TRUST Principles Mini Symposium

July 7th, 2020 | 14:00 - 17:00 UTC

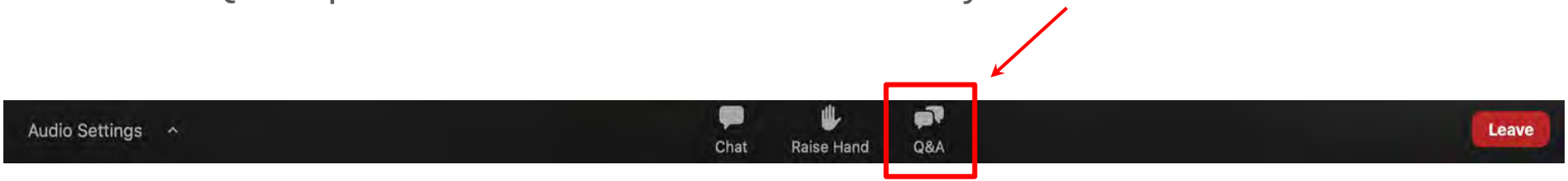
RDA
RESEARCH DATA ALLIANCE

WORLD DATA SYSTEM

RDC DRC
Research Data Canada – Données de Recherche Canada

Questions & Answers

- Please use the Q&A option to ask questions of the presenter(s). Questions will be addressed at the end of each session when possible, and also at the end of the Symposium.
- 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.

TRUST: Transparency, Responsibility, User focus, Sustainability, Technology

RDA Research Data Alliance | World Data System

RDC DRC Research Data Canada – Données de Recherche Canada

Schedule for Today

14:00-14:30 UTC	**Introduction: What are the TRUST Principles?** *Presented by Dawei Lin*
14:30-14:45 UTC	**Cautionary Tales: The Need for Digital Preservation Best Practices** *Presented by Karen Payne*
14:45-15:00 UTC	**TRUST and FAIR: Complementarity of the Principles** *Presented by Ingrid Dillo*
15:00-15:30 UTC	**Implementation Pathway: Endorsing TRUST, Core Trust Seal, and ISO Intersections** *Presented by Wim Hugo and David Giaretta*
15:30-15:40 UTC	***Break***
15:40-16:00 UTC	**Repository Manager Story: A Use Case** *Presented by John Westbrook*
16:00-16:15 UTC	**Funder Story: A Use Case** *Presented by Mark Leggott*
16:15-16:45 UTC	**Stakeholder Panel** ● *Moderator: Mustapha Mokrane* ● *TRUST author: Robert Downs* ● *Funder: Mark Leggott* ● *Research Community: Shelley Stall* ● *Repository Manager: John Westbrook* ● *Publisher: Varsha Khodiyar*
16:45-17:00 UTC	**Wrap-up and Ways Forward**

Slides

- Follow along with this slide deck here:

https://bit.ly/TRUSTSymposium

What are the TRUST Principles?

Dr. Dawei Lin, Ph.D.

Division of Allergy, Immunology, and Transplantation, NIAID, NIH
dawei.lin@nih.gov

TRUST Principles Mini Symposium, July 7th , 2020

What Are the Principles?

"Principles are smart ways for handling things that happen over and over again in similar situations"

Ray Dalio, Investor and the author of the book "Principles"

A Community Defines TRUST Principles

- Started with the CoreTrustSeal Board discussions
- Presented the TRUST White Paper at two RDA plenaries
- Received 200+ public comments
- 19 co-authors represent
  - Four continents and eight countries
  - Diverse stakeholders
    - Funders, Publishers, Certification Standards, Librarians, and Data Preservation Specialists
  - Multiple research areas
    - Social sciences, Geosciences, and Biomedical Sciences
  - Hundreds of years of experiences in digital repositories
- The 1st Trustworthy workshop at NIH

Positive Responses from the Community

About this Attention Score

In the top 5% of all research outputs scored by Altmetric

Among the highest-scoring outputs from this source (#44 of 1,252)

High Attention Score compared to outputs of the same age (97th percentile)

LESS...

Mentioned by

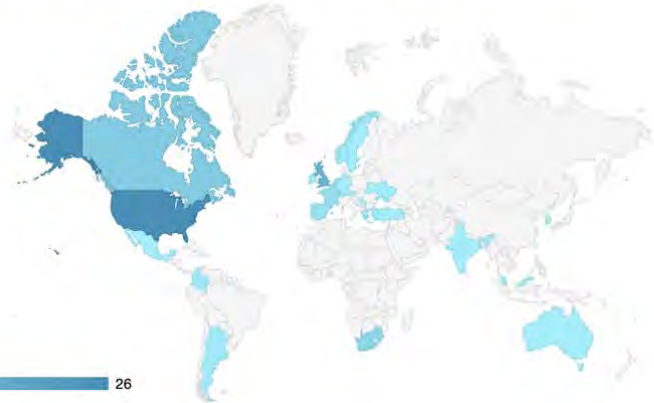
- **2** blogs
- **1** policy source
- **142** tweeters
- **5** Facebook pages
- **1** Redditor

Citations

- **2** Dimensions

TWITTER DEMOGRAPHICS | ATTENTION SCORE IN CONTEXT

The data shown below were collected from the profiles of **142** tweeters who shared this research output. Click here to find out more about how the information was compiled.



Geographical breakdown

Country	Count	As %
United States	26	18%
United Kingdom	16	11%
Canada	10	7%
South Africa	7	5%
France	5	4%
Spain	4	3%
Belgium	3	2%
Netherlands	3	2%
Colombia	2	1%
Other	21	15%
Unknown	45	32%

Demographic breakdown

Type	Count	As %
Members of the public	116	82%
Scientists	18	13%
Science communicators (journalists, bloggers, editors)	6	4%
Practitioners (doctors, other healthcare professionals)	2	1%

As of July 7, 2020 from Altmetric Attention Score

An RDA Community Effort

Organization Endorsements as of July 6, 2020

1. 4TU.ResearchData
2. The Arctic Data Center
3. Center for International Earth Science Information Network (CIESIN), The Earth Institute, Columbia University, New York, USA.
4. Comisión de Investigaciones Cientificas
5. CoreTrustSeal
6. DataONE
7. Data Archive and Network Services (DANS), The Netherlands
8. Dryad
9. Figshare
10. Finish Social Science Data Archive
11. GigaScience
12. Knowledge Network for Biocomplexity (KNB)
13. National Institute of Allergy and Infectious Diseases, NIH
14. Odum Institute UNC-Chapel Hill
15. Open Preservation Foundation
16. PANGAEA
17. Springer Nature
18. TIB Leibniz Information Centre for Science and Technology University Library
19. Universidad Nacional de La Plata, Argentina
20. Virginia Tech University Libraries
21. World Data Center for Climate (WDCC), German Climate Computing Center (DKRZ), Hamburg, Germany.
22. World Data System

+Research Data Canada (RDC)

Contact enquiries[at]rd-alliance.org

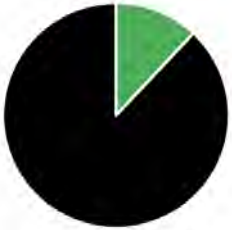
https://www.rd-alliance.org/rda-community-effort-trust-principles-digital-repositories

What are the concerns about repositories?

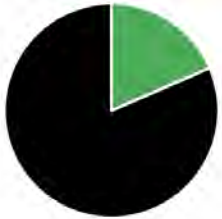
- Lost in investment and redundant expenses
- Missed opportunities to maximize the values in data

Data Needs A Trusted Home

- Dark Data
  - Only 12% of the data described in published papers were in recognized repositories.[1]



- Dark Data Repositories
  - Only 20% of the 328 biomedical data repositories in the survey were still alive or rebranded after 18 years. [2]



1. http://www.ncbi.nlm.nih.gov/pubmed/26207759
2. http://journal.embnet.org/index.php/embnetjournal/article/view/803/1209

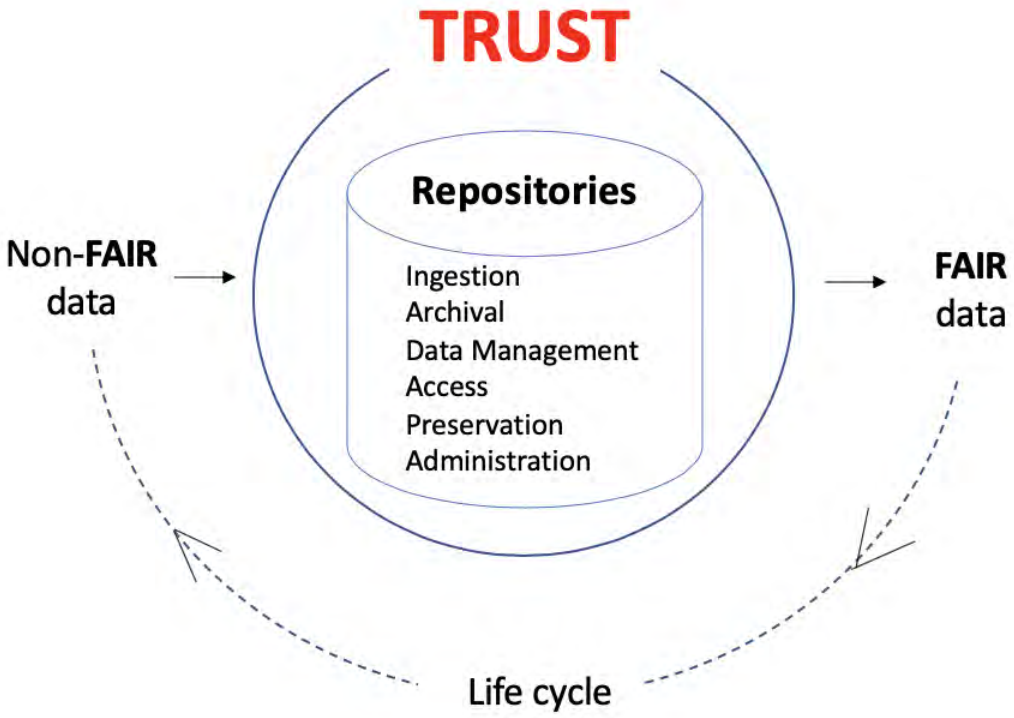
The Data Ecosystem for Open Science

OPEN *Science*

TRUST *repository*

FAIR *data*

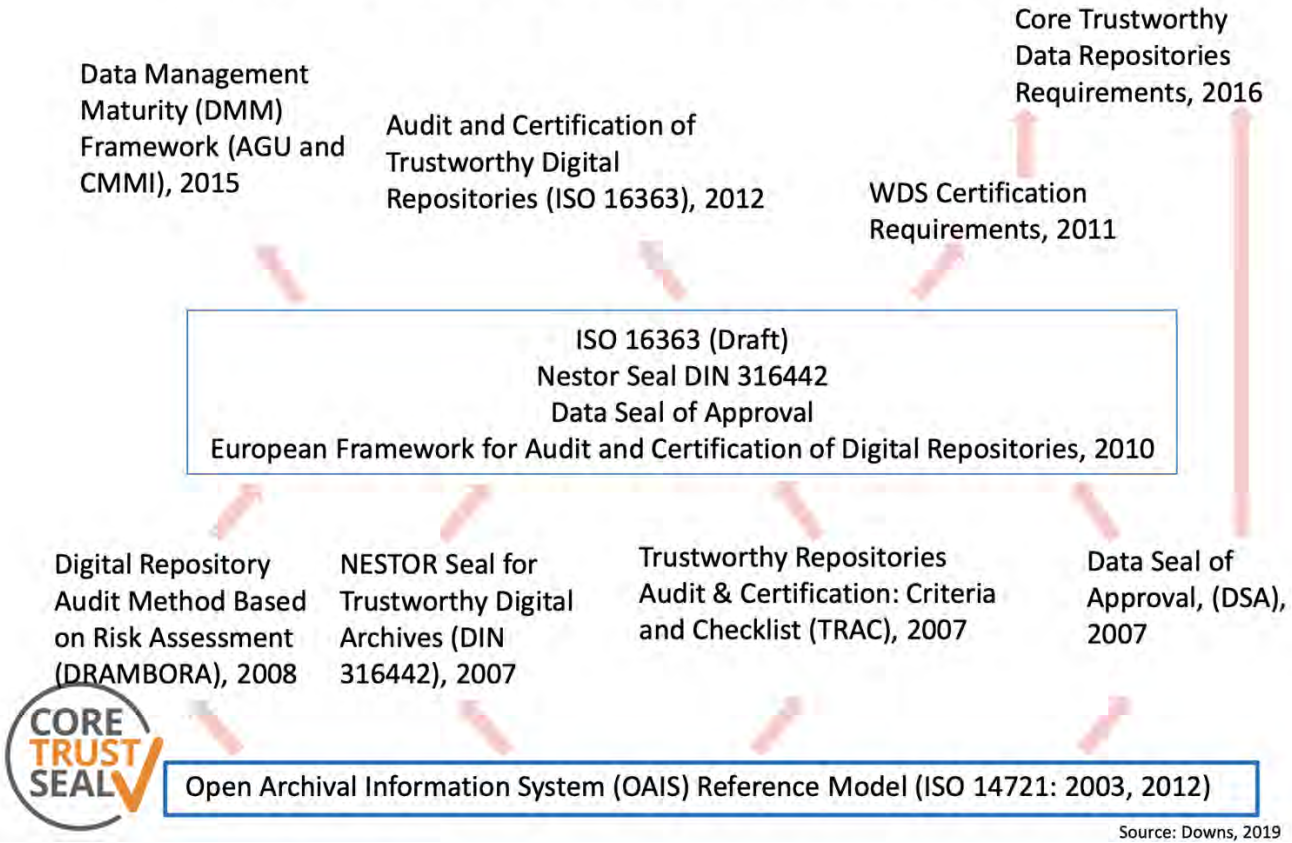
The Relationship of TRUST and FAIR



The Motivation of TRUST Principles

- Make it easy to understand the importance of Trustworthy Digital Repositories
- Develop concise and measurable approaches to achieve Trustworthy Digital Repositories
- **Not** to replace any standards, criteria, or best practices
- Provide a high-level starting point for advocating, supporting and implementing all certifications and assessments

