FAIR Data Maturity Model

Workshop #9

20th / 21st May 2020

Agenda

5'	Welcome, objectives of the meeting
5'	Roundtable
10'	State of play
10'	Public review period
15'	RDA recommendation
20'	Early adopters – Experience sharing
5'	Disseminate the FDMM to communities
5'	Thanks and resources
10'	Maintenance workplan
5'	Initiation of a feedback loop

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

FAIR

SOLUTION is to bring together **stakeholders** to build on **existing approaches** and **expertise**
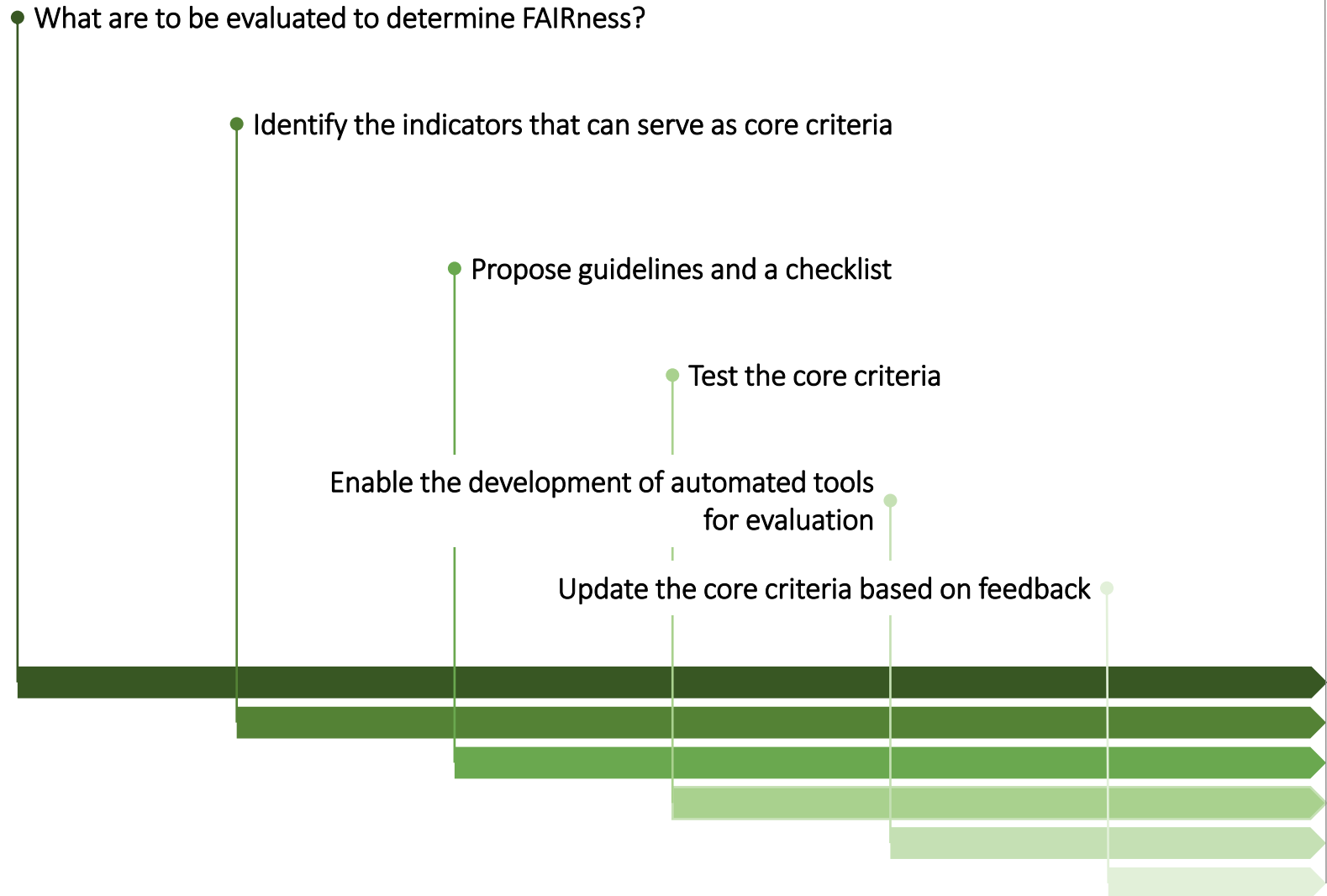
- Set of **core assessment criteria** for FAIRness
- FAIR **data maturity model & toolset**
- FAIR data **checklist**
- RDA recommendation

Join the **RDA** Working Group: RDA WG web page | GitHub

Objectives

- What are to be evaluated to determine FAIRness?

- Identify the indicators that can serve as core criteria

- Propose guidelines and a checklist

- Test the core criteria

Enable the development of automated tools
for evaluation

Update the core criteria based on feedback

Scope

BUT the Working Group does <u>NOT</u> 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**

Roundtable

In the chat window, please type…

> Your name

> Your affiliation

> Your role
>> Researcher
>> Librarian
>> Service provider
>> Policy maker
>> Funder

> Introducing the editorial team
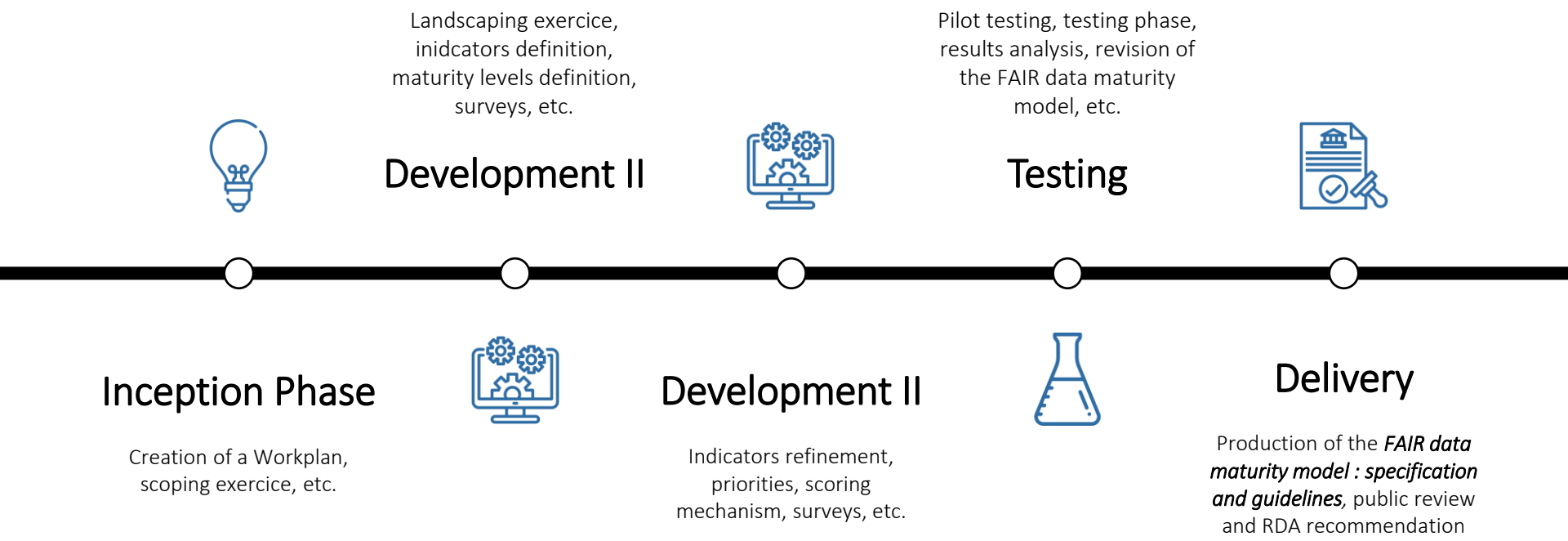
State of play

