

▶ OPEN SCIENCE ▶ FAIR ▶

SEPTEMBER 20-23, 2021

OPEN SCIENCE – TIME TO UNLOCK THE POTENTIAL OF THE DIGITAL AGE

DR ROBERT TERRY TDR, MANAGER RESEARCH POLICY

(THESE COMMENTS ARE MY OWN...)



CC BY NC IGO 3.0 NB SOME IMAGES ARE COPYRIGHTED



**World Health
Organization**



TDR For research on
diseases of poverty

UNICEF • UNDP • World Bank • WHO

#OSFAIR2021

@OpenScienceFAIR

Open Science – time to unlock the potential of the digital age

Overview

The key is the internet – digitization

Open science requires systems thinking

The value is making new knowledge from connections

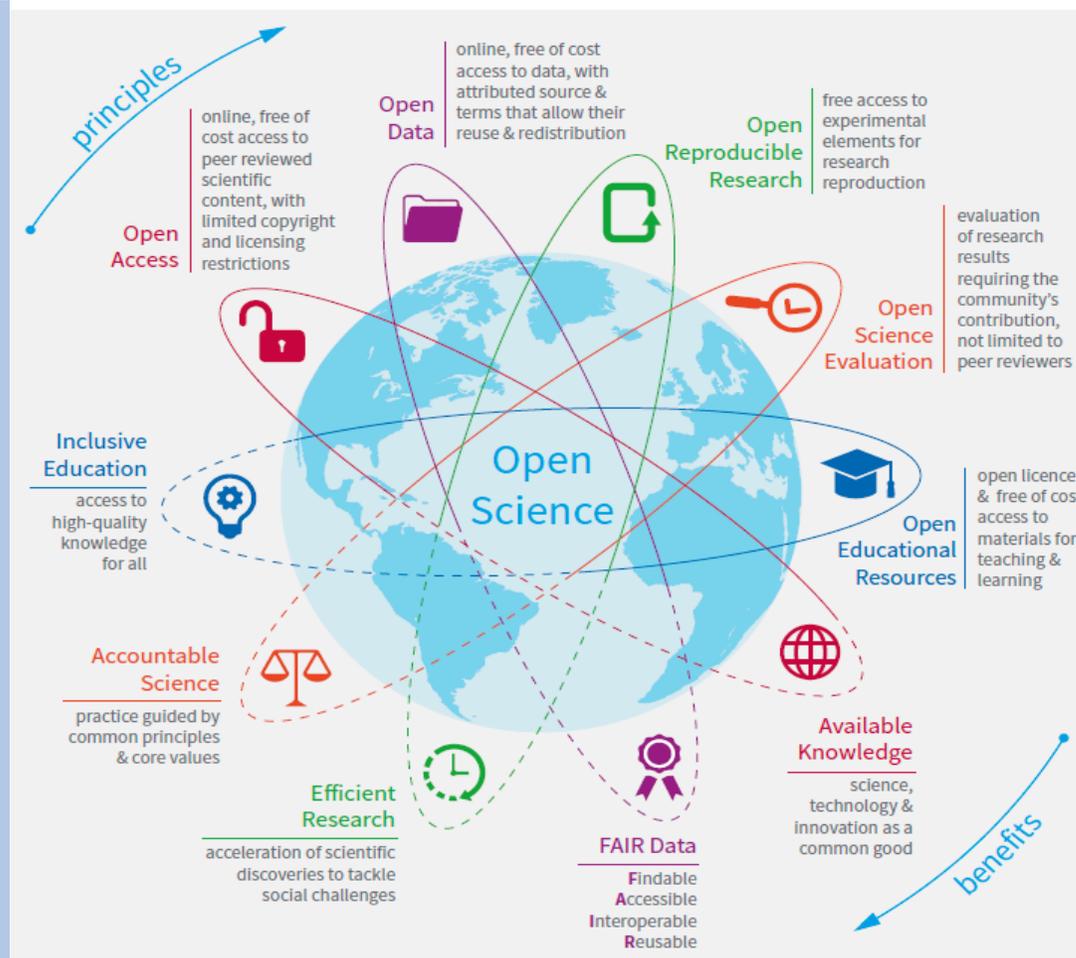
Standards to make (machine) interoperable

Minimize technical, legal, financial and linguistic barriers

Effective Ethical Equitable

Collaboration not Competition for Global Goods

A reset for scientific careers new incentive measures needed



[1] Towards a global consensus on open science: report on UNESCO's global online consultation on open science. UNESCO, 2020

[2] FOSTER portal <https://www.fosteropenscience.eu/>, accessed in March 2021

[3] Science ouverte à l'Université de Genève: feuille de route pour un partage de connaissances scientifiques 2020-2023



