THE RESEARCH DATA ALLIANCE
COVID-19 DATA SHARING
RECOMMENDATIONS AND GUIDELINES

Natalie Harrower | RDA COVID-19 WG CoChair
Timea Biro | RDA COVID-19 WG Moderator

RDA Ireland - Meet the Expert Series
29 July 2020

RDA COVID-19 Recommendations and Guidelines on data sharing https://doi.org/10.15497/rda00052
Today’s webinar

- What is the RDA COVID-19 working group and why should we pay attention to its outputs?
- Overview of the RDA COVID-19 Recommendations and Guidelines on Data Sharing https://doi.org/10.15497/rda00052
- Learnings from the collaborative writing process
- Example from a sub-group: ‘Community Participation’
- Q&A
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 Challenges Being Faced?

Critical Need for Rapid Data Sharing

- Rapid massive research response with diverse outputs challenges interoperability of data.

Lack of Harmonised Universal Standards and Context

- Lack of pre-approved sharing agreements and archaic information systems hinder rapid threat detection and evidence-based response.

A trade off between...

- Timeliness
- Precision

No universally adopted system or standard for COVID-19 research outputs.

Lack of documentation, context, and appropriate licensing challenges reusability.
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.

**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**

- Standardise terminologies, and find balance between timely data sharing and protecting privacy, confidentiality
- Organise data sharing and trial documents in trustworthy repositories

**OMICS**

- Select the best data formats and standards to fit the sub-discipline
- Promote use of domain-specific repositories to enable standardisation

**EPIDEMIOLOGY**

- 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**

- 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
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.

**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**
- Standardise terminologies, and find balance between timely data sharing and protecting privacy, confidentiality
- Organise data sharing and trial documents in trustworthy repositories

**OMICS**
- Select the best data formats and standards to fit the sub-discipline
- Promote use of domain-specific repositories to enable standardisation

**EPIDEMIOLOGY**
- 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**
- 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, EGA, and NCBI) synchronise most of their datasets daily.

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
## 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
10.4.5 Consent

Consent is the act by which a participant, patient or data subject indicates that they permit something to happen to them, or to their data, which would otherwise not be able to happen. It covers a number of different specific contexts:

1. Clinical: a patient agrees to undergoing a procedure, including taking part in a trial;
2. Data Protection: a data subject agrees to personal data being processed for specified purposes;
3. Research: a participant agrees to take part in a research study or experiment.

In both cases, the informed consent sheets for clinical or research purposes would explicitly set out how data protection will be handled, as well as samples or biobanking, rights to self-images and others.

Giving consent should be informed (e.g. the individual knows what is going to happen and why), freely given (there is no coercion or similar motivation), given by somebody with capacity, unambiguous and auditable (the consent is recorded somewhere) (See also Parra-Calderón, 2018). Depending on the jurisdiction and the research domain, there may be an additional requirement to seek consent. This may include a representative community board as well as participants themselves.

Ideally, consent should be sought for collecting, processing, sharing and publishing data. However, there are other legal bases for processing personal data. Some specific examples from the European General Data Protection Regulation (GDPR, 2016) are described below. Our recommendation would therefore be as follows:

1. Where possible, use data where the data subject has provided a valid consent that includes or is compatible with intended use of the data and complies with the requirements on consent in the specific country or region.

Where these are not possible, there are other reasons why data may be used (see Hallinan, 2020, Ó Cathasaír et al., 2020). For example, there may be a different legal basis for using personal data.

2. If using personal data, check whether there may be another basis for using the data.

In Europe, for instance, the GDPR provides other legal bases for processing personal data:
## 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: Findable, Accessible, Interoperable, Reusable.
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

29/07/20
Community participation sub-group - example

A cross-cutting theme - looking at communities, data sharing and addressing the emergency

Use cases of community generated data via apps and data challenges related to participatory disaster response strategies

- Group membership of roughly 120
- Regular group and coordination calls
- Brainstorming and several writing sprints, survey for use cases to address
- Several reiterations - with feedback

- Policy recommendations around transparency, community participation and data governance; encouraging inclusiveness and incremental and multidisciplinary approach, ethics & privacy
- Guidelines for researchers & teams around data collection and data stewardship in particular

<table>
<thead>
<tr>
<th>Community</th>
<th>Need specific guidelines for enabling citizen scientists undertaking research to contribute to a common body of knowledge</th>
<th>Encourage public and patient involvement (PPI) throughout the data management lifecycle from research question to final data sharing and usage</th>
<th>Balance between timely testing and contact tracing, emergency response, community safety and individual privacy concerns</th>
</tr>
</thead>
</table>