State of play – phases
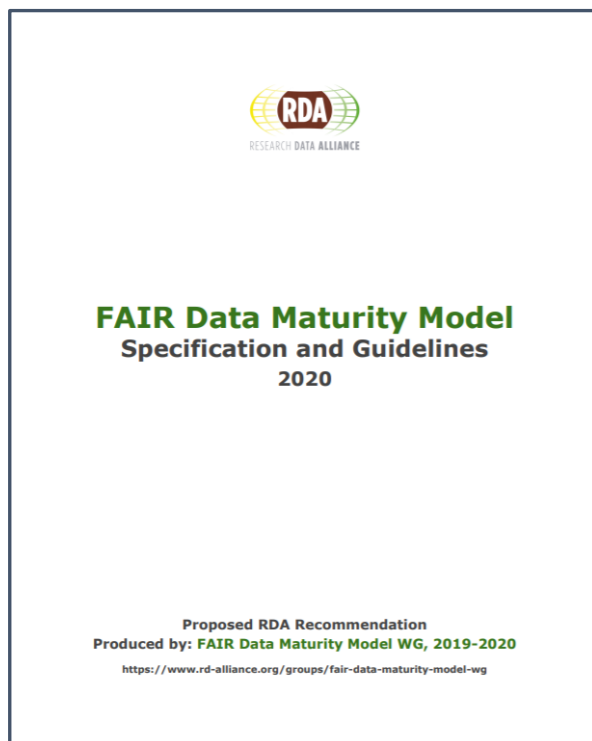
1. Definition   DONE

2. Development   DONE

    i) First phase   DONE

    ii) Second phase   DONE

3. Testing   DONE

4. Delivery   CLOSING

State of play – phases II

Landscaping exercice, inidcators definition, maturity levels definition, surveys, etc.

Pilot testing, testing phase, results analysis, revision of the FAIR data maturity model, etc.

Development II

Testing

Inception Phase

Development II

Delivery

Creation of a Workplan, scoping exercice, etc.

Indicators refinement, priorities, scoring mechanism, surveys, etc.

Production of the *FAIR data maturity model : specification and guidelines,* public review and RDA recommendation

Public review period

FAIR Data Maturity Model
Specification and Guidelines
2020

Proposed RDA Recommendation
Produced by: FAIR Data Maturity Model WG, 2019-2020
https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

THANKS TO ALL REVIEWERS

3600+ page views

14 comments

https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines

Public review period – To be fixed

› Addition of a section implementation to the section Framework

  › Clarification on the scenarios that have not been validated by the WG (e.g. in the context of data-related algorithms, tools, workflows, protocols and other data-related services)

  › Clarification of the requirement in the FAIR principles for persistent identification of both metadata and data

  › Clarification that not all of the indicators are equally important or relevant for every community

› Clarification in the section on Evaluation methods

  › Different scenarios depending on when evaluation is done: before of after data has been created

  › Necessity to highlight the importance of the responsible and careful implementation of the indicators to minimize unintended consequences

› Additions to the section Future maintenance

  › Taking into account community perspective

  › Developing use cases

  › Including examples of supporting technologies

› Editorial changes, minor rephrasing, correction of examples

Public review period – Won't fix

> Issues beyond the FAIR principles (e.g. versioning, DMP, data deletion, quality)

> Change requests as for the indicators (e.g. because overlaps, should be merged, etc.)

> Change requests as for the priorities (the WG already reached a consensus)

> Requests for clarifications concerning elements such as "what is a metadata file", "what is a data file", etc.

RDA recommendation

13 May 2020
Closing of the public review period and resolution of comments received followed by the RDA council review

14 April 2020
Beginning of the public review period. Surveying the adopters about their experience

End of May 2020
Submission of final version to RDA Council for endorsement of the *FAIR data maturity model: specification and guidelines* as an RDA recommendation

Discussion items

Early adopters – Experience sharing

- Ge Peng | NOAA
- Anusuriya Devaraju | FAIRsFAIR

… will share their relevant experience with regard to the adoption of the FDMM and answer to the following questions;

1. What is the level of adoption at your organisation? (E.g., pilot, production, …)
2. Do you plan to continue to use the Recommendation?
3. Did you need to modify the Recommendation for your use?
4. Can you give an estimate of how much time / effort you have spent on the adoption so far?
5. What's your overall experience? (E.g., Very Good, Good, Fair, Poor)
6. Would you do it again?

Evaluating the FAIRness of Environmental Data
– Application of the RDA FAIR Data Maturity Indicators

Ge Peng, PhD

Cooperative Institute for Satellite Earth System Studies (CISESS) Between
U.S. National Oceanic and Atmospheric Administration (NOAA) and North Carolina State University
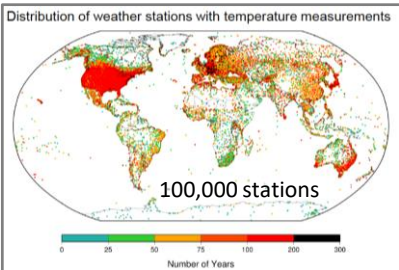at NOAA National Centers for Environmental Information (NCEI)

#9 Workshop of the RDA FAIR Data Maturity Model Working Group, May 20–21, 2020

NCEI Is an Official Archive of U.S. Environmental Data

Data Sources	Geophysical Parameters	Platforms	Scales	Archive Volume

Data Sources
- **National,**
- **International**

Distribution of weather stations with temperature measurements

100,000 stations

Number of Years

Geophysical Parameters
- **2000+**
  - Wind,
  - Temperature,
  - Ocean elevation, …

From the bottom of the Oceans to the surface of the Sun.

