An Update of TRUST principles

Dr. Dawei Lin, Ph.D.

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

IG RDA/WDS Certifications of Digital Repositories
RDA Plenary 14 , October 25th, 2019

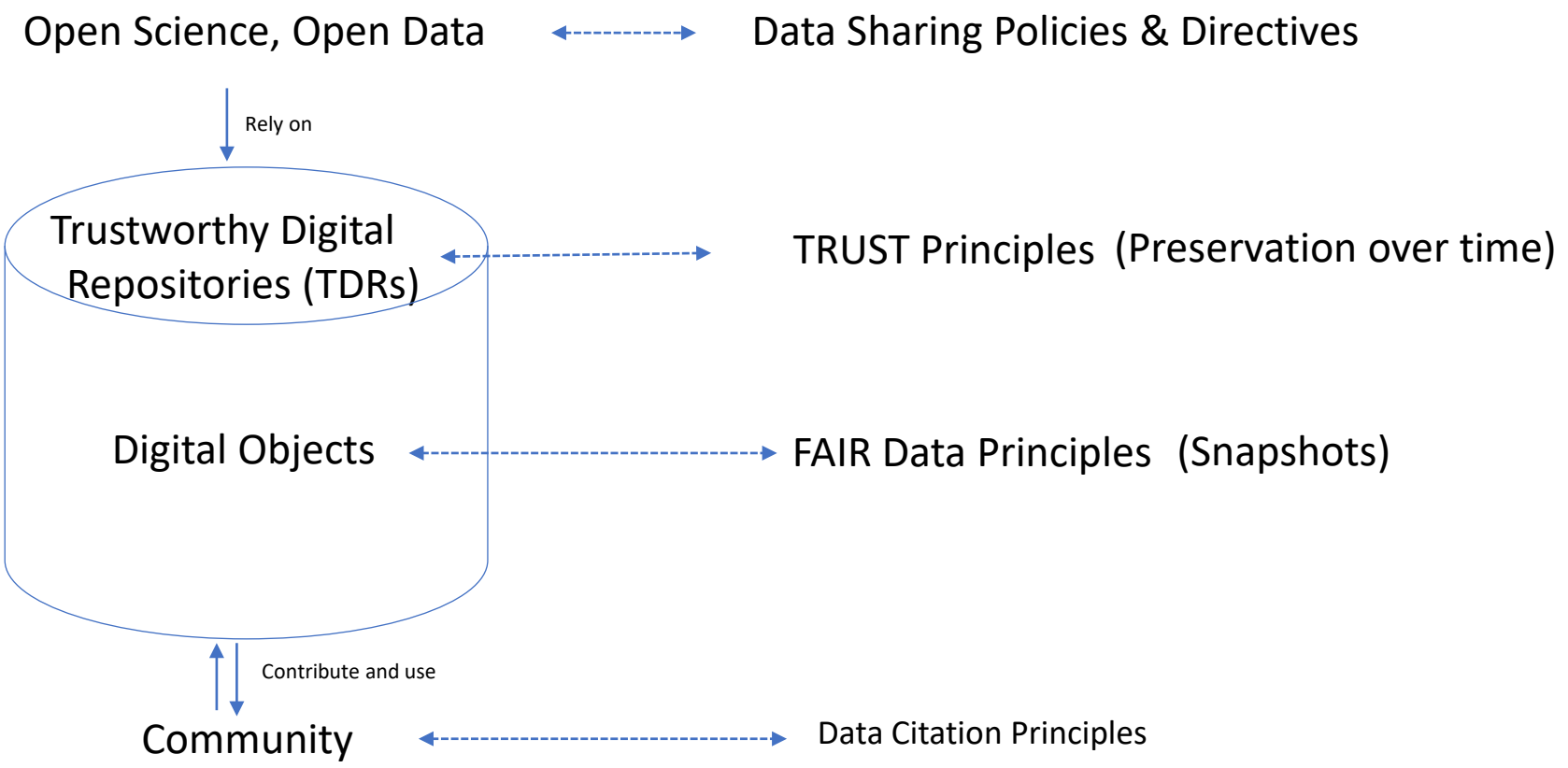
Open Science Needs Reliable Home for Data

