

Better Research Through Better
Data *Live with Konstantinos
Repanas, European Commission*
#betterdata



Driving the paradigm shift towards Open Science: FAIR data

Better Research Through Better Data

5 October 2020

Kostas Repanas

European Commission

Directorate-General for Research and Innovation

Open Science unit

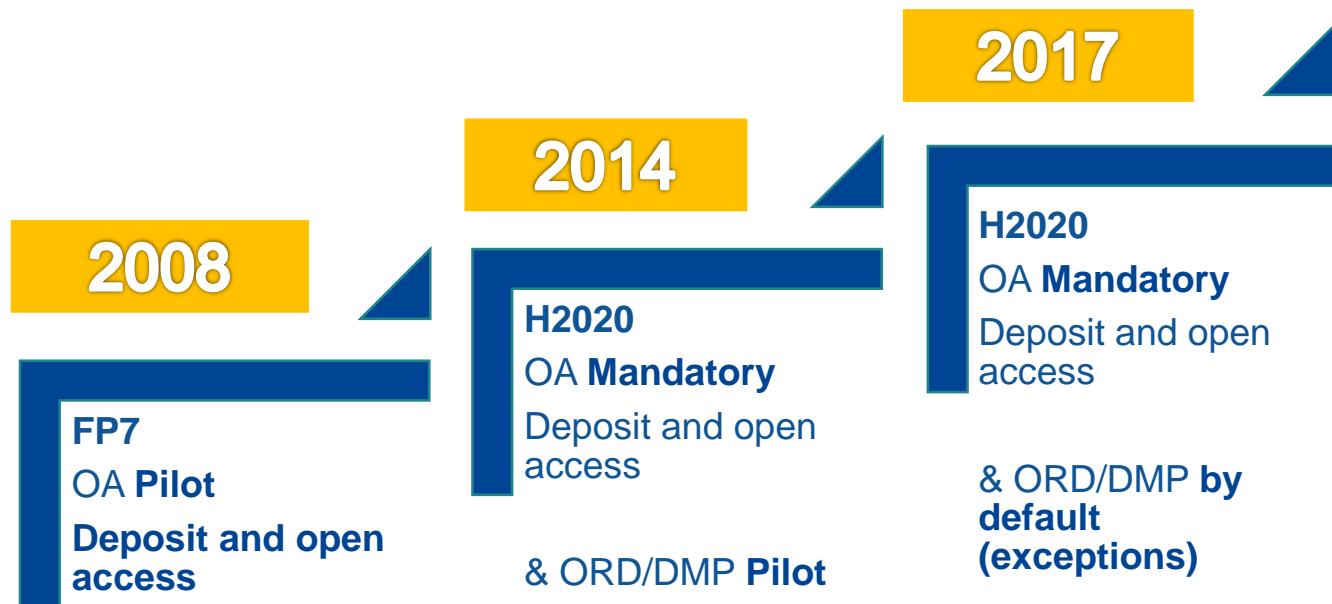
Open Science

- **Open Science** means **sharing knowledge and tools as early as possible**, not only **between researchers** and **between disciplines**, but also with **society at large**.
- Open Science improves the **quality**, **efficiency** and **creativity** of research and the **trust by society in science**. In particular, OS is beneficial for science, scientists and funders, e.g.:
 - tackles the reproducibility crisis;
 - faster response to societal challenges e.g. Coronavirus, Ebola;
 - access to and sharing results yields higher impact through collaborations;
 - generates new research findings;
 - large opportunity costs of non-FAIR data—**€10.2bn/year** (source: Cost-benefit analysis of FAIR research data, 2017).
- The Commission acts as **policy maker** (propose legislation and encourage MS), a **funder** (we set requirements to our projects) and a **capacity builder** (we fund ‘enabling’ projects).

Question - Are you aware of the FAIR data principles?

