DTL

DUTCH TECHCENTRE FOR LIFE SCIENCES

FAIR DATA

FAIR DATA OVERVIEW

Luiz Olavo Bonino - luiz.bonino@dtls.nl

SUMMARY

- What is FAIR data?
- The FAIR ecosystem
- Plans and how to realise

(FAIR) DATA STEWARDSHIP

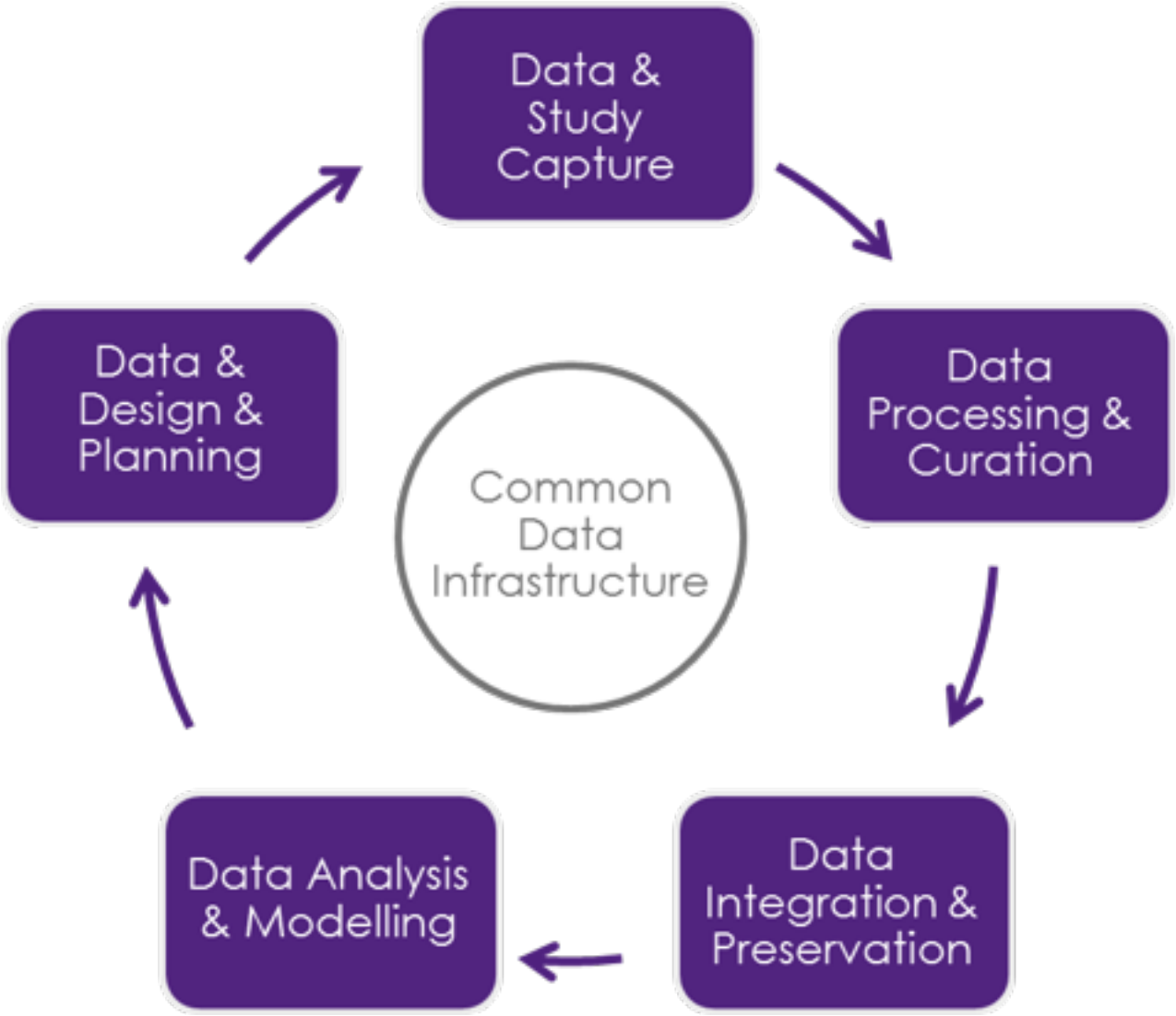
DATA STEWARDSHIP

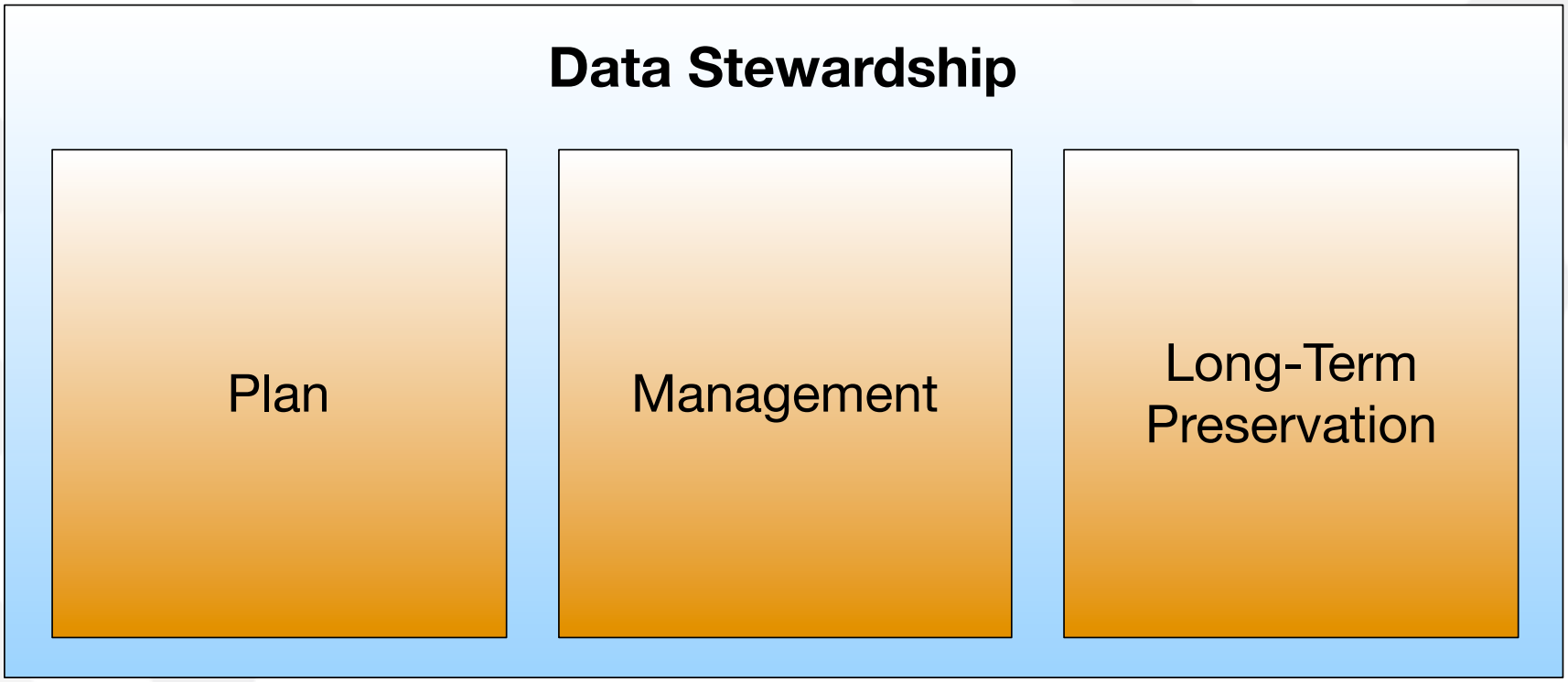
- Combination of all expertise to treat data well in a project:
  - Experiment design and data-design;
  - Re-use of existing data where possible;
  - Planning of the storage, networking and computing infrastructure;
  - Data acquisition and processing;
  - Data publishing in a format that allows functional interlinking of data(sets) as well as in a format suitable for long-term preservation.

FAIR DATA STEWARDSHIP

- **Combination of all expertise to treat data well in a project:**
  - Experiment design **and data-design**;
  - **Re-use** of existing data where possible;
  - Planning of the **storage**, **networking** and **computing** infrastructure;
  - **Data acquisition and processing**;
  - **Data publishing in a format that allows functional interlinking of data(sets)** as well as in a format suitable for long-term preservation.

FAIR Data


Data & Study Capture
Data Processing & Curation
Data Integration & Preservation
Data Analysis & Modelling
Data & Design & Planning
Common Data Infrastructure


Data Stewardship
Plan
Management
Long-Term Preservation

Data Management Plans
Mandatory for Research Projects H2020 & Member States

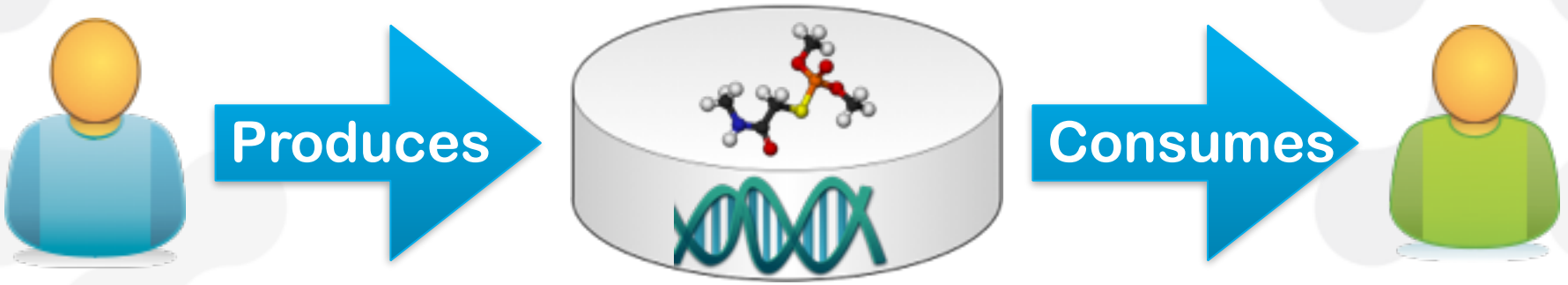
Long-Term Data Stewardship
How to finance ESFRI's and EBI SIB type + infra
Mainly private for reliability

Interoperability Backbones, Standards, Procedures
Mainly H2020 + ESFRI-type domain expertise

DTL

FAIR DATA

DTL



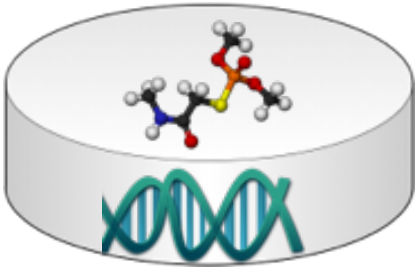
stewardship

sustainability

privacy

security

Produces



Consumes

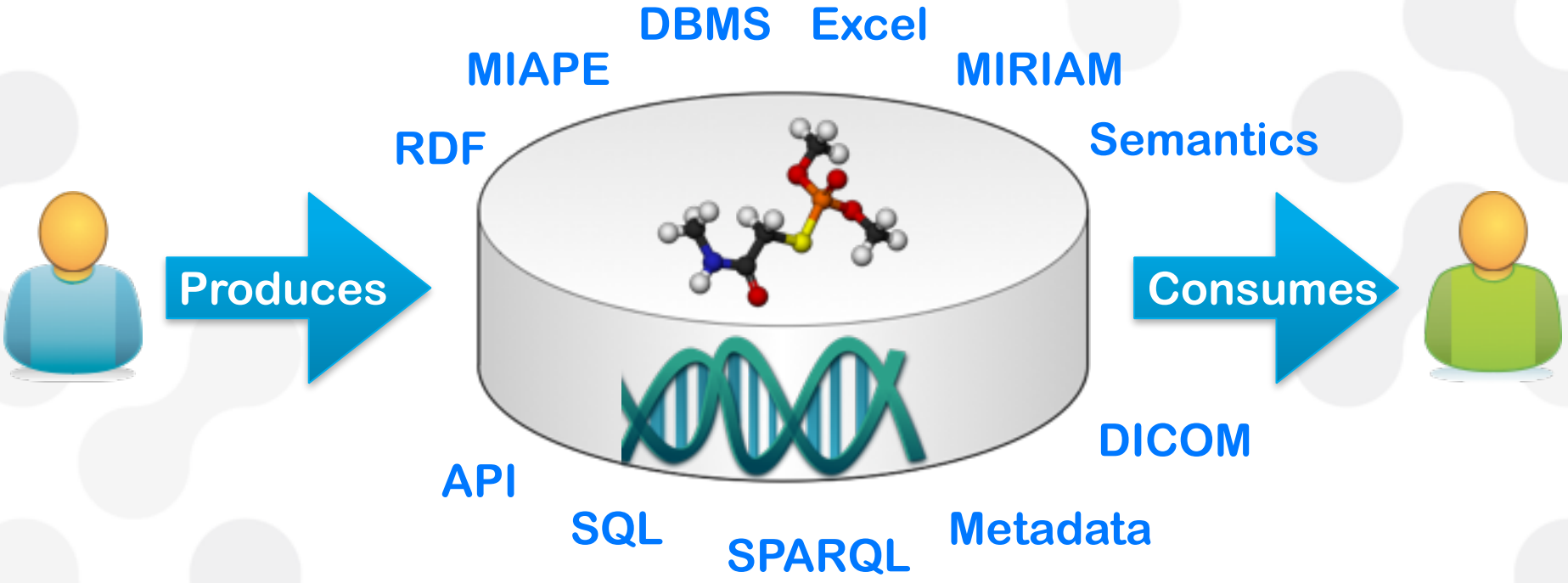
license

access

storage

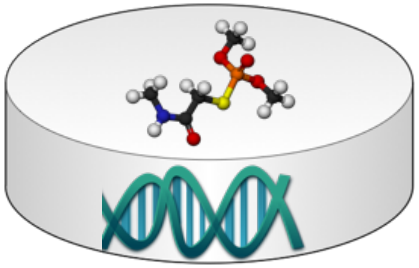
maintenance

DTL


DBMS
Excel
MIAPE
MIRIAM
RDF
Semantics
Produces
Consumes
DICOM
API
SQL
SPARQL
Metadata

DTL

Produces



Consumes

find

access

format

license

query

integrate

WHAT IS FAIR DATA?