- Dark Data
  - Only 12% of data described in published papers is in recognized repositories. [1]
- Data Repositories
  - Only 20% of 328 biomedical data repositories in the survey are still alive or rebranded after 18 years. [2]

Elixir Core Data Repository Criteria

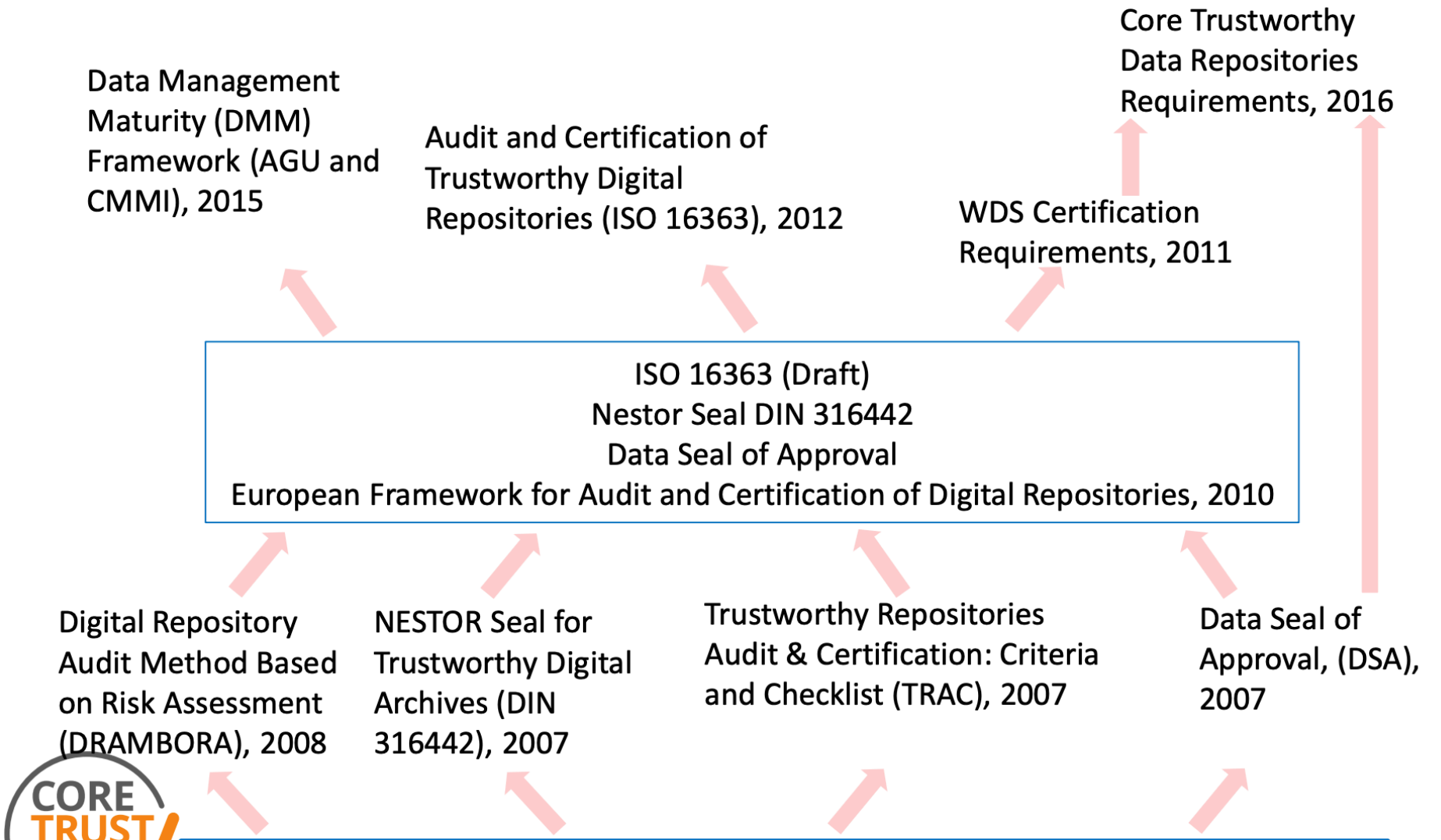
1. http://www.ncbi.nlm.nih.gov/pubmed/26207759
2. http://journal.embnet.org/index.php/embnetjournal/article/view/803/1209