Platforms
- **Stations,**
- **Buoys, Ships,**
- **Satellites,**
- **Models,**
- **Unman devices, …**

Scales
- **Temporal:**
  - Synoptic,
  - Seasonal,
  - Decadal,
  - Paleo, …
- **Spatial reso.**

From stone-age to space-age.

Archive Volume
- **35+ PB,**
- **54+ mil files.**

Projected to be 200 PB in ~ 10 years.

NCEI Is an Authoritative Source
For Environmental Data and Information

Services

Governmental Mandates and Policies

1,256 PB (April 2020)
- Government agencies;
- Businesses;
- general public.

Establishing & ensuring trustworthiness is critical.

Requirements Include:
- High-quality;
- Findable, accessible, interoperable;
- Preserved and usable – long-term

Required by law to demonstrate the compliance to federal requirements.
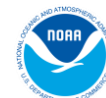
Purposes of Pilot Application

- **Examine the relevancy of the RDA FAIR DMIs (v0.04)**

- **Baseline the FAIRness of NCEI managed data**
  - In particular, *OneStop*-Ready datasets,
    - *OneStop* project was Initiated in 2015 to improve discovery and access services for NOAA datasets.
  - What worked?

- **Identify potential gaps & define path forward in NCEI data sharing practices**

OneStop-Ready Datasets

Metrics and Tiers of *OneStop* Readiness

OneStop Tier	*OneStop* Readiness Metrics				
	Collection Metadata	Data Online	Granule Metadata	DSMM	Data Formats
Gold	Y (>90%)	Y	Y	Y	Y
Silver	Y	N	N	Y	Y
Bronze	Y	N	N	N	N

NCEI Core Profiles Complete, ISO standards-based, interoperable. Completeness checked by a rubric.	**Common Framework** Interoperable, standards-based, discovery and access protocols.	**NCEI Core-lite Profile** A lite-version of NCEI Core Metadata Profiles.	**Quality Metadata** Content-rich, ISO standards-based, interoperable.	**Common Framework** standards-based, (machine-actionable preferred).

> ➢ **15 data groups, ~370 datasets with millions of granules**

The FAIRness of *OneStop*-Ready Datasets

FAIR Aspect	Total Rating (# of Criteria)	Essential (# of Criteria)	Important (# of Criteria)	Useful (# of Criteria)
F	**3.43** (7)	3.2 (5)	4 (2)	N/A (0)
A	**3.94** (16)	4 (7)	3.86 (7)	4 (2)
I	**3.67** (16)	4 (2)	3.625 (8)	3.5 (6)
R	**3.08** (12)	2.67 (3)	3.125 (8)	4 (1)

- **Evaluation was carried out manually, utilizing v0.04 RDA FAIR DMIs**
- **5 levels of compliance to each DMI:**
  - Level 4: Satisfy the criterion;
  - Level 3: Planned and implementation has started but not done;
  - Level 2: Planned but implementation has not started;
  - Level 1: In the planning stage;
  - Level 0: Not satisfy the criterion and no future plan.
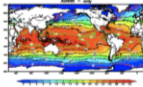
What Worked?

- **Each dataset is assigned and minted a DOI**
  - Linked to publicly accessible dataset metadata,
  - Resolved to a consistent-layout landing-page,
- **Well-curated dataset metadata**
  - Unique identifier, rich, comprehensive entities,
    - discovery metadata
- **Catalogued and accessible meta(data)**
  - Standard formats, protocols, and services.
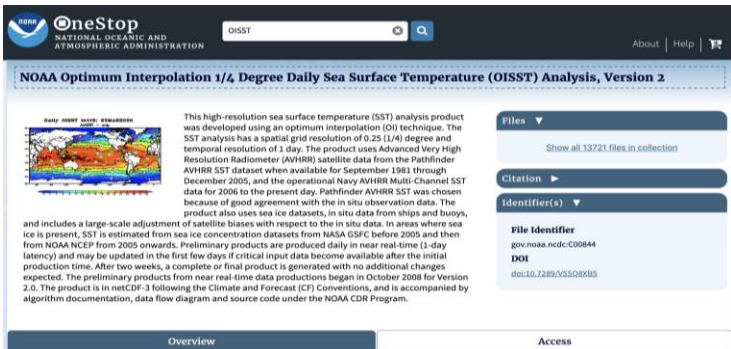
NOAA Optimum Interpolation 1/4 Degree Daily Sea Surface Temperature (OISST) Analysis, Version 2

Sea Surface Temperature - Optimum Interpolation

This high-resolution sea surface temperature (SST) analysis product was developed using an optimum interpolation (OI) technique. The SST analysis has a spatial grid resolution of 0.25 (1/4) degree and temporal resolution of 1 day. The product uses Advanced Very High Resolution Radiometer (AVHRR) satellite data from the Pathfinder AVHRR SST dataset when available for September 1981 through December 2005, and the operational Navy AVHRR Multi-Channel SST data for 2006 to the present day. Pathfinder AVHRR SST was chosen because of good agreement with the in situ observation data. The product also uses sea ice datasets, in situ data from ships and buoys, and includes a large-scale adjustment of satellite

Show more...

Sea Surface Temperature image

Dataset Citation

Dataset Identifiers
- doi:10.7289/V5SQ8XB5
- NCEI DSI 3650_01
- gov.noaa.ncdc:C00844

Unique meta(data) identifiers

ISO 19115-2 Metadata

Access | Time & Location | Documentation | Description Lineage | Credit | Keywords | Constraints

Search & Visualization Portal — GIS map application to visualize dataset layers.

Download Data

NCEI Direct Download (download)
Direct download links for dataset files.

Data Access Protocols — NCEI THREDDS Catalog (download)
THREDDS Data Service for this dataset.

NCEI ERDDAP Server (download)
ERDDAP Data Service for the dataset.

Data Format — NetCDF (Version: 3)

Distributor — Customer Engagement Branch

Use and access constraints

Including provenance

Adopting OAIS RM & DSMM Helped!

