RDA COVID-19 Recommendations and Guidelines

Main principles selected by the Open Science team at the Institut Pasteur

- The **availability of open research data** is a key component of pandemic preparedness and response. The challenge here is the trade-off between timeliness and precision. **The speed of data collection and sharing needs to be balanced with accuracy**, which takes time. The pressure to interpret results, turn studies around quickly and update statistics in almost real-time must not compromise quality and reliability.
- The consensus in this series of guidelines is that **research outputs should align with the FAIR principles**, meaning that data, software, models and other outputs should be Findable, Accessible, Interoperable and Reusable. However, there is also consensus that outputs need to be shared as quickly as possible in order to have a direct impact on the progress of the pandemic. **A balance between achieving 'perfectly' FAIR outputs and timely sharing is necessary with the key goal of immediate and open sharing as a driver**.
- In the COVID-19 situation access to data should be **as open as possible**. This does not necessarily mean completely open access, as data also need to be as closed as necessary, but measures to control and manage risk (anonymisation, aggregation, data use agreements) can be used to make access as easy as possible, while adequately protective.
- The **ethical and privacy considerations** around participant and patient data are significant in this crisis, and several guidelines note the need to **find a balance that takes into account individual, community and societal interests and benefits whilst addressing public health concerns and objectives.** Access to individual participant data and trial documents should be as open as possible and as closed as necessary, to protect participant privacy and reduce the risk of data misuse.
- Sharing data in a FAIR and timely way requires **planning for data management** early in the process of any research undertaking. Researchers should create a Data Management Plan (DMP) at the beginning of the research process so that it can be included in the work plan and the budget.
- To facilitate data quality control, timely sharing and sustained access, data should be deposited in **data repositories**. Whenever possible, these should be *trustworthy* data repositories (TDRs) that have been certified, subject to rigorous governance, and committed to longer-term preservation of their data holdings.
- The key to finding and using digital assets is **metadata**. COVID-19 research requires access to different assets for different communities. Within a given community, the commonly used metadata standards are well-known, but a researcher working across communities has more difficulty in locating relevant assets. In this case, a 'metadata element set' that is generally applicable**,** is required to be associated with each asset (e.g. see the metadata element set proposed by the RDA Metadata Interest Group).
- Research outputs need to be **well documented**, which includes documenting the following: research context, methodologies used to define, construct, and compile data, data cleaning and quality checks, data imputation, data provenance and so on. When sharing datasets, other relevant outputs (or documents) should also be made available, such as codebooks, lab journals, or informed consent form templates, so that data can be understood and potentially linked with other data sources.
- **Peer-reviewed data articles** should be treated as first-class research outputs equal in value to traditional peer-reviewed articles.

Source: RDA COVID-19 Working Group. Recommendations and guidelines. *Research Data Alliance*. DOI: https://doi.org/10.15497/rda00046