The Data Ecosystem Needs TDRs



Open *Science*, TRUST *repository*, FAIR *data*, CARE *community*

Trustworthy Data Repository Certification Standards: An Evolutionary Perspective



CORE TRUST SEAL

Source: Downs, 2019

Motivation of TRUST Principles

- Draw attention to the role of the repository in data life-cycle
- Develop concise and measurable approaches to achieve Trustworthiness of Digital Repositories (TDR) for long-term preservation and a sustainable operation
- **Not** to replace any standards, criteria, or best practices
- Provide a high-level starting point for advocating, supporting and implementing all certifications and assessments – a messenger

17 co-authors from 8 Countries for V0.02

- Dawei Lin, Division of Allergy, Immunology, and Transplantation (DAIT), National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Maryland, USA.
- Jonathan Crabtree, The Odum Institute, University of North Carolina at Chapel Hill, North Carolina, USA.
- Ingrid Dillo, Data Archiving and Networked Services (DANS), The Netherlands.
- Robert R. Downs, Center for International Earth Science Information Network (CIESIN), The Earth Institute, Columbia University, New York, USA.
- Rorie Edmunds, World Data System of the International Science Council (WDS), WDS International Programme Office, Tokyo, Japan.
- Marisa De Giusti, Universidad Nacional de La Plata, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, Argentina.
- Hervé L'Hours, UK Data Archive, UK Data Service, University of Essex, UK
- Wim Hugo, South African Environmental Observation Network, Cape Town, South Africa
- Reyna Jenkyns, Ocean Networks Canada, University of Victoria, Canada.
- Varsha Khodiya, Springer Nature, UK
- Maryann E. Martone, Ph. D. University of California, San Diego and SciCrunch Inc., USA
- Mustapha Mokrane, Data Archiving and Networked Services (DANS), The Netherlands.
- Vivek Havale, Center for Information Technology, National Institutes of Health, Bethesda, Maryland, 20817, USA.
- Jonathan Petters, Data Services, University Libraries, Virginia Tech, Blacksburg, VA 24061, USA.
- Barbara Sierman, KB National Library of the Netherlands
- Dina V. Sokolova, University Libraries, Columbia University, New York, USA
- Martina Stockhause, German Climate Computing Center (DKRZ), Germany

The TRUST Principles

- **T - Transparency** is achieved by providing publicly accessible evidence of the services that a repository does and does not offer.
- **R - Responsibility** is a commitment to provide reliable data services.
- **U - User community** is a commitment to implement and enforce the standards and norm of the user community.
- **S - Sustainability** is the capability to support long-term data preservation and use.
- **T - Technology** is the infrastructure and capabilities to support the repository operations.

TRUST principles White Paper V0.02 http://bit.ly/trustprinciples

T – Transparency

Transparency is achieved by providing publicly accessible evidence of the services that a repository can and can not offer.

Key concepts

- Clear mission statement
- Policy and governance
- Specific claims of capability

R - Responsibility

Responsibility is a commitment to provide high quality technical data services.

Key concepts

- Responsibility is shared (data producers, institutions, funders, repositories, publishers, and data users).
- Process and workflow documentations to ensure data and service quality.
- Guidance to content deposition and prospective users

U – User Community

User community is a commitment to implement and enforce the standards and norm of the user community

Key concepts

- Key user base – designated community
- Monitoring need changes overtime
- Enforcement of community norms and standards

S – Sustainability

Sustainability is the capability to support long-term data preservation and use.

Key concepts

- Plans for risk mitigation, business opportunity, disaster discovery and succession.
- Plans for long-term data preservation
- Credible contingency plans supported by public evidence such as certifications.

T – Technology

Technology is the infrastructure and capabilities to support the repository operations.