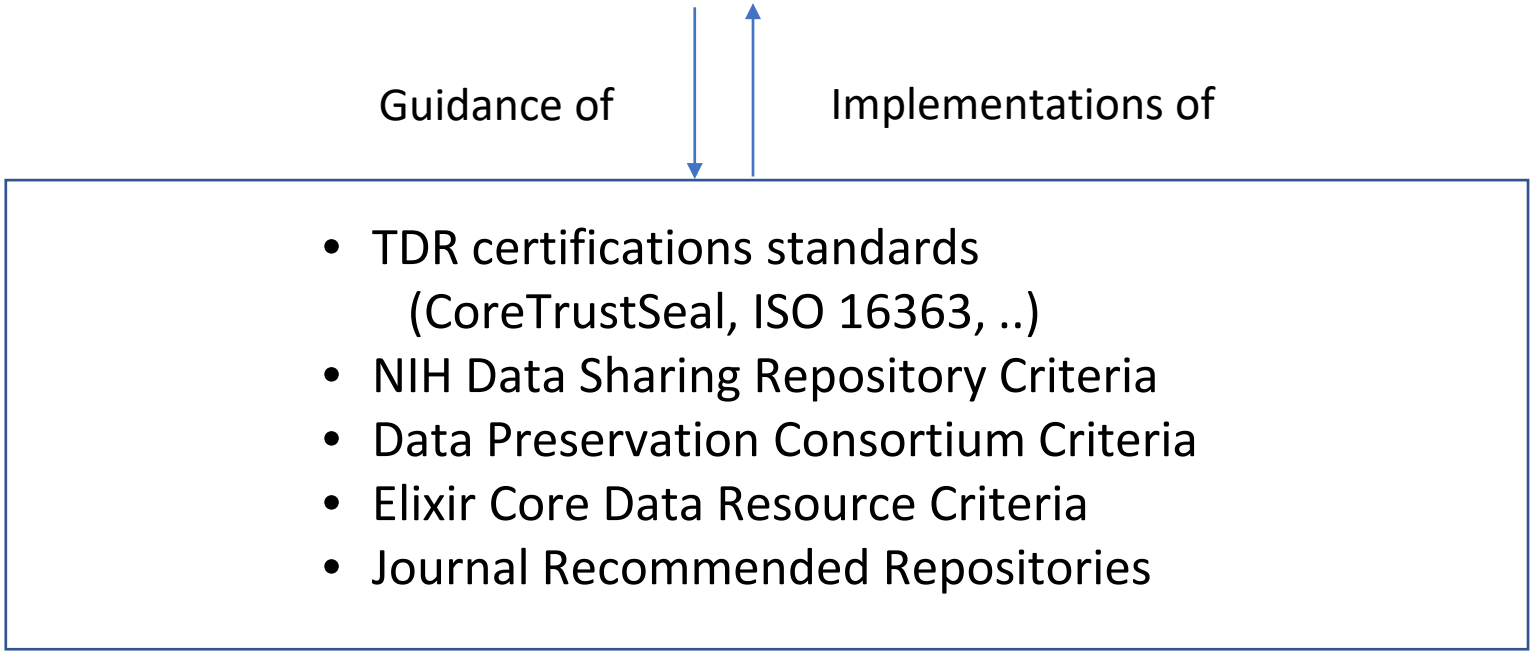
Trustworthy Data Repository Certification Standards: An Evolutionary Perspective



Source: Downs, 2019

One Voice from Digital Repository Community

TRUST Principles



The TRUST Principles

T – Transparency

Transparency: To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.

Key concepts:

- Terms of use, both for the repository and for the data holdings.
- Minimum digital preservation timeframe for the data holdings.
- Any pertinent additional features or services, for example the capacity to responsibly steward sensitive data.

R - Responsibility

Responsibility To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.

Key concepts:

- Adhering to the designated community's metadata and curation standards, along with providing stewardship of the data holdings.
- Providing data services e.g. portal and machine interfaces, data download or server-side processing.
- Managing the intellectual property rights of data producers, the protection of sensitive information resources, and the security of the system and its content.

U – User Focus

User community: To ensure that the data management norms and expectations of target user communities are met.

Key Concepts:

- Implementing relevant data metrics and making these available to users.
- Providing (or contributing to) community catalogues to facilitate data discovery.
- Monitoring and identifying evolving community expectations and responding as required to meet these changing needs.

S – Sustainability

Sustainability: To sustain services and preserve data holdings for the long-term.

Key Concepts:

- Planning sufficiently for risk mitigation, business continuity, disaster recovery, and succession.
- Securing funding to enable ongoing usage and to maintain the desirable properties of the data resources that the repository has been entrusted with preserving and disseminating
- Providing governance for necessary long-term preservation of data so that data resources remain discoverable, accessible, and usable in the future.

T – Technology

Technology: To provide infrastructure and capabilities to support secure, persistent, and reliable services.

Key Concepts:

- Infrastructure refers to the collection of people, processes, and technologies
- Hardware, software technologies, and cybersecurity measures adequate to the requirements of a robust, stable, and secure operation

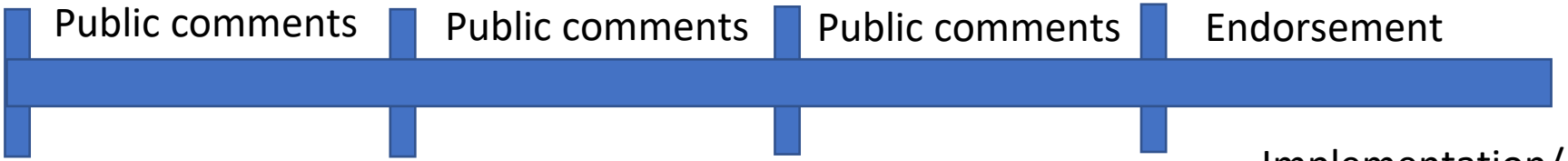
Impact of TRUST Principles

- Transparency is associated with trust of digital repositories - Donaldson, *et al*.
- Defining roles and responsibilities will help facilitate the effective stewardship – OAIS and Peng *et al.*
- Users' trust in data is also associated with their trust in the archive from which the content was obtained – Yoon *et al.*
- "Research data repositories are an essential part of the infrastructure for open science…" [and that it] "is important to ensure the sustainability of research data repositories" - OECD
- data stewardship is not just about physical and digital security: staff training, standard operating procedures, and the skills and attitudes of staff are also important - Van Staa *et al.*

Conclusions

- The TRUST Principles provide a mnemonic to remind data repository stakeholders
- The TRUST Principles, however, are not an end in themselves, rather a means to facilitate communication with all stakeholders, providing repositories with guidance to demonstrate transparency, responsibility, user focus, sustainability, and technology.

The TRUST Road Map



White Paper V0.01	White Paper V0.02	White Paper V0.03	Journal Publication
4/2019	11/2019	1/2020	5/2020
RDA	RDA	RDA	Nature Research Scientific Data

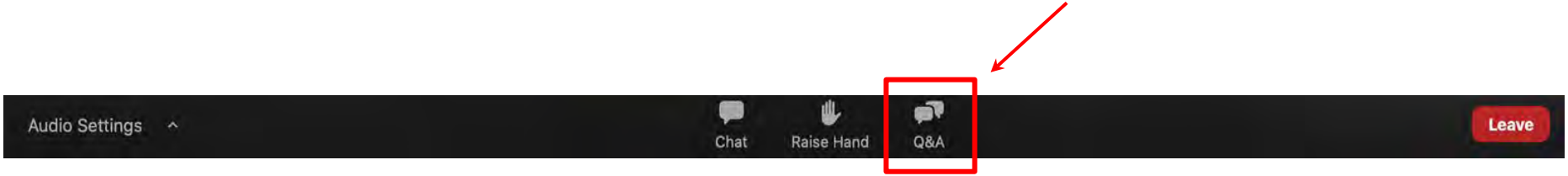
For endorsement, please contact enquiries[at]rd-alliance.org

Acknowledgements

- The members of the CoreTrustSeal Standards and Certification Board
- Participants of the Research Data Alliance Plenary 13 session, “Build TRUST to be FAIR - Emerging Needs of Certification in Life Sciences, Geosciences and Humanities”
- Participants of the NIH Workshop on Trustworthy Data Repositories for Biomedical Sciences (NIH Workshop, 2019) sponsored by NIH Office of Data Science Strategy (ODSS).
- Thoughtful discussions with Shelley Stall, Robert S. Chen, Mark Conrad, Peter Doorn, Eliane Fankhauser, Elizabeth Hull, Siri Jodha Singh Khalsa, Micky Lindlar, Limor Peer, Philipp Conzett, and Rachel Drysdale. We would like to thank Anupama Gururaj for proof-reading the article.

Questions & Answers

- Please use the Q&A option to ask questions of the presenter(s). Questions will be addressed at the end of each session when possible, and also at the end of the Symposium.
- The Q&A option can be found at the bottom of your Zoom screen:



TRUST: Transparency, Responsibility, User focus, Sustainability, Technology

RDA Research Data Alliance | World Data System

RDC DRC Research Data Canada – Données de Recherche Canada

International Technology Office

Cautionary Tales

The Importance of TRUSTworthy Repositories

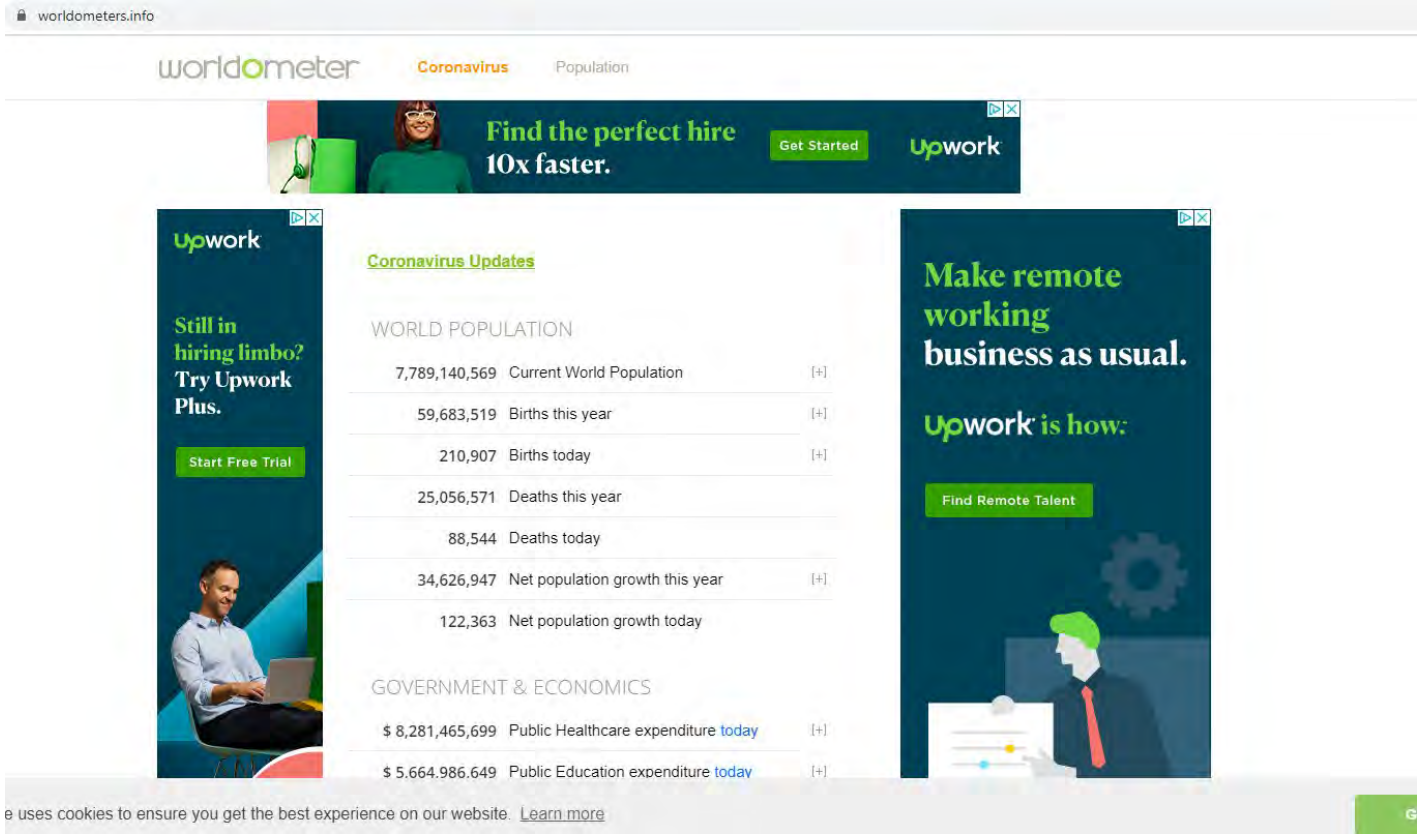
Dr. Karen Payne, Associate Director
for International Technology

Cautionary Tales - COVID19 (#1)

“We are one of the countries with the highest

number of tests carried out”

Cautionary Tales - COVID19 (#1)



An “advertising company”

Cautionary Tales - COVID19 (#2)