FAIR Data aims to support existing communities in their attempts to enable valuable scientific data and knowledge to be published and utilised 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.

THE FAIR ECOSYSTEM

FAIR Data Principles **FAIR Data Protocol**	Normative
FAIR Data Resources	Artefact
FAIR Data Core Technologies **FAIR Data Systems/Tools**	Software

FAIR ECOSYSTEM - NORMATIVE LEVEL

- **FAIR Data Principles** - general principles guiding FAIR data solutions;

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

FAIR DATA PROTOCOL

Findable - standards for describing the dataset with the relevant metadata;

Accessible - standards for represent and access the data according to the defined usage license;

Interoperable - standards for machine readable descriptions of the (meta)data and (semantic)annotation;

Reusable - standards for semantic annotation of the (meta)data supporting machine reasoning, and standards for defining data provenance and support citation;

The standards include technologies (e.g., RDF, nano pub, JSON, OWL, etc.) as well as protocols and APIs.

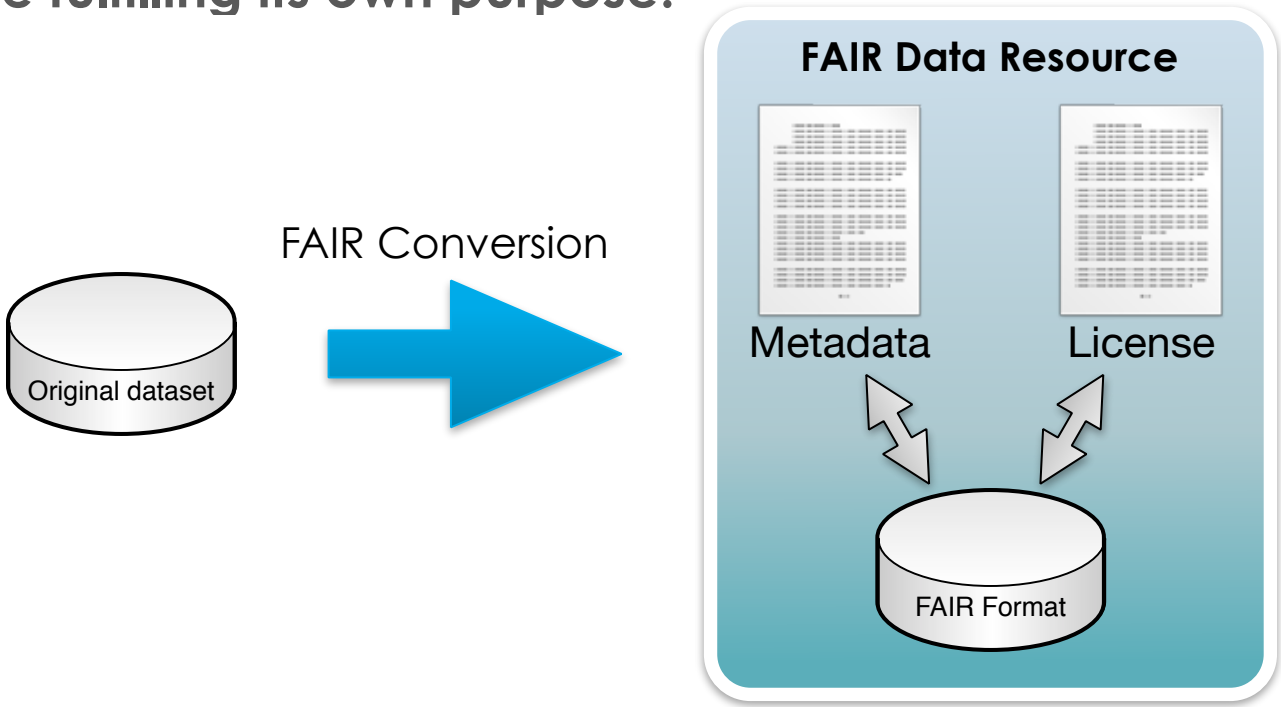
FAIR ECOSYSTEM - ARTEFACT LEVEL

- **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.
- **Controlled vocabularies, dictionaries, etc.**

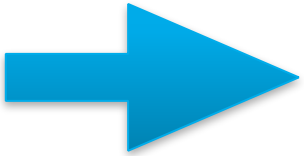
FAIR DATA RESOURCE

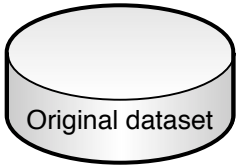
Datasets expressed using one of the prescribed standards of the FAIR Data Protocol, with metadata complying with the protocol and license. The original dataset is transformed into a FAIR format and proper metadata and license are added to produce a FAIR Data Resource. The original and the FAIR version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

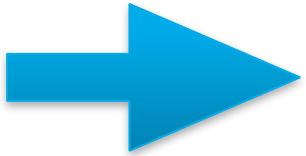


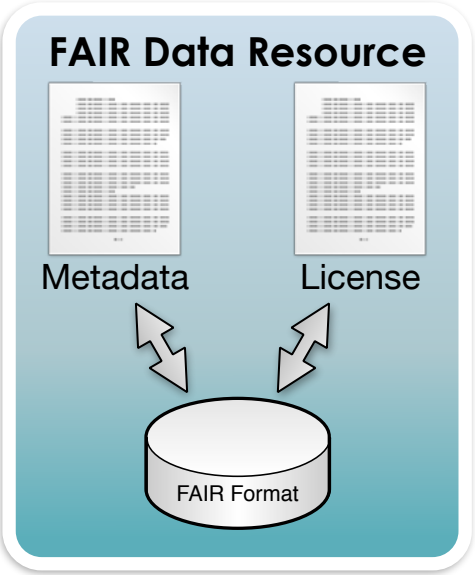
FAIR DATA RESOURCE

Data Creation

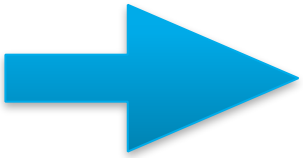


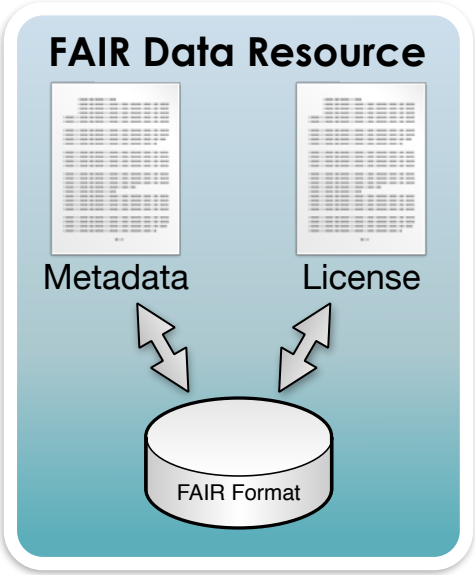






FAIR Data Creation



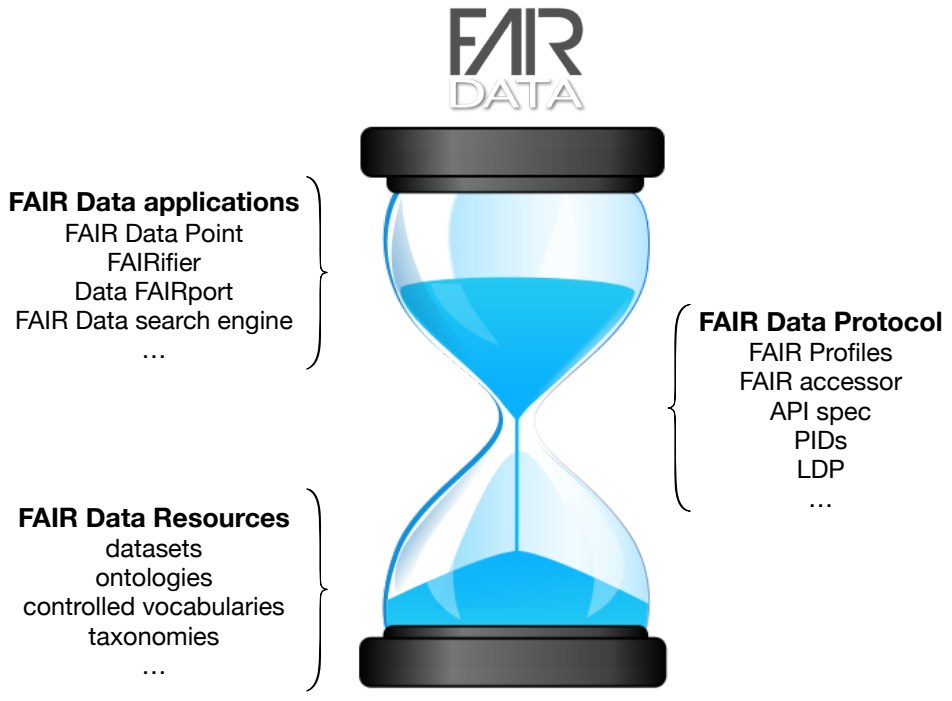


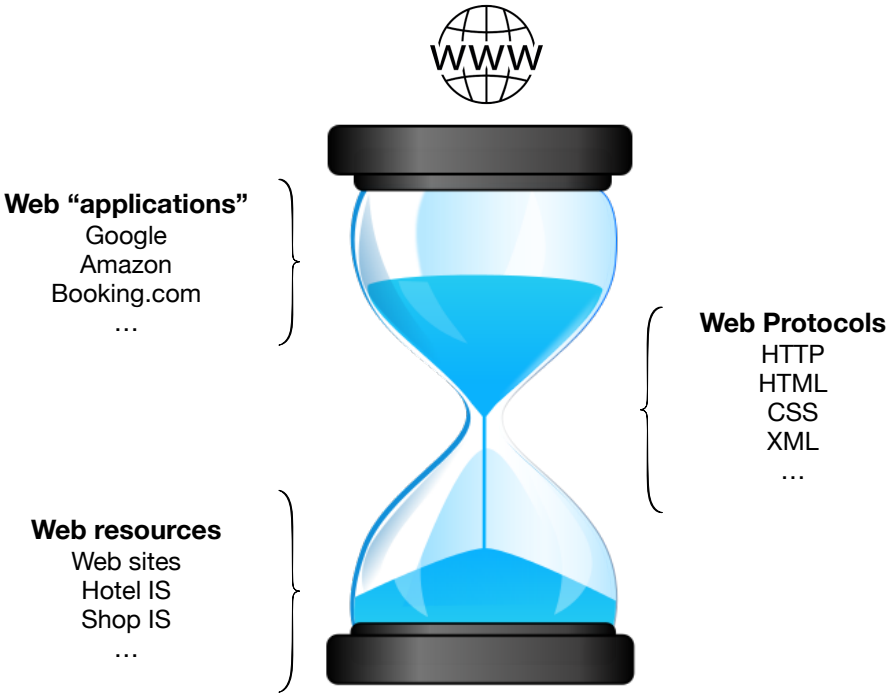
FAIR DATA

FAIR Data applications
FAIR Data Point
FAIRifier
Data FAIRport
FAIR Data search engine
…

FAIR Data Resources
datasets
ontologies
controlled vocabularies
taxonomies
…

FAIR Data Protocol
FAIR Profiles
FAIR accessor
API spec
PIDs
LDP
…


FAIR DATA
FAIR Data applications
FAIR Data Point
FAIRifier
Data FAIRport
FAIR Data search engine
…
FAIR Data Resources
datasets
ontologies
controlled vocabularies
taxonomies
…
FAIR Data Protocol
FAIR Profiles
FAIR accessor
API spec
PIDs
LDP
…


WWW
Web “applications”
Google
Amazon
Booking.com
…
Web resources
Web sites
Hotel IS
Shop IS
…
Web Protocols
HTTP
HTML
CSS
XML
…

SHARING DATA

I would like to exploit common genotype-phenotype relations between Alzheimer's Disease and Huntington's Disease…
I need to combine AD and HD data…

I can help with that!

