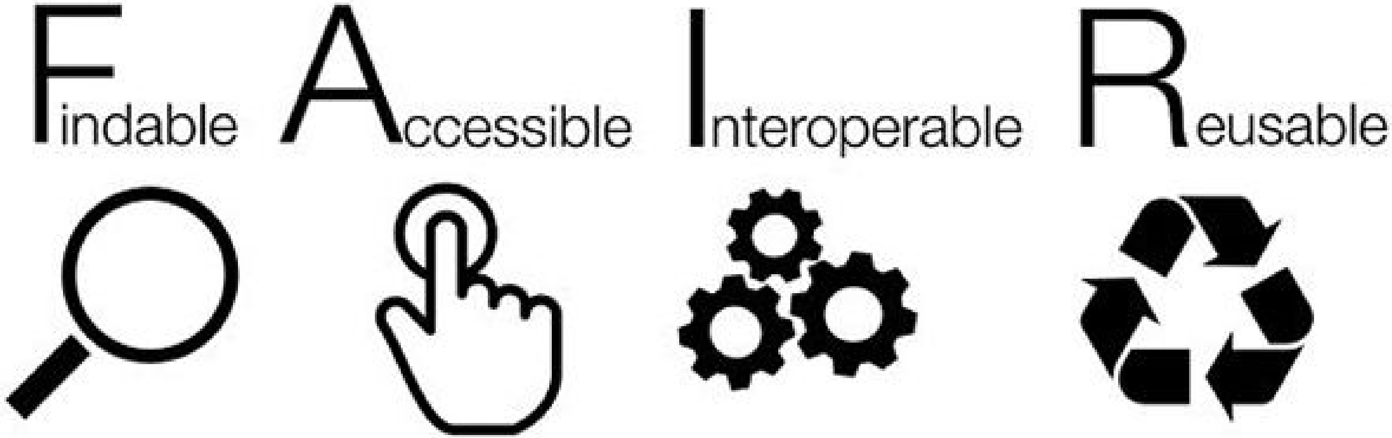


Recent Developments around the FAIR Principles

Mark D Wilkinson, CBGP UPM-INIA, 2017
ORCiD:0000-0001-6960-357X

http://tinyurl.com/FAIRDataPortugal

The Problem

64% of public data archives have flaws that partially or entirely prevent reuse*

>80% of an informatician’s time is spent finding, reformatting, and integrating data**

(only from the 36% of data that is “reusable” at all!)

* Dominique G. Roche , Loeske E. B. Kruuk, Robert Lanfear, Sandra A. Binning, November 10, 2015 https://doi.org/10.1371/journal.pbio.1002295

** G. Press, Cleaning Big Data: Most time-consuming, least enjoyable data science task, survey says, Forbes [Internet], 2016 Mar 23.

What is FAIR Data and why do we need it?

REUSELESS

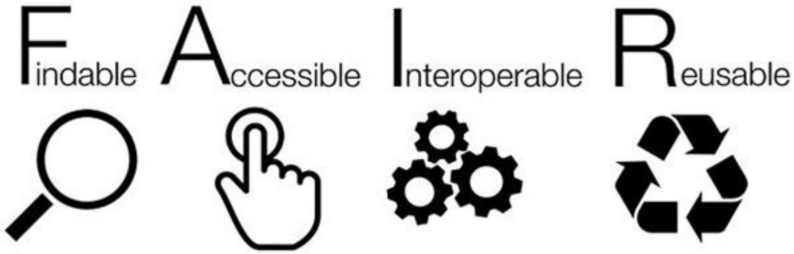
SCIENTIFIC DATA

Altmetric: 1060 Citations: 122

Comment | OPEN

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons



FINDABLE

→ Unambiguous identifiers supported by searchable metadata

ACCESSIBLE

→ Clearly-defined access protocol, preferably machine-actionable

INTEROPERABLE

→ Use shared vocabularies/ontologies in machine-accessible format

REUSABLE

→ Contextual information, allowing proper interpretation

→ Rich provenance information facilitating accurate citation

Things happened very quickly!

https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

2016

European Commission

Realising the European Open Science Cloud

First report and recommendations
of the Commission High Level Expert Group
on the European Open Science Cloud

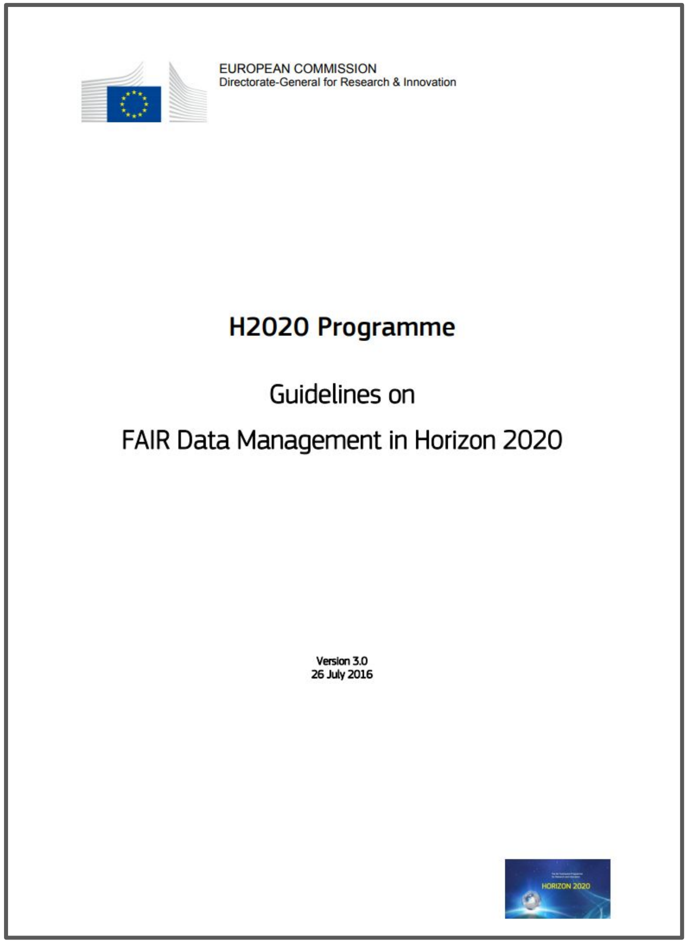
(page 18)

“Make adequate data stewardship mandatory for all research proposals.

...Horizon 2020, should only support projects that properly address Data Stewardship [and those] **that do not specify FAIR conditions...**

should not be eligible for funding.”

https://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf

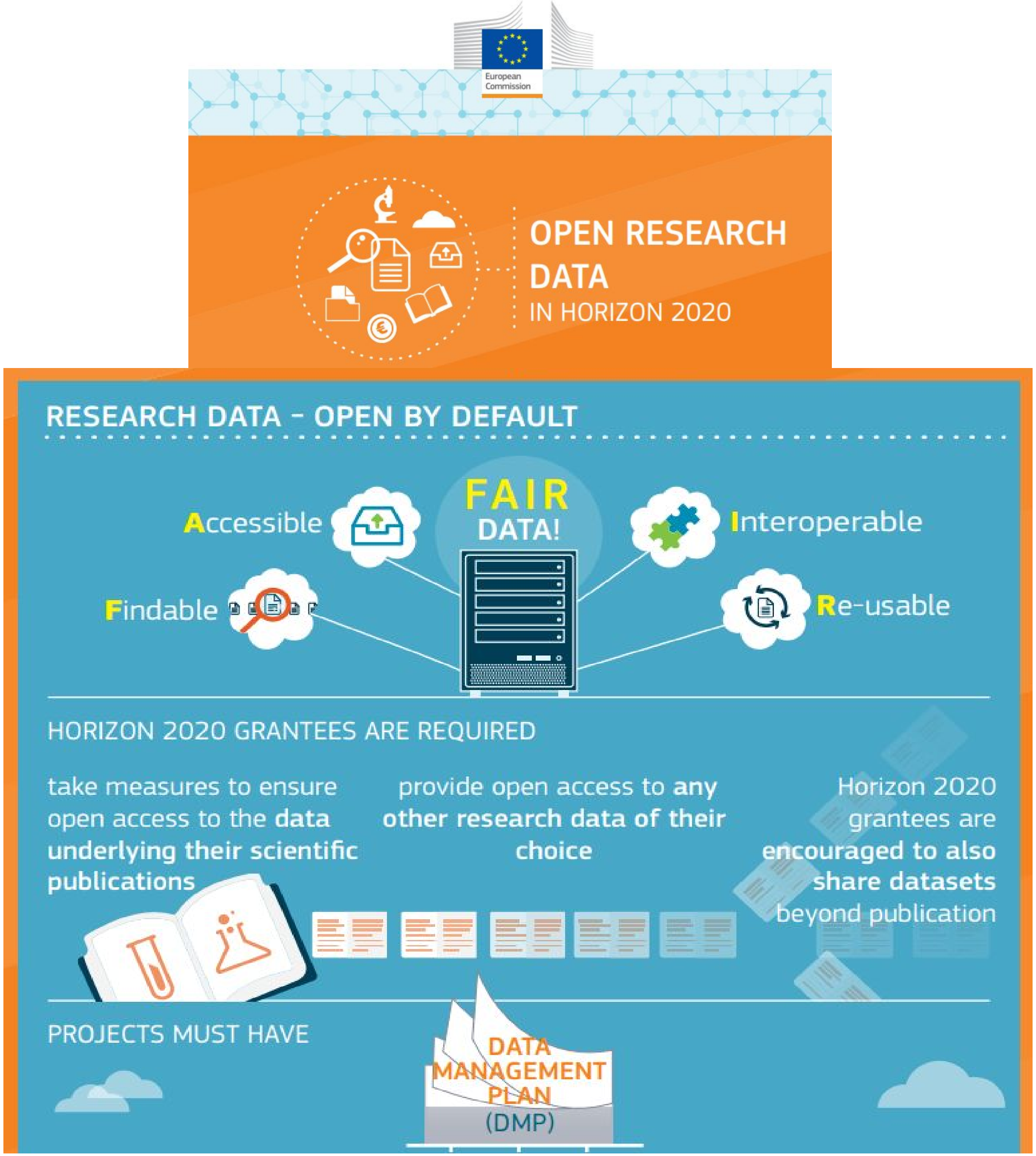


EUROPEAN COMMISSION
Directorate-General for Research & Innovation

H2020 Programme

Guidelines on
FAIR Data Management in Horizon 2020

Version 3.0
26 July 2016



European Commission

OPEN RESEARCH DATA
IN HORIZON 2020

RESEARCH DATA – OPEN BY DEFAULT

FAIR DATA!

Findable

Accessible

Interoperable

Re-usable

HORIZON 2020 GRANTEES ARE REQUIRED

take measures to ensure open access to the **data underlying their scientific publications**

provide open access to **any other research data of their choice**

Horizon 2020 grantees are **encouraged to also share datasets** beyond publication

PROJECTS MUST HAVE

DATA MANAGEMENT PLAN (DMP)

Synopsis

In the future

if you’re not FAIR

You don’t get EC funding

Research Data Canada (RDC) Update

Mark Leggott | Executive Director | RDC

canarie.ca | @canarie_inc

50

Hosted by canarie

National Institutes of Health
Office of Strategic Coordination - The Common Fund

Common Fund Programs | Common Fund Research Funding | News & Media | Common Fund High

Big Data to Knowledge

Common Fund » Common Fund Programs » Big Data to Knowledge

Big Data to Knowledge

For Public
- Highlights
- Press Releases
- In the News

For Researchers
- Funded Research
- Funding Opportunities
- Cloud Credits Model
- Data Commons Pilot Phase
  - Frequently Asked Questions
- Publication Search
- Meetings and Workshops

Using LINCS and other Big Data to Find New Cancer Treatments

Read More

Program Snapshot

As biomedical tools and technologies rapidly improve, researchers are producing and analyzing a rapidly increasing amount of complex biological data called "big data." The Big Data to Knowledge (BD2K) program, was launched in 2014 to facilitate broad use of biomedical big data, develop and disseminate analysis methods and software, enhance training relevant for large-scale data analysis, and establish centers of excellence for biomedical big data. The BD2K Program also supported initial efforts toward making data sets "FAIR" Findable, Accessible, Interoperable, and Reusable. Learn more about the FAIR principles.

African Open Science Platform Project and Research Data Alliance Workshop
Association of African Universities Conference
Palm Royal Beach Hotel, Accra, Ghana
8 June 2017

Open FAIR Data and Open Science:
Developing Partnerships, Strategies, Policies

Simon Hodson, Executive Director, CODATA
www.codata.org

European Commission - Statement

G20 Leaders' Communique Hangzhou Summit

Hangzhou, 5 September 2016

"...we support appropriate efforts to promote open science and facilitate appropriate access to publicly funded research results on findable, accessible, interoperable and reusable (FAIR) principles."

To be Findable:

F1. (meta)data are assigned a globally unique and persistent identifier

To be Findable:

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata

To be Findable:

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata

F3. metadata explicitly include the identifier of the data it describes

To be Findable:

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata

F3. metadata 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

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

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

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.

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

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

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

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

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

What does FAIR Data “look like”?

“Skunkworks”

Task: Build a prototype

“Skunkworks”

Task: Build a prototype

(build “a” solution
not “the” solution!)

Skunkworks Participants

- Mark Wilkinson
- Michel Dumontier
- Barend Mons
- Tim Clark
- Jun Zhao
- Paolo Ciccarese
- Paul Groth
- Erik van Mulligen
- Luiz Olavo Bonino da Silva Santos
- Matthew Gamble
- Carole Goble
- Joël Kuiper
- Morris Swertz
- Erik Schultes

- Erik Schultes
- Mercè Crosas
- Adrian Garcia
- Philip Durbin
- Jeffrey Grethe
- Katy Wolstencroft
- Sudeshna Das
- M. Emily Merrill

The Hourglass Concept

We want a large ecosystem of apps that use FAIR Data

The Hourglass Concept

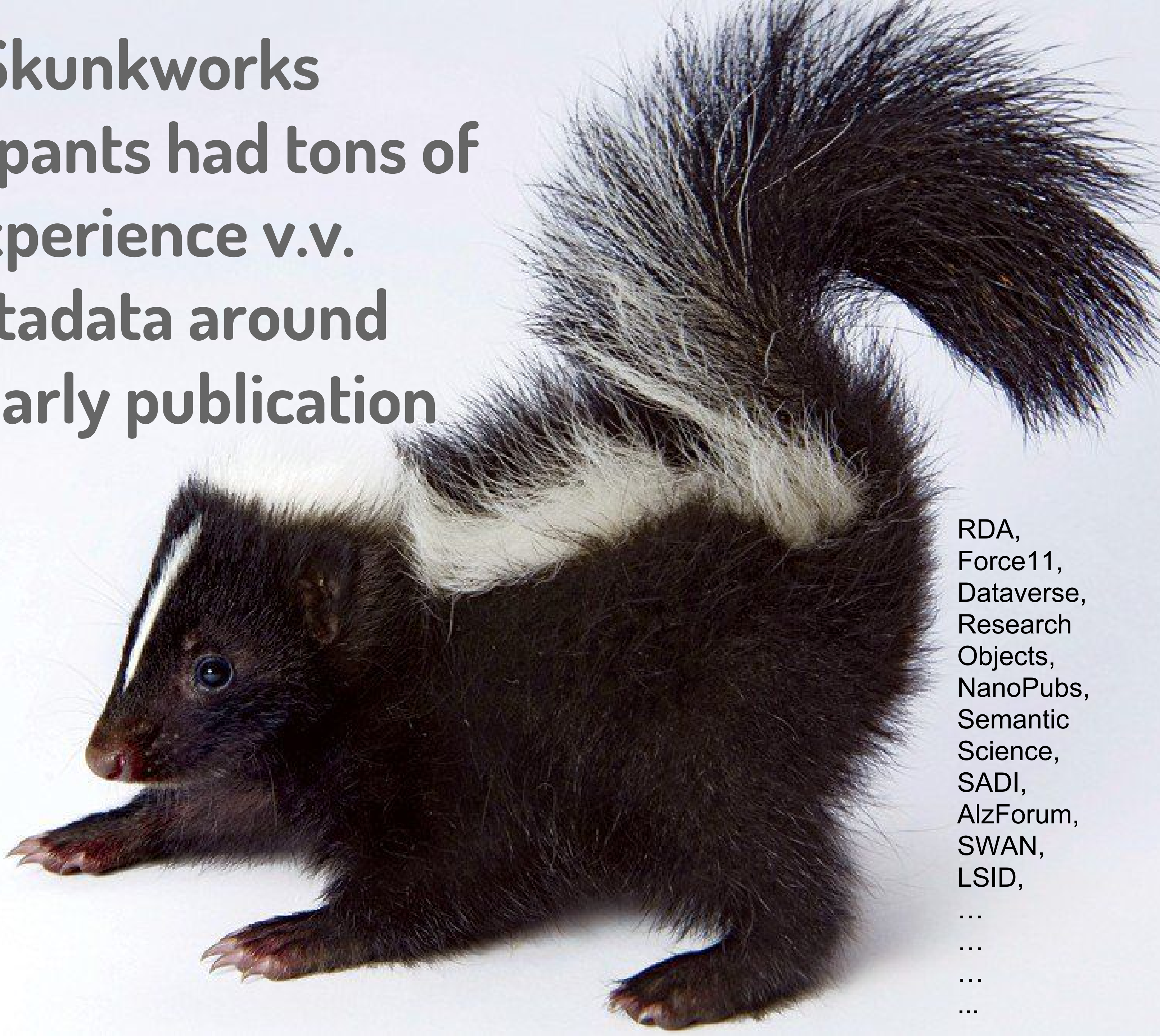
We want to support a wide range of source providers

The Hourglass Concept

The FAIR solution between them must be THIN!

