

Synergizing Biomedical Ontologies - An Industry Perspective

*Martin Romacker, Data and Information Architect
Scientific Solution Engineering and Architecture (S2EA)
Pharma Research and Early Development Informatics (pREDi)
Roche Innovation Center Basel
Workshop on Synergizing Biomedical Ontologies, 14th July 2021, virtual*

Workshop on Synergizing Biomedical Ontologies

Roche *pRED*

Digital Transformation – Prime Time for Biomedical Ontologies

Roche Data Commons, Reference Data Services & FAIR Principles

FAIR Maturity Indicators & FAIR Assessment

FAIR Maturity Indicators & FAIR APIs (JSON-LD)

Synergizing Biomedical Ontologies – Proliferation vs Convergence

Synergizing Biomedical Ontologies – Vision and Mission

Biomedical Ontologies

The Past

Management of Research Data Assets – The Strategic Importance of Biomedical Ontologies

(and noone really cares)

fine print

Practical Use of Biomedical Ontologies, EMBL-EBI, Hinxton UK March 16th-17th 2016

ISMB Conference, Bio Ontologies SIG Meeting, Orlando 8th July 2016



- Niche Players: poorly adopted in the industry
- Strategy: mainly partial solutions
- Technology Challenge: justification needed
- Perception issue: Data Quality strangely neglected
- Missing ownership: Poor interest in business
- Data Governance : Works only with pressure
- Range Issue: Only few people understand topic



Clear objective and urgent need vs poor adoption

Shift Happens – Towards a Data Driven Industry

Foundational Change: Perception of Value of Data (the Bright Future?)

Pharma Times online, 09. November 2016

British AI group licenses Janssen drug candidates

British artificial intelligence group BenevolentAI has signed an exclusive license with Johnson & Johnson group Janssen, picking up rights to a series of its novel clinical stage drug candidates. Under the deal, BenevolentAI has acquired a license to develop a select number of candidates and their extensive related portfolio of patents, after concluding that there is "strong promise" to develop them into new medicines for hard to treat diseases using its artificial intelligence technology.

[Read more](#) [Open in app](#)

Diese Verschiebung haben die Unternehmen laut PwC mit gutem Grund vorgenommen: Diejenigen Firmen, die im Branchenvergleich schneller als der Durchschnitt gewachsen seien, hätten 2015 im Schnitt 25% mehr Geld für Software-Entwicklung ausgegeben als die Unternehmen, deren Umsatz sich unterdurchschnittlich entwickelt habe, heisst es in der Studie.

IT as Key Enabler

ROCHE IST AUF RANG 7 ABGERUTSCHT

Gleichzeitig sind Unternehmen, die in der Öffentlichkeit als besonders innovativ gelten, nicht unbedingt diejenigen, die auch tatsächlich am meisten in F&E investieren. So führt wie im Vorjahr der Automobilkonzern VW die Liste mit Ausgaben von 13,2 Mrd USD an. Samsung folgt auf Platz zwei, gefolgt von Amazon.

Apple dagegen, das als innovativstes Unternehmen gilt, landet mit Ausgaben von 8,1 Mrd USD lediglich auf Platz 18. Alphabet, die Muttergesellschaft von Google, kommt mit 12,3 Mrd immerhin auf Platz vier.



«99% of the innovation is happening outside our companies»
(Severin Schwan, CEO F. Hoffmann-La Roche)

Pharma 4.0

- Digitilization
- Internet of Things
- Lab Automation
- Advanced Analytics
- Artificial Intelligence
- Big Data
- Data Science



Reimagining Novartis as a 'medicines and data science' company

Vas Narasimhan on LinkedIn

January 11, 2018

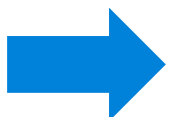
Machen Google und Co. schon bald Jagd auf Roche und Novartis?

Mit der milliarden schweren Übernahme von Whole Foods durch Amazon hat das Detailhandelssterben jenseits des Atlantiks einen traurigen Höhepunkt erreicht. Reihenweise sind kleinere Ladeninhaber gezwungen, ihre Geschäftstätigkeit aufzugeben. Gegen den mächtigen Versandhändler scheint kein Kraut gewachsen.

Der Vorstoss der Amerikaner ins Geschäft mit Nahrungsmitteln sollte auch anderen Wirtschaftszweigen eine Warnung sein. Denn immer öfter nutzen Tech-Giganten wie Amazon, Google und Co. ihre Milliarden von Dollar, um sich neue Märkte zu erschliessen.

SPIEGEL: Muss Roche mehr wie Google werden, sich zum Datenkonzern wandeln?

Franz: Absolut. Google muss erst die Pharmaerfahrung aufbauen – und wir müssen die Digitalisierung für uns nutzen. Mit den Produkten unserer Diagnostiksparte generieren wir im Jahr 15 Milliarden Tests, also Datenpunkte. Wir haben einen riesigen Datenschatz. Aber wir fangen erst an, ihn zu nutzen.



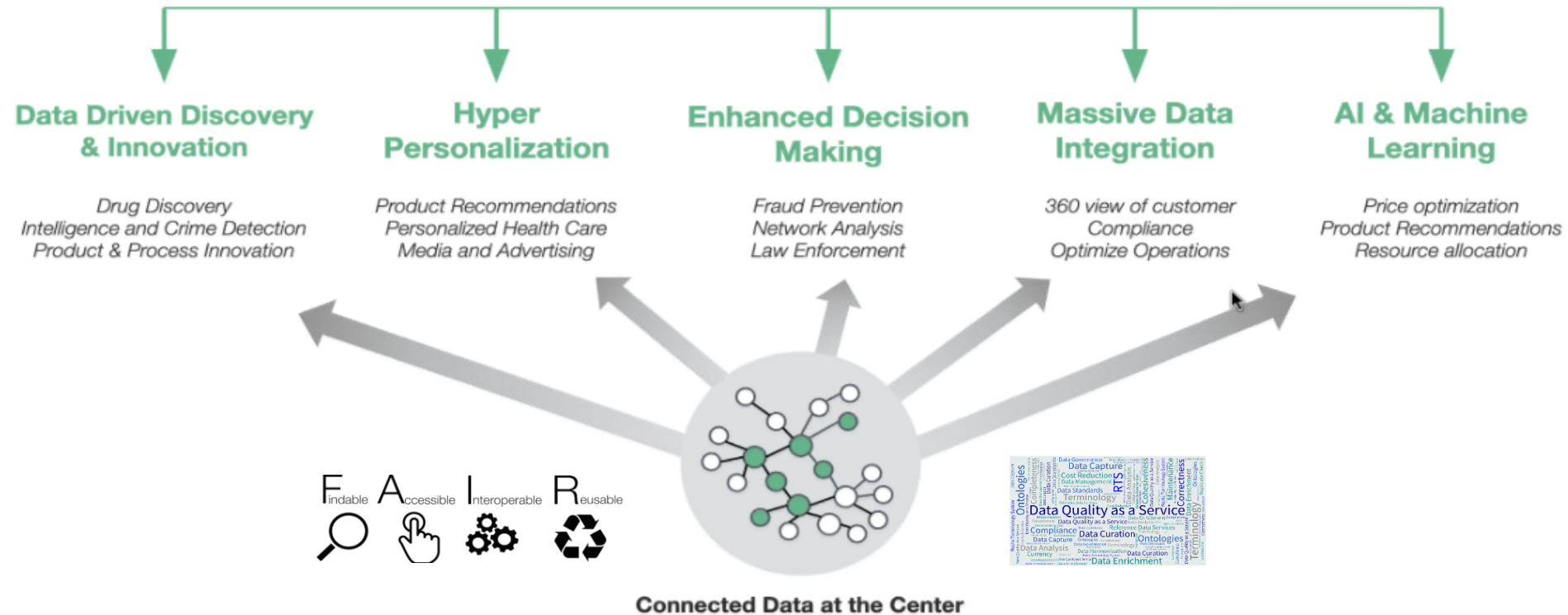
Data are no longer a by-product of business processes – business processes **are** data-driven