Source: Marcos Roos

SHARING DATA

???

DOES NOT COMPUTE

Here's my data, have fun!

米当局は、あなたの国籍故に、在米日本領事代表にあなたが
逮捕又は拘禁されたことを通報する必要があります。領事官
は通報を受けた後、あなたに電話を掛けたり、あるいはあな
たを訪問することができます。
る必要はありませんが、
あなたの家族との連絡、
置を取ってくれるかも知
事官に通報します。

Очи чёрные, очи жгучие,
Очи страстные и прекрасные,
Как люблю я вас, как боюсь я вас,
Знать увидел вас я не в добрый час.

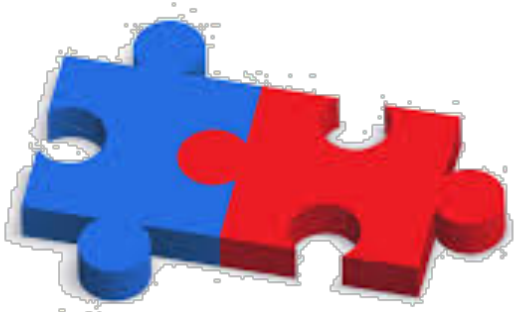
Очи чёрные, очи пламенны
И монят они в страны дальные,
Где царит любовь, где царит покой,
Где страданья нет, где вражды запрет.

Here's my data, have fun!

Source: Marcos Roos

SHARING LINKABLE DATA

I can go straight to answering my questions with data from multiple data owners! Patients will be so pleased with this speed-up!

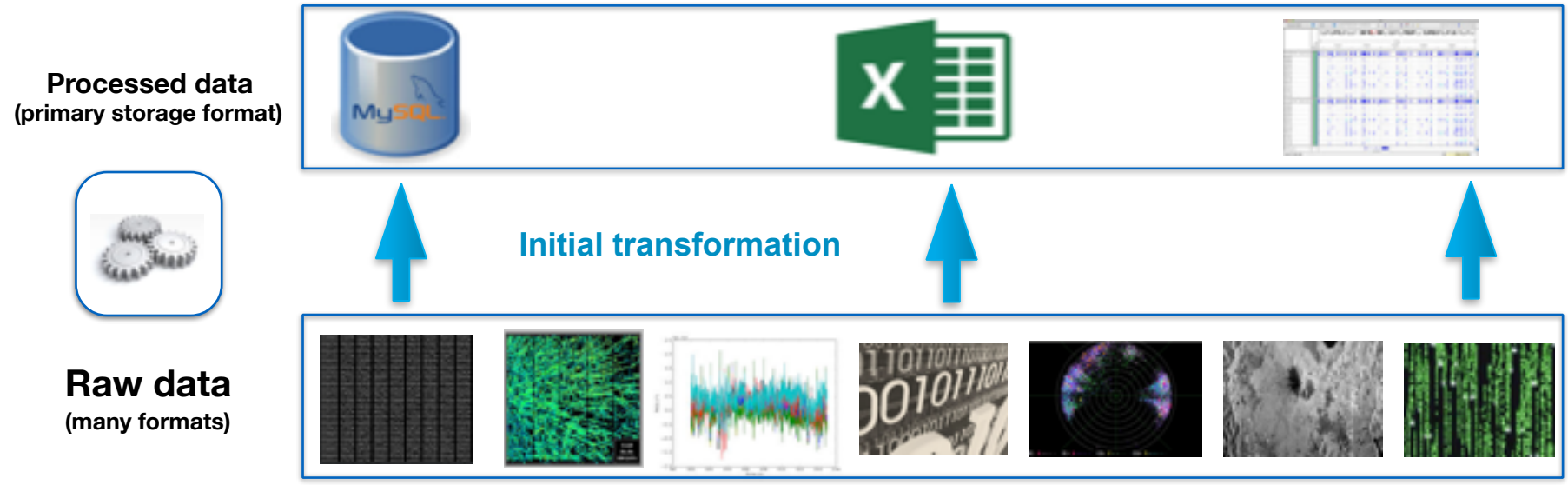


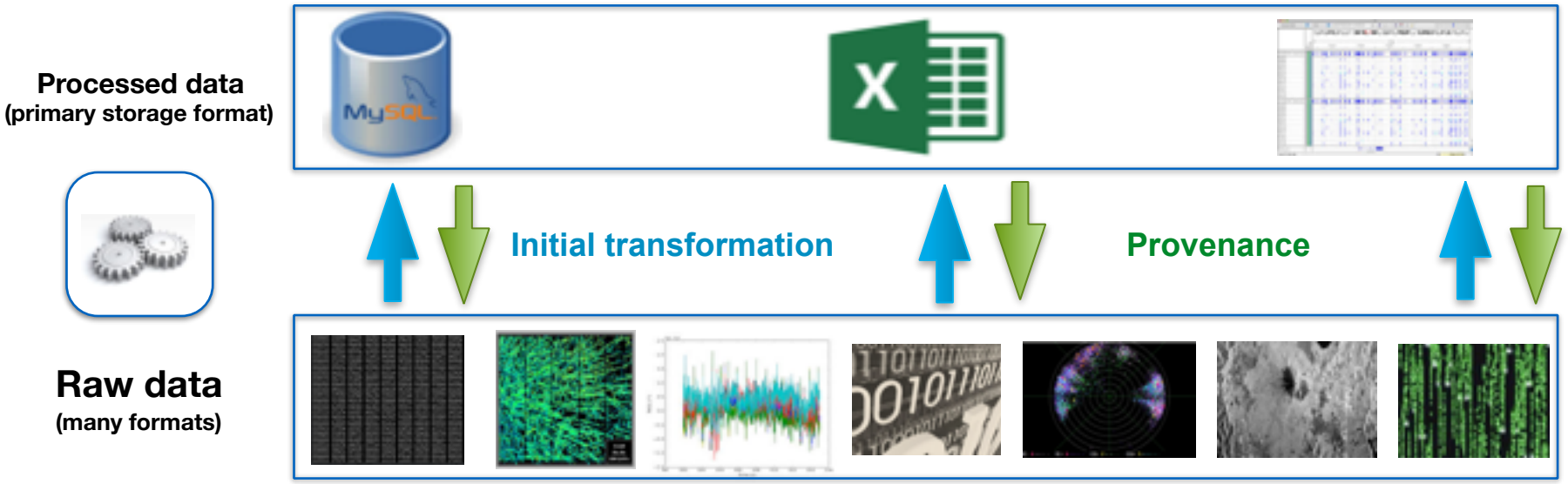
Here's my **Linked** Data, have fun!

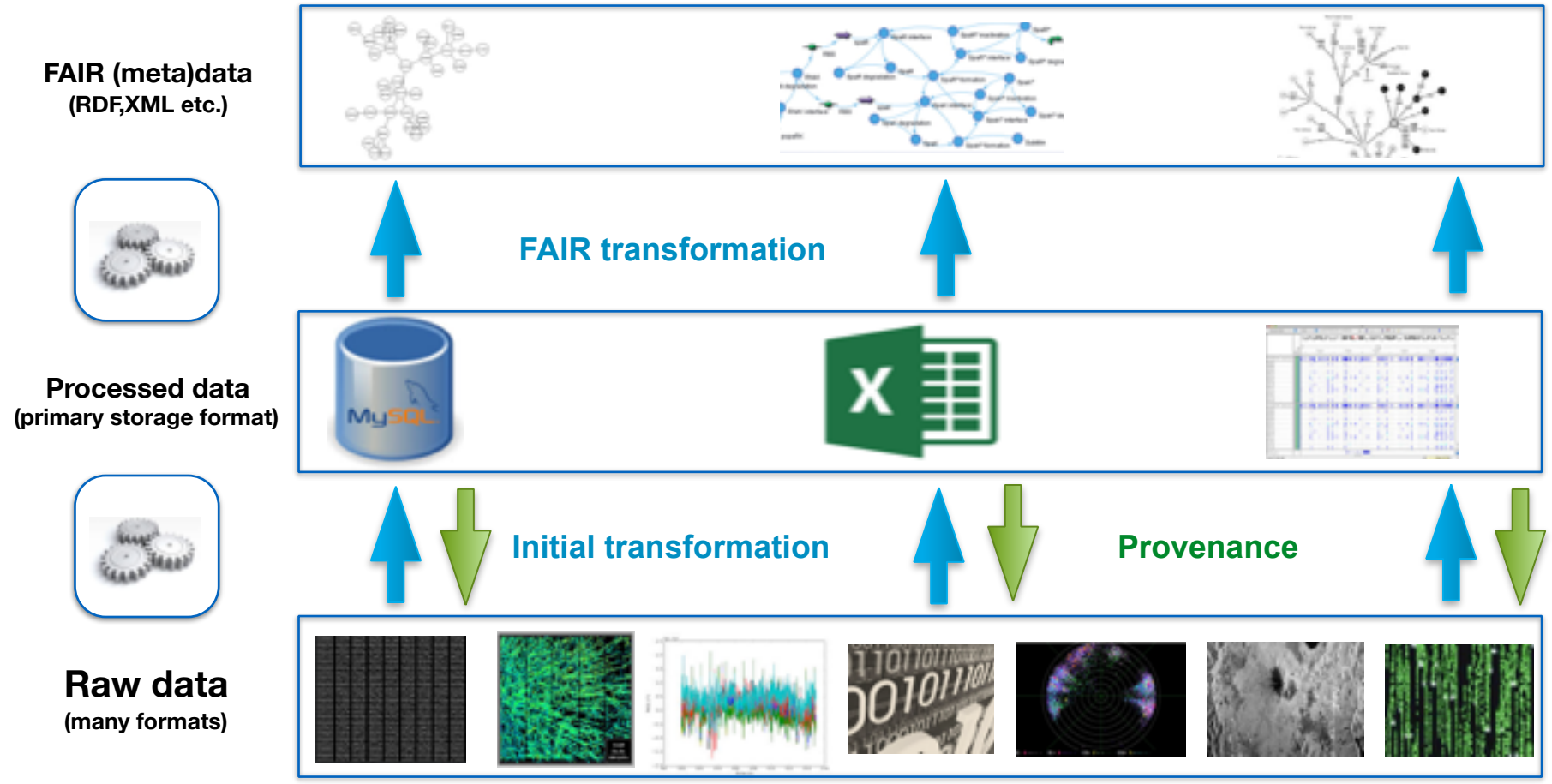
Here's my **Linked** Data, have fun!