- Raise your hand for Yes
- Lower your hand for No

Evolution of our policies across the FPs



Planned under **Horizon Europe (2021)**:

- Open Science (OA, **RDM**, Citizen Engagement, etc.) embedded throughout the FP.
- OS to play a role in the:
 - **Evaluation** of proposals (methodology)
 - **Grant Agreement**
 - **Reporting**—during the project's lifetime
- Strengthening of the obligations with respect to open access and focus **on responsible RDM in line with FAIR**

Leading by example in Horizon Europe

- Reinforce the mandate for open access to publications and data (also through the Open Research Europe publishing platform)
- Mandate research data management in line with FAIR data principles, and promote long-term preservation (Data/Output management plans)
- Develop the underpinning European Research Infrastructures and EOSC

Horizon Europe (2021-2027)

- Encourage transdisciplinary and inter-sectoral approaches that include civil society organisations, citizens and end-users
- Mainstream citizen science and user-led innovation across the programme

Research data under Horizon Europe

- The governing principle will be to manage research data **responsibly, in line with FAIR**, and for data to be “**as open as possible and as closed as necessary**”. To accomplish this, we are **likely** to see the following:
 - At proposal stage, beneficiaries will be **evaluated** on **preliminary RDM considerations**
 - All projects that **generate/collect/re-use** research data will have to establish and regularly update a **Data Management Plan** (living document and machine-actionable)
 - Beneficiaries will have to **deposit data in a repository** and link their data to publications they underpin, if applicable
 - Valid repositories will be those that provide **PIDs** for the data and ensure **rich metadata** in line with FAIR.
 - For some actions, an additional obligation to deposit in a repository that is federated under EOSC.
 - Beneficiaries to ensure **open access ASAP** under CC BY or CC0 licences (or equivalent), unless exceptions apply (duly justified in the DMP)

Other research outputs under Horizon Europe

- Data is not the only research output that should be managed → **digital**: Software, algorithms, protocols, workflows, models, and **physical**: reagents, antibodies, hardware, etc. all need to be properly managed:
 - We would like to see other research outputs also **described in the DMP**.
 - There will be a **strong encouragement** for beneficiaries to **deposit and provide open access via a repository** to these outputs, unless legitimate interests or constraints apply.
- Research outputs, either digital or physical, should also be findable, interoperable, accessible and re-usable → We need a better understanding of what **FAIR** means for digital objects, other than data (e.g. software)

Poll - In my opinion, the biggest obstacle to data sharing is....

- No credit/reward structure
- Privacy/confidentiality issues
- Effort required for data preparation/curation
- Lack of resources
- Lack of guidelines



European COVID-19 Data Platform

Launched by President Ursula von der Leyen on 20 April 2020

- “The platform is an **important part** in the **building of the European Open Science Cloud**”

The platform is a thematic priority pilot to realise the EOSC vision and to showcase the added value of FAIR data sharing to advance science and benefit researchers



Press release: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_680

Video: <https://audiovisual.ec.europa.eu/en/topnews/M-004711>

European COVID-19 Data Platform: <https://www.covid19dataportal.org/>

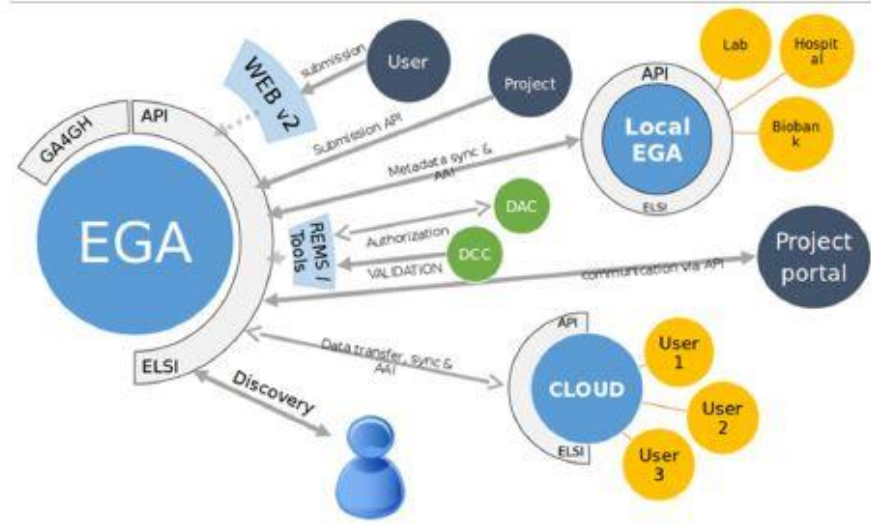
European COVID-19 Data Platform

diverse data – rapid sharing – degrees of completeness – synergy and cross-fertilisation

