RDA COVID-19 Guidelines and Recommendations on Data Sharing Part 2: Navigational Tools and Other Outputs

Rob Hooft | RDA COVID-19 Omics SG Co-Moderator, Program Manager Data Stewardship, DTL @rwwh
Mark Leggott | RDA COVID-19 WG Co-Chair, Research Data Canada @mleggott

Agenda

- Summary of the *RDA COVID-19 Recommendations and Guidelines on Data Sharing* (Rs&Gs)
- Additional “navigational tools”
  - Infographic
  - Outputs Card
  - Zotero Library
  - Data Stewardship Wizard
  - Mindmap
- Supplementary Outputs/Journal Articles

https://doi.org/10.15497/rda00052

Slide Decks

English: http://bit.ly/RDC_Webinar_Sep-16_EN

French: http://bit.ly/DRC_Webinaire_Sep-16_FR

Simultaneous Interpretation in French

> This session features both English and French language audio channels.

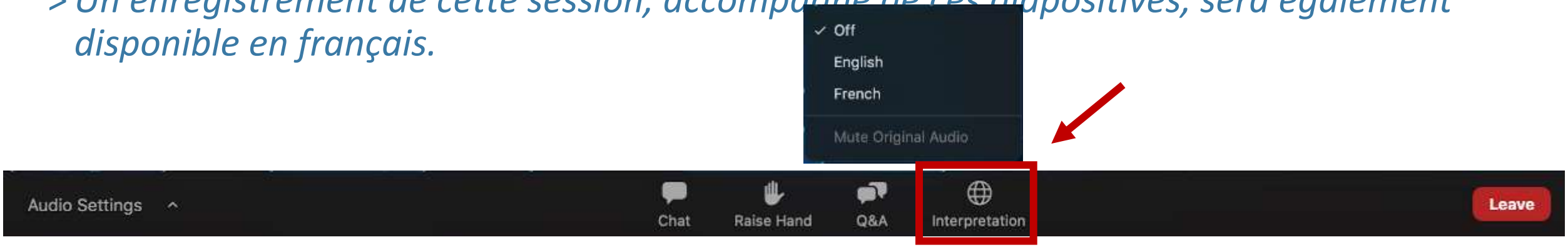
> *Cette session propose des canaux audio en anglais et en français.*

> Select the Interpretation option from your menu to select preferred language.

> *Sélectionnez l'option Interprétation dans votre menu pour sélectionner la langue préférée.*

> A recording of this session, along with these slides, will also be made available in French.

> *Un enregistrement de cette session, accompagné de ces diapositives, sera également disponible en français.*



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.

The Recommendations & Guidelines

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 Co-Moderators
- Structured through a series of teams
  - Co-Chairs, Co-Chairs + Moderators, 8 Themes, Editorial, Visualization, Zotero
- April 1 - 30 June continual sprints, webinars, etc.
  - 6 releases over 3 months
- 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.**

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.

Global Effort to Raise the Bar for Data Sharing

- 117 cross-sectoral signatories to the Wellcome Trust statement in January 2020.
- Agreement by 30 leading publishers on immediate open access to COVID-19 publications and underlying data.

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 generic as well as 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. **Expedite** article and data review processes, **prioritising** and **fast-tracking data** at all stages.
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.

A Collaborative Cross-Disciplinary Effort

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

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

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

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.

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

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)

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.

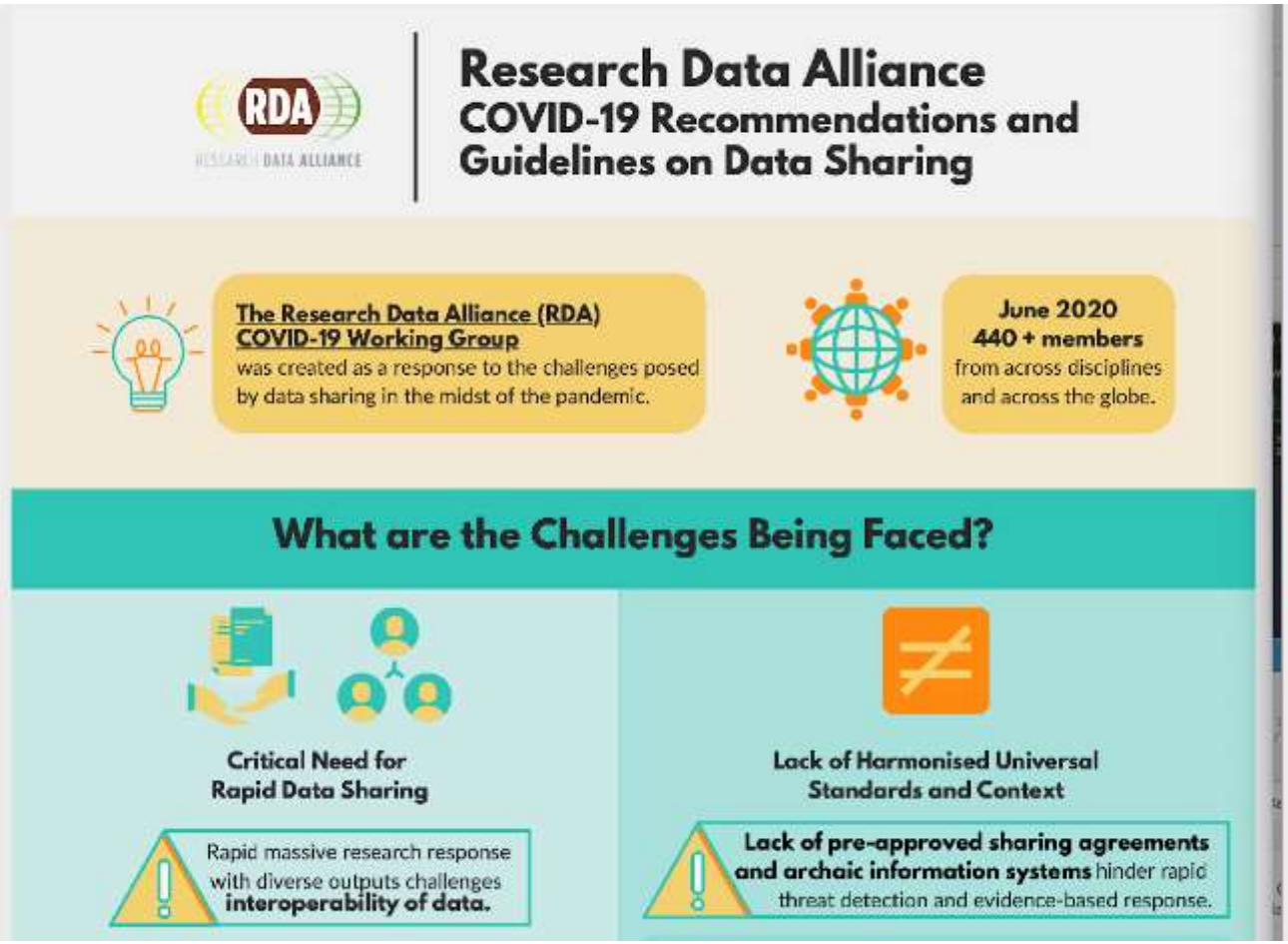
Ideally, consent should be sought for collecting, processing, sharing and publishing data. However, there are other legal bases for processing personal data. Some specific examples from the European General Data Protection Regulation (GDPR, 2016) are described below. Our recommendation would therefore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