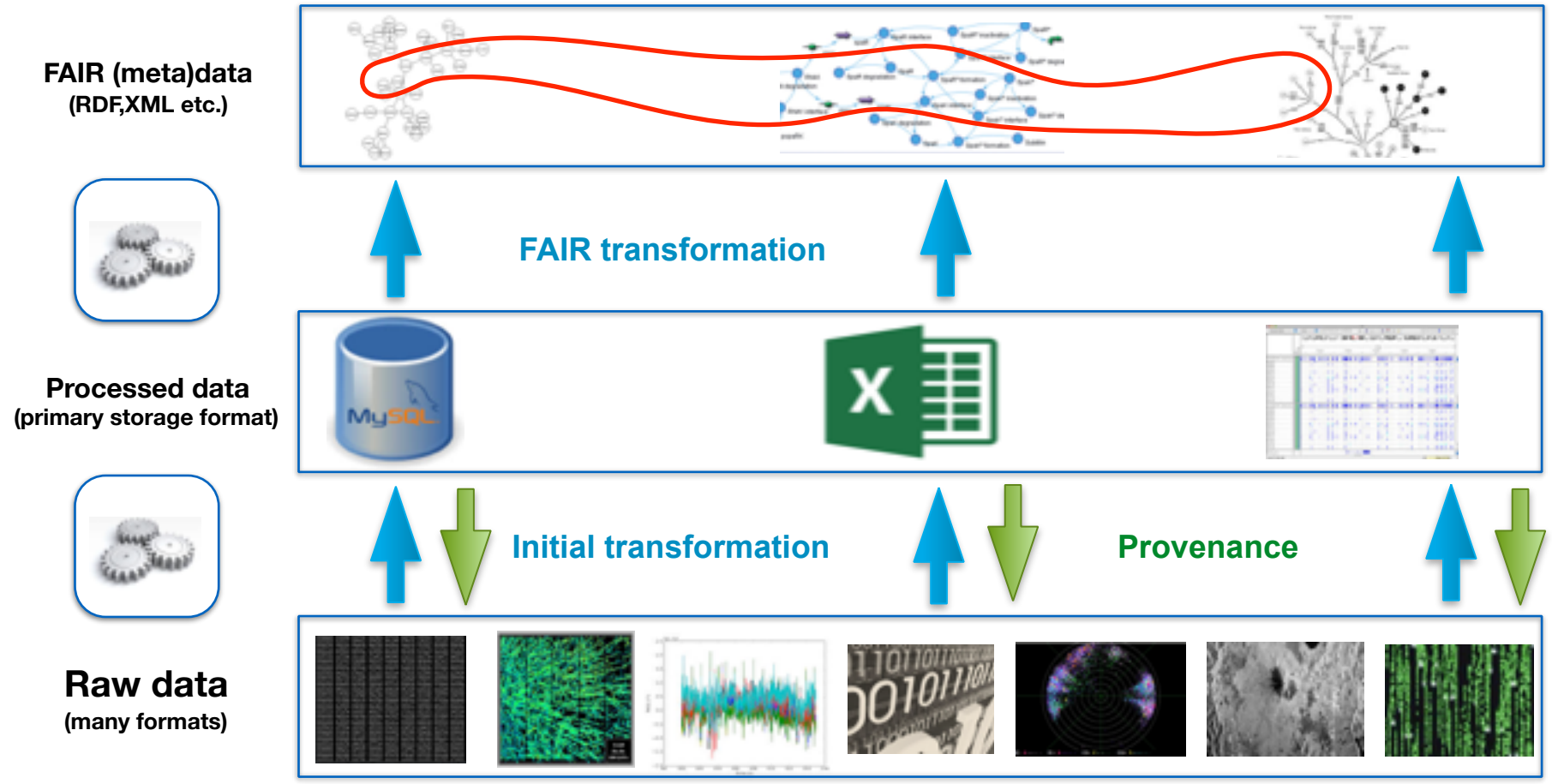
Source: Marcos Roos

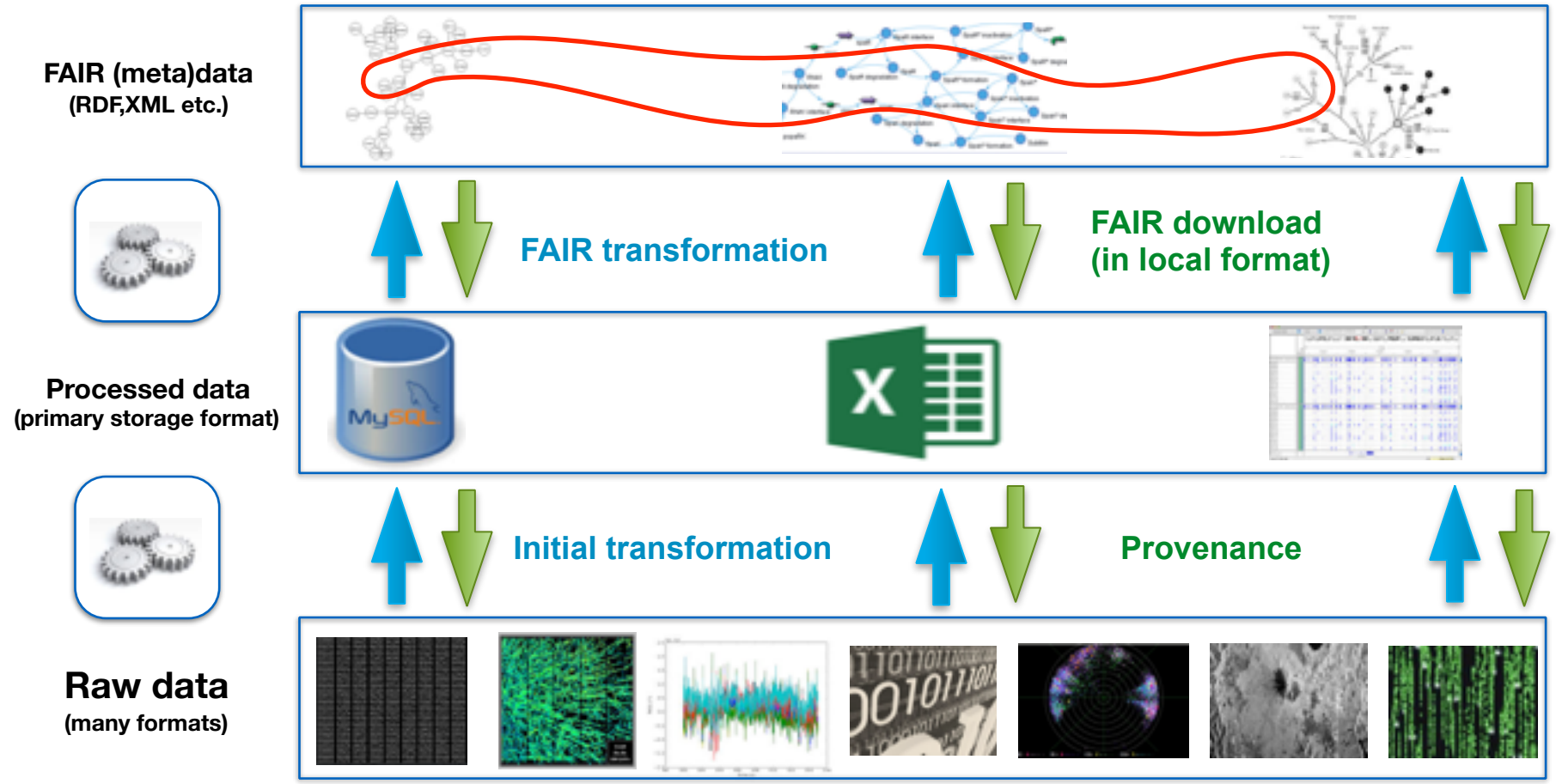
Raw data
(many formats)

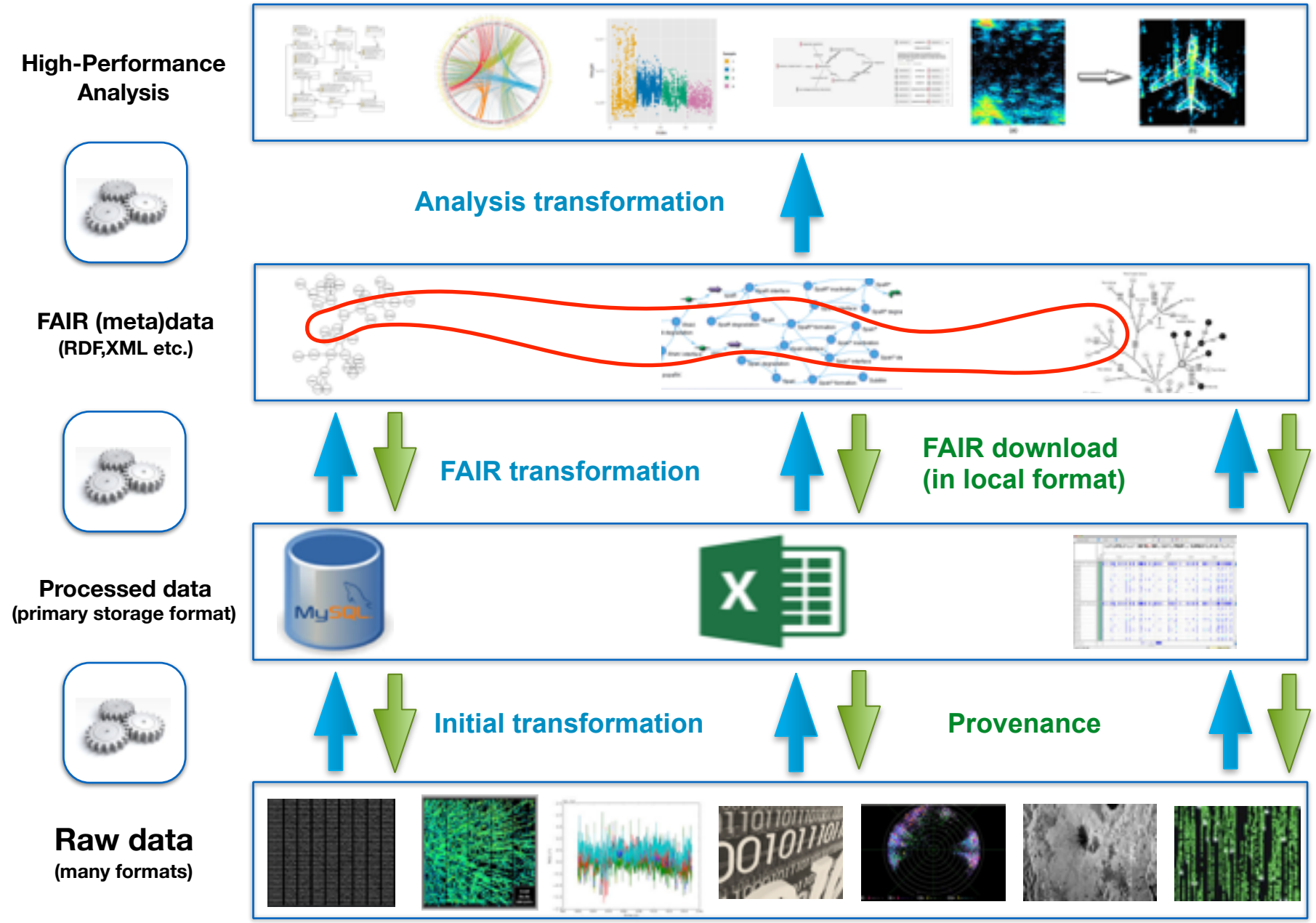

Processed data
(primary storage format)
MySQL
X
Initial transformation
Raw data
(many formats)


Processed data
(primary storage format)
MySQL
X
Initial transformation
Provenance
Raw data
(many formats)


FAIR (meta)data
(RDF,XML etc.)
FAIR transformation
Processed data
(primary storage format)
MySQL
Initial transformation
Provenance
Raw data
(many formats)

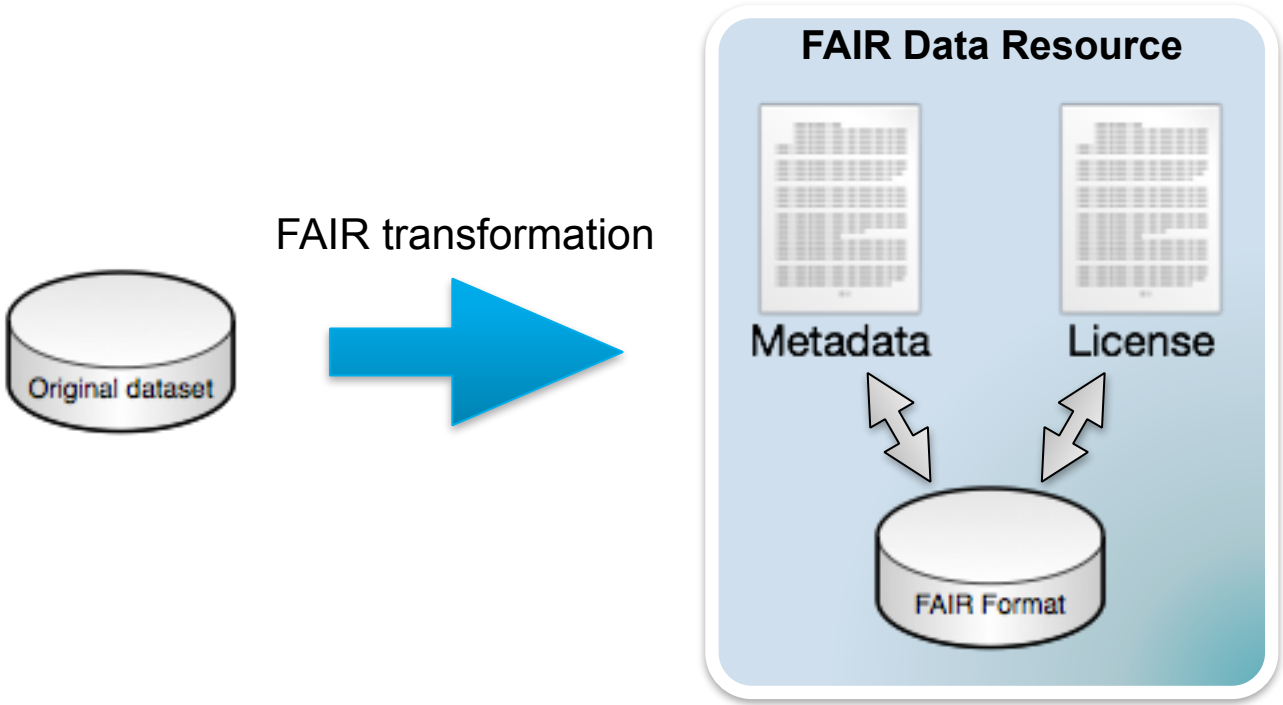

FAIR (meta)data
(RDF,XML etc.)
FAIR transformation
Processed data
(primary storage format)
MySQL
Initial transformation
Provenance
Raw data
(many formats)


FAIR (meta)data
(RDF,XML etc.)
FAIR transformation
FAIR download
(in local format)
Processed data
(primary storage format)
MySQL
X
Initial transformation
Provenance
Raw data
(many formats)


High-Performance Analysis
Analysis transformation
FAIR (meta)data
(RDF,XML etc.)
FAIR transformation
FAIR download
(in local format)
Processed data
(primary storage format)
MySQL
Initial transformation
Provenance
Raw data
(many formats)

FAIR DATA APPLICATION ECOSYSTEM (NL APPROACH)

