



Anne Cambon-Thomsen, MD, Emeritus Research director CNRS,
RDA ambassador and member of the European Group on Ethics of science and new technologies.
UMR 1027, Inserm, Université Toulouse III Paul Sabatier, Epidémiologie et analyses en santé publique;
équipe BIOETHICS (team leader : Emmanuelle Rial-Sebbag)

RDA-COVID19 Working Group

<https://www.rd-alliance.org/groups/rda-covid19>

FAPESP COVID-19 research webinar 5/8/2020

Open Data Under the COVID-19 Pandemic

An example regarding the organisation of international
cooperation around COVID-19 data.

RDA = Research data alliance

FAPESP = Sao Paulo Research Foundation



Outline of presentation

- 🌐 **Basic elements on RDA**
- 🌐 Context and characteristics of the RDA COVID-19 working group
- 🌐 Methods, organisation and steps
- 🌐 Outputs and other productions
- 🌐 Visualisation and help to use the document
- 🌐 How to foster impact?
- 🌐 Conclusion
- 🌐 Discussion





RDA in a Nutshell

Vision

Researchers and innovators openly share and re-use data across technologies, disciplines, and countries to address the grand challenges of society.

Mission

RDA builds the **social and technical bridges** that **enable open sharing and re-use** of data.

Some slides have been prepared by Hilary Hanahoe, Secretary general RDA

THE RESEARCH DATA ALLIANCE

www.rd-alliance.org

building the social and technical bridges that enable open sharing of data

45 FLAGSHIP OUTPUTS

including 12 ICT Technical Specifications

100+ ADOPTION CASES

across multiple disciplines, organisations & countries

98 GROUPS WORKING ON GLOBAL DATA INTEROPERABILITY CHALLENGES

40 Working Groups
58 Interest Groups

10,890 INDIVIDUAL MEMBERS FROM 145 COUNTRIES

69% Academia & Research
14% Public Administration
11% Enterprise & Industry

51 ORGANISATIONAL MEMBERS 11 AFFILIATE MEMBERS





Outline of presentation

- Basic elements on RDA
- **Context and characteristics of the RDA COVID-19 working group**
- **Methods, organisation and steps**
- Outputs and other productions
- Visualisation and help to use the document
- How to foster impact?
- Conclusion
- Discussion





RDA community response to COVID-19

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

by the community for the community

How is it both classical and exceptional?

5 August 2020



Objectives

Recommendations and guidelines:

- Detailed **guidelines** to help **researchers** and **data stewards** follow best practices to maximise the efficiency of their work, and to act as a blueprint for future emergencies;
- **Recommendations** to help **policymakers** and **funders** to maximise timely, quality data sharing and appropriate responses in such health emergencies.
- **Address interests** of researchers, policymakers, funders, publishers, and providers of data sharing infrastructures.





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
- 🌐 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**



Working group co-chairs



Juan Bicarregui
STFC



Anne Cambon-Thomsen
CNRS



Ingrid Dillo
DANS



Natalie Harrower
DRI



Sarah Jones
DCC



Mark Leggott
Research Data
Canada



Priyanka Pillai
University of
Melbourne

<https://www.rd-alliance.org/groups/rda-covid19>



RDA COVID-19 Guidelines and Recommendations

Clinical



Omics



Epidemiology



Social Sciences



Community Participation for Data Sharing under COVID-19

Indigenous Data under COVID-19

Legal and Ethical Considerations under COVID-19

Research Software for Data Sharing under COVID-19



Overarching Guidelines Editorial Team



Christophe Bahim
PWC



Ingrid Dillo
DANS



Natalie Harrower
DRI



Mark Leggott
Research Data
Canada



Nicolas Loozen
PWC



Priyanka Pillai
University of
Melbourne



Megan
Underwood
RDA-US/RNA



Mary O'Brien
Uhlmansiek
RDA-US/RNA



Bridget Walker
RDA
Foundation





Outline of presentation

- Basic elements on RDA
- Context and characteristics of the RDA COVID-19 working group
- Methods, organisation and steps
- **Outputs and other productions**
- Visualisation and help to use the document
- How to foster impact?
- Conclusion
- Discussion





WG outputs

- Best practice guidelines per sub-group and overall
- Recommendations for researchers and policymakers to help maximise data sharing
- Catalogue of key resources to inform guidelines (e.g. key datasets, standards, tools, repositories etc)
- A visualisation tool to facilitate navigation to specific resources by different stakeholders