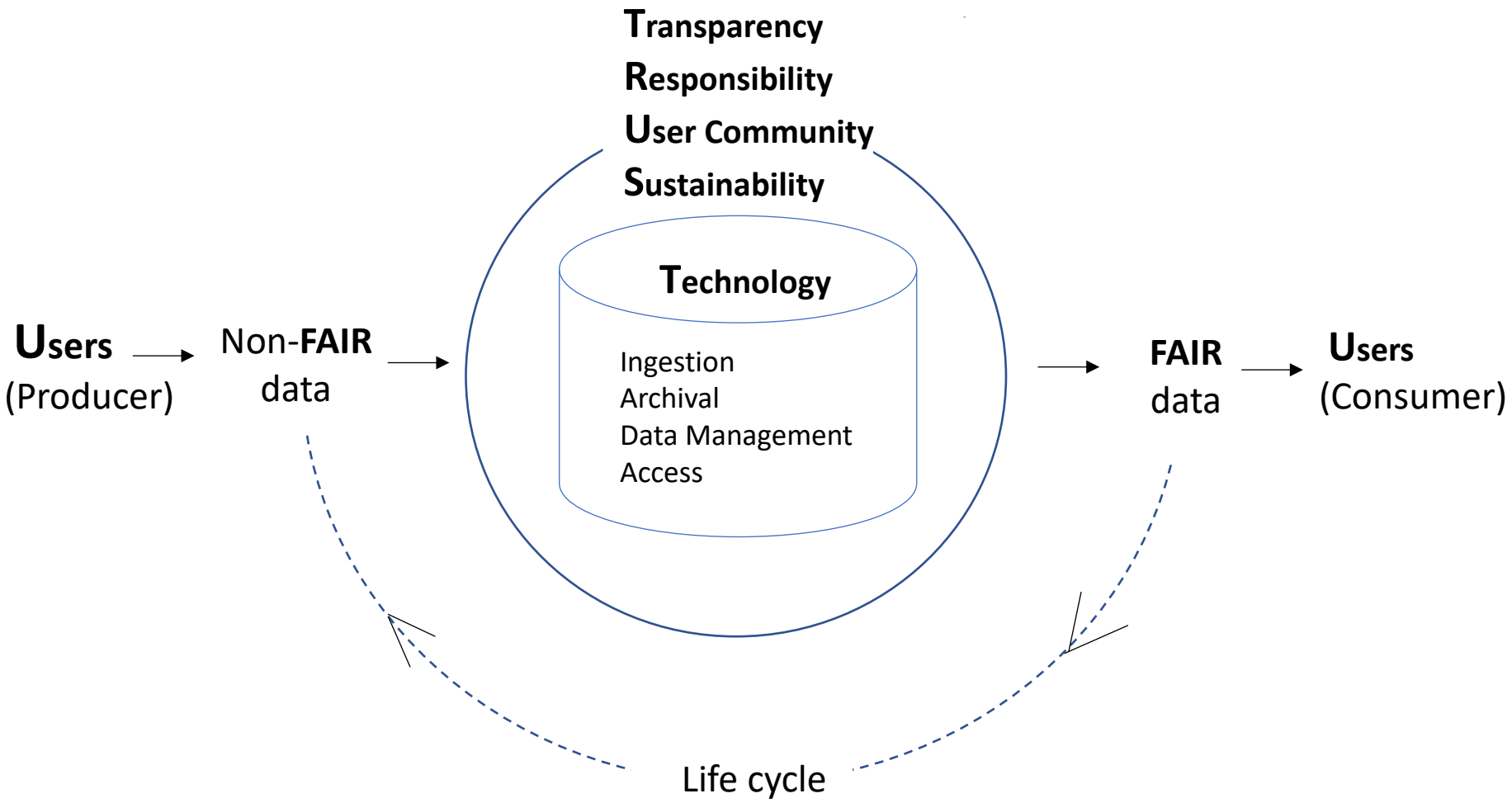
Key concepts

- 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

Road Map

- Mappings of the TRUST principles to a range of relevant standards, criteria, and practices. Such as:
  - OAIS (ISO14721)
  - ISO16363
  - CoreTrustSeal
  - Criteria for Elixir Core Data Resources
  - …….
- Write a formal publication
- Solicit endorsements

Questions/Discussions

- Does principles apply or not to your domain?
- Are any key principles missing?
- What information need to make the principle endorsable?

Functional Relationship of TRUST and FAIR