Additional Navigational Tools

Infographics



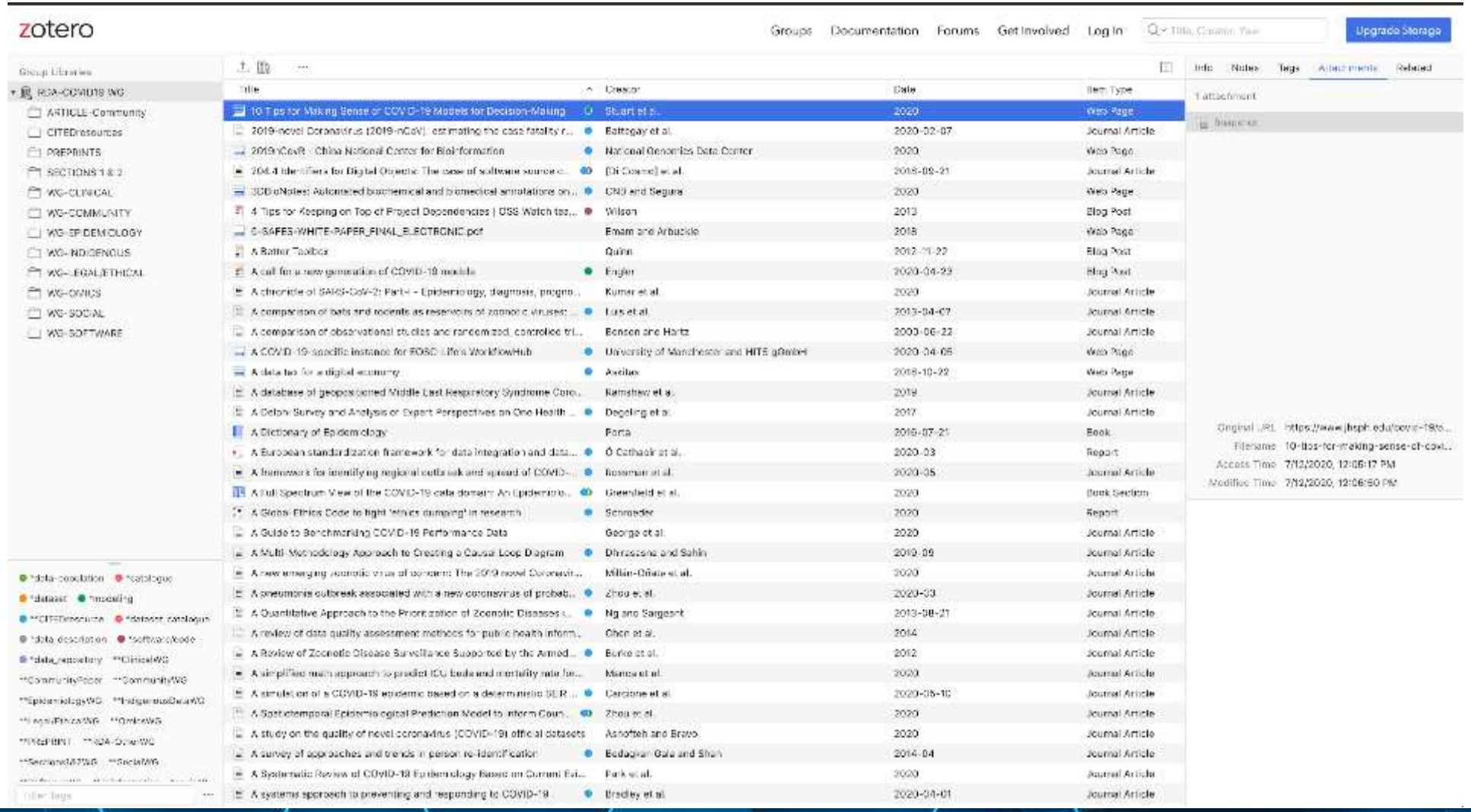
RDA Outputs Card

Zotero Library

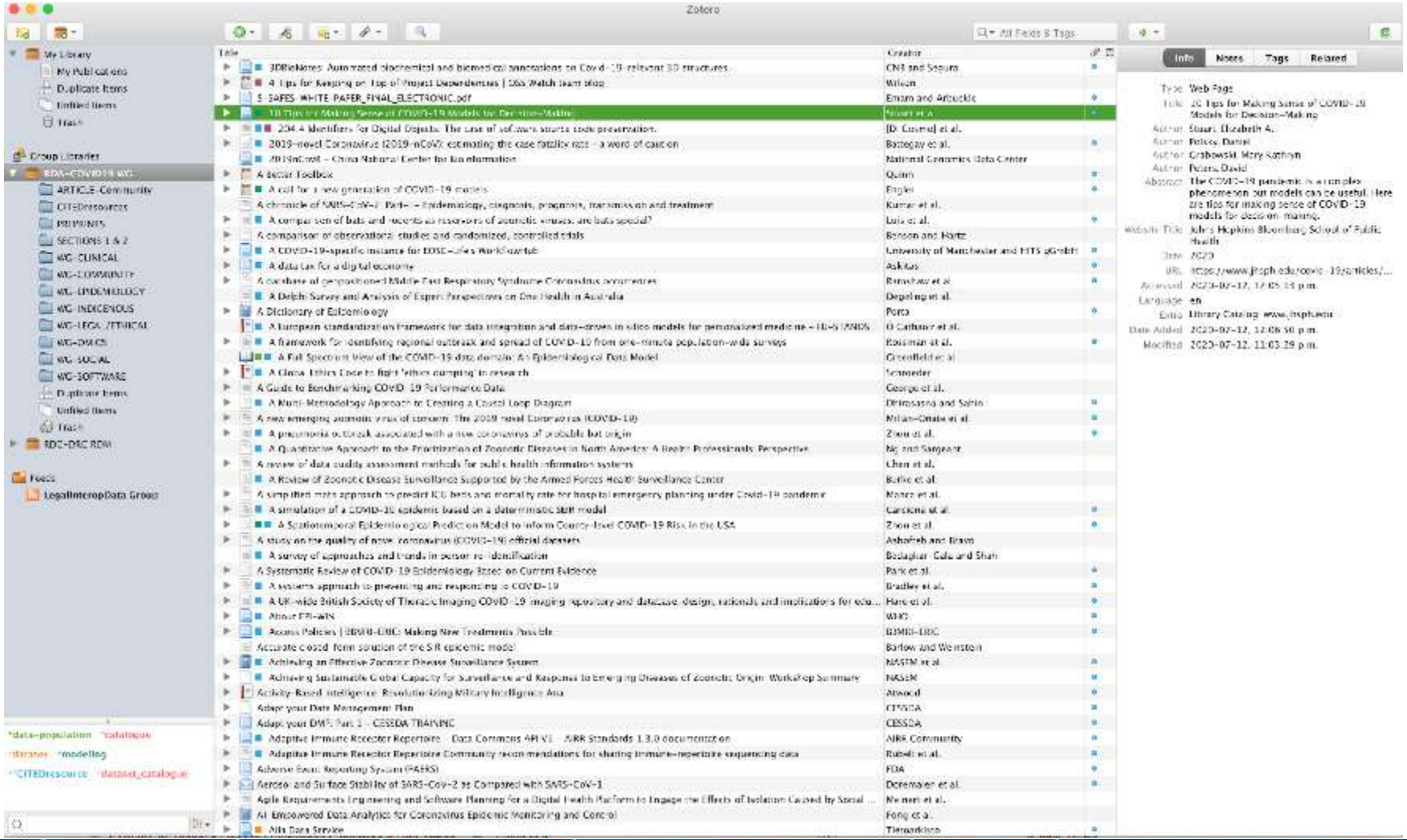
- Over 1,100 bibliographic entries
- All citations in the Rs&Gs are in the Zotero library, facilitating access to the resources, as well as formatting for inclusion in other bibliographies
- Includes separate folders for Sub-Group outputs
- Some additional resources not cited in the Guidelines are also included
- Zotero Library will be maintained/updated going forward
- Citations can be exported/imported to most reference manager tools
- Main contact is Claire Austin

https://www.rd-alliance.org/group/rda-covid19/outcomes/rda-covid19-wg-zotero-library