FAIR DATA RESOURCE



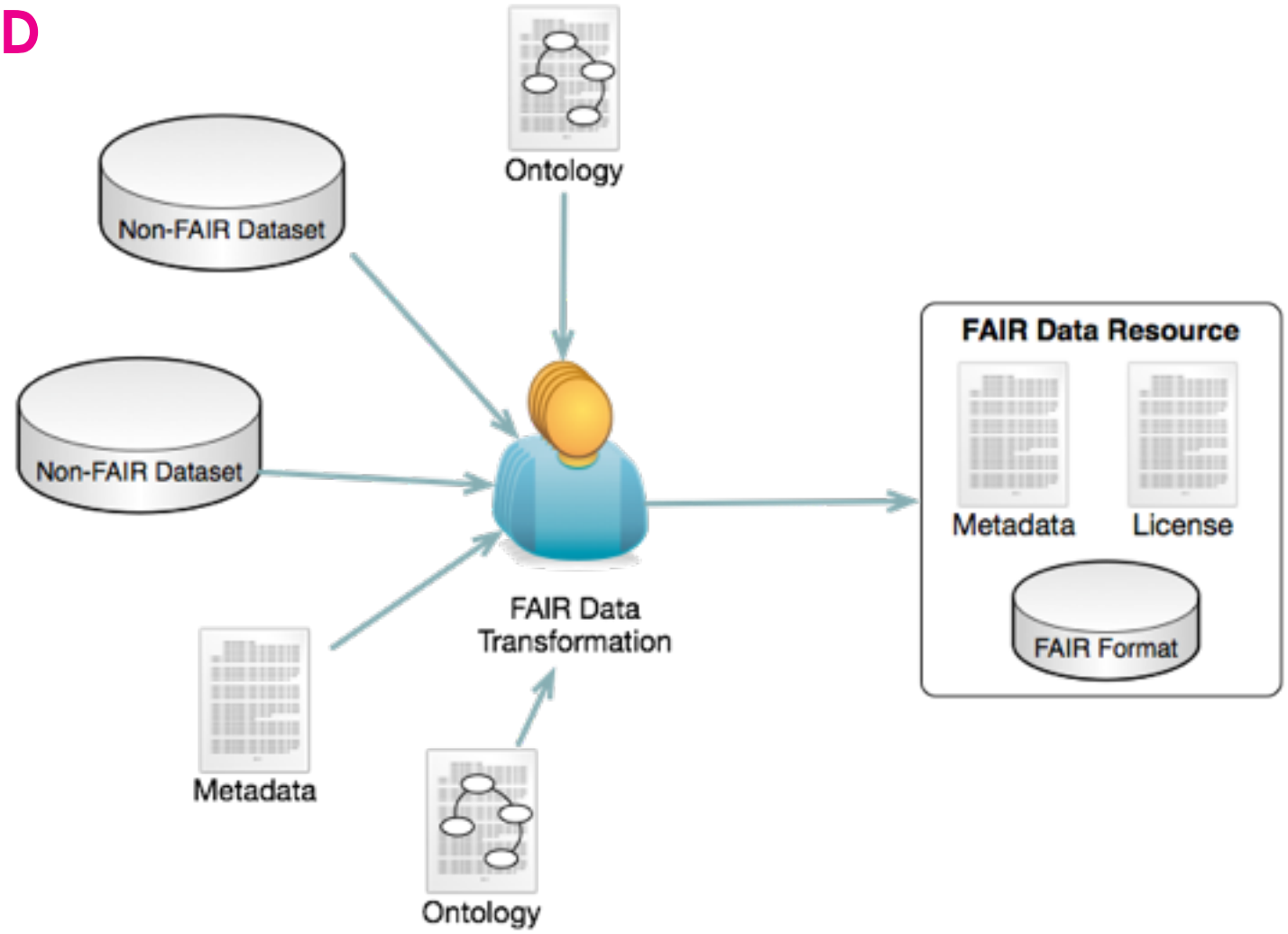
BRING YOUR OWN DATA - BYOD

- Goals:
  - Learn how to make data linkable “hands-on” with experts
  - Create a “telling story” to demonstrate its use

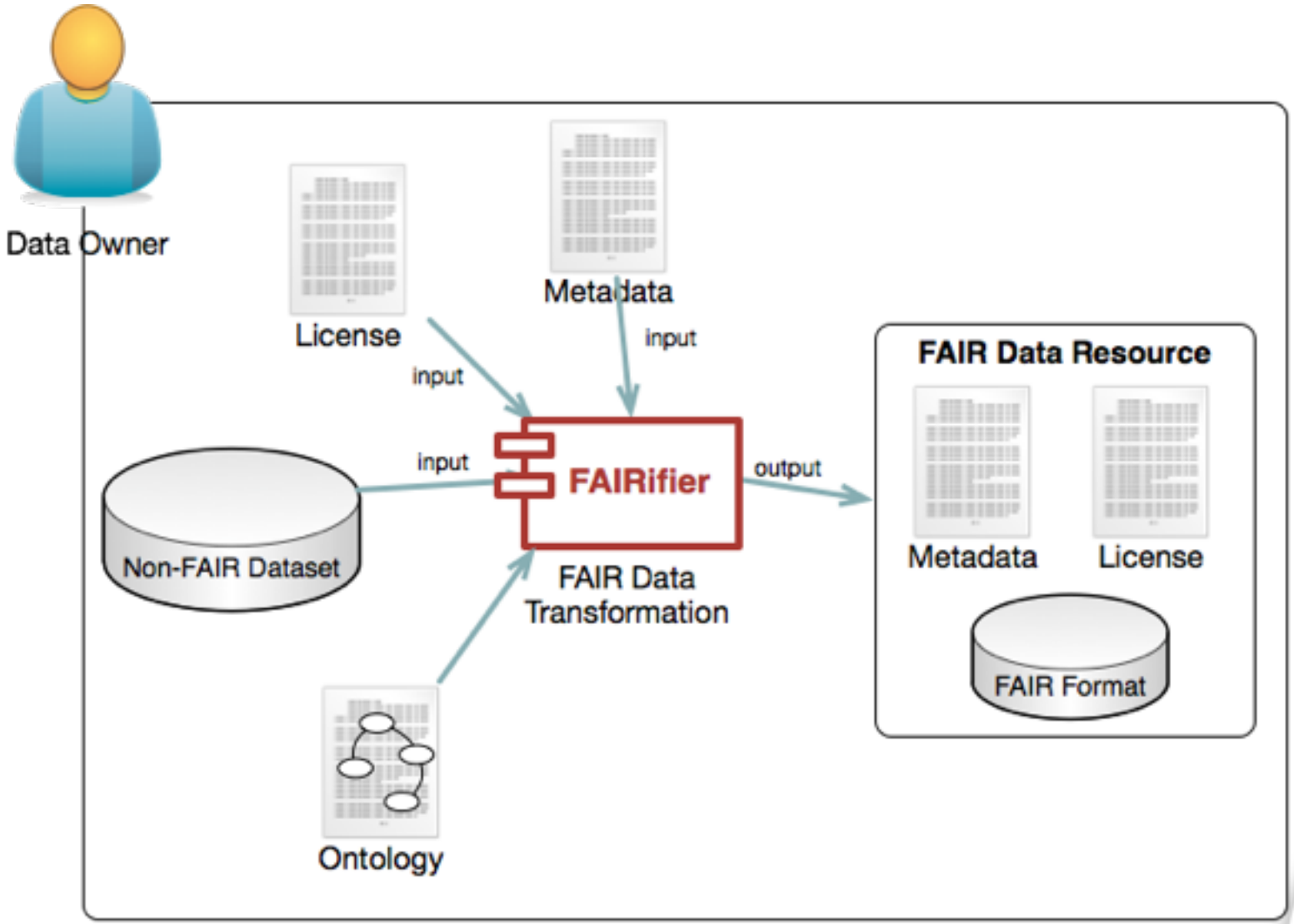
- Composition:
  - Data owners – specialists on given datasets
  - Data interoperability experts
  - Domain experts

Source: Marcos Roos

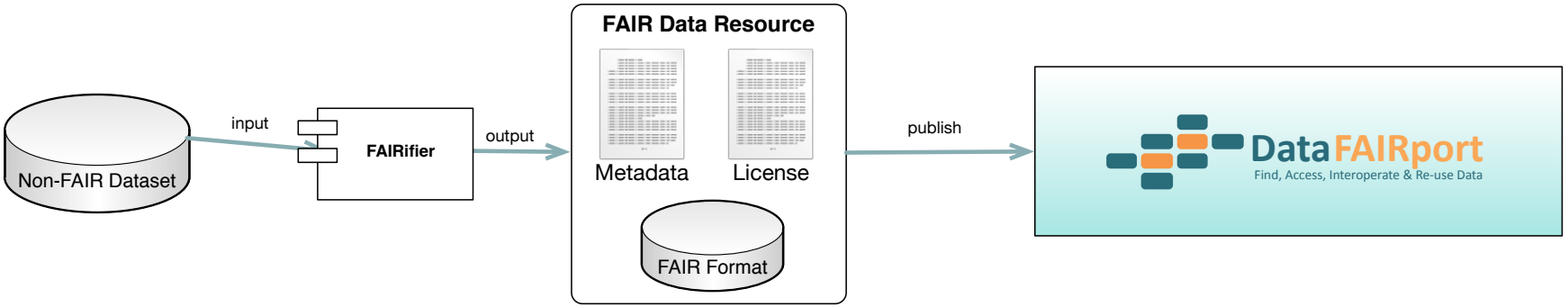
BYOD

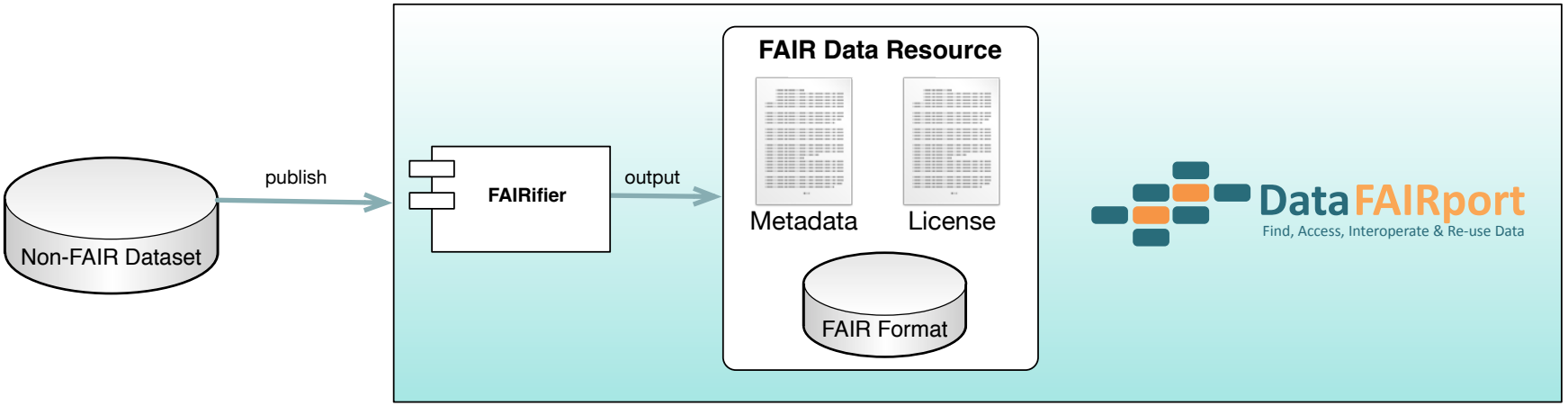

Ontology
Non-FAIR Dataset
Non-FAIR Dataset
FAIR Data Resource
Metadata
License
FAIR Format
FAIR Data
Transformation
Metadata
Ontology

