

# TRUST Principles & CoreTrustSeal

Dawei Lin  
CoreTrustSeal Board Member  
and Associate Director for  
Bioinformatics at DAIT, NIAID, NIH

*Second Latin America and Caribbean Scientific Data Management Workshop  
Certification of Scientific Data Repositories (The Why, the How, the When)  
24 February 2021*



# Data Needs A Trusted Home

---

- Dark Data

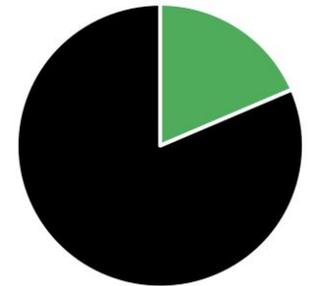
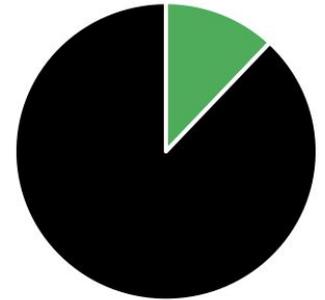
- Only 12% of the data described in published papers were in recognized repositories.<sup>1</sup>

- Dark Data Repositories

- Only 20% of the 328 biomedical data repositories in the survey were still alive or rebranded after 18 years.<sup>2</sup>