29/07/20
Cite as: "RDA COVID-19 WG (2020). Recommendations and Guidelines: Zotero library of supporting resources and references, Version 1.0. DOI: https://doi.org/10.15497/rda00051"
I. Guidelines for Researchers

The COVID pandemic does not serve to remove the basic validity of the rights and interests on which these documents and principles are based. In other words, formal protocols for conducting research are required both during a pandemic and at other times, unless otherwise modified by the relevant authorities. The emergency does, however, mandate a reconsideration of the balance between these rights and interests - in particular between a research subject's right to privacy and the public interest in the outcome of research. In some cases, this reconsideration has led to legitimate time limited adaptations of, or derogation from, normally applicable principles (Section 10.2)

1. Will you be re-using existing data?
   - Desirable: Before Submitting the Proposal
   - a. No
   - b. Yes

2. Will you be collecting data?
   - Desirable: Before Submitting the Proposal
   - External Links: FAIR principles

- Developing knowledge model of the Recommendations & Guidelines
- Tailored pathways for different stakeholders including researchers, funders, policy makers
- Customizable to different jurisdictions
- Lead-up to a mind map tool

WOMEN & MEN at work

29/07/20
Ongoing work and future steps

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

RDA Groups and RDA Plenary 16 Sessions
  • Broader efforts under RDA WGs
  • Infectious Disease BOF (goal to create a WG/CoP)
  • 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
The Value of RDA for COVID-19

Under public health emergencies, and particularly the COVID-19 pandemic, it is fundamental that data is shared in both a timely and an accurate manner. This coupled with the harmonisation of the many diverse data infrastructures is, more than ever, imperative to ensure preliminary data is made readily available. It is clear that open research data is a key component to pandemic prep, desk and response.

In late March, RDA received a direct request from one of its members, the European Commission, to create global guidelines and recommendations for data sharing under COVID-19 circumstances. Over 600 data professional and domain experts signed up and began work in early April 2020. They produce a list of detailed guidelines to help researchers and data owners follow best practices to maximise the efficiency of their work, and to act as a blueprint for future emergencies. Updated with recommendations to help policymakers and funders to maximise timely, quality data sharing and to support research into health emergencies.

On 30 June 2020, RDA published the final version of the RDA COVID-19 Recommendations and Guidelines on data sharing covering four research areas: clinical data, anxiety practice, epidemiology and social sciences. Complemented by an extensive summary focusing on legal and ethical considerations, research software, community participation and indigenous data.

The Outputs

The COVID-19 WG from April 1st through June 30th, 2020 created more than five releases of the recommendations and guidelines leading to the final endorsed version, "RDA COVID-19 Recommendations and Guidelines for Data Sharing," with ongoing efforts to add and review materials.

The Outputs

The COVID-19 WG, from April 1st through June 30th, 2020, created more than five releases of the recommendations and guidelines leading to the final endorsed version, "RDA COVID-19 Recommendations and Guidelines for Data Sharing," with ongoing efforts to add and review materials.

- RDA COVID-19 Recommendations and Guidelines for Data Sharing, Final release, published 30 June 2020
- RDA COVID-19 Recommendations and Guidelines for Data Sharing, Infographic
- RDA COVID-19 Guidelines and Recommendations - the prior 5 releases
- RDA COVID-19 WG, OAI library

Citation: RDA COVID-19 Working Group. Recommendations and Guidelines on data sharing. Research Data Alliance. 2020. DOI: https://doi.org/10.15497/rda0052

Resources

- Final executive summary

Joint Statements

- RDA COVID-19 Recommendations and Guidelines for Data Sharing: How STM Publishers can Contribute (July 2020)
- Gilda, RDA COVID-19 Guideline for Data Sharing Respecting Indigenous Data Sovereignty (July 2020)
- The Duty to Document does not Cease in a Crisis. It becomes more Essential. (May 2020)
- Data Together: COVID-19 Appeal and Action (March 2020)

RDA for COVID-19 Events

A series of weekly "RDA COVID-19 Update Webinars" occurred almost every Tuesday between April and June 2020 and provided updates on the evolving COVID-19, legal and ethical, research software, Community Participation Working Group, Indigenous Data contribution, and the four research themes (clinical, social, evidence, social sciences) along with an opportunity for members to ask questions. Recordings and presentations from these sessions are posted on the RDA event calendar.

Upcoming events include:

- RDA Ireland Meet The Experts Webinar - Data Sharing for COVID-19: Research Recommendations and Guidelines from the RDA COVID-19 Working Group - 23 July 2020
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

Dr Natalie Harrower, Director Digital Repository of Ireland
Timea Biro, RDA EU Project Manager
www.dri.ie
@dri_ireland