Towards FAIRness: some reflections from an Earth Science perspective

Maggie Hellström
ICOS Carbon Portal
(& ENVRIplus & SND & Lund University…)

Good data management in the Nordic countries
Stockholm, October 3 2018

ICOS Carbon Portal

FAIR – what is it good for?

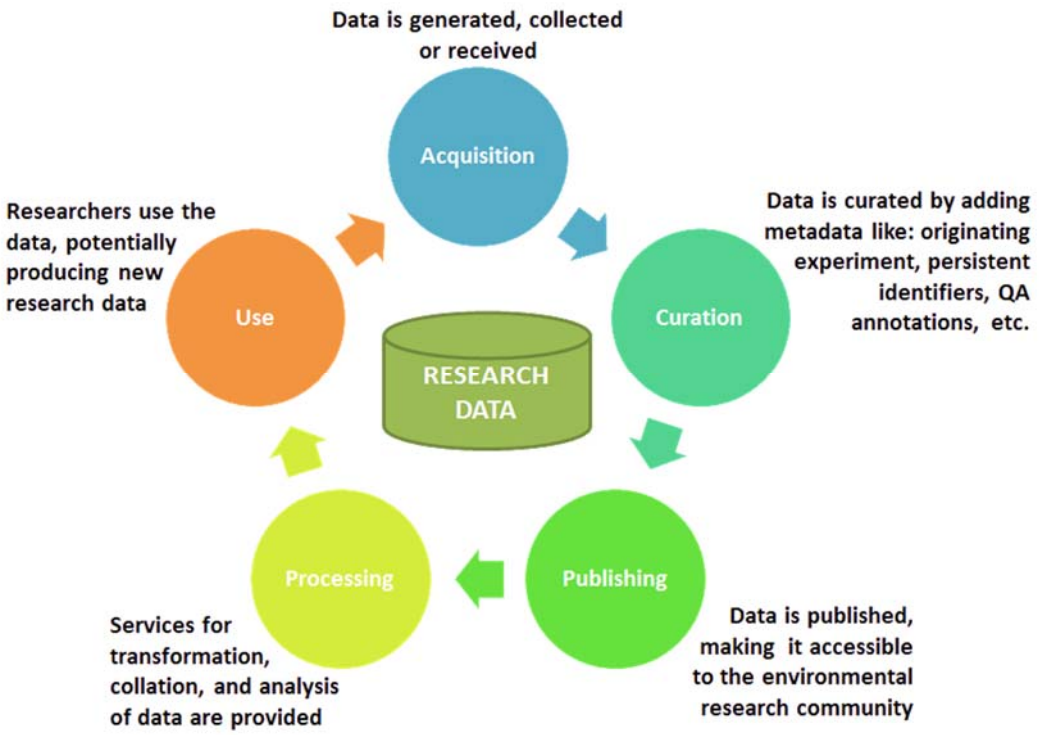
- stands for Findable, Accessible, Interoperable, Reusable
- not a standard, but a set of principles coined by FORCE11 in 2014, out of discussions in the Life Sciences community
- is increasingly called for by funders & policy makers, and is the focus of a recent report commissioned by the EC, entitled "Turning FAIR data into reality" <- check it out!
- this report contains 34 recommendations covering the topics
  - Primary Recommendations and Actions
  - FAIR data policy
  - FAIR data culture
  - Technology for FAIR
  - Skills and roles for FAIR
  - FAIR metrics
  - Costs and investment in FAIR

* FORCE11, 2014 (https://www.force11.org/fairprinciples)
* High-Level Expert Group on FAIR, see http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3464

plus a tip of the hat to Edwin Starr and his 1974 album "War & Peace" (https://www.youtube.com/watch?v=dpWmlRNfLck) …

The research data lifecycle

- FAIR has implications for all stages of the data lifecycle
- Many different actors are involved:
  - data producers
  - curators
  - repositories
  - e-service providers
  - end users
  - …
- Data Management Plans (DMPs) should cover all of the related activities



ENVRIplus project output, see e.g. deliverable D5.1 available via http://www.envriplus.eu/deliverables/

What does SND think about FAIR?

SND, the Swedish National Data service, is actively promoting FAIR

"Our vision is to be part of a global network through which researchers can easily share, find, access and reuse high-quality research data – now and in the future. We will make this happen by inspiring, energizing, and coordinating the development of a trustworthy system of research data repositories in Sweden."