Skunkworks participants had tons of experience v.v. metadata around scholarly publication

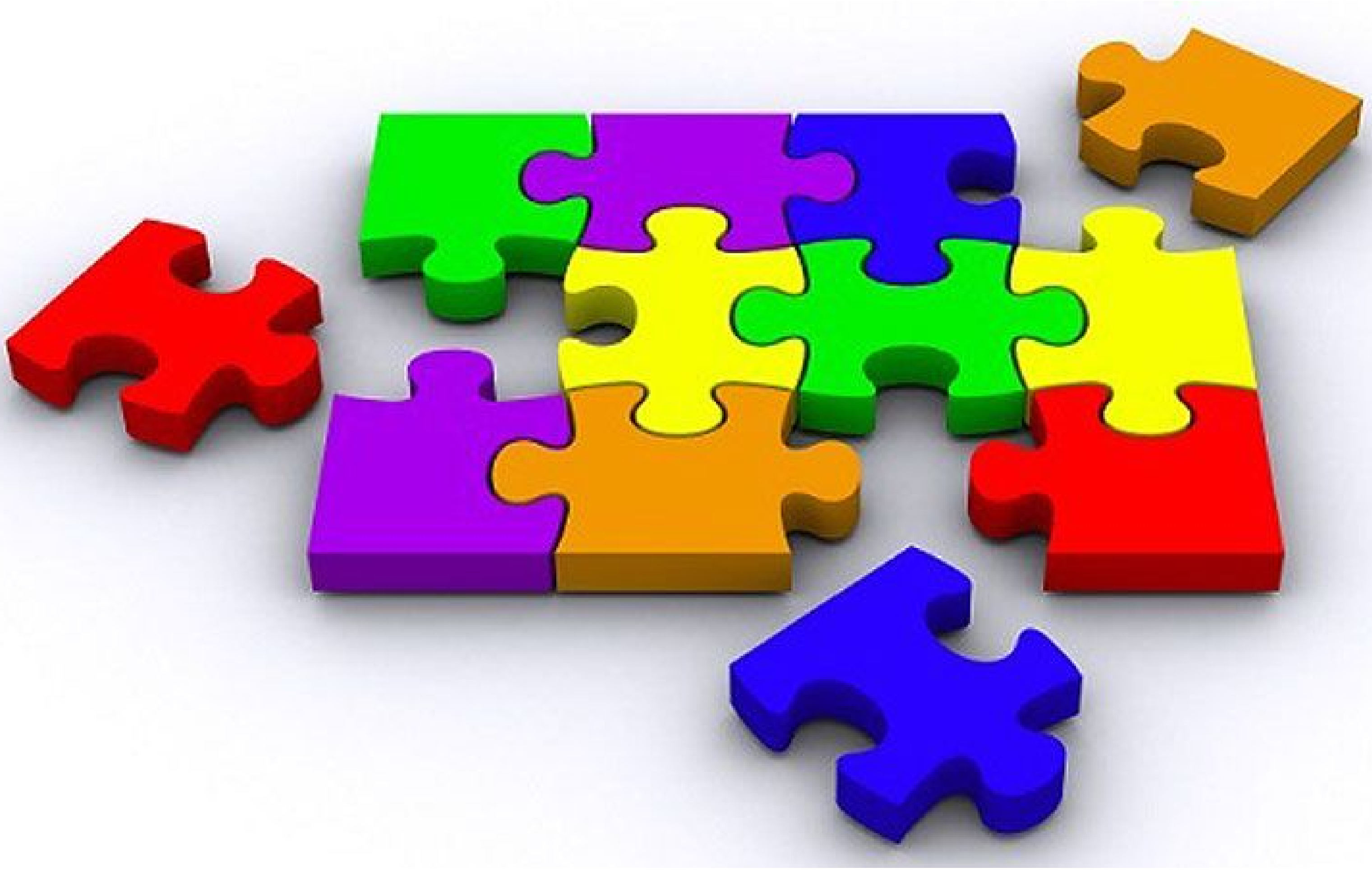
Skunkworks participants had tons of experience v.v. metadata around scholarly publication



RDA,
Force11,
Dataverse,
Research Objects,
NanoPubs,
Semantic Science,
SADI,
AlzForum,
SWAN,
LSID,
…
…
…
...

There was very little disagreement
about F,
about A,
or about R

The “I” is the big problem



The “I” is the big problem

Interoperability is Hard!!

Keeping the history brief

A series of teleconferences led to the concept of putting metadata into an iterative set of ~identical “containers”

Skunkworks Hackathons

The “containers of containers of containers” idea was elaborated by the belief that we should also **reject any solution that required a new API**

ProgrammableWeb.com already catalogues **>16,000** different Web APIs

Skunkworks Hackathons

The “containers of containers of containers” idea was elaborated by the belief that we should also reject any solution that required a new API

ProgrammableWeb.com already catalogues
>16,000 different Web APIs

APIs DO NOT MAKE YOU INTEROPERABLE!

Skunkworks Hackathons

The “containers of containers of containers” idea was elaborated by the belief that we should also reject any solution that required a new API

{ REST }

Skunkworks Hackathons

Are there existing standards that are

{ REST }

And have the properties of

?

W3C Recommendation

Linked Data Platform 1.0

W3C Recommendation 26 February 2015

This version:
  http://www.w3.org/TR/2015/REC-ldp-20150226/
Latest published version:
  http://www.w3.org/TR/ldp/
Latest editor's draft:
  http://www.w3.org/2012/ldp/hg/ldp.html
Test suite:
  https://dvcs.w3.org/hg/ldpwg/raw-file/default/tests/ldp-testsuite.html
Implementation report:
  https://dvcs.w3.org/hg/ldpwg/raw-file/default/tests/reports/ldp.html
Previous version:
  http://www.w3.org/TR/2014/PR-ldp-20141216/

LDP

Useful Features

I Uses machine-accessible standards and representations, following a ~~REST paradigm

I + R Defines the concept of a “Container” - a machine-actionable way to represent repositories, data deposits, data files, data points, and their metadata

F + A Defines HTTP-resolvable URIs for each of these containers

I Uses a widely accepted standard (DCAT) to relate metadata to data → machine-actionable data mining

The FAIR Accessor

In incremental detail

What can we describe with FAIR Accessors?

FAIR Accessors provide a machine-actionable, structured,

REST-oriented way to publish Metadata

about a wide range of scholarly “entities”

What can we describe with FAIR Accessors?

Warehouses (e.g. EBI)

Databases (e.g. UniProt)

Repositories (e.g. Zenodo, INRA-URGI Wheat Repo, UniProt)

Datasets (e.g. output from a workflow)

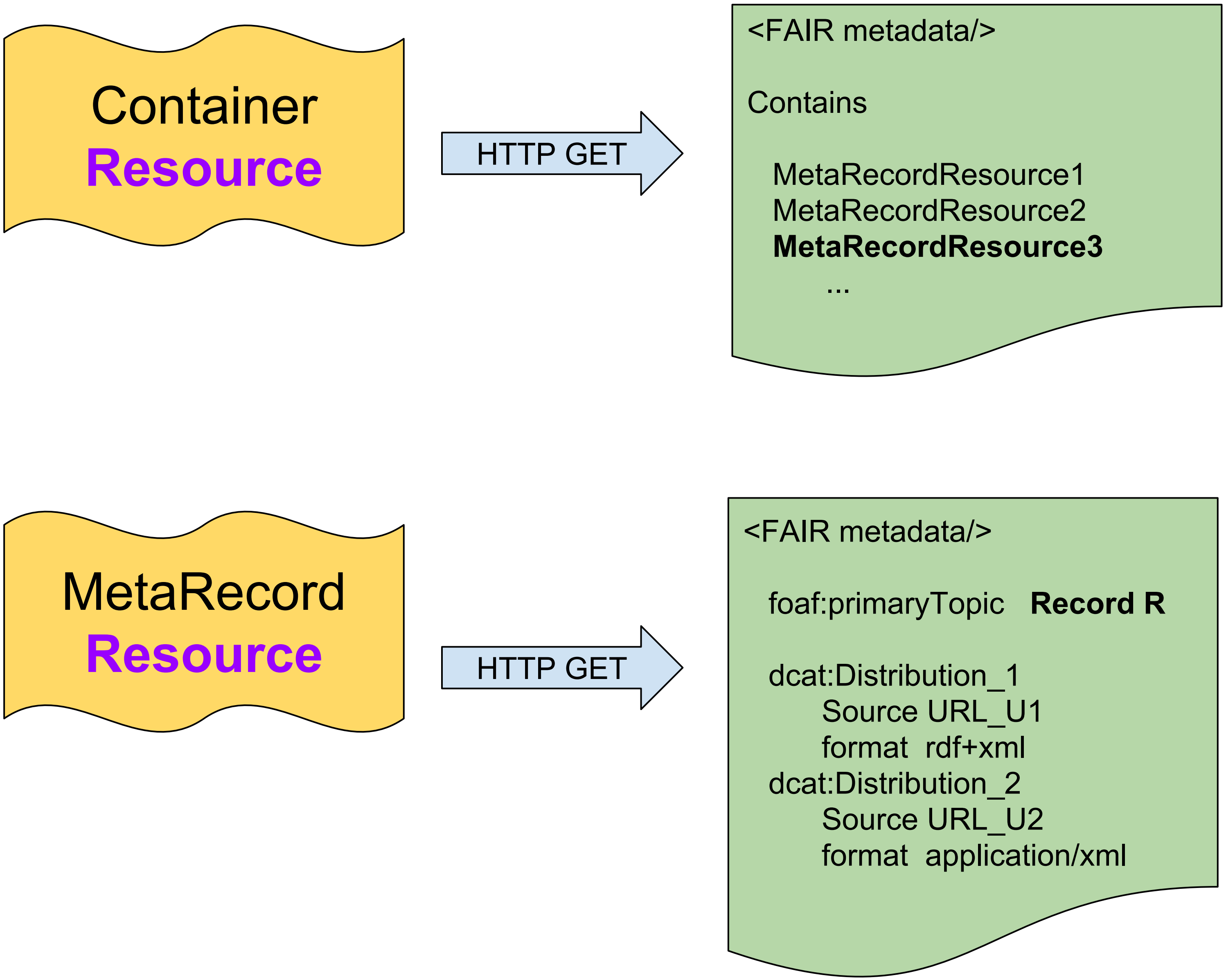
Research Objects (data a/o workflow a/o results a/o publications)

Data "slices" (e.g. the result of a database query)

Data Records (e.g. image, excel file, patient clinical record)

Other...

What does a FAIR Accessor “look like”?



Container
Resource

HTTP GET

<FAIR metadata/>

Contains

MetaRecordResource1
MetaRecordResource2
MetaRecordResource3
...

MetaRecord
Resource

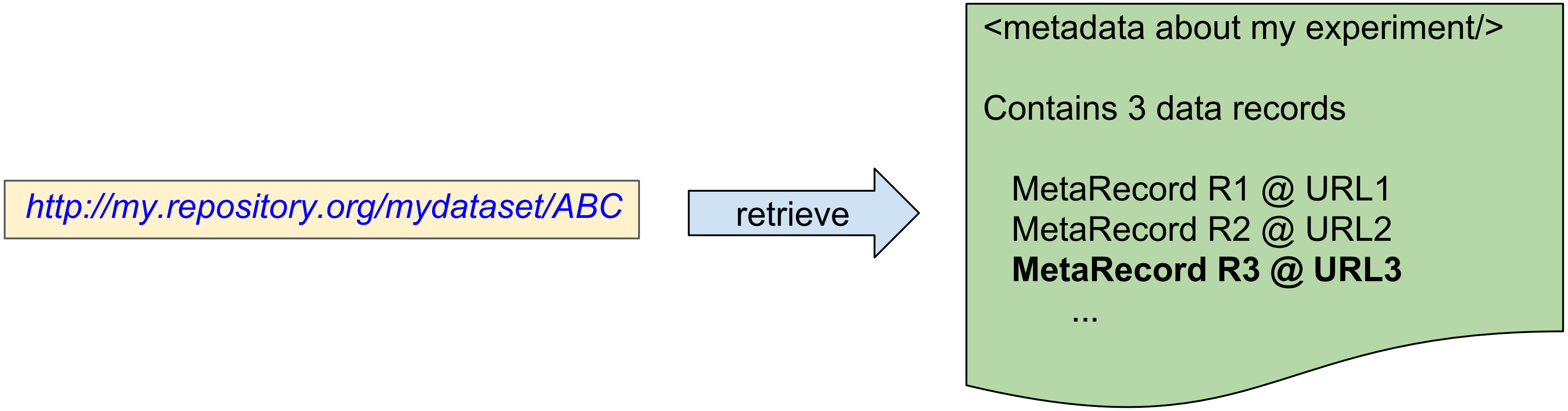
HTTP GET

<FAIR metadata/>

foaf:primaryTopic **Record R**

dcat:Distribution_1
    Source URL_U1
    format rdf+xml
dcat:Distribution_2
    Source URL_U2
    format application/xml