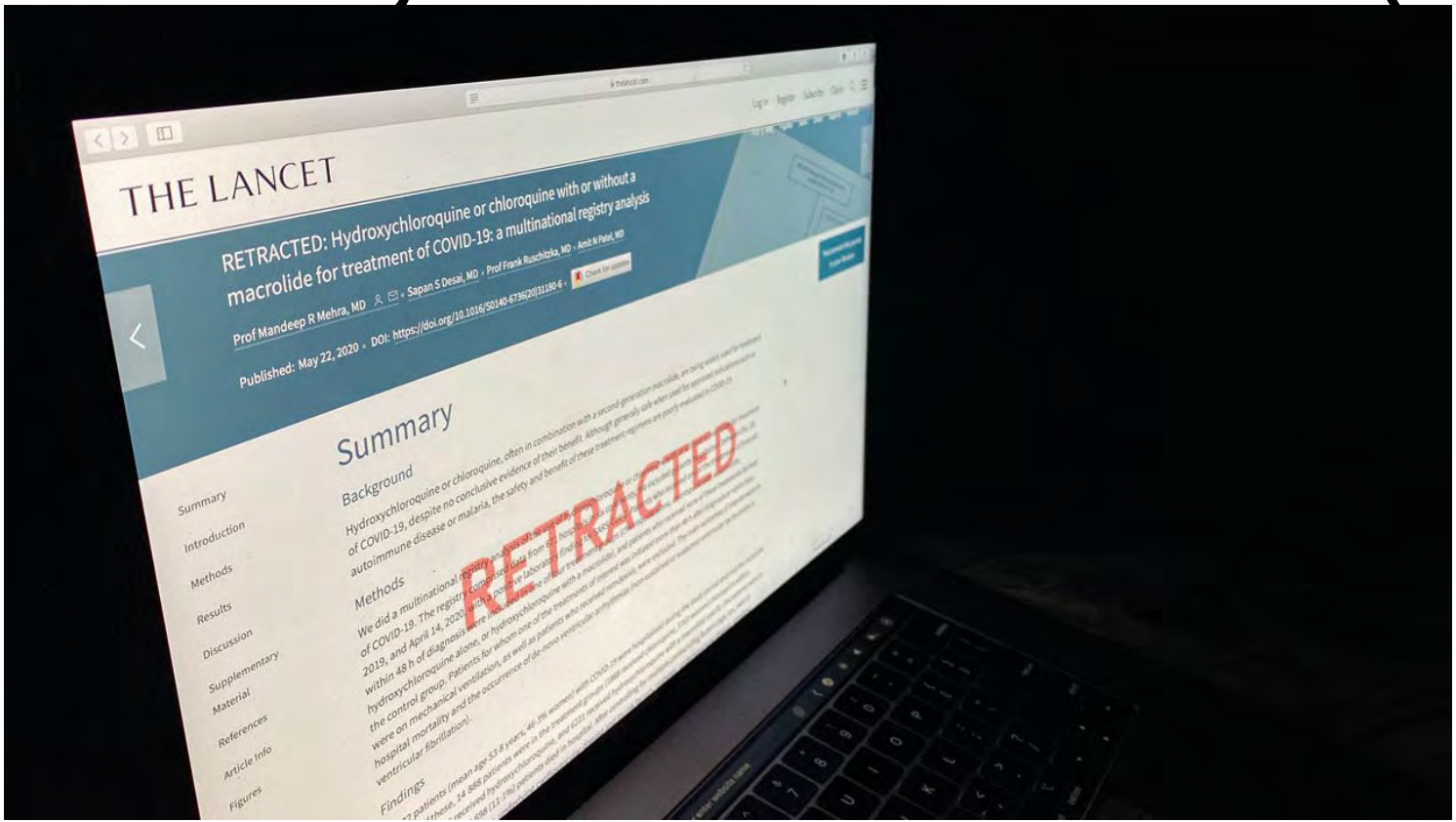
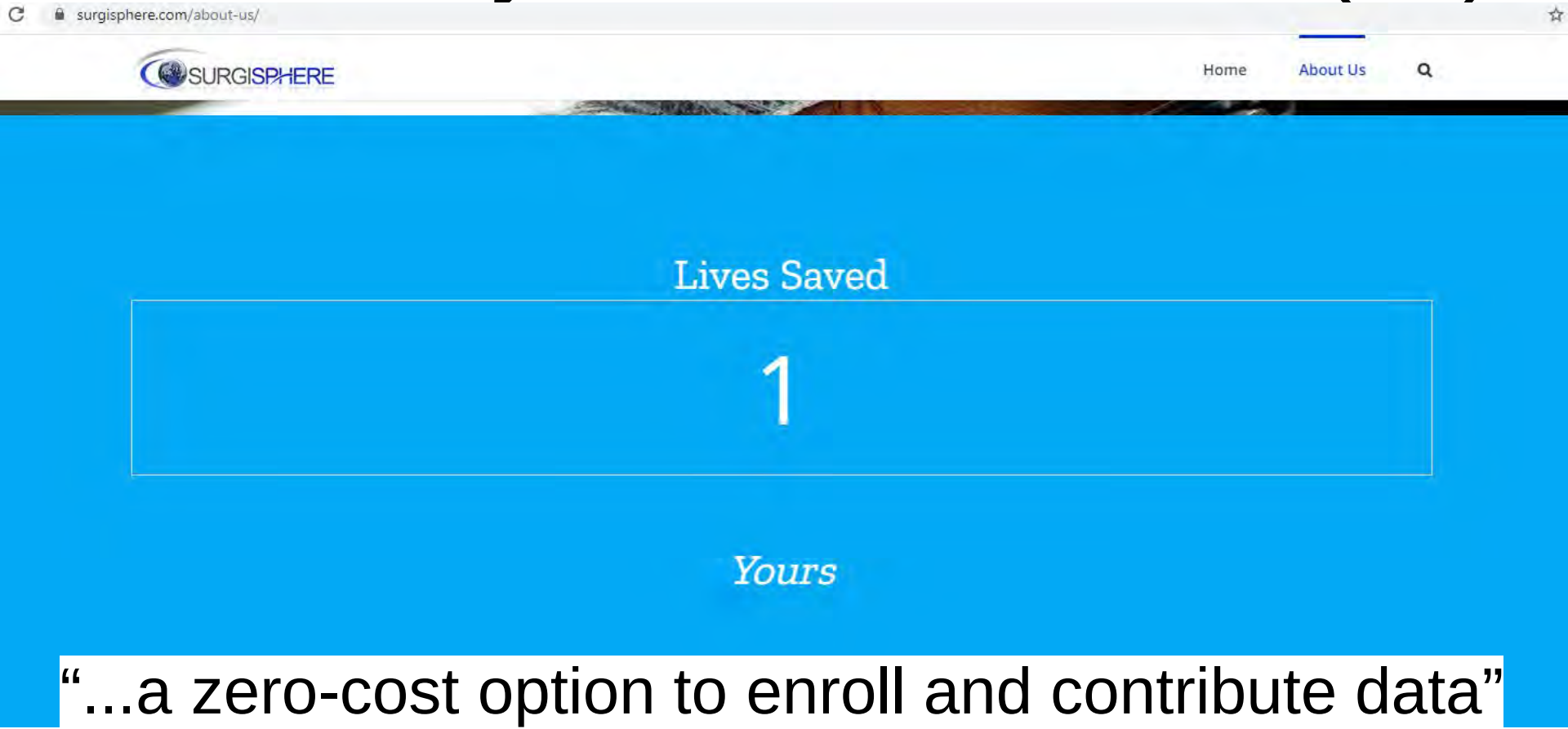


Image source: E. PETERSEN/SCIENCE https://www.sciencemag.org/news/2020/06/whos-blame-these-three-scientists-are-heart-surgisphere-covid-19-scandal

Cautionary Tales - COVID19 (#2)



"...a zero-cost option to enroll and contribute data"

Who Suffers?

1. The **general public**
2. **Governments** and the public sector
3. The parent organization or **university**
4. **Students**
5. **Colleagues**
6. The **field of study**

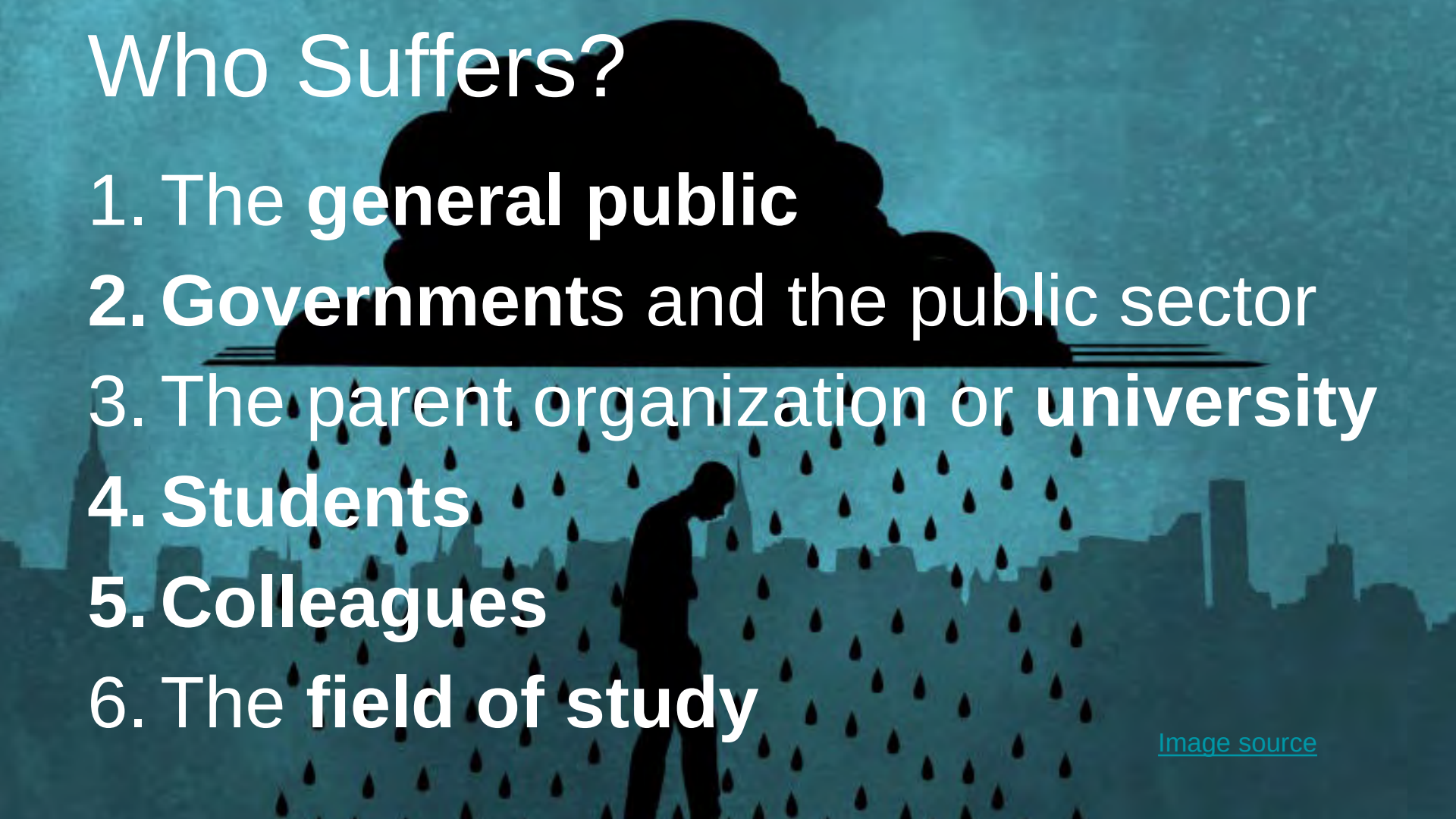
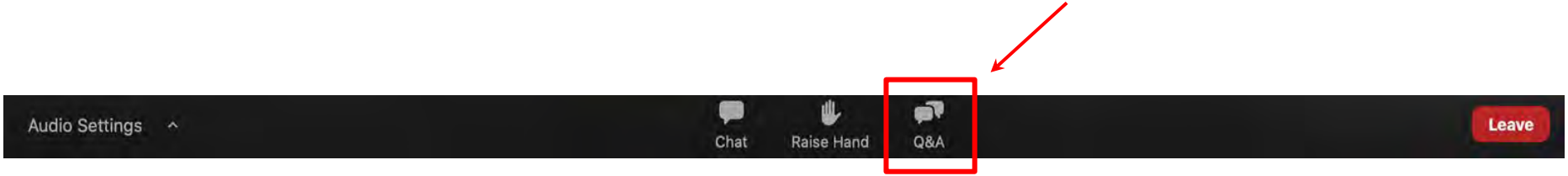


Image source

Questions & Answers

- Please use the Q&A option to ask questions of the presenter(s). Questions will be addressed at the end of each session when possible, and also at the end of the Symposium.
- The Q&A option can be found at the bottom of your Zoom screen:

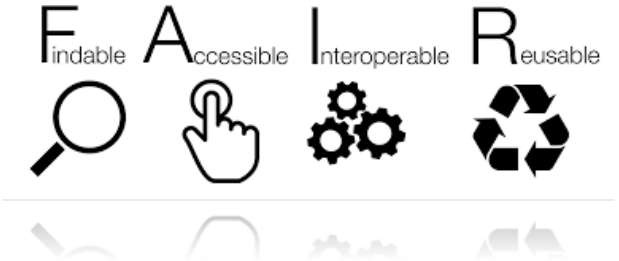


TRUST: Transparency, Responsibility, User focus, Sustainability, Technology

RDA Research Data Alliance | World Data System

RDC DRC Research Data Canada – Données de Recherche Canada

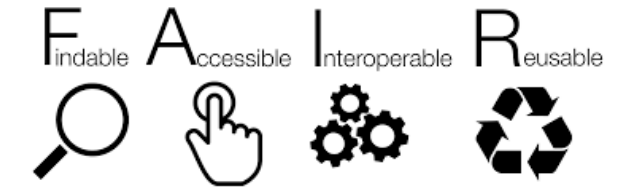
TRUST and FAIR: Complementarity of the Principles