FAIRIFIER

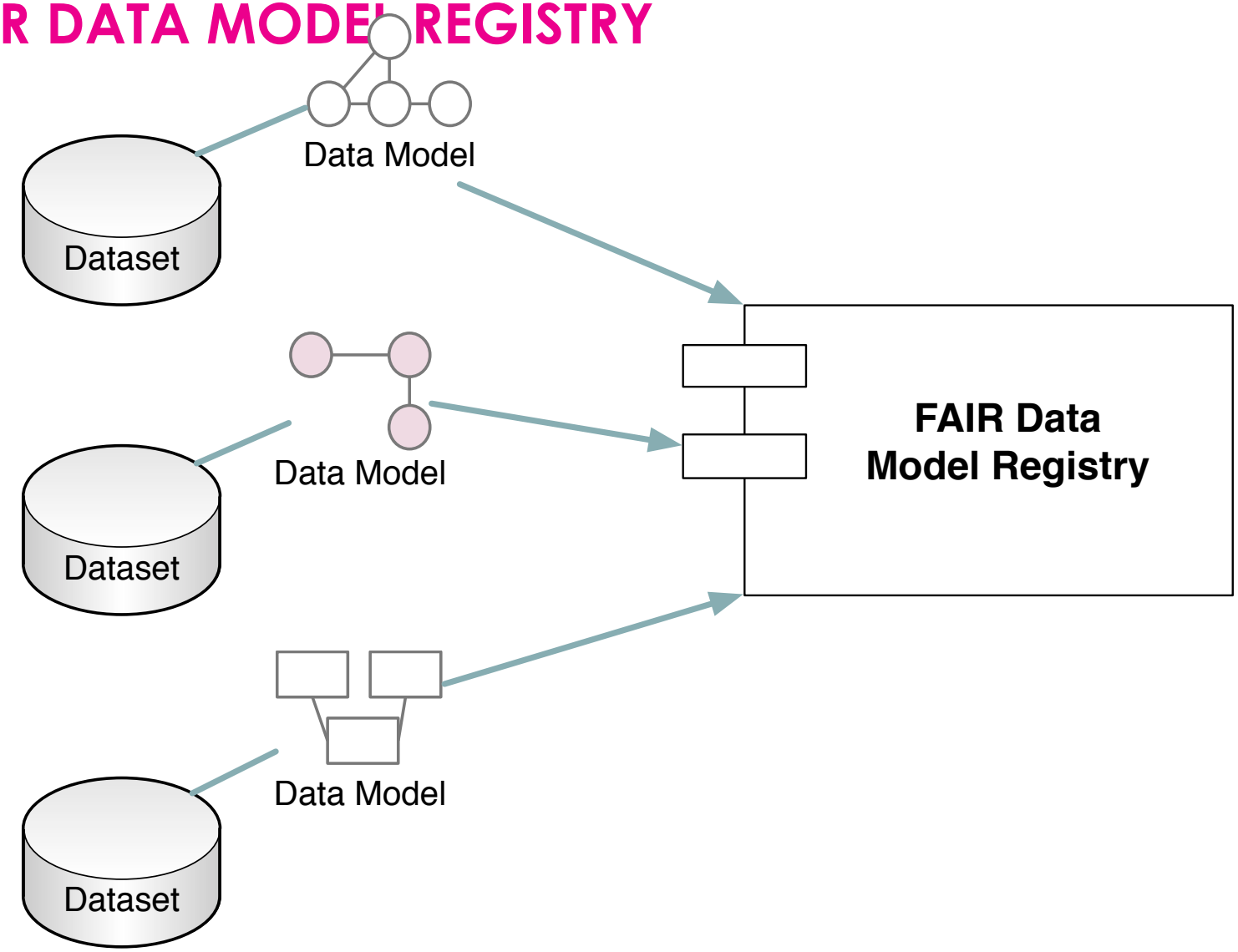

Data Owner
License
Metadata
input
input
input
FAIRifier
output
Non-FAIR Dataset
FAIR Data
Transformation
Ontology
FAIR Data Resource
Metadata
License
FAIR Format

FAIRIFIER

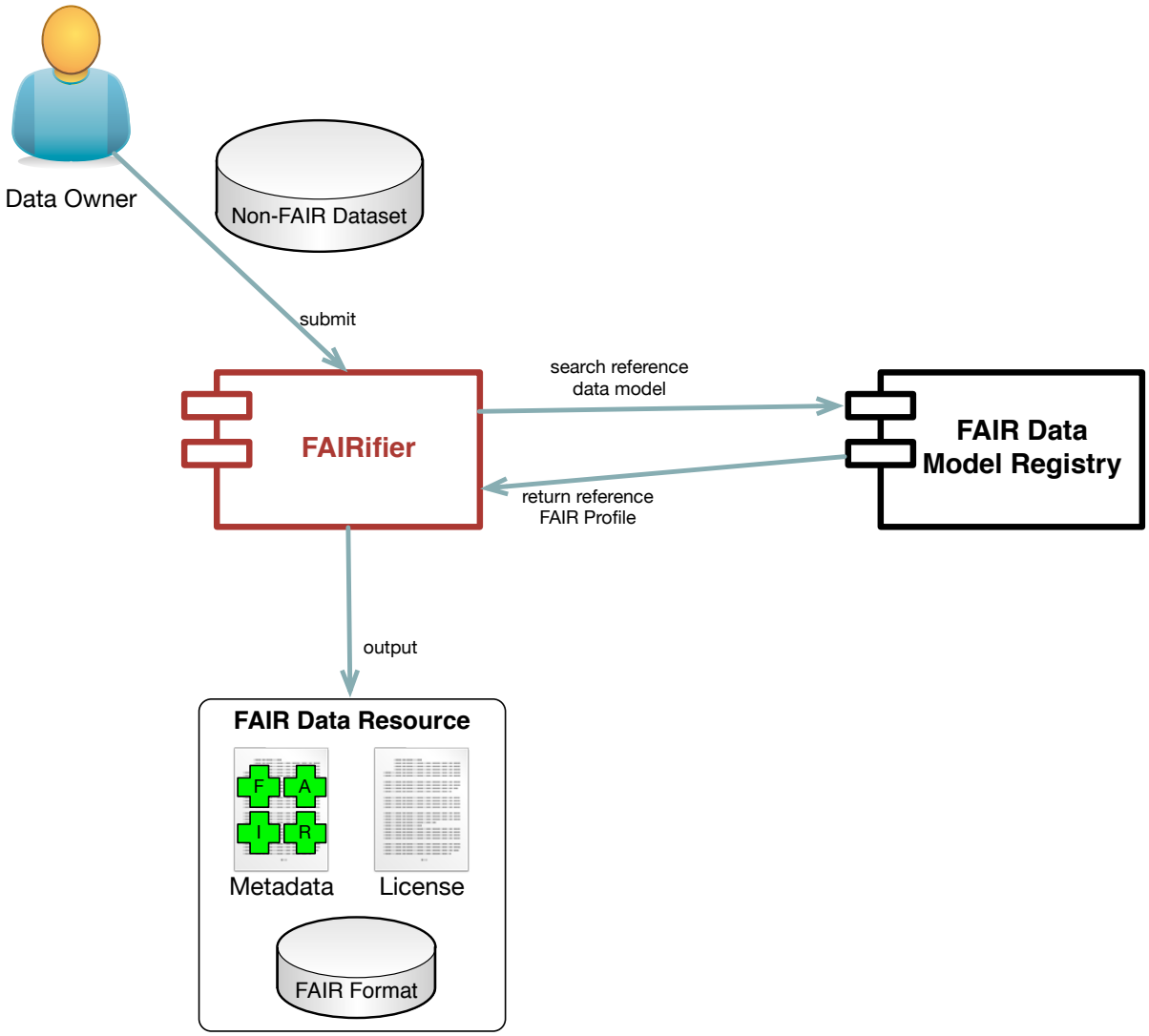

FAIR Data Resource
input
FAIRifier
output
publish
Non-FAIR Dataset
Metadata
License
FAIR Format
DataFAIRport
Find, Access, Interoperate & Re-use Data


FAIR Data Resource
publish
FAIRifier
output
Non-FAIR Dataset
Metadata
License
FAIR Format
DataFAIRport
Find, Access, Interoperate & Re-use Data

