



***A FAIRification* Roadmap for AnaEE Denmark October 2020**

**Klaus Steenberg Larsen
Associate Professor, AnaEE Denmark Coordinator
Department of Geosciences and Natural Resource Management
University of Copenhagen**



***A FAIRification* roadmap for AnaEE Denmark - 2020**

Content:

1. Strategic Rational for a FAIRification roadmap
2. Understanding FAIRification - What is FAIR data?
3. Automated Machines Access to FAIR data
4. Upholding the FAIR Principles
5. Gradually Increasing the Levels of FAIRness
6. A necessary move from *typical research cycle*
7. ... to *improved research cycle*
8. Reassessing Organization, Technology, Politics, Economy
9. Coordinating the AnaEE FAIRification roadmap
10. Aiming for community hominization and specialization
11. Aiming for an Improved Division of Labor
12. Roadmap & shorter-term FAIRification Workplans
13. Roadmap & Technical Breakdown and To-do's
14. Local and National Data Stewardship Support
15. Formulating FAIRification Success Criteria
16. Summary of AnaEE Denmark FAIRification roadmap

Strategic Rational for a FAIRification roadmap (1)

1. When analysing impacts of climate and land-use changes on ecosystems, in pursuit of the AnaEE scientific mission, it must be better recognised that:
 - a) research **data** is increasingly fundamental for longer term global scientific advancements
 - b) global community-wide **sharing** of high quality data is fundamental
 - c) consequently, state-of-the-art **data management** becomes crucial
2. As the coordinating infrastructure for Danish scientific efforts within experimental ecosystem research, **AnaEE Denmark shall support and improve the ability for researchers to collect, share and integrate data** across different ecosystems, experiments, disciplines and institutions
3. To fulfill this goal, AnaEE Denmark shall work to implement the **FAIR data principles (Findable, Accessible, Interoperable, Reusable)**.



Strategic Rational for a FAIRification roadmap (2)

4. AnaEE Denmark shall **provide open access to experimental platforms and FAIR data** for the national as well as international research community, as well as for private industry to advance science, develop new and innovative management tools and provide scientific basis for sustainable long-term use and management of our ecosystems.
5. Specifically, AnaEE Denmark shall support and develop **the production of model-targeted, high-quality FAIR data** on ecosystem processes and functioning as part of a coordinated international infrastructure in order to improve our ability to predict impacts of future climate and environmental changes on these ecosystem services and provide guidance to the sustainable management practices.



Understanding FAIRification - What is FAIR data?

