# Data policy and data governance for Open Science in crisis situations

Iryna Kuchma, Open Access Programme Manager

EOSC Symposium 2022 CODATA International Data Policy Committee Session: An International Workshop on Data policy for Open Science in crisis situations, 16 November 2022

eifl

**KNOWLEDGE
WITHOUT
BOUNDARIES**

Coordinated, Cross-jurisdictional Efforts to Foster Global Open Science

Research funders and research performing institutions around the world must coordinate with one another, and support and promote Open Science through **policy and investment** to streamline the flow of data **between local entities, and across international jurisdictions**.
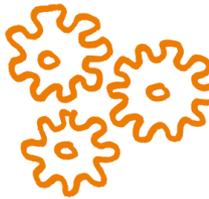
Similarly, governments, funders and policymakers should engage with big technology companies, mobile network operators, social network companies and others in the private sector who hold data that can better help understand the crisis and population behaviour. **Data sharing policies** should be adopted to encourage and facilitate data flows from data holders to the research community with the goal of protecting citizens' rights.

Incentivising the early publication/release of data outputs & software

The early publication/release of data outputs and the tools used to create them should be encouraged by **building trust, providing incentives for sharing data** (the value of sharing data for informing response during an emergency) **and providing appropriate governance**.

Research institutions and funding agencies can incentivise data stewardship, research software engineering data and software sharing by **creating structures for researchers to get credit for this work and by providing support for publishing data and software as valid research outputs**. This can include **developing research assessment systems that reward data and software outputs alongside publications and other research objects**.

Incentivising the early publication/release of data outputs & software (2)

Policymakers should **put guidelines into place that give researchers ease of mind** when licensing/sharing their data. All research data based on public research funding should be made available and exploitable in a timely manner, in particular for those of critical interest during an emergency situation. From a funding agency perspective, this could mean that increased weighting is given in the grant review process to researchers who demonstrate best practice in open data and data reproducibility with respect to their research outputs.

Peer-reviewed data articles
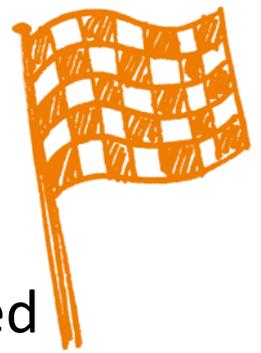
Peer-reviewed data articles should be treated as first-class research outputs equal in value to traditional peer-reviewed articles. In order to expedite reuse, data that could be used to advance research on crisis should be given top priority in the data publication process, fast-tracked by repositories, institutions, and other data publishers.

Funders need to make sure that calls for projects clearly state that for data "timely" publication means "as soon as possible after it has been collected" and not "as soon as the publication has been accepted by the journal". Publishers need to require publishing of the data, software and code underlying a study, in an even more timely manner than usual (data availability statements pointing to publicly available repository).

Infrastructure Investment & Economies of Scale

There is a need to **invest in state-of-the-art information technology (IT) and data management systems and infrastructure**, which includes hardware, networks, and the development and maintenance of critical research software. **The investment should also be directed towards people and skills to fully utilise the potential of large-scale infrastructure**: **the human resources required to maintain the infrastructure**, and the **training and support required to fully utilise the potential of large-scale infrastructure. The minimum required infrastructure for crisis response in terms of technology, skills, people and frameworks should be accessible to all jurisdictions/sectors**.
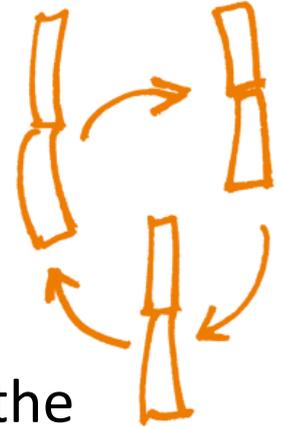
RDA COVID-19 Working Group. Recommendations and Guidelines on data sharing. Research Data Alliance. 2020. DOI: https://doi.org/10.15497/RDA00052

