# Implementation Pathway: TRUST ISO 16363/ISO 16919 intersections

Dr David Giaretta Director of PTAB <u>david@giaretta.org</u> <u>http://www.iso16363.org</u>

## TRUST

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.

## TRUST

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.

## TRUST, but VERIFY

• A Russian proverb used by President Reagan when talking about nuclear disarmament.

- While "trust, but verify," at times, can be an essential approach, often it's detrimental. Effective trust-building and leadership practices require knowing when and why to use it.
- <u>When the outcome is essential and matters more than the relationship,</u> <u>use "trust, but verify."</u> When the relationship matters more than any <u>single outcome, don't use it. https://www.psychologytoday.com/gb/blog/trust-the-new-workplace-</u> <u>currency/201507/the-problem-trust-verify-approach</u>

## The Devil is in the Details

It must be clear what the repository is claiming to do (following OAIS):

- Who is the repository guaranteeing will be able to understand/use what is being preserved? Clearly that cannot be everyone! What is its Designated Community?
  - Are such people able to understand/use the information? **Testable**.
- Usability clarified by being specific about "Preservation Objectives"
  - What should the Designated Community be able to do? Testable
- Authenticity
  - What evidence can be provided to support claims of Authenticity? **Checkable**
- Integrity
  - Fixity if bits not changed
    - Fixity Information. Testable
  - Transformation Information Properties if the bits are changed
    - Testable
- Other things are subjective and depend on the judgement of the auditors
   Who does the audits?
  - How can we be sure they are done consistently?

# Why ISO?

- ISO audits are used across the World and in vast numbers of areas on which our lives depend.
- The ISO process ensures international consistency of certification and their international recognition
- Everyone at every level is tested/evaluated every year

## **Relationship between standards**



## ISO 17021

- ...specifies requirements for bodies providing audit and certification of management systems.
- ... gives generic requirements for such bodies performing audit and certification in .....
- Observance of these requirements is intended to ensure that certification bodies operate management system certification in a competent, consistent and impartial manner, thereby facilitating the recognition of such bodies and the acceptance of their certifications on a national and international basis.
- ..... serves as a foundation for facilitating the recognition of management system certification in the interests of international trade.
- Certification of a management system provides independent demonstration that the management system of the organization:
- a) conforms to specified requirements;
- b) is capable of consistently achieving its stated policy and objectives;
- c) is effectively implemented.

Conformity assessment, such as the certification of a management system, thereby **provides value to the organization, its customers and interested parties**.

# ISO 17021 Principles for inspiring confidence include

- impartiality;
- competence;
- responsibility;
- openness;
- confidentiality;
- responsiveness to complaints;
- risk-based approach.

## **Auditor behaviour**

- a) ethical, i.e. fair, truthful, sincere, honest and discreet;
- b) open-minded, i.e. willing to consider alternative ideas or points of view;
- c) diplomatic, i.e. tactful in dealing with people;
- d) collaborative, i.e. effectively interacting with others;
- e) observant, i.e. actively aware of physical surroundings and activities;
- f) perceptive, i.e. instinctively aware of and able to understand situations;
- g) versatile, i.e. adjusts readily to different situations;
- h) tenacious, i.e. persistent and focused on achieving objectives;
- i) decisive, i.e. reaches timely conclusions based on logical reasoning and analysis;
- j) self-reliant, i.e. acts and functions independently;
- k) professional, i.e. exhibiting a courteous, conscientious and generally business-like demeanour in

the workplace;

- I) morally courageous, i.e. willing to act responsibly and ethically even though these actions may not
- always be popular and may sometimes result in disagreement or confrontation;
- m) organized, i.e. exhibiting effective time management, prioritization, planning, and efficiency.

## **ISO Standards – who can audit?**

METRIC

OAIS – ISO 14721
ISO 16363
ISO 27001
ISO 15489

STANDARD FOR REQUIREMENTS FOR AUDITORS

- None
- •ISO 16919
- •ISO 19896
- None

#### **ISO 16363**

## Audit and Certification of Trustworthy Digital Repositories

Designed for audit – self audit and independent auditors

- Hierarchy of metrics to make the auditor look at more and more specific details when required
- Metrics and their structure:
  - Statement of requirement
  - Supporting text
  - Examples of Ways the Repository can Demonstrate it is Meeting this Requirement
  - Discussion
- NUMBER of metrics at each level

Metrics⊡	Top level X	Metric X.X	Sub- metric X.X.X	Sub- sub metric X.X.X.X	Sub-sub- sub metric X.X.X.X.X
Organisational Infrastructure	1	6	21	31	31
Digital Object Management	1	7	36	62	67
Infrastructure and Security Risk Management	1	3	9	16	27
TOTAL	3	16	66	109	125



## **Examples of metrics**

#### 1 ORGANIZATIONAL INFRASTRUCTURE

#### 1.1 GOVERNANCE AND ORGANIZATIONAL VIABILITY

1.1.1 The repository shall have a mission statement that reflects a commitment to the preservation of, long term retention of, management of, and access to digital information.