Outputs, support outputs, other outputs

- 🌐 A 145 page document with context, general guidelines and recommendations, specific (per sub-group)
- 🌐 An executive summary
- 🌐 Supporting output: RDA-COVID19-WG Zotero Library
 - <https://doi.org/10.15497/rda00051>
 - The RDA-COVID19-WG Zotero Library was created to support development of the RDA-COVID19 Data Sharing Recommendations and Guidelines. It is an extensive, but not a comprehensive bibliography. The Library is publicly available on the Web at:
 - https://www.zotero.org/groups/2485086/rda-covid19_wg
- 🌐 Other outputs: Epidemiology document (endorsed), articles in preparation





Research Areas & Cross-cutting themes



CLINICAL



OMICS



EPIDEMIOLOGY



SOCIAL SCIENCES



COMMUNITY



INDIGENOUS DATA GUIDELINES



RESEARCH SOFTWARE

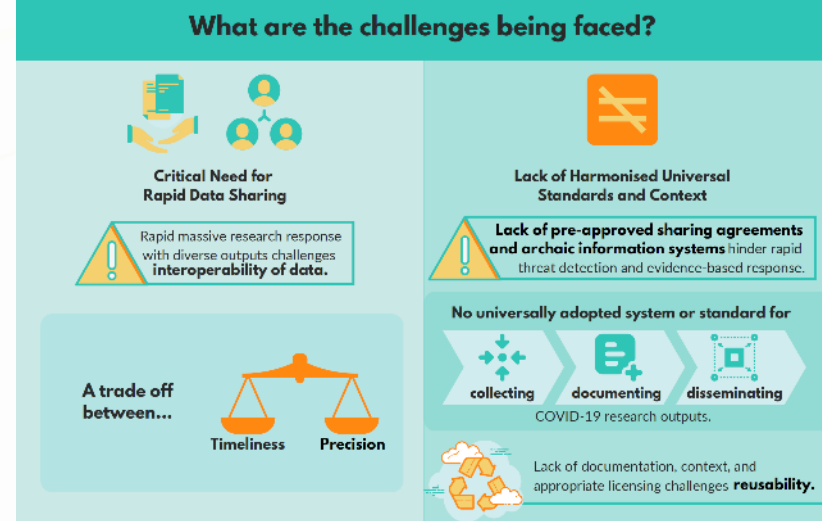


LEGAL AND ETHICAL CONSIDERATIONS



Overarching Challenges

- ✓ Rapid sharing of research efforts, findings, and data
- ✓ Timeliness vs. precision
- ✓ Lack of “pre-approved” data sharing agreements
- ✓ No universal standard / system for COVID-19 research outputs
- ✓ Cross-disciplinary reusability
- ✓ Contextualisation, licensing, documentation
- ✓ Ad-hoc Research software





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.



Cross cutting theme highlights

Cross cutting themes	Challenges	Guidelines for researchers	Recommendations for funders/policy makers
Clinical	Promotion of clinical data sharing is important due to many studies and trials being performed under enormous time pressure	Standardised clinical terminologies should be used and a fair balance achieved between timely data sharing and protecting privacy and confidentiality	Measures should be taken in order to organise the sharing of data and trial documents in a suitable, trustworthy and secure data repository
Omics	An increased need of rapid openness for omics data to gain early insights into molecular biology of the processes at cellular level	Omics research should be a collaborative effort to learn the genetic determinants of COVID-19 susceptibility, severity and outcomes	Promote use of domain-specific repositories to enable standardisation of terms and enforce metadata standards
Epidemiology	Data and models are frequently incomplete, provisional, and subject to correction under changing conditions	Data models must include clinical data, disease milestones, indicators and reporting data, contact tracing and personal risk factors	Incentivise the publication of situational data, analytical models, scientific findings, and reports used in decision-making
Social Sciences	Require equal inclusion of social and economic context with health-related information to enable evidence-based decision-making	Enable interoperable cross-disciplinary and cross-cultural data collection, data use and collaboration for managing social sciences data during pandemics	Ensure robust funding streams for social sciences research for understanding and managing the human aspects of pandemics



