TRUST in Data

Dr. Dawei Lin, Ph.D.

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

Open Data under the COVID-19 pandemic, August 5th , 2020

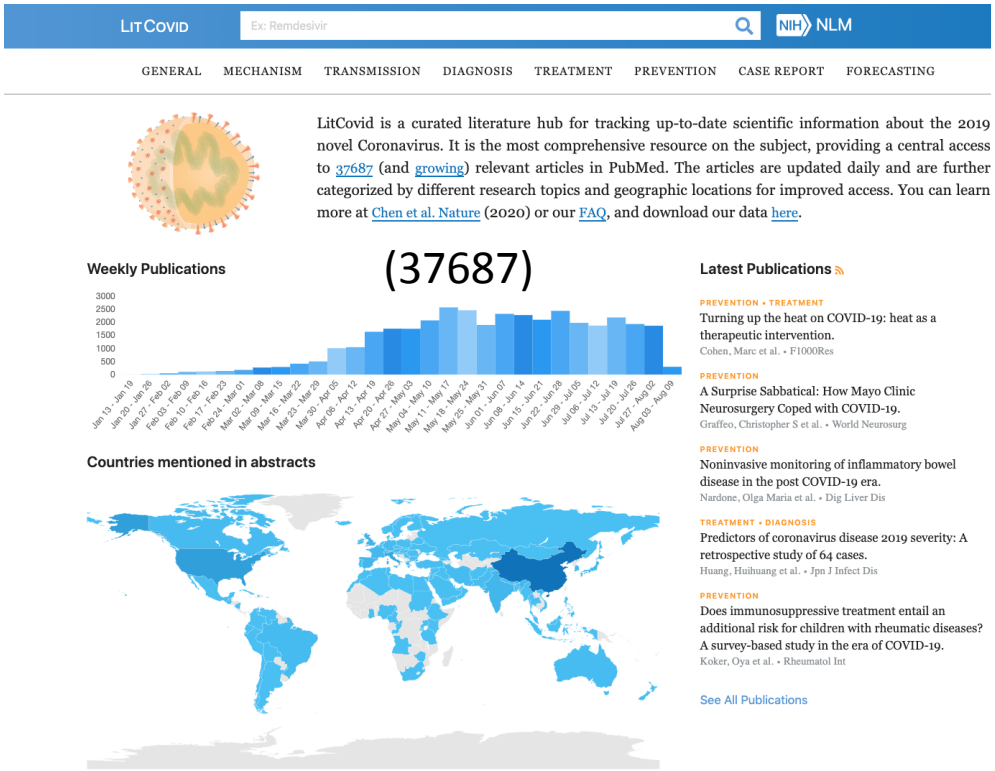
The Data Ecosystem for Open Science

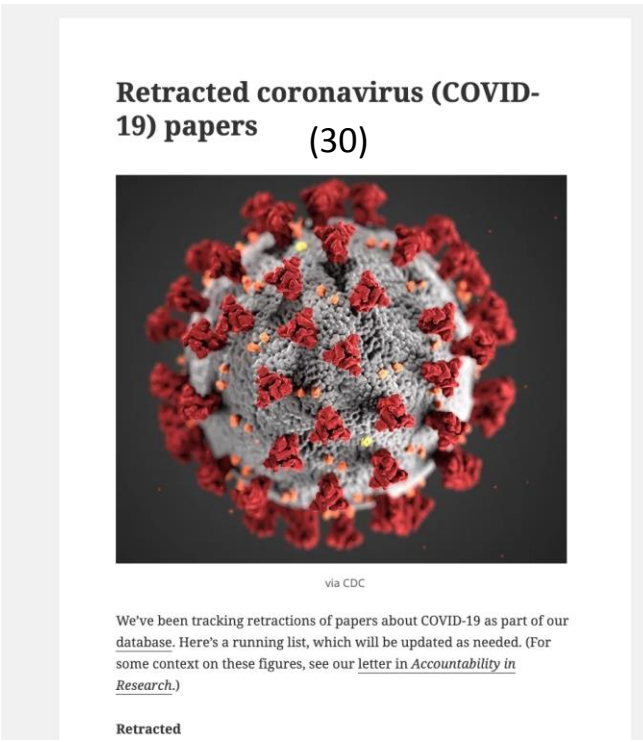
OPEN ***Science***

TRUST ***repository***

FAIR ***data***

Activities of COVID-19 Research





https://www.ncbi.nlm.nih.gov/research/coronavirus/

https://retractionwatch.com/retracted-coronavirus-covid-19-papers/

FERNANDEZ

3

Two elite medical journals retract coronavirus papers over data integrity questions