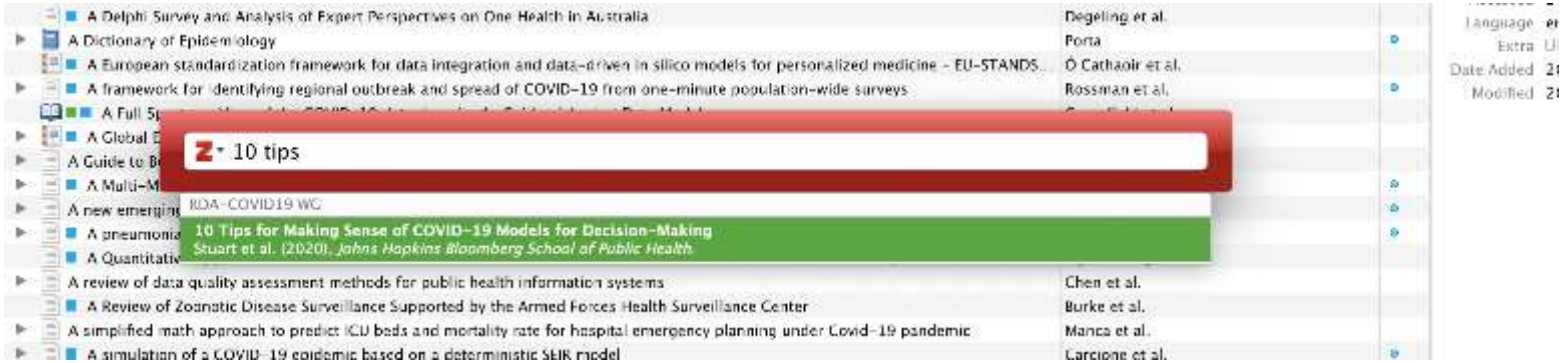
Zotero Web Interface

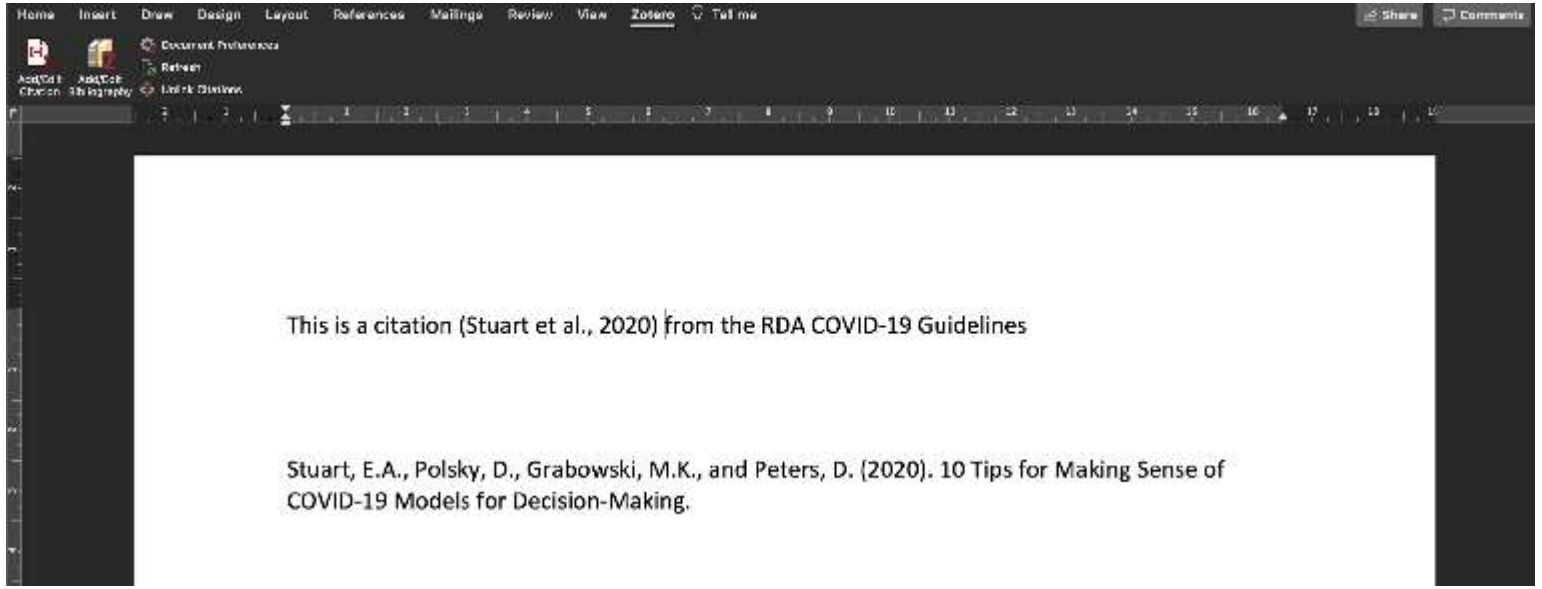


Zotero Desktop Client



Zotero Word Plugin

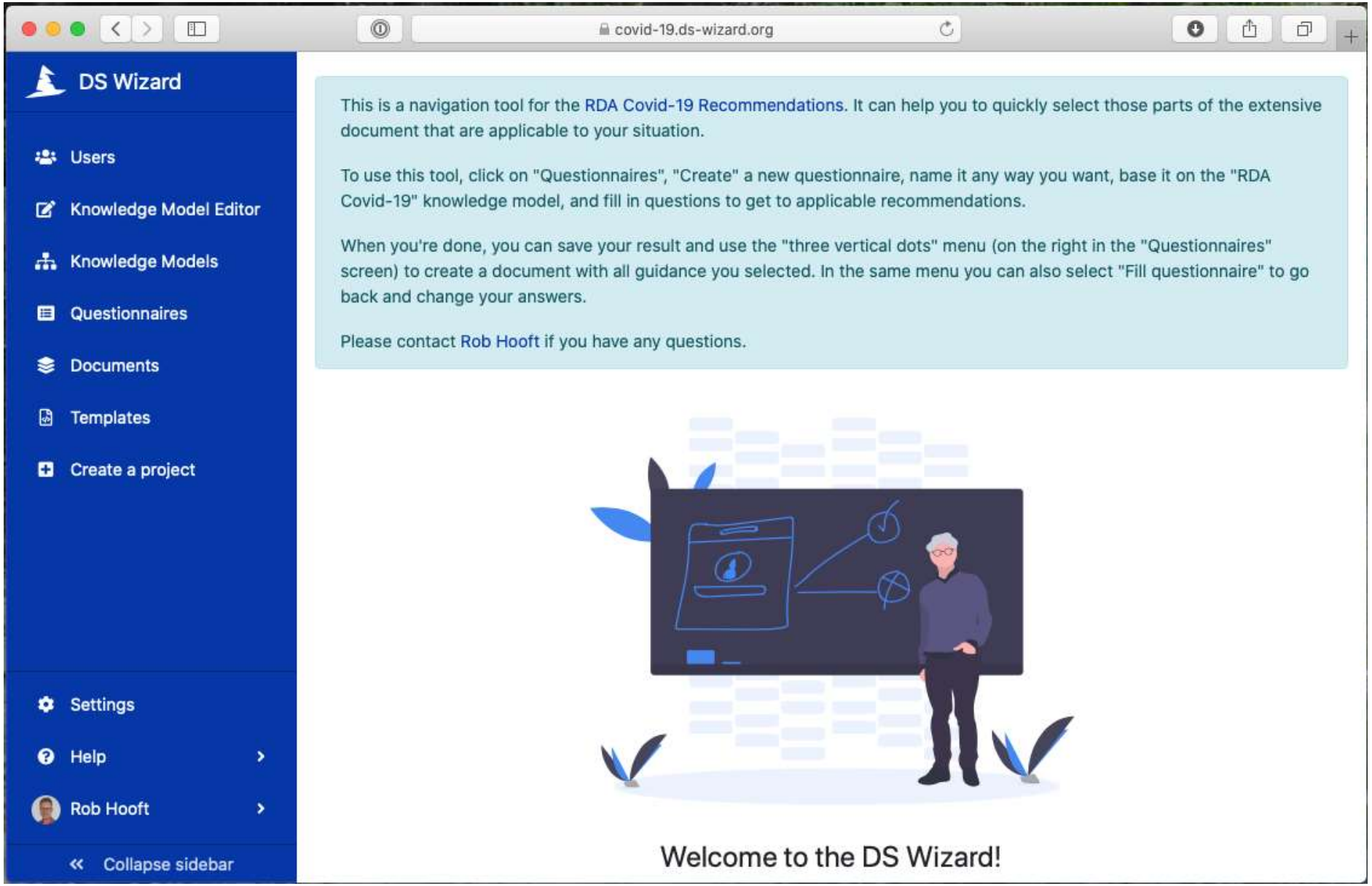


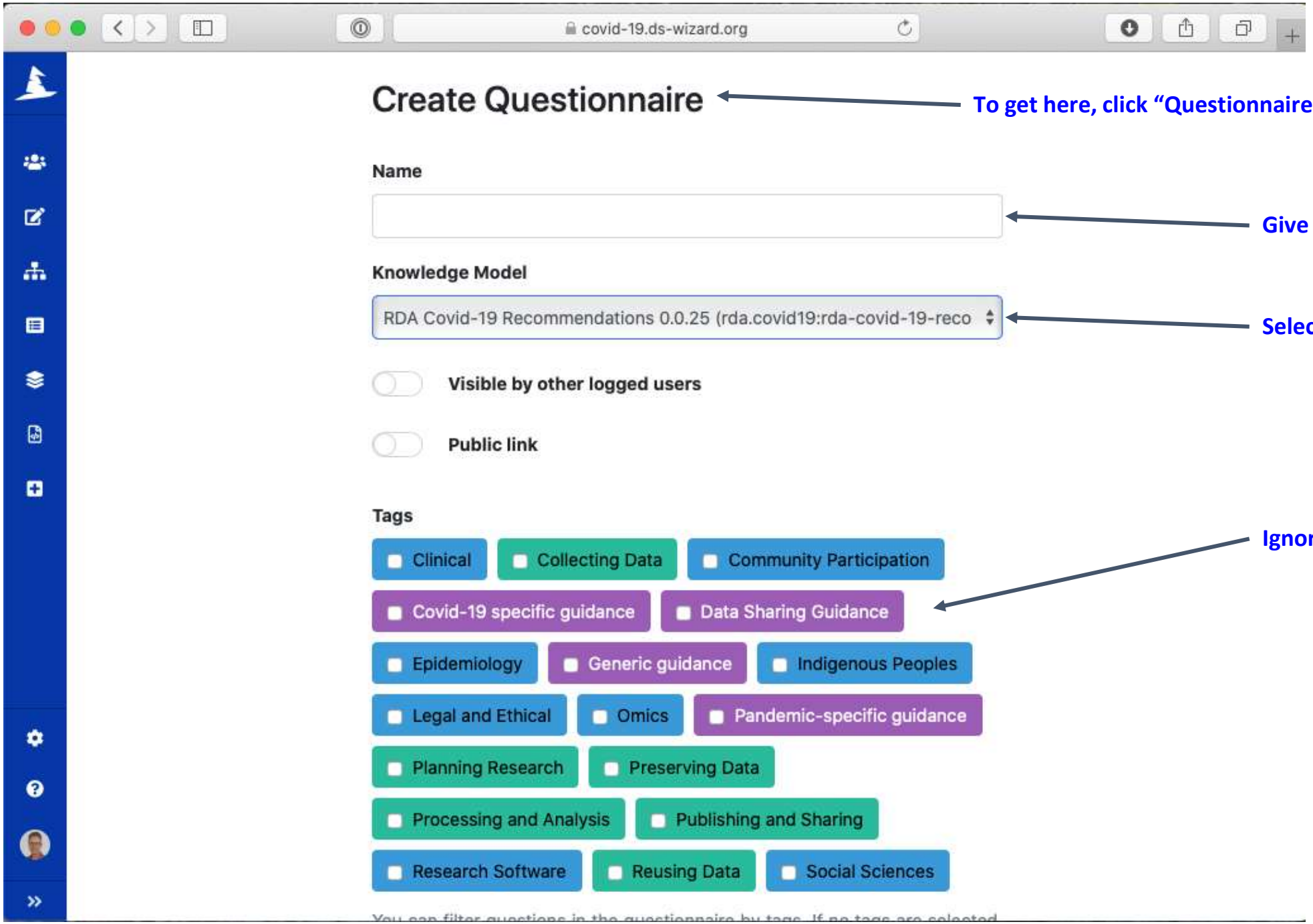


COVID-19 “Wizard”

- Select the parts of the complete 150 page document that are applicable to your own situation
- Quickly select those sections through answering a few questions
- Download a PDF with exactly that content

https://covid-19.ds-wizard.org/


covid-19.ds-wizard.org
DS Wizard
Users
Knowledge Model Editor
Knowledge Models
Questionnaires
Documents
Templates
Create a project
Settings
Help
Rob Hooft
Collapse sidebar
This is a navigation tool for the RDA Covid-19 Recommendations. It can help you to quickly select those parts of the extensive document that are applicable to your situation.
To use this tool, click on "Questionnaires", "Create" a new questionnaire, name it any way you want, base it on the "RDA Covid-19" knowledge model, and fill in questions to get to applicable recommendations.
When you're done, you can save your result and use the "three vertical dots" menu (on the right in the "Questionnaires" screen) to create a document with all guidance you selected. In the same menu you can also select "Fill questionnaire" to go back and change your answers.
Please contact Rob Hooft if you have any questions.
Welcome to the DS Wizard!