Mapping FAIR Data Principles to NCEI/CICS-NC Data Stewardship Maturity Matrix (DSMM)

FAIR Data Principles (Wilkinson et al. 2016)	DSMM Key Components								
	Preservability	Accessibility	Usability	Production Sustainability	Data Quality Assurance	Data Quality Control/Monitoring	Data Quality Assessment	Transparency /Traceability	Data Integrity
F1. (meta)data are assigned a globally unique and eternally persistent identifier								L3	
F2. data are described with rich metadata (defined by R1 below)	L3		L3					L5	
F3. metadata clearly and explicitly include the identifier of the data it describes	L3		L3					L3	
F4. (meta)data are registered or indexed in a searchable resource		L2 & L3							
A1. (meta)data are retrievable by their identifier using a standardised communications protocol		L2 & L3	L3					L3	
A1.1. the protocol is open, free, and universally implementable		L3							
A1.2. the protocol allows for an authentication and authorization procedure, where necessary		L3							
A2. metadata are accessible, even when the data are no longer available		L2							
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	L3		L3						
I2. (meta)data use vocabularies that follow FAIR principles		L4							
I3. (meta)data include qualified references to other (meta)data	L3		L3						
R1. meta(data) are richly described with a plurality of accurate and relevant attributes	L3		L3						
R1.1. (meta)data are released with a clear and accessible data usage licence	*		*						
R1.2. (meta)data are associated with detailed provenance									
R1.3. (meta)data meet domain-relevant community standards	L3		L3						

Can be easily implemented via relevant metadata entity and modified document template

(Version: v00r01 20200403; POC: gpeng@ncsu.edu; CC-BY 4.0)

Many data stewardship quality attributes are not explicitly addressed by the FAIR Data Principles.

- Most of data are open by default,
- Use agreements or use constraints,
- CC license not yet explicitly included.

Path Forward

Improving the FAIRness of NCEI & NOAA Data

Extending the Application Scope – under discussion

- **Explicitly include** a data usage license, e.g. CC-BY 4.0; CC0, in the metadata record:
  - Discussions are on-going,
  - Procedure is under development.

- **Assess:** 200+ additional NCEI datasets,
  - produced by NCEI's Center for Weather and Climate, *various stages of OneStop-ready.*
- **Examine** the scalability of the evaluation.

Integrating Assessment Results - Fairly

- Community guidelines – consistently curating and representing dataset quality information,
- Virtual workshop on July 13, 2020 – bringing together international domain experts,
- Contact me at gpeng@ncsu.edu if interested in participating or contributing.

Acknowledgement

- Support from the NOAA *OneStop P*roject, NCEI Data Stewardship Division and the Center for Weather and Climate, and CESISS.

- Beneficial information from Anna Milan, Nancy Ritchey, and Donald Collins.

- Invitation from RDA-US (Lynn Yarmey and Shelley Stall) to participate.

- RDA FAIR Data MM WG Members & Editorial Team!

THANK YOU!

Crosswalk Between FAIR Data Principles and RDA FAIR DMI

RDA FAIR Data Maturity Model Adoption (Impression and Experience)

Anusuriya Devaraju & Hervé L'Hours
(on behalf of FAIRsFAIR)
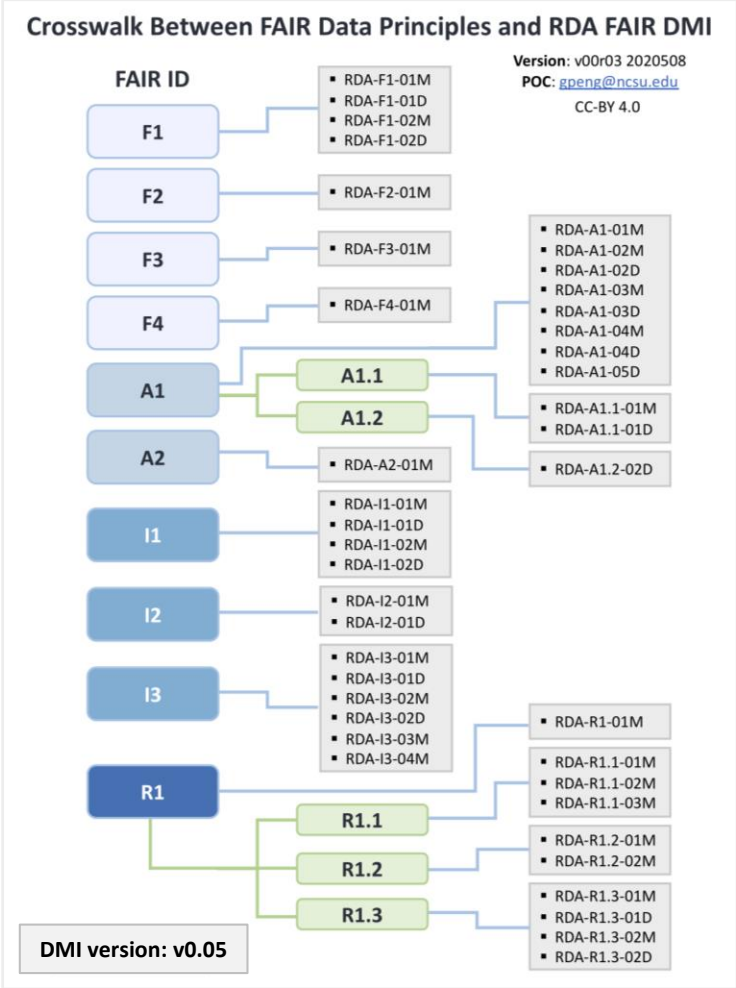
FAIRsFAIR - Fostering Fair Data Practices in Europe

- Aims to supply practical solutions for the use of the FAIR principles throughout the research data life cycle.

- Budget: €10 million

- 22 partners from 8 member states

FAIRsFAIR Involvement

- FAIRsFAIR involvement in RDA WG activities is mainly through WP4 (FAIR Certification):
  - Capability maturity models towards FAIR Certification of repositories
  - FAIR assessment of digital (data) objects : Pilots (two primary use cases)