Digital Transformation

Megatrends & Data Management Strategy

Harnessing Connections Drives Business Value

Digital Transformation Megatrends



Data Standards: Terminology, Ontology & Data Models (FAIR+Q Data)

Data as an Asset

True Costs of Data Management

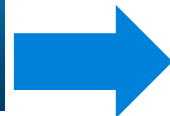


Planned/ Visible Costs

- FTEs creating Data Asset
- Material procurement (sample, reagent, compounds etc.)
- Infrastructure

Unplanned/ Invisible Costs

- ETL processes
- Searching & accessing
- Data Cleansing
- Data Curation/ Semantic Data Integration
- IT Infrastructure supporting unplanned activities



Backcharge the costs for processing to the data producers

Digital Transformation – Prime Time for Biomedical Ontologies

Roche Data Commons, Reference Data Services & FAIR Principles

FAIR Maturity Indicators & FAIR Assessment

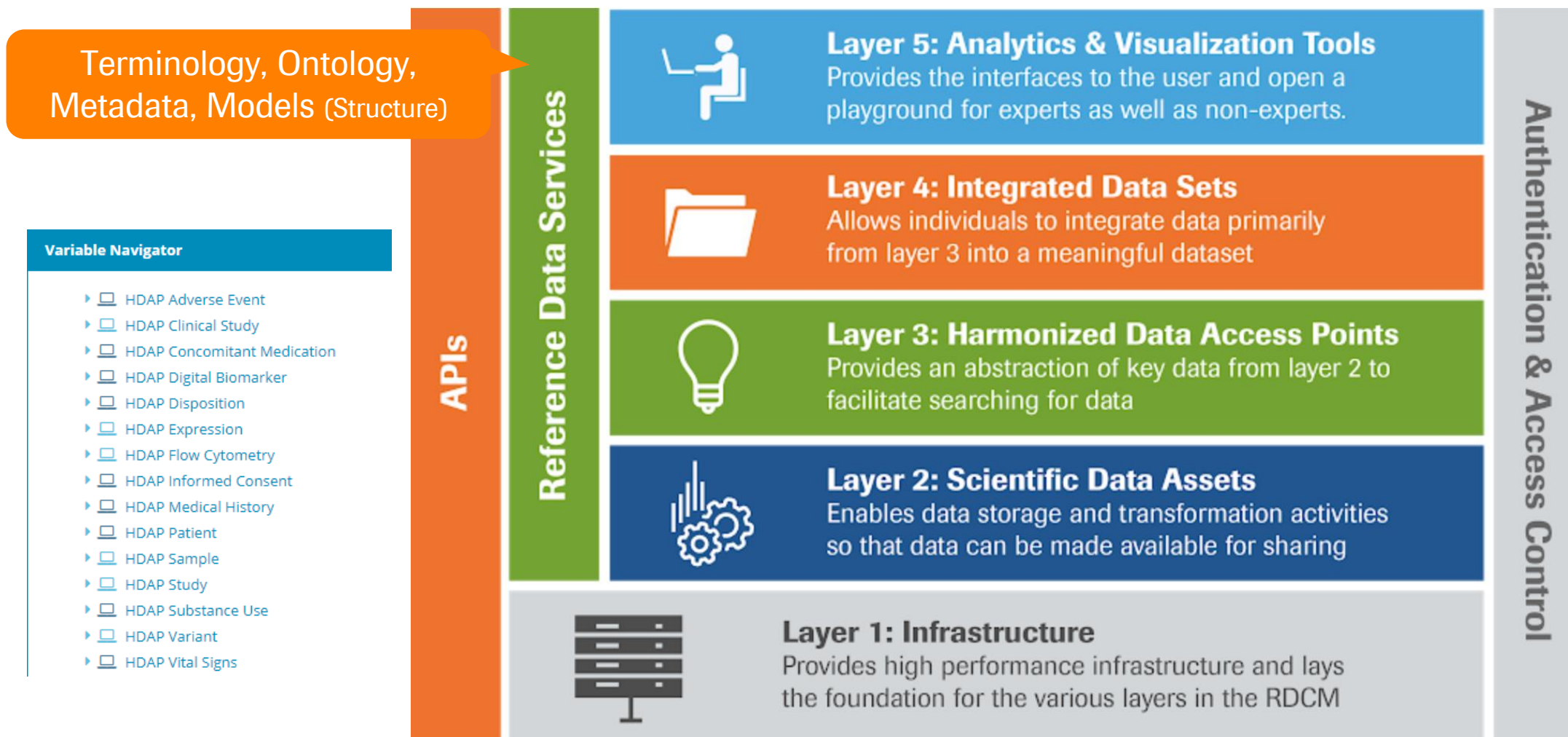
FAIR Maturity Indicators & FAIR APIs (JSON-LD)

Synergizing Biomedical Ontologies – Proliferation vs Convergence

Synergizing Biomedical Ontologies – Vision and Mission

Roche Data Commons (RDC) – Flipping the Coin in Data Mgt

Moving from an application-centric to an information-centric organization



Reference Data Services for Data Management

Terminology Management - Synergize Ontologies (FAIR)

Roche Terminology System v2.52.0 [PRD]

FAIR Metrics

Terminology Application Variable Curation Administration Information

-> More than 80 productive applications integrated

martin.romacker@roche.com

Logout

Search

✕

🔍

⌵

⌶

Scope: Indication

Terminology Navigator

Basaloid Carcinoma

Borderline Malignancy Carcinoma

Breast Carcinoma

Carcinoma ex Pleomorphic Adenoma

Carcinoma in Situ

Carcinoma of Unknown Primary

Embryonal Carcinoma

Epithelial-Myoepithelial Carcinoma

Gastric Pylorus Carcinoma

Head and Neck Carcinoma

Hepatocellular Carcinoma

Large Cell Carcinoma

Lung Mucoepidermoid Carcinoma

Lymphoepithelial Carcinoma

Mammary analogue secretory carcinoma

Metaplastic Carcinoma

Metastatic Carcinoma

Mucinous Carcinoma

Mucoepidermoid Carcinoma

Neuroendocrine Carcinoma

Non Small Cell Carcinoma

Non Small Cell Lung Cancer

Adenosquamous Cell Lung Carcinoma

ALK mutation positive non small cell lung cancer

EGFR mutation positive non small cell lung cancer

HER2 mutation positive non small cell lung cancer

Application Navigator

E-Sample Flow

e21 HCP Portal

EpiCX

Approved Indication

Country

CP Indication Mapping

Global Product Name

Product

Scientific Area Indication

Trademark Name AU

Trademark Name BR

Trademark Name CA

Trademark Name TW

Trademark Name US

Actemra

Alecensa

ANTI-HER2 TDC

astegolimab

Avastin

Cervical Cancer

Colorectal Cancer

Glioblastoma

Malignant Mesothelioma

Non-Small Cell Lung Cancer

Ovarian Cancer

Renal Cell Carcinoma

balovaptan

basmisanil

Boniva

Cadherin-11 mAB

Cathflo Activase

CEACAM5 CD3 TCB

Concept Entity Properties

General information

Label: Non Small Cell Lung Cancer

Status: Active

Terminology: Indication

Identifier: ROX1305277804386

Definition

A group of at least three distinct histological types of lung cancer, including squamous cell carcinoma, adenocarcinoma, and large cell

Comment

References

Relations

Application References

Landing Page

Lifecycle

Changes

Application

Concept

Master Terminology

Application Terminology

Label	Language	Source	Label Type	Lexical Type	
Non Small Cell Lung Cancer	en	Roche	Synonym	prefLabel	
Cancer, lung, non small cell	en	PIP	Synonym	altLabel	
Cancer, non small cell lung	en	Roche	Synonym	altLabel	
Carcinoma, Non Small Cell Lung	en	Roche	Synonym	altLabel	
Carcinoma, non small cell lung	en	Roche	Synonym	altLabel	
Carcinoma, non small cell lung cancer	en	Roche	Synonym	altLabel	
Carcinoma, Non-Small-Cell Lung	en	Roche	Synonym	altLabel	
Non small cell lung cancer	en	ADIS, TPP	Synonym	altLabel	
Non small cell lung cancer (NSCLC)	en	Roche	AcroDefinition	altLabel	