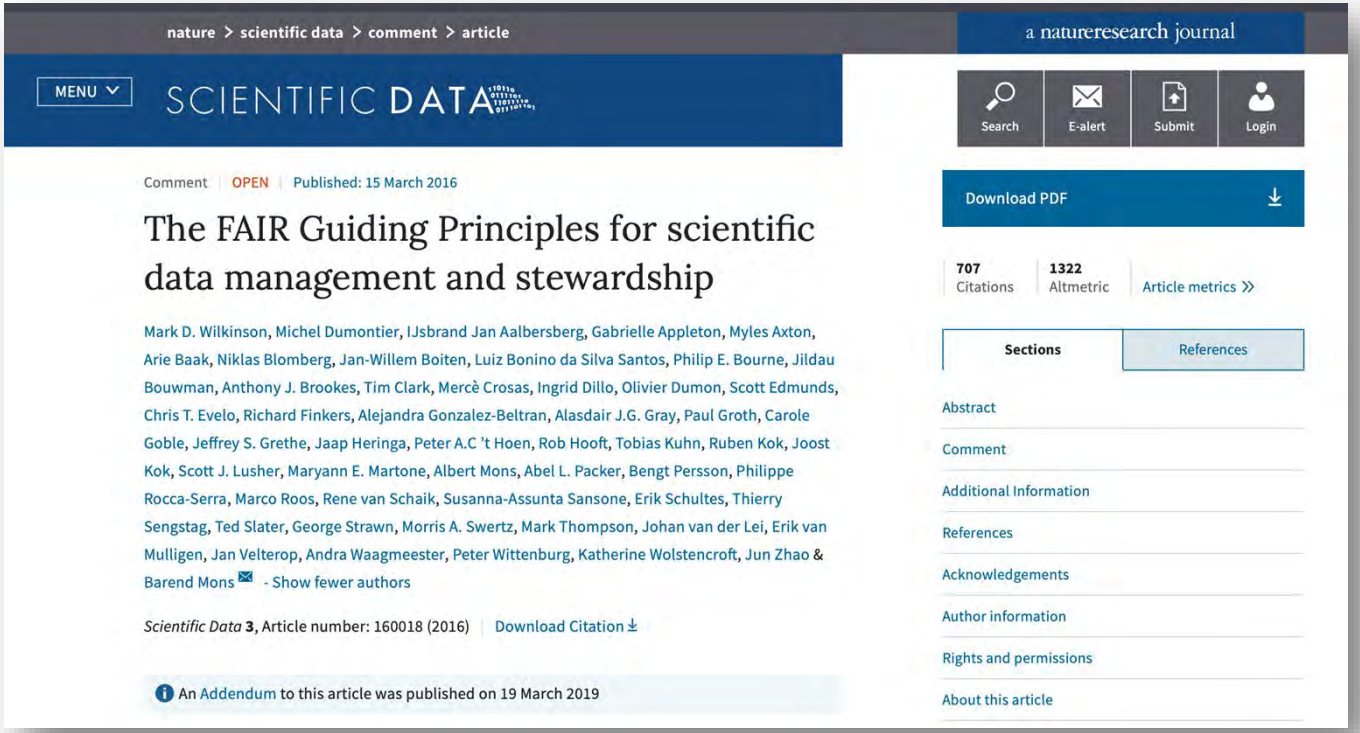
Dr. Ingrid Dillo
deputy director DANS, Netherlands
ingrid.dillo@dans.knaw.nl
https://dans.knaw.nl/en

FAIR principles 2014

During the 2014 workshop “Jointly designing a Data FAIRPORT” for the life sciences in Leiden a minimal set of community-agreed guiding principles were formulated.



FAIR Guiding Principles 2016



nature > scientific data > comment > article

a natureresearch journal

MENU

SCIENTIFIC DATA

Search | E-alert | Submit | Login

Comment | OPEN | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons - Show fewer authors

Scientific Data **3**, Article number: 160018 (2016) | Download Citation

An Addendum to this article was published on 19 March 2019

Download PDF

707 Citations | 1322 Altmetric | Article metrics »

Sections | References

- Abstract
- Comment
- Additional Information
- References
- Acknowledgements
- Author information
- Rights and permissions
- About this article

https://www.nature.com/articles/sdata201618

To facilitate reuse, data need to be:

- **Findable**: the first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers.
- **Accessible**: once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.
- **Interoperable**: the data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.
- **Reusable**: the ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

The success

- Well known among policymakers, funders, data service providers
- Less known among researchers

Science, Digital; Fane, Briony; Ayris, Paul; Hahnel, Mark; Hrynaszkiewicz, Iain; Baynes, Grace; et al. (2019): The State of Open Data Report 2019. figshare. Report. https://doi.org/10.6084/m9.figshare.9980783.v2

The confusion

continuum

The characteristics

- Focus on the data and metadata

- Evolution in time from principles to metrics/criteria to implementation standards/tools:
  - Producing FAIR data
  - Assessing the FAIRness of existing data
  - FAIRification of existing data
  - FAIR ecosystem including FAIR tools and services

- Provides a “snapshot” of a digital object
  in isolation of its context

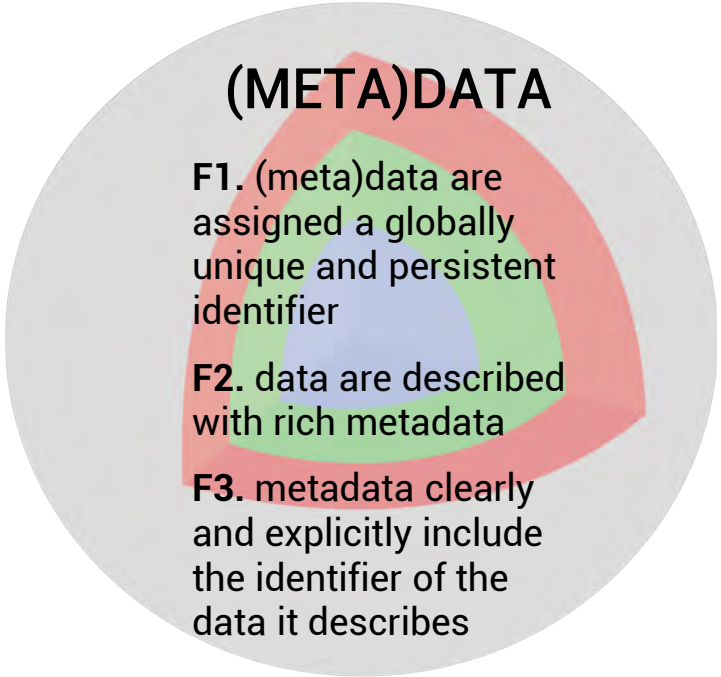
A FAIRytale?

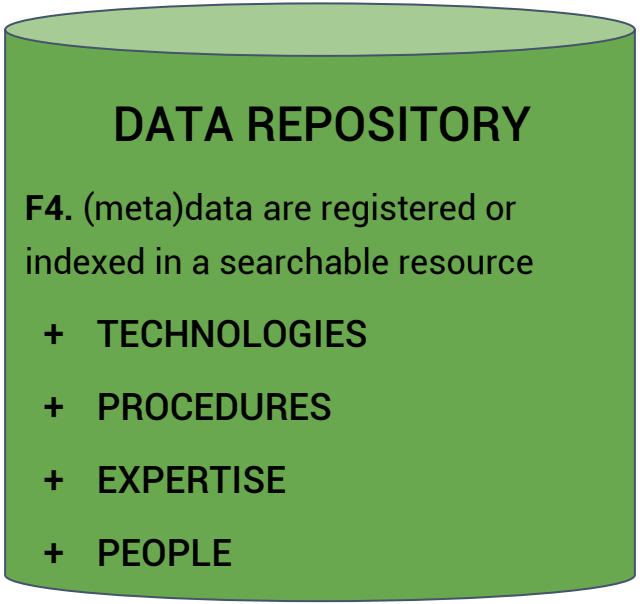
"Research data will not become nor stay FAIR by magic. We need skilled people, transparent processes, interoperable technologies and collaboration to build, operate and maintain research data infrastructures."

Mari Kleemola, Finnish Social Science Data Archive

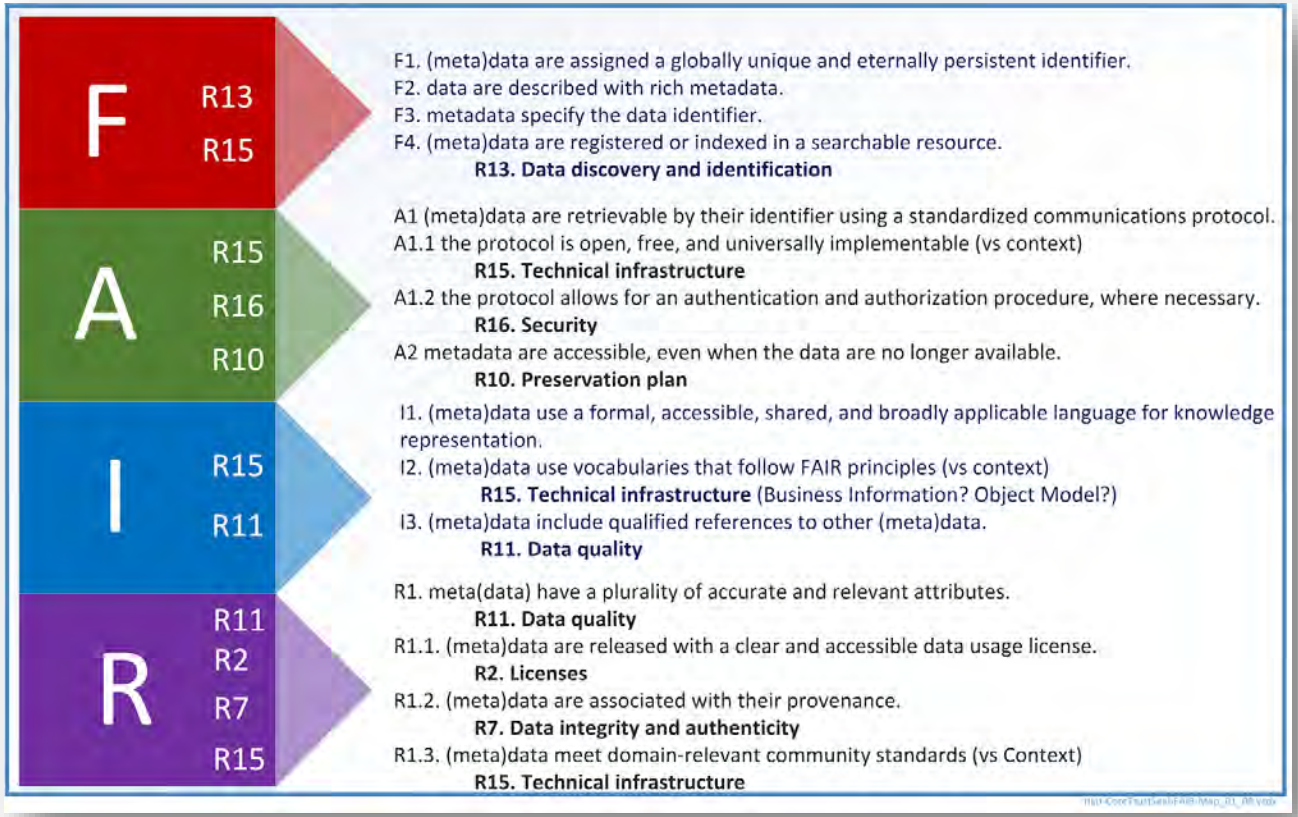
https://tietoarkistoblogi.blogspot.com/2018/11/being-trustworthy-and-fair.html

FAIR data assessment: findable





Mapping CoreTrustSeal and FAIR



Majority of CoreTrustSeal requirements (indirectly) refer to the FAIRness of the repository holdings

FAIRSFAIR
Fostering Fair Data Practices in Europe

D4.2 Repository Certification Mechanism: a Recommendation on the Extended Requirements and Procedures
https://doi.org/10.5281/zenodo.3835697

Importance of the environment in which data live

Trustworthy repositories add value because they:

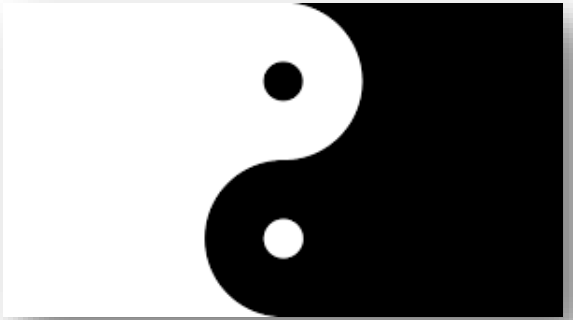
- enable a baseline FAIRness level to the datasets they hold (although some will be more FAIR than others);
- contribute to maintain or even increase the level of FAIRness over time through appropriate data curation and stewardship services.

https://ipres2019.org/static/pdf/iPres2019_paper_74.pdf

FAIR and TRUST are complementary

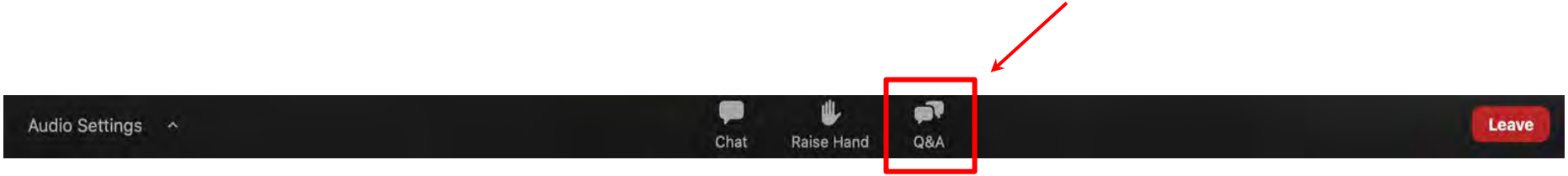
- FAIR data in Trustworthy Data Repositories (TDRs)

- TRUST principles provide the basis for building this trustworthiness into repositories:
  - Endorsement as a first step and an acknowledgement of the importance of trustworthiness
  - Certification as a possible next step



Questions & Answers

- Please use the Q&A option to ask questions of the presenter(s). Questions will be addressed at the end of each session when possible, and also at the end of the Symposium.
- The Q&A option can be found at the bottom of your Zoom screen:



TRUST: Transparency, Responsibility, User focus, Sustainability, Technology

RDA Research Data Alliance | World Data System

RDC DRC Research Data Canada – Données de Recherche Canada

TRUST, FAIR, CARE, and CTS

Wim Hugo
July 7th, 2020

Some Basic Ideas

Why do we need trust?