Before

After

Researchers
FAIR awareness and education

Image: Parkjisun, Noun Project

Data Publication

Image: Michael Thompson, Noun Project

Data Repositories
Assessment of published datasets

Repository Certification

- CoreTrustSeal follows a self-assessment and peer review model

- FAIRsFAIR is offering support with a CoreTrustSeal+FAIR angle

- Map object characteristics to where repositories can enable FAIR

Repository Certification

Apply → Self-Assessment → Submit → Reviewers Assigned → Peer Review → Board Review → Request Revisions → Revision
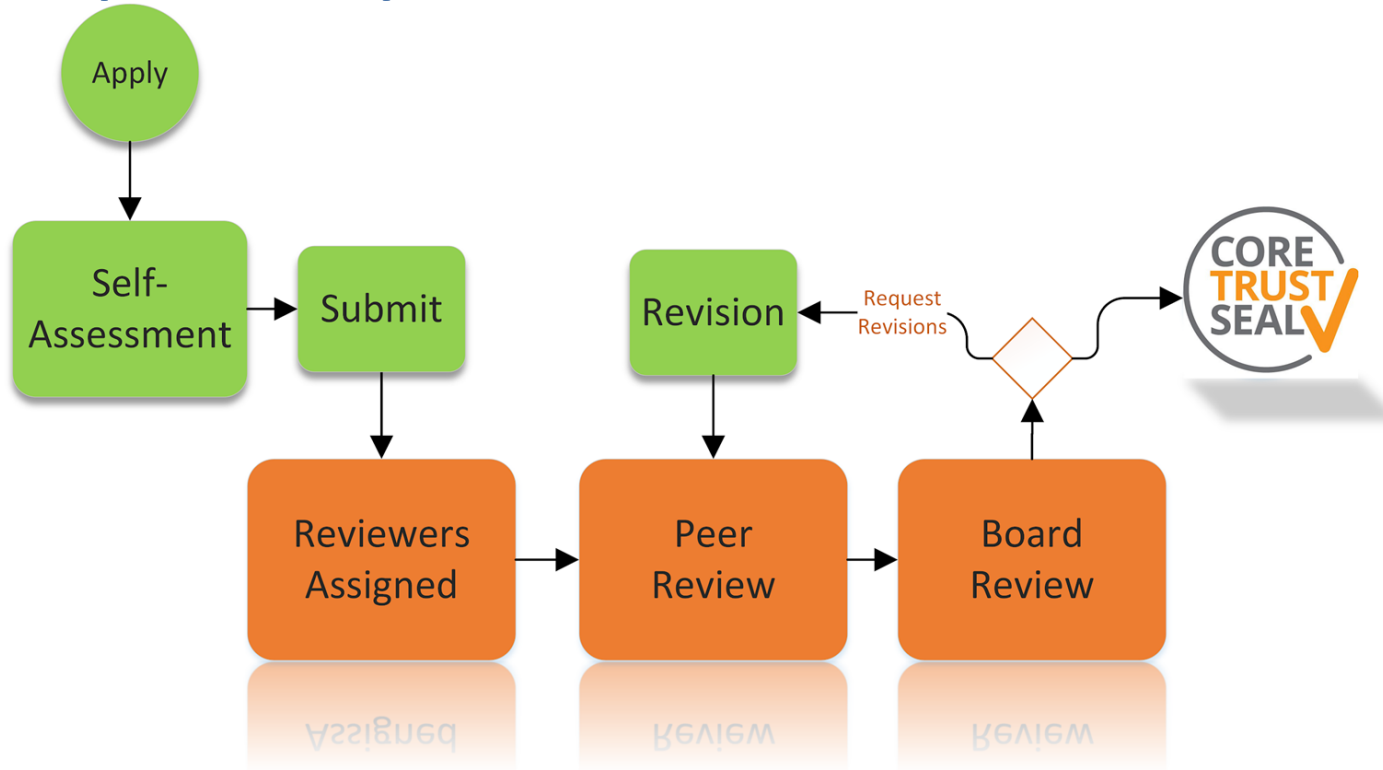
CORE TRUST SEAL ✓

HLH-CoreTrustSeal-Process_00_03.vsdx

Repository Certification

Repository Certification

Repository Certification

Repository Certification

F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. metadata specify the data identifier.
F4. (meta)data are registered or indexed in a searchable resource.
R13. Data discovery and identification

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable (vs context)
R15. Technical infrastructure
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
R16. Security
A2 metadata are accessible, even when the data are no longer available.
R10. Preservation plan

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles (vs context)
R15. Technical infrastructure (Business Information? Object Model?)
I3. (meta)data include qualified references to other (meta)data.
R11. Data quality

R1. meta(data) have a plurality of accurate and relevant attributes.
R11. Data quality
R1.1. (meta)data are released with a clear and accessible data usage license.
R2. Licenses
R1.2. (meta)data are associated with their provenance.
R7. Data integrity and authenticity
R1.3. (meta)data meet domain-relevant community standards (vs Context)
R15. Technical infrastructure