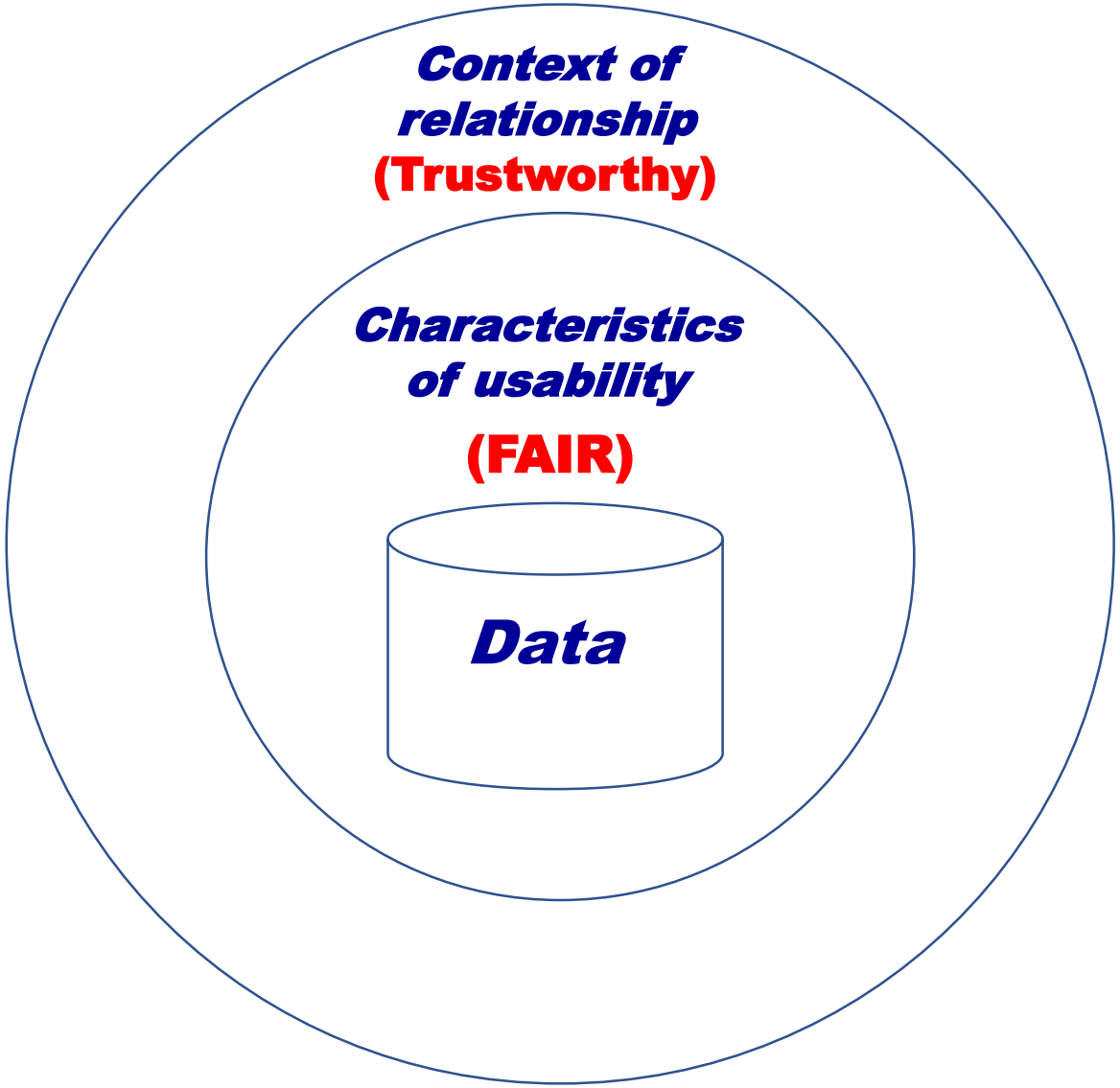
By Charles Piller, Kelly Servick | Jun. 4, 2020 , 5:30 PM