- Reproducibility and Credibility of Science
- Maximising Efficiency (investment, virtuous circle, …)

Increasingly, we will rely on:

- Third-party verification of trust
- Automated verification of trust and provenance of the elements of scientific work (data, methodology and protocols, code and algorithms, vocabularies and thesauri, samples and specimens, …)

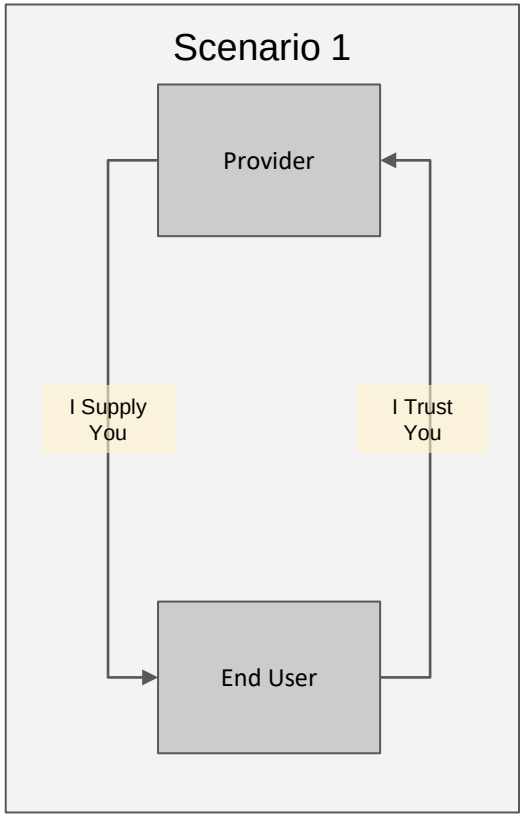
Trust is a Contract

It involves a **provider** and a **consumer**.

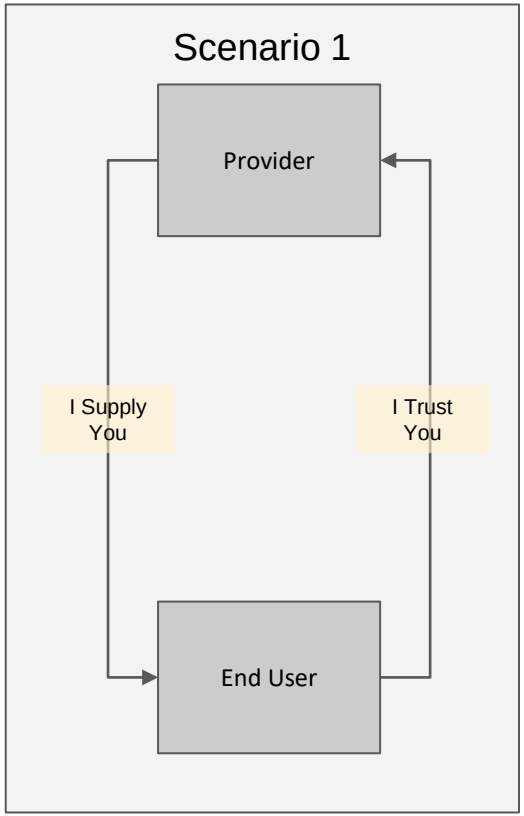
The consumer typically sets the requirements or expectations, and the provider aspires to fulfill those expectations or requirements.

The degree to which the consumer believes that the provider will comply is measured as a level of trust.

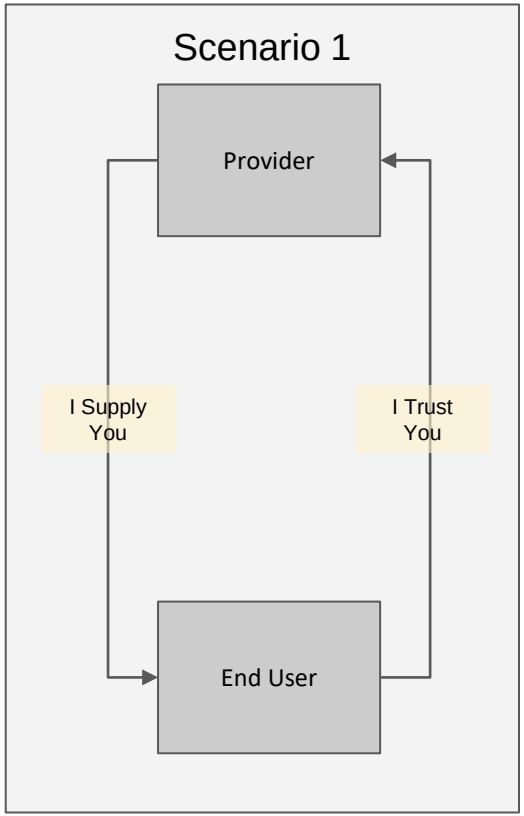
Options for Establishment of Trust

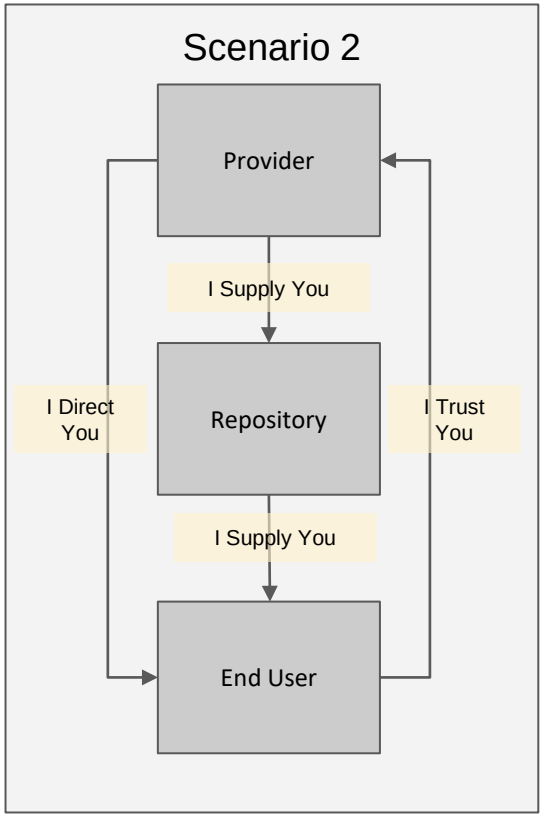


Options for Establishment of Trust



Options for Establishment of Trust



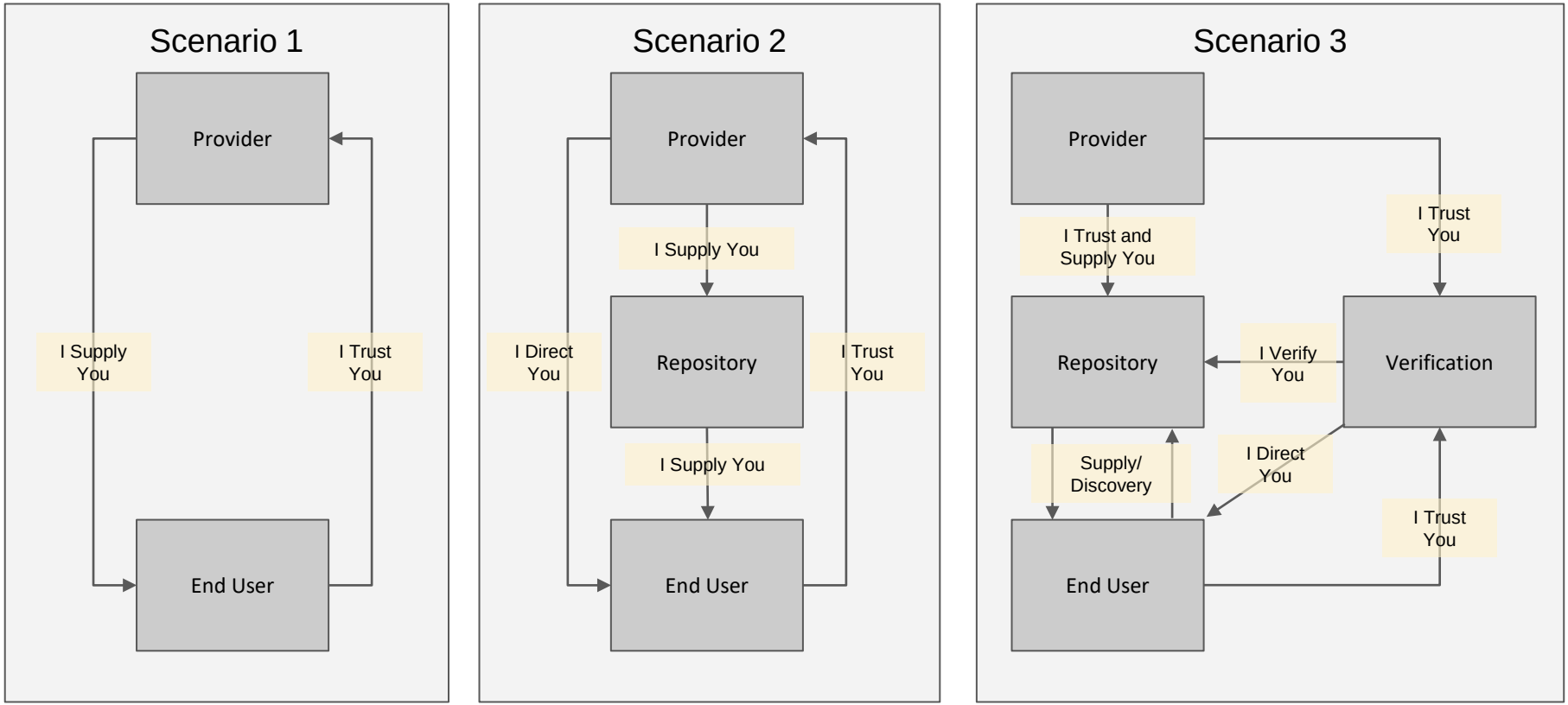


zenodo
The place to share your research results

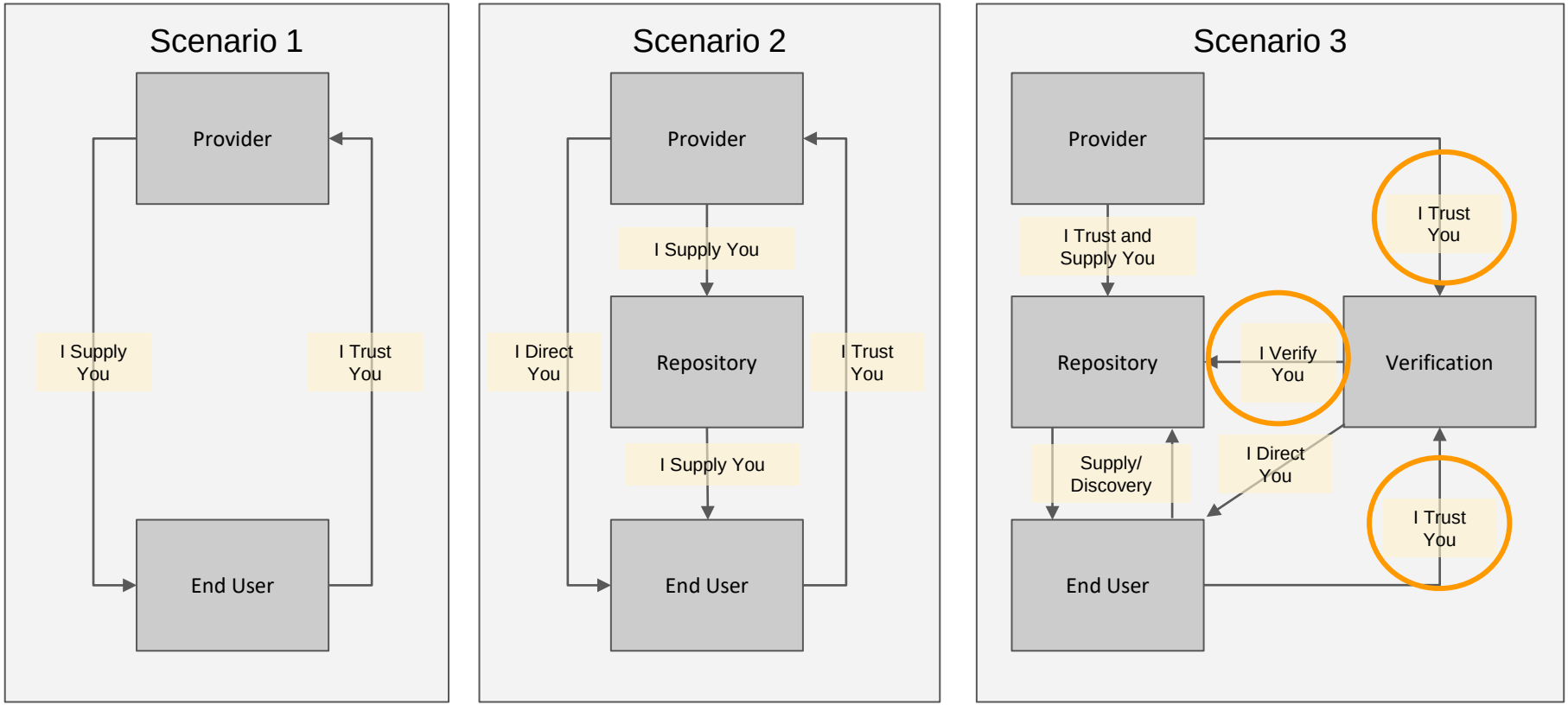
figshare
credit for all your research

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Options for Establishment of Trust