- Concerns

- Lost in investments
- Miss opportunities to maximize the values in data



1. <http://www.ncbi.nlm.nih.gov/pubmed/26207759>

2. <http://journal.embnet.org/index.php/embnetjournal/article/view/803/1209>

# 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 “Principles”

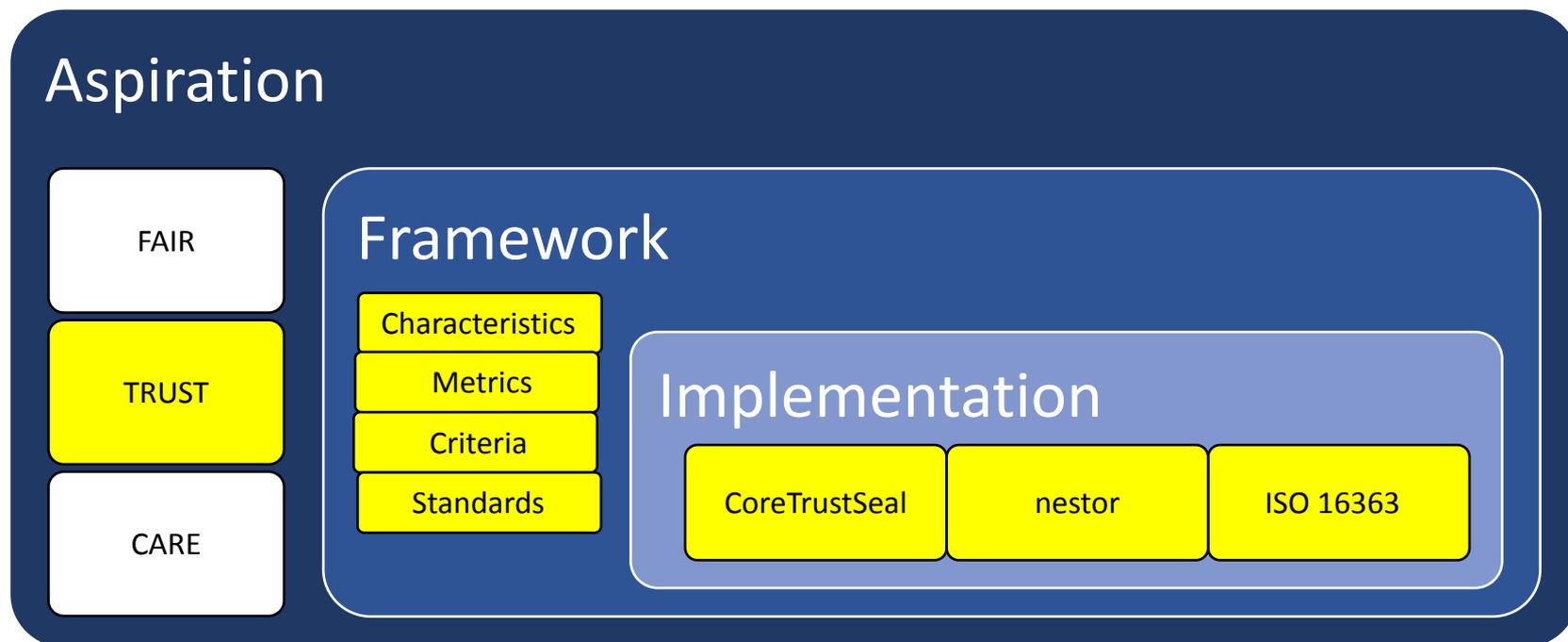
# The Motivation of TRUST Principles

---

High level principles  
to facilitate stakeholder  
discussion  
and guide repositories

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

# The Relationship of TRUST to Others



[FAIR](#): Findable, Accessible, Interoperable, Reusable

[TRUST](#): Transparency, Responsibility, User focus, Sustainability, Technology

[CARE](#): Principles for Indigenous data governance

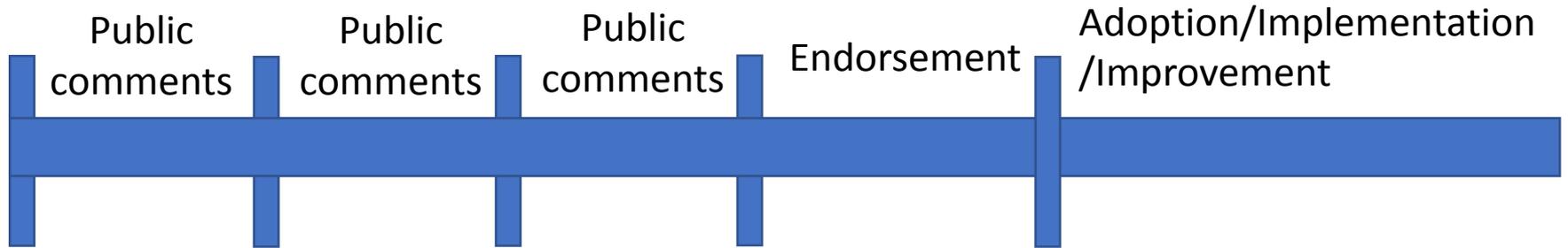
[OAIS-RM](#): Open Archival Information System Reference Model

[CTS](#): CoreTrustSeal data repository certification

[ISO 16363](#): International Organization for Standardization

# TRUST is a Community Effort

---



White Paper V0.01	White Paper V0.02	White Paper V0.03	Journal Publication	TRUST mini Symposium	Blogs, social media, citations,
4/2019	11/2019	1/2020	5/2020	7/2020	
RDA	RDA	RDA	Nature Scientific Data		

- Received 200+ public comments, 8+ countries, 4+ continents
- Stakeholders include funders, publishers, standards, librarians, and Data Preservationists
- Disciplines include social sciences, geosciences, and biomedical Sciences

