to replicate the INFN Open Access Repository (based on Invenio v3 and Zenodo)

M. Fargetta¹, R. Rotondo¹, R. Barbera^{2,1}

1) Italian National Institute of Nuclear Physics, Division of Catania - Italy
2) Department of Physics and Astronomy "E. Majorana" of the University of Catania - Italy

533 views 291 downloads [See more details...](#)

Publication date: November 23, 2020
DOI: [10.15161/oar.it/73283](https://doi.org/10.15161/oar.it/73283)
Keyword(s): [infn](#) [openaccessrepository](#) [invenio](#) [zenodo](#)
Related identifiers: [References: 10.15161/oar.it/21260](#)
Communities: [INFN](#)

Based on **INVENIO** and



with some specific add-on's

A success story...



January 2018 - NADRE mission established

- The National Academic Digital Repository of Ethiopia (NADRE) intends to provide researchers, lecturers, students and stakeholders access to all research works published by Ethiopian universities and research institutions
- All public and private universities as well as research institutions will contribute to the NADRE and will provide all citizens access to it
- In order to complement and feed the NADRE each Ethiopian university is encouraged to build its own repository and to provide the Ethiopian Research and Education Network (EthERNet) access to that university repository in order to harvest and amalgamate publications from these repositories to a compiled National Academic Digital Repository of Ethiopia
- Those universities that are not currently able to establish their own repository will get support from EthERNet to deploy their own Institutional Repository (IR). However, each university will be responsible for the curation of contents stored its own IR



NADRE objectives

1. To increase the access of academics and citizens in Ethiopia and worldwide to research works published by Ethiopian researchers at one of the Ethiopian universities or research institutions
2. To foster the dissemination of research outcomes and make them, as well as their authors, more visible inside and outside the country
3. To spread completed research to other corners of the country to allow other researchers to build on. For the time being, the NADRE will mainly include Master theses, dissertations, journals, articles, conference proceedings as well as all works published by researchers from Ethiopian universities and research institutions as well as from Ethiopian researcher living abroad. All artefacts will be published under open licenses (e.g., Creative Commons licenses) and will be tagged with Digital Object Identifiers (DOIs)

NADRE stakeholders



CEARL
Consortium of
Ethiopian Academic
and Research Libraries



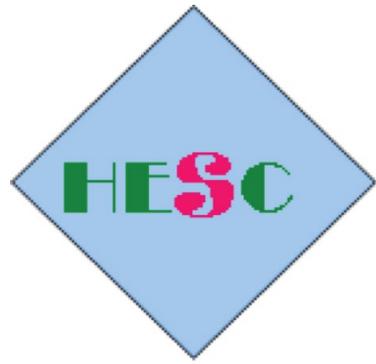
german
cooperation
DEUTSCHE ZUSAMMENARBEIT

