RDA FAIR data maturity model

11th of November 2020

RDA Plenary

Agenda

- Welcome and introduction to the meeting [5 min]
- History of the working group [10 min]
- Survey on bridging the gap between funders and communities [5 min]
- Presentations and discussions [60 min]
- Outline of the maintenance phase [5 min]
- Closing [5 min]

Tour de table

In the chat window, please type…

- Your name
- Your affiliation
- Your role
  - Researcher
  - Librarian
  - Service provider
  - Policy maker
  - Funder

Context

The principles are **NOT** strict

- **Ambiguity**
- Wide range of **interpretations** of FAIRness

Different **FAIR Assessment** Frameworks

- Different metrics
- No comparison of results
- No benchmark

SOLUTION is to bring together **stakeholders** to build on **existing approaches** and **expertise**

- Set of **core assessment criteria** for FAIRness
- FAIR **data maturity model & toolset**
- FAIR data **checklist**
- RDA recommendation

Join the **RDA** Working Group: [RDA WG web page]

Scope

BUT the Working Group does **NOT** have the purpose to …

- **develop yet-another-evaluation-method**: the core criteria are intended to provide a common 'language' across evaluation approaches, not to be applied directly to datasets.
- **define how the core criteria need to be evaluated**. The exact way to evaluate data based on the core criteria is up to the owners of the evaluation approaches, taking into account the requirements of their community
- **revise and re-design the FAIR principles**

History of the FAIR data maturity model WG

2019

- Establishement of the RDA Working Group with a clear mandate
- Definition of the scope and the methodology
- Landscaping exercice comparing all publicily availble (FAIR) assessments tools and methodologies
- Draft set of indicators based on the decomposition of the FAIR principles
- Improvement of the indicators based on iterations and addition of priorities
- Birth of the FAIR data maturity model

History of the FAIR data maturity model WG

2020

- Testing phase of the indicators against data sets and methodologies which led to a revision of the FAIR data maturity model
- Publication of the FAIR data maturity model as an RDA Recommendation
- Turning the Working Group into maintenance mode
- Investigating maintenance and governance aspects
- Identification of topics to be addressed for the next version of the FAIR data maturity model
- Survey to bridge the gap between funders and communities on the understanding of FAIR assessments
- Retrospective publication in the Data Science Journal on the outcome of the FAIR data maturity model

Welcome

State of play

- Publication of the FAIR Data Maturity Model: Specification and Guidelines as an RDA recommendation, 25 June 2020
- EOSC-SYNERGY Intermediate report on technical framework for FAIR principles implementation, 2 Sept. 2020
- Publication of The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments as a paper in the Data Science Journal, 27 Oct. 2020

Roadmap (remainder 2020)

- Report on Bridging the gap between Funders and Communities (survey)
- Establish work plan for 2021
- CODATA FAIR symposium

Survey on FAIR assessments

Problem statement

Improve the understanding of benefits and challenges of the FAIR assessments from the perspective of the funders and communities.

Objectives

1. Formulate conclusions and recommendations on the level of policy, (i.e. better understanding of the perspectives of both sides)
2. Finding out how the research community and the funders' community might want to use the model and what changes they would want to see

Format and output

- E-survey
- Voluntary basis
- 8-point questionnaire

Anonymise, compile and analyse the results, draw conclusions and propose recommendations for further actions to improve understanding

Survey on FAIR assessments - results

Policy & adoption

- Funders
  - Funders to set requirements for assessment
  - Professional societies and infrastructure providers to drive adoption of FAIR assessment methodologies
  - Cross-discipline interoperability through awareness raising across communities with community-specific approaches
- Communities
  - Funding for global standards, regional implementation
  - Communities to share best practices and develop community policies and competence centres
- Both:
  - RDA as a neutral platform to bring stakeholders together and create cross-community understanding

Future work

- Funders
  - Set the bar for 'FAIR enough'
  - Select most relevant indicators in specific context
- Communities
  - Gather experience and success stories that show impact
  - Pay attention to choices to be made before data is produced
- Both
  - Consider scoring in context of community targets and practices
  - Make assessment approach more practical
  - Create actionable guidelines

Guest speakers

15’

Dr. Chris Seal
University of Auckland

15’ — Q&A

15’

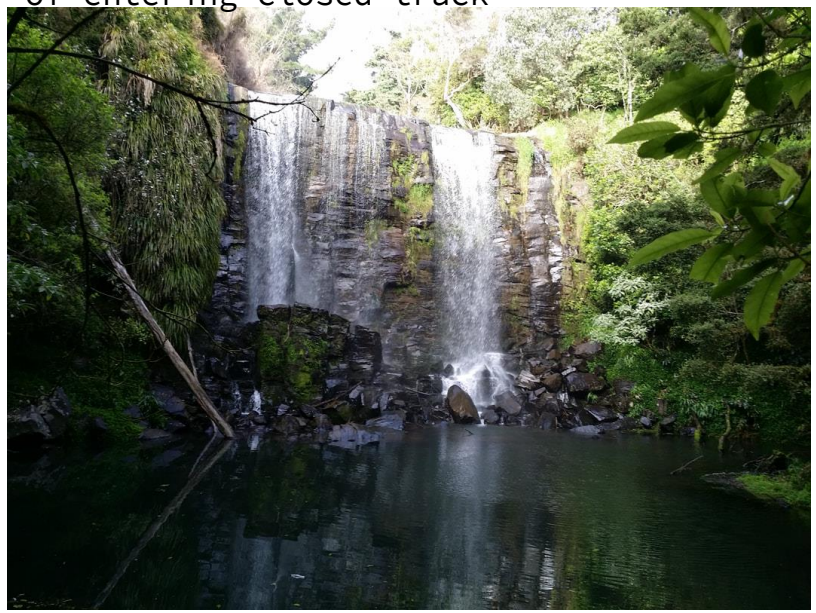
Dr. Pedro Luiz Pizzigatti Corrêa
University of São Paulo

15’ — Q&A

Dr. Chris Seal
University of Auckland

Pepeha

https://www.stuff.co.nz/environment/117703029/waitakere-ranges-auckland-council-files-charges-against-person-accused-of-entering-closed-track



https://www.kerikeriwalks.kiwi/tracks.html

Ko Waitakere te maunga

Waitakere ranges are the mountains

Ko Wairoa te awa

Wairoa is the river

Nō Tāmaki Makaurau ahau

I am from Auckland

Ko Chris tōku ingoa

My name is Chris

Tēnā koutou, tēnā koutou, tēnā tatou katoa.