Open Science requires systems thinking done well it can democratize science

Open Access

1. Digital free-to-access online
2. Free to reuse (rights)
3. Archived digital standards

Open Data to Managed Access

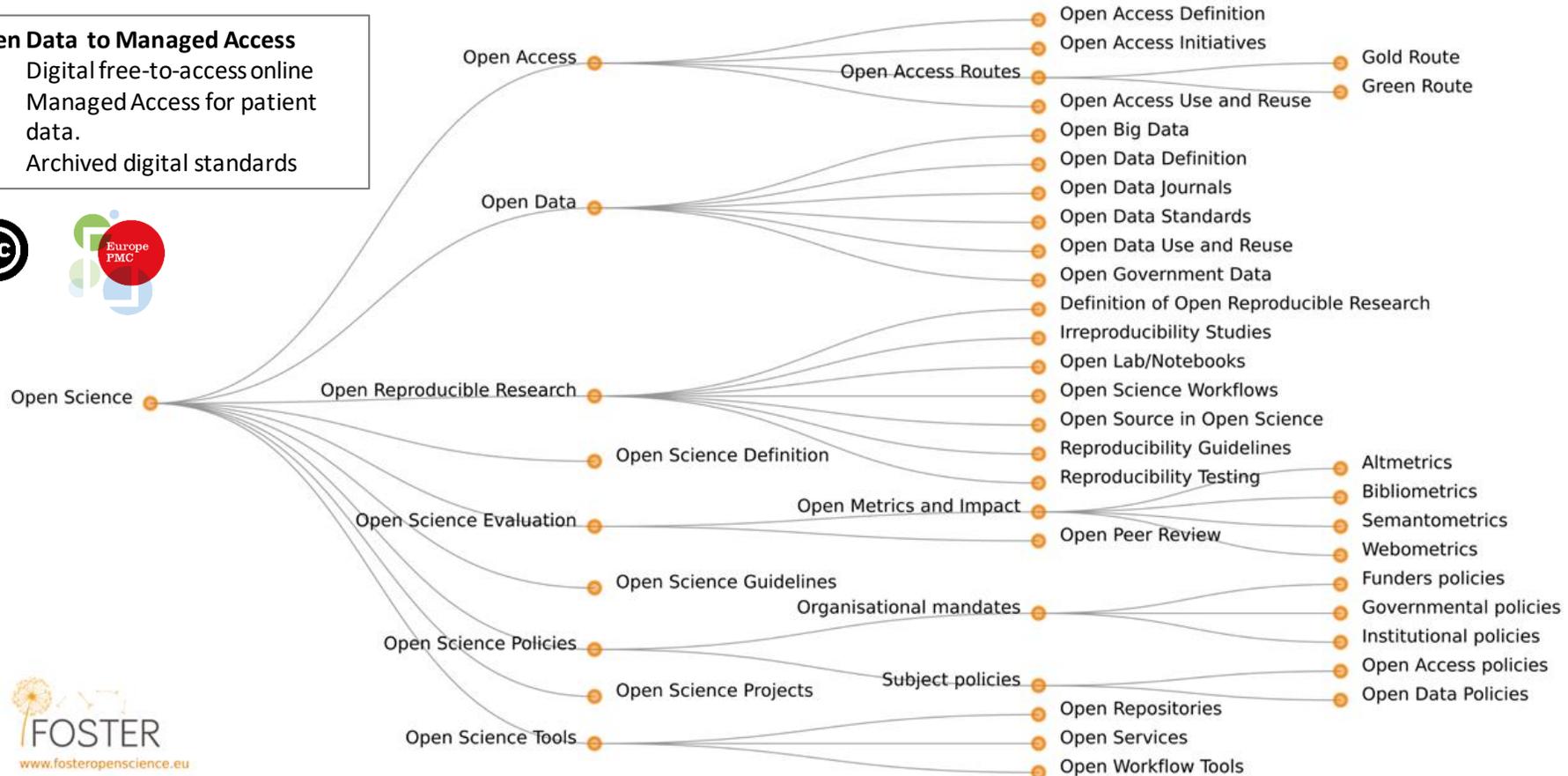
1. Digital free-to-access online
2. Managed Access for patient data.
3. Archived digital standards