1.1.2 The repository shall have a Preservation Strategic Plan that defines the approach the repository will take in the long-term support of its mission.

1.1.2.1 The repository shall have an appropriate succession plan, contingency plans, and/or escrow arrangements in place in case the repository ceases to operate or the governing or funding institution substantially changes its scope.

1.1.2.2 The repository shall monitor its organizational environment to determine when to execute its succession plan, contingency plans, and/or escrow arrangements.

1.1.3 The repository shall have a Collection Policy or other document that specifies the type of information it will preserve, retain, manage, and provide access to.

## **ISO Accreditation and Certification**



## **PTAB Process following ISO 17021**



## **TRUST PRINCIPLES ARE A START**

• Very easy to claim that one's repository follows the TRUST principles

BUT

- the details matter
- the auditors matter
- the process matters

Trust but Verify - because our digitally encoded intellectual capital is important

## References

- ISO/IEC 17021-1:2015 Conformity assessment -- Requirements for bodies providing audit and certification of management systems -- Part 1: Requirements, available from <u>http://www.iso.ch</u>
- Audit and Certification of Trustworthy Digital Repositories. Magenta Book. Issue 1. September 2011., available from <u>https://public.ccsds.org/Pubs/652x0m1.pdf</u> also known as ISO 16363:2012
- Reference Model for an Open Archival Information System (OAIS). Magenta Book. Issue 2. June 2012, available from <u>https://public.ccsds.org/Pubs/650x0m2.pdf</u> also known as ISO 14721:2012,
- Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories. Magenta Book. Issue 2. March 2014, available from <a href="https://public.ccsds.org/Pubs/652x1m2.pdf">https://public.ccsds.org/Pubs/652x1m2.pdf</a> also known as ISO 16919:2014

## **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:



# BREAK

TRUST Principles Mini Symposium | July 7th, 2020





Protein Data Bank A Community Archival Data Repository Example





John Westbrook RCSB PDB

#### rcsb.org

#### Outline

- Brief overview of the PDB resources and services
- PDB support of the TRUST principles
- Some challenges for repository maintainers



### **Protein Data Bank Established 1971**

- PDB 1<sup>st</sup> Open Access digital data resource in all of biology
- Founded with 7 X-ray crystal structures of proteins
- Single global archive for protein and DNA/RNA experimental structures
- Today, Open Access to >165,000 structures
- wwPDB collaboration US (RCSB PDB), EU (PDBe), and Japan (PDBJ), and BMRB (EMDB to join early 2020)



Structures that Inspired Launch of the PDB

#### wwPDB Partnership Established 2003

- RCSB PDB, PDBe, and PDBj (all locally funded) manage PDB according to FAIR
- Collaborate on Data In



- wwPDB governed by Memorandum of Understanding mandating
  - Policies/Procedures
  - No charge for Data Deposition
  - No charge for Data Download
- RCSB PDB, PDBe, and PDBj each provide open access to complementary views of identical PDB Data for Public/Industry Research and Education

RCSB PDB is Archive Keeper



# TRUST

### Transparency

What do you do? How do you do it? Can I get a copy? Is what you do fit for purpose?



#### rcsb.org

### **Articulation of Mission and Scope**

- wwPDB maintains a single archive of macromolecular structural data that are freely and publicly available to the global community
- wwPDB maintains these organizational details on the wwpdb.org resource portal along with other key project documentation



#### Licenses and Terms of Use

- PDB primary data are free of all copyright restrictions and made fully and freely available for both non-commercial and commercial users
- This PDB license pre-dates contemporary open source licenses
- Some additional conditions on adaptation of data protect authenticity of repository data files
- Compliance issues with primary data are rare
- Other PDB software and educational materials are provided under standard open source licenses (e.g., Apache and Creative Commons)

#### **CoreTrustSeal Certification**

- Strong commitment and tradition within our scientific community for support of data and process standards
- Expectations of both our repository contributors and users to adopt and maintain best practices in archiving and data management
- Increasing focus of funders on supporting FAIR data management practices
- Certification documents the resource investment required to responsibly manage the full life cycle of archival data
- Relatively low barrier and modest effort certification process
- Good balance between rigor and certification effort



# TRUST

## Responsibility

What data do you have? What do you call it? Who says the data are any good? How can I find it? Can my program read and process it?