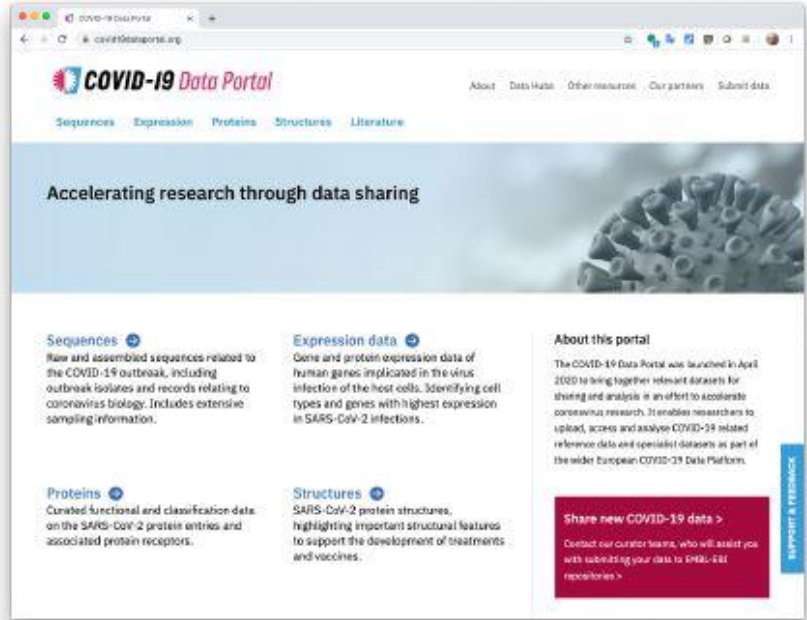
SARS-CoV-2 Data Hubs



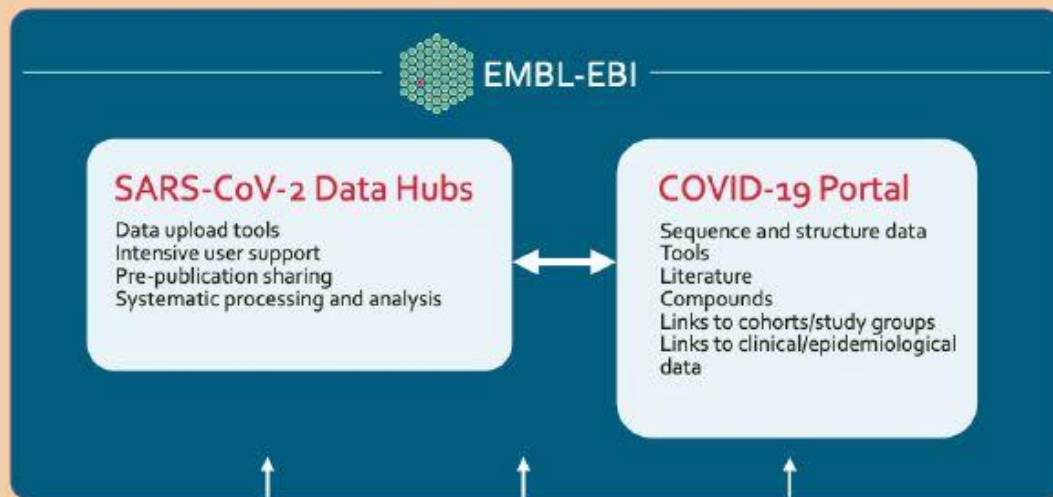
Federated European Genome-phenome Archive