Open Science



Open Science Taxonomy



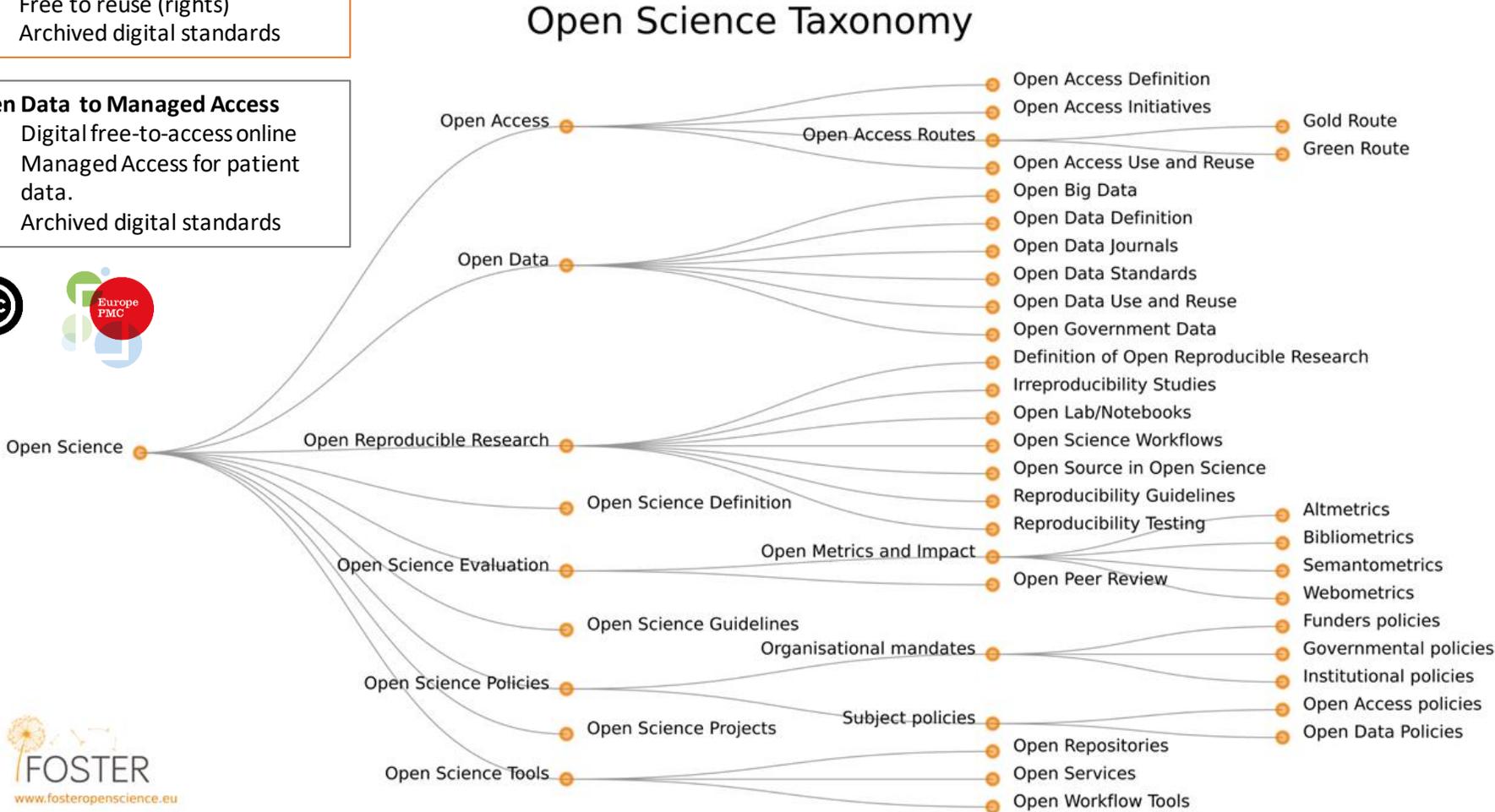
Open Science resources



Open Science requires systems thinking done well it can democratize science

- Open Access**
1. Digital free-to-access online
 2. Free to reuse (rights)
 3. Archived digital standards

- Open Data to Managed Access**
1. Digital free-to-access online
 2. Managed Access for patient data.
 3. Archived digital standards