#### rcsb.org

### **Data Integrity and Authenticity**

- Community metadata and data standards
- Metadata and data change management policies
- Expert biocuration
- Maintaining consistency through retrospective repository remediation
- ORCID identification for depositors

### **Data Quality Standards**

- All PDB deposition, validation and biocuration tools support and enforce Community data standards
- PDB provides validation reports describing compliance with Community data quality standards
- Validation reports tailored for a range of audiences: depositors, editorial reviewers, and general users
- PDB validation reports required by most scientific journals describing 3D structure determinations

### **Data Appraisal**

- Pre-deposition validation services
- wwPDB OneDep Deposition System
- Expert biocuration
- Data and data diagnostics delivered in well-defined community data formats



#### **Expert Community Guidance**

#### wwPDB Method-specific Community Task Forces

Task Force	Meeting	Chair(s)/Membership	Outcomes
X-ray Validation Task Force	2008 2015	Randy Read (Univ of Cambridge) 17 members	(2011) Structure 19: 1395-1412
NMR Validation Task Force	2009- 2019	Gaetano Montelione (Rutgers) Michael Nilges (Institut Pasteur) 10 members	(2013) Structure 21: 1563-1570
3DEM Validation Task Force	2010	Richard Henderson (MRC-LMB) Andrej Sali (UCSF) 21 members	(2012) Structure 20: 205-214
Small-Angle Scattering Task Force	2011 2014	Jill Trewhella (Univ Sydney) 6 members	(2013) Structure 21: 875-881 (2017) Acta Cryst D73
Hybrid Methods Task Force	2014	Andrej Sali (UCSF), Torsten Schwede (Univ Basel), Jill Trewhella (Univ Sydney) 27 members	(2015) Structure 23: 1156-1167
Ligand Validation Workshop	2015		(2016) <i>Structure</i> 24: 502-508
PDBx/mmCIF Working Group	2011 -	Paul Adams (LBL) 13 members	Regular virtual meetings and workshops







#### Data Discovery, Identification, and Integrated Access





# TRUST

## **User Focus**

Who is using your data? How do you know (not just clicks and downloads)? Is anything useful being done with your data? What are you doing to make that easier?



#### rcsb.org

#### **RCSB PDB Services Support Full Data Lifecycle**



Burley et al. (2019) Nucleic Acids Research 47, D464-D474.

#### Impact on US FDA New Drug Approvals<sup>1</sup>



- 1. Westbrook & Burley (2019) Structure 27, 211-217
- 2. Galkina Cleary et al. (2018) PNAS 115, 2329-2334; Value in 2016 US\$
# Impact on Anti-neoplastic Drug Approvals<sup>1</sup> 79 NEW ANTI-CANCER DRUGS Approved 2010-2018 74 of these new drugs had a total of **2412** unique structures in the PDB explaining target biology and facilitating discovery/development Structure-guided drug discovery $\Box >70\%$ of small-molecule

1. Westbrook *et al.* (2020) *Drug Discovery Today* 25, in press

#### **Supporting Reusability and Extensibility**

- Data and metadata requirements for deposition
- Content and format extensibility
- Maintaining repository content and format consistency through retrospective biocuration
- Repository metadata and data content documentation (mmcif.wwpdb.org)



#### **Outreach and Educational Resources**

03/25

PDB-101	Molecule of the Month	- Browse	Learn -	Global Health -	Teach +	SciArt -	Events+	More -	-
PDB-101 Molecular explorations						Search Molecule of the Month articles and more Go			
Educational portal of TRDR						Celebrating 20 YEARS OF Molecule of the Month			
Molecule of the Month						By Ca	ategory	By Date	By Title

#### **Coronavirus Proteases**

Coronavirus proteases are attractive targets for the design of antiviral drugs.

In this work of that and easy travel, enwaging viruses are increasingly becoming a major darget to work health. Concreatives are a notable example. Particularly virunits forms have enverged from their natural animal hosts and pose a threat to human communities, for 2503, the BARS virus energied in China from bait populations, moving to civits and finally to humans. They ears later, the MERS virus also enverged from baits, transferring in the Middle East to dromadary camels and their to humans. Recently, another community a fait enregied in China to ywo of animism in a five market. Structural bodogi is helping us undestituat these dangenous flews, and hopefully will help us develop new ways to fight them.