COVID-19 Data Portal



European COVID-19 Data Platform



Clinical/epidemiological data where infrastructures do not exist

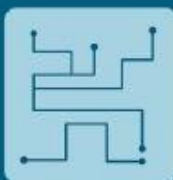
Sequence Data

Data links

National Infrastructures



Data Resources



Software & Workflows



Cloud & Compute Capabilities

erinha

European Research Infrastructure on Highly Pathogenic Agents

instruct
ERIC

eatris

EURO-BIOIMAGING
Image life, discover the future

EMBRC
EUROPEAN PAHORE BIOLOGICAL RESOURCE CENTRE

EU-openscreen
Chemical keys for life's locks

ECRIN
EUROPEAN CLINICAL RESEARCH INFRASTRUCTURE NETWORK

INFRAFONTIER
Infrastructures for the future

MIRRI
Molecular Imaging Resource for Research and Innovation

ISBE

EMPHASIS

EMBL-EBI



EUROPEAN OPEN SCIENCE CLOUD



Re CoD ID



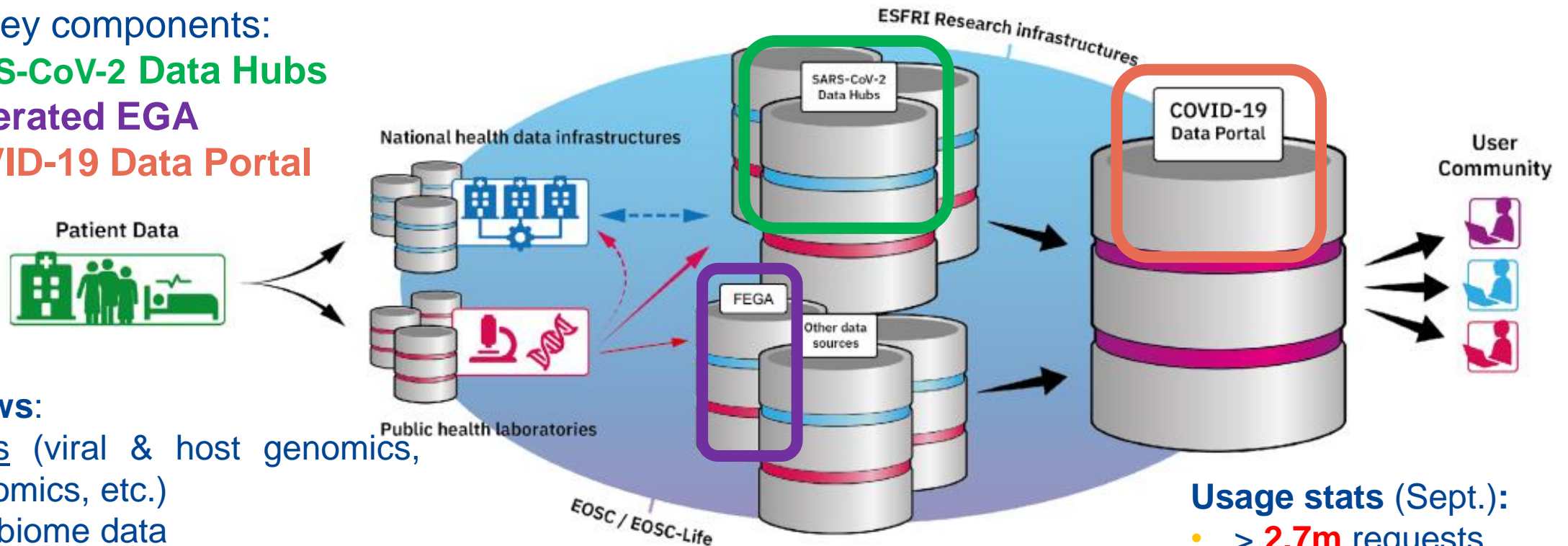
FAIR and open data sharing against COVID-19

- This effort represents a practical use case of **FAIR data in action**.
- The European Commission launched on 20 April the **European COVID-19 Data Platform** together with EMBL-EBI, ELIXIR, and other partners, and with the support of Member States.
- The Platform is a thematic priority pilot to realise the **EOSC vision** and to showcase the added value of **FAIR data sharing** to advance science and benefit researchers
- It responds to the need to capitalise on the quick and wide **sharing, re-use, processing** of and **access** to **research data** and **metadata** on the SARS-CoV-2, and the related COVID-19 disease.
- A very strong focus is placed on ensuring that data and metadata on this Platform are as **open** and as **FAIR** as possible (not always possible, because of the types of data, or because of the limited curation at the submission end). To maximise the FAIRness of the data, e.g. ELIXIR-NORWAY is performing manual curation on the viral sequences received.
- Crucial lesson: Researchers need **quick** and **unrestricted** access to multiple data sources to accelerate their research. **FAIR data** is an essential component in this.