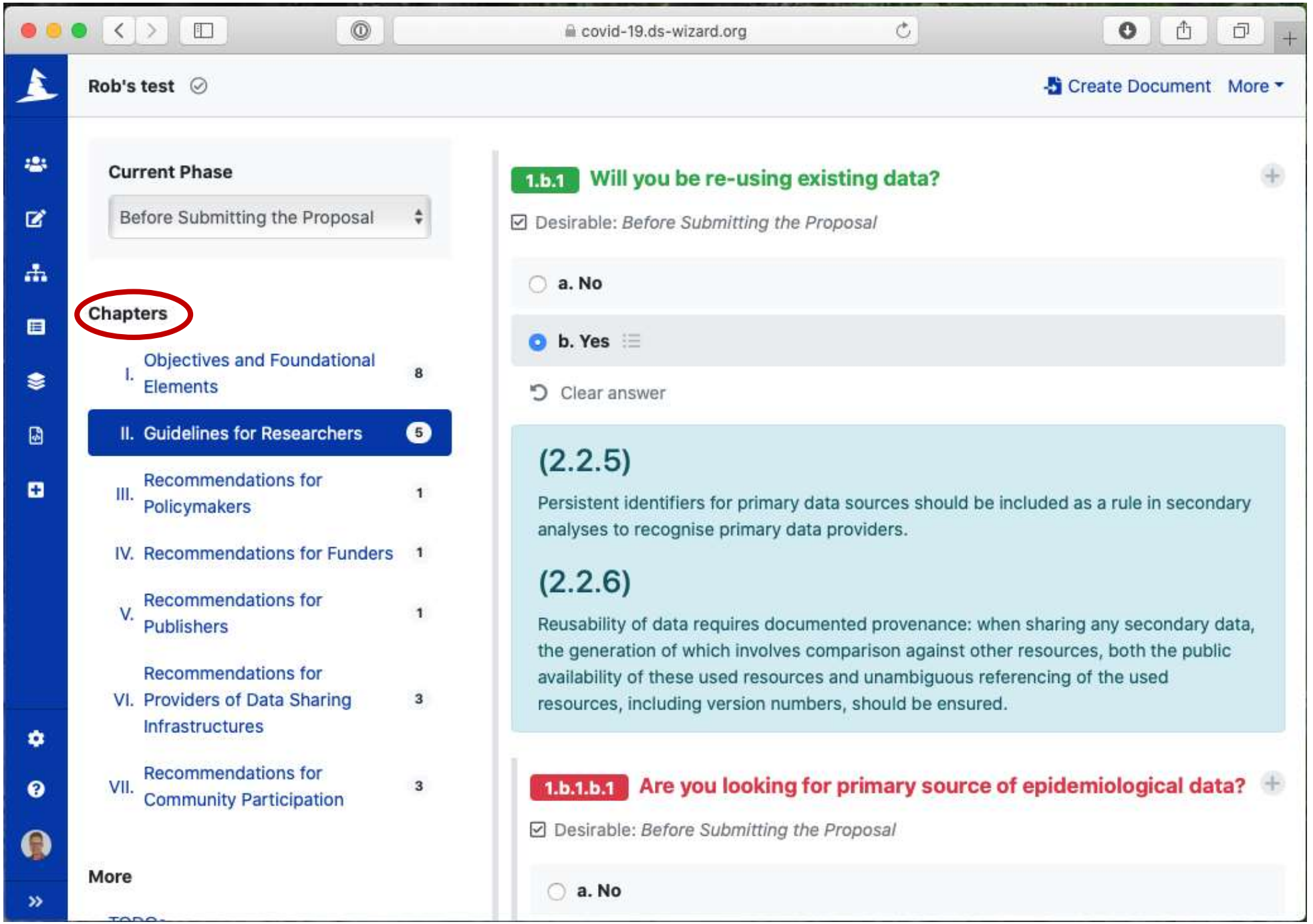


To get here, click “Questionnaires” and then “Create”

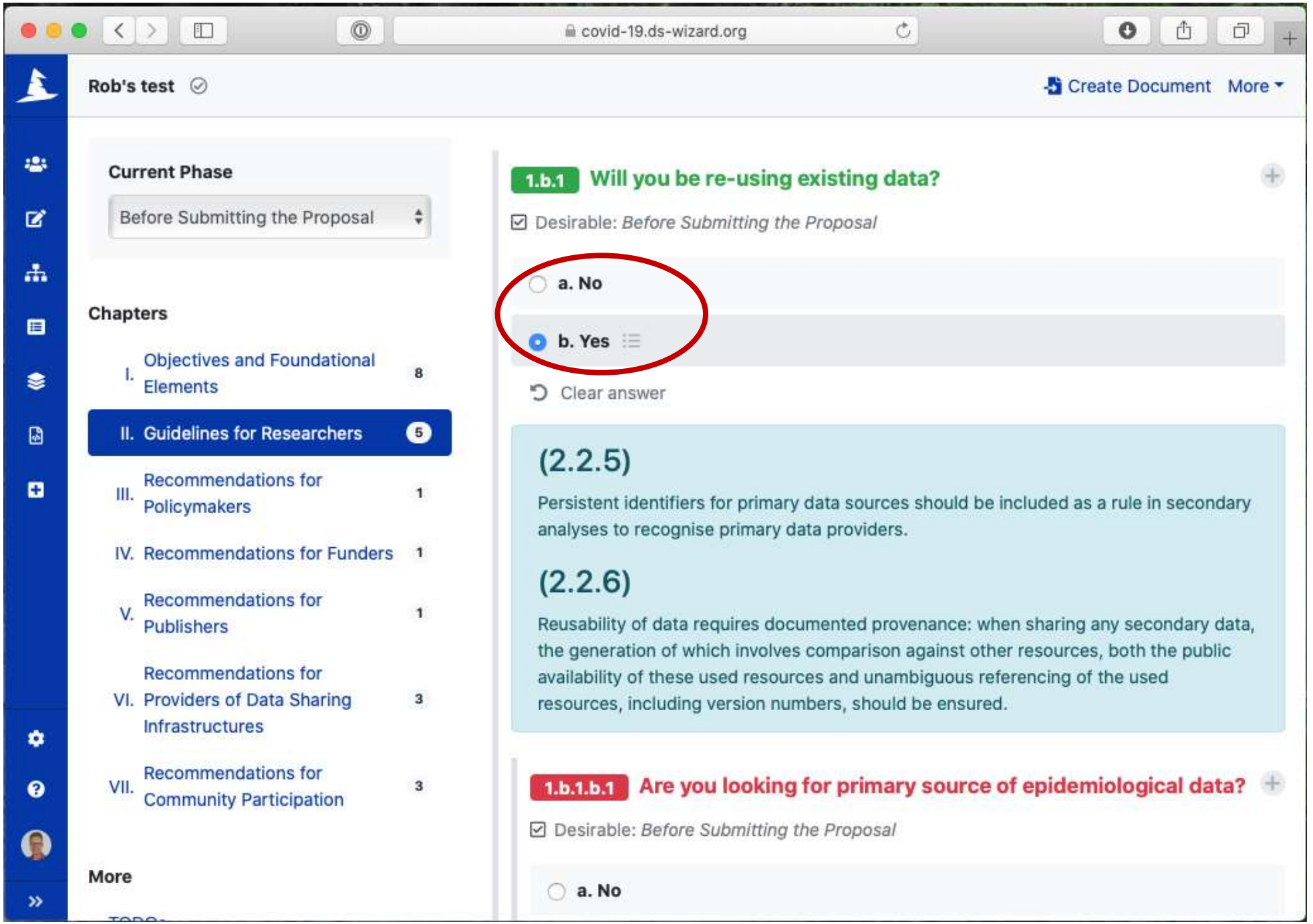
Give it a recognizable name

Select the “Covid recommendations”