Infrastructure Investment & Economies of Scale (2)

Funding decisions should **prioritise projects where the data being produced can be used across domains and are linkable and interoperable**.
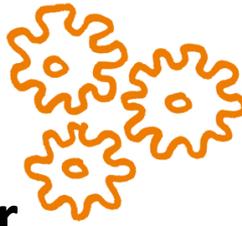
Funding should **require data sharing and provide support for infrastructure for data archiving and preservation**. This includes striving for funding models that are applied equitably across projects, researchers, and countries. This is also a **mandate for covering costs for infrastructure in the broadest sense** (e.g. ensuring open access to data, curation services, research data management costs across the lifecycle, and long-term preservation, among others).

FAIR and Timely

A **balance between achieving 'perfectly' FAIR outputs and timely sharing** is necessary with the key goal of immediate and open sharing as a driver.

What can be done to make the data as FAIR as possible with a reasonable time investment to enable availability and reusability of research data in order to prevent unnecessary duplication of work.

Researchers should be paired with data stewards to facilitate FAIR sharing, and data management should be considered at the start of a study or trial. FAIRifying data saves valuable time, as it facilitates a level of trustworthiness in research outputs.

RDA COVID-19 Working Group. Recommendations and Guidelines on data sharing. Research Data Alliance. 2020. DOI: https://doi.org/10.15497/RDA00052

FAIR and Timely (2)

Researchers should also be encouraged to **share what they have as-is** without fear of it being insufficient, **and signal that help is needed**. **The reusability of data can be increased with consistent preprocessing**. This would be a second-phase step that should not unnecessarily slow down researchers collecting data. In the crisis situation access to data should be as open as possible. This does not necessarily mean completely open access, as data must also be protected as necessary, but measures to control and manage risk (e.g. encryption, anonymisation, de-identification, aggregation, data use agreements) can be used to ease authorised access as much as possible, while still offering adequate protections.
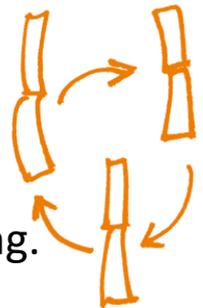
FAIR and Timely (3)

Finally, it is important to note that a lot of data that are very relevant to the crisis are kept exclusively on websites and are therefore extremely fragile. This data should be deposited in trusted data repositories, but to address existing gaps, the **websites should be web-archived systematically** (and permit doing so by way of their robots.txt), so as to ensure persistent availability of the information and to facilitate retrospective analyses. Preference should be given to public web archives that are created and stored by archival organisations.

5:00 AM

COALITION

OHCHR | Berkeley Protocol on

https://www.ohchr.org/en/publications/policy-and-methodological-publications/berkeley-protocol-d

United Nations

UNITED NATIONS
HUMAN RIGHTS
OFFICE OF THE HIGH COMMISSIONER

| What are human rights? ▼ | Topics ▼ | Countries ▼ | Instruments & mechanisms ▼ | Latest ▼ | About us ▼ |

Publications

POLICY AND METHODOLOGICAL PUBLICATIONS

Berkeley Protocol on Digital Open Source Investigations: A Practical Guide on the Effective Use of Digital Open Source and Information in Investigating Violations of International Criminal, Human Rights and Humanitarian Law

https://www.ohchr.org/en/publications/policy-and-methodological-publications/berkeley-protocol-digital-open-source

National Open Science Action Plan adopted by the Cabinet of Ministers of Ukraine on October 8, 2022

- FAIR RDM in national science/open science policies

- DMPs, support & training

- FAIR data repositories

- Monitoring

- EOSC & European data spaces

Thank you! Questions? Your thoughts?

Contact: iryna.kuchma@eifl.net

@irynakuchma

eifl **KNOWLEDGE WITHOUT BOUNDARIES**

www.eifl.net