Greetings to you all

RDA FAIR maturity model

Chris Seal, Senior eResearch Solutions Specialist
& Yvette Wharton, eResearch Solutions Lead

Aotearoa, New Zealand Research Sector

National scene - FAIR, CARE and Maori Data Sovereignty

What’s happening at the University of Auckland

Aotearoa New Zealand Research Sector

8 Universities, and 7 Crown Research Institutions

Funders - Royal Society - Marsden, Health Research Council, MBIE

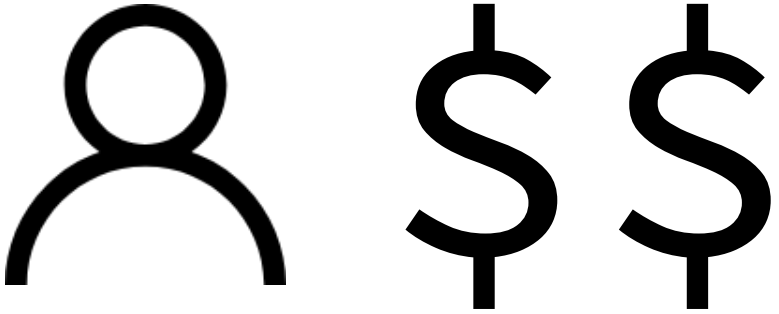
$2.5 billion NZD

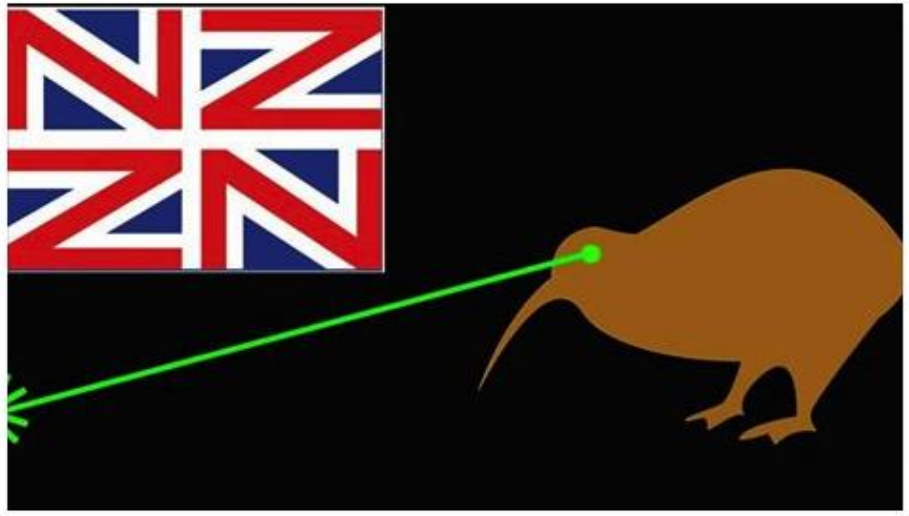
Open data - public good, societal impact - values, limited incentives

Researchers - Royal Society Code of Professional Standards and Ethics (2019) and Research Charter (2020) - principles

esri - Watercolour Map

AOTEAROA NZ - VALUES AND PRINCIPLES DRIVEN

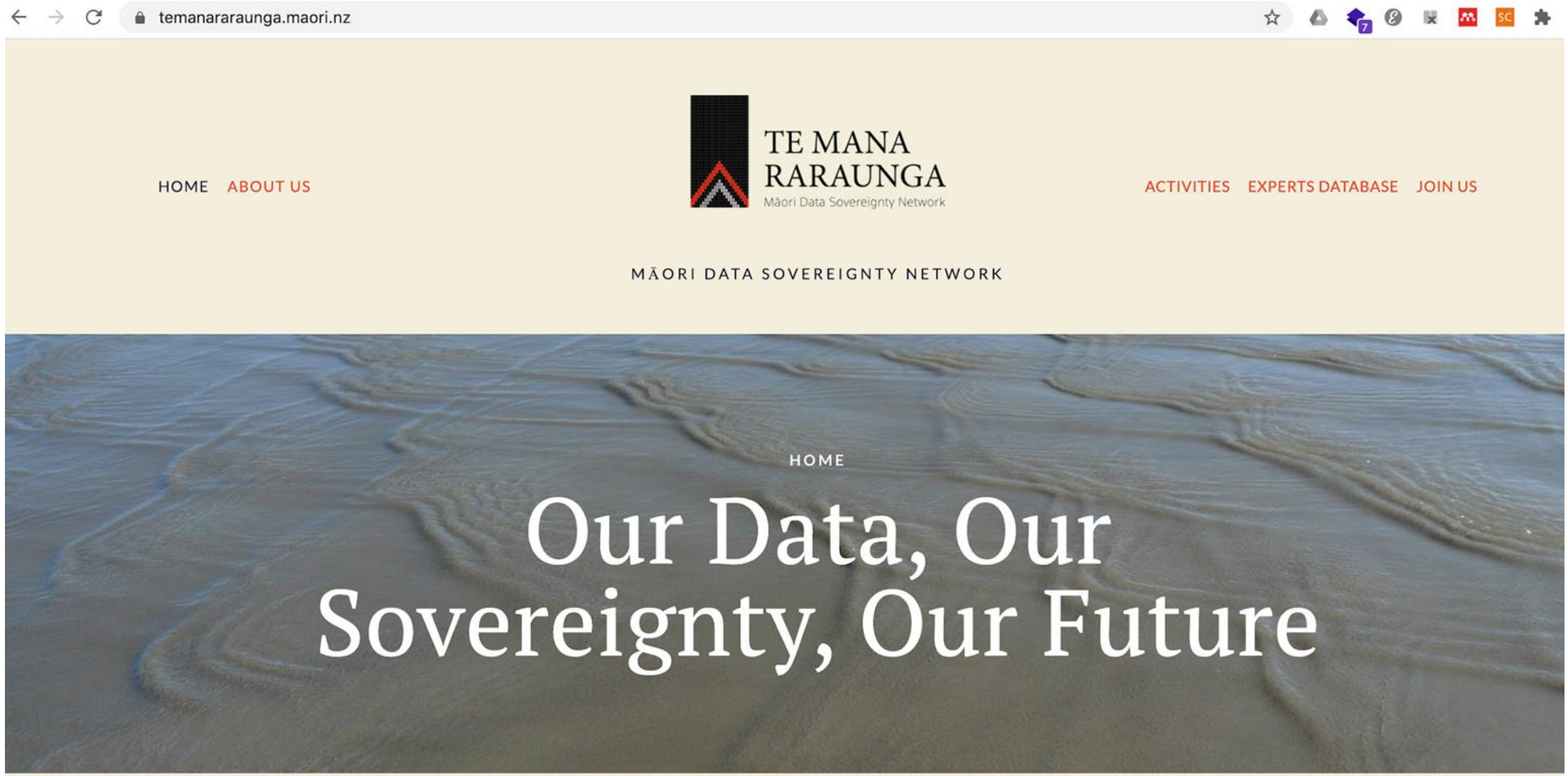




Privacy Principles

Privacy Act 2020, 1 December

Further details: https://privacy.org.nz/assets/Privacy-Act-2020-content/2020-A-quick-tour-of-the-privacy-principles-Oct-2020.pdf


temanararaunga.maori.nz
HOME ABOUT US
TE MANA RARAUNGA
Māori Data Sovereignty Network
ACTIVITIES EXPERTS DATABASE JOIN US
MĀORI DATA SOVEREIGNTY NETWORK
HOME
Our Data, Our Sovereignty, Our Future
Nga mihi ki a koutou katoa
Contact us