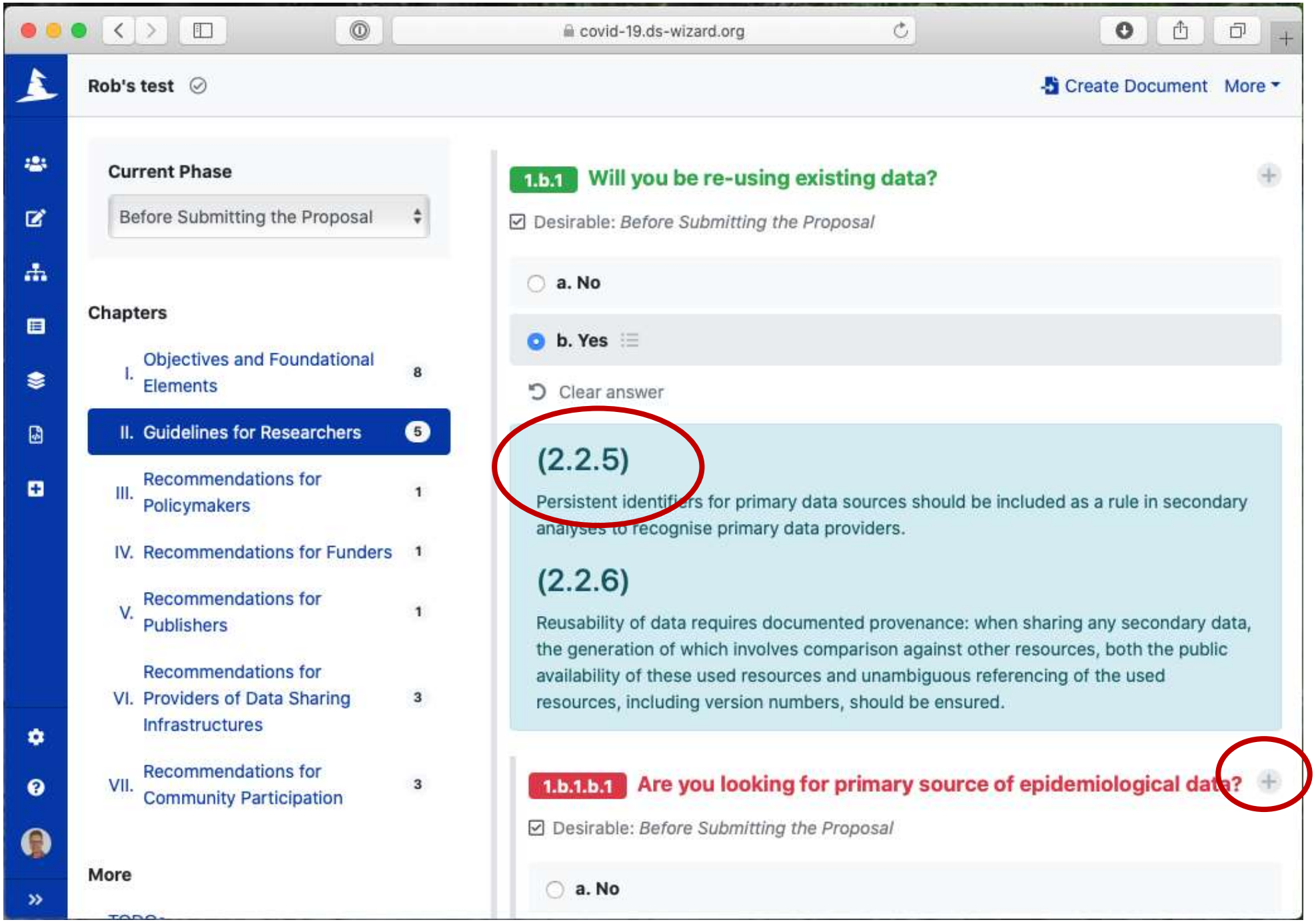
Ignore the rest of the options for now



Chapters for each target audience show how many open questions are left

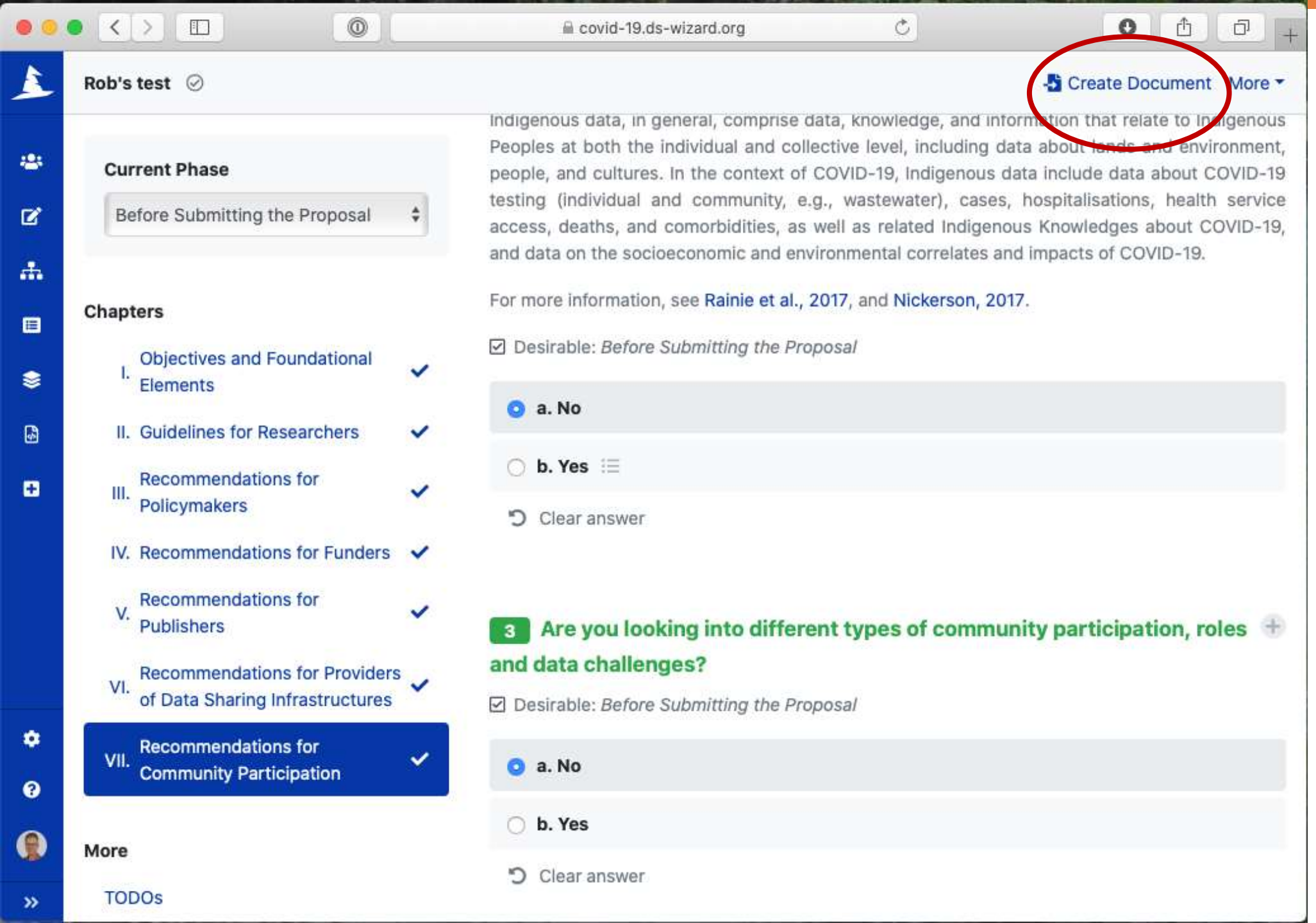


Answer all open questions; new questions may appear when you select an answer.

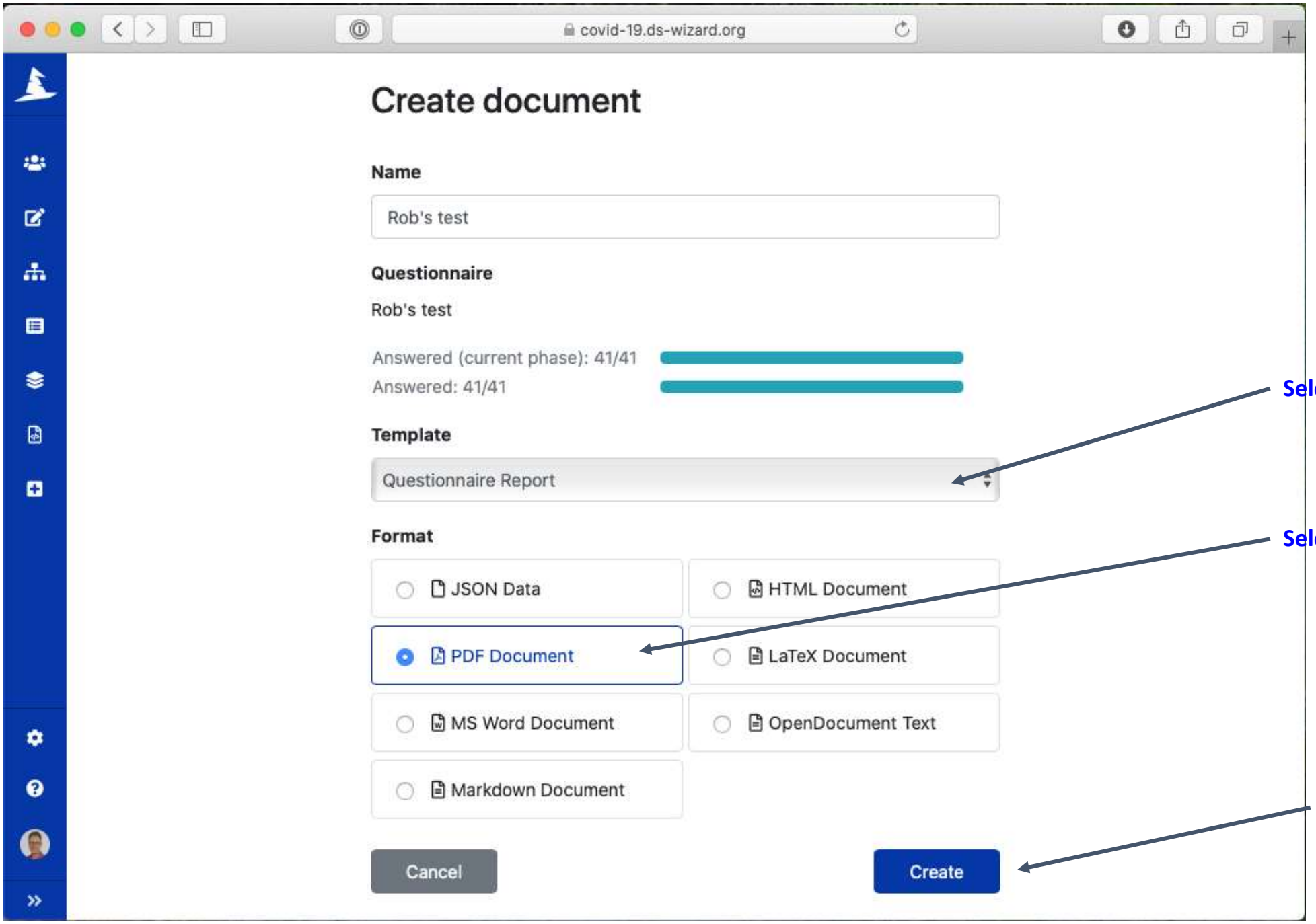


Applicable sections from the guidance will appear in shaded text boxes

Use “+” to give us feedback



When you are done, use "Create document" to assemble all guidance into a PDF tailored for you.