The FAIR Accessor



http://my.repository.org/mydataset/ABC

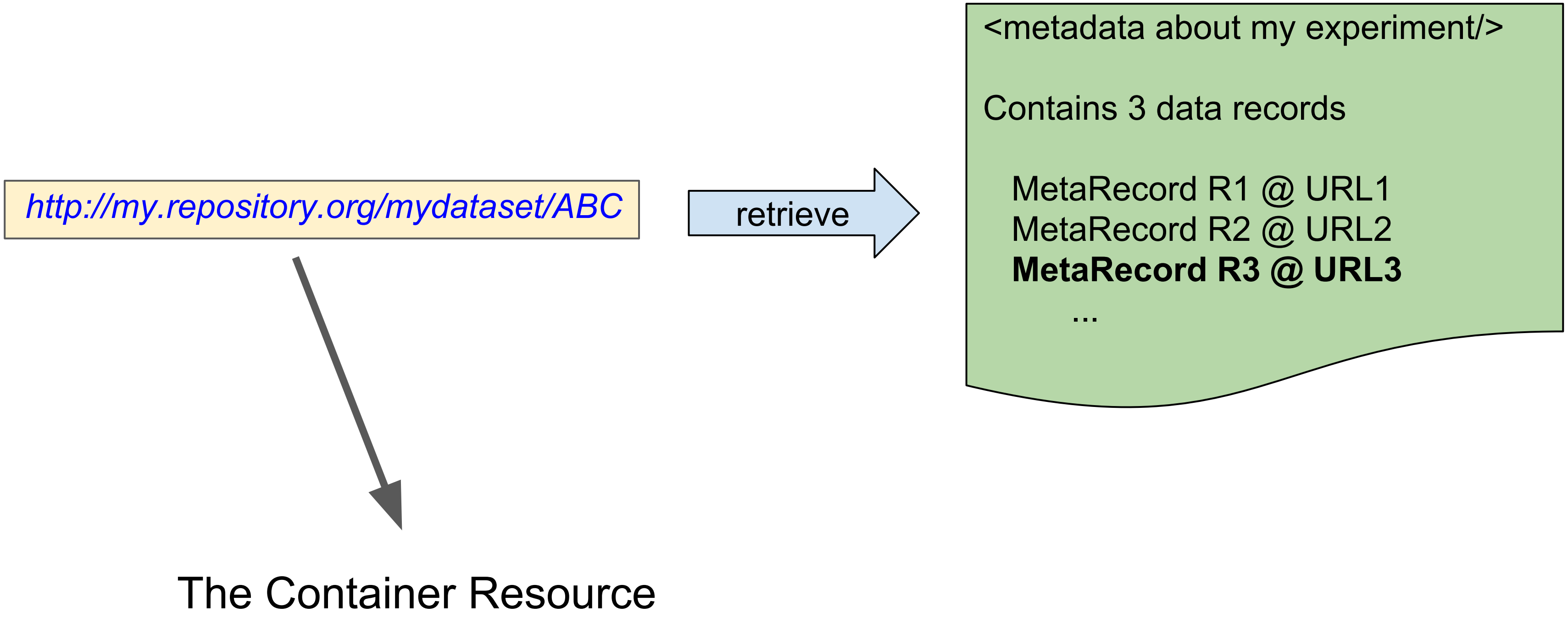
retrieve

<metadata about my experiment/>

Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

The FAIR Accessor



http://my.repository.org/mydataset/ABC

retrieve

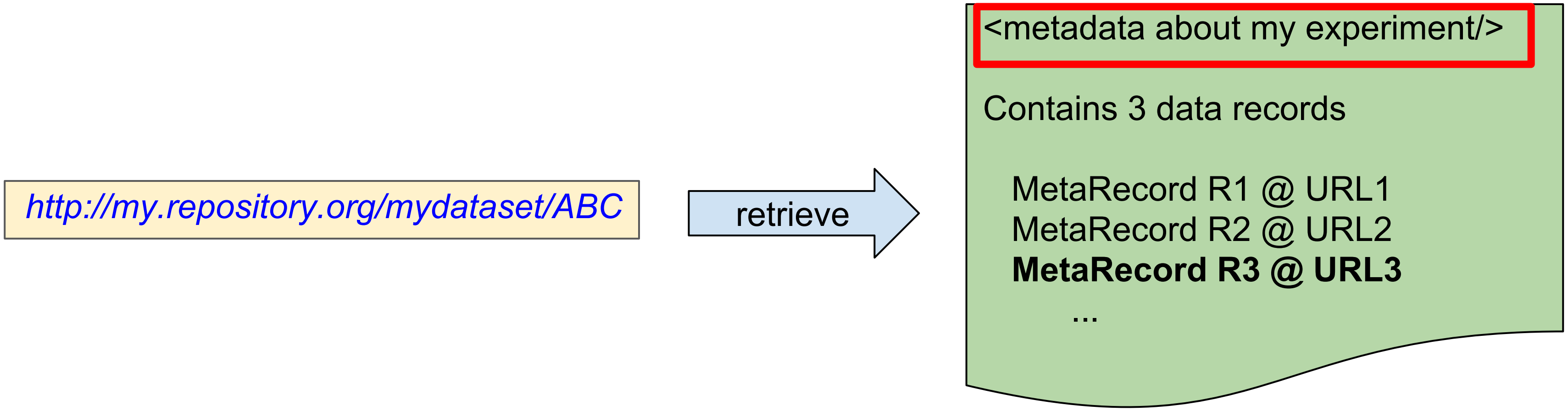
<metadata about my experiment/>

Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

The Container Resource

The FAIR Accessor



http://my.repository.org/mydataset/ABC

retrieve

<metadata about my experiment/>

Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

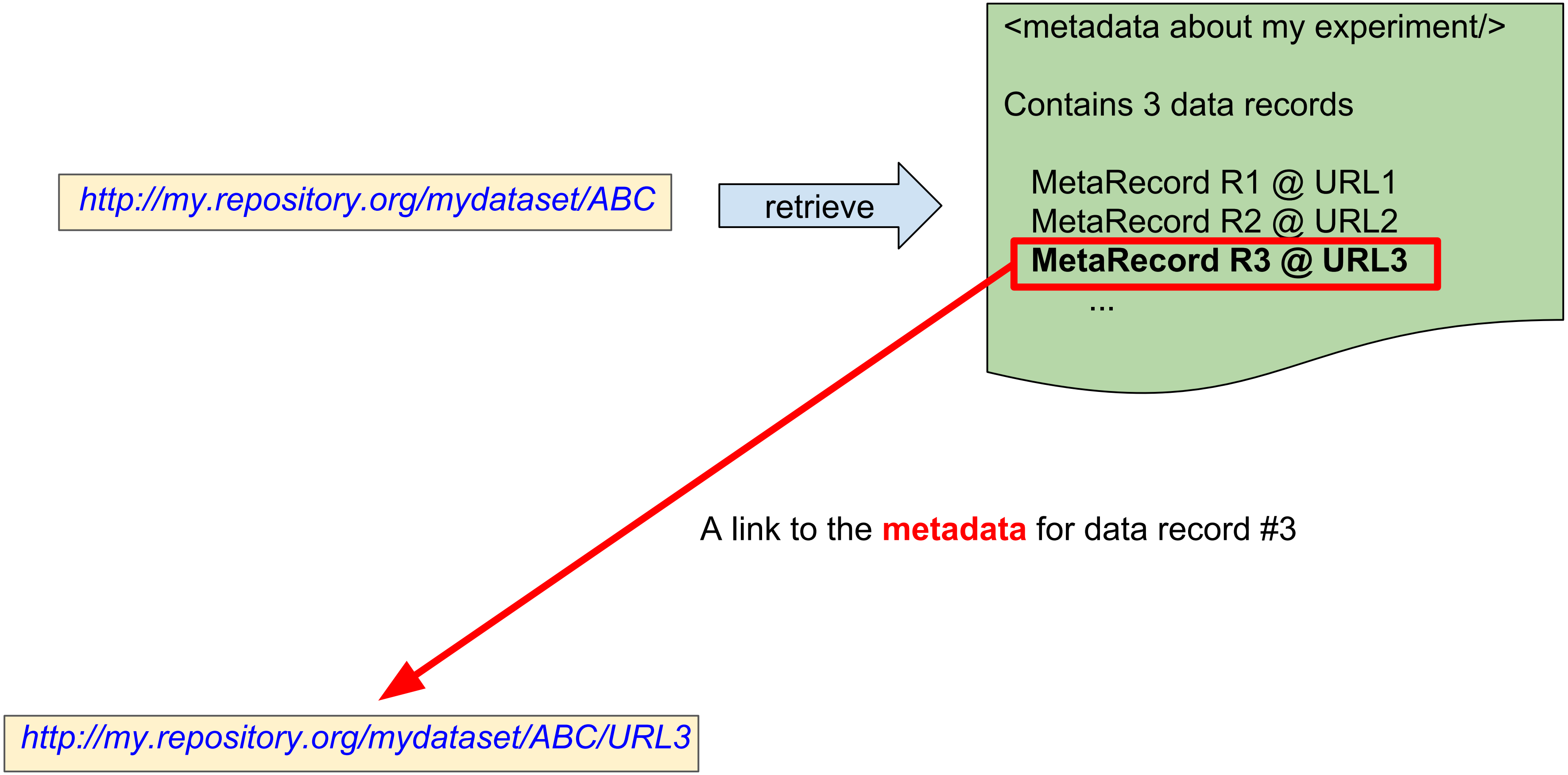
The container “Metadata” describes the content of the container

For example, when describing an experiment, it is information like:

What was the hypothesis? What technologies did I use?
What protocols? What organism/strain?
Who to cite

(i.e. the materials and methods… written in a formal,
machine-readable manner)

The FAIR Accessor



http://my.repository.org/mydataset/ABC

retrieve

<metadata about my experiment/>

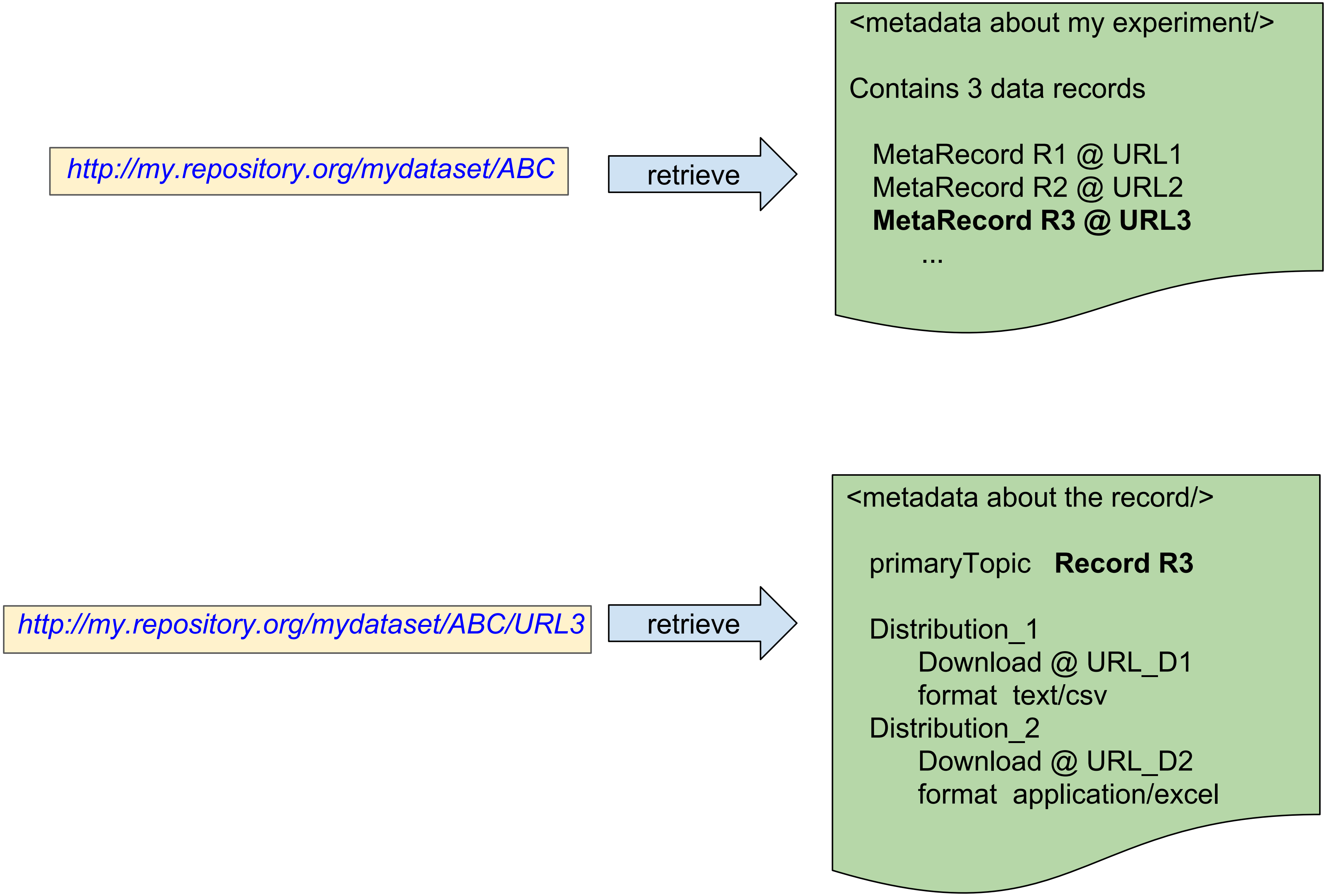
Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

A link to the **metadata** for data record #3

http://my.repository.org/mydataset/ABC/URL3

The FAIR Accessor



http://my.repository.org/mydataset/ABC

retrieve

<metadata about my experiment/>

Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

http://my.repository.org/mydataset/ABC/URL3

retrieve

<metadata about the record/>

primaryTopic **Record R3**

Distribution_1
Download @ URL_D1
format text/csv
Distribution_2
Download @ URL_D2
format application/excel

The FAIR Accessor

http://my.repository.org/mydataset/ABC

retrieve

<metadata about my experiment/>

Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

http://my.repository.org/mydataset/ABC/URL3

retrieve

<metadata about the record/>

primaryTopic **Record R3**

Distribution_1
  Download @ URL_D1
  format text/csv
Distribution_2
  Download @ URL_D2
  format application/excel

e.g. What machine/device? What filters/cutoffs?
What algorithms with what settings/parameters?
Data Usage License

The FAIR Accessor

http://my.repository.org/mydataset/ABC

retrieve

<metadata about my experiment/>

Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

http://my.repository.org/mydataset/ABC/URL3

retrieve

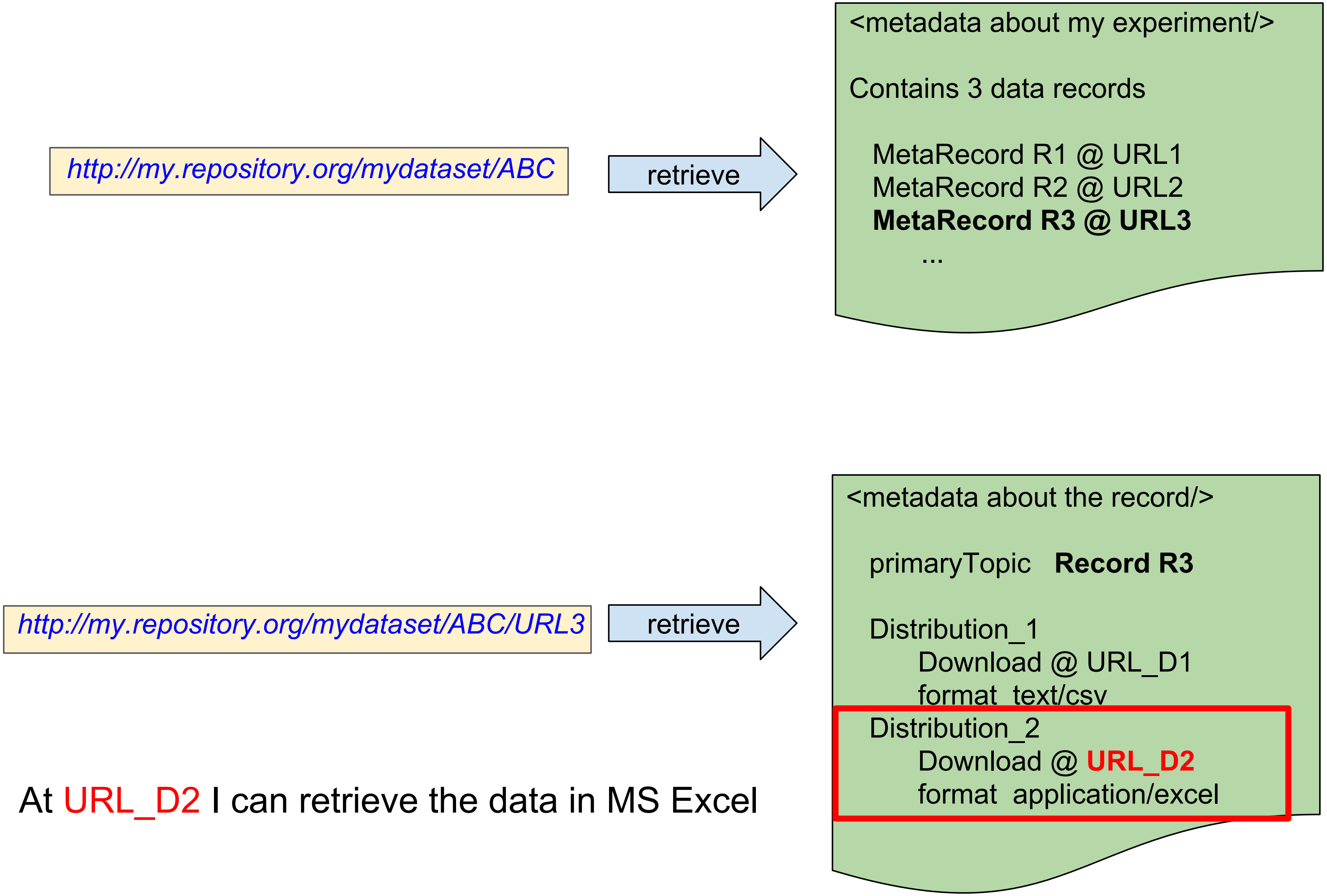
<metadata about the record/>

primaryTopic **Record R3**

Distribution_1
Download @ **URL_D1**
format text/csv
Distribution_2
Download @ URL_D2
format application/excel