University of Auckland: enabling FAIR

Tāmaki Makaurau/Auckland

City of Sails

Pacific city

1.6 million people

Ethnically diverse

Monogenetic volcanic field
(52 volcanic cones)

Makaurau
The University of Auckland

Founded in 1883

Aotearoa NZ’s largest University

Over 40,000 students, 5,500 FTE staff

World’s highest ranking for sustainable impact
(THE University impact ratings - 2019)

University of Auckland : RDM Framework (2020-)

– – – –

“Towards an integrated RDM Framework”

Our drivers:

- Domestic & international funder, publisher and ethics committee requirement\
- Responsiveness to and engagement with Māori
- Research impact
- Best tools & practices - enabling FAIR, CARE & Māori Data Sovereignty principles
- Data as a strategic asset

“Where is our research data maturity currently and

what do we need to do to get to where we aspire to be?”

RESEARCH DATA MANAGEMENT

Research Plan

Identify and record:
- grant and contact information
- types/sizes of data created/collated
- how data will be organised
- documentation and metadata
- who needs access to what data
- where data will be stored
- copyright, IP and licensing
- public dissemination
- retention or deletion requirements
- long term archiving, preservation
- ongoing roles and responsibilities

Ensure scope of ethical consent enables future re-use of data.

Consider data types, sources, volume, and file formats. Include descriptive and technical metadata, and use open, machine readable file formats. Use logical file names and data organisation strategies. Build in data checking and quality assurance during data collection. Keep data in raw format whenever possible to facilitate future re-analysis and analytical reproducibility.

Data Creation/ Collation

Instrument Data Storage

Raw data and associated metadata output from some instruments will be preserved for posterity.

Active Data Storage

Have a systematic backup scheme. Storage method depends on size and nature of data, costs of storage, how the data will be used, time to transfer, who needs access, and privacy concerns. Raw data may need cleaning or reducing.

Processed Data

Analysis and Modelling

How computationally intensive are any analytical processes? Should intermediate data be saved or can it be regenerated? Conduct analyses with a particular level of reuse in mind. Track processes used to generate data and versions. Keep an electronic lab notebook to record metadata that will later be packaged with final data that is stored, reused and shared.

Retained Output Data

Completed or publishable research data to be prepared to enable easy interpretation by third parties (Publishing and/or Archival) and for retention/deletion according to University policy and DMP.

Preservation

How long should the data be accessible for? Consider preservation and curation issues, how and where the data will be stored or accessed, and the need for active migration of data to different formats or media through time. Also consider when and how data should be deleted or destroyed.

IP, Copyright and Licensing

Enable re-use and increase citation and impact by choosing open licences and formats or structures to facilitate easy combination with other data.

Report

Ensure compliance with all funder, government, and institutional policies on how data will be managed and shared.

Impact

Impact, operationalised using a variety of tracking and reporting metrics. e.g. Altmetrics, BiblioInformatics, and Research Outputs.

Visibility

Increase visibility of research and researchers e.g. VIVO, researcher profile tools and social media.

Publish

Publish open data or metadata only. Establish copyright and licensing of data. Give data a permanent, unique identier (DOI and publish in institutional, discipline, or journal repositories (obeying any data restrictions or privacy concerns). Create discovery metadata along with user documentation or links to provide the context needed to interpret the data. Cite and link data in publications.

Data Discovery

Data Catalogue

University/National catalogue/registry (metadata only) of all data outputs

Study concept and design

Projectisation to link research to University systems, and service provisioning and reporting.

Data lifecycle model adapted from McLean, C. A. (2016): The Research Data Lifecycle. figshare.https://dx.doi.org/10.17608/k6.auckland.3102922.v1