Options for Establishment of Trust



On a High Level, we have TRUST Principles

Transparency	To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.
Responsibility	To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.
User Focus	To ensure that the data management norms and expectations of target user communities are met.
Sustainability	To sustain services and preserve data holdings for the long-term.
Technology	To provide infrastructure and capabilities to support secure, persistent, and reliable services

Lin et al., 2020. The TRUST Principles for Digital Repositories. Scientific Data
https://doi.org/10.1038/s41597-020-0486-7

Also, FAIR and CARE

Findable	Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services
Accessible	Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation
Interoperable	The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.
Reusable	The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

Collective Benefit	Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data.
Authority to Control	Indigenous Peoples' rights and interests in Indigenous data must be recognised and their authority to control such data be empowered.
Responsibility	Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples' self-determination and collective benefit.
Ethics	Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Research Data Alliance International Indigenous Data Sovereignty Interest Group. (September 2019). "CARE Principles for Indigenous Data Governance." The Global Indigenous Data Alliance. GIDA-global.org

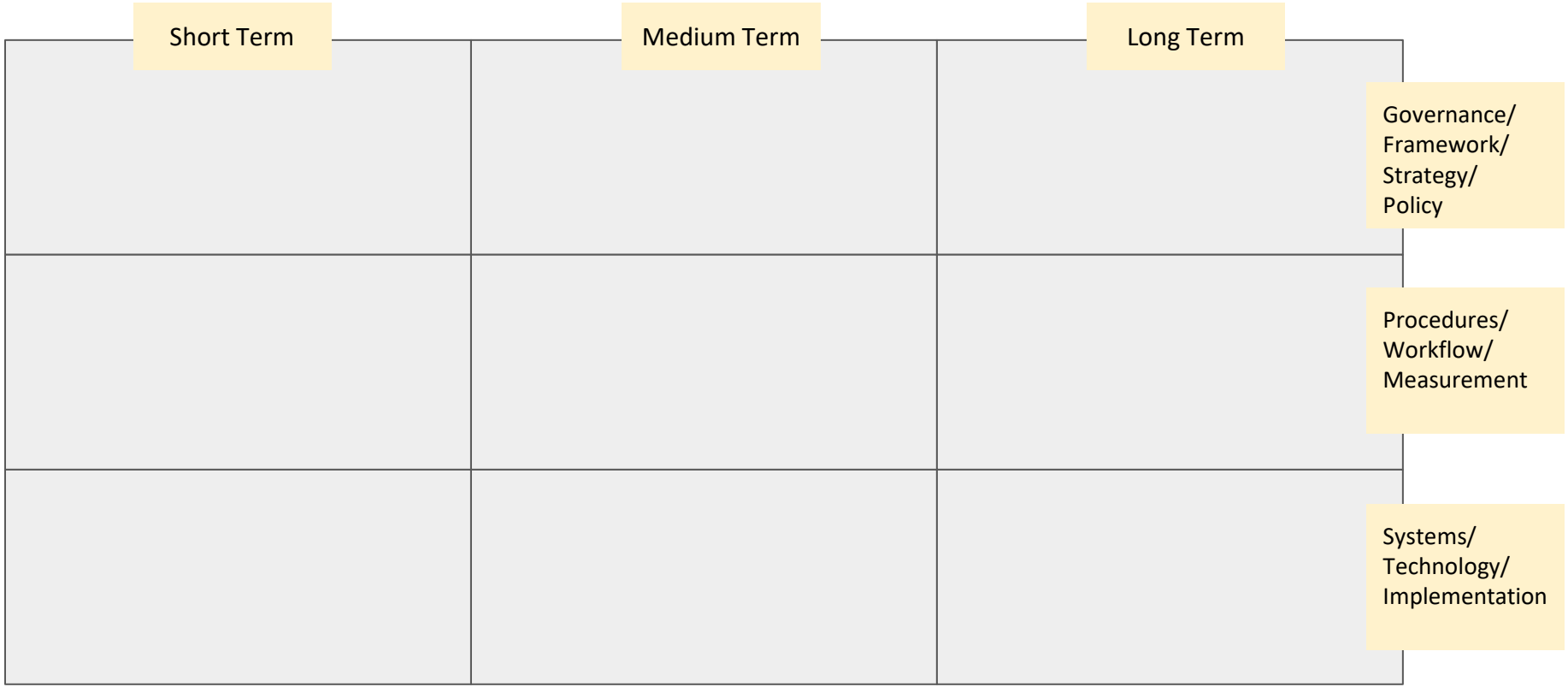
In the following slides: 3 Quick Perspectives