At URL_D1 I can retrieve the data in CSV format

The FAIR Accessor



http://my.repository.org/mydataset/ABC

retrieve

<metadata about my experiment/>

Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

http://my.repository.org/mydataset/ABC/URL3

retrieve

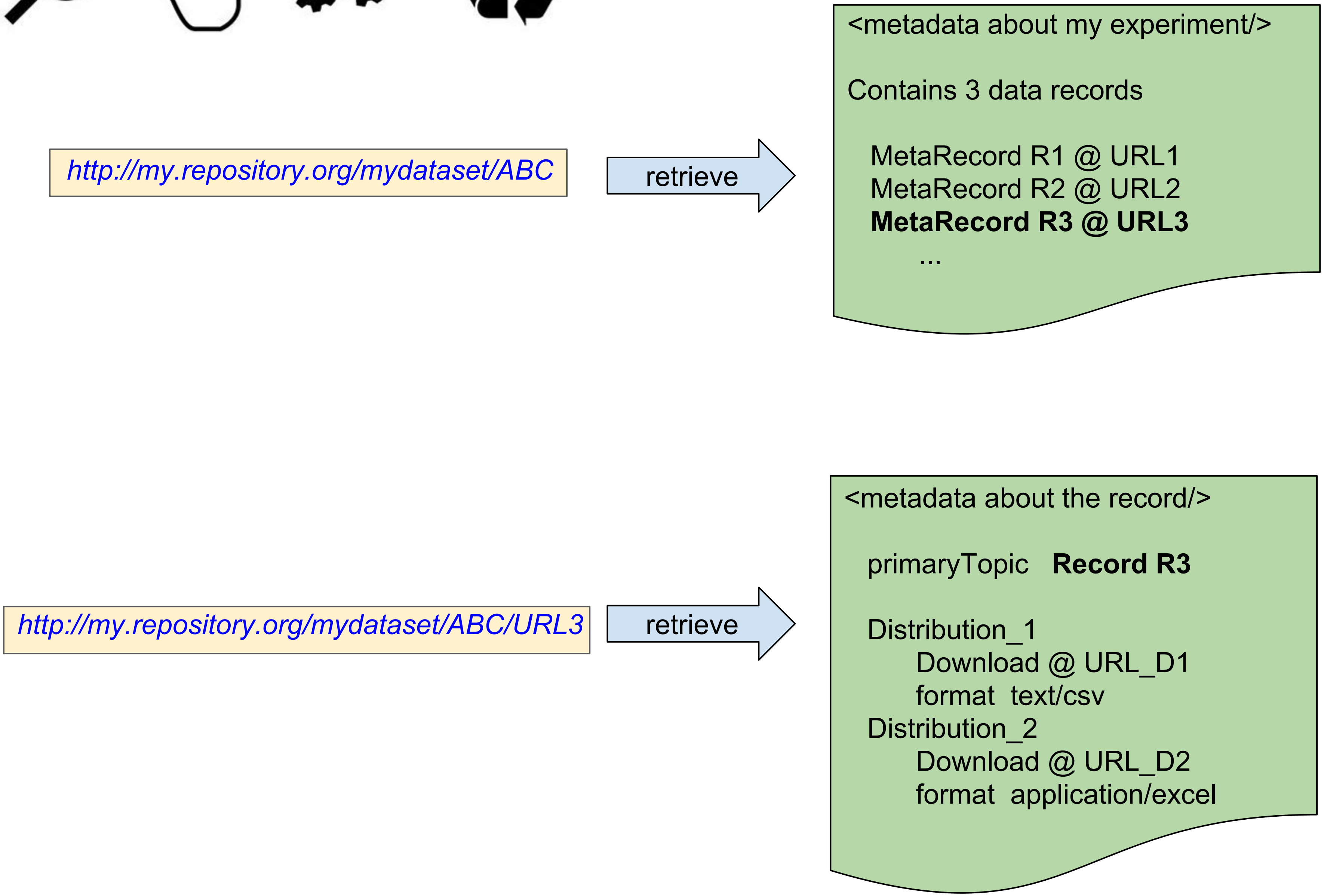
<metadata about the record/>

primaryTopic **Record R3**

Distribution_1
Download @ URL_D1
format text/csv
Distribution_2
Download @ **URL_D2**
format application/excel

At URL_D2 I can retrieve the data in MS Excel

Findable **A**ccessible **I**nteroperable **R**eusable



http://my.repository.org/mydataset/ABC

retrieve

<metadata about my experiment/>

Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

http://my.repository.org/mydataset/ABC/URL3

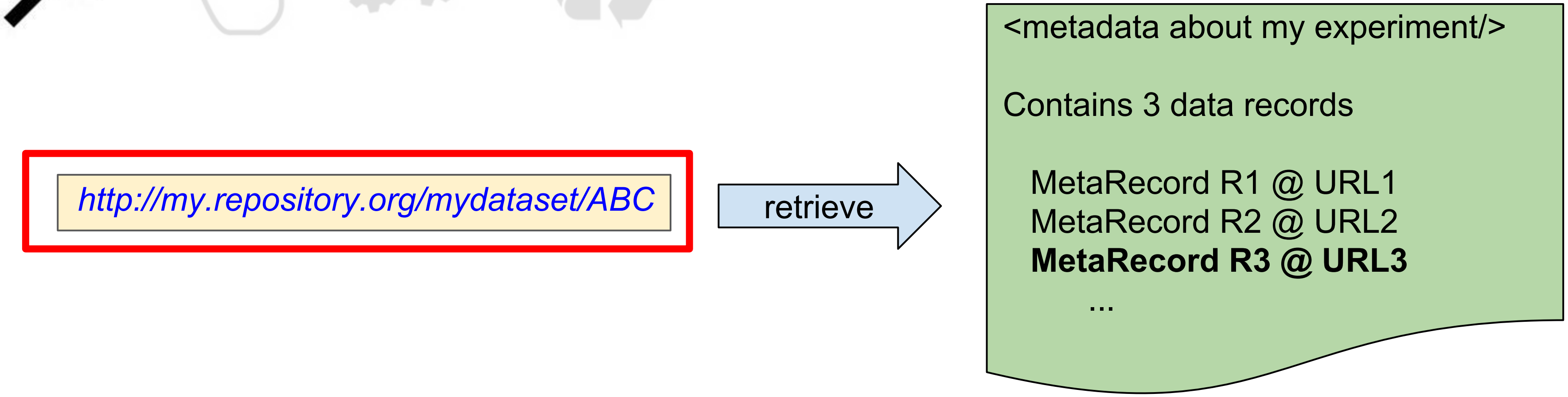
retrieve

<metadata about the record/>

primaryTopic **Record R3**

Distribution_1
  Download @ URL_D1
  format text/csv
Distribution_2
  Download @ URL_D2
  format application/excel

Findable Accessible Interoperable Reusable



http://my.repository.org/mydataset/ABC

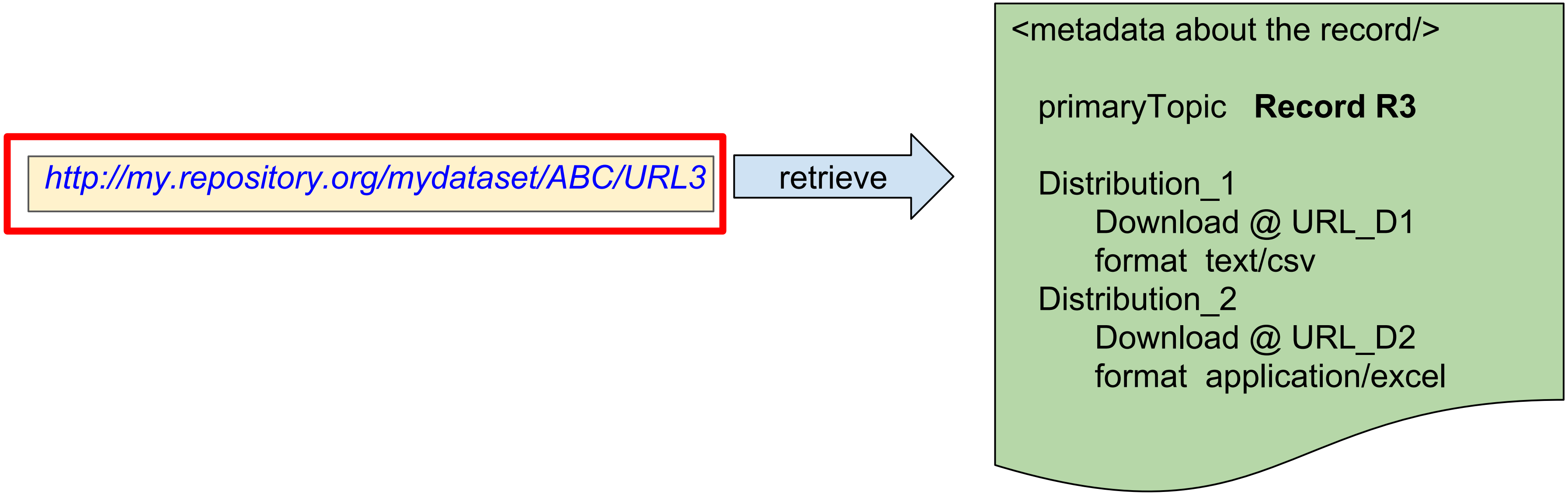
retrieve

<metadata about my experiment/>

Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

Globally-unique Identifiers for data and metadata



http://my.repository.org/mydataset/ABC/URL3

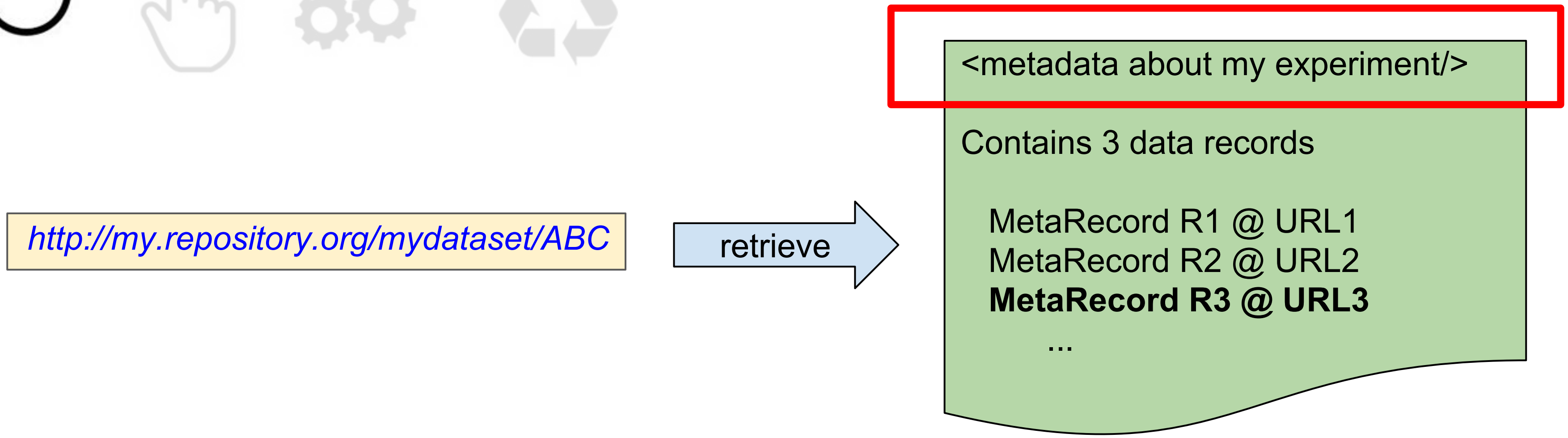
retrieve

<metadata about the record/>

primaryTopic **Record R3**

Distribution_1
Download @ URL_D1
format text/csv
Distribution_2
Download @ URL_D2
format application/excel

Findable Accessible Interoperable Reusable



http://my.repository.org/mydataset/ABC

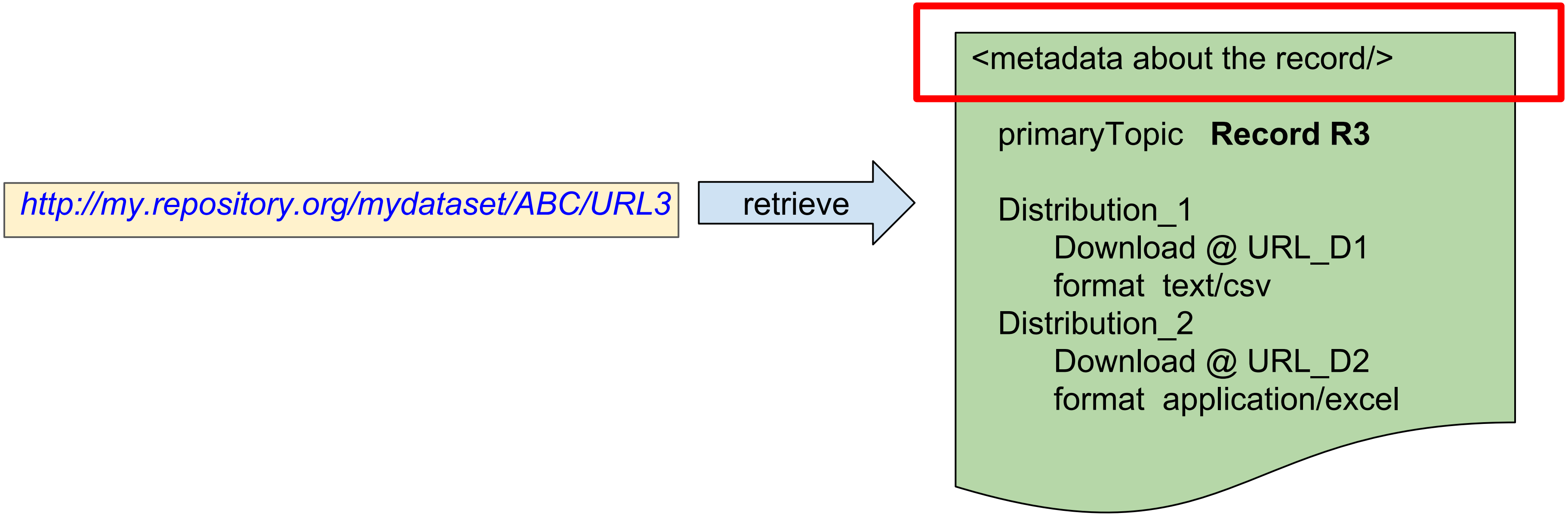
retrieve

<metadata about my experiment/>

Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

Metadata that can be searched



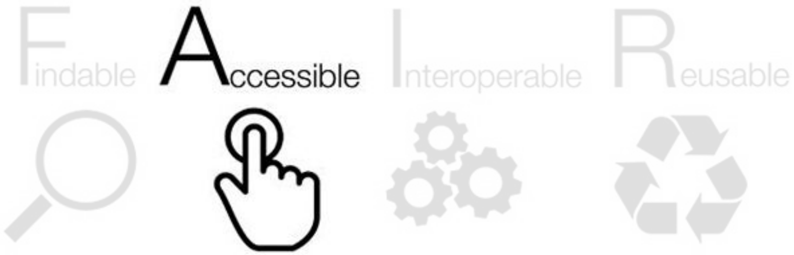
http://my.repository.org/mydataset/ABC/URL3

