RDC DRC
Research Data Canada – Données de recherche Canada

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

Natalie Harrower | RDA COVID-19 WG Co-Chair, Director Digital Repository of Ireland @natalieharrower
Mark Leggott | RDA COVID-19 WG Co-Chair, Research Data Canada @rdc_drc

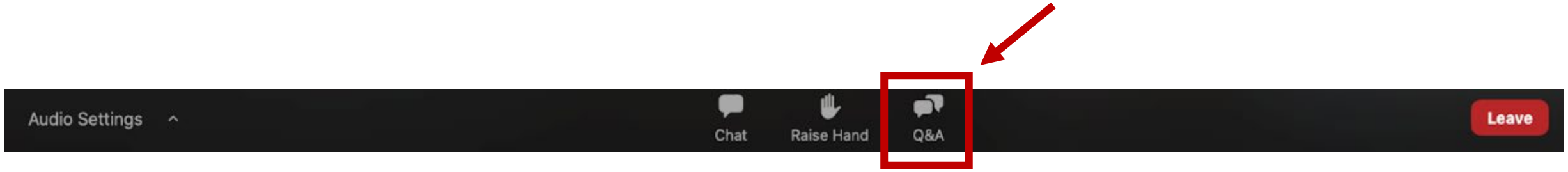
Agenda

- 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*
- Learnings from the collaborative writing process
- Example from a sub-group: ‘Indigenous Data’
- Q&A

https://doi.org/10.15497/rda00052

Questions & Answers

- Please use the Q&A option to ask questions of the presenters. Questions will be monitored throughout the session and addressed at the end.
- The Q&A option can be found at the bottom of your Zoom screen:



- Please note that this **event is being recorded**, including questions and answers.

Background and Why

- Request from the European Commission to the Research Data Alliance (RDA)
- Working Group setup within weeks
  - 4 Research Areas, 4 Cross-cutting themes, each with Moderators
- Structured through a series of teams
  - Co-Chairs, Co-Chairs + Moderators, 8 Themes, Editorial, Visualization, Zotero
- April 1 - 30 June continual sprints, consultations, webinars,
  - 6 releases over 3 months
  - 143 pages in the end; 4-page Executive Summary, Infographic +
- Exhausting and exhilarating!

Overview of the Guidelines

What are the Challenges Being Faced?

Critical Need for Rapid Data Sharing

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

A trade off between...

Timeliness **Precision**

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.

No universally adopted system or standard for

collecting **documenting** **disseminating**

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

Omics Example: 4.4.2 Guidelines for Host Genomics Data

Several different types of host genomics data are being collected for COVID-19 research. Some suitable repositories for these are:

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)

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

Legal/Ethics Example:10.4.5 Consent Guidelines

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.