***Science*'s COVID-19 reporting is supported by the Pulitzer Center**.

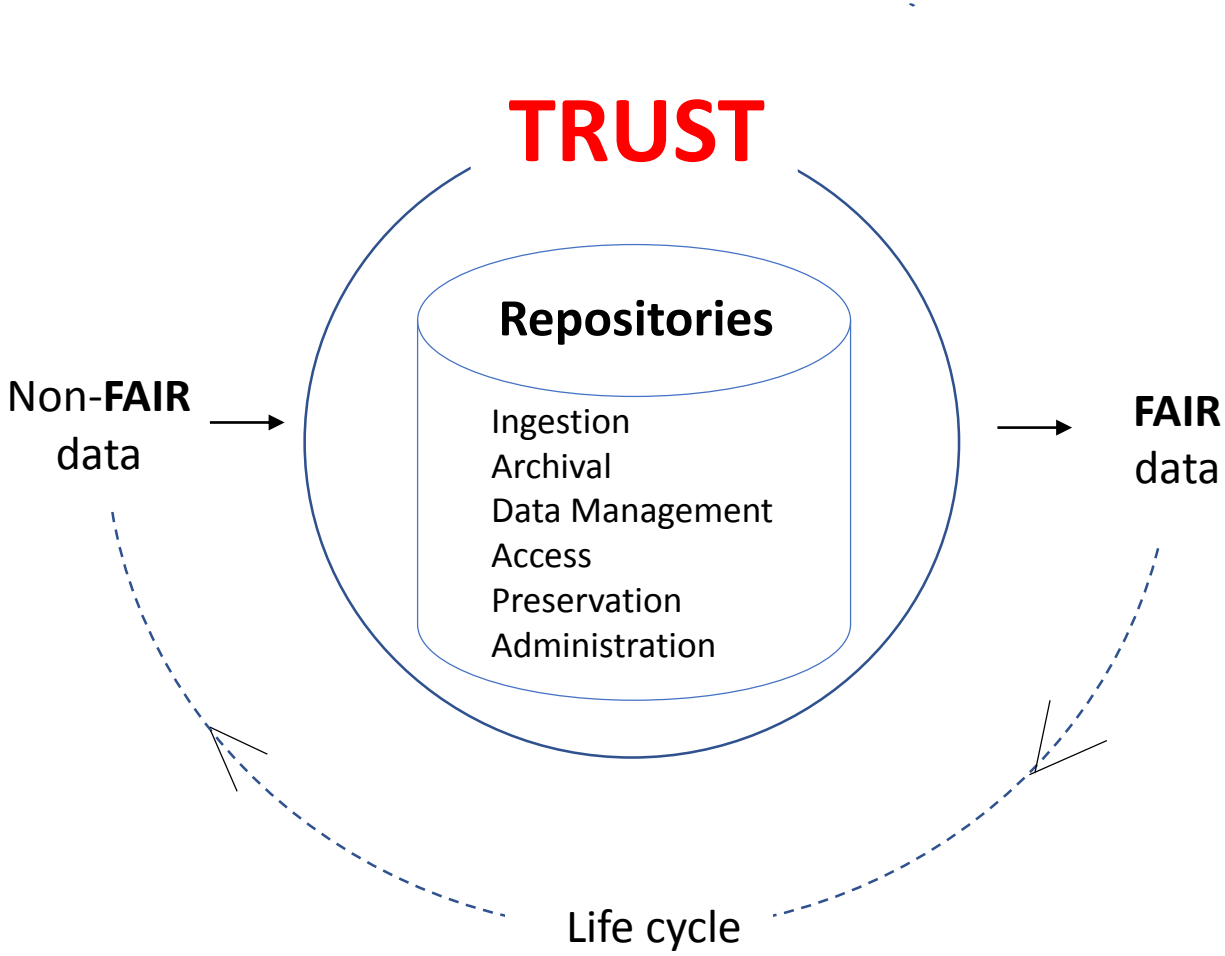
In the first big research scandal of the COVID-19 era, *The Lancet* and *The New England Journal of Medicine* (*NEJM*) today retracted two high-profile papers after a company declined to make the underlying data for both available for an independent audit, following **questions being raised about the research**. The *Lancet* paper, which claimed an antimalarial drug touted by President Donald Trump for treatment of COVID-19 could cause serious harm without helping patients, had had a global impact, halting trials of one of the drugs by the World Health Organization (WHO) and others.