retrieve

<metadata about the record/>

primaryTopic **Record R3**

Distribution_1
  Download @ URL_D1
  format text/csv
Distribution_2
  Download @ URL_D2
  format application/excel



http://my.repository.org/mydataset/ABC

retrieve

<metadata about my experiment/>

Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

Clear access protocol (in this case, simply the Web)

http://my.repository.org/mydataset/ABC/URL3

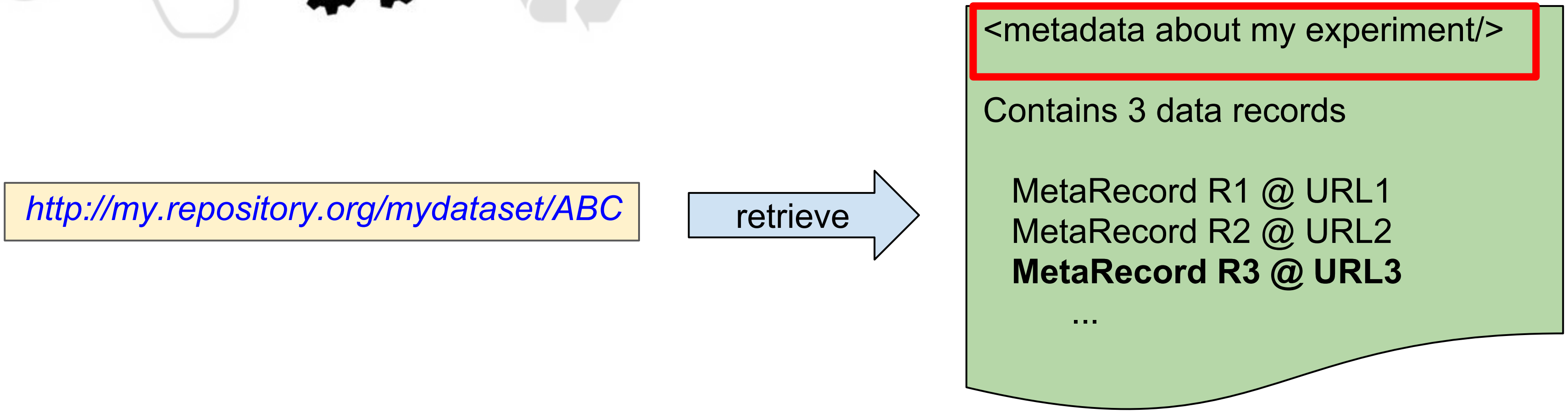
retrieve

<metadata about the record/>

primaryTopic **Record R3**

Distribution_1
    Download @ URL_D1
    format text/csv
Distribution_2
    Download @ URL_D2
    format application/excel

Findable Accessible **Interoperable** Reusable



http://my.repository.org/mydataset/ABC

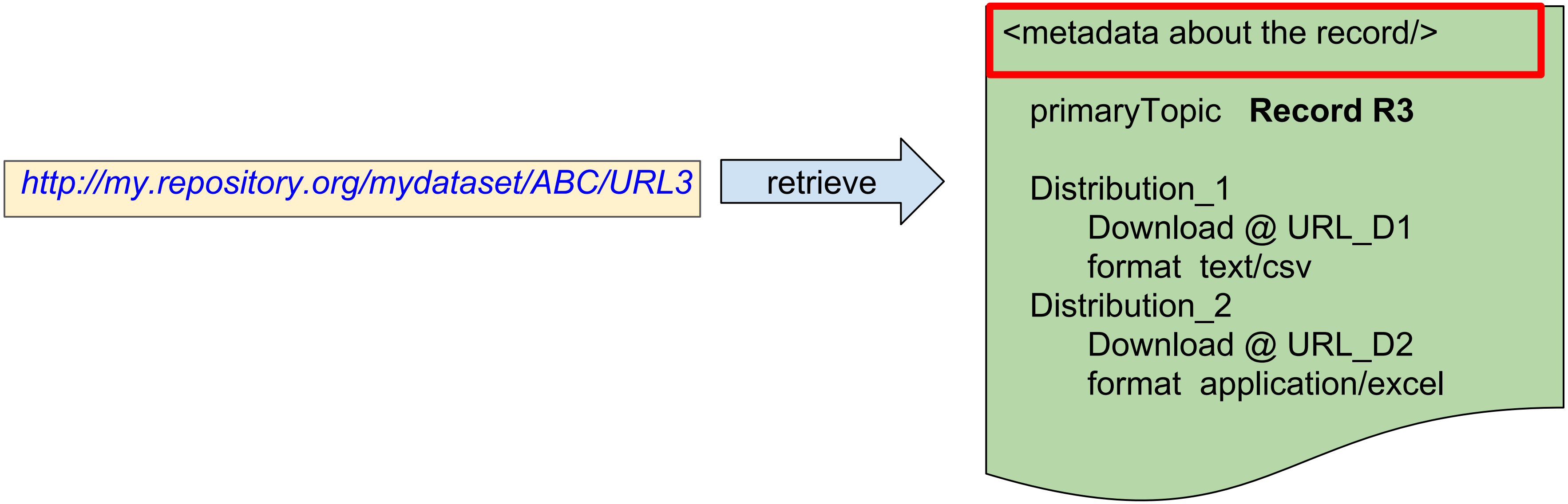
retrieve

<metadata about my experiment/>

Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

Metadata uses a machine-readable syntax and ontologies
(in our prototype, we used RDF + OWL)



http://my.repository.org/mydataset/ABC/URL3

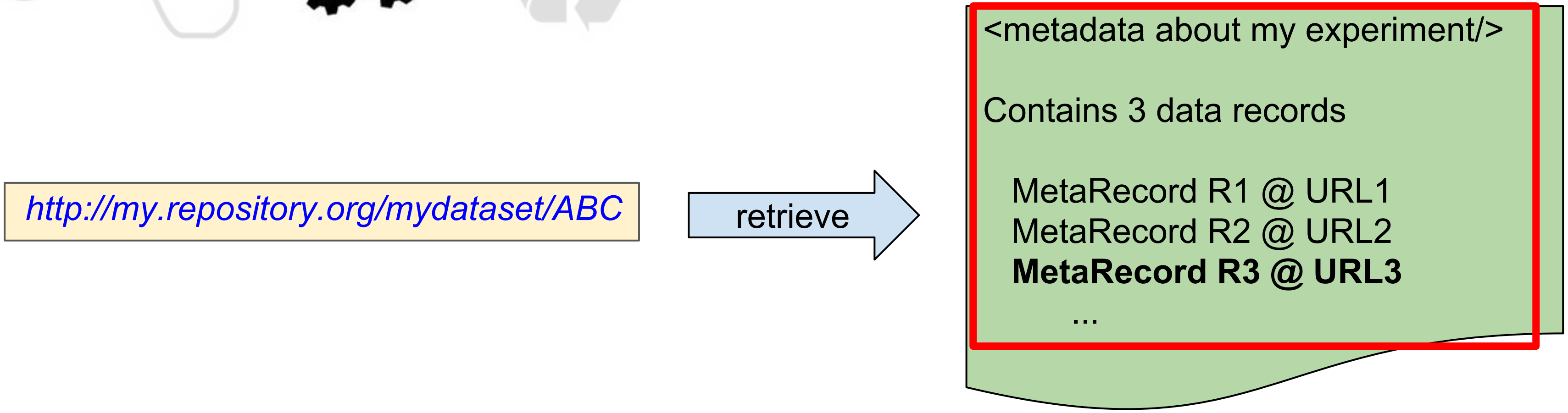
retrieve

<metadata about the record/>

primaryTopic **Record R3**

Distribution_1
    Download @ URL_D1
    format text/csv
Distribution_2
    Download @ URL_D2
    format application/excel

Findable Accessible **Interoperable** Reusable



http://my.repository.org/mydataset/ABC

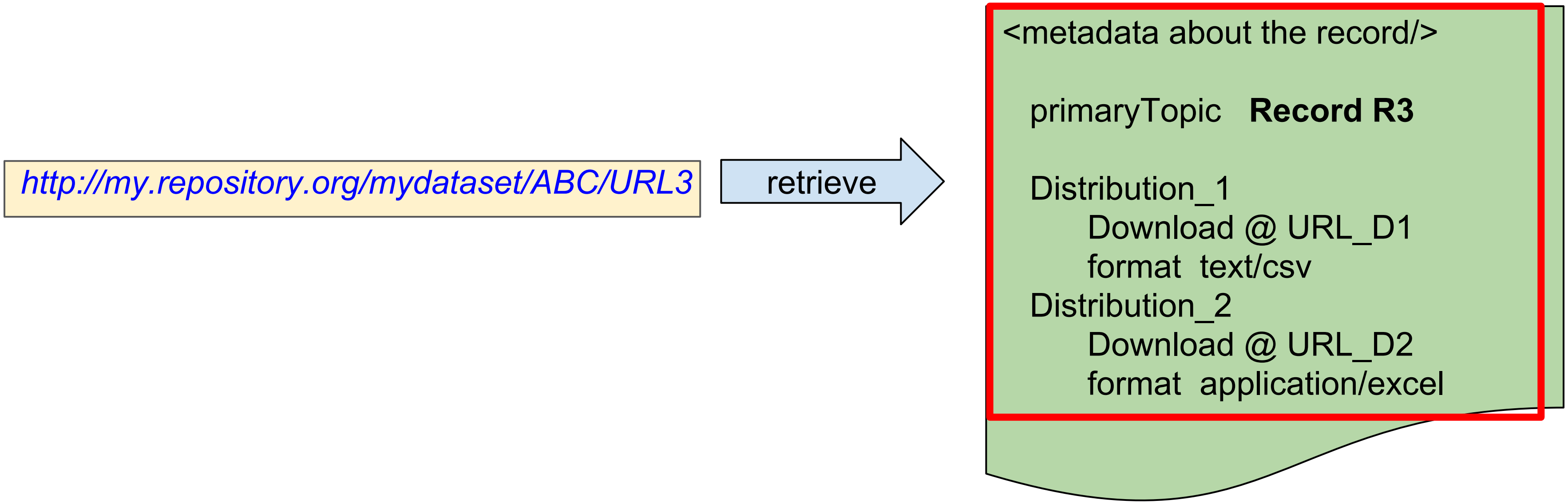
retrieve

<metadata about my experiment/>

Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

These data structures conform to a global standard
Called "DCAT" (the Data Catalog standard)
A lot of software can interpret these data structures.



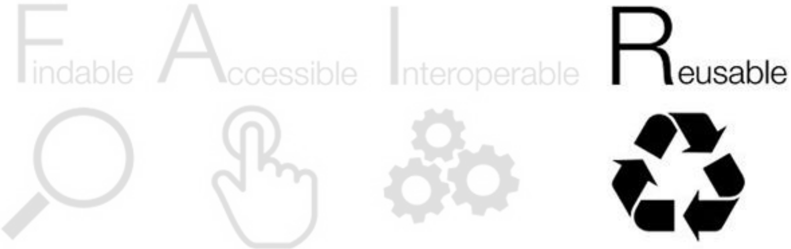
http://my.repository.org/mydataset/ABC/URL3

retrieve

<metadata about the record/>

primaryTopic **Record R3**

Distribution_1
Download @ URL_D1
format text/csv
Distribution_2
Download @ URL_D2
format application/excel



http://my.repository.org/mydataset/ABC

retrieve

<metadata about my experiment/>

Contains 3 data records

MetaRecord R1 @ URL1
MetaRecord R2 @ URL2
MetaRecord R3 @ URL3
...

Metadata includes:

- **Contextual information**
- **License**
- **Citation**
- **Provenance**

http://my.repository.org/mydataset/ABC/URL3

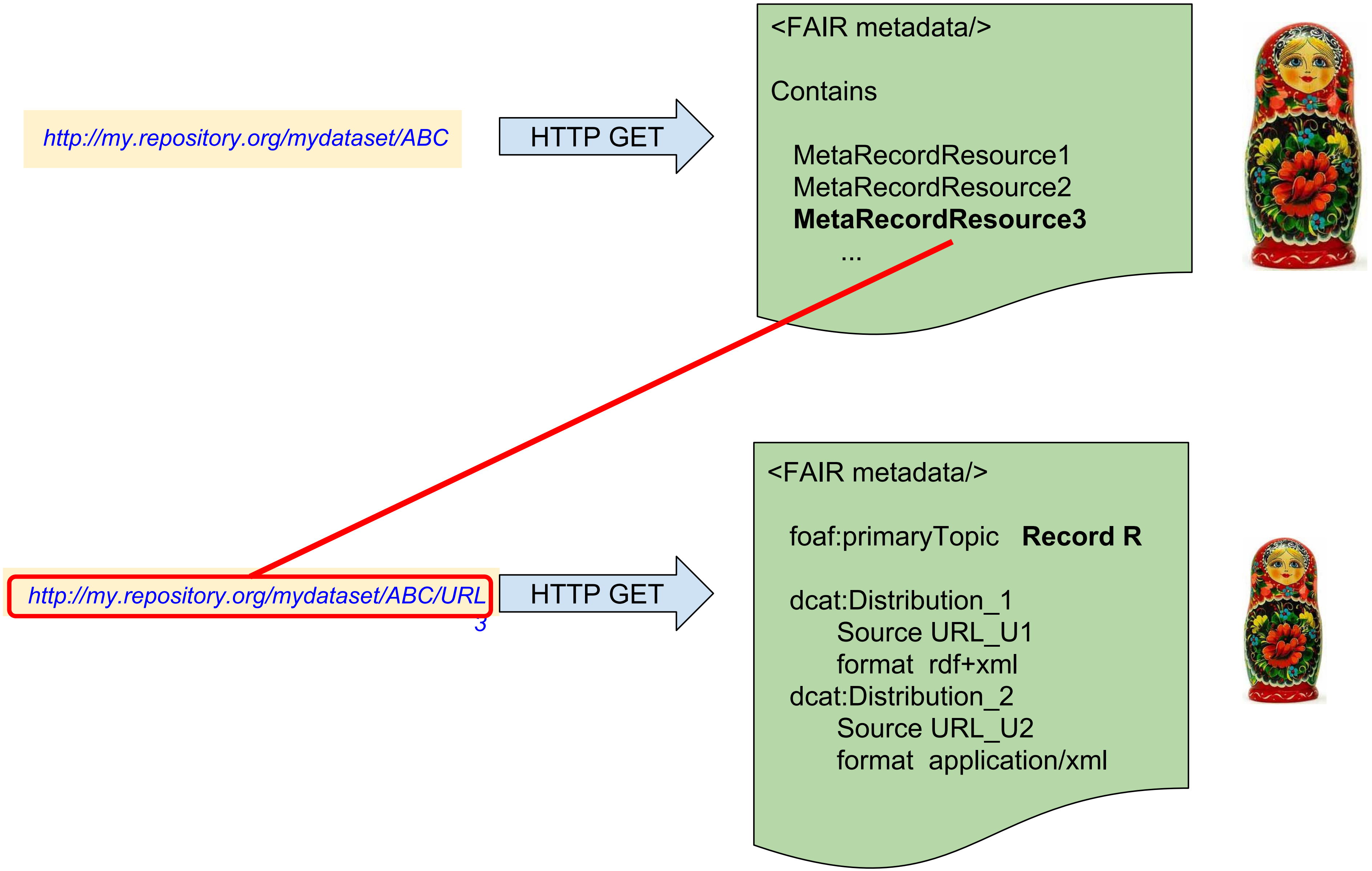
retrieve

<metadata about the record/>

primaryTopic **Record R3**

Distribution_1
    Download @ URL_D1
    format text/csv
Distribution_2
    Download @ URL_D2
    format application/excel

What does a FAIR Accessor “look like”?



http://my.repository.org/mydataset/ABC

HTTP GET

<FAIR metadata/>

Contains

MetaRecordResource1
MetaRecordResource2
MetaRecordResource3
...