UoA DATA ECOSYSTEM

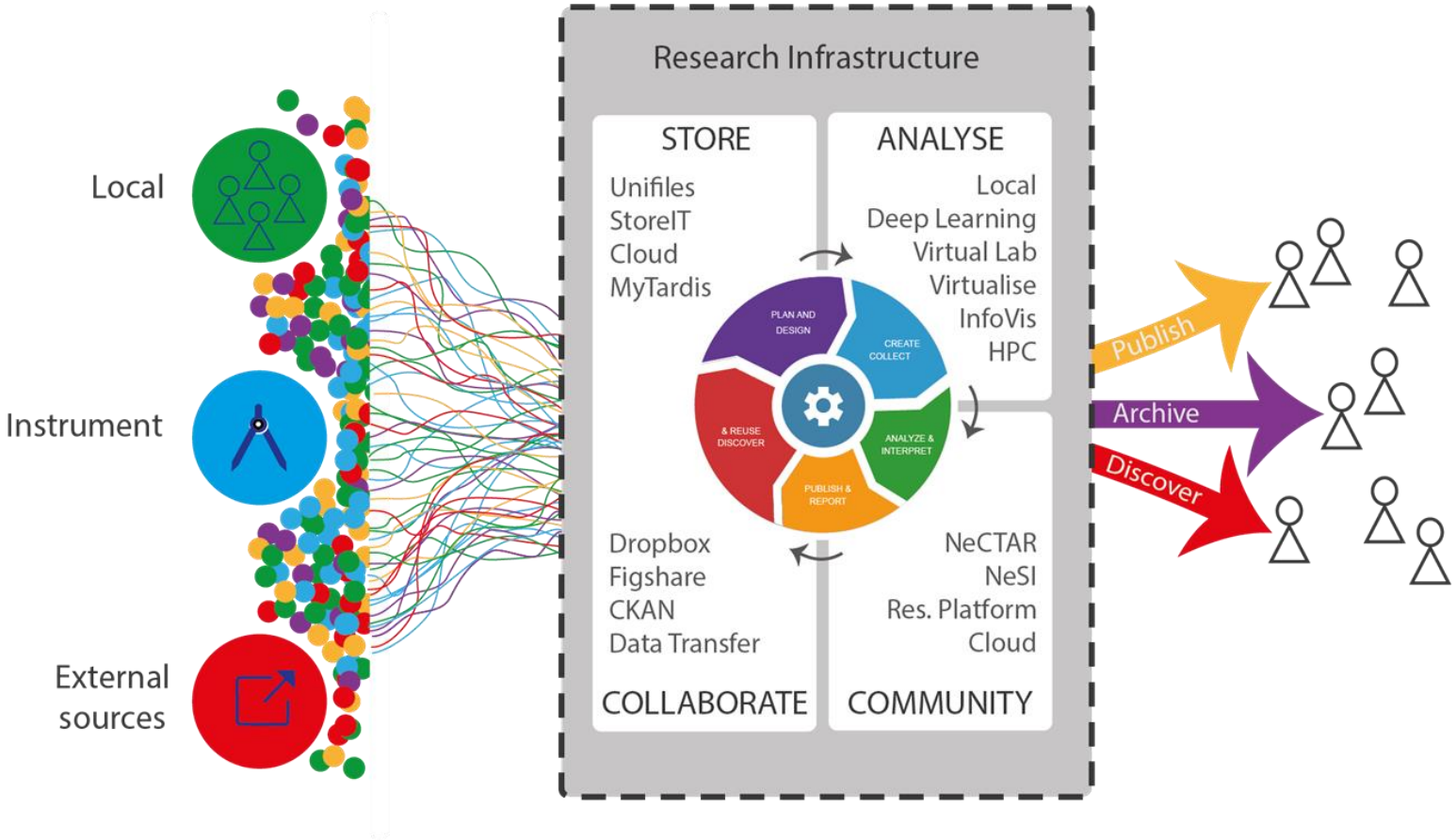
TRAINING / DIGITAL SCHOLARSHIP

DATA RECORD
DMP

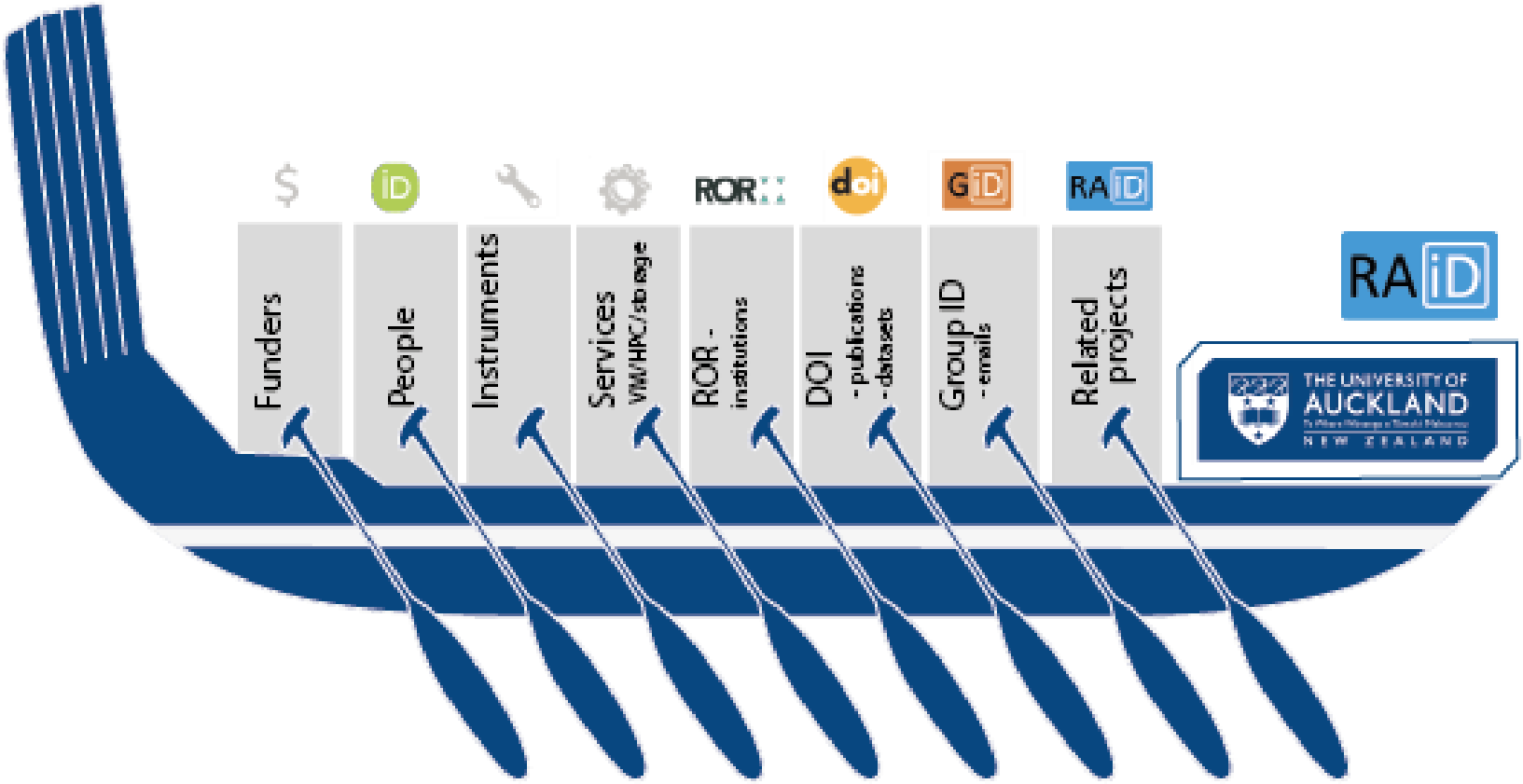
MANAGED DATA
METADATA

DATA ARCHIVE, DISCOVERY
& FORECASTING

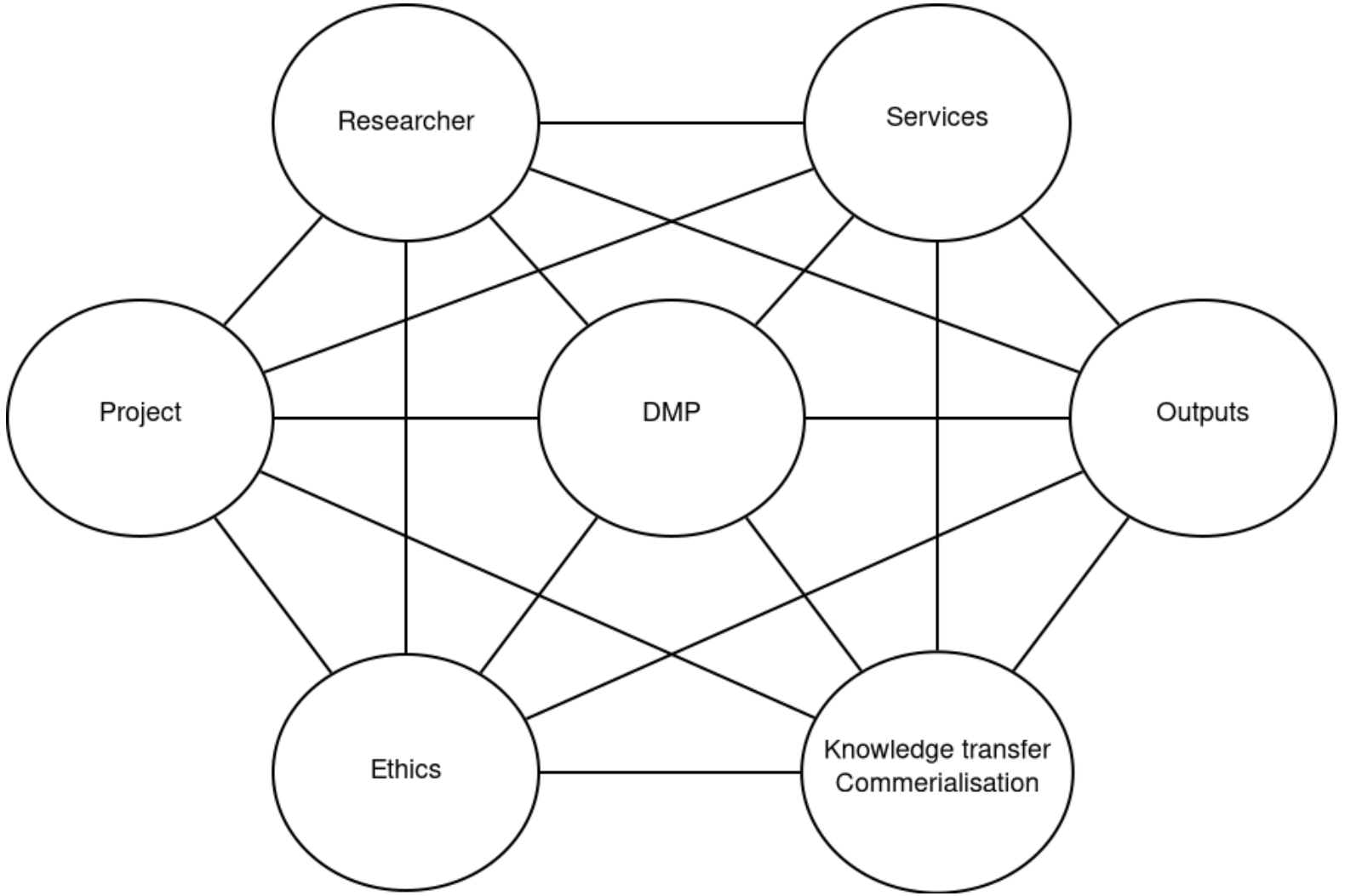
DOCUMENTATION



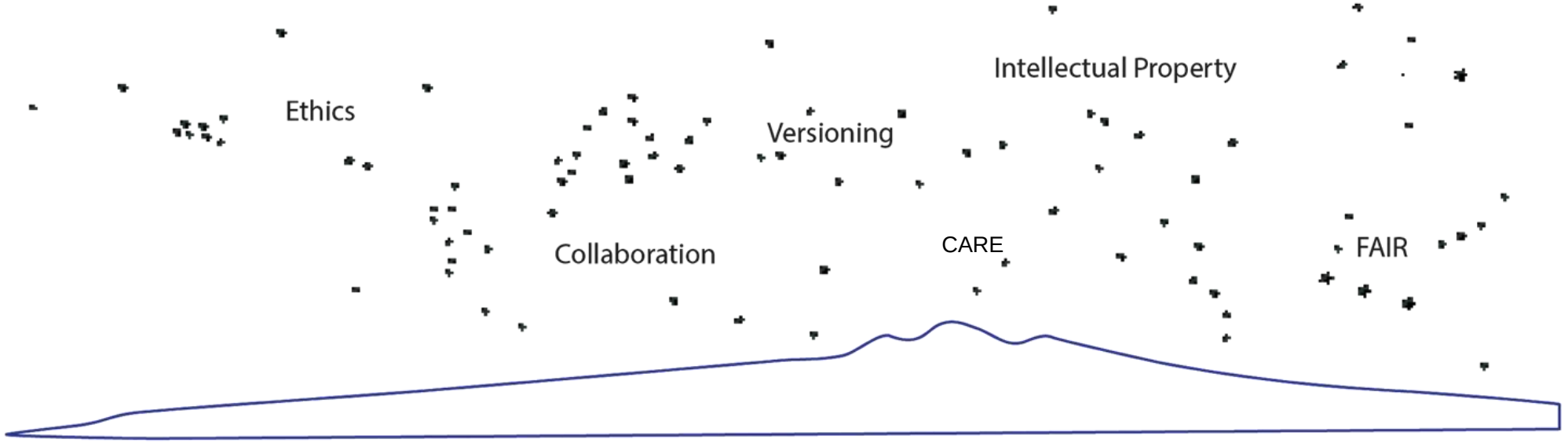