Cross cutting theme highlights



Cross cutting themes	Challenges	Guidelines for researchers	Recommendations for funders/policy makers
Community	Need specific guidelines for enabling citizen scientists undertaking research to contribute to a common body of knowledge	Encourage public and patient involvement (PPI) throughout the data management lifecycle from research question to final data sharing and usage	Balance between timely testing and contact tracing, emergency response, community safety and individual privacy concerns
Indigenous Data Guidelines	Indigenous data rights, priorities and interests must be recognised in all COVID-19 research and surveillance activities	Indigenous governance of data collection, ownership, sharing and use priorities is the central principle of Indigenous data sovereignty	CARE Principles of Indigenous Data Governance set minimum standards for collectors, users and stewards of data
Legal and Ethical Considerations	Achieve a balance between rights of people and interests of researchers and policymakers	Ethical instruments should be interpreted with the law, and can guide the interpretation of the law if the law does not address a particular issue	During a pandemic, ethical review and approval for legally sharing data should be expedited
Research Software	Need systems in place for sharing of research software and accelerated and reproducible research during a pandemic	It is critical for software that is used in data analysis to produce results that can, if necessary, be reproduced	Funders must allocate financial resources to support the development and maintenance of new research software



Table of content main document

Executive Summary 7

1. Objectives and Use of This Document 11

2. Foundational Elements 14

2.1 Challenges 14

2.2 Recommendations 15

2.2.1 Coordinated, Cross-jurisdictional Efforts to Foster Global Open Science 15

2.2.2 Infrastructure Investment & Economies of Scale 16

2.2.3 FAIR and Timely 17

2.2.4 Data Management Planning 17

2.2.5 Metadata 18

2.2.6 Documentation 19

2.2.7 Use of Trustworthy Data Repositories 19

2.2.8 Publications / Data Publications 20

3. Data Sharing in Clinical Medicine 21

3.1 Focus and Description 21

3.2 Scope 21

3.3 Policy Recommendations 21

3.3.1 Trustworthy Sources of Clinical Data 21

3.4 Guidelines 23

3.4.1 Data and Metadata Standards for Clinical Data 23

3.4.2 Clinical Trials on COVID-19 24

3.4.3 Immunological, Imaging and Healthcare Data 25

4. Data Sharing in Omics Practices 27

4.1 Focus and Description 27

4.2 Scope 27

4.3 Policy Recommendations 27

4.3.1 Researchers Producing Data 27

4.3.2 Policymakers & Funders 27

4.4 Guidelines 28

4.4.1 Guidelines for Virus Genomics Data 28

4.4.2 Guidelines for Host Genomics Data 29

4.4.3 Guidelines for Structural Data 31

4.4.4 Guidelines for Proteomics 33

4.4.5 Guidelines for Metabolomics 34

4.4.6 Guidelines for Lipidomics 35



5. Data Sharing in Epidemiology 37



- 5.1 Focus and Description 37
- 5.2 Scope 37
 - 5.2.1 Supporting Output 37
- 5.3 Policy Recommendations 37
 - 5.3.1 Information Technology and Data Management 37
 - 5.3.2 COVID-19 Epidemiological Data, Analysis and Modelling 38
- 5.4 Guidelines 38
 - 5.4.1 COVID-19 Population Level Data Sources 39
 - 5.4.2 Interoperable COVID-19 Epidemiological Surveillance: Clinical and Population-based Instruments 40
 - 5.4.3 Preservation of Individuals’ Privacy in Shared COVID-19 Related Data 41
 - 5.4.4 Full Spectrum View of the COVID-19 Data Domain: An Epidemiological Data Model 41
 - 5.4.5 Epi-TRACS: Rapid Detection and Whole System Response for Emerging Pathogens 42
 - 5.4.6 COVID-19 Emergency Public Health and Economic Measures Causal Loops: A Computable Framework 42
 - 5.4.7 Common Data Models and Full Spectrum Epidemiology: Epi-STACK Architecture for COVID-19 Epidemiology Datasets 42

6. Data Sharing in Social Sciences 44

- 6.1 Focus and Description 44
- 6.2 Scope 44
- 6.3 Policy Recommendations 45
- 6.4 Guidelines 46
 - 6.4.1 Data Management Responsibilities and Resources 46
 - 6.4.2 Documentation, Standards, and Data Quality 46
 - 6.4.3 Storage and Backup 47
 - 6.4.4 Legal and Ethical Requirements 47
 - 6.4.5 Data Sharing and Long-term Preservation 48