F R13 R15

A R15 R16 R10

I R15 R11

R R11 R2 R7 R15

HLH-CoreTrustSeal-FAIR-Map_01_00.vsdx
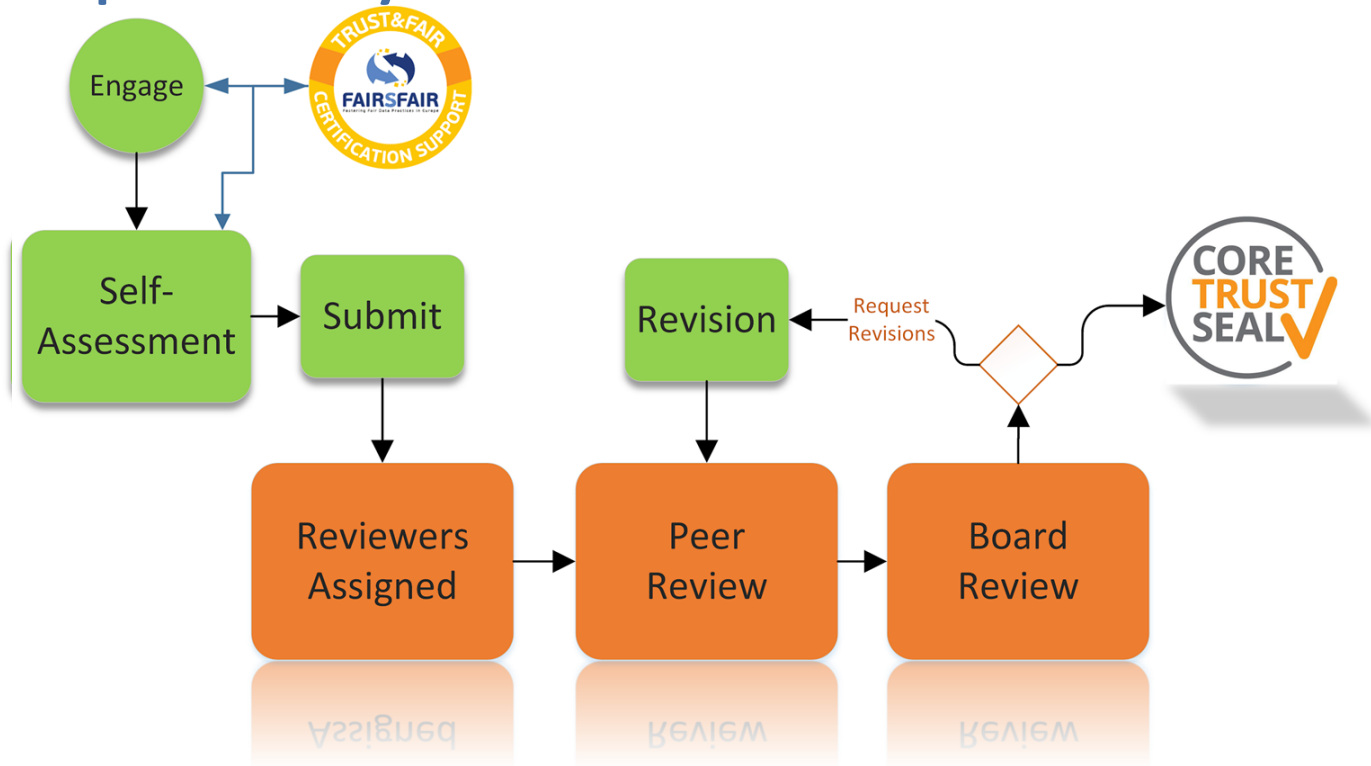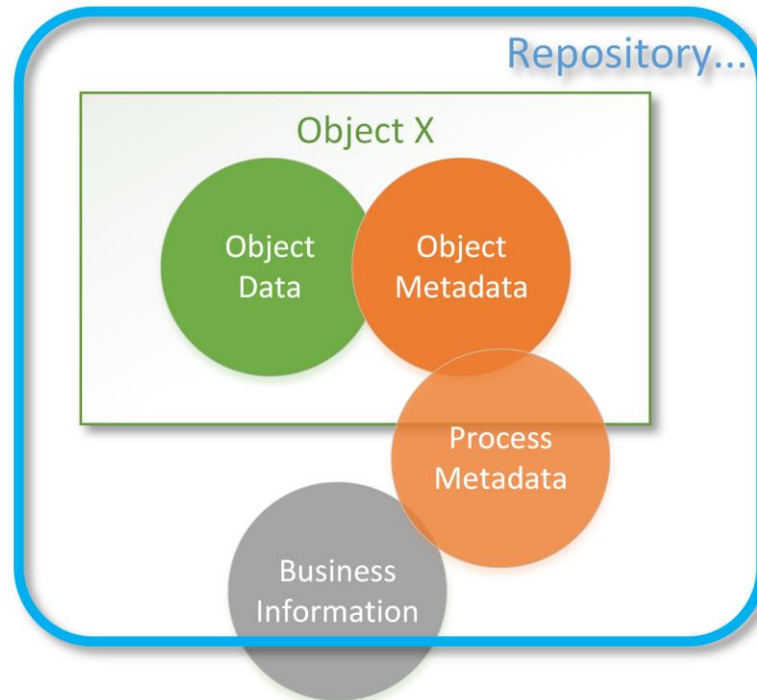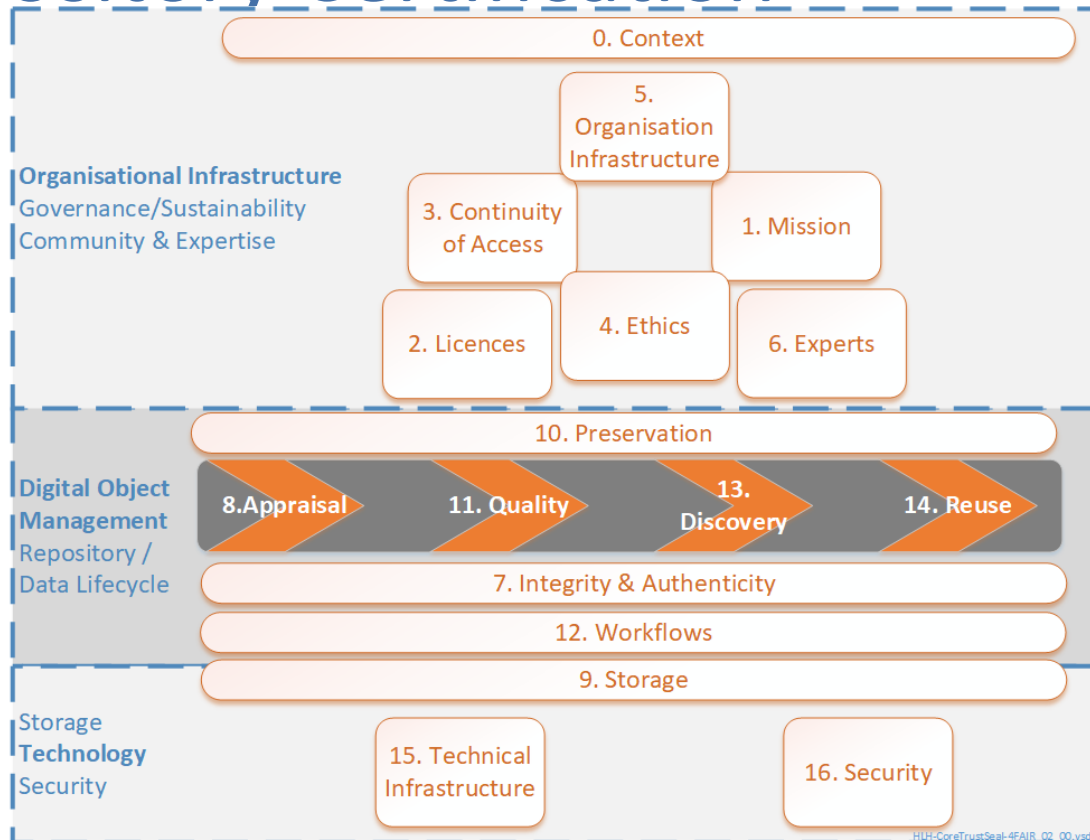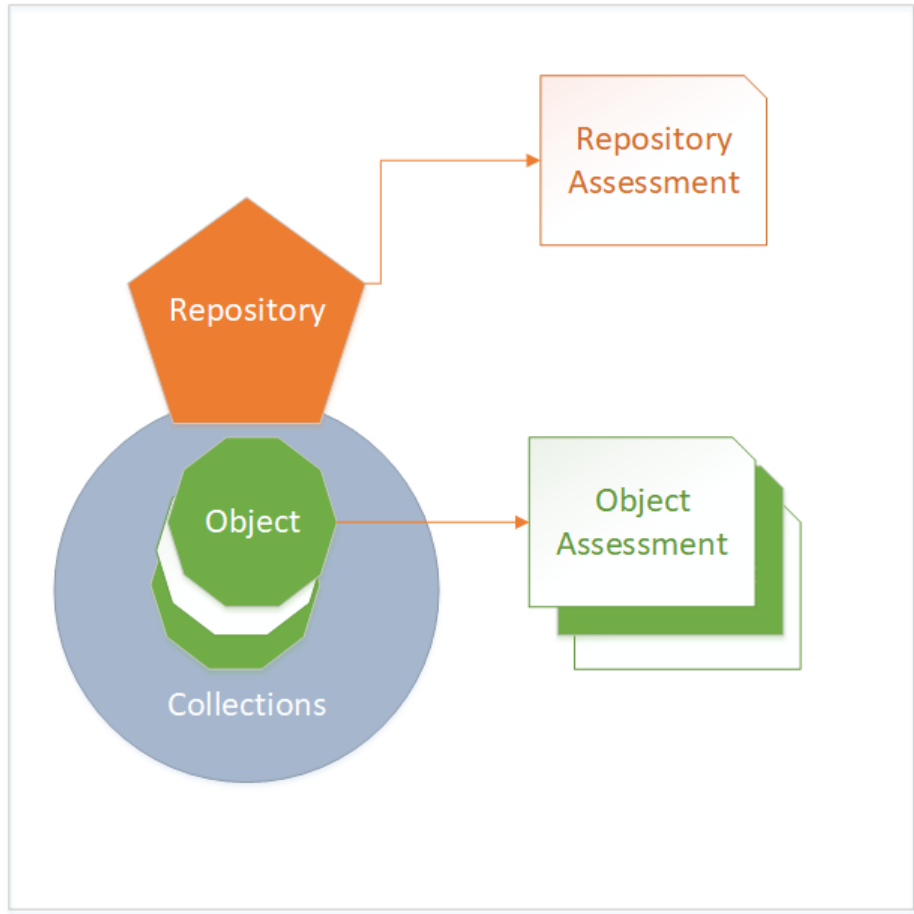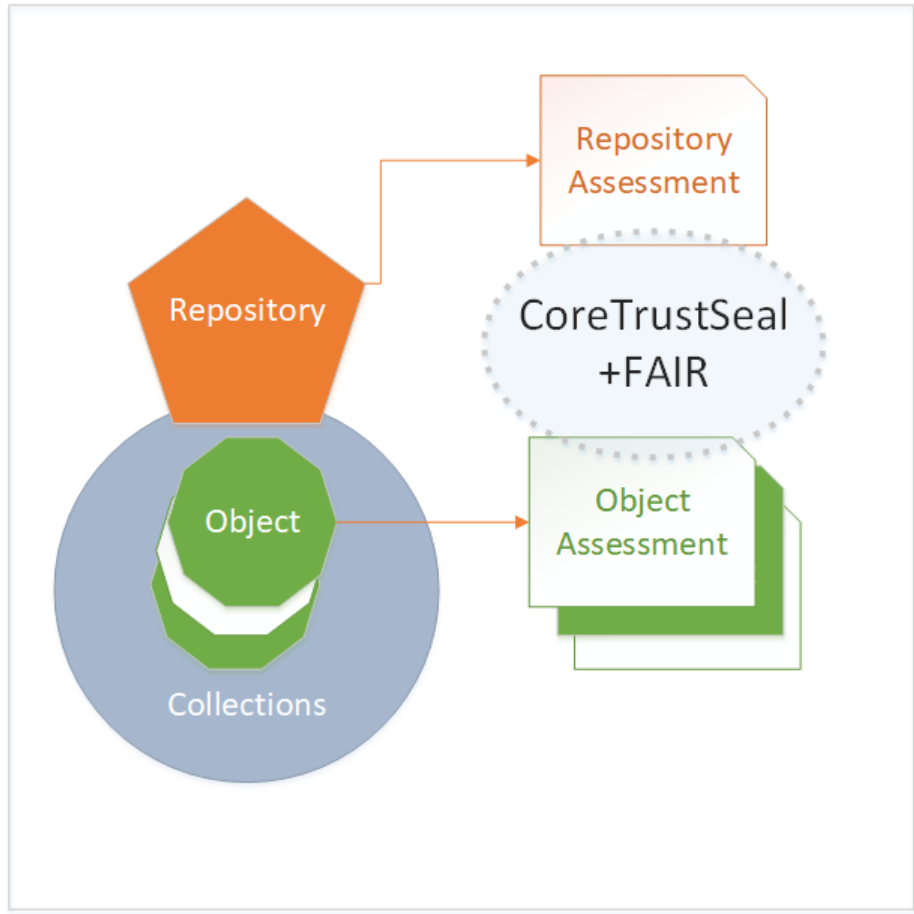
Repository Certification