Important steps towards this include:

- Facilitate and Secure Research Data and Metadata Flows
- Assist and Benefit Users
- Accumulate and Maintain Expertise

Maggie Hellström, 2018-10-03

ICOS Carbon Portal

What do the (ENV)RIs think about FAIR?

The European community of environmental & Earth science research infrastructures recognizes the importance of making their data – and services – FAIR as soon as possible

A new cluster project, ENVRI-FAIR, was proposed and will receive H2020 funding from January 2019. Some overarching goals are:

- While recognizing that the participant RIs are at different levels of maturity, it is still possible to define common requirements and collaborate towards implementing increased FAIRness for Earth science data
- Well defined community policies and standards are needed for all steps of the data life cycle; these policies and standards must be aligned with both European policies and international developments.
- Each participating RI will aim to set up and operate sustainable, transparent and auditable data services, for each step of data life cycle, compliant to the FAIR principles.

Maggie Hellström, 2018-10-03

What do scientists think about FAIR?

The reactions from the scientific communities to increased calls for FAIRness of their data are mixed. Most are cautiously positive, but there are also some quite reactionary views. Unfortunately, there is a lot of confusion…

Some recent reactions to the proposed Lund University policy for research data:

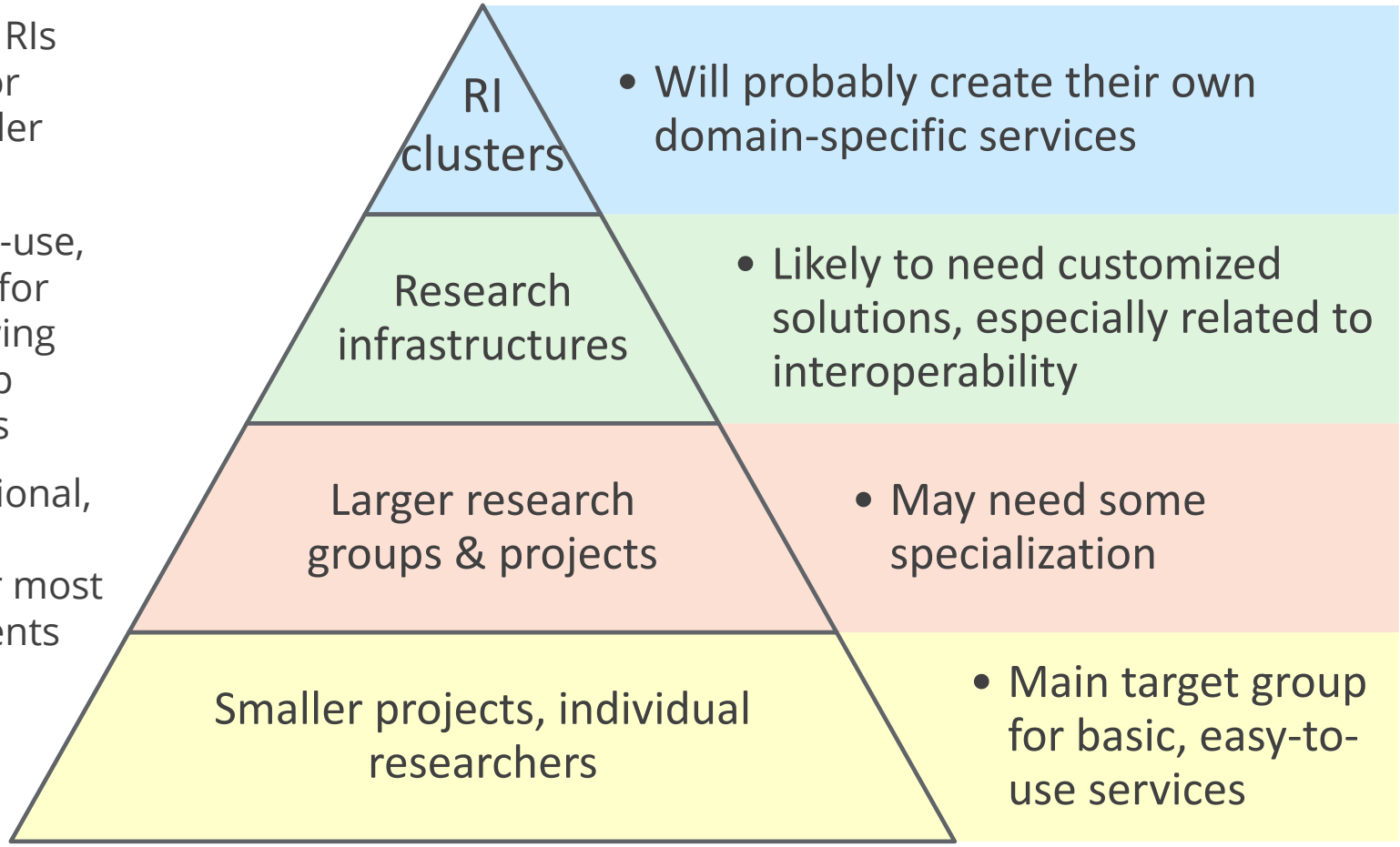
- FAIR and Open Science are probably good in principle, but…
- Is it really necessary to make *all* data FAIR?
- Is there an overlap between requirements for FAIR and archiving?
- How should we/the universities/Europe implement FAIR in practice?
- From when should it start? (I.e. does it apply to “old data”, and do we need to redo or update previous actions?)
- Who pays for the associated “extra” work?
- Who is responsible for getting the work done?
- Is there going to be a penalty if we fail?