Reference Data Services for Data Management

Metadata Registry/ Dataset Models – Synergize Ontologies (FAIR)

Roche Terminology System v2.52.0 [PRD]

FAIR Metrics

martin.romacker@roche.com [Logout](#)

Terminology

Application

Variable

Curation

Administration

Information

Search

Country

Scope: SP Variable

Terminology Navigator

- DM variable
 - Actual Arm Code
 - Age
 - Age Units
 - Animal Status
 - Birth Delivery Procedure
 - Country
 - Date and Time of Death
 - Date Time of Birth
 - Date Time of Data Collection
 - Date Time of End of Participation
 - Date Time of First Study Treatment
 - Date Time of Informed Consent
 - Date Time of Last Study Treatment
 - Description of Actual Arm
 - Description of Planned Arm
 - Domain Abbreviation
 - Ethnicity
 - Globally Unique Subject Identifier
 - Investigator Identifier
 - Investigator Name
 - Organism Species Subspecies

Variable Navigator

- HDAP Subject
 - DM Domain
 - Age
 - Age in Days
 - Analysis Age
 - Baseline Body Mass Index (kg per m2)
 - Country
 - Date of Death
 - End Date Time of Treatment
 - End Date of Last Treatment
 - Ethnicity
 - Intent-To-Treat Population Flag
 - Link to Layer 2 dataset
 - Race
 - Safety Population Flag
 - Sex
 - Start Date Time of Treatment
 - Start Date of First Treatment
 - Subject Class Identifier
 - Time from Diagnosis to Rnd (years)
 - Unique Subject Identifier
- HDAP Substance Use
- HDAP Variant
- HDAP Vital Signs
- HDPA Tumor Identification
- HGDI
- HTAg
- I2O Knowledge Base
- IDMP

Application Entity Property

General information

Variable name: Country

Value Domain type: Application Terminology

App Terminology: Country Code (Alpha 3)

Variable Multiplicity: single-valued

Variable Policy: Required Variable

Curation Policy:

Variable Context:

Definition

Country of the investigational site in which the subject participated in the trial (GDSR).

Comment

ISO 3166 format.

Concept Reference

Concept

Country

Link

Data Dictionary

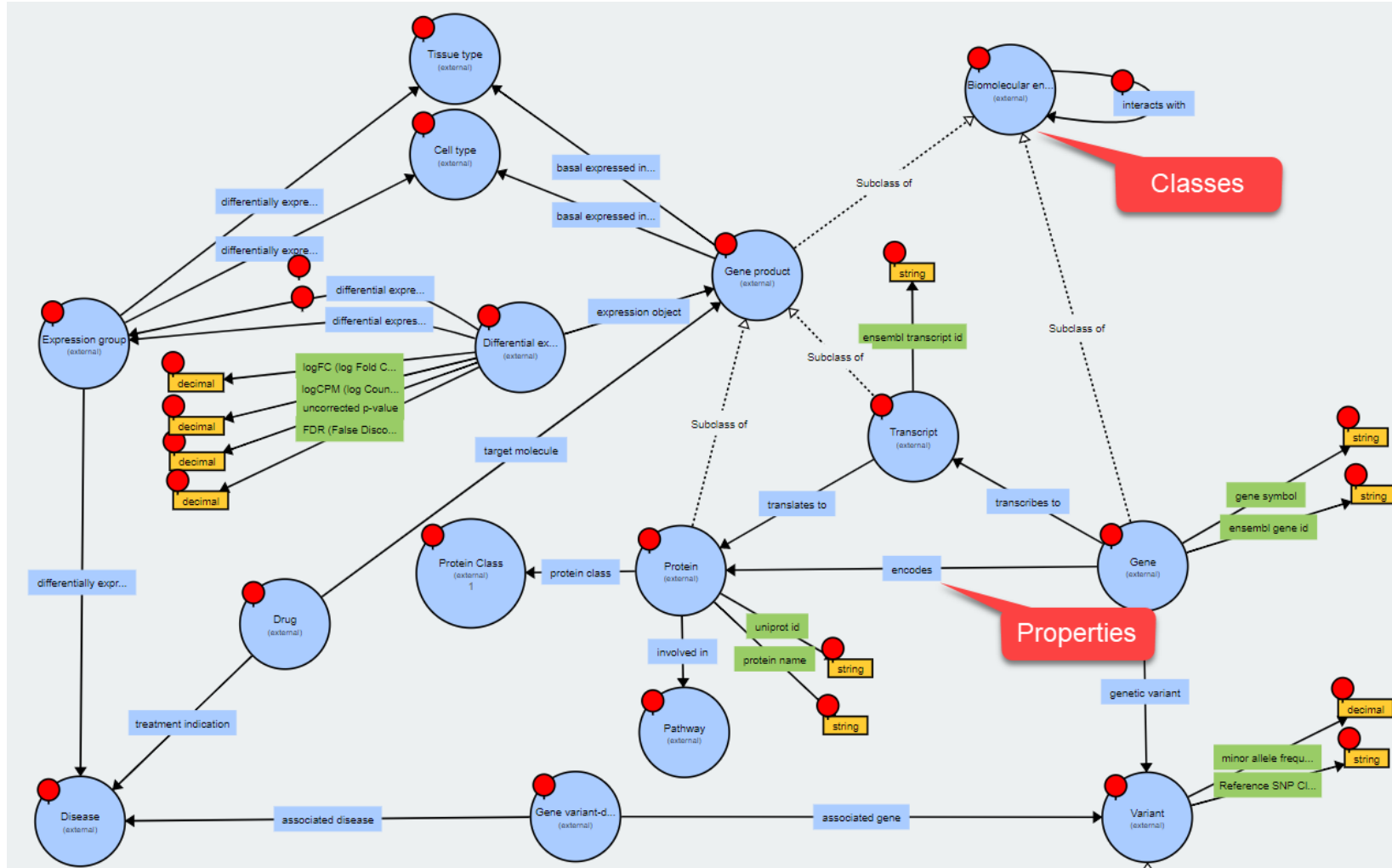
Application

Variable Properties

Variable

Data Domains & purpose-driven Ontologies

Building Knowledge Graphs



- FAIR: fully harmonized
- Linked to standards
- Linked to ontologies
- Instantiated with data
- Federated queries

Roche Data Commons (RDC)