FAIR DATA MODEL REGISTRY



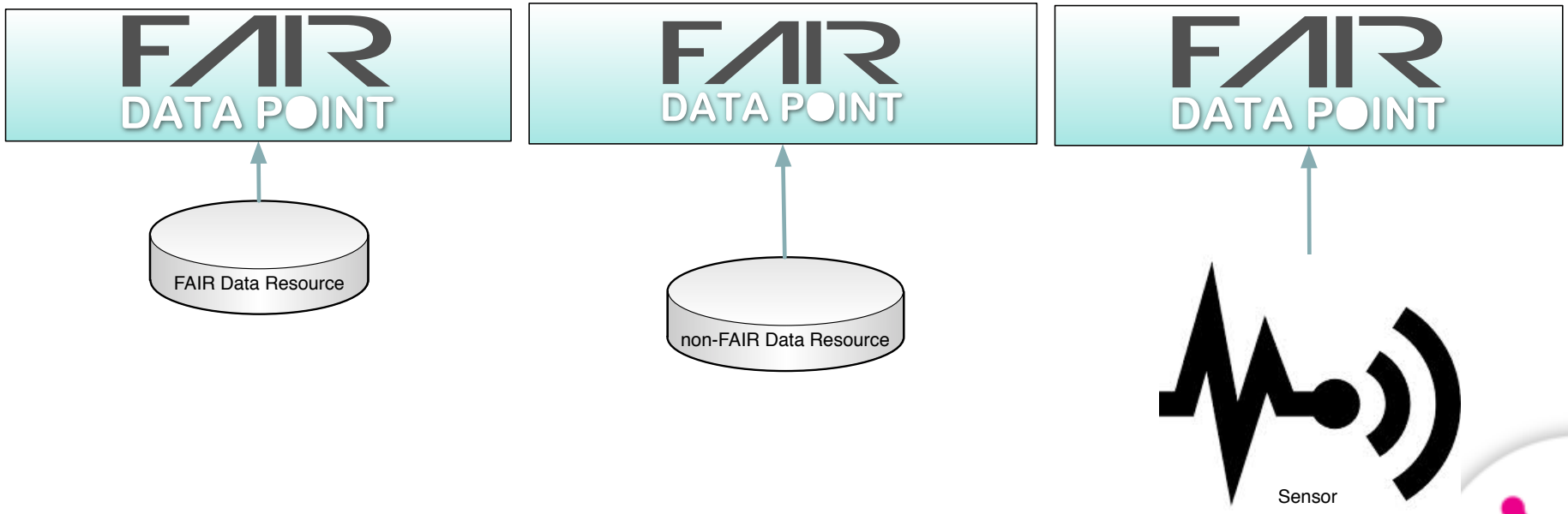
FAIRIFIER AND FAIR DATA MODEL REGISTRY



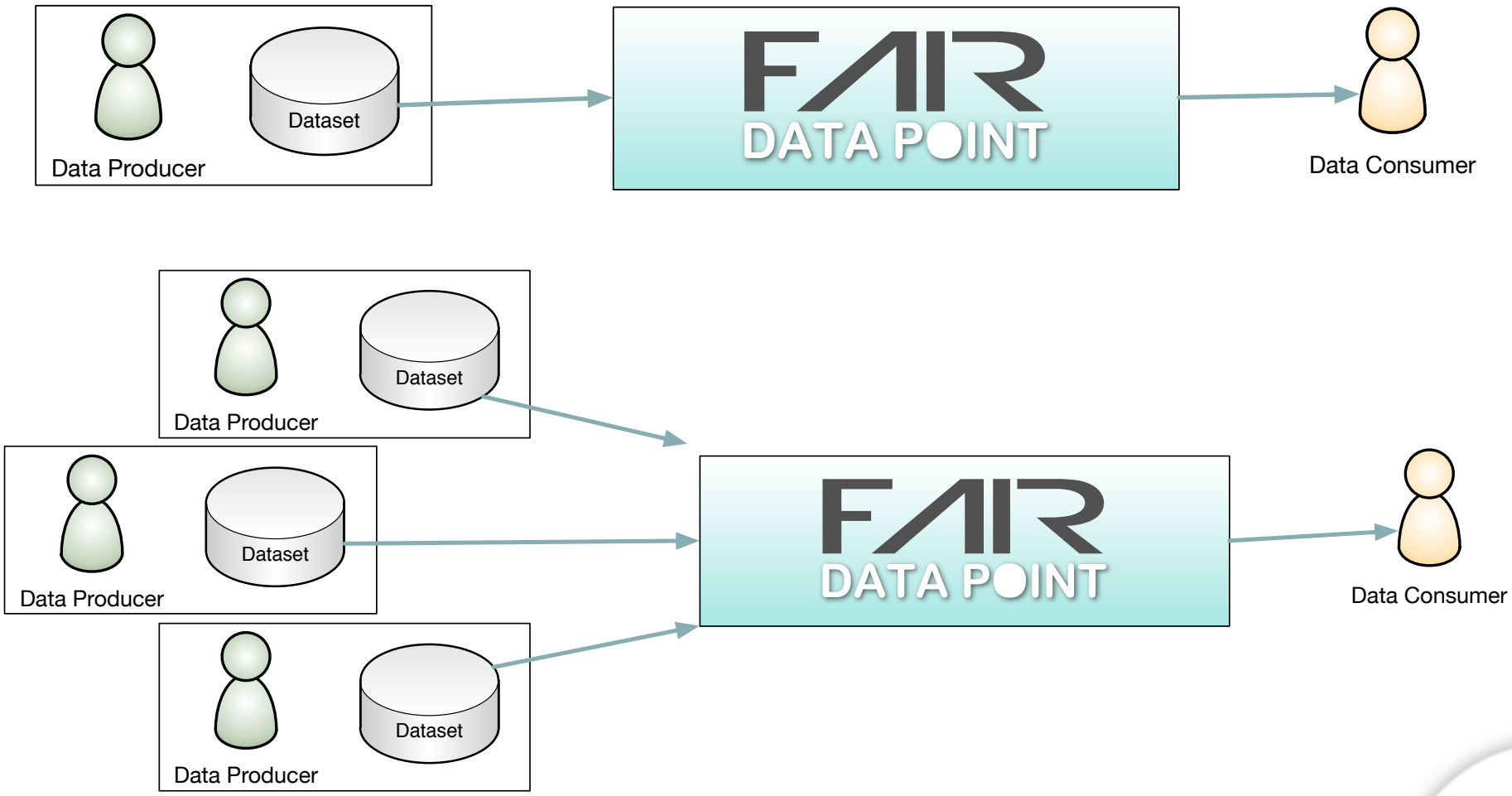
DTL

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



DataFAIRport

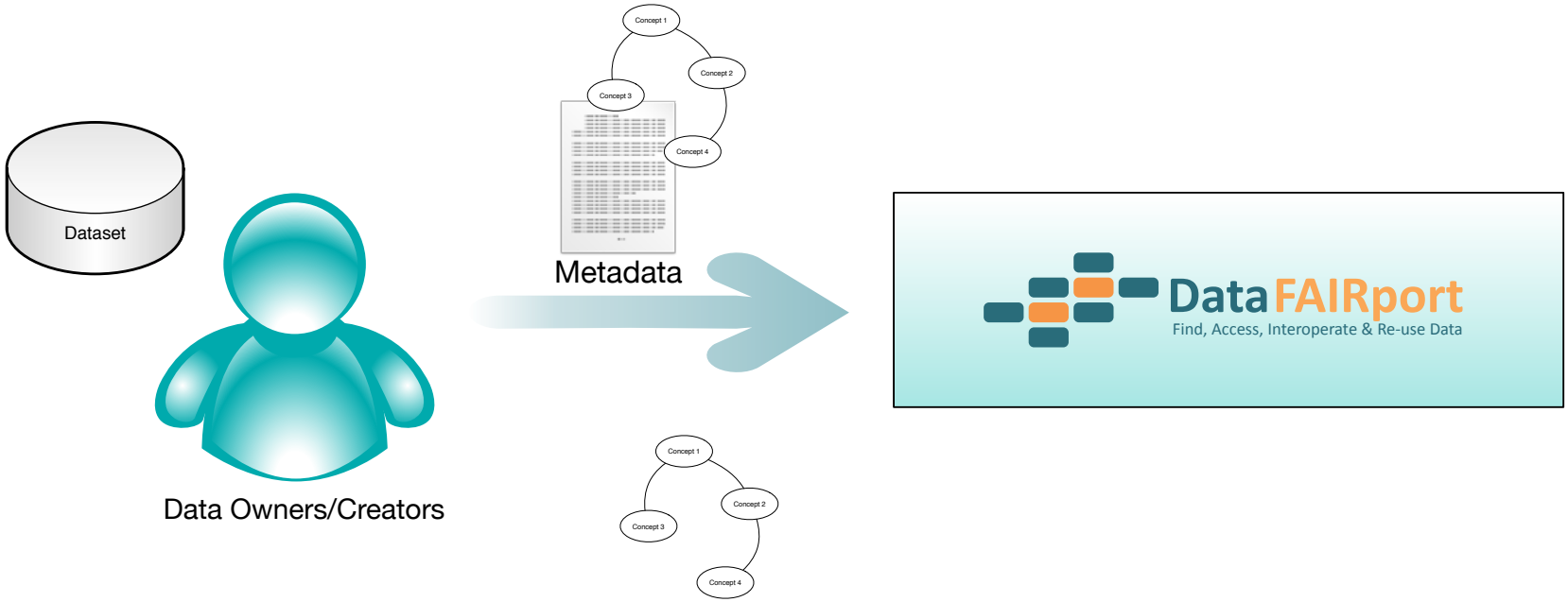
Find, Access, Interoperate & Re-use Data

DTL

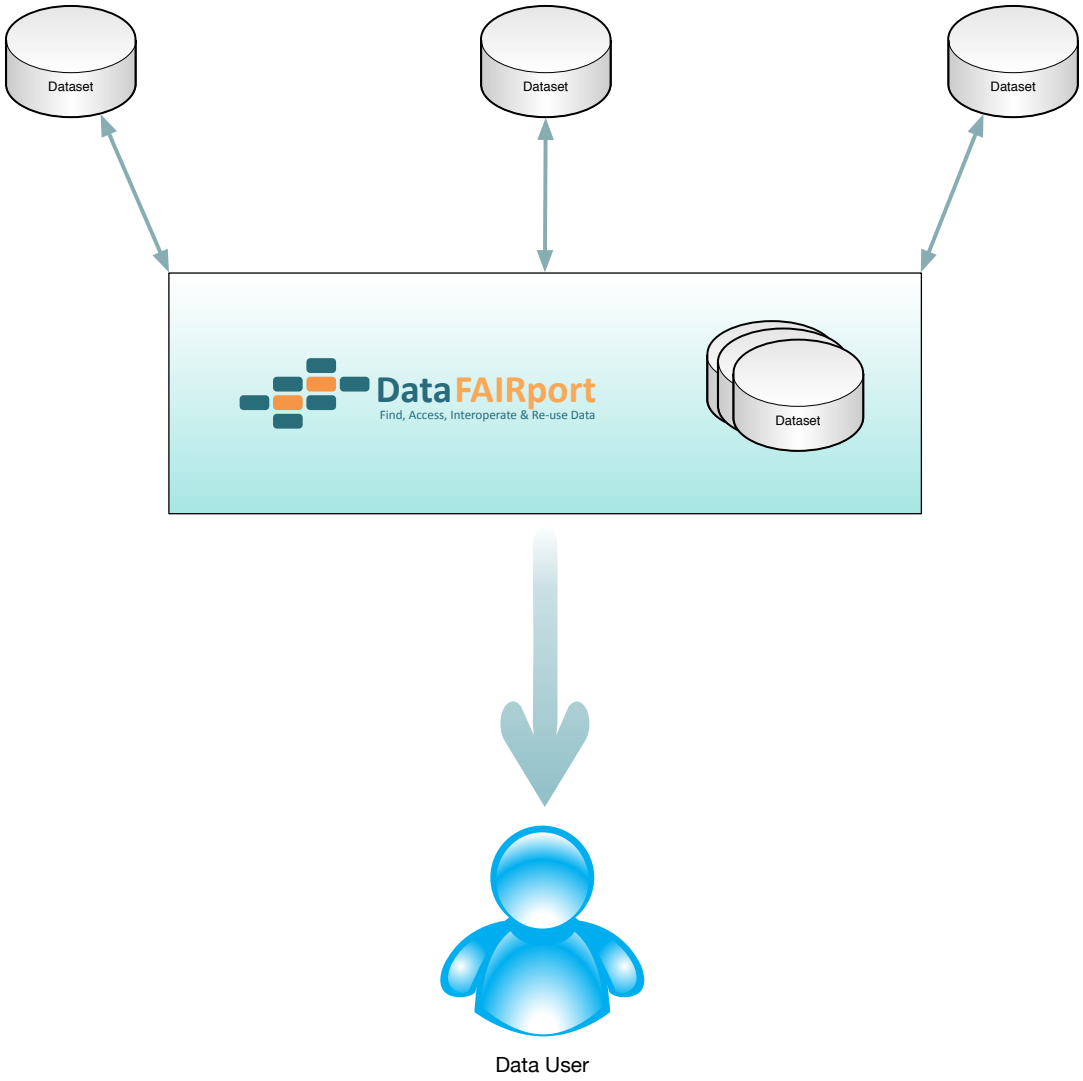
DataFAIRport

- 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;