More Important to End Users

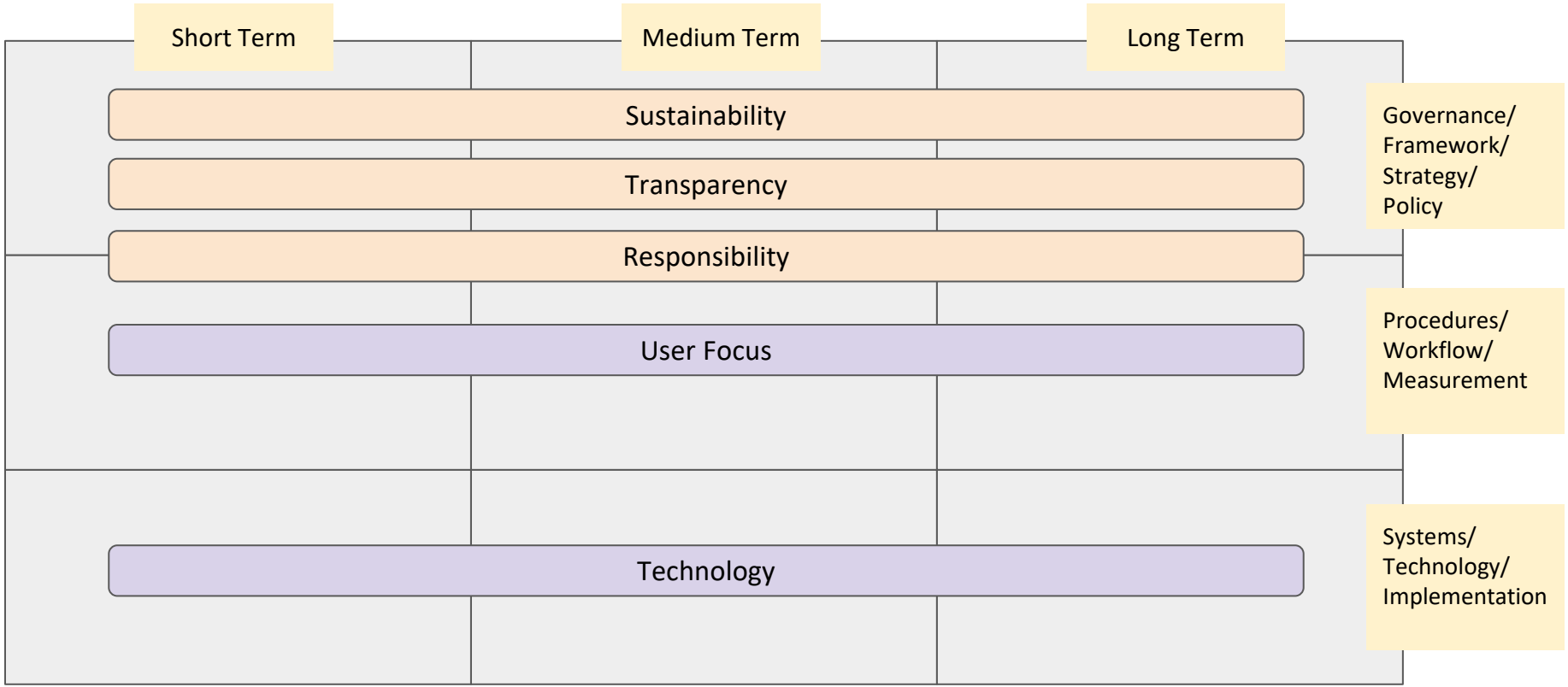
More Important to Depositors

Equally Important to Both

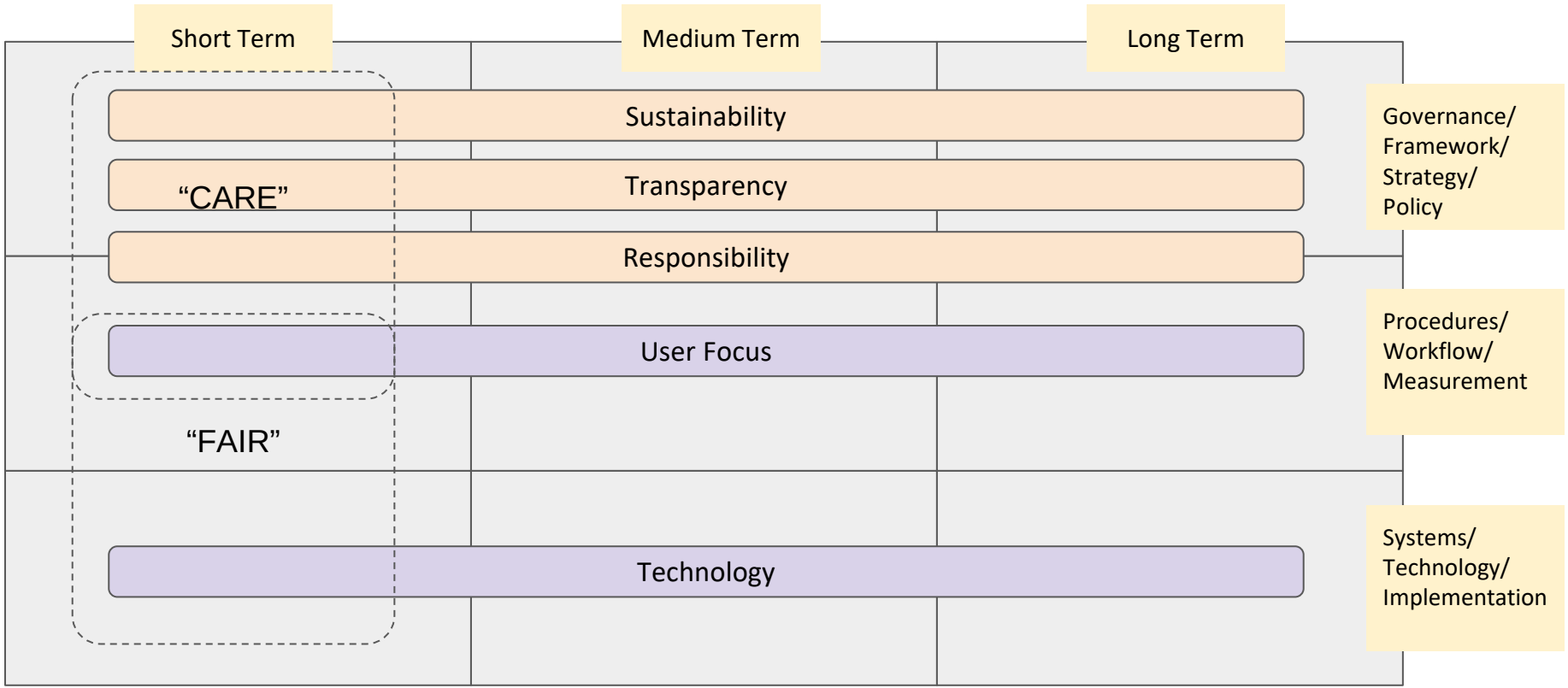
A Continuum of TRUST



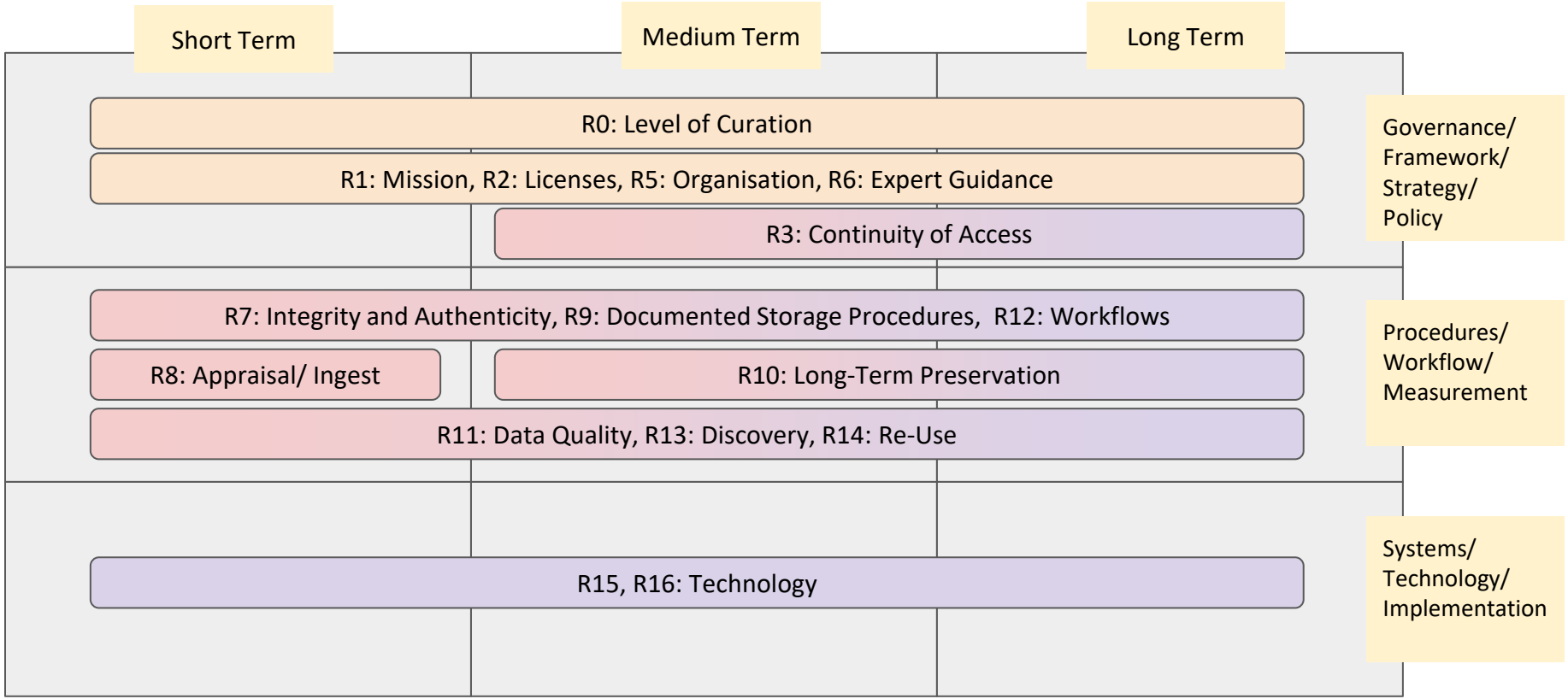
Perspective 1: A Continuum of TRUST



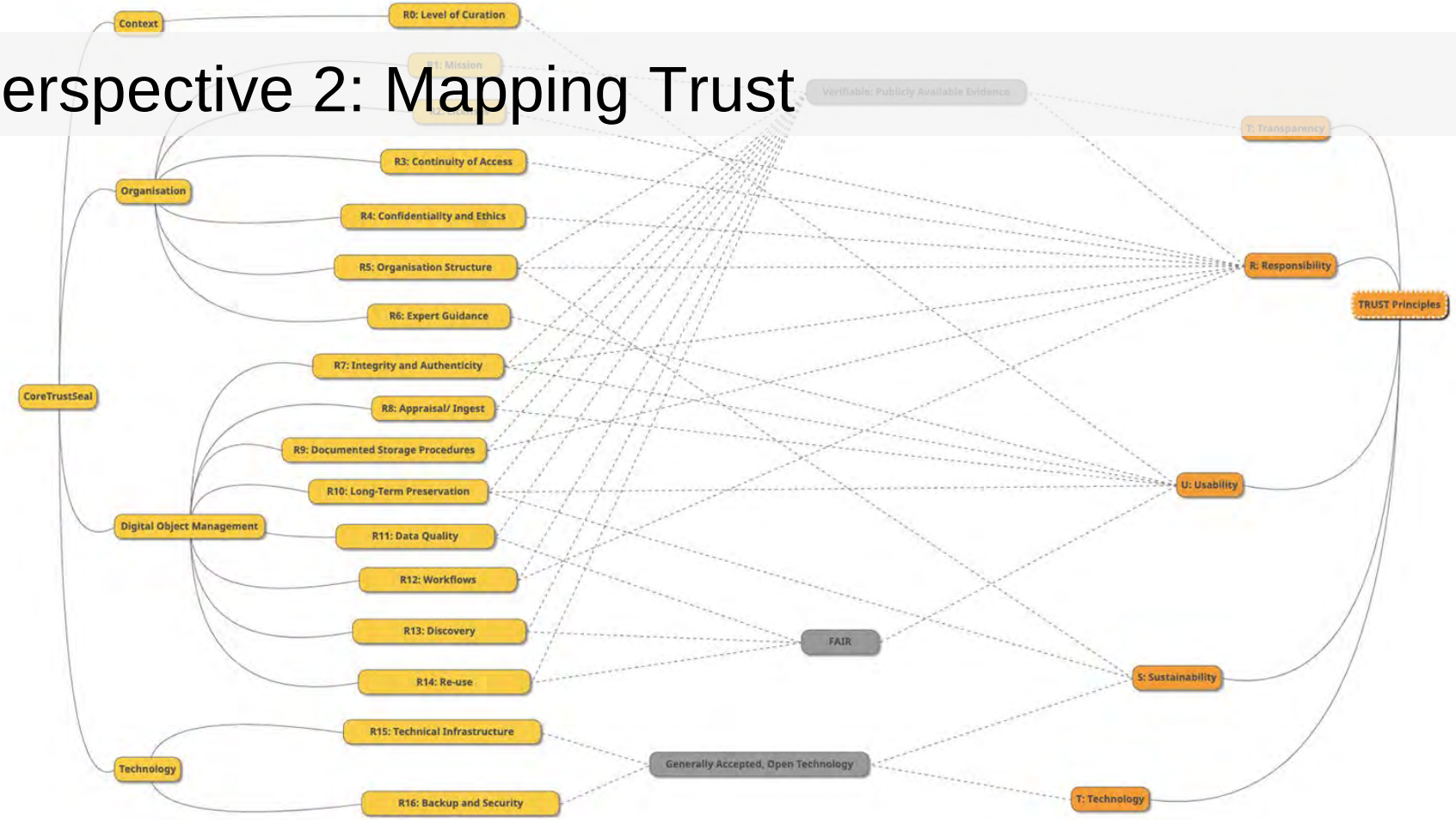
Perspective 1: A Continuum of TRUST

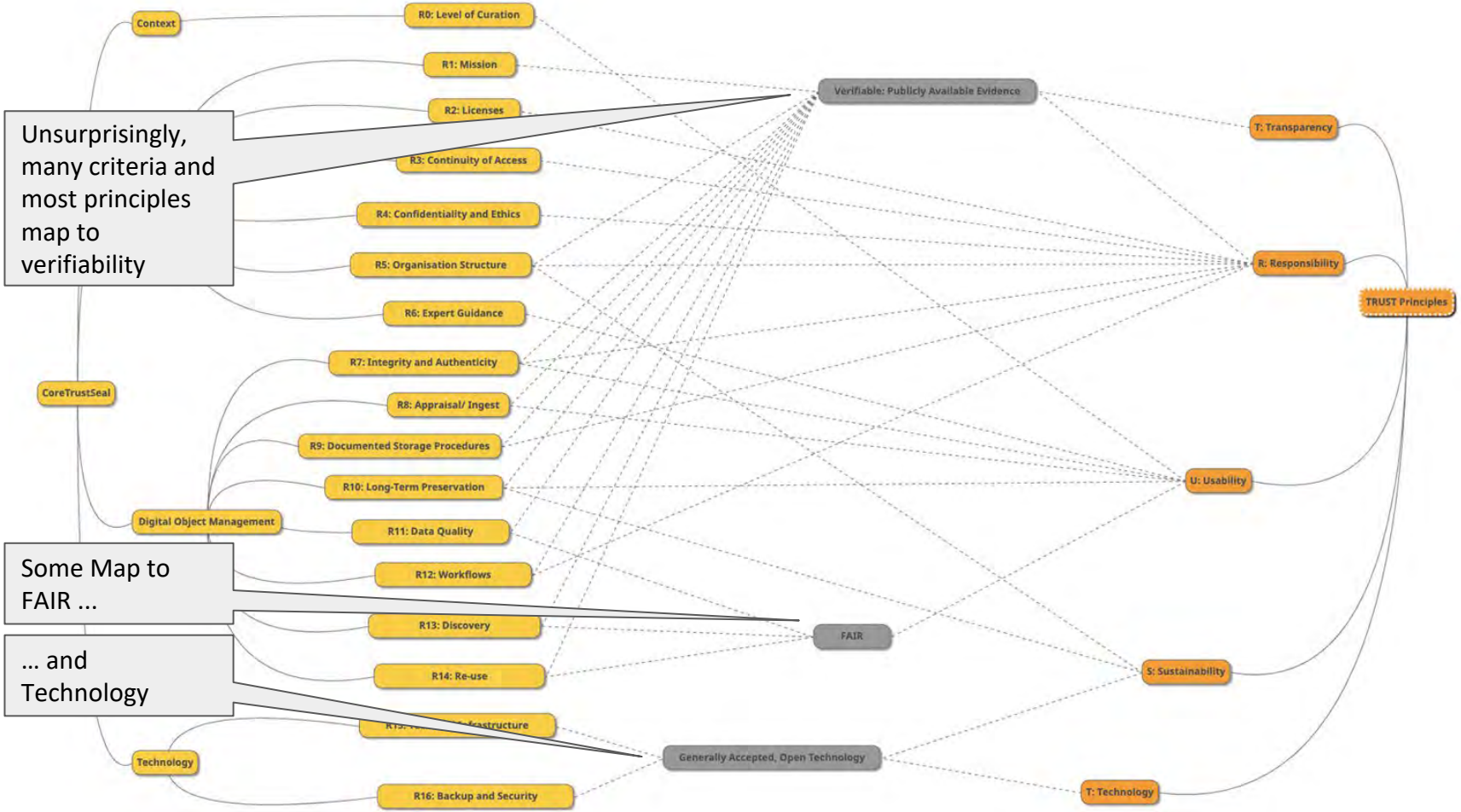


Perspective 1: A Continuum of TRUST (CTS)

