Open Science
The Internet for Social Machines
The end of data sharing as we know it

FAIR > The Machine Knows what I Mean

Barend Mons
Gdansk, October 2019
The Future belongs to Data Stewards ??
Box 2 | The FAIR Guiding 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

DOI: 10.1038/sdata.2016.18
Implementation areas for data stewardship

**Policy**
- Define policies at institutions, for specific data/use cases, FAIR data, GDPR

**Research**
- Adopt the appropriate data workflows, tools, standards, infrastructure

**Infrastructure**
- Facilitate software, hardware, services, technical infrastructure

- **Aligning researcher’s data handling and data policies**
- **Aligning data policies and features of data services and infrastructures**
- **Aligning researcher’s needs and required data infrastructure**

- Policy makers
  - Funders
  - EU
  - (Applied) University boards
  - (Applied) University deans

- Scientists
- Data scientists

- Application managers
- Technicians
- Infrastructure providers
- IT personnel
The seven capital sins of Open Science
Age factor...Reward only narrative.com
2: Ignore complexity and existing data
Disrespect other disciplines
4: publish data without a supplementary paper
5: create a nightmare for machines
refuse to invest in research - infrastructure
7: Create Data without a Data Stewardship plan
Machines ‘know’ what it means

FAIR
Advances in Digital Objects

Kahn & Wilensky papers on DO  OECD Guidelines  Riding the Wave Report  Start RDA RDA DFT  DAITF Workshop 2012


Concept Web Alliance  Nano Publication Model

Start RDA DF DFT Report  start C2CAMP  DOIP V2 launched  DO testbed

GEDE

C2CAMP

Data Fabric

RDA DFT

EOSC

FAIR at FORCE11

Lorentz Workshop FAIR

FAIR in Scientific Data  Start GOFAIR  EOSC Gov

GO FAIR

EOSC IN Meeting

Advances in FAIR Principles

<<DFObject>> Metadata

<<DFType>>

FAIR Digital Record

Globally Unique, Persistent and Resolvable Identifier

Resource
FAIR and GO FAIR

Lorentz

EOSC

FAIRdICT

IFDS

Birth
2014

Infancy
2015

Adolescence
2016

Maturity
2017
2018…
The Road to FAIRness

From a few cars
The Road to FAIRness

To congestion?
FAIR

Machines know what it means

1. what (type) is it?
2. what operations are possible?
3. what operations are allowed?

Semantic descriptions
Types/operations

<<FDTType>>
FAIR Digital Record

<<FDF Object>>
Metadata

<<FDF Object>>
Globally Unique, Persistent and Resolvable Identifier

Resource

Digital/physical Resource/Object/Construct
Many different DO’s of the type ‘metadata’ can be stored in many Repositories and point to the same DO.
Digital Twins are a type of Digital Objects

Physical Objects Can have Digital Twins

Metadata Are Digital Objects

FAIR Digital Objects are stored In FAIR enabled repositories (FAIR data points/stations)

Can be a physical Object

Can be a Digital Object
Primary publication of (FAIR Digital Twins of) Research Objects (RO)

Independently Citable/awardable

All metadata also linked to article
Community Adoption

1. Many domain specific communities developing novel open science practices

2. Proposed Good Practices (standards, protocols, etc.)

3. Convening experts, community review of schema

4. Transparent procedures to formalise, document

5. Endorsed as Good Practice

SDG’s

1-5: community driven
6-7: voluntary community adoption

1. many domain specific communities developing novel open science practices
2. Early Implementation
   - Proposed Good Practices (standards, protocols, etc.)
   - Intellectual Design
3. consensus building
   - transparent procedures to formalise, document
4. convening experts community review of schema
5. proposed as Good Practice
   - good practice (data related)
6. Authority/funder recommendation
   - feed back and advocacy to the scientific and innovation community
7. Community Adoption
   - convergence matrix
many domain specific communities developing novel open science practices

1. Proposed Good Practices (standards, protocols, etc.)

2. Early Implementation

3. Transparent procedures to formalise, document

4. Convening experts, community review of schema

5. Good practice (data related)

6. Authority/funder recommendation

7. Feedback and advocacy to the scientific and innovation community

8. International Science Council endorsed as Good Practice

9. Schema based on Best Practice

10. License

11. Accreditation bodies

12. Certification partners

WG/IG outcomes

learning by doing

funders require certified services in DS plans

Community Adoption

Convergence matrix
End users are domain scientists

DWARF

- applications and services
- VMNs, Workflows Algorithms
- Data (formats)
- Long tail metadata
- Domain
- Core
- FDF
- Metadata
- AAI
- Network (routing)
- Repositories
- Files, Cloud, FDO storage
- Flash, DNA, Tapes, (HP)C

Geant
End users are domain scientists

DWARF

- applications and services
- VMNs, Workflows Algorithms
- Data (formats)
- Long tail metadata
- Domain
- Core
- Metadata
- AAI
- Network (routing)
- Repositories
- Files, Cloud, FDO storage
- Flash, DNA, Tapes, (HP)C

Geant
End users are domain scientists

NPOS+
- applications and services
- VMNs, Workflows Algorithms
- Data (formats)
- Long tail metadata
- Domain
- Core
- Metadata
- AAI
- Network (routing)
- Repositories
- Files, Cloud, FDO storage
- Flash, DNA, Tapes, (HP)C

Geant

Example NL (Health)

> 10 DSCC’s
End users

Metadata = FDO

Generic templates

Basic genomics

Bioinformatics
The Future belongs to Data Stewards !!
FAIR made easy