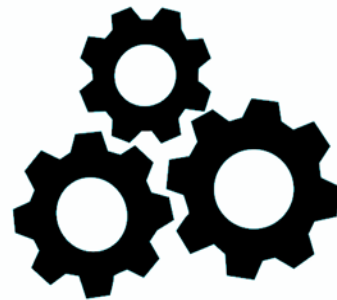
F
Findable



A
Accessible



I
Interoperable



R
Reusable

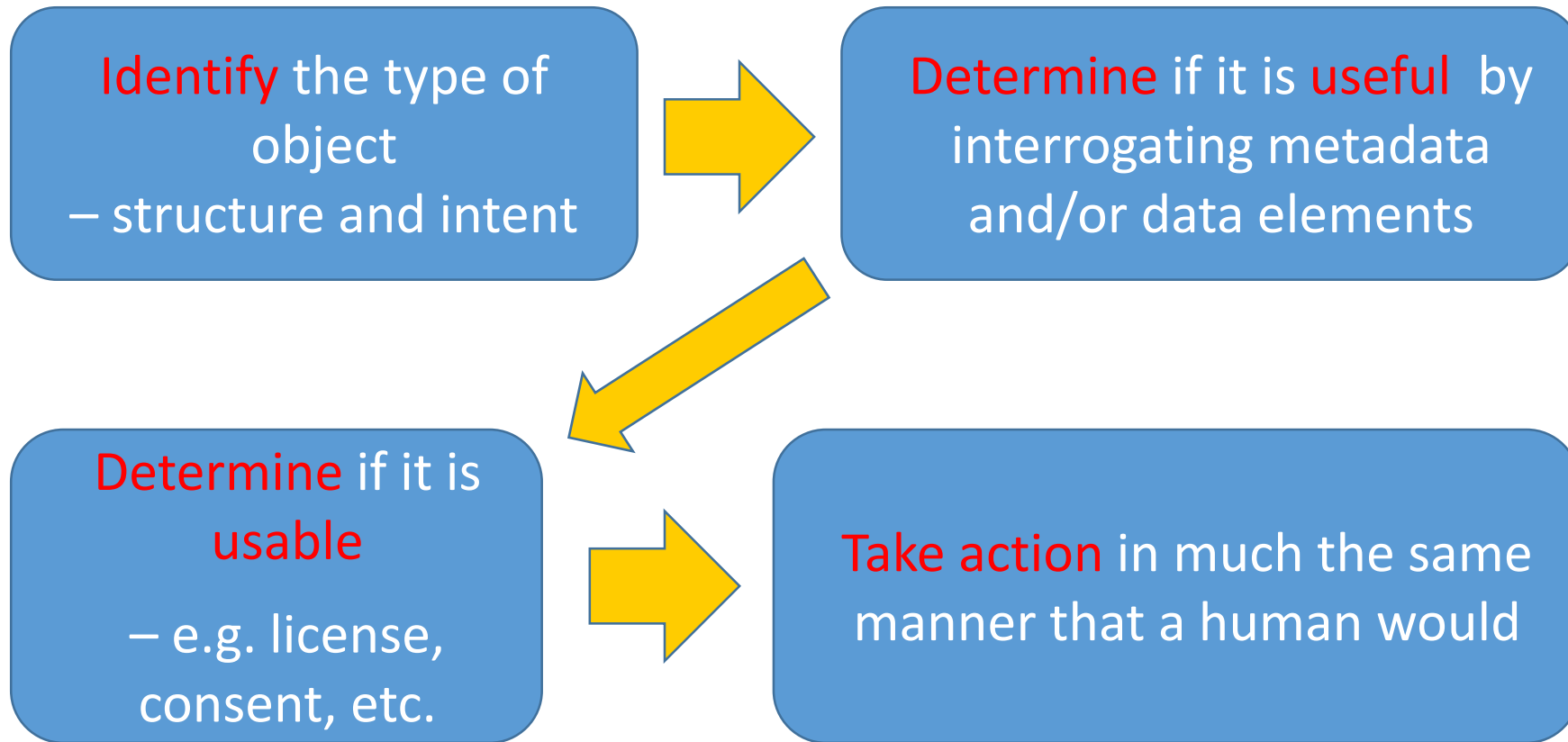


FAIR \neq Open but Open \implies FAIR
As open as possible, as closed as necessary



Automated Machine Access to FAIR data

... will allow computers and models to:



Upholding the FAIR Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

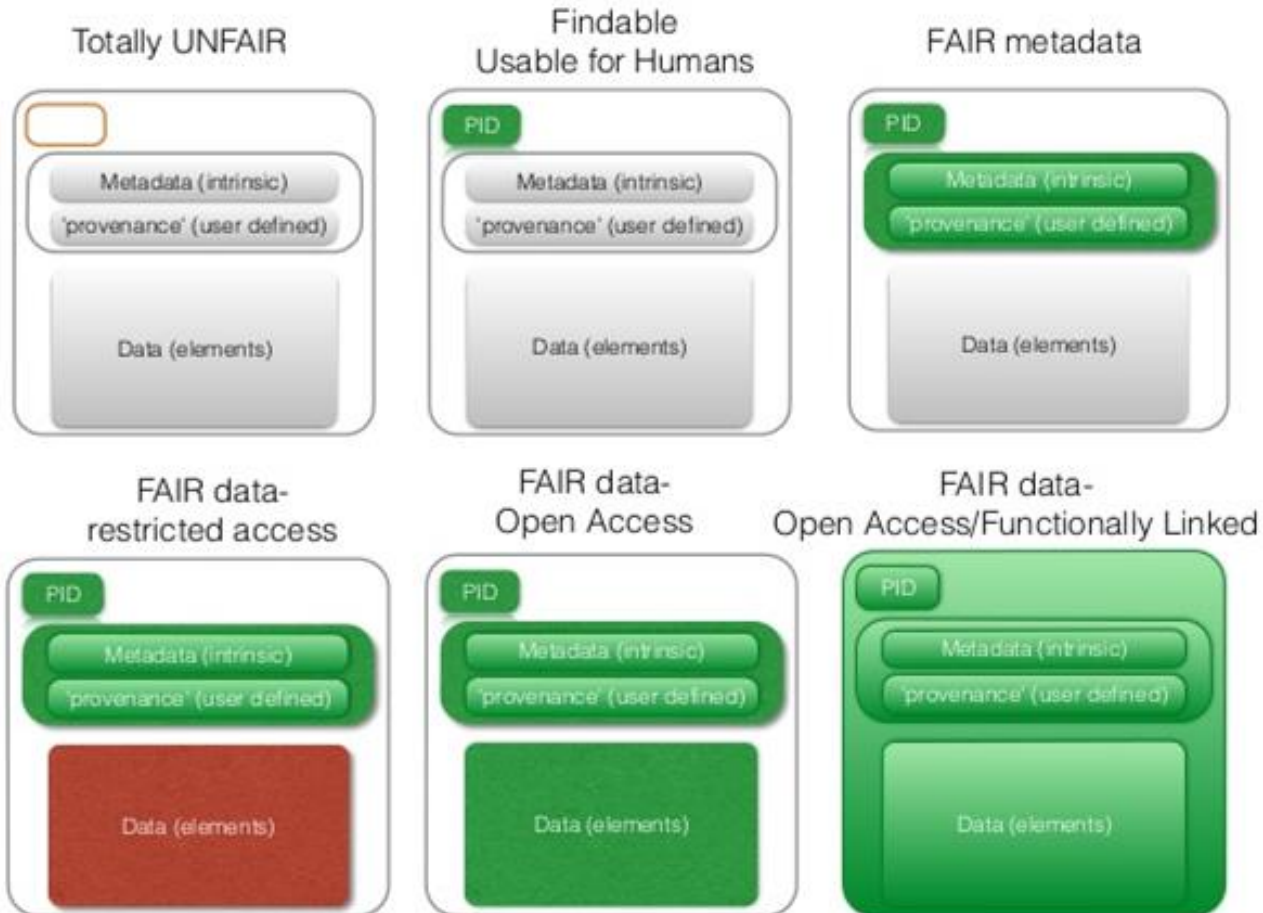
- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards



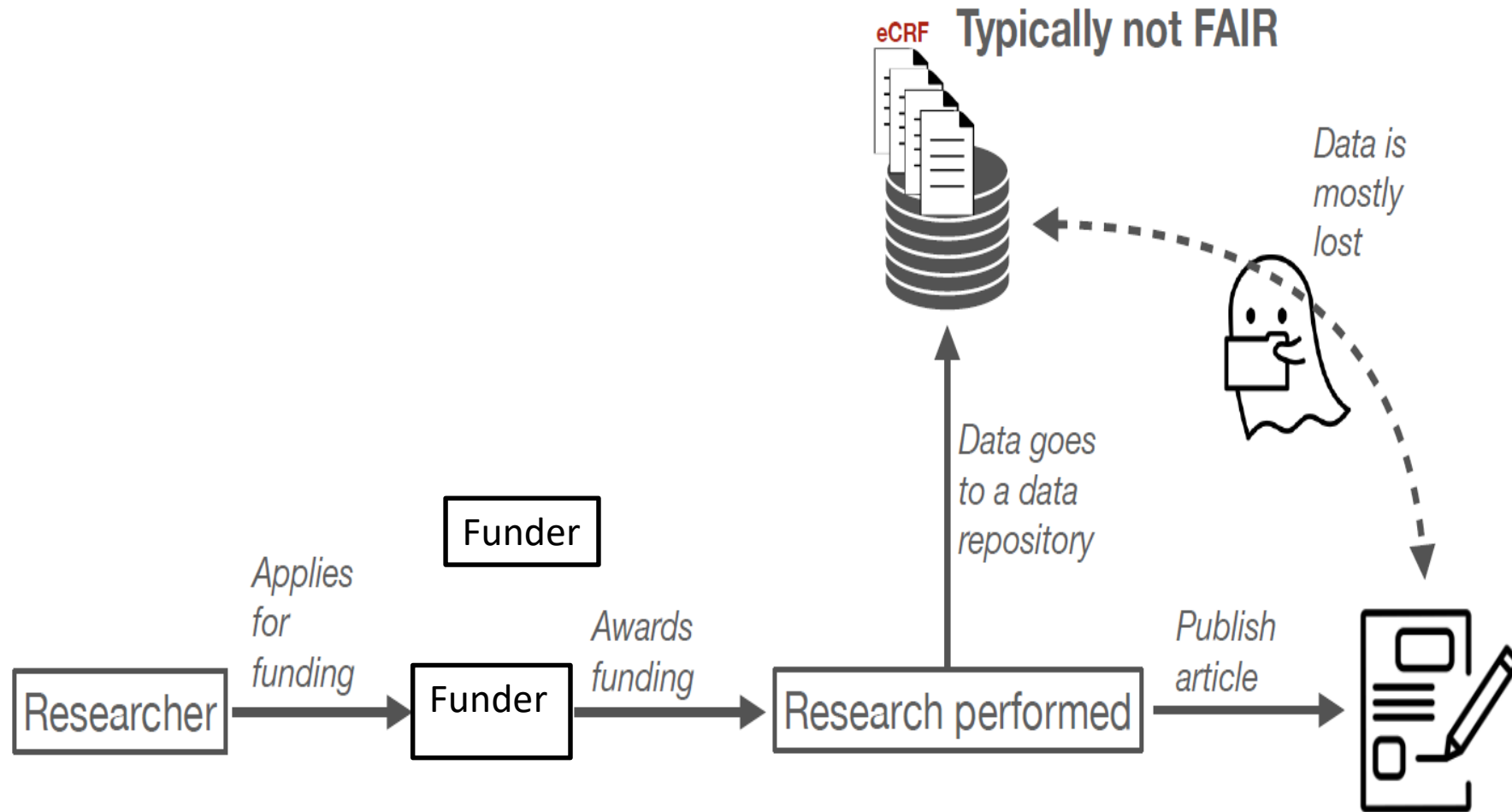
Gradually Increasing the Levels of FAIRness