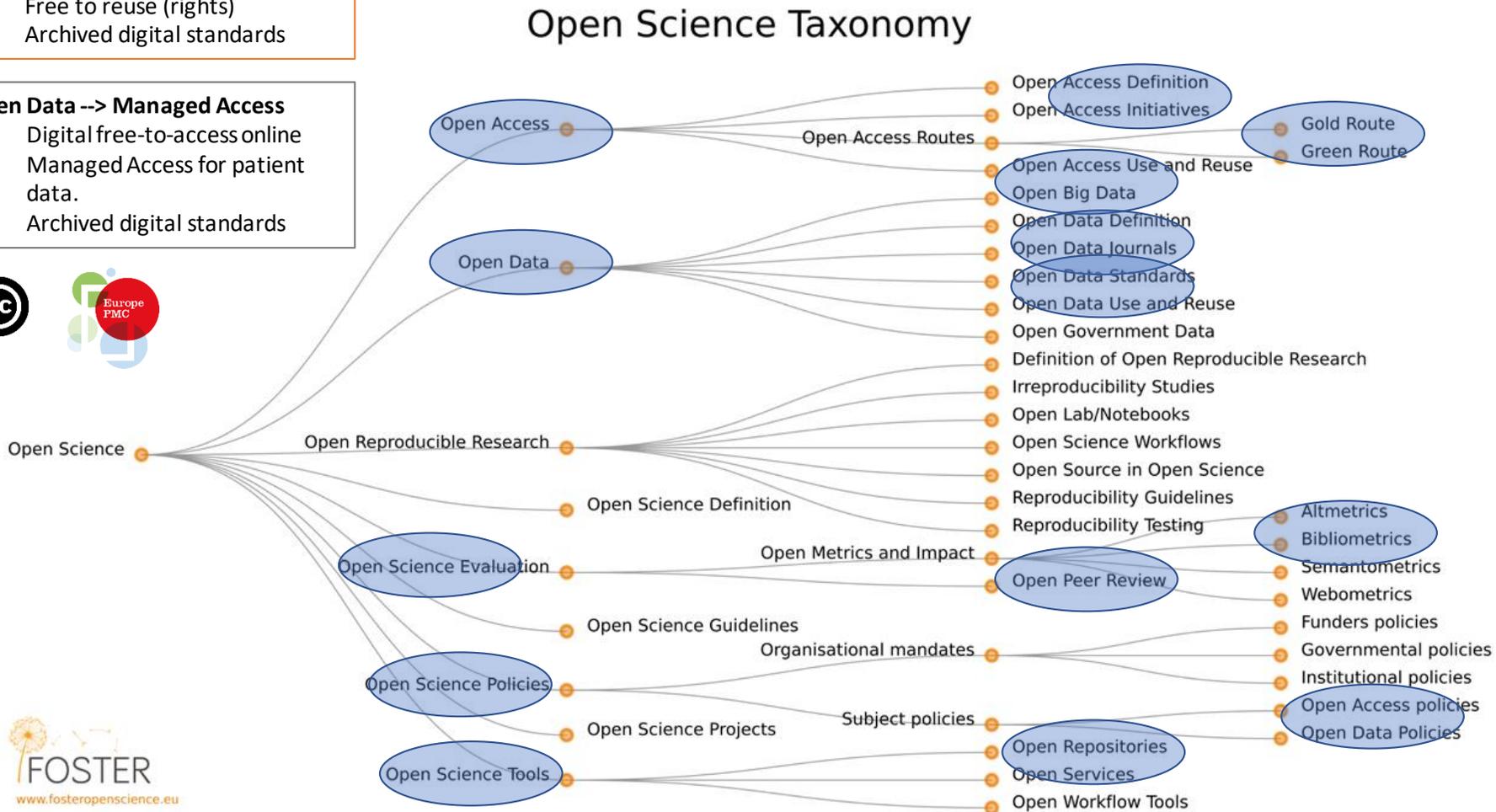


TDR & Open Science

Organizations and individuals should map themselves to this taxonomy

- Open Access**
1. Digital free-to-access online
 2. Free to reuse (rights)
 3. Archived digital standards

- Open Data --> Managed Access**
1. Digital free-to-access online
 2. Managed Access for patient data.
 3. Archived digital standards



Why is open science so important...?

+ve Open science enables:

- Creation of new knowledge
- Automated sharing and collaboration
- Interoperability within a discipline and across disciplines (FAIR)
- Transparency
- Quality assurance
- Provenance ->->-> Benefit sharing
- Better peer review
- Faster results

-ve Open science enables:

- Scooping no attribution
- Exploitation
- Anti-commercial
- Wrong messages
- Unsustainable
- Chaos
- Inequity
-

Further work:

- What are the benefits individuals, nations, global goods vs products
- Define fair benefit sharing community resources
- New business models
- New metrics to motivate and reward
- Reduce sovereignty increase globalization
- Build trust through strong governance
-

Human Genome Project - a new paradigm in sharing ?



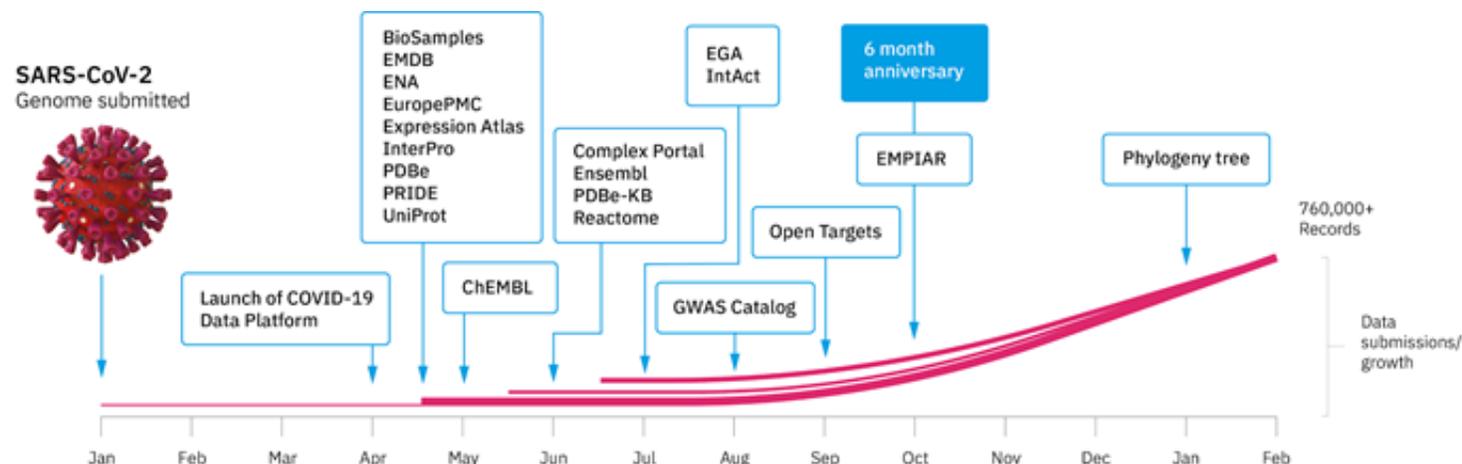
© BBC http://news.bbc.co.uk/2/hi/uk_news/politics/806819.stm

.....We applaud the decision by scientists working on the Human Genome Project to release raw fundamental information about the human DNA sequence and its variants rapidly into the public domain, and we commend other scientists around the world to adopt this policy.

Joint statement President Clinton and Prime Minister Blair March 2000
https://clintonwhitehouse4.archives.gov/WH/New/html/20000315_2.html

Use case: COVID-19 Data Portal

- Over 2,500,000 records across molecular platforms and literature
- Access to data resources and tools
- 78 linked “related” resources
- Web, API and download



Viral Sequences

All (2,055,747)
 Sequences (635,498)
 Reference sequences (2)
 Raw reads (703,077)
 Sequenced samples (704,064)
 Studies (392)
 Genes (22)
 Browser (1)
 Variants (12,691)

Host Sequences

All (16,981)
 Human studies (controlled access) (14)
 Human reads (consented for full access) (11,372)
 Other species reads (5,555)
 Association studies (40)

Expression

All (97)
 Gene expression (4)
 Gene expression experiments (24)
 Single cell expression (4)
 Single cell expression experiments (22)
 Protein expression experiments (43)

Proteins

All (1,800)
 Protein sequences (106)
 Protein families (336)
 Protein structures - Knowledge Base (12)
 Protein structures (773)
 Electron microscopy density maps (555)
 Electron microscopy public image archive (18)

Biochemistry

All (4,717)
 Pathways (16)
 Interactions (3,297)
 Complexes (31)
 Compound document (10)
 Drug targets (1,361)
 Metabolomics experiments (2)