European COVID-19 Data Platform

Three key components:

- SARS-CoV-2 Data Hubs
- Federated EGA
- COVID-19 Data Portal



Data flows:

- Omics (viral & host genomics, proteomics, etc.)
- Microbiome data
- Serology data
- Epidemiological data + SSH data
- Clinical research & patient data
- Observational studies data

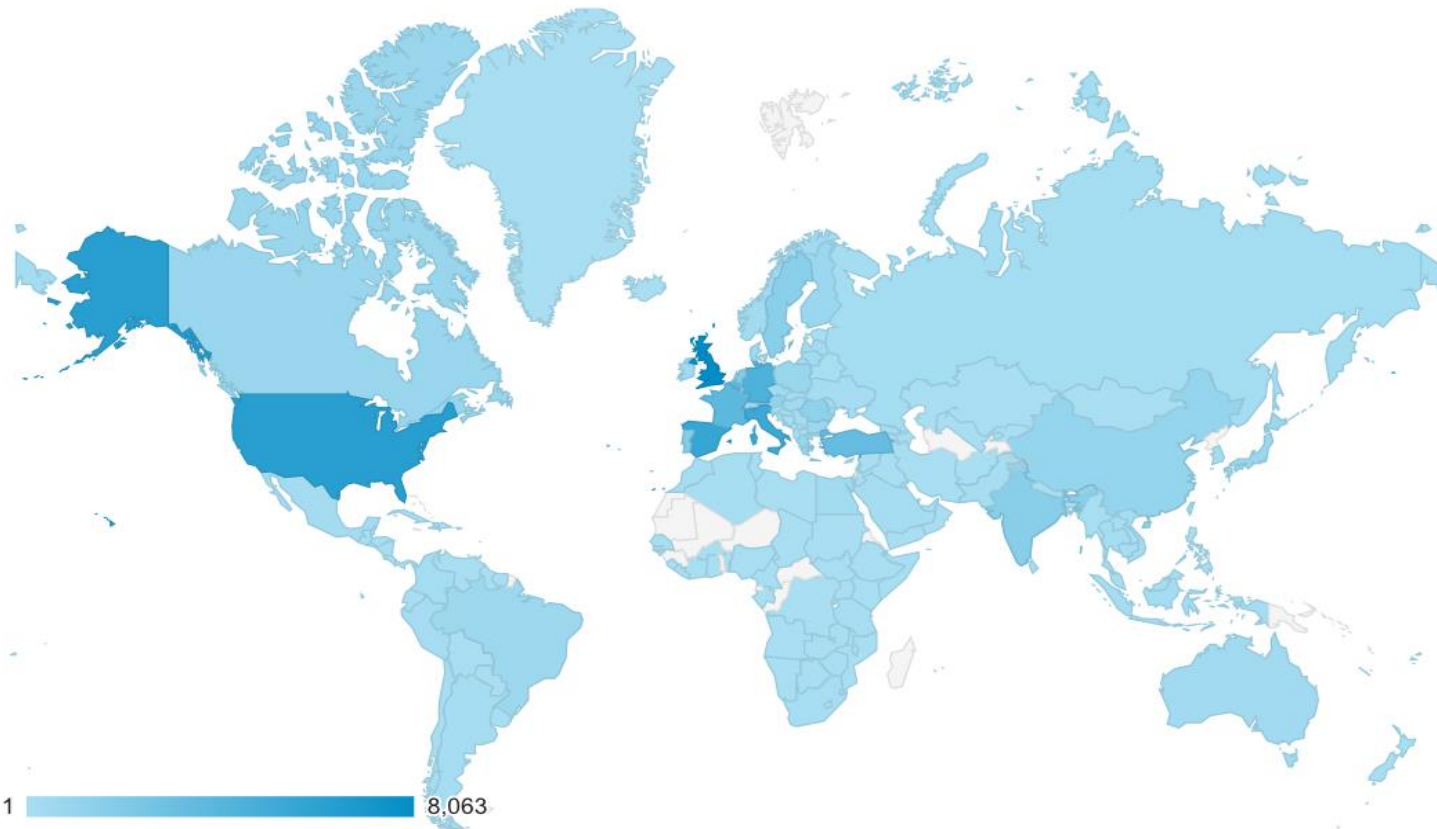
Portal stats (Sept.):

- ~63K raw viral sequence data sets
- ~23K viral sequences
- ~360 protein structures
- ~90K scientific literature publications...