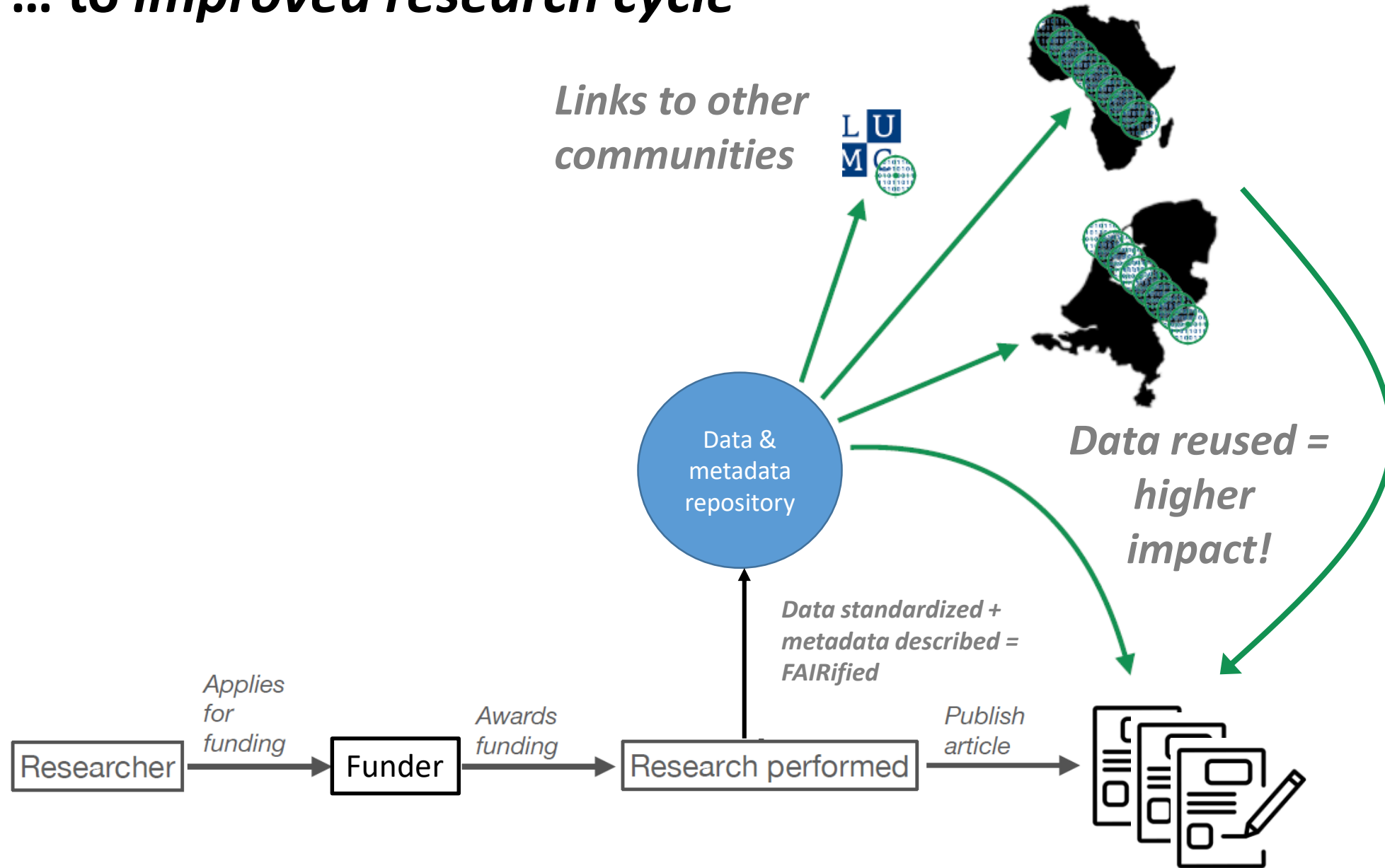
Goal: AnaEE Denmark will strive to achieve a continuous development of higher and higher levels of FAIRness in data produced on its experimental platforms.



A necessary move from *typical research cycle*



... to improved research cycle

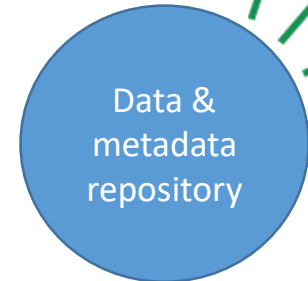


... to improved research cycle:

Goal: Reuse and INCREASE scientific usage of AnaEE Denmark generated data by creating FAIR data!

Requirements: generation of standard data and metadata templates, change in culture (clarify benefits for individual researchers)

Links to other communities



Data reused = higher impact!

Data standardized + metadata described = FAIRified



Organizational, Technological and political aspects

Organizational:

AnaEE Denmark is a **distributed research infrastructure**, organized as a project¹. Data storage, curation and access is also distributed, but shall aim for common protocols, vocabularies and metadata descriptions, and shall be anchored in AnaEE international.