Imaging

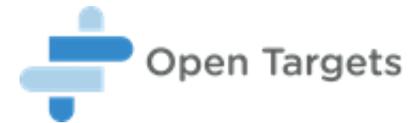
All (20)
 Images (2)
 Electron microscopy public image archive (18)

Literature

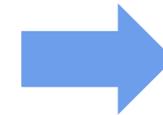
All (421,836)
 Coronaviruses (213,284)
 Diseases (197,931)
 Related viruses and diseases (2,550)
 Genes, receptors and antibodies (8,066)
 Supplementary material (5)



Use case: therapeutics discovery



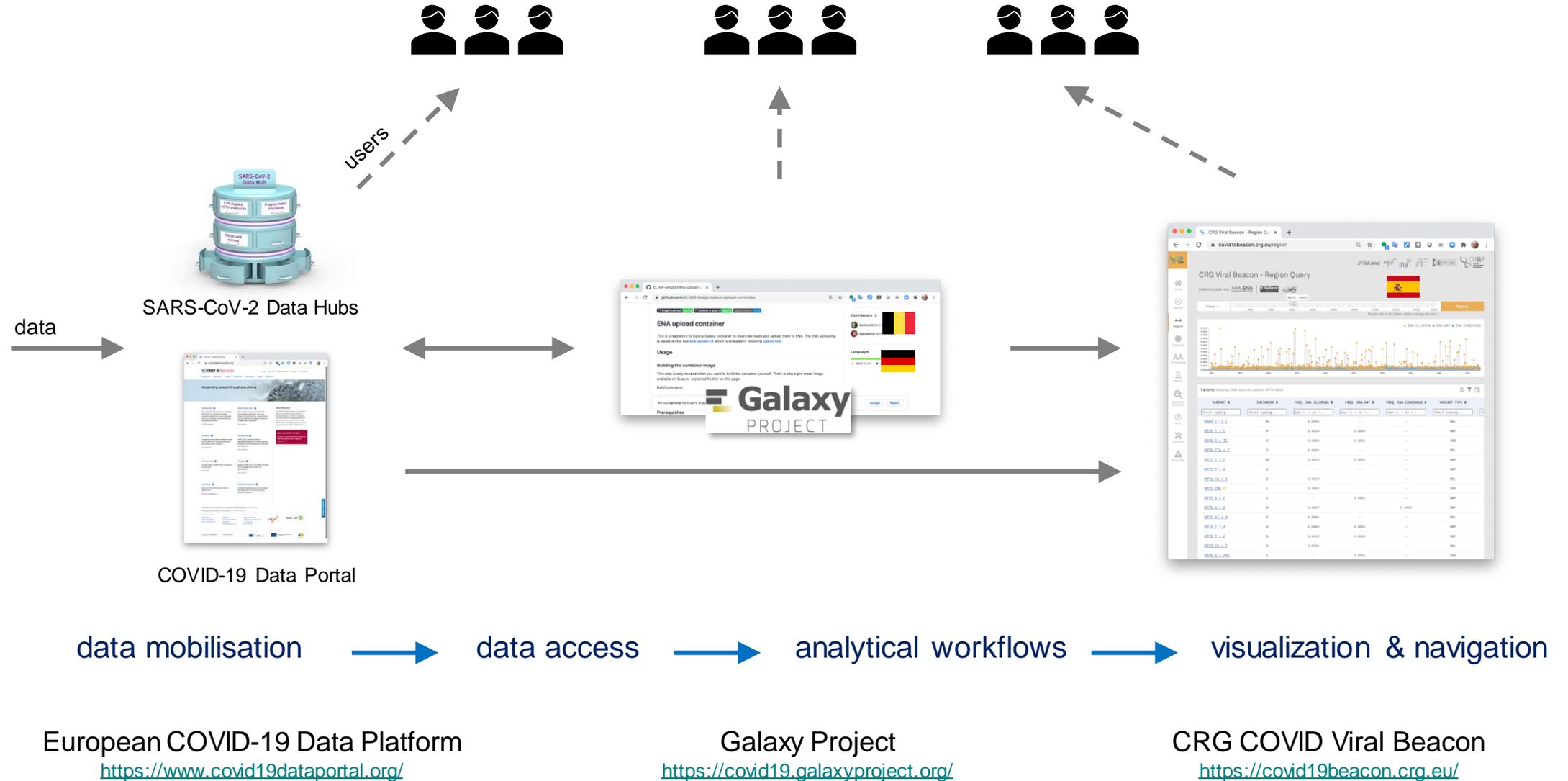
Vaccines



Treatment

<https://covid19.opentargets.org/>

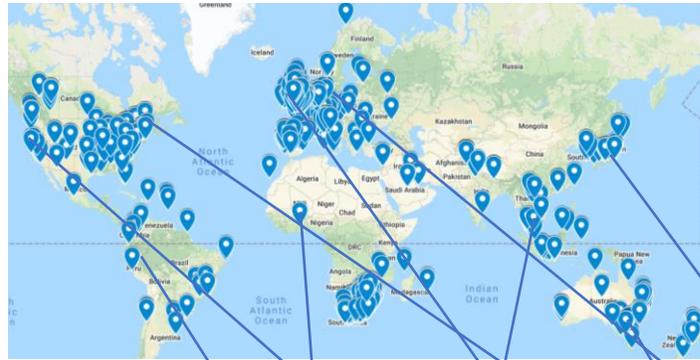
Use case: viral variation, tracking and planning vaccination



Use case: by collating global clinical data ISARIC rapidly showed that young children and older adults presented with unique symptoms, not captured by current COVID-19 case definitions.