# The TRUST Principles

---

**T**



Transparency

Foundation

**R**



Responsibility

Action

**U**



User focus

Motivation

**S**



Sustainability

Duration

**T**



Technology

Operation

# 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

# An RDA Community Effort

## Endorsements

1. 4TU.ResearchData
2. American Geophysical Union (AGU)
3. The Arctic Data Center
4. Canadian Association of Research Libraries (CARL)
5. Carnegie Mellon University Libraries
6. Center for International Earth Science Information Network (CIESIN), The Earth Institute, Columbia University, New York, USA.
7. CODATA Center for Excellence in Data for Society at the University of Arizona
8. Comisión de Investigaciones Científicas
9. CoMSES
10. CoreTrustSeal
11. DataONE
12. Data Archive and Network Services (DANS), The Netherlands
13. Digital Repository of Ireland
14. Dryad
15. Dutch Digital Heritage Network
16. Figshare
17. Finish Social Science Data Archive
18. GigaScience
19. ICPSR
20. Knowledge Network for Biocomplexity (KNB)
21. National Institute of Allergy and Infectious Diseases, NIH
22. Nestor
23. Ocean Networks Canada
24. Odum Institute UNC-Chapel Hill
25. Open Preservation Foundation
26. PANGAEA
27. Portage
28. Springer Nature
29. Research Data Access and Preservation Association (RDAP)
30. Research Data Canada / Données de recherche Canada
31. TIB Leibniz Information Centre for Science and Technology University Library
32. University of Arizona Libraries
33. Universidad Nacional de La Plata, Argentina
34. Virginia Tech University Libraries
35. World Data Center for Climate (WDCC), German Climate Computing Center (DKRZ), Hamburg, Germany.
36. World Data System



<https://www.rd-alliance.org/rda-community-effort-trust-principles-digital-repositories-0>

(As of Feb. 23, 2021)

# Positive Responses from the Community

## The TRUST Principles for digital repositories

Overview of attention for article published in Scientific Data, May 2020



### About this Attention Score

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

MORE...

### Mentioned by

- 3 blogs
- 2 policy sources
- 159 tweeters
- 5 Facebook pages
- 1 Redditor

### Citations

- 12 Dimensions

### Readers on

- 82 Mendeley

What is this page?

### SUMMARY

Blogs

Policy documents

Twitter

Facebook

Reddit

Dimensions citations

Title	The TRUST Principles for digital repositories
Published in	Scientific Data, May 2020
DOI	10.1038/s41597-020-0486-7 <a href="#">↗</a>
Pubmed ID	32409645 <a href="#">↗</a>
Authors	Dawei Lin, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giaretta, Marisa... <a href="#">[show]</a>

[View on publisher site](#)

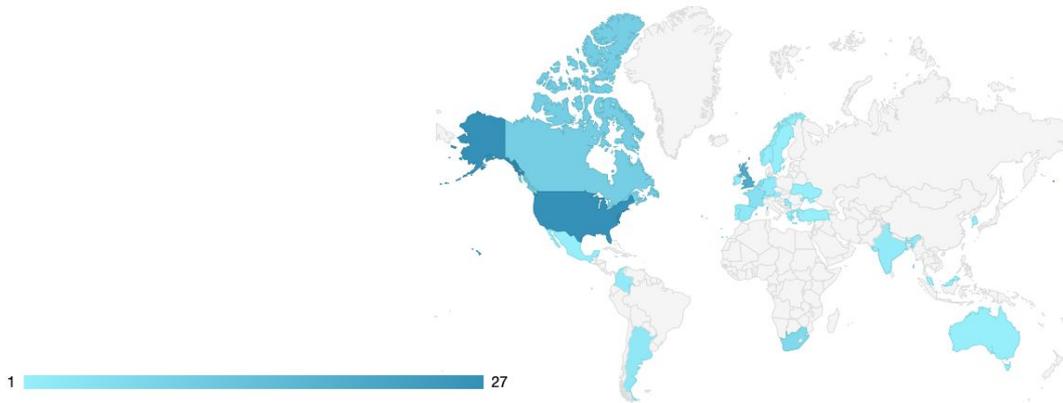
[Alert me about new mentions](#)

### TWITTER DEMOGRAPHICS

### MENDELEY READERS

### ATTENTION SCORE IN CONTEXT

The data shown below were collected from the profiles of 159 tweeters who shared this research output. [Click here to find out more about how the information was compiled.](#)



Geographical breakdown

Demographic breakdown

14K access, 8 citations by Scientific data,  
25 citations by Google Scholar

As of Feb. 23, 2021 from Altmetric Attention Score at Nature Research Data

# Responses and Suggestions



**Caroline Coward**  
@tenorclef530

335  
FOLLOWERS

A timely and insightful article for all the data heads out there. If you like your data FAIR, you'll love a repository you can TRUST: <https://t.co/llKxGkYUDT>



**FDM-Hildesheim homeoffice Research Data Management**  
@fdmhildesheim

Fully agree: "What the FAIR Principles lack, however, is systematic attention to digital preservation."