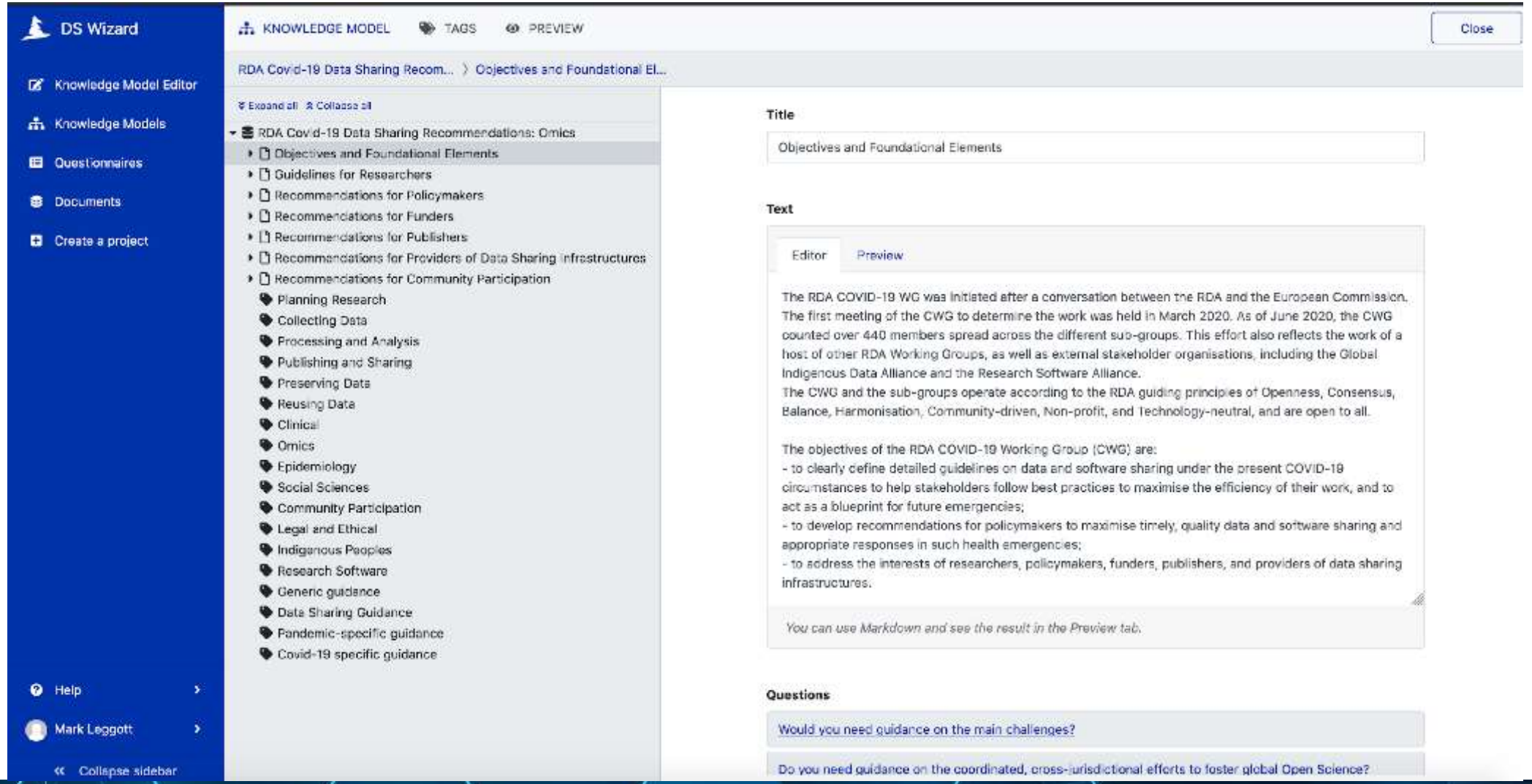


Select the “Questionnaire report”

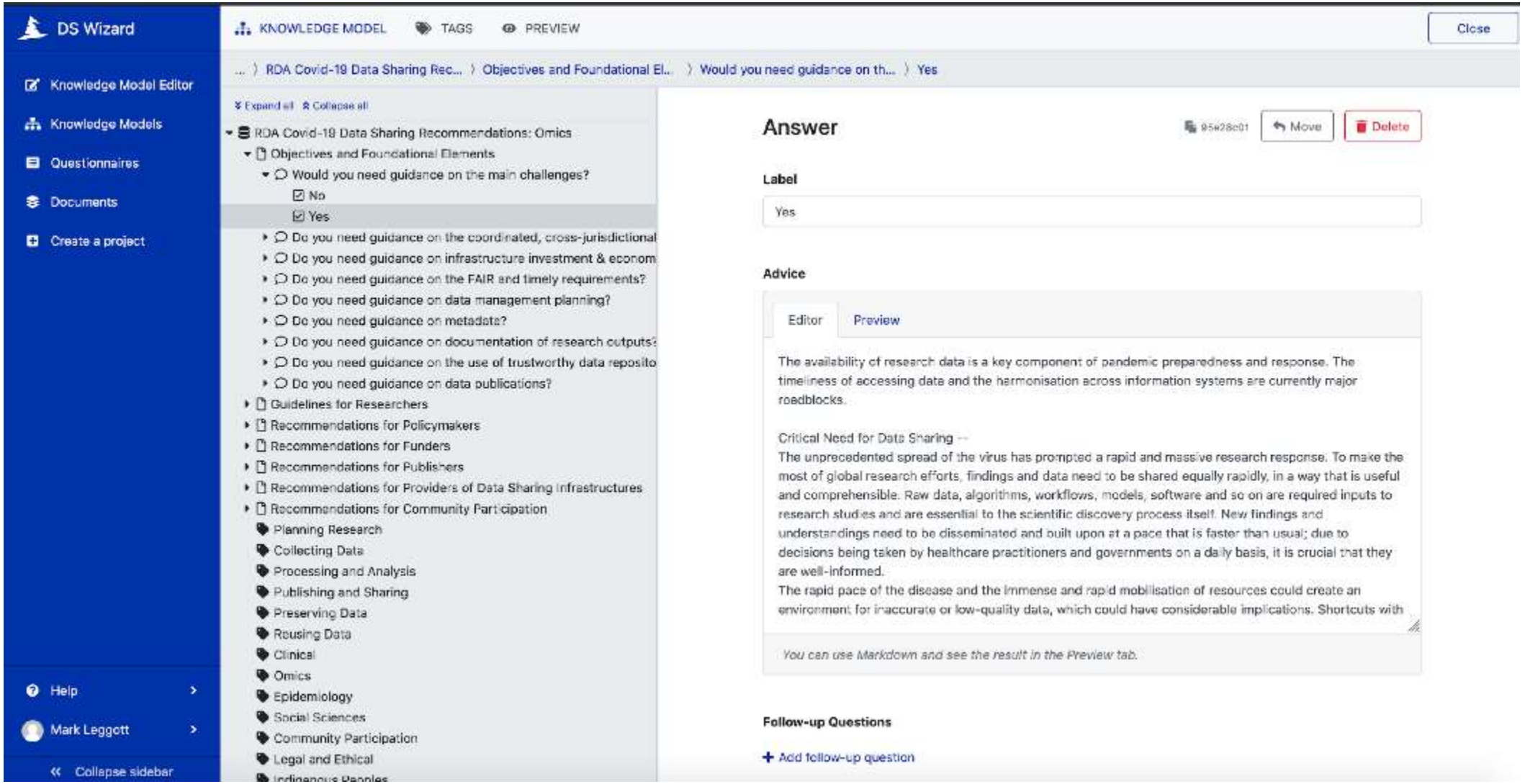
Select “PDF Document”

Click “Create”

DS Wizard Editing Mode - Structure



DS Wizard Editing Mode – Questions & Text



About the DS Wizard

- DS Wizard was originally created to help making data management choices
- Started from a 600 question mind map
- An “Expert system” to **help researchers** find the right information
- Not primarily “Data Management Plan” for the **funder**!
- Suitable to update data management during a whole project
- Supporting (RDA standards for) machine actionable Data Management Plans