http://my.repository.org/mydataset/ABC/URL3

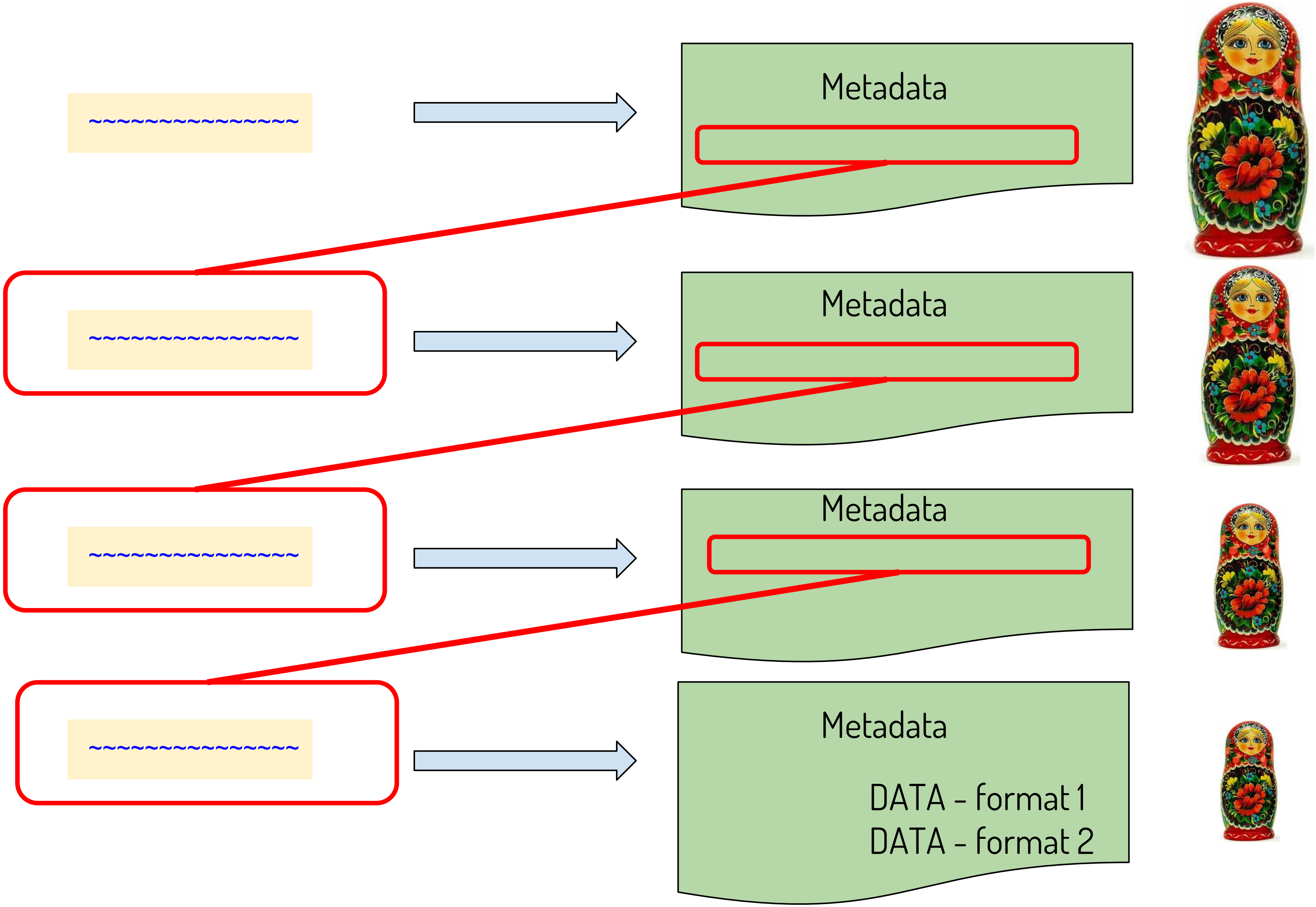
HTTP GET

<FAIR metadata/>

foaf:primaryTopic **Record R**

dcat:Distribution_1
Source URL_U1
format rdf+xml
dcat:Distribution_2
Source URL_U2
format application/xml

Or you may add additional layers...



~~~~~~~~~~~~~~~

Metadata

~~~~~~~~~~~~~~~

Metadata

~~~~~~~~~~~~~~~

Metadata

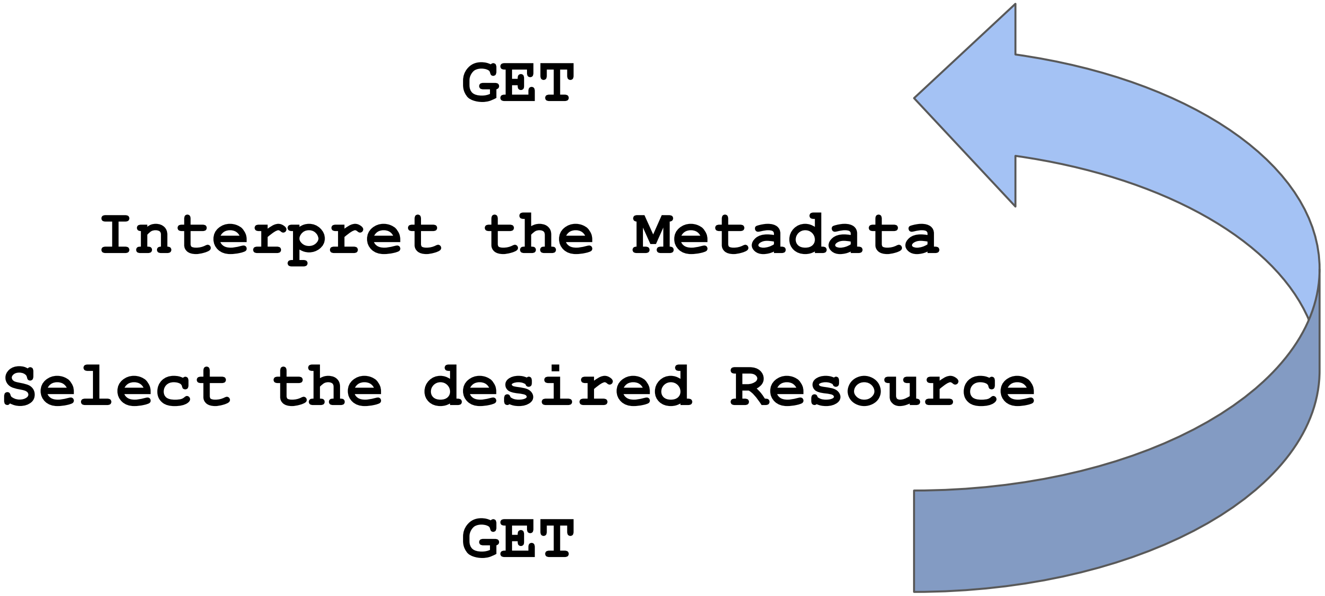
~~~~~~~~~~~~~~~

Metadata

DATA – format 1
DATA – format 2

Features of the FAIR Accessor

1: There is no API



GET

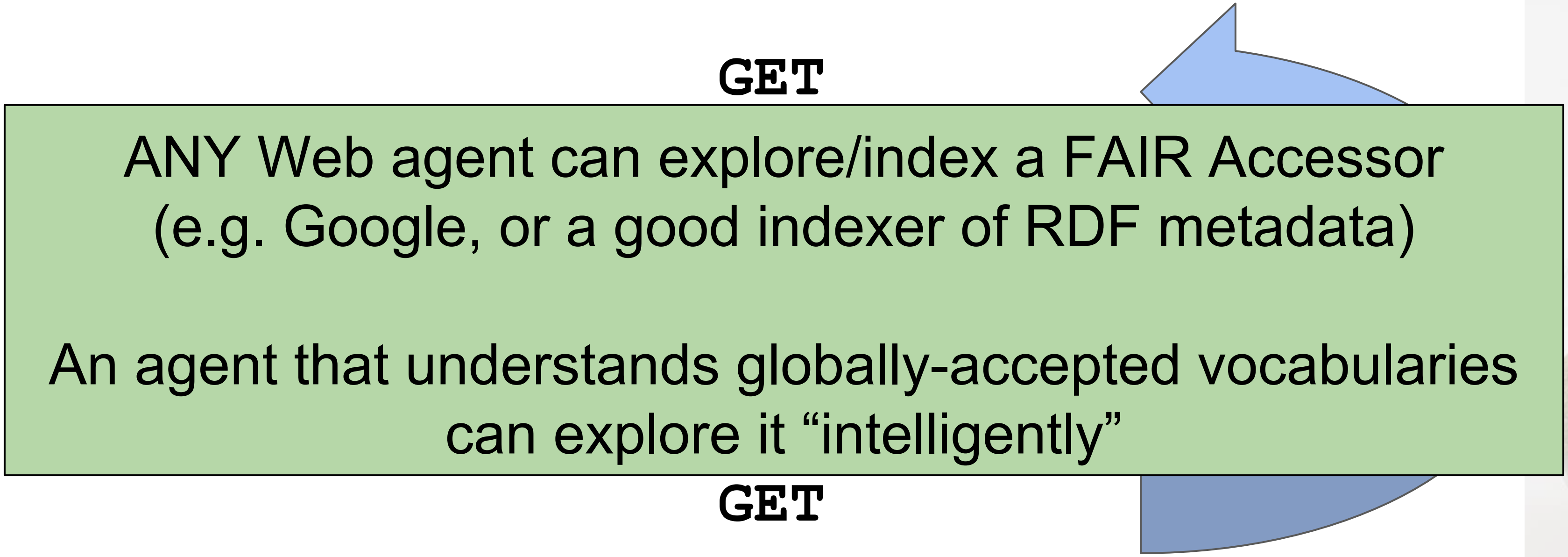
Interpret the Metadata

Select the desired Resource

GET

Features of the FAIR Accessor

1: There is no API



ANY Web agent can explore/index a FAIR Accessor
(e.g. Google, or a good indexer of RDF metadata)

An agent that understands globally-accepted vocabularies
can explore it “intelligently”

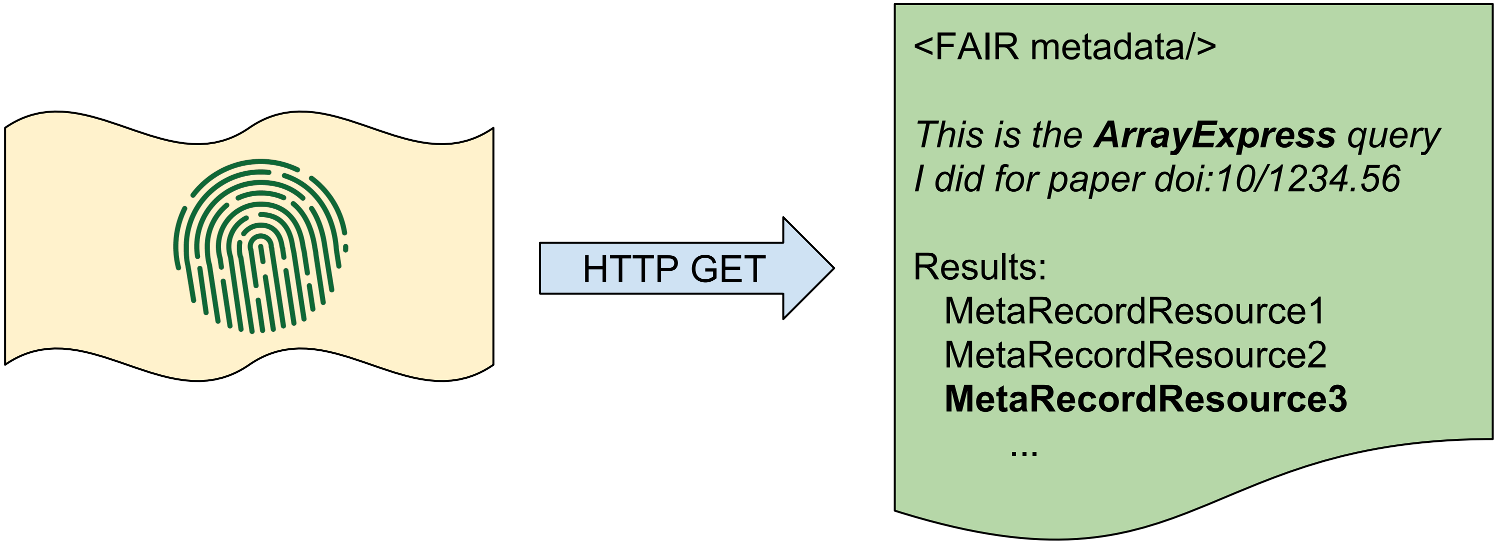
Features of the FAIR Accessor

1: There is no API

It's difficult to get thinner than nothing...

Features of the FAIR Accessor

2: Identifiers for unidentifi-ed/-able things



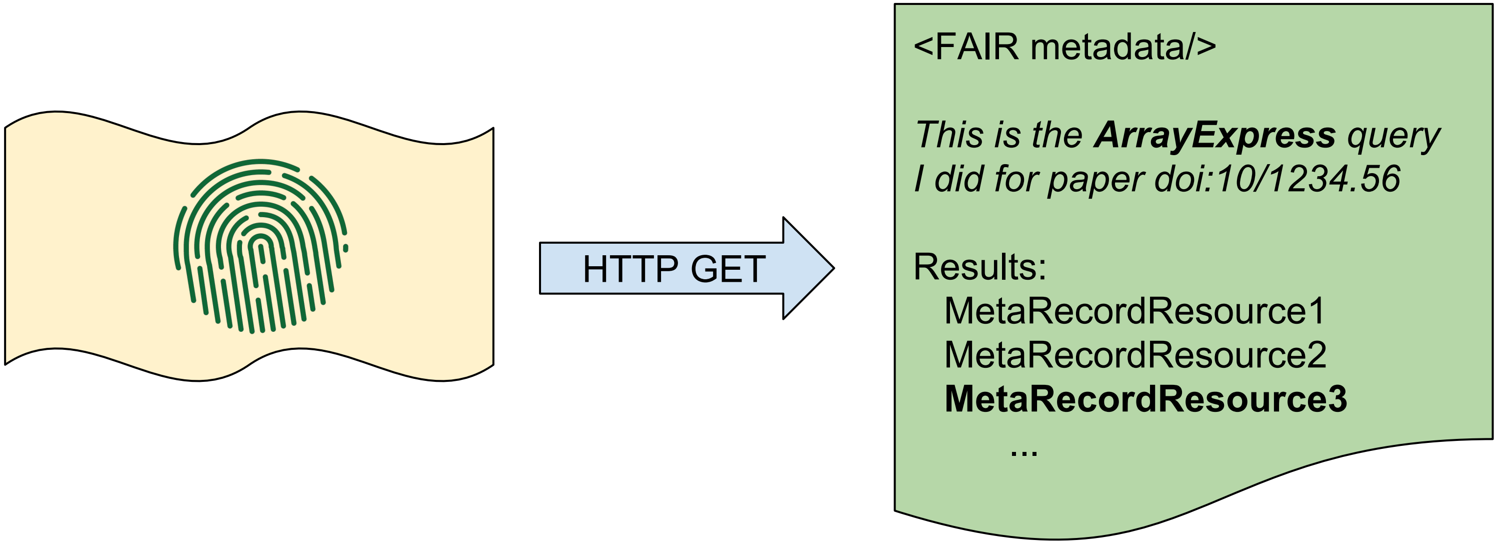
<FAIR metadata/>

*This is the **ArrayExpress** query I did for paper doi:10/1234.56*

Results:
  MetaRecordResource1
  MetaRecordResource2
  MetaRecordResource3
  ...

