LET'S BE FAIR WITH DATA TOWARDS A DATA INTEROPERABILITY PLATFORM

Luiz Olavo Bonino
Dutch Techcentre for Life Science / Elixir NL
Elixir SME - Wageningen - March 19th 2015
FAIR PLATFORM
WHAT IS FAIR DATA?

The FAIR Data initiative aims to support existing communities in their attempts to enable valuable scientific data and knowledge to be published and utilized in a ‘FAIR’ manner.

- **Findable** - (meta)data is uniquely and persistently identifiable. Should have basic machine readable descriptive metadata.
- **Accessible** - data is reachable and accessible by humans and machines using standard formats and protocols.
- **Interoperable** - (meta)data is machine readable and annotated with resolvable vocabularies/ontologies.
- **Reusable** - (meta)data is sufficiently well-described to allow (semi)automated integration with other compatible data sources.
## FAIR ECOSYSTEM LEVELS

<table>
<thead>
<tr>
<th>FAIR Data Principles</th>
<th>NORMATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAIR Data Protocol</td>
<td></td>
</tr>
<tr>
<td>FAIR Data Resource</td>
<td>ARTEFACT</td>
</tr>
<tr>
<td>FAIR Data Core Technologies</td>
<td></td>
</tr>
<tr>
<td>FAIR Data Systems/Tools</td>
<td>SOFTWARE</td>
</tr>
</tbody>
</table>
FAIR ECOSYSTEM - NORMATIVE LEVEL

- **FAIR Data Principles** - general principles guiding FAIR data solutions; ([www.force11.org/group/fairgroup/fairprinciples](http://www.force11.org/group/fairgroup/fairprinciples))

- **FAIR Data Protocol** - complying with the FAIR Data Principles, provide guidelines for implementing FAIR data solutions, e.g., standards, APIs, technologies, ...;
FAIR ECOSYSTEM - ARTEFACT LEVEL

- **FAIR Data Resource** - datasets expressed using one of the prescribed standards of the FAIR Data Protocol and with metadata complying with the protocol.

- **Annotation Ontology** - reference conceptual model used to provide semantics to elements of FAIR Data Resources through annotation.
FAIR DATA RESOURCE

Datasets expressed using one of the prescribed standards of the FAIR Data Protocol, with metadata complying with the protocol and license. The original dataset is transformed into a FAIR format and proper metadata and license are added to produce a FAIR Data Resource. The original and the FAIR version can co-exist, each one fulfilling its own purpose.
FAIR ECOSYSTEM - SOFTWARE LEVEL

- **FAIR Data Core Tech** - core FAIR standards, e.g., FAIR Profiles;
- **FAIR Data Software** - *software systems* implementing FAIR data APIs and complying with standards;
- **FAIR Data Point** - a lightweight FAIR Data Software functioning as an end-point for accessing published FAIR Data Resources;
- **Data FAIRport** - a FAIR Data Software platform providing functionality to publish and access FAIR Data Resources (the FAIR data “Amazon/eBay“);
- **FAIRifier** - a FAIR Data Software that transform a non-FAIR dataset into a FAIR Data Resource, i.e., a dataset in a FAIR format together with its related metadata and license.

...
FAIR DATA POINT

A particular class of FAIR Data System that provides access to published datasets. The datasets can be external or internal to the FAIR Data Point. Also, the source data can be a regular (non-FAIR) dataset or a FAIR Data Resource. If the source data is non-FAIR, the FAIR Data Point needs to made the necessary FAIR transformations on the fly.
FAIR DATA POINT

Data Producer

Dataset

FAIR Data Point

Data Consumer

Data Producer

Dataset

Data Producer

Dataset

Data Producer

Dataset

FAIR Data Point

Data Consumer
A particular class of FAIR Data System to provide support for data interoperability;

Supports publication and access to FAIR data.

Fosters an ecosystems of applications and services;

Federated architecture: different FAIRports (and other FAIR Data Systems) are interconnectable;

Supports citations of datasets and data items;

Provides metrics for data usage and citation;
FAIR DATA PUBLICATION

Dataset

Metadata

Data Owners/Creators

DataFAIRport
Find, Access, Interoperate & Re-use Data
FAIR DATA ACCESS
DISTRIBUTED ARCHITECTURE

Organization X
DataFAIRport

Plant FAIR Data Point

Open PHACTS DataFAIRport

RD FAIR Data Point

VLPB/WUR DataFAIRport

Organization Y
DataFAIRport

DTL DataFAIRport
Fairport
FAIRIFIER

Data Owner

License

Metadata

input

input

output

Non-FAIR Dataset

FAIRifier

FAIR Data Resource

Metadata

License

FAIR Format
ROADMAP 2015

• FAIR Profiles first draft - March/April 2015;
• Data Model Registry Beta 1 - Q2/Q3 2015;
• FAIRifier Beta 1 - Q3 2015;
• FAIR Data Point Beta 1 - Q4 2015;
HOW TO REALISE

DTL
National FAIR data engineering team

Elixir

Core FAIR Technology

FAIR Data V.T.

BBMRI 2.0

FAIR Data V.T.

ENPADASI

FAIR Data V.T.

ODEX 4ALL

FAIR Data V.T.

Local “DTLs”/projects
data

FAIR Data IG
(P.I.s with data)

Data SAC

Data Core Team

FAIR Data Executive Team
Questions?

Luiz Olavo Bonino - luiz.bonino@dtls.nl
FAIRIFIER

Data Owner

Non-FAIR Dataset

License

Metadata

FAIR Data Transformation

FAIR Data Resource

Metadata

License

FAIR Format
FAIRIFIER

Essential
- Metadata
- License

Desirable
- ARTA Service
- Ontology Service
- Identifier Service

Non-FAIR Dataset

FAIRifier
FAIRIFIER

Non-FAIR Dataset → FAIRifier → FAIR Data Resource (Metadata, License) → URL → WWW

Non-FAIR Dataset → FAIRifier → FAIR Data Resource (Metadata, License) → accessible through FAIR Data Point
FAIRIFIER

Non-FAIR Dataset → FAIRifier → FAIR Data Resource
- Metadata
- License
- FAIR Format

Published by DataFAIRport: Find, Access, Interoperate & Re-use Data

Non-FAIR Dataset → FAIRifier → FAIR Data Resource
- Metadata
- License
- FAIR Format

Published by DataFAIRport: Find, Access, Interoperate & Re-use Data
Find & Access & Interoperate & Reuse Data

FAIR Data Protocol

Data FAIRport

VLPB/WUR Data FAIRport

Private Data

Public Data

VLPB Data

Private Data

Shareable Data

Private Data

Shareable Data

Private Data

Private Data

Shareable Data

Private Data

Data FAIRport

Data FAIRport

Data FAIRport

Data FAIRport

Data FAIRport

Data FAIRport

NCB naturalis Data FAIRport

Meijer Data FAIRport

Averis Seeds Data FAIRport

HZPC Data FAIRport

FAIR Data Point

VMP Tomatoes Genomics

Data FAIRport