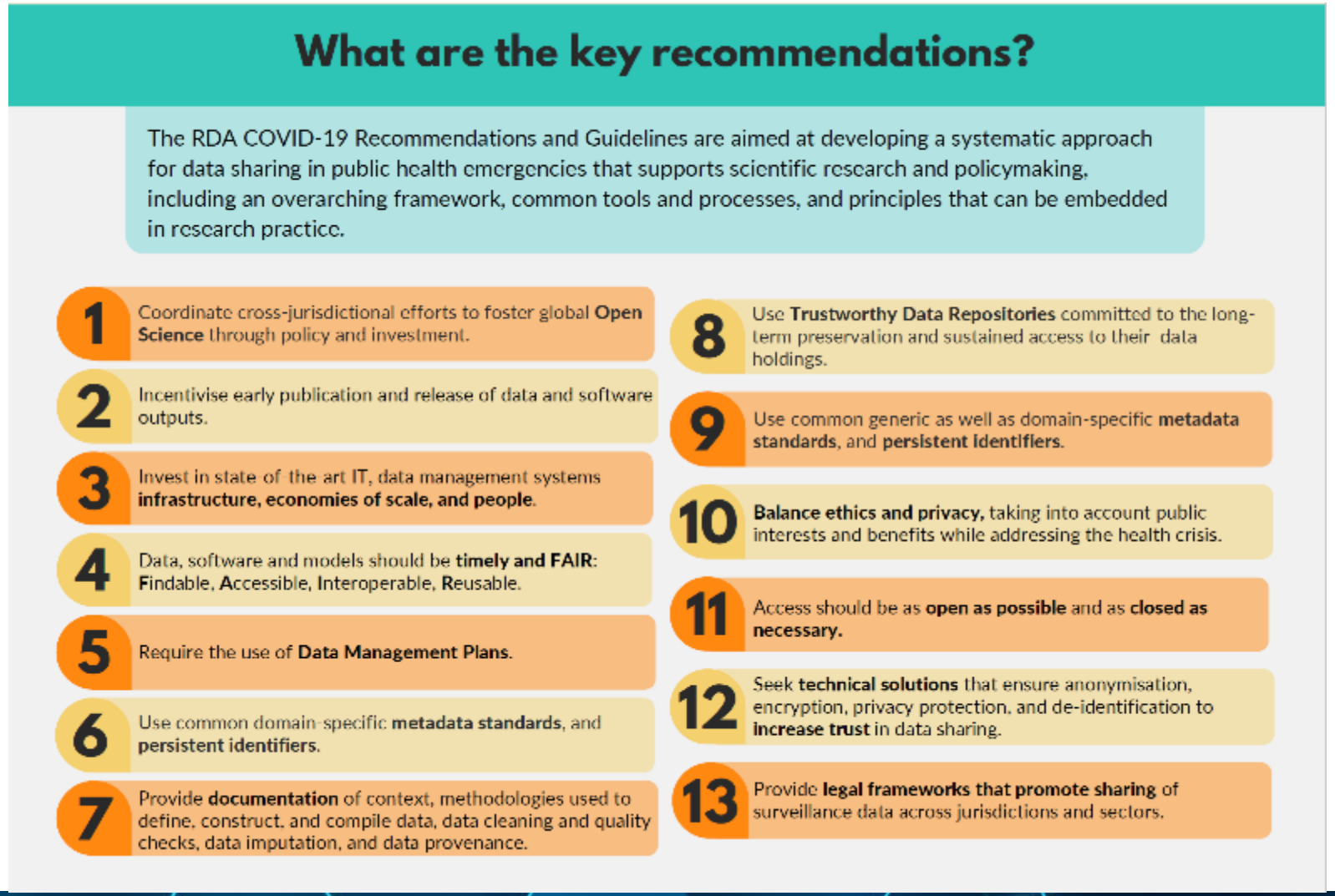
Giving consent should be informed (e.g. the individual knows what is going to happen and why), freely given (there is no coercion or similar motivation), given by somebody with capacity, unambiguous and auditable (the consent is recorded somewhere) (See also Parra-Calderón, 2018). Depending on the jurisdiction 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

Foundational Elements



Foundational Elements

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**

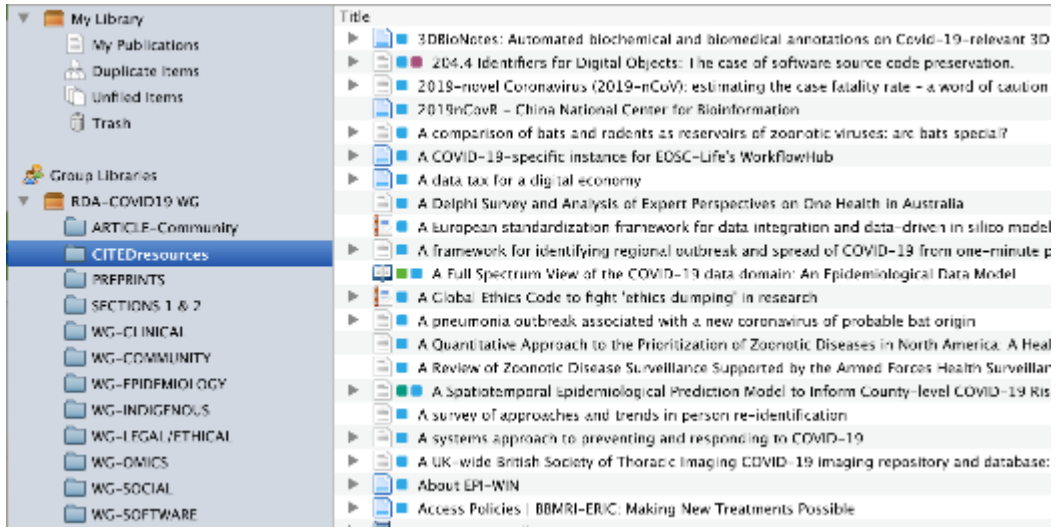
Indigenous Data sub-group Example

- Group membership of a dozen
- Produced the output at the final stages, in 2-3 weeks
- Based on the CARE Principles previously written by the Global Indigenous Data Alliance (GIDA)
- COLLECTIVE BENEFIT: “Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data.”