7. Community Participation and Data Sharing 50

7.1 Focus and Description 50

7.2 Scope 50

7.3 Policy Recommendations 52

7.3.1 Transparency, Community Participation and Data Governance 52

7.3.2 Inclusive, Incremental and Multidisciplinary Approach 52

7.3.3 Legal and Ethical Aspects 53

7.3.4 Software Development 53

7.4 Guidelines 54

7.4.1 Data Collection 54

7.4.2 Data Quality and Documentation 54

7.4.3 Data Storage and Long-term Preservation 55

8. Indigenous Populations and Data Sharing 56

8.1 Focus and Description 56

8.2 Scope 57

8.3 Policy Recommendations and Guidelines 57

9. Research Software Sharing for Data Analysis 61

9.1 Focus and Description 61

9.2 Scope 61

9.3 Policy Recommendations 61

9.4 Guidelines for Publishers 63

9.5 Guidelines for Researchers 65

10. Legal and Ethical Considerations 67

10.1 Focus and Description 67

10.2 Scope 67

10.3 Policy Recommendations 68

10.3.1 Initial Recommendations 68

10.3.2 Relevant Policy and Non-Policy Statements 68

10.4 Guidelines 69

10.4.1 Cross-Cutting Principles 69

10.4.2 Hierarchy of Obligations 69

10.4.3 Seeking Guidance 71

10.4.4 Anonymisation 72

10.4.5 Consent 74

10.4.6 Licensing Data and Licensing Software 75

10.4.7 The 5 Safes Model 76

10.4.8 Vulnerable Groups 77





Table of Figures 6

Table of Tables 6

Executive Summary 7

1. Objectives and Use of This Document 11

11. Glossary 78

12. Acronyms 89

13. Additional Resources 95

14. References 100

15. Contributors 141

Legend of Figures

Figure 1 - RDA COVID-19 WG sub-groups including research areas and cross-cutting themes 11

Title of Tables

Table 1 - Summary of challenges, guidelines and recommendations 9

Table 2 - COVID-19 population level data sources 37

Table 3 - Questionnaire instruments: Reference studies 37

Table 4 - Questionnaire instruments: Resources 38





Outline of presentation

- 🌐 Basic elements on RDA
- 🌐 Context and characteristics of the RDA COVID-19 working group
- 🌐 Methods, organisation and steps
- 🌐 Outputs and other productions
- 🌐 **Visualisation and help to use the document**
- 🌐 **How to foster impact?**
- 🌐 Conclusion
- 🌐 Discussion





How to use the document?

- 🌐 No one will generally need to explore the full document
- 🌐 The need will be for a rather specific question or domain
- 🌐 How to quickly find what you need?
- 🌐 A « visualisation group » is working on 3 tools
 - An infographics
 - A mindmap
 - A wizard





How to foster impact?

- 🌐 Recommendations and guidelines that are not implemented are useless!
- 🌐 Dissemination to communities and institutions is essential
- 🌐 These recommendations are broad enough to be of interest for numerous stakeholders across the world





Outline of presentation

- 🌐 Basic elements on RDA
- 🌐 Context and characteristics of the RDA COVID-19 working group
- 🌐 Methods, organisation and steps
- 🌐 Outputs and other productions
- 🌐 Visualisation and help to use the document
- 🌐 How to foster impact?
- 🌐 **Conclusion**
- 🌐 **Discussion**





Highlights of general policies

1. Promote FAIR sharing of COVID-19 related Data
2. Encourage expedited review processes for pandemic related research
3. Encourage researchers to apply generic metadata for COVID-19 related data
4. Request authors to add available contextual documentation
5. Promote the use of trustworthy repositories
6. Respect the ground rules for Ethics, Privacy and sound Legal Frameworks in COVID-19 data sharing



Take home messages

- Data availability in time of a pandemic is a priority
- A variety of well documented data is needed, not just numbers of cases and deaths, with good metadata (interdisciplinary oriented)
- The Research data alliance (RDA) has issued guidelines and recommendations to increase data sharing, especially regarding standards and trustworthy data sources, as well as policy oriented
- These recommendations are in principle applicable in a broad geographical coverage, (not specific to a given jurisdiction)
- As interdisciplinarity is needed, referring to community accepted standards was chosen, even if already well known within a specific community.





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

5 August 2020

Contact Anne Cambon-Thomsen :
anne.cambon-thomsen@univ-tlse3.fr

RDA

Email - enquiries@rd-alliance.org

Web - www.rd-alliance.org

Twitter - [@resdatall](https://twitter.com/resdatall)

LinkedIn -

www.linkedin.com/in/ResearchDataAlliance

Slideshare -

<http://www.slideshare.net/ResearchDataAlliance>

Hilary Hanahoe - Contact

Secretary General, Research Data Alliance

Email - hilary.hanahoe@rda-foundation.org

Twitter - [@hilaryhanahoe](https://twitter.com/hilaryhanahoe)