Data from cases around the world shared with ISARIC

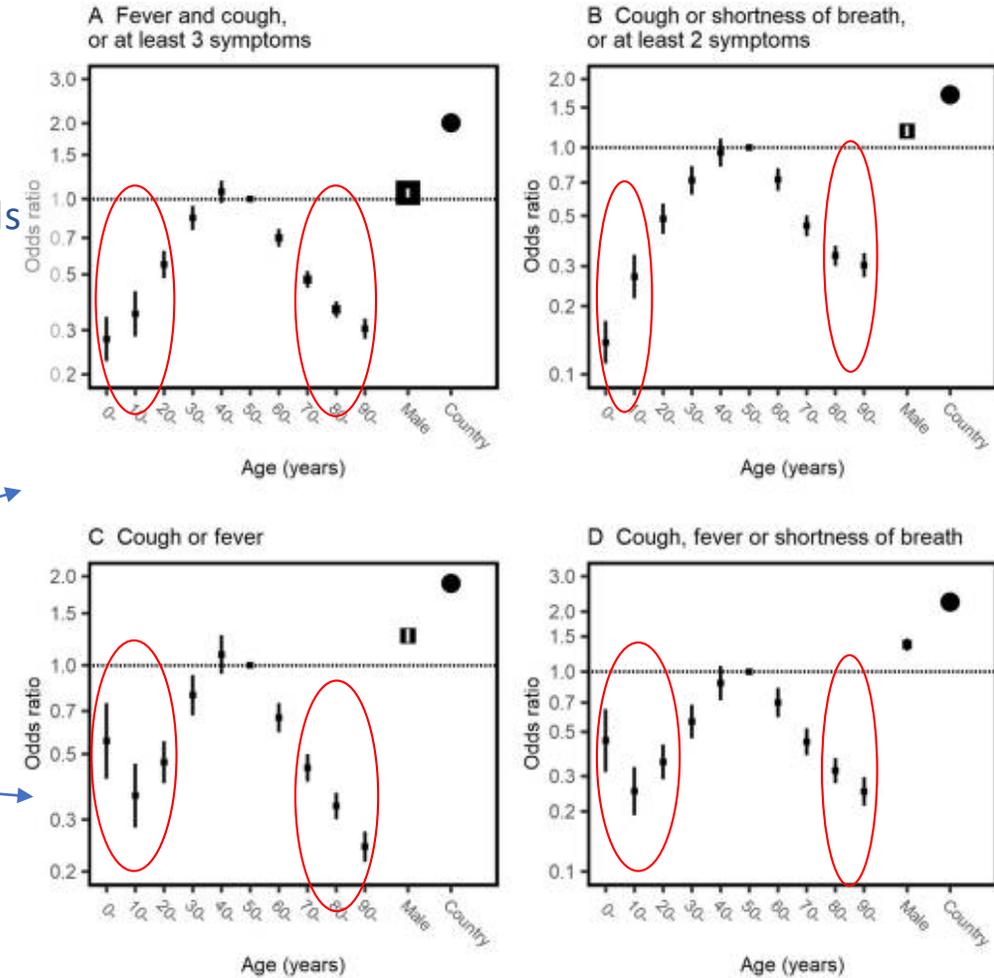


Excluded
 SARS-CoV-2 not confirmed, n = 24 336
 Missing all symptoms, n = 1930
 Missing age, n = 1220
 Missing sex, n = 131
 Missing country, n = 9
 Admitted before onset of Covid-19, n = 5794
 Asymptomatic, n = 6094

All patients in dataset
 n = 99 623

Included in analysis
 n = 60 109
 (ISARIC Core CRF, n = 6483
 ISARIC Rapid CRF, n = 446
 ISARIC 4C, n = 47 275
 COVID-19 CCC, n = 1041
 Non-REDCap, n = 4864)