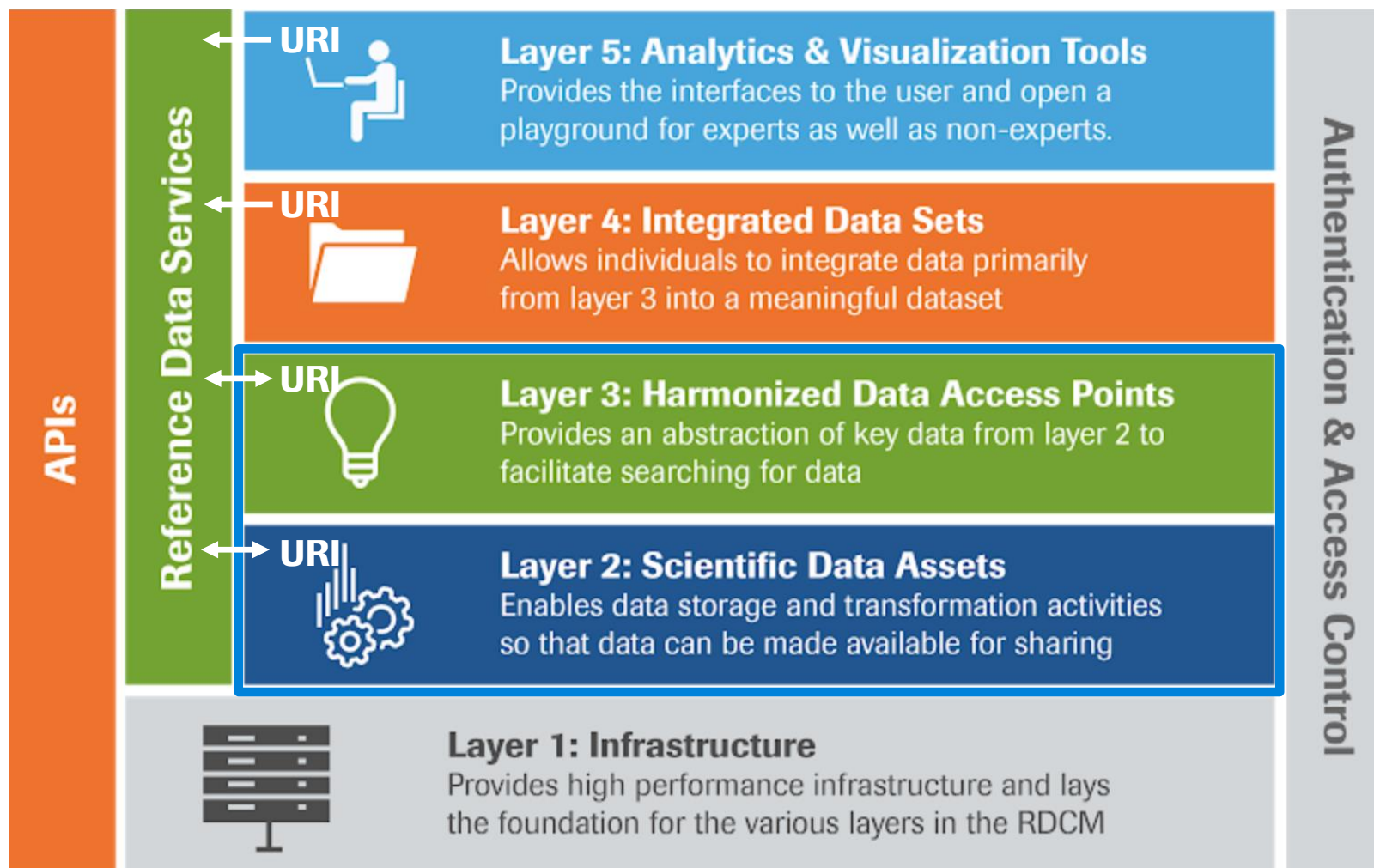
Data FAIRification – Everything is a Resource (URI)!

HDAPs organize data in Information Types

Interoperability (URIs):
semantic data dictionary
semantic models

Data FAIRification
only in layer 2 & 3

No more transformation
between layer 3 & 4,5



Digital Transformation – Prime Time for Biomedical Ontologies

Roche Data Commons, Reference Data Services & FAIR Principles

FAIR Maturity Indicators & FAIR Assessment

FAIR Maturity Indicators & FAIR APIs (JSON-LD)

Synergizing Biomedical Ontologies – Proliferation vs Convergence

Synergizing Biomedical Ontologies – Vision and Mission

FAIR scientific data management

FAIR guiding principles

FAIR

Ability for scientist/data consumer to find, access and understand the data
(without the presence of the data owner)



Ability for a machine to automatically find and use the data
(machine actionable)

by Olivier Roche (pREDi)

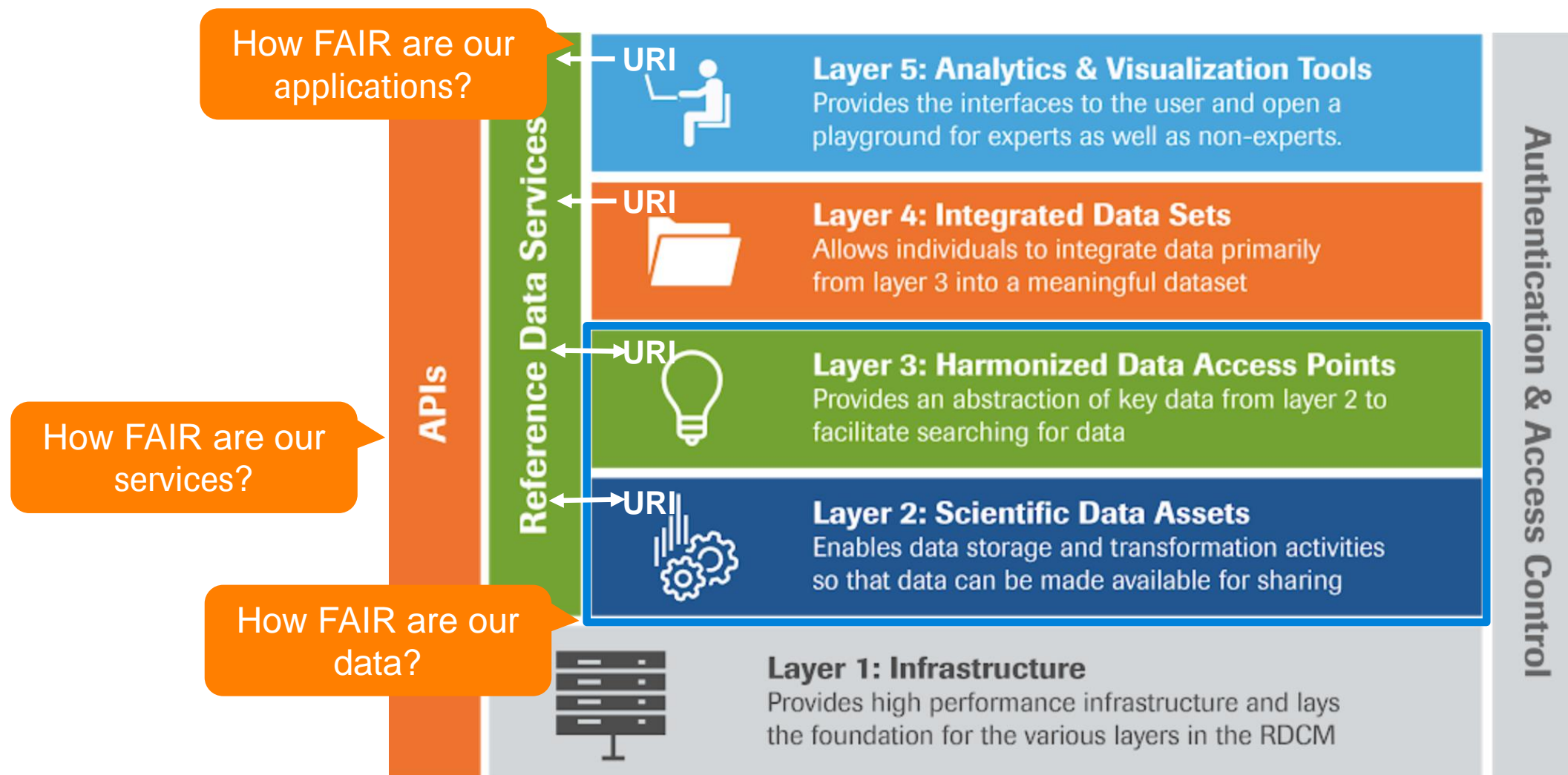
Dilution of the FAIR Principles:

FAIR is not primarily about the *THAT*

FAIR is above all about the *HOW*

Roche Data Commons (RDC)