Goal: To strengthen the national and international links for sharing FAIR data

Technological:

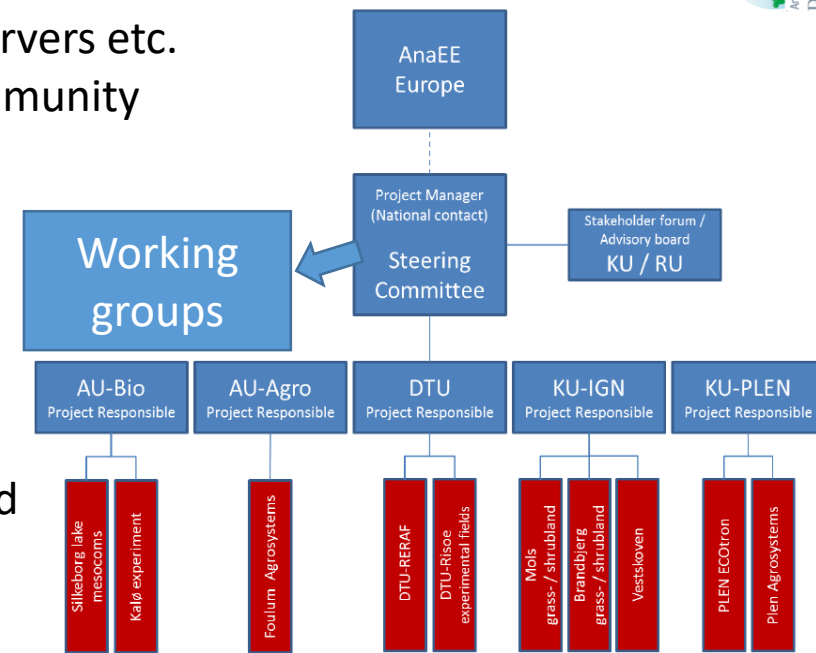
AnaEE Denmark is depended on technologies made available by other parties in terms of data repositories, DOI providers, metadata servers etc. but shall commit a Working Group to continuously enhance community competences in data management and optimal FAIRification.

Goal: The use of FAIR technologies must be enhanced, formalized and standardized

Politics and Economy:

The AnaEE Denmark partners are obliged to participate in the development of common data management, tools, protocols and templates and to and open access to AnaEE data

Goal: Obligations must be more formalized and Internationalized – politically and economically



¹ <https://anaee.dk/about/funding>

Coordinating the AnaEE FAIRification Roadmap

AnaEE Denmark is a national Danish research Infrastructure, but is also the national node for AnaEE International, which is again a partner in ENVRIFAIR together with other environmental research infrastructures.

National level: AnaEE Denmark will when feasible seek collaborations, shared knowledge and converge FAIRification of data and infrastructures with:

- other adjacent research communities (i.e. ICOS, ACTRIS, eLTER and others)
- infrastructure initiatives, like NUFI, DeiC, NeIC, EOOSC etc.

International level: AnaEE Denmark will:

- commit itself to follow the internationally agreed data policies under AnaEE International
- take active part in the development of common data management tools at the international level

Goal: AnaEE must anchor a FAIRification roadmap within its scientific mission



Aiming for Community Hominization and Specialization

AnaEE Denmark must strengthen its goal of **complying with already existing FAIR data ontologies and vocabularies**, to ensure compatibility with data from other research communities, where similar data may be produced (E.g. ICOS and WMO)

Due to the close relationship between the long-term observational data series in ICOS and the experimental data from AnaEE, AnaEE Denmark can e.g. benefit from adapting to ICOS vocabulary for commonly shared data variables

For community specific data AnaEE must aim to develop data vocabulary – including adapting as much as possible from the already existing AnaEEThes (developed by AnaEE France)



Aiming for an Improved Division of Labor

Institutional level: Each partner institution should support the development of FAIR data in projects and infrastructures by providing data and metadata repositories as well as FAIR data stewards to support the researchers.

Goal: Institutions must commit to developing better *Data Stewardship* know-how

AnaEE Denmark level: The project manager coordinates AnaEE Denmark and the Data Management Working Group, and alignment of national policies with AnaEE International. AnaEE Denmark partners representatives in the Data Management working group are responsible for each partners activity, including data management and roll-out of agreed policies and technologies.