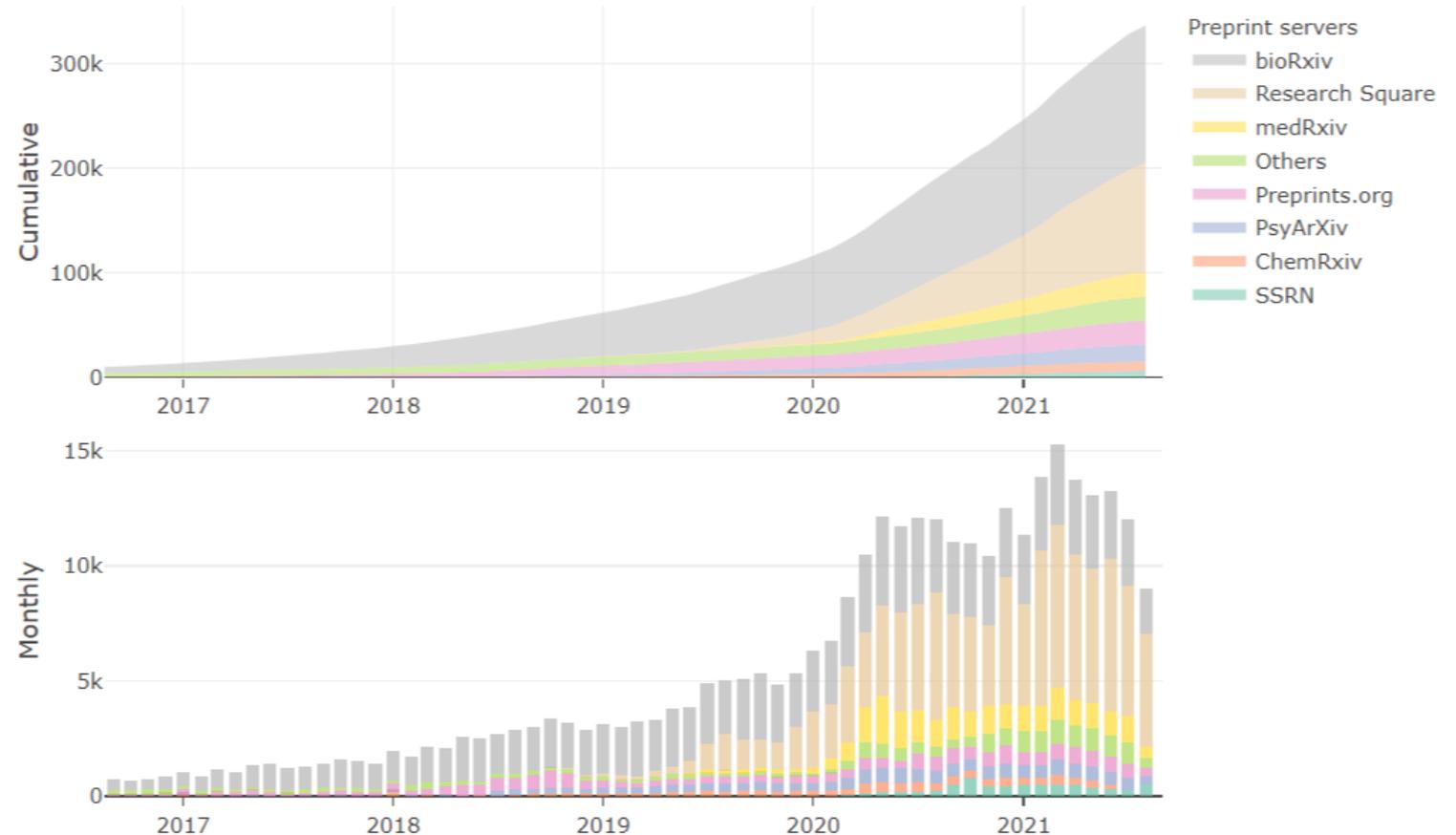
Rapid analysis showed new trends in symptoms at presentation – alerting health care workers to identify cases outside of the published case definitions



Open science and publication – time to reorder the process

- (Nearly) all Covid-19 publications free-to-read in the pandemic – admission that subscription impedes progress.
- In 2017 only a handful of life science pre-prints less than 25,000
- In 2021 EPMC indexes ++ 300,000 mirrored by page views
- How are these used?

Preprints in Europe PMC



Open science and publication: standards = usability

“As the Chief Scientist of WHO I welcome the huge increase in the use of pre-prints by researchers to rapidly share the emerging evidence from the many studies on Covid-19. However, these are published as .pdf documents and I recognise that the information they contain could be more rapidly searched and linkages made between the results and data they contain if they were converted to the standard publishing language XML. I therefore support this initiative by Europe PMC to take on this task.”

—Dr. Soumya Swaminathan
Chief Scientist, WHO
4 May 2020



<https://europepmc.org/Preprints>

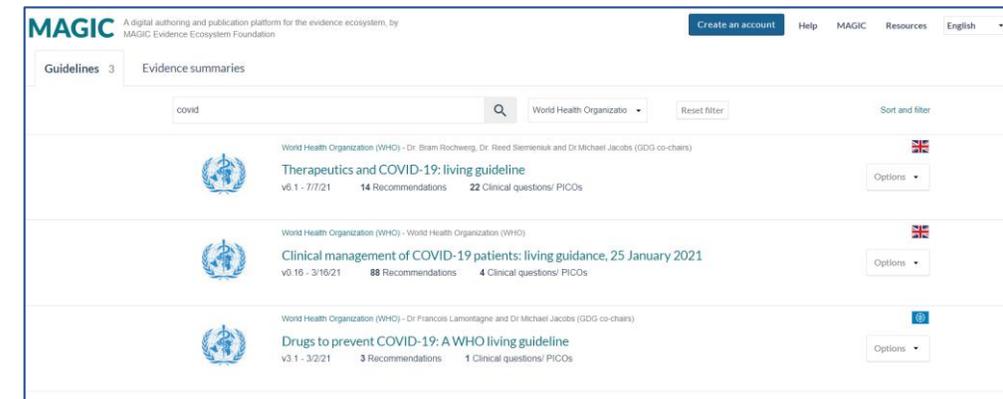
Use of pre-prints – calling time on subscriptions

- WHO repository IRIS 150 publications relating to Covid-19 - 25% referencing pre-prints
- NEW development WHO [Living Guidelines](#) available online via the MAGICapp
- 3 WHO Living guidelines for Covid-19. Therapeutics 6 versions since November 2020.



Analysis of version 5 March 2021

- 44% of its references as pre-print
- 33% unpublished results shared with WHO
- Therefore < 25% from traditional published literature.....



<https://app.magicapp.org/#/guidelines>

Use of pre-prints – calling time on subscriptions

Dr Maria van Kerkhove, WHO Covid-19 Technical Lead at
HDR UK conference June 2021

'...publication in a high impact journal does not equal quality..... it is important we need to receive data from chemistry, engineering, architecture not just medicine...'