Implemented by: **giz** Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH



የሳይንስና ከፍተኛ ትምህርት ሚኒስቴር
Ministry of Science and Higher Education

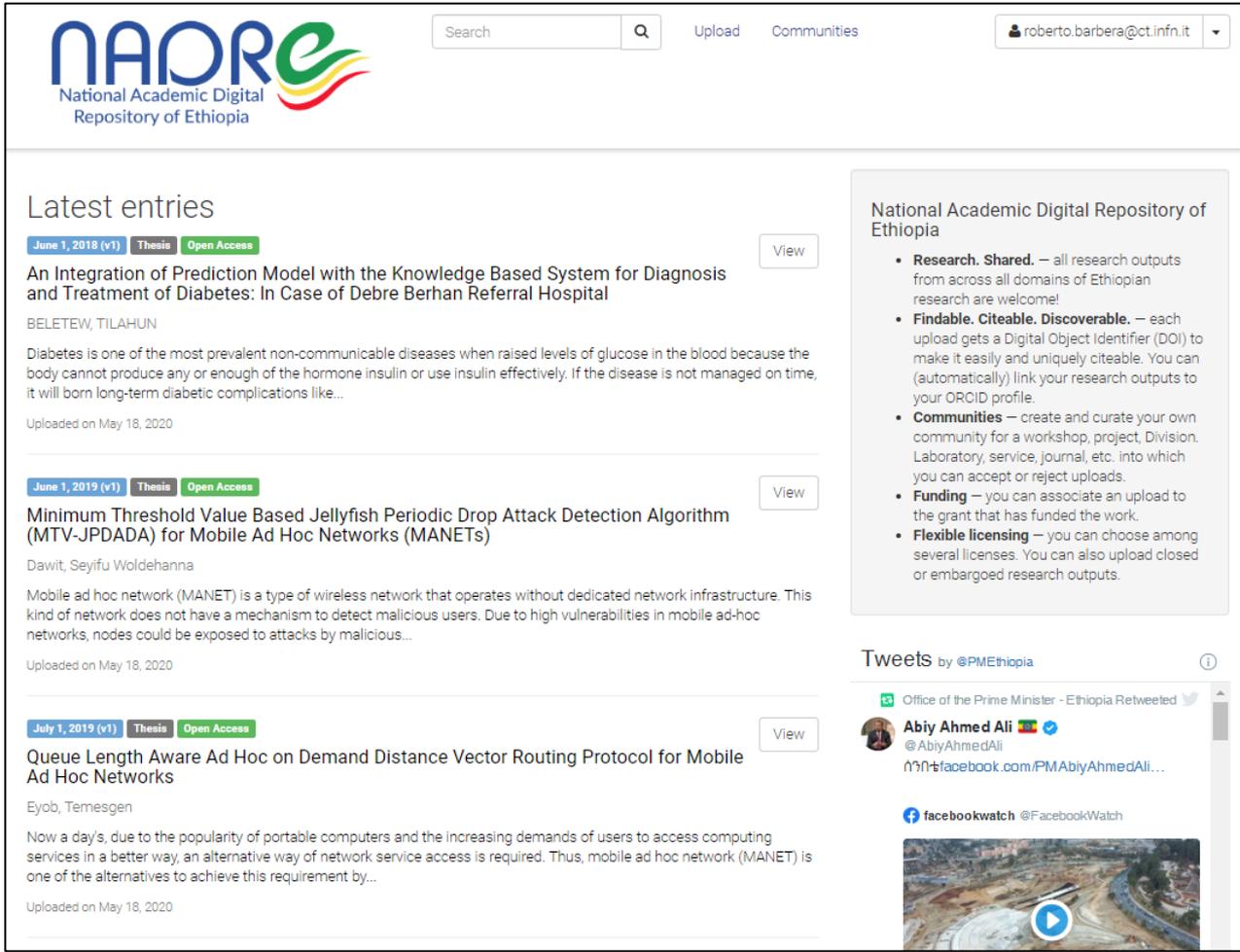
Current implementation Task Force



**Ethiopian
Universities**

September 2019 – NADRE is online

(<https://nadre.ethernet.edu.et>)



- Adopts leading-edge technologies in the domain of digital repositories
- Exploits the concept of “communities”, which is central in the Zenodo architecture, to cope with several aggregation of contents: about 80 communities from many Ethiopian universities have already been created

- Supports DOI versioning
 - MOSHE owns a DOI prefix

from 

- **Is compliant with FAIR principles**
- **Is compliant with Plan S requirements**

Based on the INFN Open Access Repository

The National Open Access Policy of Ethiopia for Higher Education (1/2)

(<https://doi.org/10.20372/nadre/4192>)