Goal: More clarity in data management obligations and division of labor

National level: AnaEE Denmark collaborates with other related communities to converge continuous development – with DeIC facilitating national Data Stewardship competence.

Goal: More clarity in obligations and division of labor

International level: AnaEE International guides the research agenda with input from national nodes, and coordinates with other international communities in ENVRIFAIR.

Goal: AnaEE International must pursue a FAIRification within division of labor



Roadmap & Shorter-Term FAIRification Workplans

Aimed for Milestones < primo 2021	Date
AnaEE Denmark launched	1 Jan 2018
First FAIR data project initiated (with DeiC)	1 Jul 2019
3 meteorological data sets FAIRified	18 Nov 2019
M4M workshop with GOFAIR, DeiC and NEST	30 Sept 2020
Machine-readable metadata template finalized	1 Oct 2020
Metadata template (as csv) for greenhouse gas fluxes	1 Oct 2020
Make new metadata available on landing pages (or elsewhere)	20 Oct 2020
CEDAR template of GHG fluxes	25 Oct 2020
Autofilling of metadata based on ORCID (with NEST, JP, KSL, John Greyball)	1 Nov 2020
Roll-out use of template with student project data and own data (All, 1 Nov 2020)	1 Nov 2020
More data variables in vocabulary	primo 2021 - 2022
Alignment with development of AnaEE Int	primo 2021 -



Roadmap & Technical Breakdown and To-do's

1. Finalize current metadata as CEDAR template (JP, 1 Oct. 2020)
2. Metadata template (as csv) for greenhouse gas fluxes (KSL, 1 Oct 2020)
3. Mini-project on KU using Dataverse software (Richard Dennis, Falco Hüser, KSL, 15 Oct 2020)
4. Integration of new metadata on existing data repositories (or elsewhere) (JP, KSL, John Graybeal, 20 Oct 2020)
5. Help with CEDAR when needed for the development of a template for greenhouse gas flux data (KSL, JP, AT, John Graybeal, 25 Oct 2020)
6. Continued development of ontologies, ICOS and AnaEE DK specific where needed, for proper implementation in BioPortal and full usability with CEDAR (JP, John Graybeal, 25 Oct 2020)
7. Auto filling of metadata based on ORCID (with NEST, JP, KSL, John Graybeal, 1 Nov 2020)
8. Roll-out use of template with student, PhD, and project data (All, 1 Nov 2020)
9. More data variables in vocabulary (primo 2021 - 2022)
10. Alignment with development of AnaEE Int. (primo 2021 -)



Local and National Data Stewardship Support

AnaEE Denmark has through partner projects and workshops (DeiC nationally and GOFAIR internationally) trained a number of scientists who can now assist other researchers within the AnaEE Denmark community with FAIRifying their data. These people will facilitate further FAIRification of more data in AnaEE Denmark.

This FAIRification progress can aid AnaEE Denmark in acquiring a stronger international role and greater scientific impact.

However, for faster and larger scientific impact through FAIRification AnaEE must:

- Request university support functions in the form of competent **Data Stewards** that can continuously support the generation of new data, metadata templates and data infrastructure
- Request additional even **higher specialized Data Stewardship** most efficiently applied at the level beyond the individual project, departments and universities, but rather pooled at the national or international level (i.e. data stewards should have critical mass of competence and serve multiple projects for optimal convergence between projects and communities)
- **push its institutions to provide this organizational support.**



Formulating FAIRification Success Criteria

Key Performance indicators, could be:

- Number of data variables with standardized vocabularies and metadata descriptions
- Number of data sets made openly accessible with DOI's
- Number of data sets made openly accessible with DOI's from PhD projects
- Number of data sets made openly accessible with machine-readable metadata
- Number of ORCID's among data publishers using AnaEE Denmark tools and templates
- Number of times datasets are referenced



Summary of AnaEE Denmark FAIRification Roadmap