“Scale” of the research context matters!

What is adequate for RIs may not be useful – or achievable – for smaller groups

Need to offer easy-to-use, streamlined services for the latter, while allowing the former to develop their own alternatives

A combination of national, European and global services should cover most needs and requirements



ICOS Carbon Portal

So what about these DMPs…?!

Wikipedia says:

"A **data management plan** or **DMP** is a formal document that outlines how data are to be handled both during a research project, and after the project is completed. The goal of a data management plan is to consider the many aspects of data management, metadata generation, data preservation, and analysis before the project begins; this ensures that data are well-managed in the present, and prepared for preservation in the future."

https://en.wikipedia.org/wiki/Data_management_plan

So what about these DMPs…?!

Wikipedia says:

"A **data management plan** or **DMP** is a formal document that outlines how data are to be handled both during a research project, and after the project is completed. The goal of a data management plan is to consider the many aspects of data management, metadata generation, data preservation, and analysis before the project begins; this ensures that data are well-managed in the present, and prepared for preservation in the future."

https://en.wikipedia.org/wiki/Data_management_plan

A slightly less formal definition:

"A DMP is a plan for organizing, storing and sharing data."

Recommendations from SUHF

SUHF (the Association of Swedish Higher Education Institutions) recently published a set of recommendations related to Data Management Plans (DMPs).

The overarching goal is to ensure that management, storage, accessibility and archiving of research data – as supported by the HEI infrastructures and support services – will follow the FAIR principles

They propose that the following topics be included by researchers based in Sweden applying for grants:

- Description of the data and its acquisition and/or reuse of previously collected data
- Documentation of the data and its quality
- Storage and backups
- Ethical and legal aspects
- Accessibility and archiving

SUHF

REK 2018-1 Rekommendation för datahanteringsplan, available from http://www.suhf.se/publicerat/rekommendationer

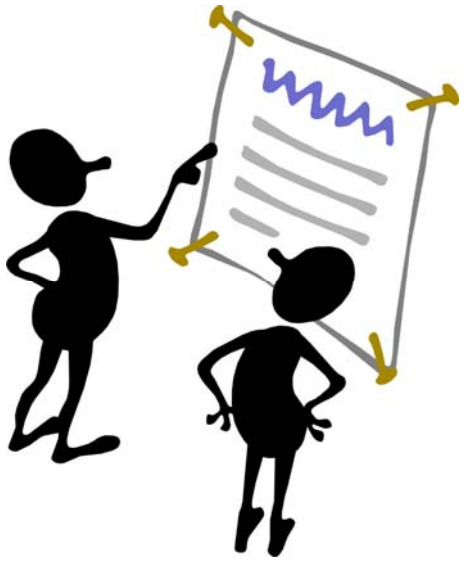
The views of the EC on DMPs & FAIR…

- Data Management Plans (DMPs) are a key element of good data management.
- A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project.
- DMPs should be revised periodically throughout the project
- As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on:
  - the handling of research data during and after the end of the project
  - what data will be collected, processed and/or generated
  - which methodology and standards will be applied
  - whether data will be shared/made open access and
  - how data will be curated and preserved (including after the end of the project).