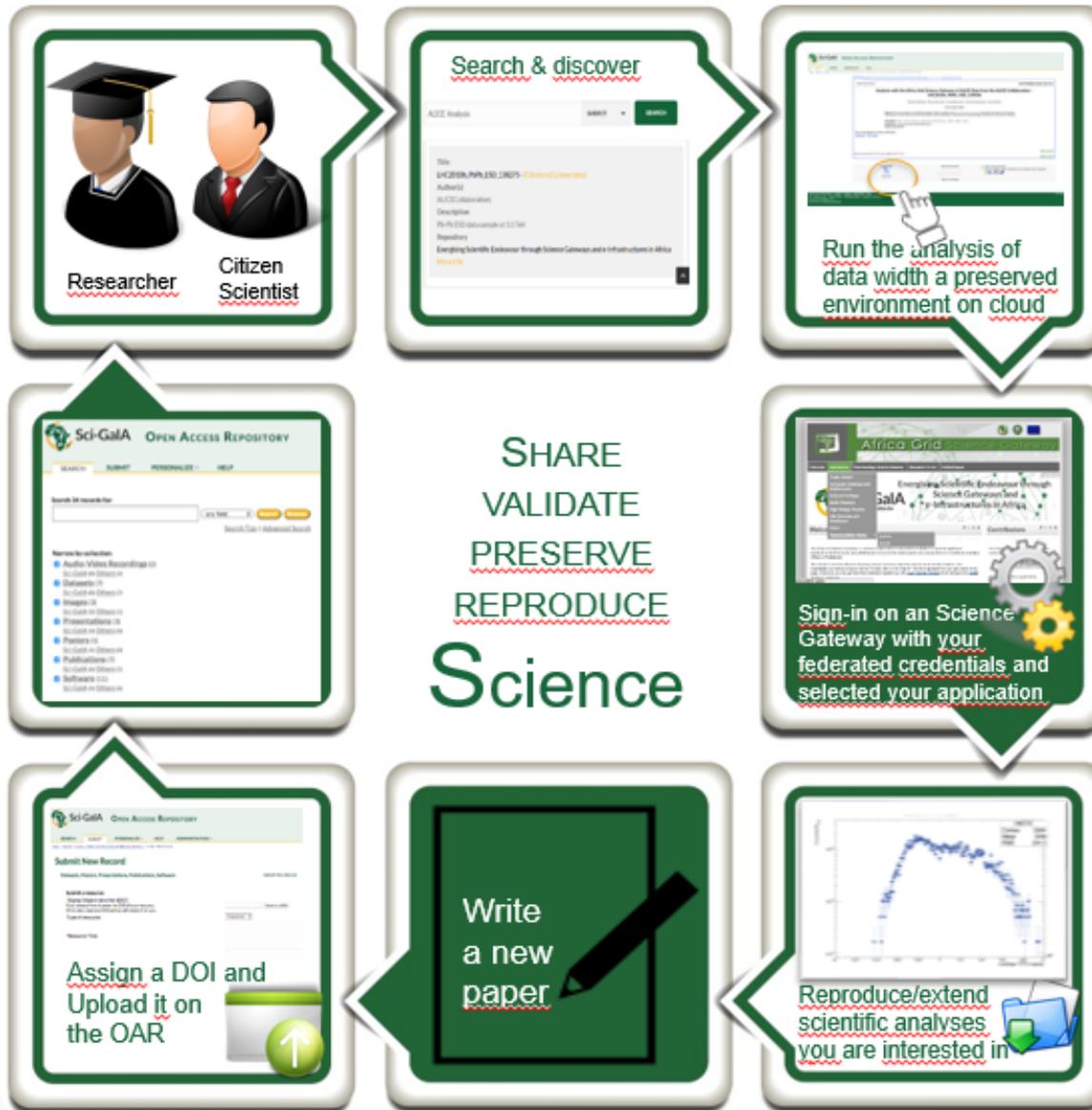
The screenshot shows the NAADRE website interface. At the top left is the NAADRE logo with the text "National Academic Digital Repository of Ethiopia". To the right of the logo is a search bar and navigation links for "Upload" and "Communities". A user profile dropdown menu shows "roberto.barbera@ct.infn.it". Below the header, the date "November 6, 2019" is circled in red. The main title of the document is "NATIONAL OPEN ACCESS POLICY OF ETHIOPIA FOR HIGHER EDUCATION". Below the title, it identifies the author as the "Minister of Science and Higher Education of Ethiopia" and provides a brief description: "The document contains the Ethiopia national Open Access policy for higher education." A preview window is open, showing the first page of the document. The document header includes the logo of the Ministry of Science and Higher Education (MSHE) and the text "የሥነ ምግባርና ከፍተኛ ትምህርት ሚኒስቴር" and "Ministry of Science and Higher Education - Ethiopia". The main heading of the document is "NATIONAL OPEN ACCESS POLICY OF ETHIOPIA FOR HIGHER EDUCATION", followed by the section "1. Aims and Scope of the Policy". The text in the preview states: "The Federal Democratic Republic of Ethiopia (FDRE) Ministry of Science and Higher Education is committed to supporting research of high quality and to ensuring that public research spending will lead to a maximum economic and social return. The FDRE Ministry of Science and Higher Education supports the principles of OPENNESS to research outputs and processes as integral to research excellence as well as the sharing and creation of new knowledge. The policy applies to all research outputs authored/created, or co-authored/co-created, by..." On the right side of the page, there are several interactive elements: an "Edit" button, a "New version" button, an "Admin" dropdown, a "Communities" section listing "National Academic Digital Repository of Ethiopia" with a "Remove" button, a statistics box showing "410 views" and "279 downloads" with a "See more details..." link, and a "Publication date" section showing "November 6, 2019". At the bottom right, there is a "DOI" section with the DOI "10.20372/nadre/4192" and a "Keyword(s)" section with tags for "Ethiopia", "Open Access", "Policy", and "Higher Education". A "Communities" section at the very bottom right lists "National Academic Digital Repository of Ethiopia".