- CoreTrustSeal follows a self-assessment and peer review model

- FAIRsFAIR is offering support with a CoreTrustSeal+FAIR angle

- Map object characteristics to where repositories can enable FAIR

Later:

- **Integrate object evaluation outcomes**

Overall Adoption Experience

- The recommendation should be used as a starting reference point for data FAIRness assessment.

- Presentation - specification and guidelines are well structured!

- 'What' aspect of FAIR assessment
  - Descriptions of indicators are very helpful!
  - Suggestion - Include priority level next to each of the indicators.
  - Essential I-indicators missing (needs further work or not important?)

- 'How' aspect of FAIR assessment
  - Context matters (e.g., practices, data types)
  - Assessment details not always provide sufficient detail to implement tests.
  - Potential supporting technologies and services should be described.

Overall Adoption Experience

- Future maintenance - How can adopters feed the results back to the WG?

- For more detailed feedback, see FAIRsFAIR Comments Response on RDA FAIR Data Maturity Model Working Group (2020), https://doi.org/10.5281/zenodo.3827108

To remember when applying/implementing the indicators

- Practices of identifying and locating 'objects'

  - A data object is assigned with a persistent identifier, which resolves to a landing page that includes metadata and links to access the content.

  - Persistent identifiers for all (data, metadata)

  - Data and metadata in a self-describing format.

- Indicators and priorities may be changed/extended depending on **community practices**, **users** (evaluators) and at which **stage** of the data cycle the assessment is performed.

- How can we communicate FAIR assessment results to different stakeholders in meaningful ways?