- Very suitable to make local derivatives pointing to local resources

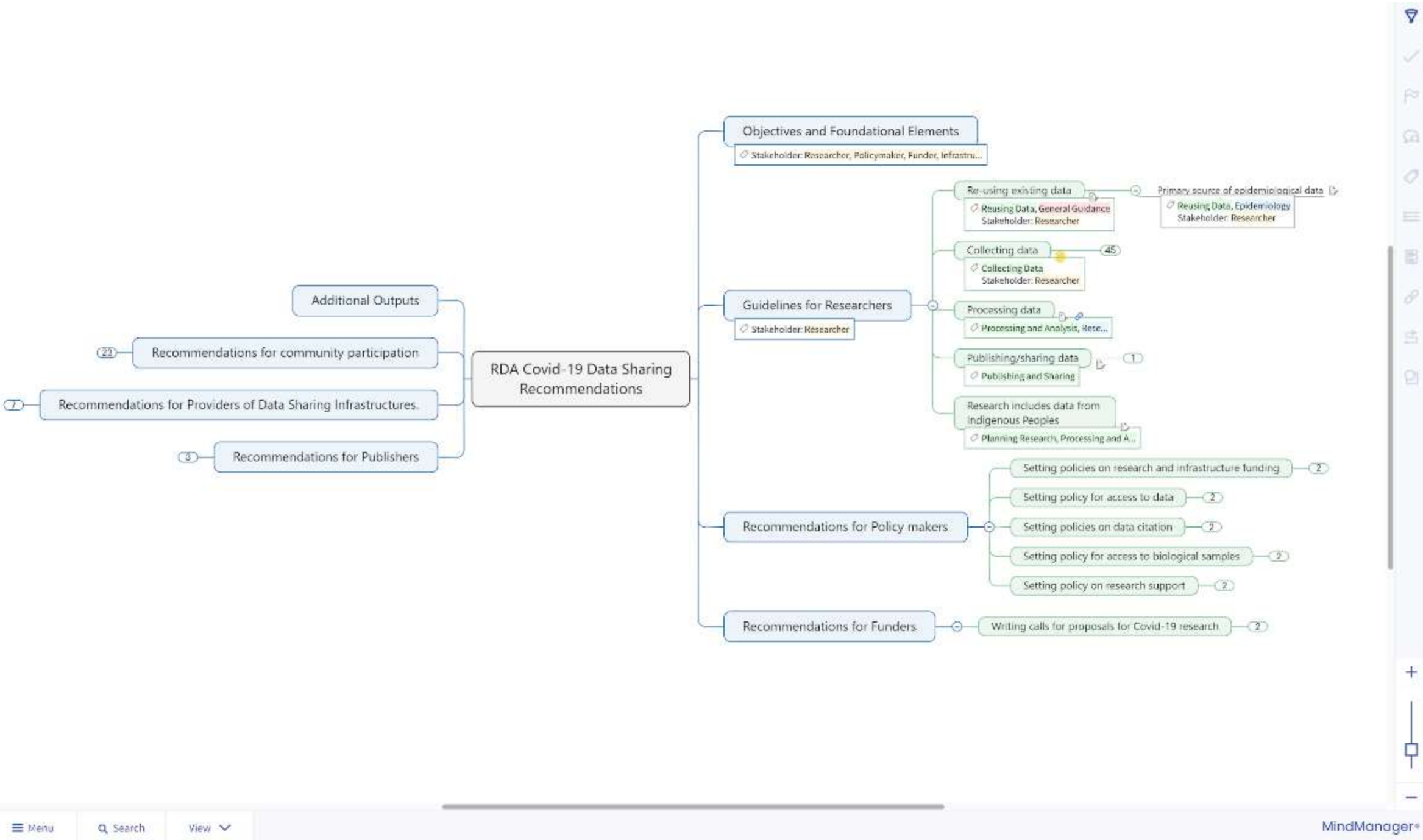
- Also available to host your own expert systems; open source

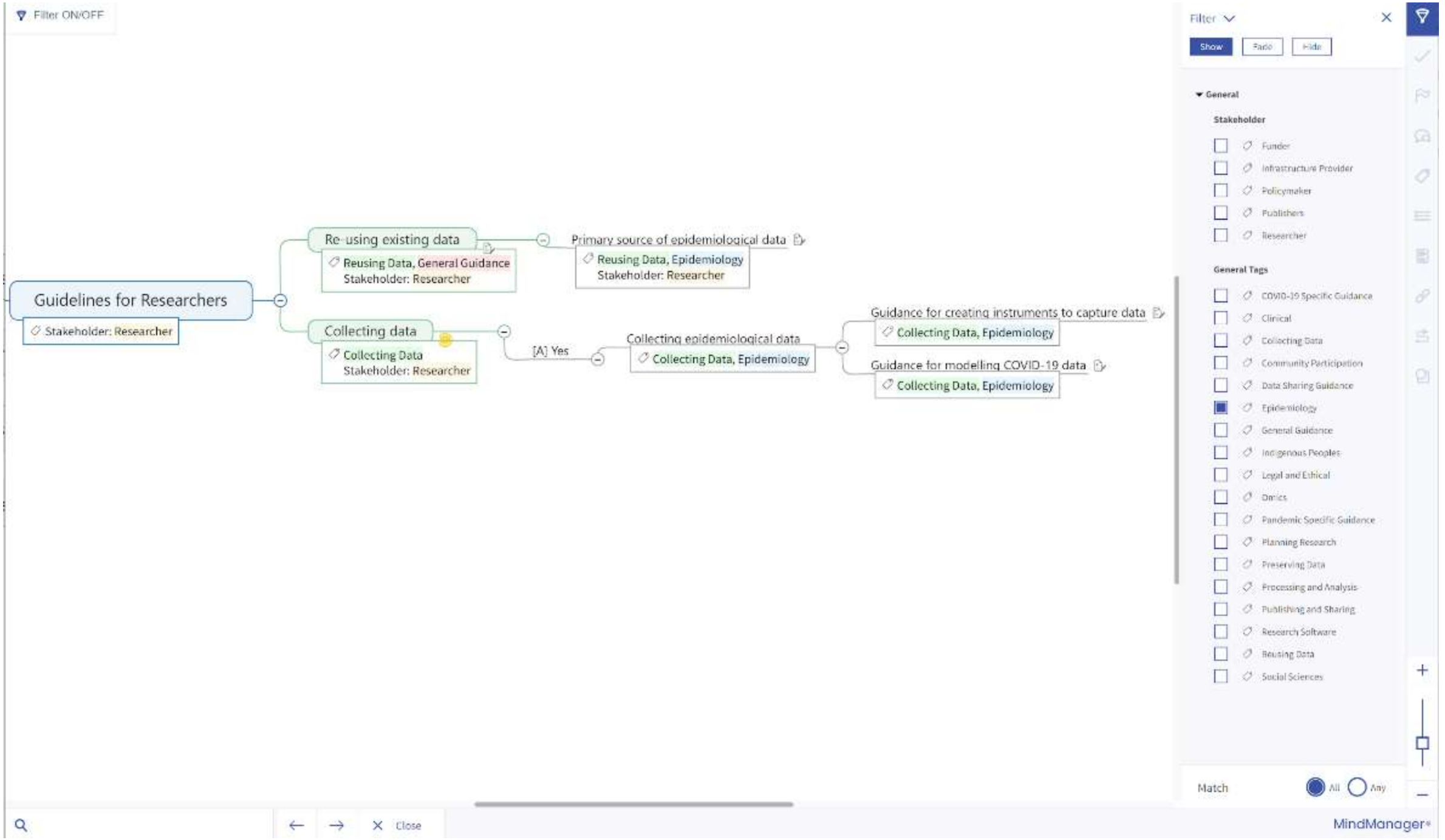
DSW was created in cooperation of

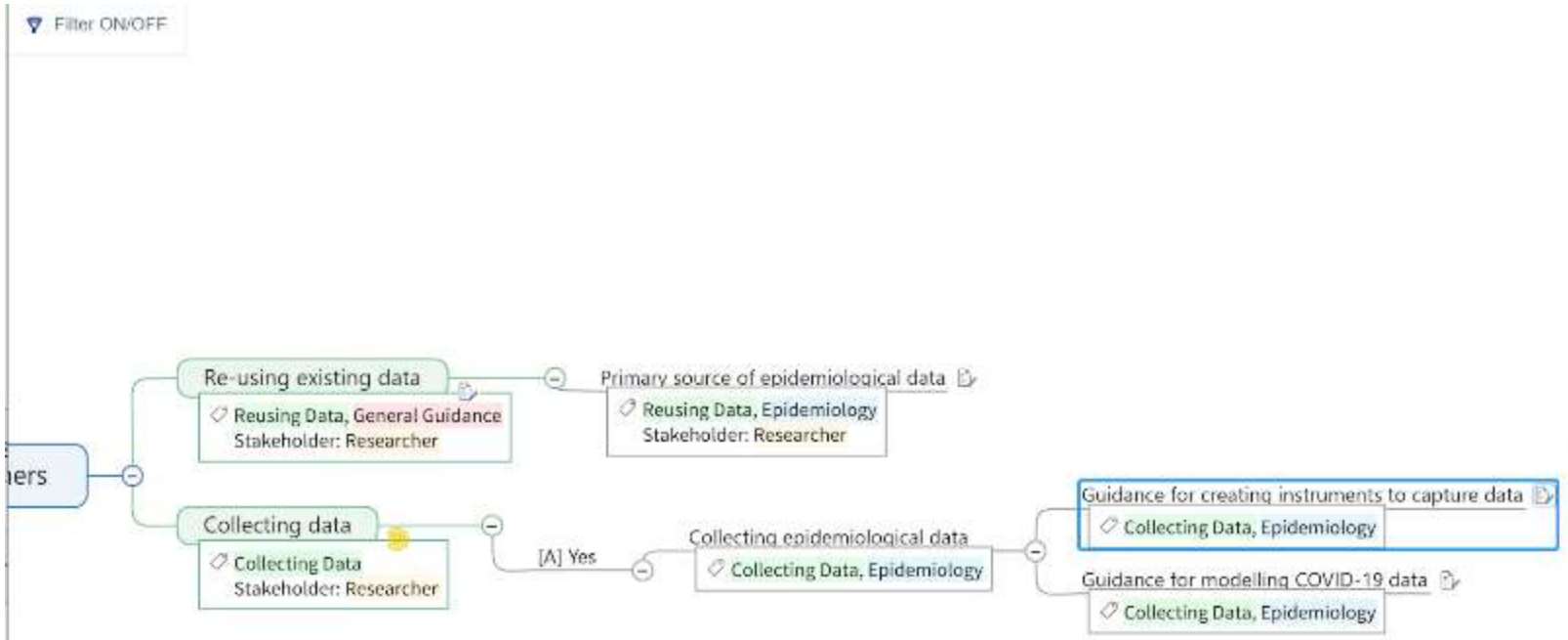
Mindmap

- A single view of the Rs&Gs using a mindmap approach
- Includes the full text of the Rs&Gs as notes attached to nodes that correspond to Questions from the DS Wizard version
- Intended to provide a high-level view, and the ability to drill down
- Can filter nodes by tags that include
  - stakeholder group
  - section ans topic
  - fulltext search of all nodes and notes
  - web-accessible view

https://bit.ly/RDA-COVID19-Guidelines


RDA Covid-19 Data Sharing Recommendations
Objectives and Foundational Elements
Stakeholder: Researcher, Policymaker, Funder, Infrastru...
Guidelines for Researchers
Stakeholder: Researcher
Re-using existing data
Reusing Data, General Guidance
Stakeholder: Researcher
Primary source of epidemiological data
Reusing Data, Epidemiology
Stakeholder: Researcher
Collecting data
45
Collecting Data
Stakeholder: Researcher
Processing data
Processing and Analysis, Rese...
Publishing/sharing data
1
Publishing and Sharing
Research includes data from Indigenous Peoples
Planning Research, Processing and A...
Recommendations for Policy makers
Setting policies on research and infrastructure funding
2
Setting policy for access to data
2
Setting policies on data citation
2
Setting policy for access to biological samples
2
Setting policy on research support
2
Recommendations for Funders
Writing calls for proposals for Covid-19 research
2
Additional Outputs
Recommendations for community participation
23
Recommendations for Providers of Data Sharing Infrastructures
7
Recommendations for Publishers
3
Menu
Search
View
MindManager