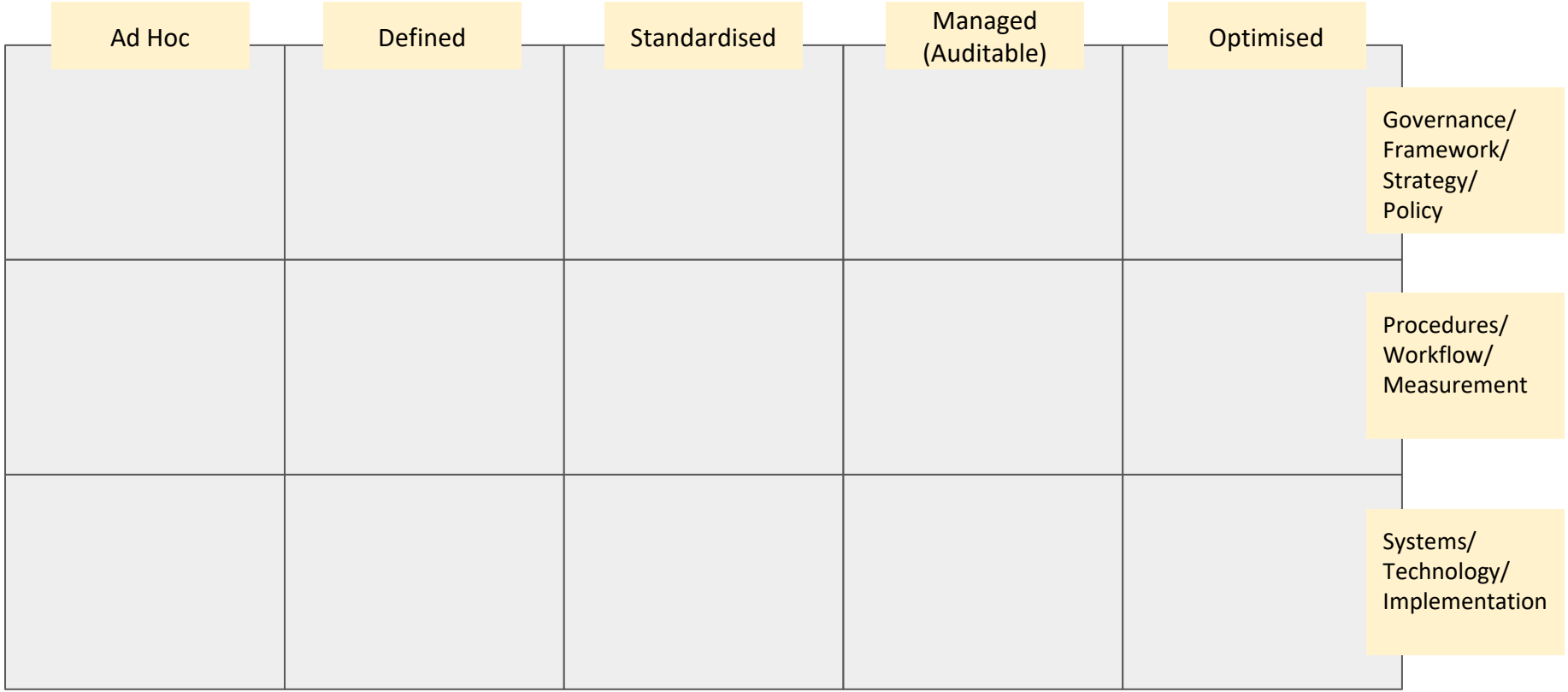


Perspective 2: Mapping Trust



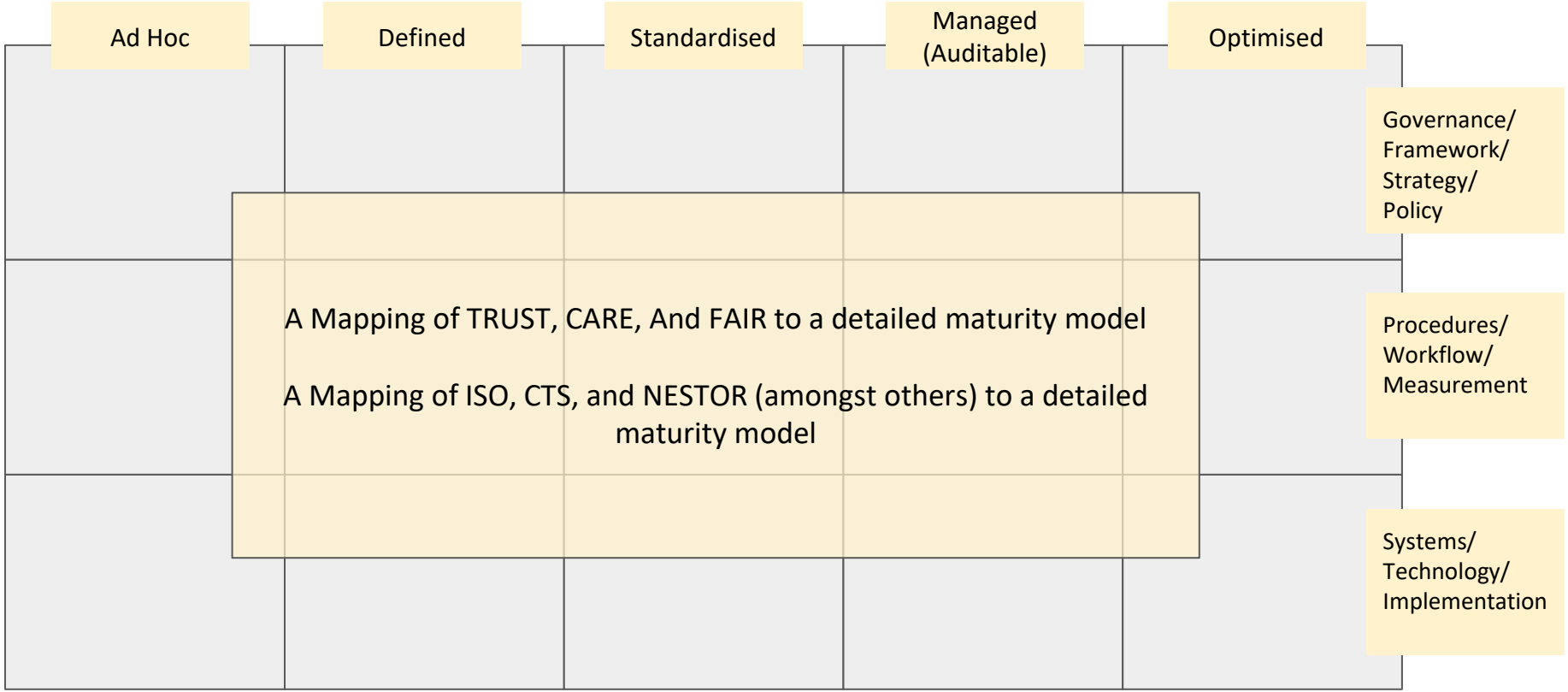

Unsurprisingly, many criteria and most principles map to verifiability
Some Map to FAIR ...
... and Technology
Context
CoreTrustSeal
Digital Object Management
Technology
R0: Level of Curation
R1: Mission
R2: Licenses
R3: Continuity of Access
R4: Confidentiality and Ethics
R5: Organisation Structure
R6: Expert Guidance
R7: Integrity and Authenticity
R8: Appraisal/ Ingest
R9: Documented Storage Procedures
R10: Long-Term Preservation
R11: Data Quality
R12: Workflows
R13: Discovery
R14: Re-use
R16: Backup and Security
Verifiable: Publicly Available Evidence
FAIR
Generally Accepted, Open Technology
T: Transparency
R: Responsibility
TRUST Principles
U: Usability
S: Sustainability
T: Technology

Perspective 3: Maturity and Measurement



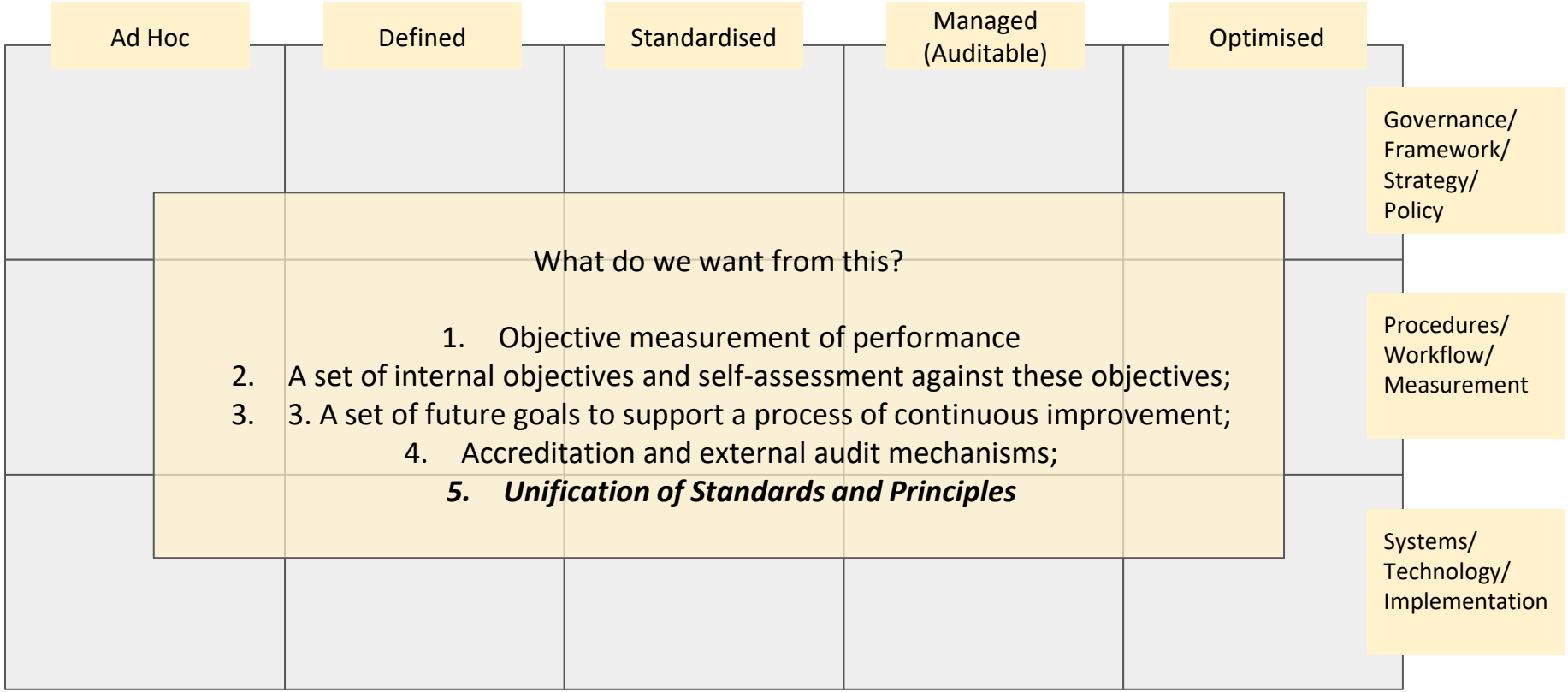
Hugo, Wim, "A Maturity Model for Digital Data Centers," Data Science Journal, no. 12 (2013): WDS189–WDS192, doi:10.2481/dsj.WDS-032 .

Perspective 3: Maturity and Measurement



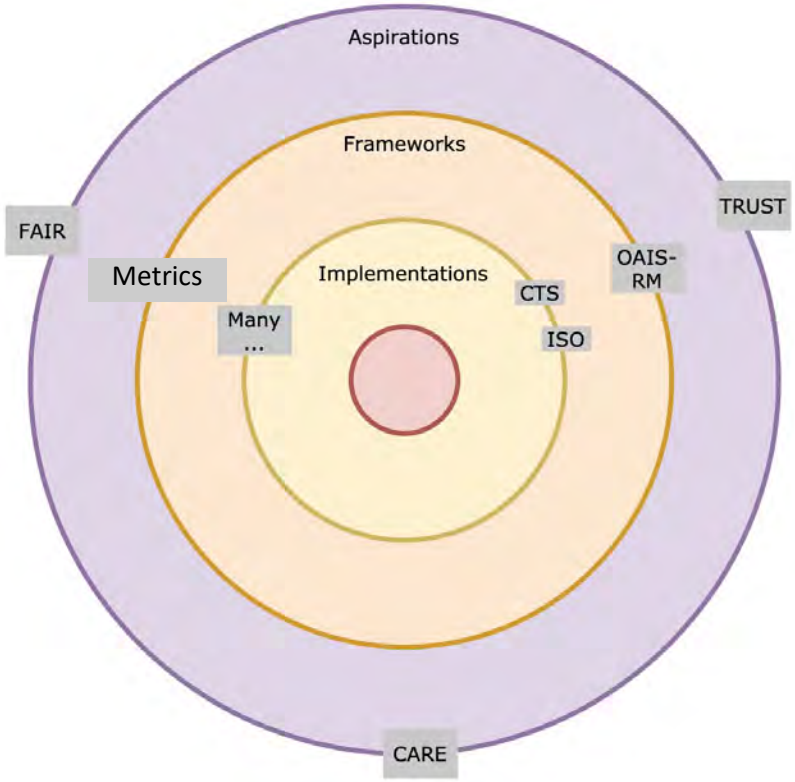
Hugo, Wim, “A Maturity Model for Digital Data Centers,” Data Science Journal, no. 12 (2013): WDS189–WDS192, doi:10.2481/dsj.WDS-032.

Perspective 3: Maturity and Measurement



Hugo, Wim, “A Maturity Model for Digital Data Centers,” Data Science Journal, no. 12 (2013): WDS189–WDS192, doi:10.2481/dsj.WDS-032.

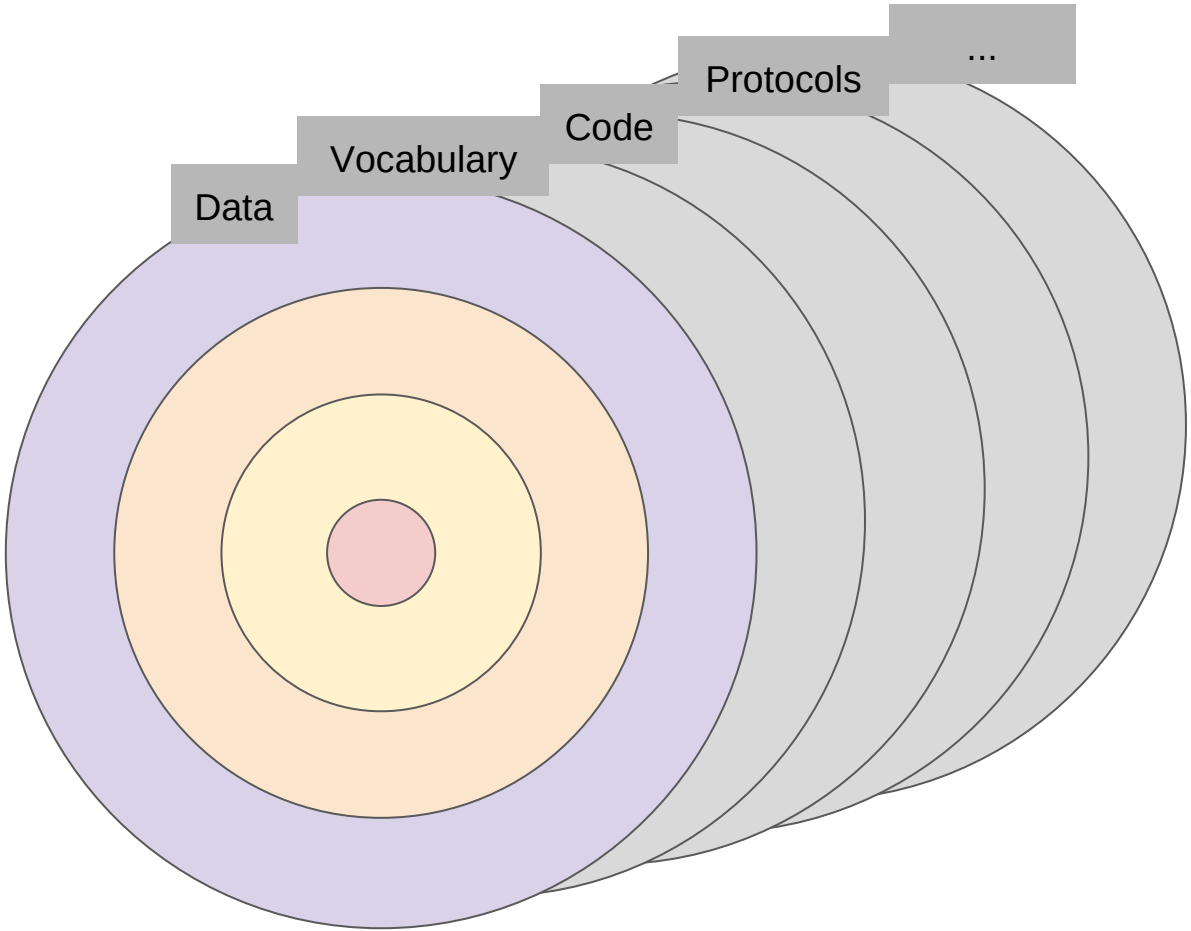
Call to Action



Call to Action

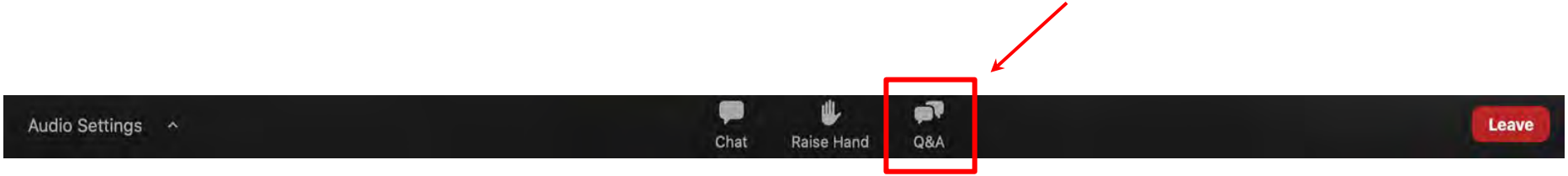
Developing an **Architecture** for TRUST, FAIR, and CARE that is integrated, measurable, covers all of shared scientific endeavour, and can be automated.

Covers Governance, Standards, Processes, Technology



Questions & Answers

- Please use the Q&A option to ask questions of the presenter(s). Questions will be addressed at the end of each session when possible, and also at the end of the Symposium.
- The Q&A option can be found at the bottom of your Zoom screen:



TRUST
Transparency Responsibilty User focus Sustainability Technology
RDA RESEARCH DATA ALLIANCE
WORLD DATA SYSTEM
RDC DRC
Research Data Canada – Données de Recherche Canada