**George Mensah, MD**  
@NHLBI\_Translate

ICYMI: Read the community-developed TRUST principles for promoting the adoption of [#Transparency](#) [#Responsibility](#) [#User-focused](#) [#Sustainability](#) and [#Technology](#). They are complementary to the [#FAIR](#) Principles & aim to promote [#TRUST](#) for data [#repositorie](#)



**Philipp Konzett**  
@PhilippKonzett

Good to see @dataverseorg community members represented in this work! Looking forward to seeing more and more Dataverse repositories being CoreTrustSeal-certified.



**UCT Digital Library**  
@DigitalUct

880  
FOLLOWERS

An important read for ilifu: 'Repositories must earn the trust of the communities they intend to serve and demonstrate that they are reliable and capable of appropriately managing the data they hold.' <https://t.co/3mhax0m3XD> @UCT\_eResearch



**dkNET**  
@dkNET\_Info

494  
FOLLOWERS

RT @ScientificData: A diverse set of authors present TRUST, a community-developed set of principles designed to help data repositories supp...



**Springer Nature**  
@SpringerNature

51,483  
FOLLOWERS

We have endorsed TRUST, a community-developed set of principles designed to help data repositories support [#FAIRdata](#) and build trust with their user community. Read more in [@ScientificData](#): <https://t.co/KUV2NuLPNA> <https://t.co/Kkbb1drtkM>



**force11.org**  
@force11rescomm

4,554  
FOLLOWERS

RT @MokraneMA: Very happy to see our paper 'The TRUST Principles for digital repositories' published. We hope this to be a useful guide and...

## Recommendations for further development

In order to fulfill their full potential, we suggest that the TRUST Principles be developed further, paying attention to the following aspects:

While succeeding at establishing an overall framework for TRUSTedness, the TRUST Principles do not convey a sufficiently comprehensive picture of the requirements that data holding organizations need to fulfill to be considered trustworthy long-term archives. While organizational aspects such as business continuity and disaster recovery are addressed, measures relating to the actual digital objects – e.g. preservation planning and the implementation of suitable long-term preservation strategies – are not. This bears a risk that, despite their importance, these aspects of digital preservation do not receive sufficient attention.

To counter this, the TRUST Principles must be linked with established and accepted criteria suited to measuring trustworthiness. These criteria are already referenced by the TRUST Principles and may serve as inspiration as to which additional, preservation action-related requirements may be relevant indicators of TRUSTedness. This can also help mitigating the risk that claims of "compliance" with the TRUST Principles are used to communicate TDR status without the application of agreed metrics. At the same time, from our perspective, the creation of another standard to measure a repository's capacity to provide reliable long-term preservation of digital objects, further lowering the minimum ('core') requirements for trustworthiness, as expressed by the CoreTrustSeal, for example, should be avoided.

## Benefits and recommendations at a glance

- provides a common framework to facilitate discussion by all stakeholders
- mnemonic helps to raise awareness
- provides a low-threshold entry point
- principles do not convey a sufficiently comprehensive picture of the requirements
- preservation planning and suitable long-term preservation strategies are missing
- TRUST Principles must be linked with established and accepted criteria suited to measuring trustworthiness



**Ma'n H. Zawati**  
@MHZawati

"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." h

# Using TRUST as an Outline

## Establishing TRUST at ONC: Transparency

### Established

- Clear mandate to serve as a data repository over the long-term
- **User documentation** for ONC data access tools (Oceans 2.0 interfaces and APIs), data product descriptions, data citation overview, and more
- **Device workflows** are publicly available for all instrument deployments, maintenance and recovery processes
- **Data agreements** with data partners to outline attributions and restricting sensitive data access



## Establishing TRUST at ONC: Responsibility

### Established

- Metadata and data are curated by staff dedicated to specific instrument and source data types, following internally documented processes which are informed by community practices
- ONC's Oceans 2.0 digital infrastructure includes tools and APIs for **data discovery, access and annotation**.
- Data is served in various formats, including non-proprietary formats where possible.