Features of the FAIR Accessor

2: Identifiers for unidentifi-ed/-able things



<FAIR metadata/>

*This is the **ArrayExpress** query*
I did for paper doi:10/1234.56

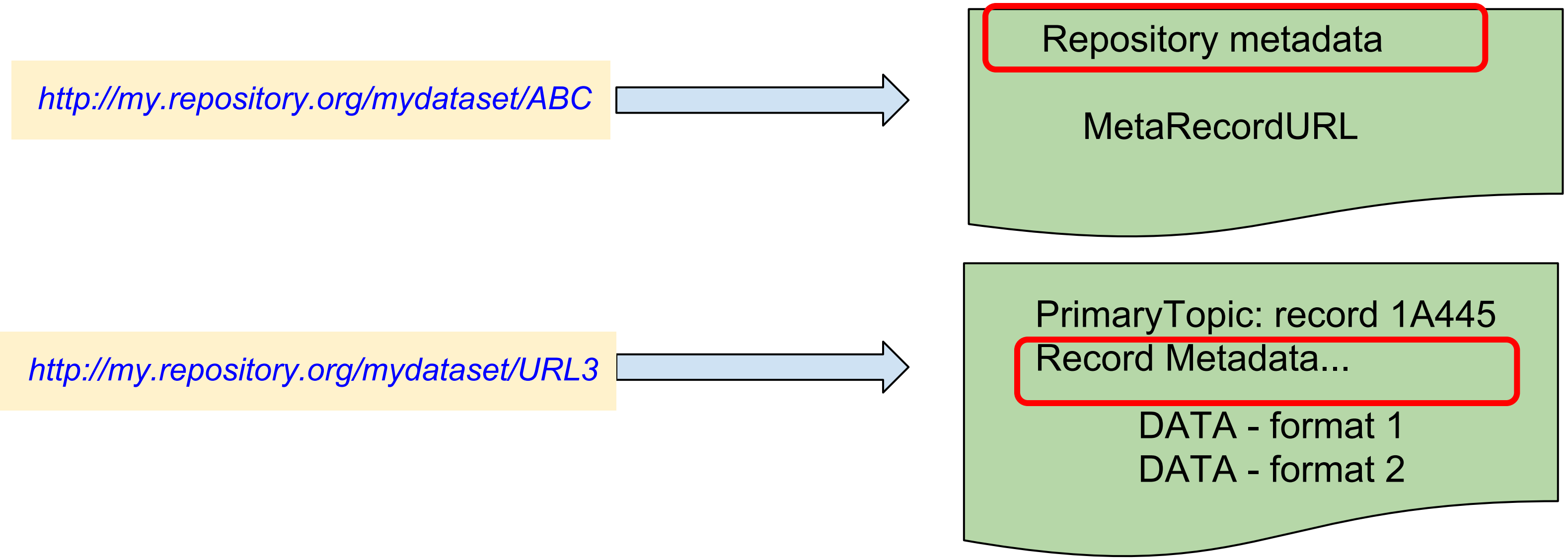
Results:
  MetaRecordResource1
  MetaRecordResource2
  MetaRecordResource3
  ...

Should assist with reproducibility and transparency

Features of the FAIR Accessor

3: A predictable "place" for metadata

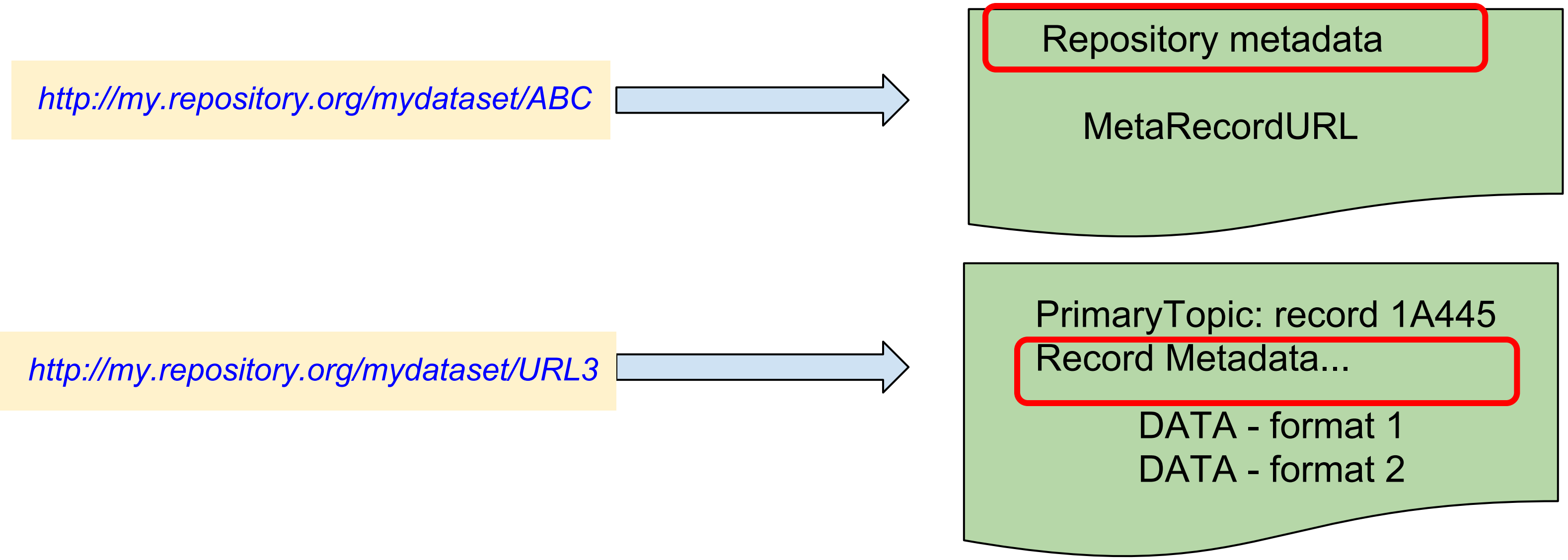
Different "kinds" of metadata have distinct ontological types, and distinct document structures. There is no ambiguity regarding what the metadata is describing - a repository or a record.



Features of the FAIR Accessor

3: A predictable "place" for metadata

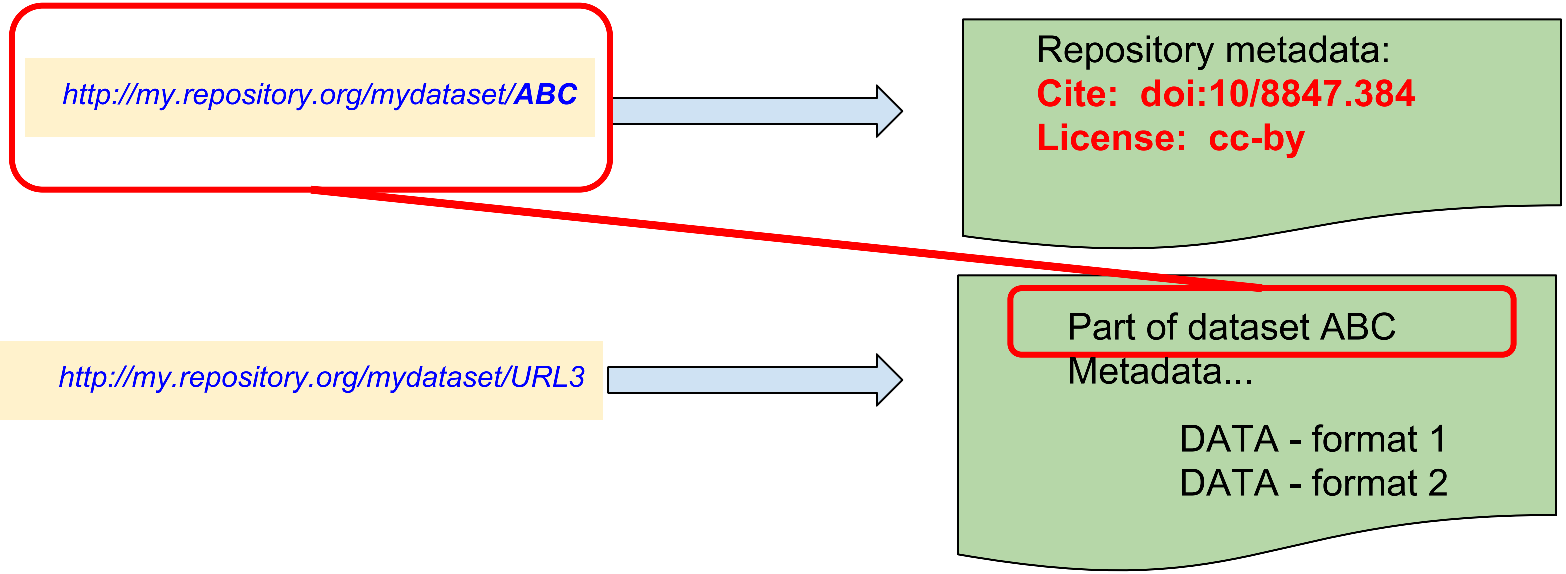
Also encourages an explicit separation of data from metadata. This aligns with Principle A2: "metadata are accessible, even when the data are no longer available"



Features of the FAIR Accessor

4: Symmetry & predictable path to citation

The record metadata contains an "upward" link to the Repository-level metadata, which should contain license and citation information



*http://my.repository.org/mydataset/**ABC***

Repository metadata:
Cite: doi:10/8847.384
License: cc-by

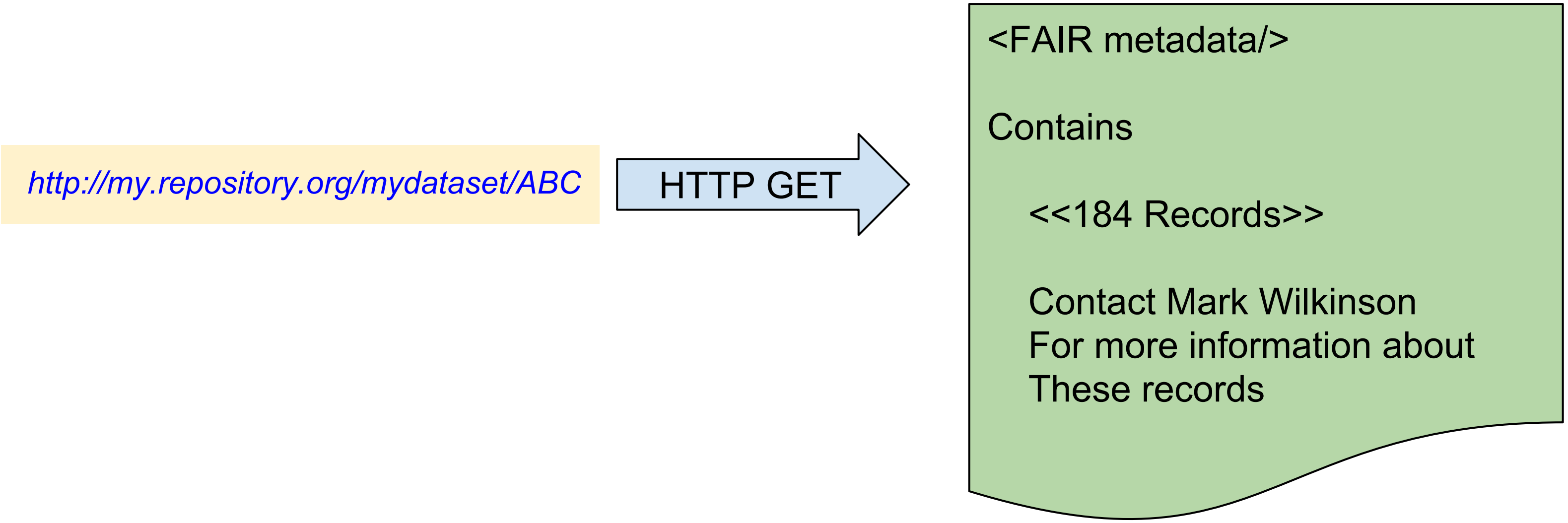
http://my.repository.org/mydataset/URL3

Part of dataset ABC
Metadata...