Implications of pandemic for publications

- Need to rethink publishing
 - 1st Publish
 - 2nd Open (meta) peer review
 - 3rd Earn impact
- Why have impact factors?! - Like awarding the medals BEFORE the race has run
- Traditional publishing model is no longer fit for purpose too slow and no guarantee of quality
- It feels like we're running electric cars on steam train tracks



Impact Factor is a
toxic indicator



Implications of pandemic for publications

- No such thing as the Version of Record – science is dynamic, changing and evolving
- The concept of the ‘Journal’ is dead = wasteful and biased
- Role for post-published aggregations perhaps Papers of the month
- Open science must create the interoperable links across all stages and disciplines. Links between the paper and the data are indivisible

All public science should be open access

Citizens should demand this

Pre-prints encouraged recognized and rewarded



João Batista Neto, CC BY 3.0

<https://creativecommons.org/licenses/by/3.0/>, via Wikimedia Commons

Open Science – time to unlock the potential of the digital age

Recap

Open science requires systems thinking

We all need to map ourselves to the open science ecosystem

The value is making new knowledge from connections

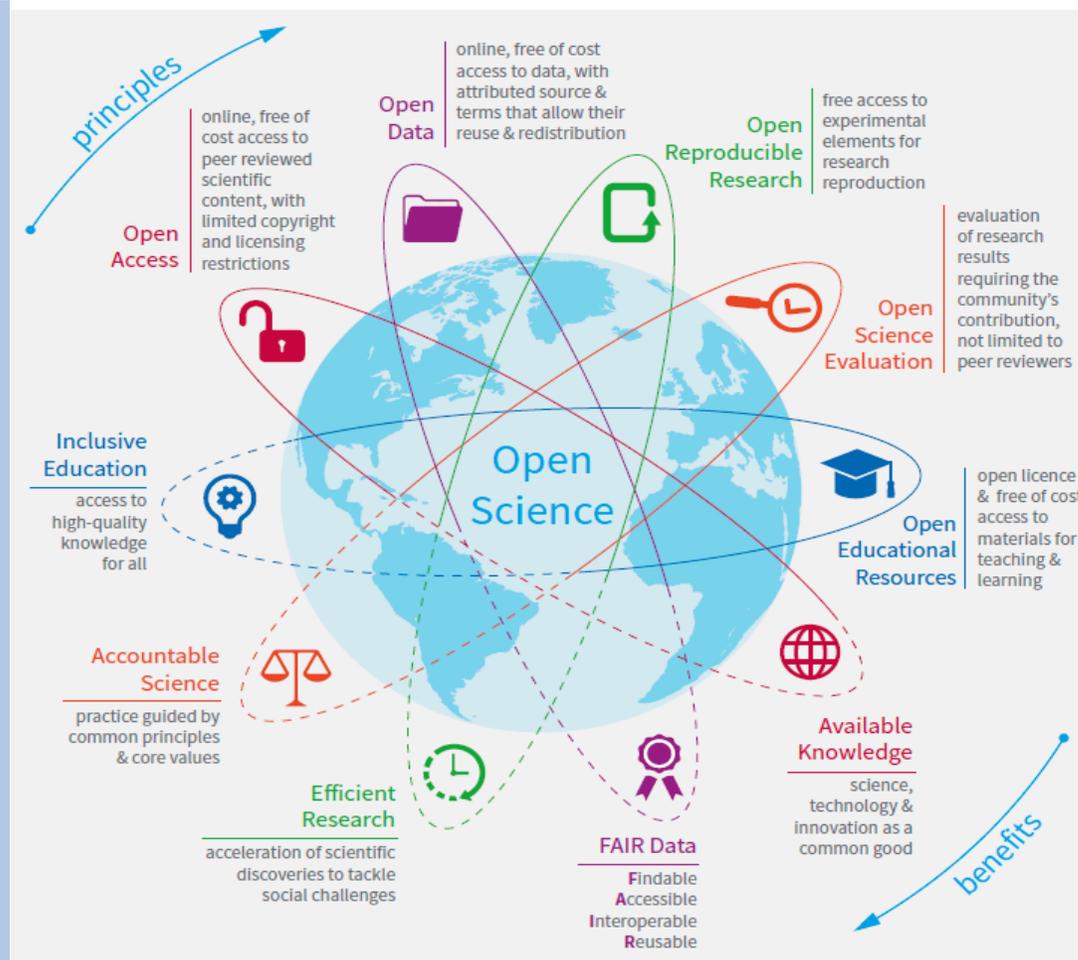
Standards to make (machine) interoperable

Minimize technical, legal, financial and linguistic barriers

Effective Ethical Equitable

Collaboration not Competition for Global Goods

A reset for scientific careers new incentive measures needed



[1] Towards a global consensus on open science: report on UNESCO's global online consultation on open science. UNESCO, 2020

[2] FOSTER portal <https://www.fosteropenscience.eu/>, accessed in March 2021

[3] Science ouverte à l'Université de Genève: feuille de route pour un partage de connaissances scientifiques 2020-2023



Acknowledgements

Contributions from:

Tomas Allen, WHO Library

Laura Merson, Associate Director IDDO

Guy Cochrane, Head of European Nucleotide Archive EBI

Robert Terry

TDR, Manager Research policy

@Terryr364

(these comments are my own....)

CC BY NC IGO 3.0

*'Research is born free
and everywhere is in
chains'*

(apologies to Rousseau)