PIDs to Facilitate Connectivity

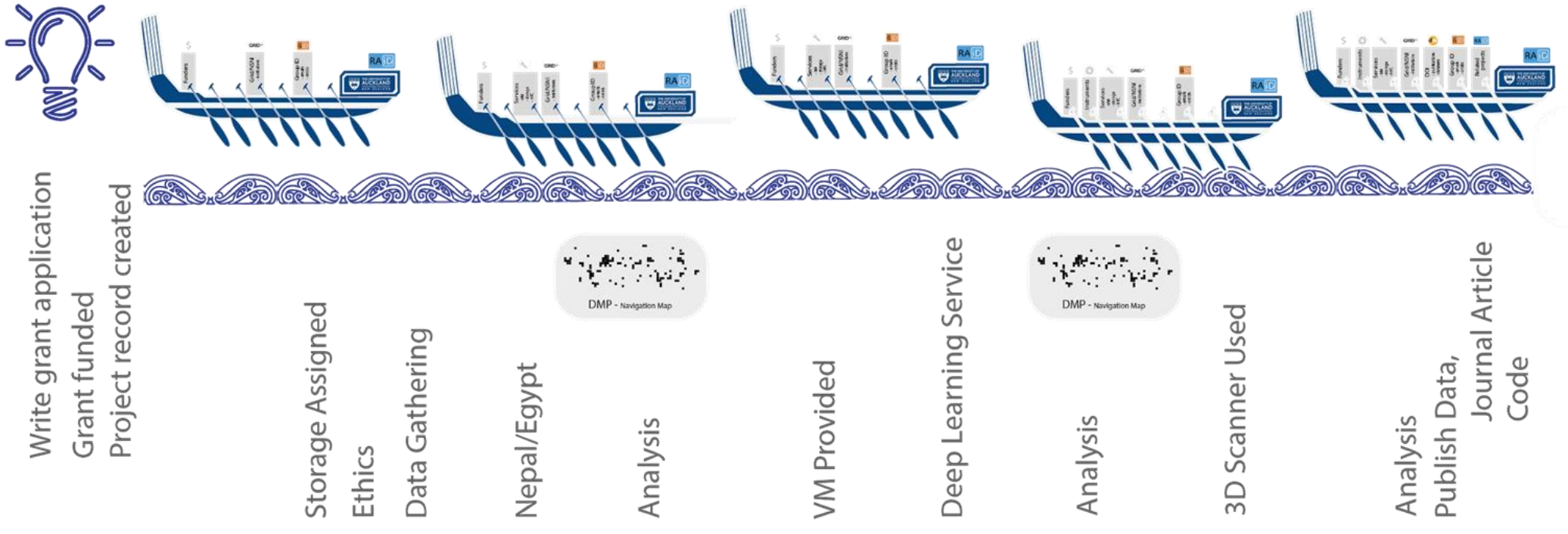


Towards a Connected Ecosystem



A vision :
Automated
Connected
FAIR
CARE





Dr. Pedro Luiz Pizzigatti Corrêa
University of São Paulo

FAIR Data maturity model session
November, 10th – 2020
FAIR Development in Brazil – Research Actions