Usage stats (Sept.):

- > 2.7m requests
- > 78K unique users
- > 170 countries

Top 10 countries accessing the portal

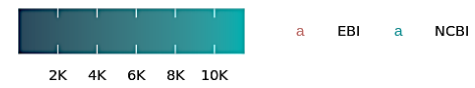
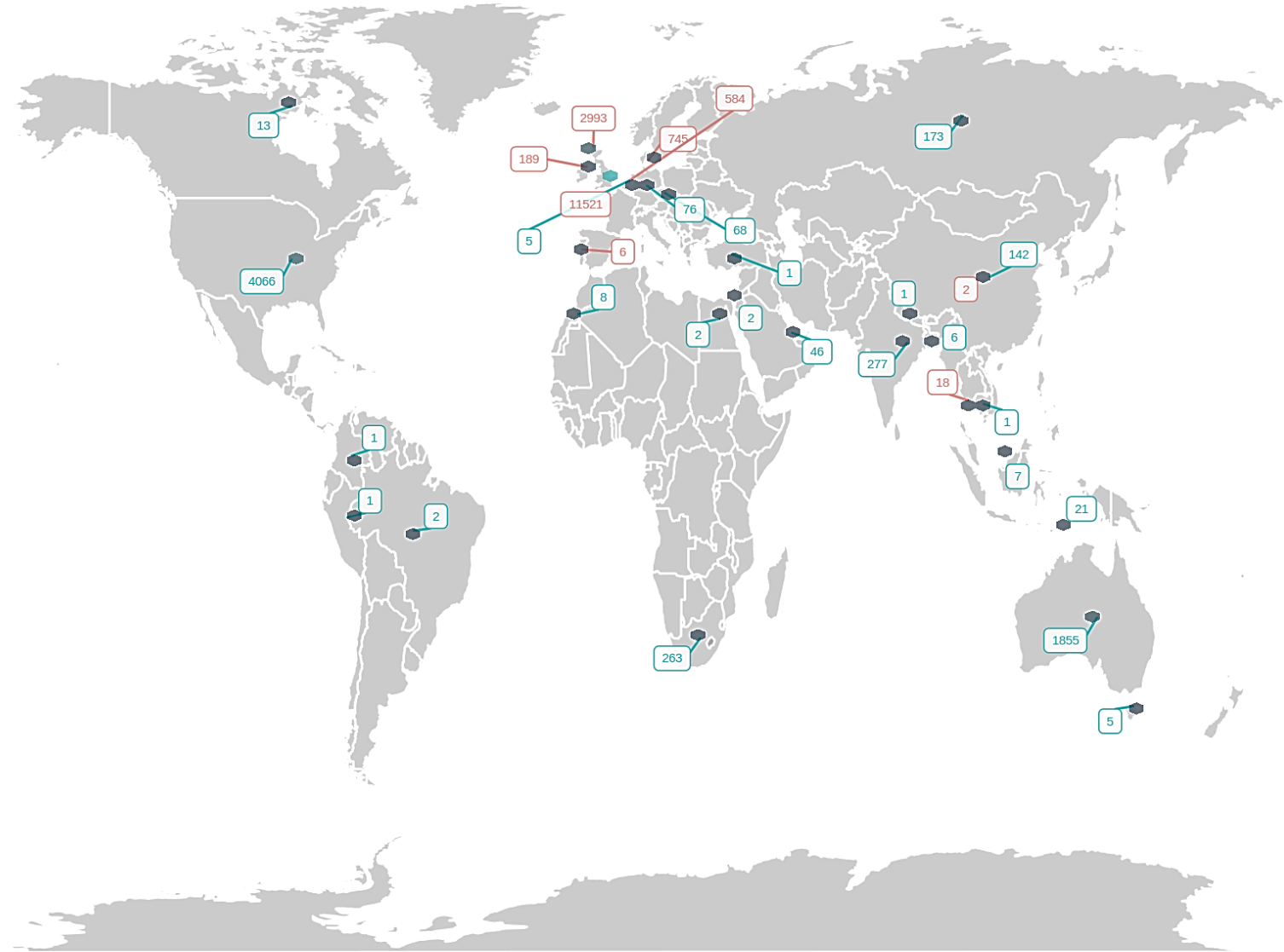