Disseminate the FDMM to communities

Working group members are encouraged to **actively distribute and promote the recommendation** – to their communities
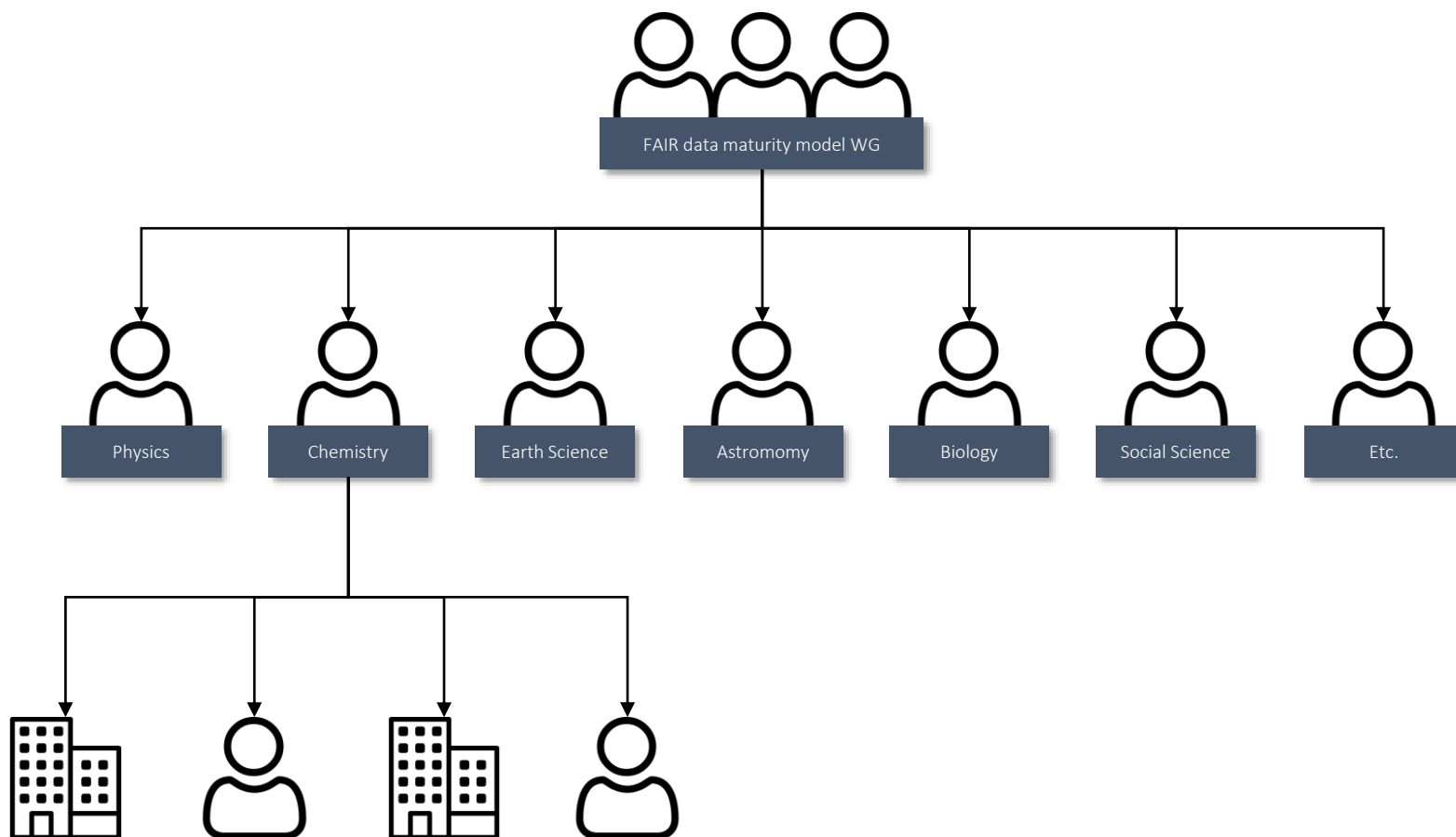
Acknowledgment

SPECIAL THANKS to the following members for their unfailing commitment and critical contribution to this WG

- Alejandra González Beltrán
- Alicia Fátima Gómez Sánchez
- Andras Holl
- Anusuriya Devaraju
- Barbara Sierman
- Carole Globle
- Françoise Genova
- Ge Peng
- Gerry Coen
- Helen Parkinson
- Hervé L'hours
- Keith Jeffery
- Kerry Levett
- Kevin Long
- Jean-Eudes Hollebecq
- Jolanda Strubel
- Jonathan Petters
- Leyla Garcia
- Marco Molinaro
- Maggie Hellström
- Mark Wilkinson
- Marta Teperek
- Michel Dumontier
- Nichola Burton
- Nick Juty
- Mustapha Mokrane
- Oya Deniz Bayan
- Peter McQuilton
- Rob Hooft
- Romain David
- Susanna Sansone
- Yann Le Franc

Resources - Updated
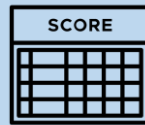
RDA Webpage	Evaluation mechanism	Public Review	RDA recommendation

> RDA FAIR data maturity model WG – [GitHub](#)

> RDA FAIR data maturity model WG – [Mailing list](#)

Action items

&

Next steps

Working Group
RDA FAIR data maturity
model Working Group
→
RDA recommendation

Maintenance Working Group
with different aim and
possibly broader *(i.e.
platform to maintain and
agree indicators and services
to support FAIR data)*

June   July

Turning the **Working Group** into an
Maintenance Working Group

RDA FDMM <u>Maintenance</u> WG | Workplan

PHASE 1 | Coordinate efforts to ensure the adoption of the FAIR data maturity model

PHASE 2 | Maintain the deliverables produced upon feedback received by the communities

PHASE 3 | Proposing governance model for handling maintenance activities in 2021 and beyond

June 20'   September 20'   December 20'

- Seek new adopters
- Liaise with other groups
- Provide assistance to communities putting in place the recommendation

- Formalise the maintenance phase
- Define requirements and tasks
- Initiate a feedback loop and derive maintenance activities
- Create a communication plan

- Seek new project sponsors (e.g. chairs, editorial team, contributors)
- Update the project workplan
- Create a hand-over plan

Action item and next steps

Working Group members are invited to:

> Reach out to their communities as for the publishing of the **FAIR data maturity model: specification and guidelines** (i.e. RDA recommendation)

> Continuously provide feedback to the Editorial Team and pass on information with regards to the use of the **FAIR data maturity model: specification and guidelines** (i.e. RDA recommendation)

The editorial team will look into a release calendar and change management schedule

---

WORKSHOP #10

Possibly **September 2020**