Prof. Dr. Pedro Luiz Pizzigatti Corrêa - pedro.correa@usp.br
Digital Systems and Computer Engineering Department
Escola Politécnica da Universidade de São Paulo - EPUSP
Big Data and Data Science Research Group of EPUSP **wds.poli.usp.br**

School of Engineering (Escola Politécnica – POLI) of University of São Paulo is the most complete and important engineering School in Latin America

"Train professionals comitted to the sustainable development of the country, with social, economic and environmental responsibilities (...)"

Poli Mission

Fonte: Website da Escola Politécnica (www.poli.usp.br), Times Higher Education

Institutional Data

- 15 Departments of Teaching and Research
- Built area: 141,500m² -
- 9 buildings Library: collection of 590,000 documents
- UnderGrad: ~ 4,500 students enrolled PhD: ~ 840 Master's and ~730 PhD students

Leadership position

- Poli/USP is the 105th best technology school in the world and the best in Latin America
- Largest graduate center in engineering in Brazil
- One of the largest trainers of entrepreneurs and executives in the country
- USP is responsible for more than 20% of the total national scientific production

2

Research on management and analysis of large volumes of scientific data

Pedro Luiz Pizzigatti Corrêa:

- Associate Professor (Univeristy of São Paulo - USP) – Department of Computer and Digital Systems Engineering and Coordinator Big Data and Data Science Research Group - Engineering.
- Education: Bachelor and Master of Computer Science (ICMC/USP). PhD in Electrical Engineering (USP) and Post-doctorate in data science focusing on distributed databases – Univesity of Tem

Research projects involving scientific data management:

- Devlopment of new tools for sharing and reuse of data through transnational research on the socioeconomic impact of Conservation Units (PARSEC) - FAPESP/NSF/ANR/JST - BELMONT FORUM - Result – Data Science and Computational Models (that uses satellite images to generate socioeconomic indicators of communities close to Protect Areas) - https://parsecproject.org/

https://parsecproject.org/

- FAPESP Thematic Project in the Climate Change program focusing on Data Management (Coordination Prof. Dr. Paulo Artaxo). Result under development: Model for Aerosol Data Quality Management Report (DQMR), Data Portal and Big Data Analysis based on Cloud Infrastructure (Partnership with ARM/ORNL/DoE/USA)
- E-Science Program - FAPESP - "Enabling Integrated Research through monitoring of biodiversity and climate measurements" – Result: Infrastructure of Big Data Analytics bioclimatic data that integrates biodiversity observation data and aerosols collected at different sites near the city of Manaus (Amazon - Brazil) – finished
- Brazilian Biodiversity Data Portal – Minitry of Environment - Brazil, 2015 in colaboration with Atlas of Living Autralian (ALA) https://portaldabiodiversidade.icmbio.gov.br/

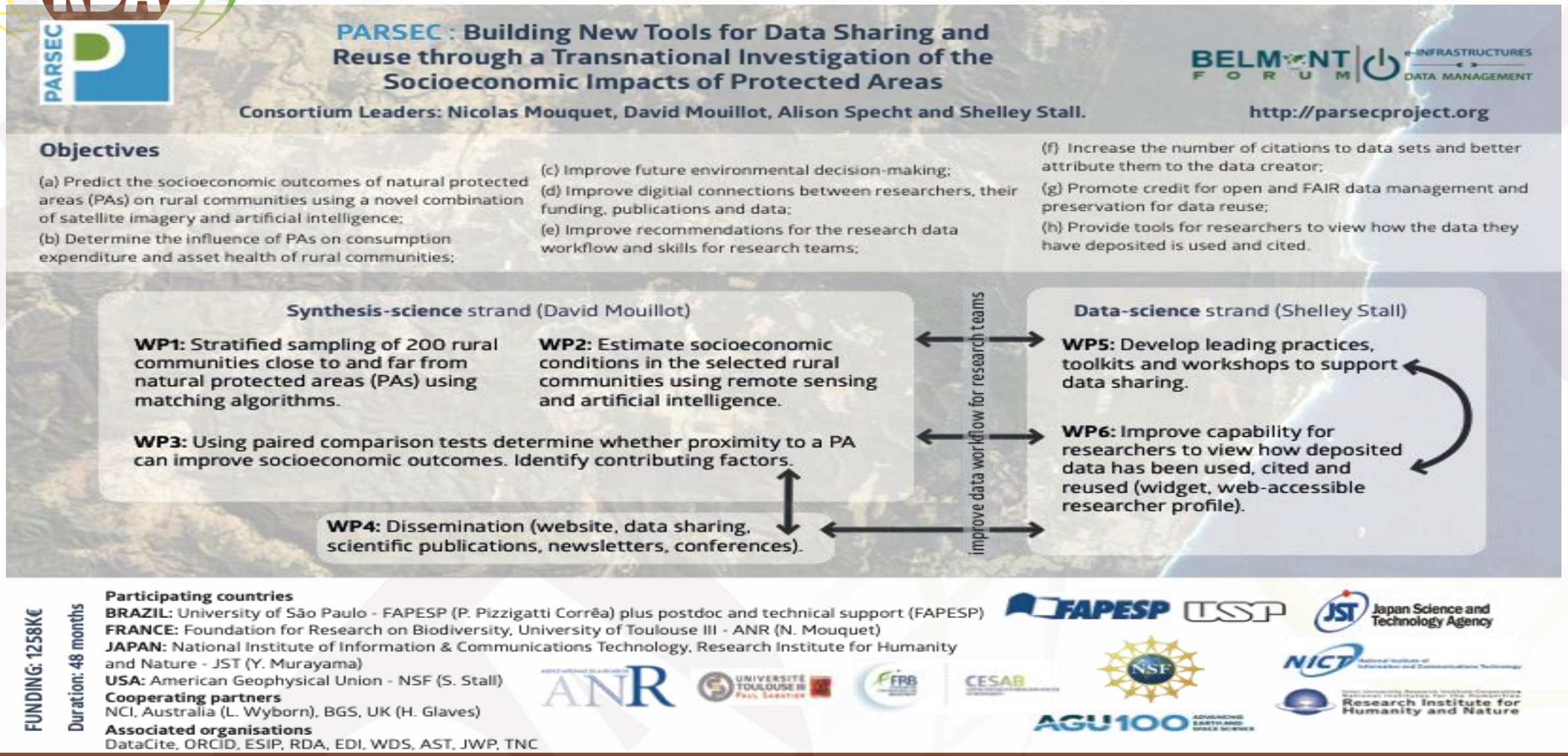
https://portaldabiodiversidade.icmbio.gov.br/