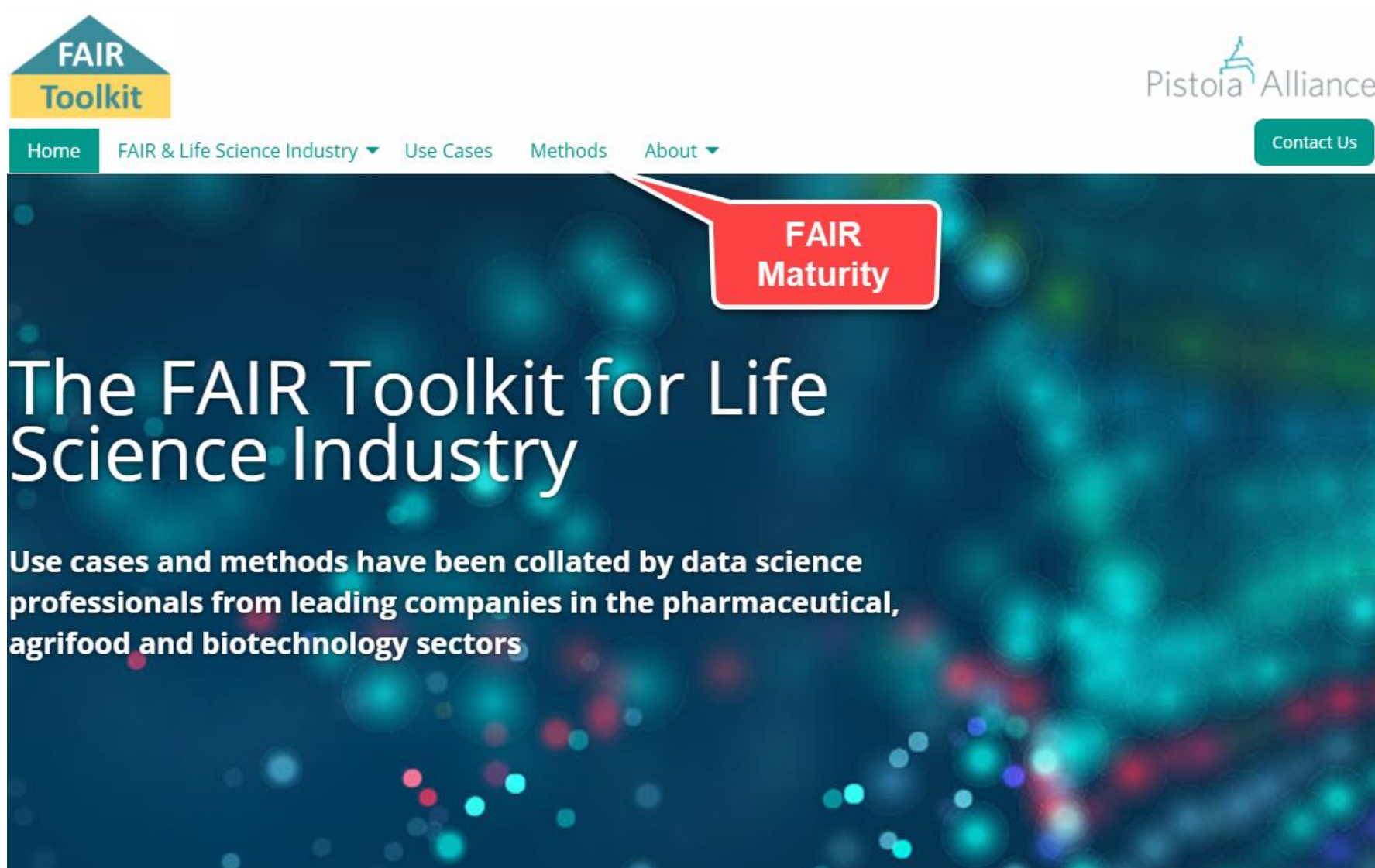
FAIR Maturity Indicators (FAIR Metrics)



FAIR Assessment

Pistoia Alliance

Roche



[FAIR Toolkit](#)

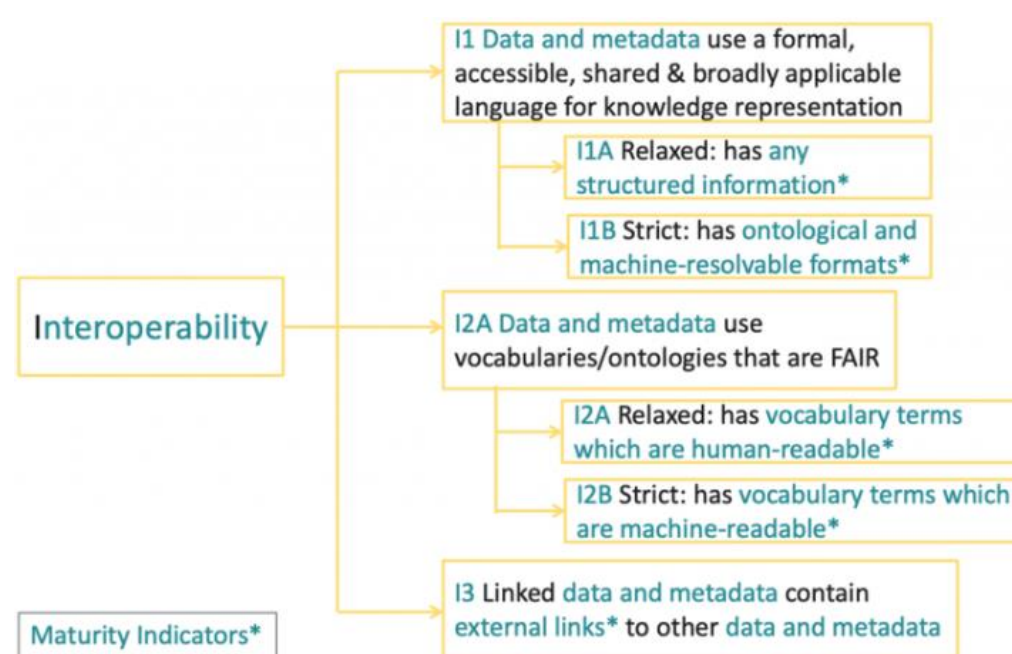
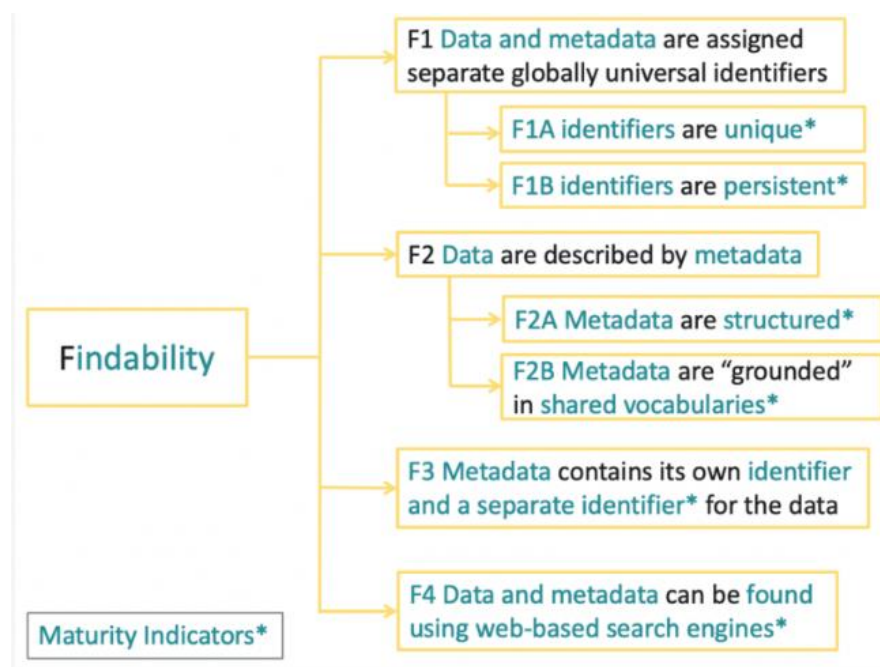
Implementation for FAIR Data Principles in Life Science R&D

Maturity Indicators: FAIR Metrics



<https://fairtoolkit.pistoiaalliance.org/>

Home FAIR & Life Science Industry ▼ Use Cases Methods About ▼



➡ Cross-site & cross- functional project: FAIR Maturity Assessment

FAIR Principle & FAIR Metrics

5 Star Rating

FAIR Principles

Home > FAIR Principles

> FAIR Principles

- > **F1: (Meta) data are assigned globally unique and persistent identifiers**
- > **F2: Data are described with rich metadata**
- > **F3: Metadata clearly and explicitly include the identifier of the data they describe**
- > **F4: (Meta)data are registered or indexed in a searchable resource**
- > **A1: (Meta)data are retrievable by their identifier using a standardised communication protocol**
- > **A1.1: The protocol is open, free and universally implementable**
- > **A1.2: The protocol allows for an authentication and authorisation where necessary**
- > **A2: Metadata should be accessible even when the data is no longer available**
- > **I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation**
- > **I2: (Meta)data use vocabularies that follow the FAIR principles**
- > **I3: (Meta)data include qualified references to other (meta)data**
- > **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**
- > **How to GO FAIR**
- > **FAIRification Process**

In 2016, the 'FAIR Guiding Principles for scientific data management and stewardship' were published in *Scientific Data*. The authors intended to provide guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets. The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.