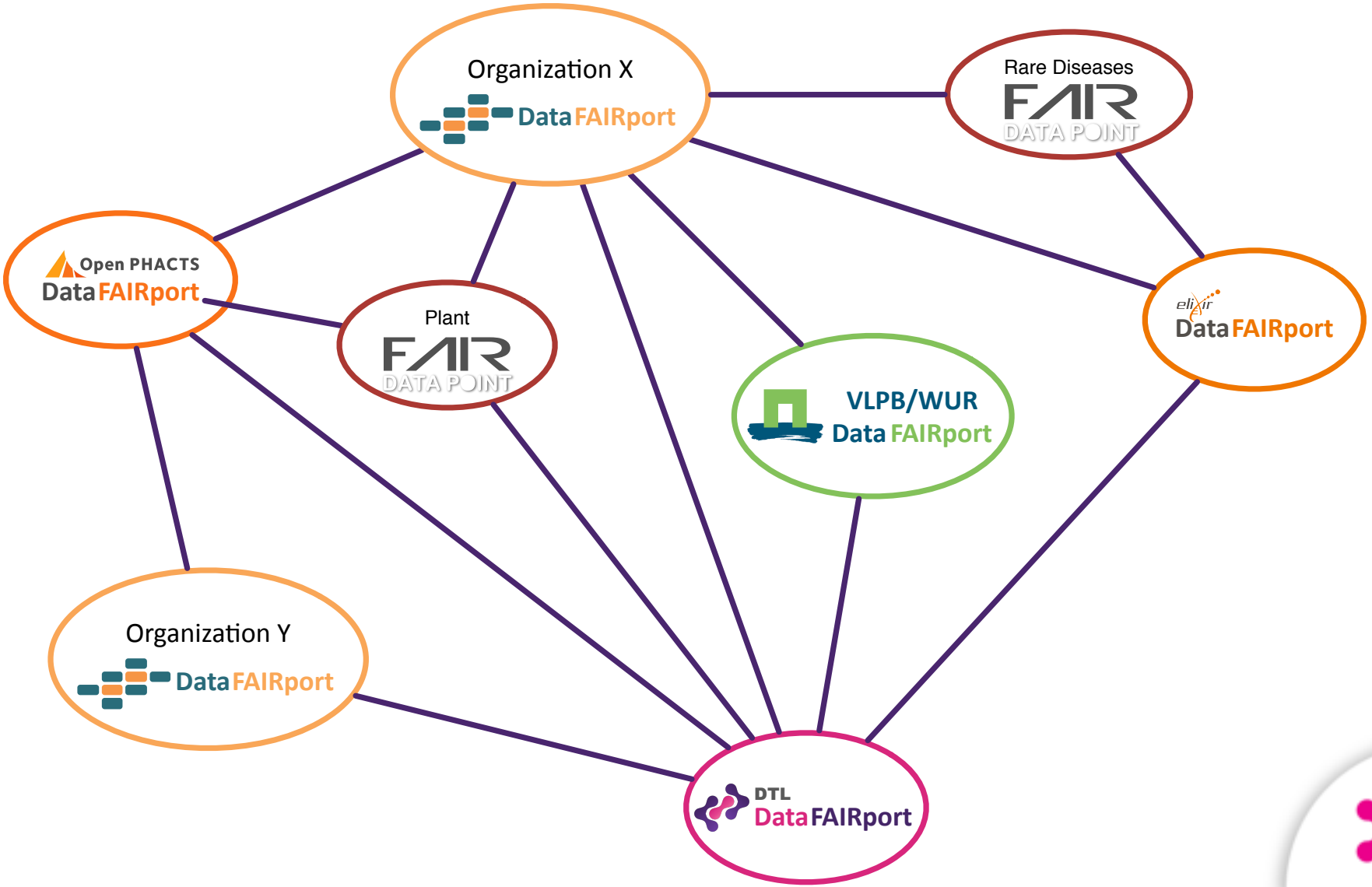
FAIR DATA PUBLICATION

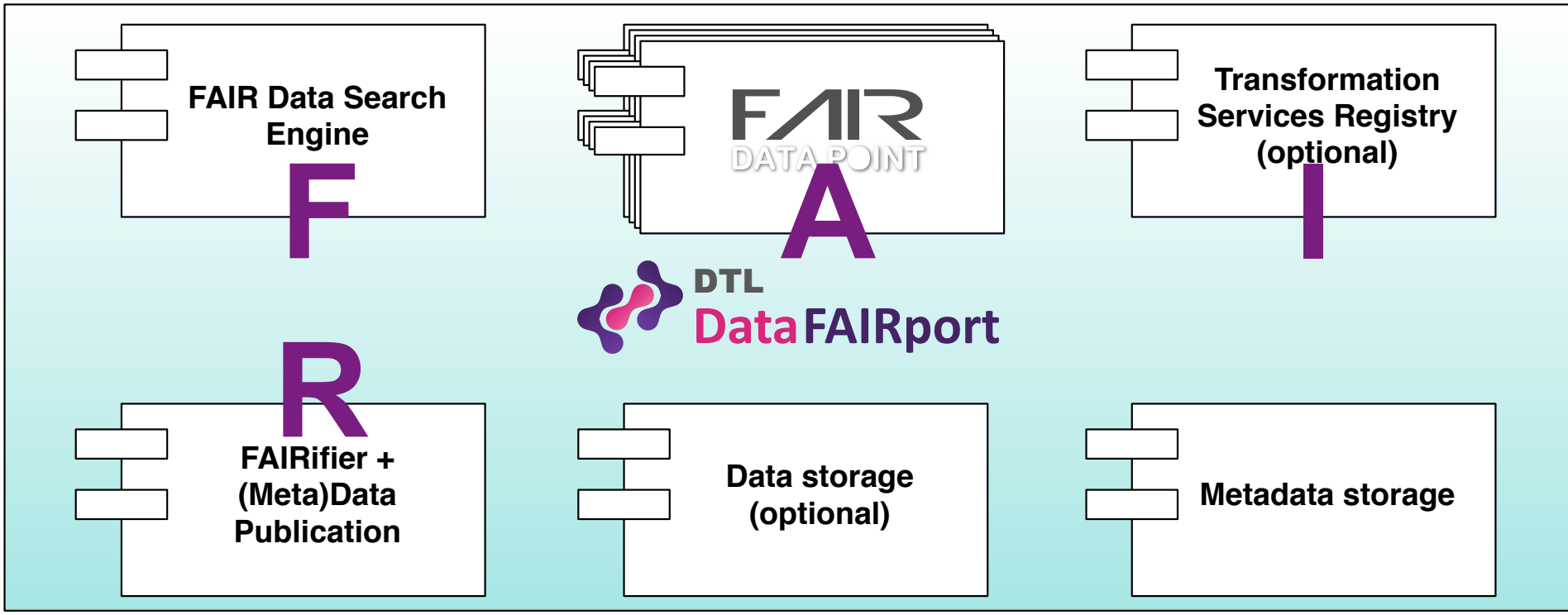


FAIR DATA ACCESS



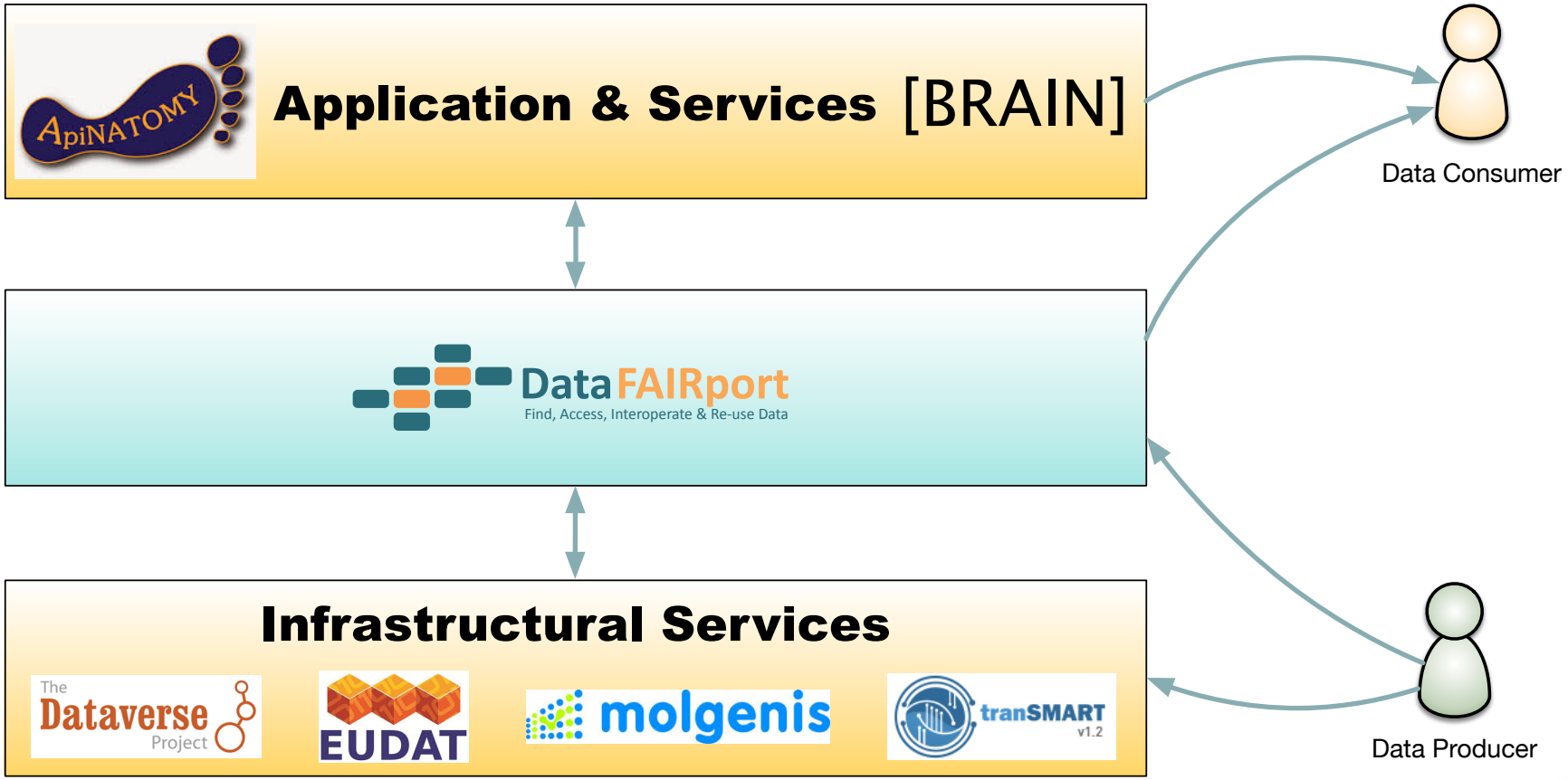
DISTRIBUTED ARCHITECTURE



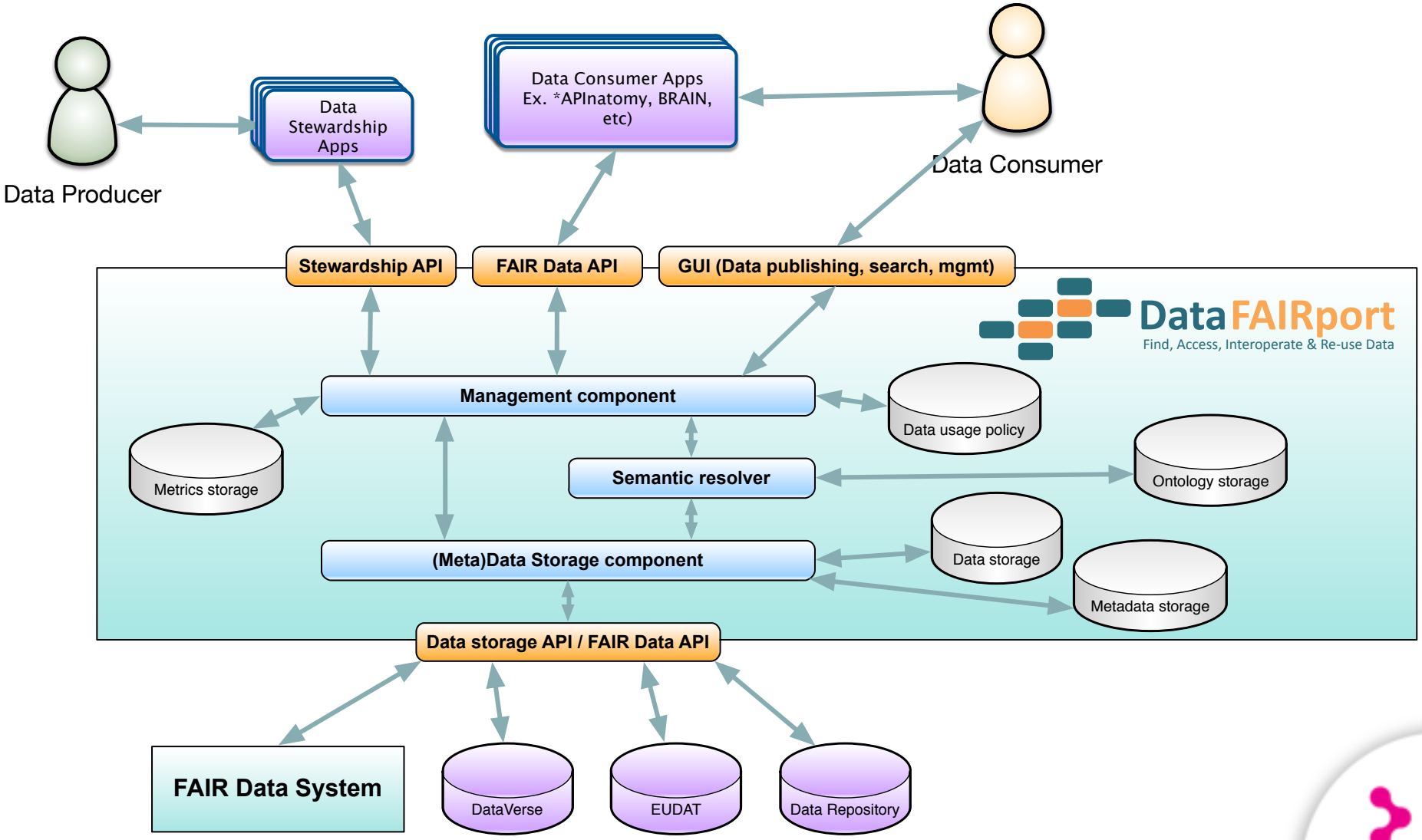


elixir NETHERLANDS

FAIRPORT ECOSYSTEM



FAIRPORT


Data Producer
Data Stewardship Apps
Data Consumer Apps
Ex. *APInatomy, BRAIN, etc)
Data Consumer
Stewardship API
FAIR Data API
GUI (Data publishing, search, mgmt)
DataFAIRport
Find, Access, Interoperate & Re-use Data
Management component
Data usage policy
Metrics storage
Semantic resolver
Ontology storage
(Meta)Data Storage component
Data storage
Metadata storage
Data storage API / FAIR Data API
FAIR Data System
DataVerse
EUDAT
Data Repository

WORK ORGANISATION (NL APPROACH)

HOW TO REALISE

projects

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

Netherlands Metabolomics Centre

BBMRI-NL

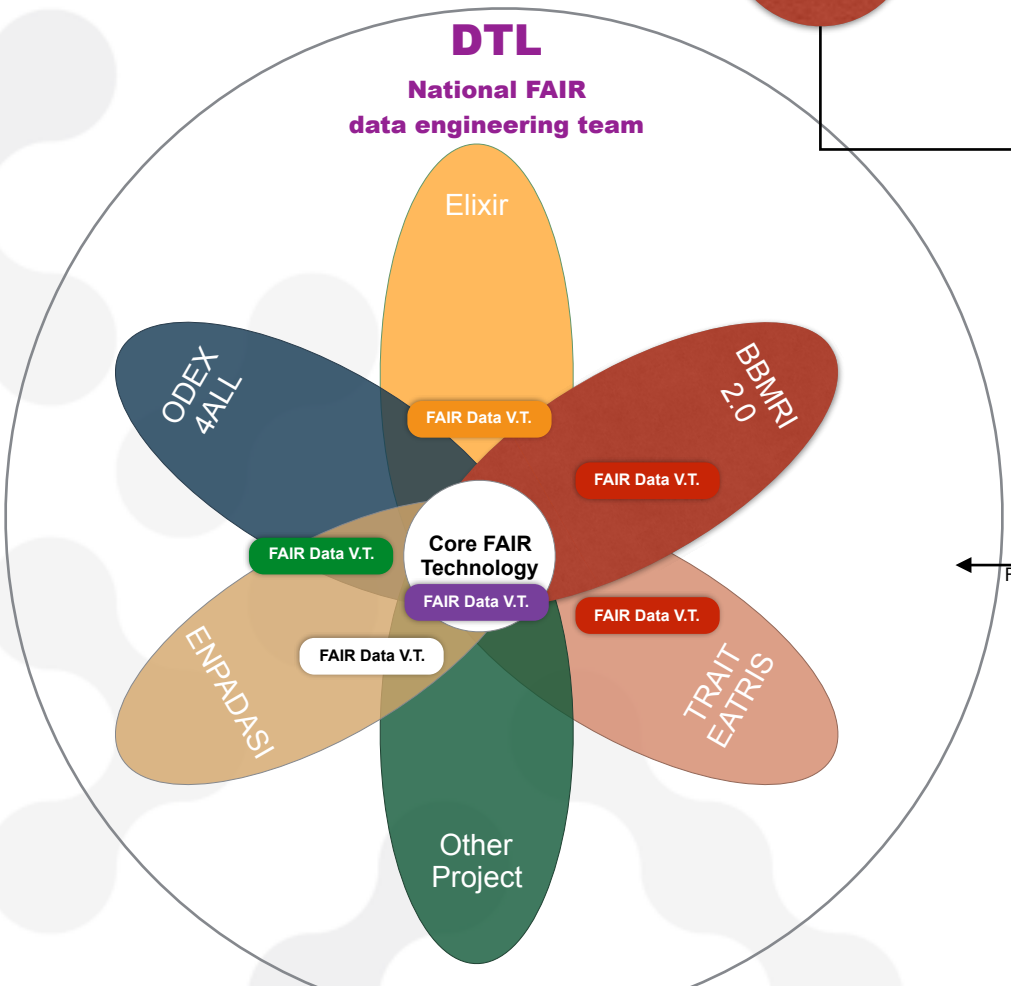
NFU NEDERLANDSE FEDERATIE VAN UNIVERSITAIR MEDISCHE CENTRA

Data SAC
G.M., B.M., M.G.

Data Core Team

FAIR Data Executive Team
L.B. (CTO), R.H., P.B., M.S., J.B., J.W.

engineers in the FAIR Data virtual team



QUESTIONS?