The Research Data Alliance COVID-19 Recommendations and Guidelines for Data Sharing

Fotis Psomopoulos (INAB|CERTH)

RDA COVID-19 WG Moderator

on behalf of the RDA COVID19 WG

**ELIXIR-GREECE All hands meeting
22 September 2020**

Adapted from the COVID-19 RDA IE webinar slides

Background

- Request from the European Commission to the Research Data Alliance

- Working Group set up with 4 Research Areas, 4 Cross-cutting themes

- Structured through a series of teams - chairs & moderators, research areas, editorial, visualisation team, Zotero library team

- April 1 - 30 June continual sprints, consultation, webinars, 6 releases

- 143 pages in the end; 4 page Executive Summary, Infographic +

- Exhausting and exhilarating

What are the Objectives?

1.0
Clearly define detailed guidelines on data and software sharing for COVID-19 research.

1.1
Help stakeholders follow best practices to **maximise efficiency.**

1.2
Act as a **blueprint** for future emergencies to maximise the efficiency of their work.

2
Develop **recommendations** for funders and policymakers to maximise timely, quality data and software sharing and appropriate responses in health emergencies.

3
Address interests of researchers, policymakers, funders, publishers, and providers of data sharing infrastructures.

A Collaborative Cross-Disciplinary Effort

The work has been divided into **four research areas** with **four cross-cutting themes.**

The guidelines and recommendations listed here are highlights. Please find more detailed information in the full-length publication.

(i) **Guidelines** - detailed practical advice aimed at researchers, data stewards, research software engineers, and public health officials.

(★) **Recommendations** - higher level generic advice aimed at policymakers, funders, publishers, and infrastructure providers.

CLINICAL

(i) Standardise terminologies, and find balance between timely data sharing and protecting privacy, confidentiality

(★) Organise data sharing and trial documents in trustworthy repositories

OMICS

(i) Select the best data formats and standards to fit the sub-discipline

(★) Promote use of domain-specific repositories to enable standardisation

EPIDEMIOLOGY

(i) Data models must include clinical data, disease milestones, indicators, reporting data, contact tracing and personal risk factors

(★) Incentivise publication of situational data, analytical models, scientific findings and reports

SOCIAL SCIENCES

(i) Enable interoperable cross-disciplinary, cross-cultural data use and collaboration

(★) Ensure robust funding streams for research aimed at understanding and managing the human aspects of the pandemic

Example from Omics
4.4.2 Guidelines for Host Genomics Data

1. **Gene expression data** should in general be retrieved from or deposited in the repositories listed below (Blaxter et al., 2016). To achieve load balancing, it is recommended to choose the respective regional repository. It should be noted that INSDC resources (i.e., DDBJ, ENA and NCBI) synchronise most of their datasets daily[2].
   1.1. Transcriptomics of human subjects (requiring authorised access):
      1.1.1. Database of Genotypes and Phenotypes (dbGaP) (Mailman et al., 2007)
      1.1.2. European Genome-Phenome Archive (EGA) (Lappalainen et al., 2015); the corresponding non-sensitive metadata will be available through EBI ArrayExpress (Athar et al., 2019)
      1.1.3. Japanese Genotype-phenotype Archive (JGA) (Kodama et al., 2015)
   1.2. Transcriptomics (from cell lines/animals):
      1.2.1. ArrayExpress (Athar et al., 2019)
      1.2.2. Gene Expression Omnibus (Barrett et al., 2013)
      1.2.3. Genomic Expression Archive
   1.3. Underlying reads can be retrieved from/will automatically be deposited to the corresponding read archive:
      1.3.1. DDBJ Sequence Read Archive (DRA) (Kodama et al., 2012), for submission documentation see here
      1.3.2. European Nucleotide Archive for submission documentation see here
      1.3.3. NCBI Sequence Read Archive (SRA) for submission documentation see here
   1.4. Microarray-based gene expression data:
      1.4.1. ArrayExpress (Athar et al., 2019)
      1.4.2. Gene Expression Omnibus (Barrett et al., 2013)
      1.4.3. Genomic Expression Archive
   1.5. Data on the originating sample can be retrieved from/will automatically be deposited to the corresponding sample archive:

👥 COMMUNITY

ⓘ Encourage public and patient involvement throughout data management lifecycle

★ Balance between timely testing and contact tracing, emergency response, community safety, and individual privacy concerns

🧭 INDIGENOUS DATA GUIDELINES

ⓘ Indigenous governance of data collection, ownership, and sharing and use priorities is the central principle of Indigenous data sovereignty

★ CARE Principles set minimum standards for collectors, users, and stewards of Indigenous data.

‹/› RESEARCH SOFTWARE

ⓘ Software used in data analysis must be able to reproduce results, if necessary

★ Allocate financial resources to support development and maintenance of new research software

🔨 LEGAL AND ETHICAL CONSIDERATIONS

ⓘ Although the law provides the foundation for data handling, ethical frameworks should also inform expedited approval to maximise data use and sharing

★ Expedite ethical review and approval for legal data sharing during a pandemic

Foundational Elements

What are the key recommendations?