A practical 'how to' guidance to go FAIR can be found in the **Three-point FAIRification Framework**.

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the **FAIRification process**.

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 they describe

F4. (Meta)data are registered or indexed in a searchable resource

Accessible

Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

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

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

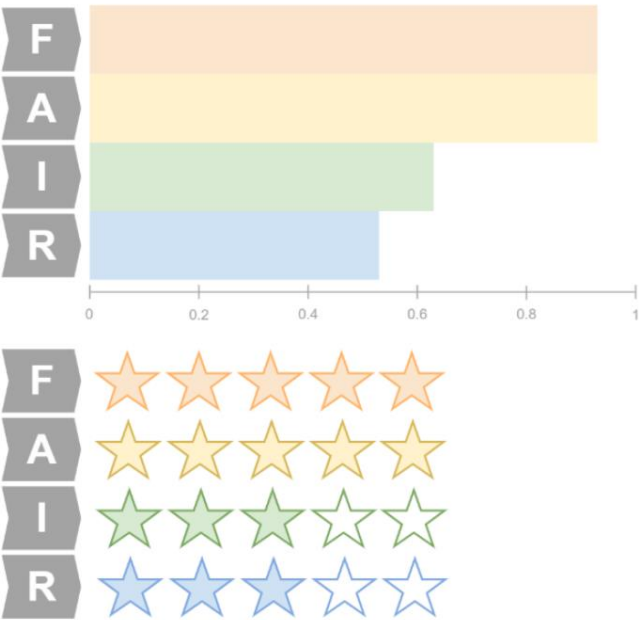
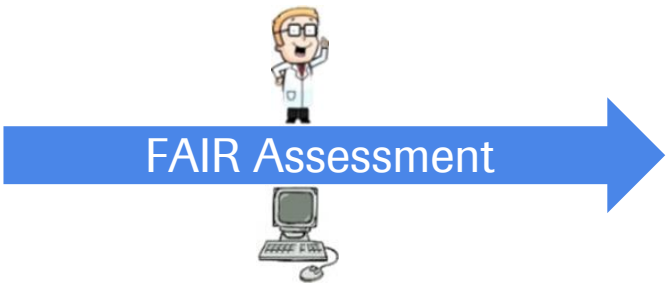
R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

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

The principles refer to three types of entities: data (or any digital object), metadata (information about that digital object), and infrastructure. For instance, principle F4 defines that both metadata and data are registered or indexed in a searchable resource (the infrastructure component).



Digital Transformation – Prime Time for Biomedical Ontologies

Roche Data Commons, Reference Data Services & FAIR Principles

FAIR Maturity Indicators & FAIR Assessment

FAIR Maturity Indicators & FAIR APIs (JSON-LD)

Synergizing Biomedical Ontologies – Proliferation vs Convergence

Synergizing Biomedical Ontologies – Vision and Mission

Digital Data Assets & Data Management

The Hamster Wheel



Data Transformation (map & merge)

clinical-study HDAP :

```
{...  
StudyIndication:  
"Non-small cell lung cancer",  
...  
}
```

pRED-study hdap :

```
{...  
TargetDisease :  
"NSCLC",  
...  
}
```

clinical-standard respository :

```
{...  
TherapeuticIndication:  
"Carcinoma, non-small cell, lung",  
...  
}
```


JSON-Linked Data (JSON-LD)

Leveraging on a Semantic Infrastructure

```

{
  "@graph" : [ {
    "@id" : "ROX1305277804386",
    "contributor" : "JIMENES6",
    "broader" : [ "ROX1305277804385", "ROX1305277805920", "ROX1394550342848" ],
    "definition" : "A group of at least three distinct histological types of lung cancer, including squamous cell carcinoma, adenocarcinoma, and
large cell carcinoma. Non-small cell lung carcinomas have a poor response to conventional chemotherapy.",
    "status" : {
      "@id" : "ROX11410222618619111",
      "prefLabel" : "Active"
    },
    "rdfs:label" : {
      "@id" : "ROX32426970969993323",
      "labelTypeConcept" : {
        "@id" : "ROX32508475213363140",
        "prefLabel" : "Synonym"
      },
      "languageConcept" : {
        "@id" : "ROX32410222618619687",
        "prefLabel" : "en"
      },
      "sourceConcept" : {
        "@id" : "ROX32508475213363138",
        "prefLabel" : "Roche"
      }
    },
    "literalForm" : "Non Small Cell Lung Cancer"
  } ]
}

```

Subject

Predicate

Object



In a universe of FAIR applications, data and services **everything** should be considered as a resource

JSON-Linked Data (JSON-LD)

Context provides Model for unambiguous interpretation

Model

```
"@context" : {
  "@base": "http://ontology.roche.com/" ,
  "prefLabel" : {
    "@id" : "http://www.w3.org/2004/02/skos/core#prefLabel"
  },
  "broader" : {
    "@id" : "http://www.w3.org/2004/02/skos/core#broader",
    "@type" : "@id"
  },
  "contributor" : {
    "@id" : "http://purl.org/dc/terms/contributor"
  },
  "definition" : {
    "@id" : "http://www.w3.org/2004/02/skos/core#definition"
  },
  "status" : {
    "@id" : "http://ontology.roche.com/status",
    "@type" : "@id"
  },
  "sourceConcept" : {
    "@id" : "http://ontology.roche.com/sourceConcept",
    "@type" : "@id"
  },
  "languageConcept" : {
    "@id" : "http://ontology.roche.com/languageConcept",
    "@type" : "@id"
  },
  "labelTypeConcept" : {
    "@id" : "http://ontology.roche.com/labelTypeConcept",
    "@type" : "@id"
  },
  "literalForm" : {
    "@id" : "http://www.w3.org/2008/05/skos-xl#literalForm"
  },
  "rts" : "http://ontology.roche.com/",
  "dct" : "http://purl.org/dc/terms/",
  "skosxl" : "http://www.w3.org/2008/05/skos-xl#",
  "xsd" : "http://www.w3.org/2001/XMLSchema#",
  "skos" : "http://www.w3.org/2004/02/skos/core#",
  "dc" : "http://purl.org/dc/elements/1.1/"
}
```



What is a **SmartAPI**?

The SmartAPI project aims to maximize the FAIRness (Findability, Accessibility, Interoperability, and Reusability) of web-based Application Programming Interfaces (APIs). Rich metadata is essential to properly describe your API so that it becomes discoverable, connected, and reusable. We have developed an openAPI-based [specification](#) for defining the key API metadata elements and value sets. SmartAPI's leverage the [Open API specification v3](#) and [JSON-LD](#) for providing semantically annotated JSON content that can be treated as [Linked Data](#).