The National Open Access Policy of Ethiopia for Higher Education (2/2)

- The policy applies to all research outputs authored/created, or co-authored/co-created, by employees of Ethiopian universities; this includes research outputs of graduate students and PhD students
- Researchers and research students have to make themselves visible and findable, using persistent identifiers such as ORCID IDs, and their outputs available on an Open Access basis, if their research resulting entirely or partly from public funding
- The FDRE Ministry of Science and Higher Education supports the National Academic Digital Repository of Ethiopia (NADRE) as the aggregated national repository for publications, underlying research data that proof the validity of the related publication and open educational resources. Institutional repositories for publication and research data should be open for harvesting through NADRE
- Universities will implement and support Institutional Repositories and respectively NADRE as the Open Access publishing platform for online journals which are created and managed by University departments or research centers
- Universities encourage the uptake of Open Science practices (in addition to open access to publications and data) such as the involvement in collaborative community science projects, the use of open educational resources etc. - by including "openness" as one of the criteria during research assessment and evaluation

What will be possible: The Knowledge Workflow

(to enable a Knowledge Society)



The ultimate goal: the Knowledge "Nexi" (to enable a Knowledge Society)



Summary and conclusions

- FAIR principles compliant digital repositories and Digital Object Identifiers are key Open Science enablers
- NADRE is a successful example of service that exploits leading-edge software technologies for standard-based repositories:
 - It triggered the establishment in Ethiopia of a national Open Access Policy for the first time ever
 - Its governance and leadership model is meant to involve all the actors and stakeholders of research and higher education in the country
- **The Somali Research and Education Repository (SORER) is based on the same software framework of NADRE and the INFN Open Access Repository and it can really help to improve the visibility of Somali researchers and their scientific results**
- **SORER can trigger the establishment of an Open Access/Open Science Policy in Somalia, which would be a tremendous result**
- **The know-how acquired by SomaliREN is very important not only within the country but also in the rest of the continent**
 - **A collaboration is about to start with Palestine in the context of the Arab States Research and Education Network (ASREN)**

Thank you !
Mahadsanidiin!