## Establishing TRUST at ONC: User Focus

### Established

- **Expert guidance** via Board of Directors, International Science Advisory Board, and Ocean Observatory Council
- Consultations with stakeholders via an **annual survey** to gauge satisfaction and identify areas of improvement.
- On a more individual basis, data contributors and Principal Investigators provide their specific requirements and give feedback on resulting products and services.



## Establishing TRUST at ONC: Sustainability

### Established

- **Strong relationships** with funders and partners, which have secured funding and influenced ONC's services to align with stakeholder needs
- Provision of services to **transform data into information and knowledge** for evidence-based decision making (e.g., contribute data, process metadata, create metadata)

## Establishing TRUST at ONC: Technology

### Established

- **Oceans 2.0 infrastructure** to support metadata and data curation activities, as well as control and monitoring of cabled instruments. This infrastructure using Linux OS machines, Active MQ, software written mainly in Java, Matlab and Javascript, etc.
- Software tools for executing staff work such as Atlassian products (JIRA, Confluence, Bitbucket), Alfresco CMS, Splunk, Graylog, Prometheus, Slack, and many more.
- Procedures and agreements for **cyber-security** of our digital infrastructure, including the networks that support our cabled observatories

### In Progress

- Adoption of Insightly as a tool for managing contacts and contractual agreements
- Improving remote working capabilities for staff to efficiently execute work and communicate with one another
- Migration of Oceans 2.0 user interfaces to React javascript library
- **Replacement of hardware** nearing life expectancies



## iReceptor TRUST – Part 1

### • T for Technology:

- OpenStack cloud resources for deployment
- Docker containers for software deployment
- Compute Canada as a cloud service provider (reliability and longevity)
- Best practices for security, archival and backup
- Performance monitoring for service continuity
  - <https://sfu-ireceptor.github.io/ADC-API-Plots/ADC-API-Plots/index.html>

## iReceptor TRUST – Part 2

### • AIRR Community drives Transparency, Responsibility, and User Focus

- Standard terminology, file formats, query APIs
- How does **iReceptor** (or any group) provide more TRUST?

### • T for Transparency:

- **Curation process and policy:** [www.ireceptor.org/curation](http://www.ireceptor.org/curation)

### • R for Responsibility:

- **Data provenance:** [www.ireceptor.org/repositories/provenance](http://www.ireceptor.org/repositories/provenance)

### • U for User Focus:

- iReceptor Gateway as a **registry** for AIRR Data Commons

## iReceptor TRUST – Part 3

### • S for Sustainability - the hard part!

- iReceptor is a relatively **long-lived** research project
  - Under development since 2013, in production since 2014
- **Broad funding** over many years from CANARIE (x3), CFI, CIHR, EU H2020
- Broad **AIRR Community** support
- **International** user base (100s of users)
- **Relevant:** providing important **domain specific data for COVID-19** research
  - COVID-19 immunology papers cite iReceptor as source of data reuse (FAIR)

### • So what is the problem?

- **Sustainability** - We don't know what will happen when our current funding runs out!
- **Funding treadmill?** Community support? Commercial support? Institutional support?
- If we are truly **user focused**, sustainable should imply sustaining relevance
  - The platform/repository has to evolve as well...

Screenshots are from the slides of Brian Corrie, iReceptor and Reyna Jenkyns, Ocean Networks Canada (ONC), TRUST Principles In the Canadian Context, November 17, 2020

# 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

# 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.
- Endorsers, bloggers, tweeters, etc.