Digital Data Assets & Data Management

Breaking up the Vicious Circle

Instantaneous Integration of Data & Metadata



```
clinical-study HDAP: {...  
StudyIndication: {  
  @id : ROX1305277804386,  
  prefLabel :  
    "Non-small cell lung cancer"}  
... }
```

```
"@context" : {...  
"StudyIndication" : {  
  "@id" : ROX37603872443814754,  
  "@type" : "@id"}  
... }
```

```
pRED-study: {...  
TargetDisease: {  
  @id : ROX1305277804386,  
  prefLabel :  
    "NSCLC"}  
... }
```

```
"@context" : {...  
"TargetDisease" : {  
  "@id" : ROX37603872443814754,  
  "@type" : "@id"}  
... }
```

```
clinical-standard respository: {...  
TherapeuticIndication: {  
  @id : ROX1305277804386,  
  prefLabel :  
    "Carcinoma, non-small cell, lung"}  
... }
```

```
"@context" : {...  
"TherapeuticIndication" : {  
  "@id" : ROX37603872443814754,  
  "@type" : "@id"}  
... }
```

JSON-Linked Data (JSON-LD)

RDF Serialization – immediate usage

Expanded

Compacted

Flattened

Framed

N-Quads

Normalized

Table

Visualized

Signed with RSA

Signed with Bitcoin

```

<http://ontology.roche.com/ROX11410222618619111> <http://www.w3.org/2004/02/skos/core#prefLabel> "Active" .
<http://ontology.roche.com/ROX1305277804386> <http://ontology.roche.com/status> <http://ontology.roche.com/ROX11410222618619111> .
<http://ontology.roche.com/ROX1305277804386> <http://purl.org/dc/terms/contributor> "JIMENES6" .
<http://ontology.roche.com/ROX1305277804386> <http://www.w3.org/2004/02/skos/core#broader> <http://ontology.roche.com/ROX1305277804385> .
<http://ontology.roche.com/ROX1305277804386> <http://www.w3.org/2004/02/skos/core#broader> <http://ontology.roche.com/ROX1305277805920> .
<http://ontology.roche.com/ROX1305277804386> <http://www.w3.org/2004/02/skos/core#broader> <http://ontology.roche.com/ROX1394550342848> .
<http://ontology.roche.com/ROX1305277804386> <http://www.w3.org/2004/02/skos/core#definition> "A group of at least three distinct histological types
of lung cancer, including squamous cell carcinoma, adenocarcinoma, and large cell carcinoma. Non-small cell lung carcinomas have a poor response to
conventional chemotherapy." .
<http://ontology.roche.com/ROX1305277804386> <http://www.w3.org/2008/05/skos-xl#prefLabel> <http://ontology.roche.com/ROX32426970969993323> .
<http://ontology.roche.com/ROX32410222618619687> <http://www.w3.org/2004/02/skos/core#prefLabel> "en" .
<http://ontology.roche.com/ROX32426970969993323> <http://ontology.roche.com/labelTypeConcept> <http://ontology.roche.com/ROX32508475213363140> .
<http://ontology.roche.com/ROX32426970969993323> <http://ontology.roche.com/languageConcept> <http://ontology.roche.com/ROX32410222618619687> .
<http://ontology.roche.com/ROX32426970969993323> <http://ontology.roche.com/sourceConcept> <http://ontology.roche.com/ROX32508475213363138> .
<http://ontology.roche.com/ROX32426970969993323> <http://www.w3.org/2008/05/skos-xl#literalForm> "Non Small Cell Lung Cancer" .
<http://ontology.roche.com/ROX32508475213363138> <http://www.w3.org/2004/02/skos/core#prefLabel> "Roche" .
<http://ontology.roche.com/ROX32508475213363140> <http://www.w3.org/2004/02/skos/core#prefLabel> "Synonym" .

```


Digital Transformation – Prime Time for Biomedical Ontologies

Roche Data Commons, Reference Data Services & FAIR Principles

FAIR Maturity Indicators & FAIR Assessment

FAIR Maturity Indicators & FAIR APIs (JSON-LD)

Synergizing Biomedical Ontologies – Proliferation vs Convergence

Synergizing Biomedical Ontologies – Vision and Mission

FAIRsharing Catalog of Biomedical Resources

Proliferation and Fragmentation of Standards

FAIRsharing.org standards, databases, policies

Search all of FAIRsharing

Standards Databases Policies Collections Add/Claim Content Stats Log in or Register

Standards

Contribute by adding a standard Any problems? Please tell us!

The standards in FAIRsharing are manually curated from a variety of sources, including [BioPortal](#), [MIBBI](#) and the [Equator Network](#).

Manually done-
no smart interfaces

Search Standards Search Search Reset Advanced

Showing records 1 - 50 of 1299.

View as Table View as Grid

Sort by Name

Recommended Records

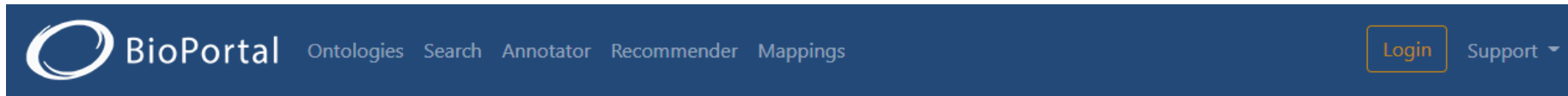
Recommended

Associated Publication?

Registry	Name	Abbreviation	Type	Subject	Related Database	Related Standard	Related Policy	In Collection/Recommendation	Status
ABA Adult Mouse Brain	ABA	Standard	None		None	None	None	None	R
Access to Biological Collection Data	ABCD	Standard	Biodiversity Biology Life Sciences	None	GBIF Atlas of Living Australia IPT - GBIF Australia	ABCD EFG ABCDDNA	None	TDWG Biodiversity Information Standards	R

Synergizing Biomedical Ontologies

Proliferation vs Convergence



Class Search

Enter a class, e.g. Melanoma

[help](#)

[Show advanced options](#)

Search

of hits

Matches in 24 ontologies

Malignant melanoma - Medical Dictionary for Regulatory Activities Terminology (MedDRA) (MEDDRA)

<http://purl.bioontology.org/ontology/MEDDRA/10025650>

[details](#) - [visualize](#) - [13 more from this ontology](#)

Malignant melanoma - SNOMED CT (SNOMEDCT)

<http://purl.bioontology.org/ontology/SNOMEDCT/372244006>

[details](#) - [visualize](#) - [27 more from this ontology](#)

mapping

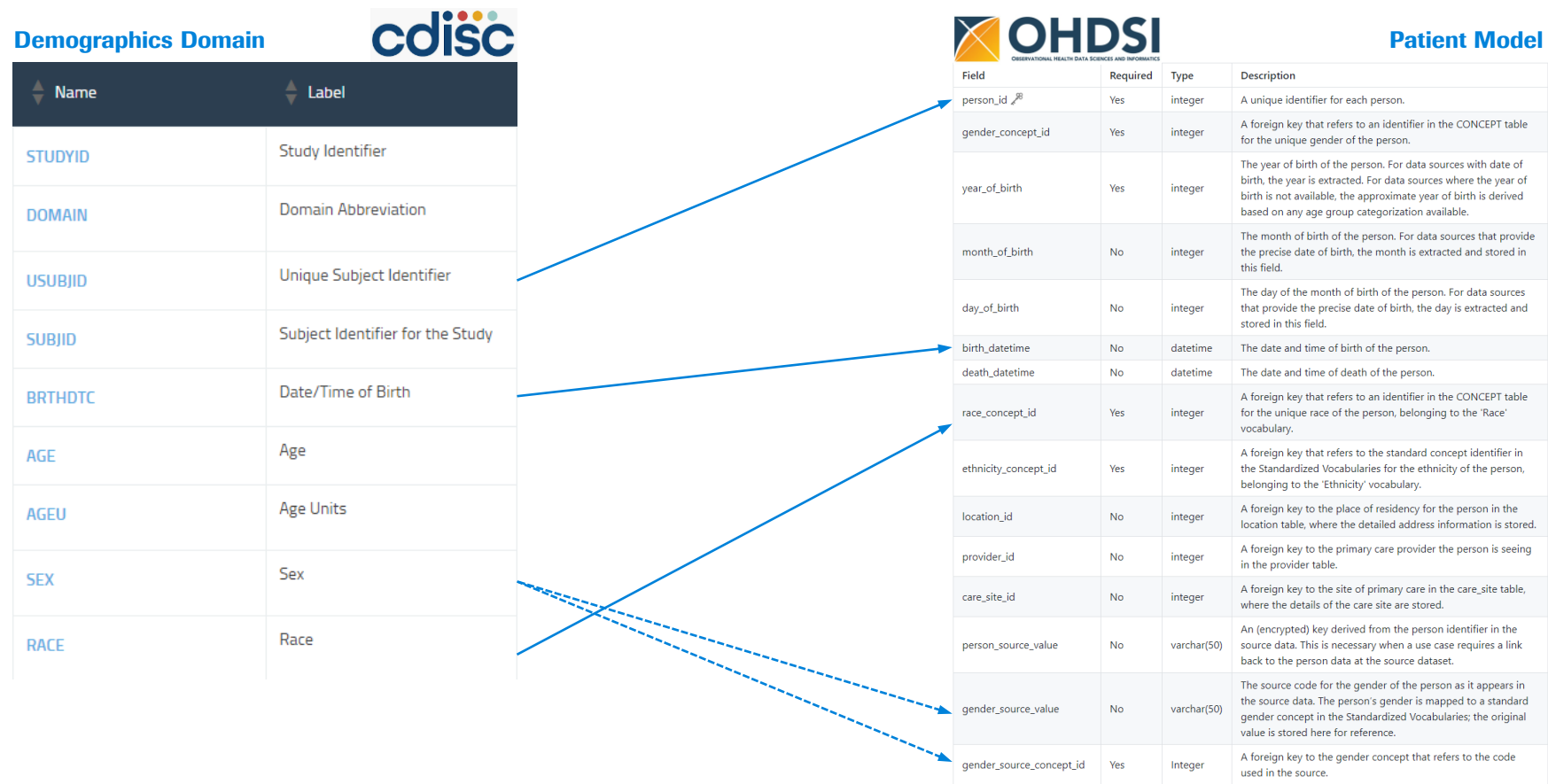
F1 Data and metadata are assigned separate globally universal identifiers