Three authors on the *Lancet* paper requested the retraction, after initiating an independent review of the raw hospital patient data summarized and provided by Surgisphere, a small Chicago-based company operated by Sapan Desai, the fourth author of the study. Desai had previously said he and his co-authors—cardiologist Mandeep Mehra of Harvard University and Brigham and Women's Hospital, Frank Ruschitzka of University Hospital Zürich, and Amit Patel, an adjunct faculty member at the University of Utah—were getting such an audit of the data, but the agreement apparently fell apart.

Layers of Data Properties



The Relationship of TRUST and FAIR




T
R
U
S
T
01010
10101
01010
Transparency
Responsibilty
User focus
Sustainability
Technology

The TRUST Principles for Digital Repositories

Principle	Guidance for repositories
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.

https://www.nature.com/articles/s41597-020-0486-7

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.*

An RDA Community Effort

Endorsements

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

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

3. Data Sharing in Clinical Medicine

3.1 Focus and Description

Health care measures and clinical research are at the forefront of combating the COVID-19 pandemic. Promotion of clinical data sharing is of utmost importance because many studies and trials are performed under enormous time pressure, with weaknesses in the methodology (e.g. no control) and preliminary results published without any review. Sharing of data, and related documentation (e.g. protocols) will reduce duplication of effort and improve trial design, when many similar studies are being planned or implemented in different countries (Sharing and re-use of individual participant data from clinical trials: principles and recommendations, BMJ Open 2017). Clinical data outside clinical trials (e.g. case studies, descriptive cohorts of patients, etc.) may also be of high value and should be reported.

3.3 Policy Recommendations

3.3.1 Trustworthy Sources of Clinical Data

During a pandemic like COVID-19, it is important to concentrate efforts on scrutinising reliable data sources that provide data and metadata of high quality and guarantee the authenticity and integrity of the information. The recommendations are:

1. Measures should be taken in order to organise the transferral of data and trial documents to a suitable and secure data repository to help ensure that the data are properly prepared, available in the longer term, stored securely (with respect to access control, confidentiality, and integrity) and subject to rigorous governance. Repositories that explicitly support data sharing for COVID-19 trials should be announced.
2. Trustworthy repositories should be leveraged as a vital resource for providing access to and supporting the depositing of research data. However, as an emerging and evolving area in biomedical domains, trustworthiness assessment should not be limited to certification or accreditation (Consultative Committee for Space Data Systems, 2011; CoreTrustSeal Standards and Certification Board, 2019). A wide range of community-based standardised quality criteria, best practices, and principles (e.g. TRUST Principles (Lin et al., 2020)) should also be considered.
3. If analysis environments that allow in situ analysis of datasets are available, but prevent downloads, they should be provided to the end-user researchers in a pandemic situation, without fees if possible.
4. Tools allowing different datasets from different repositories to be analysed together on a temporary basis should be provided.

One Voice from Digital Repository Community

TRUST Principles

Guidance of

Implementations of

- TDR certifications standards
  (CoreTrustSeal, ISO 16363, ..)
- NIH Data Sharing Repository Criteria
- Data Preservation Consortium Criteria
- Elixir Core Data Resource Criteria
- Journal Recommended Repositories

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

Thank You

Email: dawei.lin@nih.gov