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

The H2020 DMP model

1. Data summary
2. FAIR Data
   - Making data findable, including provisions for metadata
   - Making data openly accessible
   - Making data interoperable
   - Increase data re-use (through clarifying licences)
3. Allocation of resources
4. Data security
5. Ethical aspects
6. Other national/funder/sectorial/departmental procedures for data management



To support researchers, a number of European services are proposed as good options

- Metadata Standards Directory provided by the Research Data Alliance
- license wizard included in the EUDAT B2SHARE tool
- re3data.org registry of research data repositories
- Zenodo repository service (operated by OpenAIRE and Cern)
- ScienceMatters scientific publishing platform

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

H2020 DMP “problem areas”?!

1. Data summary
2. FAIR Data
   - Making data findable, including provisions for metadata
   - Making data openly accessible
   - Making data interoperable
   - Increase data re-use (through clarifying licenc
3. Allocation of resources
4. Data security
5. Ethical aspects
6. Other national/funder/sectorial/departmen

To support researchers, a number of European s

- Metadata Standards Directory provided by the Research Data Alliance
- license wizard included in the EUDAT B2SHARE tool
- re3data.org registry of research data repositories
- Zenodo repository service (operated by OpenAIRE and Cern)
- ScienceMatters scientific publishing platform

- Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of […] persistent and unique identifiers [..]?
- Will search keywords be provided that optimize possibilities for re-use?
- What metadata will be created?

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

H2020 DMP “problem areas”?!

1. Data summary
2. FAIR Data
   - Making data findable, including provisions for metadata
   - Making data openly accessible
   - Making data interoperable
   - Increase data re-use (through clarifying licenc
3. Allocation of resources
4. Data security
5. Ethical aspects
6. Other national/funder/sectorial/departmen

To support researchers, a number of European s

- Metadata Standards Directory provided by th
- license wizard included in the EUDAT B2SHAR
- re3data.org registry of research data reposito
- Zenodo repository service (operated by Open
- ScienceMatters scientific publishing platform

- Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared […] explain why […].
- What methods or software tools are needed to access the data [and is it] possible to include the relevant software (e.g. in open source code)?
- Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories […]

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

H2020 DMP “problem areas”?!

1. Data summary
2. FAIR Data
   - Making data findable, including provisions for metadata
   - Making data openly accessible
   - Making data interoperable
   - Increase data re-use (through clarifying licenc
3. Allocation of resources
4. Data security
5. Ethical aspects
6. Other national/funder/sectorial/departmen

To support researchers, a number of European s

- Metadata Standards Directory provided by th
- license wizard included in the EUDAT B2SHAR
- re3data.org registry of research data repositc
- Zenodo repository service (operated by Open
- ScienceMatters scientific publishing platform

- Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. […]?
- What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?
- In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

H2020 DMP “problem areas”?!

1. Data summary
2. FAIR Data
   - Making data findable, including provisions for metadata
   - Making data openly accessible
   - Making data interoperable
   - Increase data re-use (through clarifying licenc
3. Allocation of resources
4. Data security
5. Ethical aspects
6. Other national/funder/sectorial/departmen

- How will the data be licensed to permit the widest re-use possible?
- Are the data produced and/or used in the project useable by third parties, in particular after the end of the project?
- How long is it intended that the data remains re-usable?

To support researchers, a number of European services are proposed as good options

- Metadata Standards Directory provided by the Research Data Alliance
- license wizard included in the EUDAT B2SHARE tool
- re3data.org registry of research data repositories
- Zenodo repository service (operated by OpenAIRE and Cern)
- ScienceMatters scientific publishing platform

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

So can DMPs help dispel the worries & confusion?

Yes, probably!

But other things will also be necessary:

- training on data management (not just FAIR)
- concrete examples & “success stories”
- supporting (e-)services helping with
  - DMPs
  - metadata (profiles, cataloguing, maintenance)
  - sustainable storage
  - data discovery
- clear formulation of expectations and “rules of engagement”
- definitions of metrics for evaluation of FAIRness & Open Science
- Merit systems in place, allowing quantifiable professional credit for data sharing

Figure courtesy of Dexlab Analytics (http://www.dexlabanalytics.com)

Thanks for listening!