#### Coronavirus Code

Consistivations contain a genome composed of a long RNA strind—one of the largest of all RNA viruses. This genome active statilise a messanger RNA when it indices a cell, and directs the synthesis of two long polyproteins that include the machinery that the virus midda to replicate new viruses. These profession includes a replication/themosynthese complex that makes more RNA, serviced structural proteins that includes a replication/themosynthese or professes. The professes give essential roles in outting the polyproteins into all of times functional pieces.

#### Main Protease

The main protesse of coronavirus makes most of these cuts. The one shown here (PDB entry 5iii) is from the SARS-CaV-2 (2019-rCoV) coronavirus that is currently posing dangers in WMass. It is a dimer of the isolatical subcutinitis that together form two active sites. The protein fold is smith to service protesses like trypsin, but a cysterine amino acid and a nearby histidine perform the protein-cutting reaction and an extra domain stabilizes the dimer. This starturus heas appoltai-like inhibitor bound in the active site.

SARS-CoV-2 (2019 mCoV) coronavitus main p Downood high quality TVP image (A

#### COVID-19/SARS-CoV-2 Resources



Coronavirus CellPAINT Contest Winners in Art and Science at PDB-101

#### PDB Structures (as of July 1, 2020)

#### Access all SARS-CoV-2 PDB structures

#### Table of PDB structures complexed with Ligands of Interest (LOI); LOI Molecular Weight; LOI InCHIKey (TSV)

- · New this week
- Main protease PDB structures
- · Spike protein and spike receptors
- Papain-like SARS-CoV-2 structures
- · Other SARS-CoV-2 structures
- · PanDDA analysis structures and main protease with unliganded active site
- · Contains SARS-CoV-2-reactive human antibody



# truSt

#### **Sustainability**

Who is looking after things? Will that last?



#### rcsb.org

#### **Organizational Infrastructure**

- Regional wwPDB partner data centers
- Global load-balancing and failover of deposition services
- Complimentary data access services



Regional partners responsible for data from: RCSB PDB (US): Americas and Oceania PDBe (UK): Europe and Africa PDBj (Japan): Asia and Middle East

#### **Continuity of Access**

- 40+ year track record of funding support in US
- wwPDB organization provides for continuity of data and service access if a regional partner site should become unavailable
- Versioned repository delivery
- Snapshotting of primary data





Regional partners responsible for data from: RCSB PDB (US): Americas and Oceania PDBe (UK): Europe and Africa PDBj (Japan): Asia and Middle East



# TRUS

## Technology

Are you using cool tools? Are they cool enough? Are you in the cloud? Can you get hacked?



#### rcsb.org

#### **Technical Infrastructure**

- Data reference standards and ontologies in use
  - Lengthy and requiring consolidation from many sources
- Automated software development and deployment process
- Managing community software tools
- Workflow driven data processing and infrastructure management
- Capacity monitoring and management tools



INTERNATIONAL UNION OF BIOCHEMISTRY AND MOLECULAR BIOLOGY



#### Security

- Service availability, redundancy, disaster recovery
- Institutional security protocols and resources
- Application security protocols
  - Coding standards
  - Code review
  - Testing and deployment protocols
  - Version control





## **Challenges from a Domain Perspective**



#### rcsb.org

#### FAIR and FACT are in our GENES

# Findability (phenylalanine) + Accessibility (alanine) + Interoperability (isolucene) + Reusability (arginine)

FQRSTPAITLESPDIKYPLRLIDREIISHDTRRFRFALPSPQHILGLPVGQHIYLSARIDGNLVVRPYTPISSDDDKGFVDLV IKVYFKDTHPKFPAGGKMSQYLESMQIGDTIEFRPSGLLVYQGKGK<mark>FAIR</mark>PDKKSNPIIRTVKSVGMIAGGTGITPMLQVI RAIMKDPDDHTVCHLLFANQTEKDILLRPELEELRNKHSARFKLWYTLDRAPEAWDYGQGFVNEEMIRDHLPPPEEEPLVLMC GPPPMIQYACLPNLDHVGHPTERCFVF

Human Erythrocyte NADH-cytochrome b5 Reductase (PDB 1UMK UniProt P00387)

# Fair (phenylalanine) + Accurate (alanine) + Confidential (cysteine) + Transparent (threonine)

gpgpeyqaqglamylqengidcpkckfsyalarggcmh<mark>FACT</mark>qcrhqfcsgcynafyaknkcpepncrvkkslhghhprdc lfylrdwtalrlqkllqdnnvmfnteppagaravpgggcrvieqkevpnglrdeacgketpagyaglcqahykeylvslinah sldpatlyeveeletaterylhvrpqplagedppayqarllqklteevplgqsiprrk

Human E3 ubiquitin-protein ligase RNF31 (PDB 4LJP UniProt Q96EP0)