Big Data and Data Science Research Group of Engineering wds.poli.usp.br

Center of Data Science (C2D) – Itaú-Unibanco - http://c2d.poli.usp.br/

Collaboration:

PARSEC Project



WDS 2021 Plan

VI Workshop on Data Science: best practices on data sharing and data synthesis

(previous Workshop: wds.poli.usp.br/wds5)

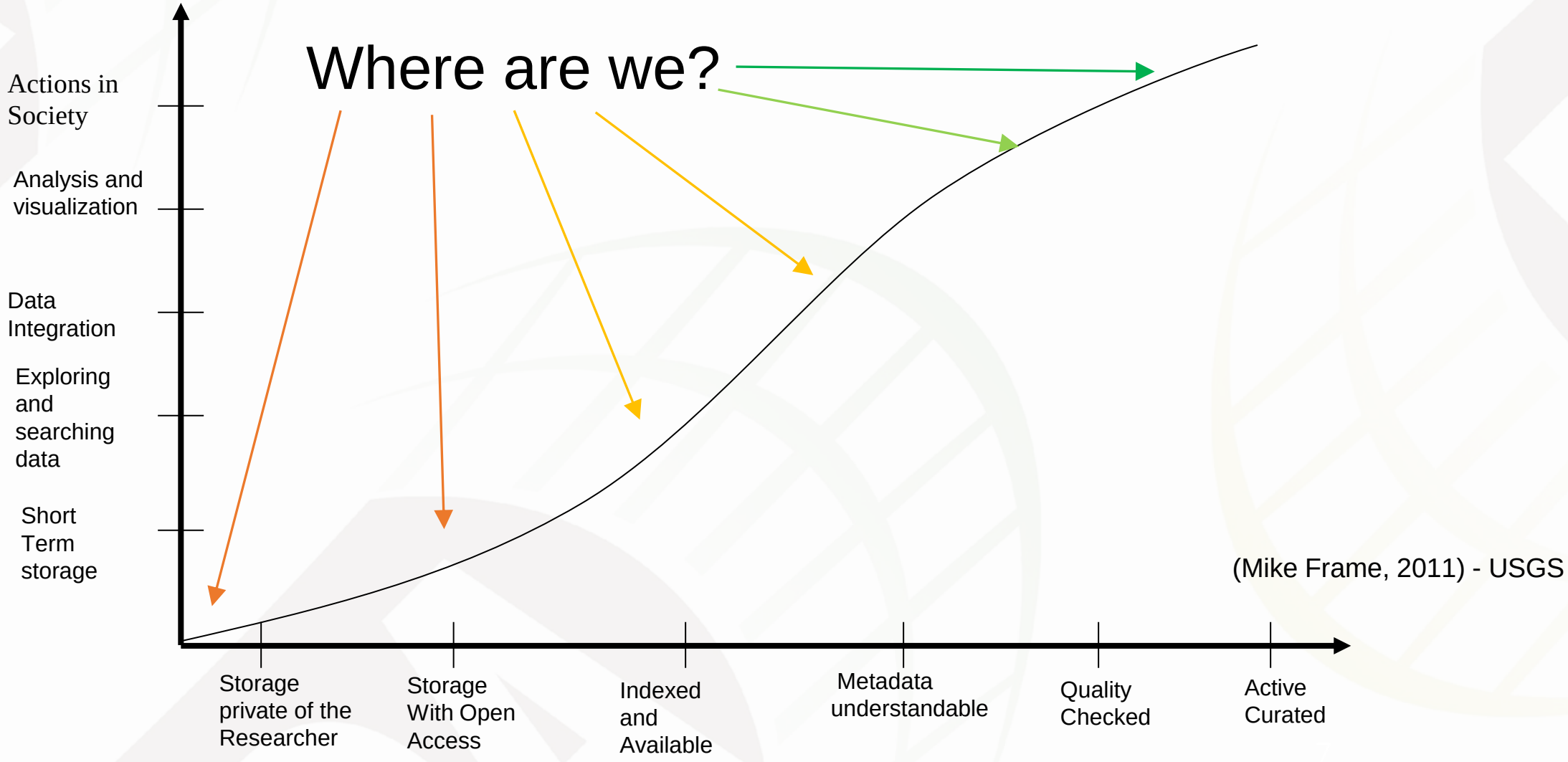
Goal: Present the advances of PARSEC on leading practice for data sharing, attribution, credit, reuse and synthesis as a tool and toolkit for Brazilian and international community, interested in open data management and analysis.

Date/Place: 1st week of September/2021 in São Paulo, Brazil.

General Recomendations to improve FAIR in Brazil

- Be transparent in methods, platforms and infrastructure - Clear Data Policy ...
- International collaborations, USGS, ORNL, NCI, (Workshops, visiting researchers), active participation in international projects, international forums and Communities (RDA, CODATA, ESIP, EGU)
- Be prepared for the transition to the next generations of hardware
- Continuous understanding of nature in our business model and value chain;
- Multidisciplinary teams, also involving areas of knowledge in Computing and Information Science
- International collaboration in research, software development, training and qualification of people - Engaged more people!
- People don’t scale, systems do

Scientific Data Management - FAIR



FAIR Data maturity model session
November, 10th – 2020
FAIR Development in Brazil – Research Actions

Prof. Dr. Pedro Luiz Pizzigatti Corrêa - pedro.correa@usp.br
Digital Systems and Computer Engineering Department
Escola Politécnica da Universidade de São Paulo - EPUSP
Big Data and Data Science Research Group of EPUSP **wds.poli.usp.br**