Filter ON/OFF
Guidelines for Researchers
Stakeholder: Researcher
Re-using existing data
Reusing Data, General Guidance
Stakeholder: Researcher
Primary source of epidemiological data
Reusing Data, Epidemiology
Stakeholder: Researcher
Collecting data
Collecting Data
Stakeholder: Researcher
[A] Yes
Collecting epidemiological data
Collecting Data, Epidemiology
Guidance for creating instruments to capture data
Collecting Data, Epidemiology
Guidance for modelling COVID-19 data
Collecting Data, Epidemiology
Filter
Show
Fade
Hide
General
Stakeholder
Funder
Infrastructure Provider
Policymaker
Publishers
Researcher
General Tags
COVID-19 Specific Guidance
Clinical
Collecting Data
Community Participation
Data Sharing Guidance
Epidemiology
General Guidance
Indigenous Peoples
Legal and Ethical
Omics
Pandemic Specific Guidance
Planning Research
Preserving Data
Processing and Analysis
Publishing and Sharing
Research Software
Reusing Data
Social Sciences
Match
All
Any
Close
MindManager



Notes

5.4.2Interoperable COVID-19 Epidemiological Surveillance: Clinical and Population-based Instruments

International efforts are currently underway to create COVID-19 instruments/ questionnaires (Tables 3-4). These COVID-specific tools are concentrated at person-level for clinic/hospital surveillance (e.g. Case Report Forms-CRFs), or community surveillance (e.g. questionnaire for general population), and do not necessarily collect the same data. Adherence of new studies to already introduced instruments will strongly enhance the comparability of results.

Table 3 - Questionnaire instruments: Reference

1. CLINICAL
    1. Australia: NSW Case questionnaire
    2. Austria: EMS
    3. Europe: TESSy
    4. Germany: Covid-19 research dataset
    5. Uganda: Perinatal COVID-19 Uganda
    6. US: Human Infection with 2019 Novel CoronavirusPerson Under Investigation (PUI) and Case Report Form
    7. Worldwide: Global COVID-19: clinical platform: novel coronavius (COVID-19): rapid version] - NEED CORRECT URL
2. POPULATION BASED
    1. Brazil: Brazil Prevalence of Infection Survey
    2. Europe: Questionnaire by WHO Europe
    3. Germany: GESIS Panel Special Survey on the Coronavirus SARS-CoV-2 Outbreak in Germany
    4. Germany: NAKO COVID-19 Survey tool
    5. Israel: One-minute population wide survey
    6. LMICs: LMIC Covid Questionnaire
    7. South Africa: South African Population Research Infrastructure (SAPRIN) COVID-19 Screening Form
    8. South Asian Countries: National Institute for Health Research (NIHR) Global Health Research Unit
    9. UK: UK COVID-19 Questionnaire
    10. Worldwide (WHO): Population-based age-stratified sero-epidemiological investigation protocol for COVID-19 virus infection

Table 4 - Questionnaire instruments: Resources

1. NIH: Public Health Emergency and Disaster Research Response (DR2)
2. NIH: COVID-19OBSSR Research Tools
3. PhenX: PhenX COVID-19 Toolkit

Some of the questionnaire initiatives shown in Tables 3 and 4 are currently feeding into the construction of a COVID-19 demographic and epidemiological surveillance question bank that can be used to form locality specific surveys with both common and distinct questions by domains and cohorts (Wellcome Trust). Some, such as the UK COVID-19

Ongoing Work and Future Steps – Outputs and Articles

> Supplementary Outputs

- A number of the COVID-19 Sub-Groups have published more detailed documents that build on the primary Rs&Gs document
  - Data Sharing in Epidemiology
- Other Sub-Group documents are available in Google folders

> Journal Articles and Endorsements

- A number of articles in preparation by COVID-19 WG Members
  - Summary of the findings of the Rs&Gs
  - Description of the process used to create the Rs&Gs
  - Community Participation article
- Summaries of the Rs&Gs in various journals (e.g. Cell Patterns, HealthCare IT News, etc.)

Ongoing Work and Future Steps – Endorsements/Statements

- Stakeholder support happening in various ways
  - Adoption and implementation of the recommendations and guidelines
  - Promotion by policymakers, funders and publishers
  - Use and promotion by researchers
- Statements
  - STM Publishers
  - Global Indigenous Data Alliance (GIDA)
  - Duty to Document Statement

Ongoing Work and Future Steps – Events/Survey

- Webinars
  - RDM organizations
    - Research Data Alliance and National Nodes (e.g. Ireland)
    - Research Data Canada
  - Other Organizations
    - FAPESP (Brazil)
    - European Open Science Forum
    - Scottish Council on Archives
- Impact Survey
  - Survey to measure use and impact of the Rs&Gs is under development
  - Will be distributed broadly when completed

Ongoing Work and Future Steps – RDA Plenary Sessions

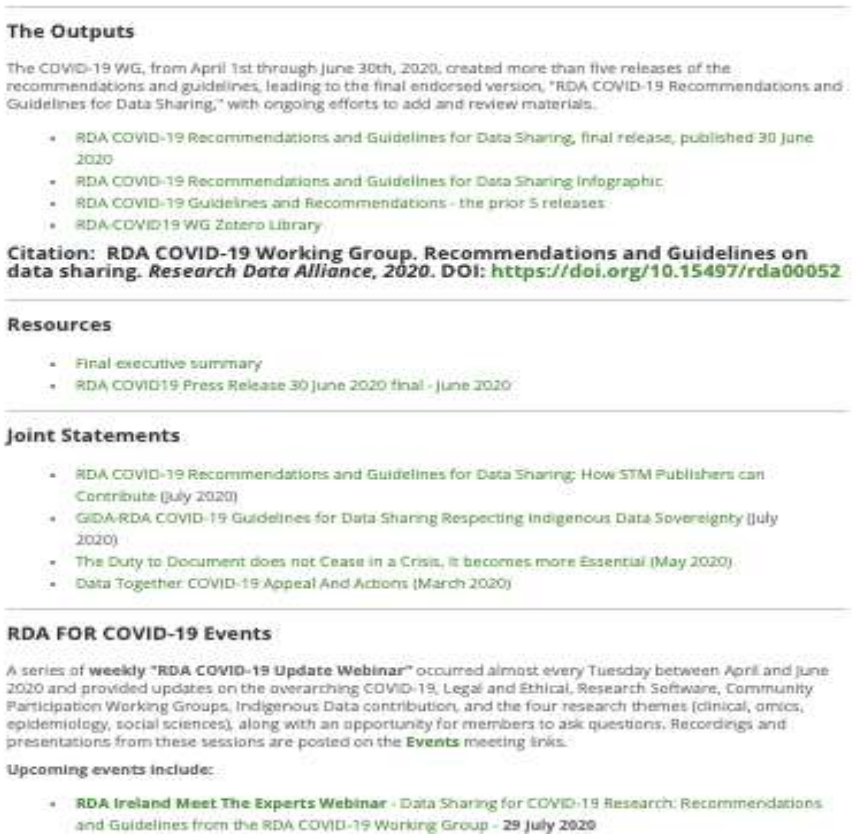
> VP16 (Virtual Plenary P16) was going to be in Costa Rica, will now be virtual

> Various sessions will discuss elements of COVID-19 and/r infectious disease

- 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

https://www.rd-alliance.org/plenaries/rda-16th-plenary-meeting-costa-rica

Value of RDA for COVID-19



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

RDA as a Community Platform

- The COVID-19 Rs&Gs demonstrate the core RDA value
  - Ability to gather a knowledgeable grassroots group
- Easy to get engaged!
  - Membership is free
  - Engagement opportunities are rich and diverse
- 100+ Interest Groups and Working Groups
- New Communities of Practice group type
- Virtual Plenary 16 a good chance to engage

RDA's guiding principles:

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

…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

Robyn Nicholson, Intern, Research Data Canada, @robyncecelia
Rob Hooft, Program Manager Data Stewardship, DTL, @rwwh
Mark Leggott, Executive Director, Research Data Canada @rdc_drc

Visualisation Team Contributors

Claire C Austin (0000-0001-9138-5986)

Timea Biro (0000-0002-8900-8978)

Anne Cambon-Thomsen (0000-0001-8793-3644)

Piotr Wojciech Dabrowski (0000-0003-4893-805X)

Kheeran Dharmawardena (0000-0002-4292-7475)

Leyla Jael Garcia Castro (0000-0003-3986-0510)

Hilary Hanahoe (0000-0002-0328-3419)

Rob Hooft (0000-0001-6825-9439)

Dawei Lin (0000-0002-5506-0030)

Mark Leggott (0000-0003-1392-7799)

Robyn Nicholson (0000-0002-1621-2425)

Christian Ohmann (0000-0002-5919-1003)

Brian Pickering (0000-0002-6815-2938)

Amy Pienta (0000-0003-1174-6118)

Mary Uhlmansiek (0000-0002-7949-2057)

Meghan Underwood (0000-0001-6538-9617)

RDC DRC

Research Data Canada – Données de recherche Canada

Thank you! Questions?