#### ... But Only TRST is Common

Transparency (threonine) + Responsibility (arginine) + User Focus (selenocysteine) + Sustainability (serine) + Technology (threonine)

SMLKRLSTEEATRWADSFDVLLSHKYGVAAFRAFLKTEFSEENLEFWLACEEFKKTRSTAKLVSKAHRIFEEFVDVQAPRE VNIDFQTREATRKNLQEPSLTCFDQAQGKVHSLMEKDSYPRFLRSKMYLDLLS Human Regulator of G-protein signaling RGS8 (PDB 5DO9 UniProt P57771)

We all need to work harder to master - User Focus -

#### **RCSB PDB Team**

#### Funding

RCSB PDB is funded by the National Science Foundation (DBI-1832184), National Institute of General Medical Sciences, National Institute of Allergy and Infectious Disease,

and National Cancer Institute, (R01GM133198), and the US Department of Energy (DE-SC0019749)

Follow us f У 🕨 💭



#### Management

RCSB PDB is hosted by:

RUTGERS



SDSC SAN DIEGO SUPERCOMPUTER CENTER

UC San Diego



RCSB PDB is a member of the Worldwide Protein Data Bank partnership (wwPDB; **wwpdb.org**)



#### **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:



# Funder Story: A Use Case

#### Presented by Mark Leggott



## CANARIE RDM Funding (2018 Program)

- CANARIE Research Data Management Funding Program Launched 2018
- Funding call based on Community Consultation
- Identified 8 key "gaps and opportunities" in the ecosystem
- Also highlighted adherence to
  - the FAIR Principles
  - the National Data Services Framework (NDSF)
- Applicants were required to address one or more of the gaps, as well as FAIR and NDSF
- 9 Projects selected, completed in March 2020

#### **RDM Call 1 Guidance**



## **RDM Call 1 Projects**





















#### **RDM Call 1 Project Intersections**





## CANARIE RDM Call 2 (2019)

- Call 2 launched late 2019
- Built on parameters from Call 1, reinforcing FAIR+NDSF focus
- Added requirement to create/enhance interoperability between multiple disparate systems
- Encouraged integration between research platforms and repositories
- 4 projects selected, development Apr 2020-Mar 2022



#### Research Data Canada's NDSF

- National Data Services Framework
  - a conversation with all stakeholders at all levels;
  - agreement on best-practices, standards and protocols;
  - a suite of interoperable services and resources.
- Summits held 2017, 2019, 2020
- Facilitates the development of national research infrastructures
- Intersects with CANARIE RDM funding, as well as efforts of the emerging New Digital Research Infrastructure Organization (NDRIO), other Canadian funders and data management/research software organizations

#### **Building Synergies with National Efforts**



#### **Evolution Towards TRUSTed Services**

- Goal is to continue to evolve the national conversation, and work with funders to integrate support for TRUST and national data repositories into ongoing efforts
- Portage Data Repositories Expert Group is working on a multi-stakeholder proposal (e.g. Portage, NDRIO, RDC, WDS-ITO) to define a new call for support and funding



#### **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:



# **Stakeholder Panel**

Moderated by Mustapha Mokrane

Sustainabili

User focus

Responsibility

10101

Technology

Research Data Canada – Données de Recherche Canada

Panelists:

- TRUST author Robert Downs
- Funder Mark Leggott
- Research Community Shelley Stall
- Repository Manager John Westbrook
- Publisher Varsha Khodiyar

# **Stakeholder Panel**

# **Question 1:**

# What would you see as the main challenges for implementation within your stakeholder community?



# Stakeholder Panel Question 2:

What would you consider a sign of success and what would be the impact of the TRUST Principles in the short and longer term?

TR

UST

01010

Technolog

Research Data Canada – Données de Recherche Canada

# **Stakeholder Panel**

## **Question 3:**

# How can we as a community help the TRUST Principles succeed?



# Wrap-up and Ways Forward

#### Any remaining questions?

- 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:



# Thank you!

#### For further information:

- Read 'The TRUST Principles for digital repositories' in *Nature Research's Scientific Data*: <u>https://www.nature.com/articles/s41597-020-0486-7</u>
- DOI: <u>https://doi.org/10.1038/s41597-020-0486-7</u>
- For further details and endorsements, see the RDA website: <u>https://www.rd-alliance.org/rda-community-effort-trust-principles-digital-repositories-0</u>
- Find this slide deck at <a href="https://bit.ly/TRUSTSymposium">https://bit.ly/TRUSTSymposium</a>









Research Data Canada – Données de Recherche Canada