The RDA COVID-19 Recommendations and Guidelines are aimed at developing a systematic approach for data sharing in public health emergencies that supports scientific research and policymaking, including an overarching framework, common tools and processes, and principles that can be embedded in research practice.

1 Coordinate cross-jurisdictional efforts to foster global **Open Science** through policy and investment.

2 Incentivise early publication and release of data and software outputs.

3 Invest in state-of-the-art IT, data management systems **infrastructure, economies of scale, and people**.

4 Data, software and models should be **timely and FAIR**: **F**indable, **A**ccessible, **I**nteroperable, **R**eusable.

5 Require the use of **Data Management Plans**.

6 Use common domain-specific **metadata standards**, and **persistent identifiers**.

7 Provide **documentation** of context, methodologies used to define, construct, and compile data, data cleaning and quality checks, data imputation, and data provenance.

8 Use **Trustworthy Data Repositories** committed to the long-term preservation and sustained access to their data holdings.

9 Use common generic as well as domain-specific **metadata standards**, and **persistent identifiers**.

10 **Balance ethics and privacy,** taking into account public interests and benefits while addressing the health crisis.

11 Access should be as **open as possible** and as **closed as necessary.**

12 Seek **technical solutions** that ensure anonymisation, encryption, privacy protection, and de-identification to **increase trust** in data sharing.

13 Provide **legal frameworks that promote sharing** of surveillance data across jurisdictions and sectors.

With thanks to the team at CANARIE and Research Data Canada for the Infographic
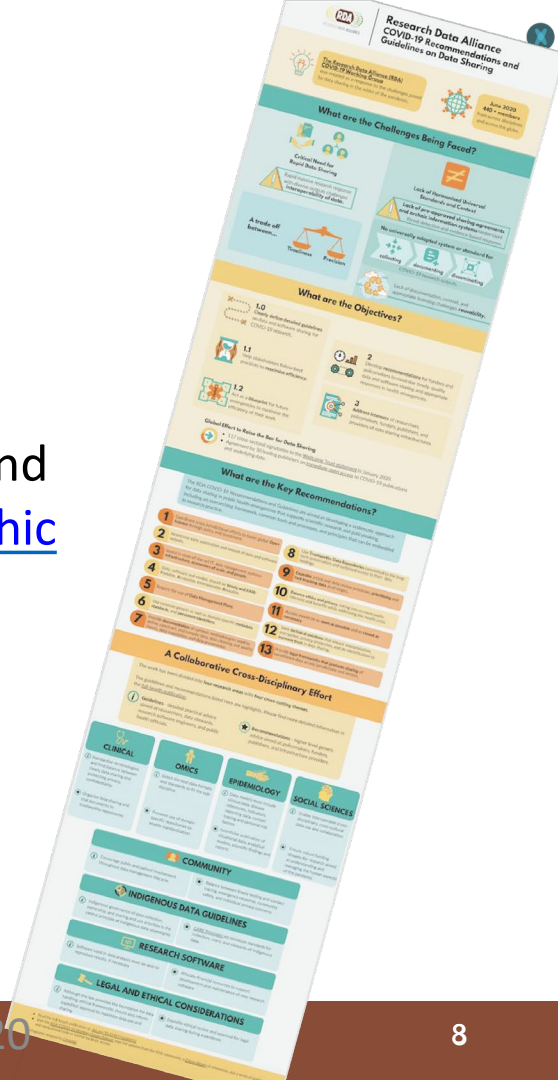
RDA Community response

Call to action to create a fast track Working Group aimed at developing a system for **data sharing in public health emergencies, specifically COVID-19**

- Around 600 RDA members and newcomers registered for the different groups
- 165 active contributors to the documents
- 6 Co-Chairs + Secretariat
- Experts in different fields as group moderators
- Regular calls and iterations
- Weekly webinars, requests for comments
- 5 releases produced (April – June 2020)
- Final release – 30 June 2020

RDA's guiding principles:

✔ **Openness**
✔ **Consensus**
✔ **Balance**
✔ **Harmonization**
✔ **Community-driven**
✔ **Non-profit and technology-neutral**

by Mark Leggott | rdc-drc.ca

Ongoing work and future steps

Journal Articles and Endorsements
- o 4-5 articles completed or in preparation by COVID-19 WG Members

RDA Groups and RDA Plenary 16 Sessions
- o Broader efforts under RDA WGs
- o Infectious Disease BOF (goal to create a WG/CoP)
- o Community Participation BoF / Citizen Science

Stakeholder support

- Adoption and implementation of the recommendations and guidelines;

- Policymakers, funders and publishers have a major influence on the behaviour of researchers and data stewards;

Adapted from Mark Leggott | rdc-drc.ca

Value of RDA for COVID-19

https://www.rd-alliance.org/value-rda-covid-19-0

The Research Data Alliance COVID-19 Data Sharing Recommendations and Guidelines
by the community
for the community

RDA

Email - **enquiries@rd-alliance.org**

Web - **www.rd-alliance.org**

Twitter - @resdatall

LinkedIn - **www.linkedin.com/in/ResearchDataAlliance**

Slideshare - **http://www.slideshare.net/ResearchDataAlliance**