- United States 6,396 (10.30%)
- United Kingdom 5,325 (8.57%)
- Spain 4,887 (7.87%)
- Italy 4,887 (7.87%)
- Germany 4,171 (6.72%)
- Turkey 3,436 (5.53%)
- France 3,408 (5.49%)
- Belgium 3,097 (4.99%)
- India 2,201 (3.54%)
- Netherlands 1,611 (2.59%)

(Data from Sep 2020)

Data Contribution

- National contributions of raw viral sequence data
- Red boxes show data mobilised via SARS-CoV-2 Data Hubs and blue boxes show direct deposition into the European Nucleotide Archive (ENA) via the International Nucleotide Sequence Database Collaboration (INSDC).



Poll - In times of emergency such as the current pandemic, early data sharing should be the norm for all relevant results

- Agree
- Disagree
- Don't know / unsure

FAIR and data sharing, modelling and processing

- In addition to the platform, the Commission promotes FAIR and data sharing through:
 - The drafting of detailed [guidelines](#) on open access to publications and data **for Horizon 2020 projects**, encouraging beneficiaries to go beyond their contractual obligations
 - The guidelines share best practices on FAIR, open access, data management plans, and also provide links to standards, relevant repositories, and an overview of ongoing data-related efforts under the umbrella of EOSC
 - The **Research Data Alliance (RDA)** quickly responded to our request and set up a COVID-19 Working Group to deliver detailed [guidelines on data sharing under a health emergency](#)
 - **OpenAIRE** has recently launched a COVID-19 [gateway](#) to **aggregate and enable discovery** of relevant publications, datasets, other research outputs
- Furthermore: linking platform with **supercomputing** centres in Europe, for **data modelling and processing**

How do we move forward?

- **FAIR data** are an essential component in **Open Science policies** to use the wealth in data to generate new and impactful research and tackle the reproducibility crisis
- As a funder, the European Commission, will impose **stringent requirements** on all Horizon Europe projects with respect to publications, data, and other research outputs
- Specifically, with respect to data, the focus will be on **responsible research data management in line with the FAIR principles**
- The community-backed work developed under the **RDA FAIR Data Maturity Model WG** represents a blueprint on which further **evaluation mechanisms of FAIRness** can be built.
- **Assessment of FAIRness**, either by funders, or supported by FAIR-enabling repositories (certified or trustworthy), will raise awareness and increase levels of compliance across the board.

RoRI and FAIRsFAIR to partner on 'FAIRware' initiative | Funders across UK, Europe and Canada join forces to develop new open research tools.

22.09.2020



FAIRware
initiative



FAIRsFAIR
Fostering Fair Data Practices in Europe

22 September 2020

FAIRsFAIR actions



**FAIR Competences
for Higher
Education Design**

Online WORKSHOP

8-9 October 2020

The background of the poster features a repeating pattern of blue graduation caps and white rolled-up diplomas with red ribbons on a blue background.

- Organised by the University of Amsterdam as part of [FAIRsFAIR work on skills and training](#).
- The workshop will gather projects and experts on research data management skills in higher education to advance a “FAIR for Higher Education” framework.
- Participants will discuss methods, standards, and implementation aspects of advancing FAIR within higher education curricula.

[**REGISTER HERE**](#)

Poll - What else could funders do to help you manage and share your research data?

- Provide precise guidelines on FAIR requirements, repositories, etc
- Reward data sharing and other open science practices
- Provide funding for hiring data professionals/stewards
- Implement stricter data sharing policies

Keep in touch



ec.europa.eu/



europa.eu/



[@EU_Commission](https://twitter.com/EU_Commission)



[@EuropeanCommission](https://www.facebook.com/EuropeanCommission)



[European Commission](https://www.linkedin.com/company/european-commission)



[europeancommission](https://www.instagram.com/europeancommission)



[@EuropeanCommission](https://www.medium.com/@EuropeanCommission)



[EUTube](https://www.youtube.com/EUTube)



[EU Spotify](https://www.spotify.com/eu)

Thank you for listening

Question time



© European Union 2020

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