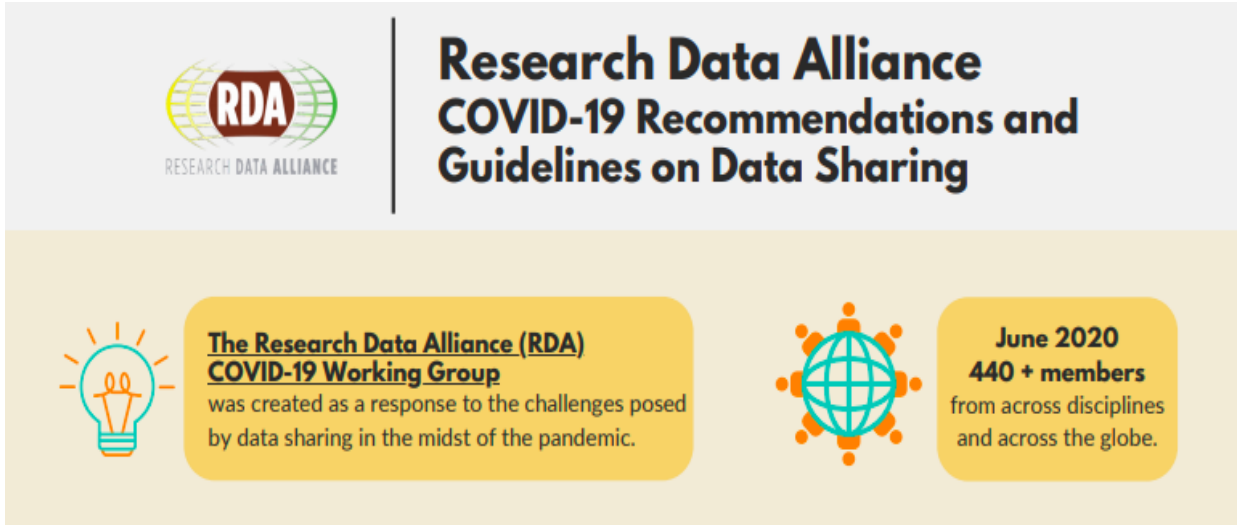
Indigenous Data Guidelines	Indigenous data rights, priorities and interests must be recognised in all COVID-19 research and surveillance activities	Indigenous governance of data collection, ownership, sharing and use priorities is the central principle of Indigenous data sovereignty	CARE Principles of Indigenous Data Governance set minimum standards for collectors, users and stewards of data