If you have questions, comments or criticisms, don't hesitate to get in touch – just drop me a line at margareta.hellstrom@nateko.lu.se !

The FAIR principles

TO BE FINDABLE:

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

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 RE-USABLE:

R1. meta(data) have 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 their provenance.
R1.3. (meta)data meet domain-relevant community standards.

https://www.force11.org/group/fairgroup/fairprinciples

The research data lifecycle

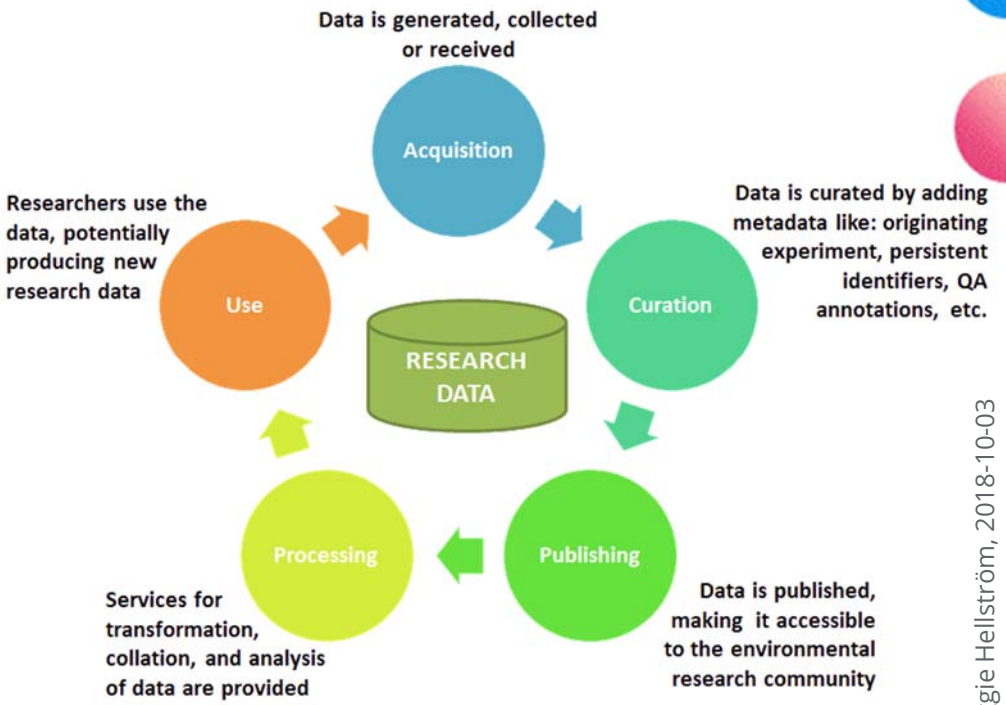
To be able to make good choices supporting efficient and sustainable management of data (and other research-related resources), some knowledge and understanding of underlying "theory" is needed:

Basic "supporting pillars"

- Curation
- Identification & citation
- Processing
- Cataloguing
- Provenance

Cross-cutting "beams"

- Architecture design
- Common data models
- Common vocabularies
- Provisioning of compute, storage, networking



It is not realistic to expect individual researchers, or even research groups & projects, to dive into details – but data management specialists at the national and HEI level should do so

Maggie Hellström, 2018-10-03

ENVRIplus project output, see e.g. deliverable D5.1 available via http://www.envriplus.eu/deliverables/

ICOS Carbon Portal

Maggie's hat collection…

ICOS (Integrated Carbon Observation System) is a pan-European research infrastructure with a focus on greenhouse gas observations The data center of ICOC, the Carbon Portal, is hosted by Sweden and located at Lund University. I'm one of the Carbon Portal "data officers".

ENVRIplus is a H2020 cluster project, bringing together 20+ Ris in the Earth Science sector. The project aims to find common solutions and services addressing ENVRI needs in research data management. I co-lead a work package on identification & citation in the "Data for science" theme (and will lead the WP on training in the upcoming ENVRI-FAIR project).

The Swedish National Data Service (SND) now has a remit to cover also environmental, climate and geosciences. Since 2018, I'm one of the SND domain specialists responsible for this domain, working mainly with a national focus but also for Lund University