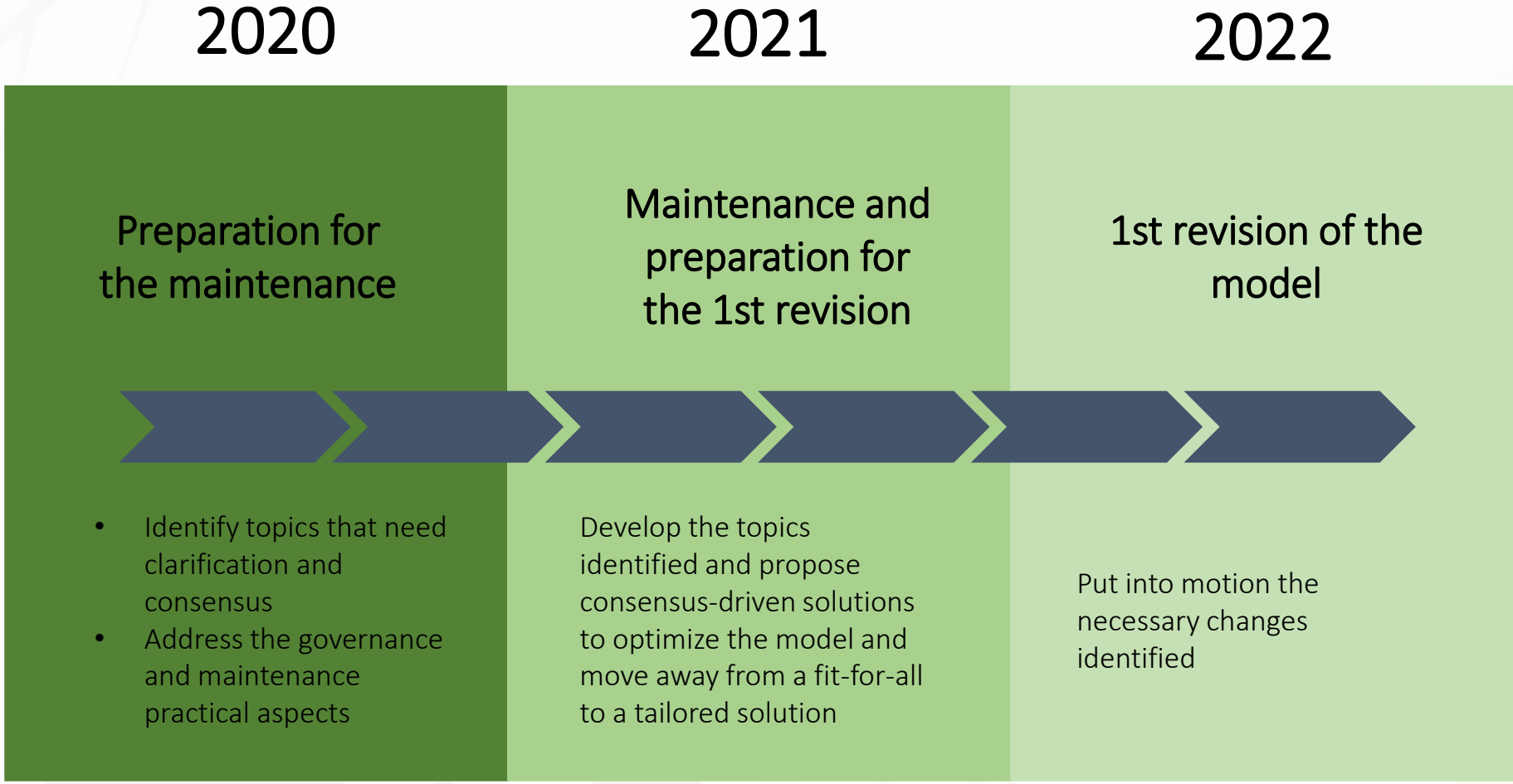
OPEN

DISCUSSION

Topics for the next version of the model

- 44% • Approaches toward evaluation of FAIR assessment tools and services, taking into account community aspects
- 25% • Consensus within the communities for the priorities of the indicators, respecting the different targets for FAIRness and speeds of implementation
- 31% • Cross-community interoperability and evolution of standards
- 31% • Role of landing pages and human readable documentation
- 44% • PID practices across communities (identifiers for metadata, data, separately, combined)
- 63% • Metadata practices:
  - Role of generic platforms / repositories in improving domain-specific metadata
  - Metadata at several levels (collection level, individual level)
  - Consensus on minimum level for 'Rich' metadata
  - Shared understanding of knowledge representation
- 50% • Data granularity (collection, dataset, data item)

Maintenance and governance



Next steps

11th of November 2020

Report on the survey on benefits and challenges of FAIR assessments

19th of November 2020

Webinar on how to enhance the FAIRness of the science within communities and the use of the FAIR Data Maturity Model as a framework for comparing results

maintenance and governance plan

3rd of December 2020

CODATA FAIR symposium
Wrap-up of the FAIR data maturity model and definition of priorities for 2021

https://conference.codata.org/FAIRconvergence2020/

Upcoming Webinar

19 November, 2 – 3:30 pm UTC (9-10:30am EST)

This webinar will convene a group of experts representing national science organizations to discuss their current initiatives to enhance the FAIRness of the science within their communities and the use of the FAIR Data Maturity Model as a framework for comparing results.

The presentations and discussions will particularly interest policy makers and organizational leaders interested in learning about successful strategies for integrating the FAIR Data Maturity Model within overarching organizational strategies for promoting open science.

Speakers:

- EOSC FAIR WG and the FAIR Data Maturity Model
  - Sarah Jones, Geant, Oya Beyan, Fraunhofer Institute for Applied Information Technology FIT; Aachen University.
- Driving the paradigm shift towards Open Science: FAIR data and the Role of the FAIR Data Maturity Model
  - Carlos Casorrán and Konstantinos Repanas, European Commission, Directorate-General for Research and Innovation, Open Science Unit
- The US Geological Survey FAIR Roadmap and the role of the FAIR Data Maturity Model
  - Viv Hutchison, USGS Branch Chief for the Science Data Management in the Science Analytics, and Synthesis (SAS) program, Frances Lightsom, USGS FAIR Roadmap Architect and Project Manager
- The NIH Strategic Plan Data Science advocating for FAIR Principles
  - Susan Gregurick, Associate Director for Data Science and Director of the Office of Data Science Strategy (ODSS)
- The National Science Foundation / EarthCube
  - Pending

Joint Session: Implementing the CARE Principles: The CARE-full Process

Thursday, 12 Nov, 5:00 - 6:30 PM UTC | Breakout 7

Groups Leading:

- International Indigenous Data Sovereignty IG
- FAIR Data Maturity Model WG

Meeting description:

- This Joint Session meeting will include presentations and discussions of what a process for implementing the CARE Principles within research, government, non-profit and other institutions might look like, including whether or not there is a need for criteria, metrics, or assessments.
- The approach taken by the FAIR Data Maturity Model WG and the supporting FAIR community involved in its development is a relevant use case for how the CARE Principles can be implemented and assessed.

FAIR data maturity model

Stay in touch!
https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

RDA Webpage

Evaluation mechanism

RDA recommendation

- RDA FAIR data maturity model WG – GitHub
- RDA FAIR data maturity model WG – Mailing list

images: Flaticon.com