RDA COVID-19 Zotero Library & Infographic

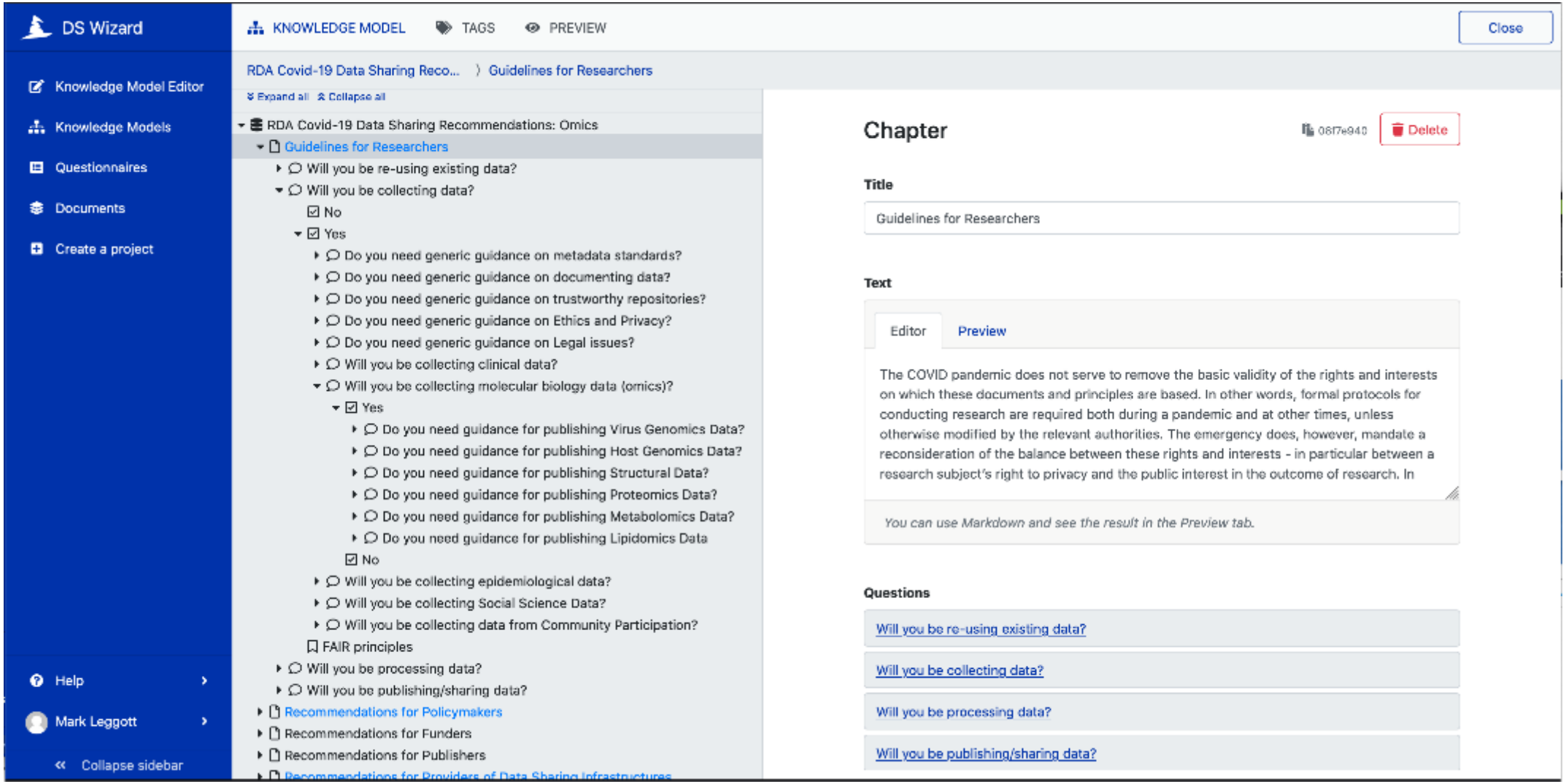
RDA COVID-19 Zotero Library



RDA COVID-19 Infographic



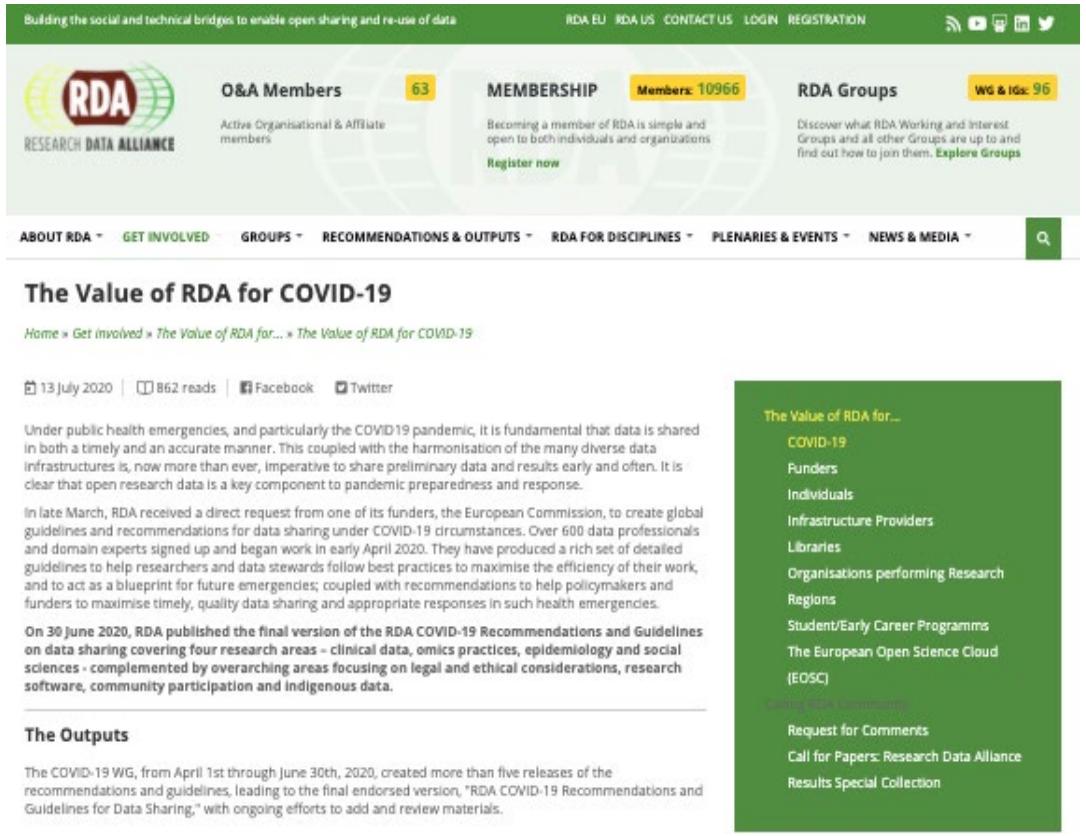
RDA COVID-19 Data Stewardship Wizard

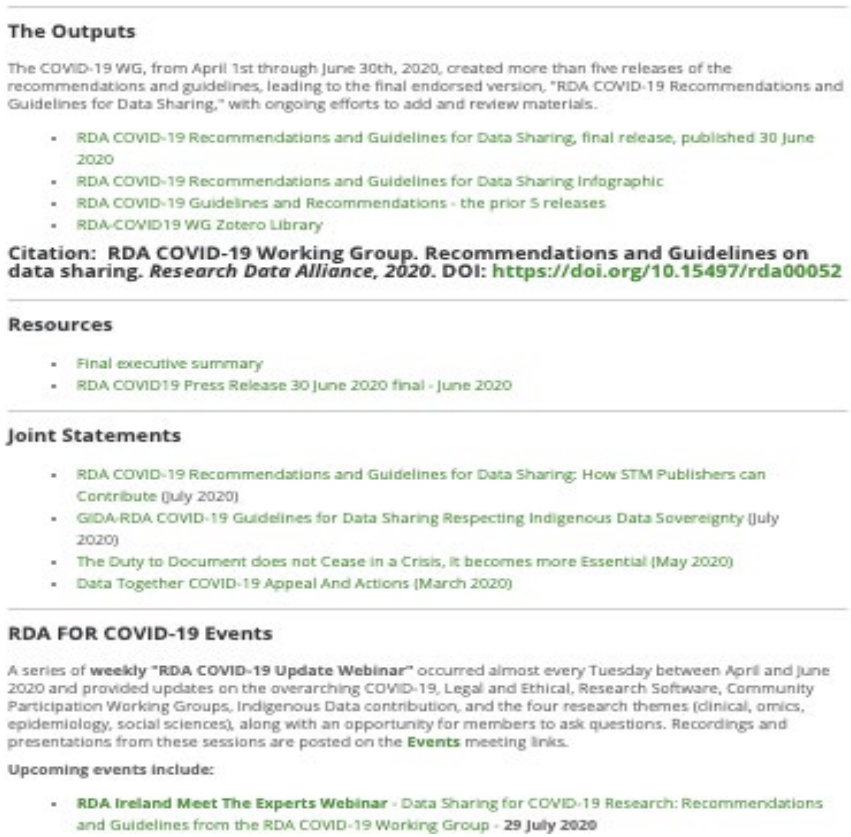


Ongoing Work and Future Steps

- 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
  - COVID-19 Epidemiology WG / Epidemiology WG
- 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.

Value of RDA for COVID-19





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

…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

Natalie Harrower, Director, Digital Repository of Ireland @natalieharrower
Mark Leggott, Executive Director, Research Data Canada @rdc_drc

RDC DRC

Research Data Canada – Données de recherche Canada

Thank you! Questions?