F1A identifiers are unique*

F1B identifiers are persistent*

Data Standards & Interoperability Challenges

CDISC vs OMOP/ OHDSI



Creation of insights & analytics blocked: different model, variables and values

Convergence of Biomedical Ontologies

Cellosaurus – the winner takes it all (1)

Expasy

Cellosaurus



Cellosaurus - a knowledge resource on cell lines

excellent
coverage

Release information: Version 38 (May 2021)
128806 cell lines (96820 human, 21791 mouse, 2444 rat)

[Description of the Cellosaurus](#)

[Browse by cell line group](#)

[Browse by cell line panel](#)

[Browse problematic \(contaminated/misidentified\) cell lines](#)

[CLASTR - STR similarity search](#)

[Download complete Cellosaurus data in various formats](#)



Convergence of Biomedical Ontologies

Cellosaurus – the winner takes it all (2)

Cell line name	COLO 741																
Synonyms	COLO-741; Colo741; COLO741; Colorado 741																
Accession	CVCL_1133																
Resource Identification Initiative	To cite this cell line use: COLO 741 (RRID:CVCL_1133)																
Comments	<p>Problematic cell line: Misclassified. Originally thought to be a colon adenocarcinoma (PubMed=25926053).</p> <p>Part of: AstraZeneca Colorectal cell line (AZCL) panel.</p> <p>Part of: Cancer Cell Line Encyclopedia (CCLE) project.</p> <p>Part of: COSMIC cell lines project.</p> <p>Part of: MD Anderson Cell Lines Project.</p> <p>Doubling time: 30 hours (PubMed=25984343).</p> <p>Microsatellite instability: Stable (MSS) (Sanger).</p> <p>Omics: Deep exome analysis.</p> <p>Omics: Deep proteome analysis.</p> <p>Omics: Deep quantitative proteome analysis.</p> <p>Omics: Deep RNAseq analysis.</p> <p>Omics: DNA methylation analysis.</p> <p>Omics: Protein expression by reverse-phase protein arrays.</p> <p>Omics: shRNA library screening.</p> <p>Omics: SNP array analysis.</p> <p>Omics: Transcriptome analysis.</p>																
Sequence variations	<p>Mutation; HGNC; 1097; BRAF; Simple; p.Val600Glu (c.1799T>A); ClinVar=VCV000013961; Zygosity=Heterozygous (PubMed=12068308; CCLE; Cosmic-CLP).</p> <p>Mutation; HGNC; 11730; TERT; Simple; c.1-146C>T (c.250C>T) (C250T); Zygosity=Unspecified; Note=In promoter (PubMed=31068700).</p> <p>Mutation; HGNC; 11998; TP53; Simple; p.Pro322fs*24 (c.963_964insAA); Zygosity=Heterozygous (CCLE; Cosmic-CLP).</p>																
Genome ancestry	<p>Source: PubMed=30894373</p> <table border="1"> <thead> <tr> <th>Origin</th><th>% genome</th></tr> </thead> <tbody> <tr> <td>African</td><td>0</td></tr> <tr> <td>Native American</td><td>0.14</td></tr> <tr> <td>East Asian, North</td><td>1.84</td></tr> <tr> <td>East Asian, South</td><td>0</td></tr> <tr> <td>South Asian</td><td>2.44</td></tr> <tr> <td>European, North</td><td>63.98</td></tr> <tr> <td>European, South</td><td>31.59</td></tr> </tbody> </table>	Origin	% genome	African	0	Native American	0.14	East Asian, North	1.84	East Asian, South	0	South Asian	2.44	European, North	63.98	European, South	31.59
Origin	% genome																
African	0																
Native American	0.14																
East Asian, North	1.84																
East Asian, South	0																
South Asian	2.44																
European, North	63.98																
European, South	31.59																
Disease	<p>Melanoma (NCIt: C3224)</p> <p>Derived from metastatic site: Pelvic wall.</p>																
Species of origin	Homo sapiens (Human) (NCBI Taxonomy: 9606)																
Sex of cell	Female																
Age at sampling	69Y																
Category	Cancer cell line																
	<p>Source(s): Cosmic-CLP; ECACC; PubMed=25877200; PubMed=25926053</p> <p>Markers:</p> <table border="1"> <tbody> <tr> <td>Amelogenin</td><td>X</td></tr> <tr> <td>CSF1PO</td><td>11</td></tr> </tbody> </table>	Amelogenin	X	CSF1PO	11												
Amelogenin	X																
CSF1PO	11																

unique identifier

high quality annotations

exhaustively curated

rich metadata



Examples:

Antibody: RRID:AB_2178887

Cell Line: RRID:CVCL_0033

Organism: RRID:MGI:3840442

Digital Transformation – Prime Time for Biomedical Ontologies

Roche Data Commons, Reference Data Services & FAIR Principles

FAIR Maturity Indicators & FAIR Assessment

FAIR Maturity Indicators & FAIR APIs (JSON-LD)

Synergizing Biomedical Ontologies – Proliferation vs Convergence

Synergizing Biomedical Ontologies – Vision and Mission

Synergizing Biomedical Ontologies

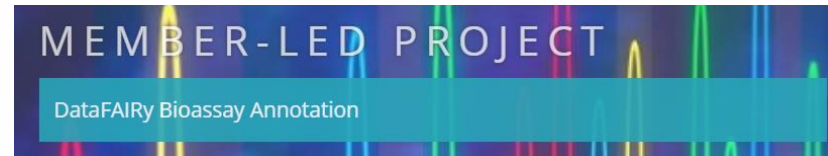
Vision and Mission

Open public-private semantic infrastructure of FAIR applications, services and data

- **Applications** based on FAIR data models and metadata models (eg DCAT for data catalogs).
Open interfaces for import of FAIR terminologies and FAIR data schemas.
Open interfaces for data access, model readout and semantically enables microservices
(connectors and integration layers are old-fashioned)
- **Services**: Smart APIs - semantically harmonized, FAIR APIs (data values and attributes as GUPRIs)
Data contracts based on JSON-LD supporting seamless data integration (no transformation)
- **Metadata and data** represented with semantic standards (GUPRIs)
Biomedical Ontologies converge: shared engineering principles, shared semantics, consolidated reference space (see Cellosaurus – less is more)
Public and commercial content provider adhere to FAIR principles and implement data standards

DataFAIRy BioAssay Annotation

Pre-competitive Data FAIRification



Day 2: Thursday, July 15

all times are in EDT

3:20pm – 3:40pm

CDD Annotator and Perspectives from the Data FAIRy Initiative

Samantha Jeschonek



Doing now what patients need next