DATA - format 1
DATA - format 2

Features of the FAIR Accessor

4: Granularity of Access/Privacy/Security



http://my.repository.org/mydataset/ABC

HTTP GET

<FAIR metadata/>

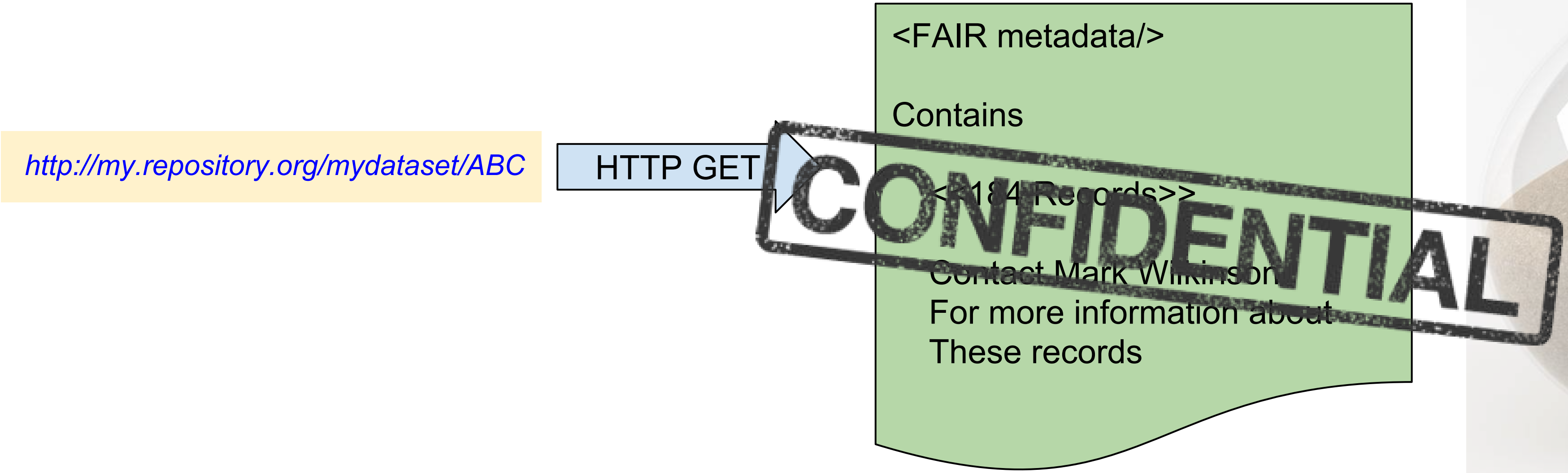
Contains

<<184 Records>>

Contact Mark Wilkinson
For more information about
These records

Features of the FAIR Accessor

4: Granularity of Access/Privacy/Security



http://my.repository.org/mydataset/ABC

HTTP GET

<FAIR metadata/>

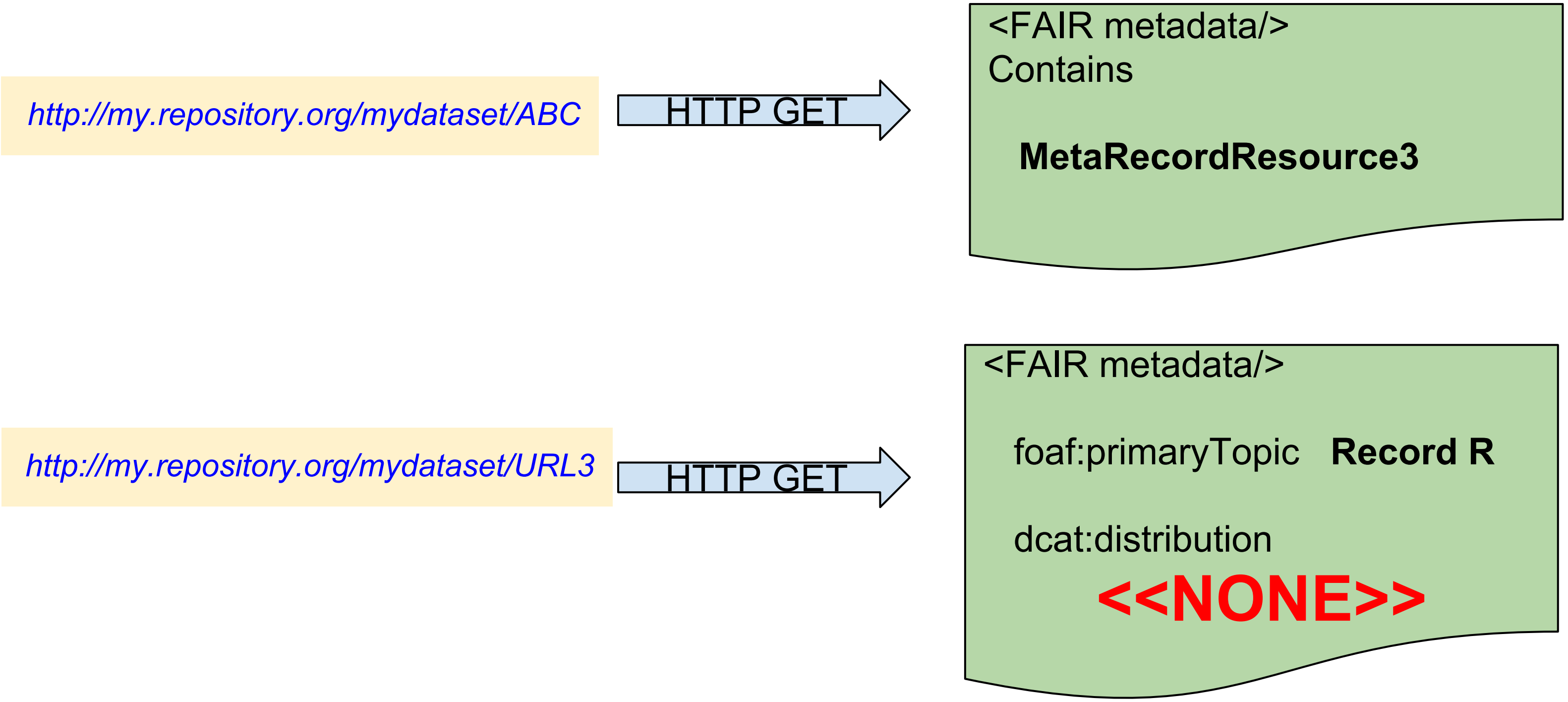
Contains

<<184 Records>>

Contact Mark Wilkinson
For more information about
These records

Features of the FAIR Accessor

4: Granularity of Access/Privacy/Security



http://my.repository.org/mydataset/ABC

HTTP GET

<FAIR metadata/>
Contains

MetaRecordResource3

http://my.repository.org/mydataset/URL3

HTTP GET

<FAIR metadata/>

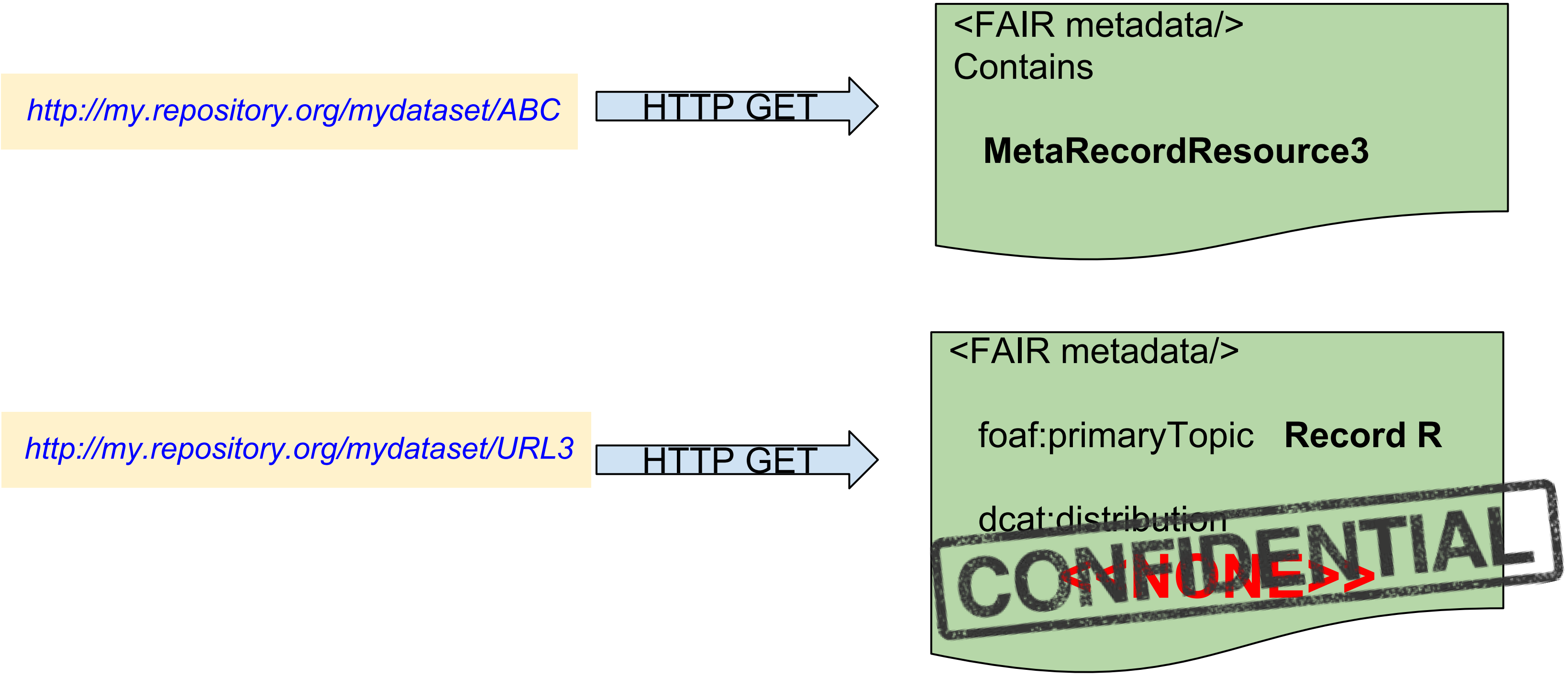
foaf:primaryTopic **Record R**

dcat:distribution

<<NONE>>

Features of the FAIR Accessor

4: Granularity of Access/Privacy/Security



http://my.repository.org/mydataset/ABC

HTTP GET

<FAIR metadata/>
Contains

MetaRecordResource3

http://my.repository.org/mydataset/URL3

HTTP GET

<FAIR metadata/>

foaf:primaryTopic **Record R**

dcat:distribution

<<NONE>>

CONFIDENTIAL

Features of the FAIR Accessor

4: Granularity of Access/Privacy/Security

http://my.repository.org/mydataset/ABC → HTTP GET →

<FAIR metadata/>

Contains
MetaRecordResource3
...

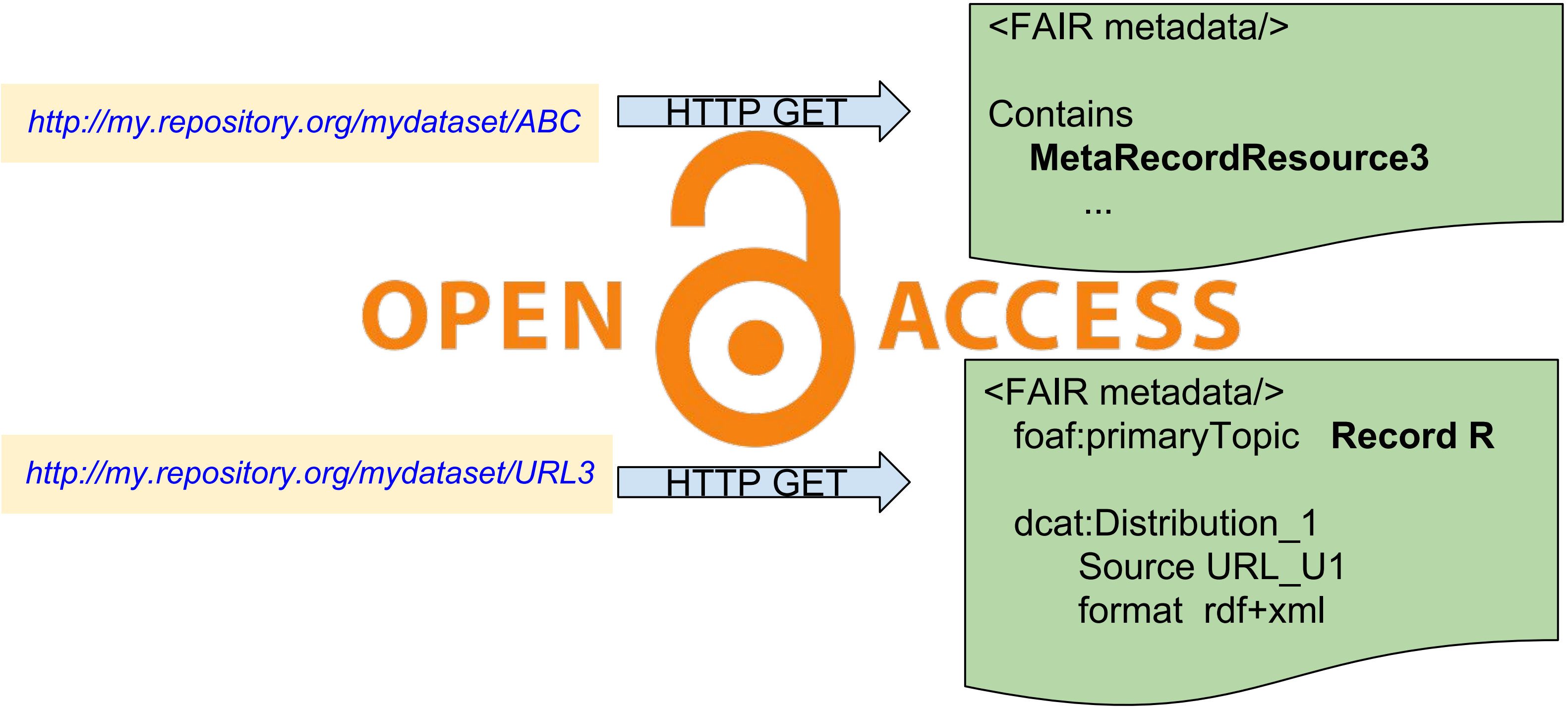
http://my.repository.org/mydataset/URL3 → HTTP GET →

<FAIR metadata/>
foaf:primaryTopic **Record R**

dcat:Distribution_1
Source URL_U1
format rdf+xml

Features of the FAIR Accessor

4: Granularity of Access/Privacy/Security



http://my.repository.org/mydataset/ABC

HTTP GET

<FAIR metadata/>

Contains
MetaRecordResource3
...

http://my.repository.org/mydataset/URL3

HTTP GET

<FAIR metadata/>
foaf:primaryTopic **Record R**

dcat:Distribution_1
Source URL_U1
format rdf+xml

Features of the FAIR Accessor

4: Granularity of Access/Privacy/Security

Thin solution - if it’s private, ***do nothing!***

The Real Thing

A working FAIR Accessor
Serving a “Slice” of UniProt

A real-world scenario...

You are publishing a paper describing the evolution of proteins in the RNA Processing machineries of the fungus *Aspergillus nidulans.*

You want to be a good scholarly publisher interested in transparency and reproducibility

So you must describe, in detail, the inclusion/exclusion criteria for selecting proteins for your dataset

(today, this is generally done either in the text of the paper, or not at all...)

The query that returns the relevant proteins

```
WHERE
{
    ?protein a up:Protein .
    ?protein up:organism ?organism .
    ?organism rdfs:subClassOf taxon:162425 .
    ?protein up:classifiedWith ?go .
    ?go rdfs:subClassOf* <http://purl.obolibrary.org/obo/GO_0006396> .
    bind(replace(str(?protein),
                  "http://purl.uniprot.org/uniprot/", "", "i") as ?id)
}
```

The query that returns the relevant proteins

```
WHERE
{
    ?protein a up:Protein .
    ?protein up:organism ?organism .
    ?organism rdfs:subClassOf taxon:162425 .
    ?protein up:classifiedWith ?go .
    ?go rdfs:subClassOf* <http://purl.obolibrary.org/obo/GO_0006396> .
    bind(replace(str(?protein),
                  "http://purl.uniprot.org/uniprot/", "", "i") as ?id)
}
```

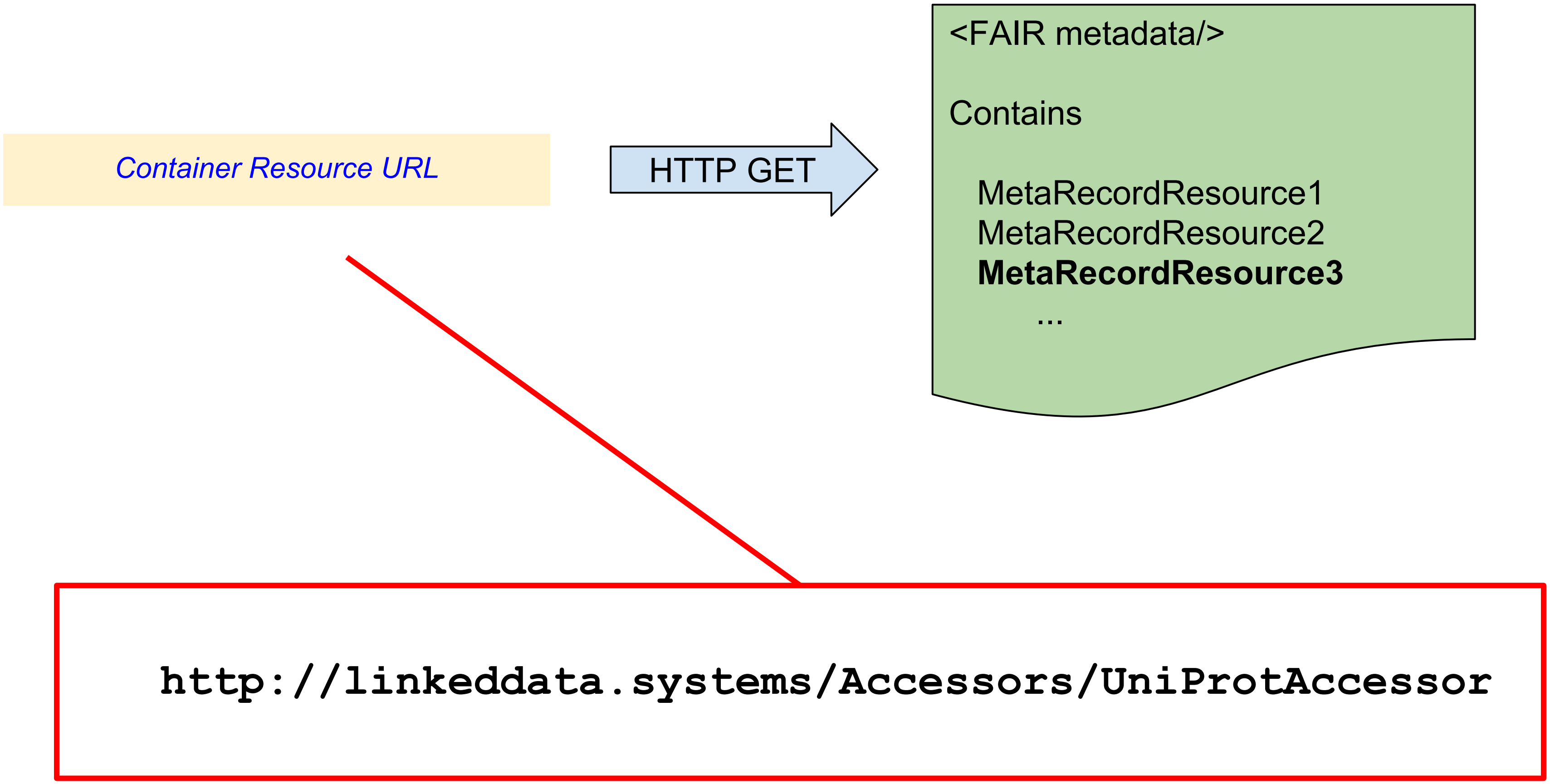
NCBI Taxonomy:
Aspergillus nidulans

The query that returns the relevant proteins

```
WHERE
{
    ?protein a up:Protein .
    ?protein up:organism ?organism .
    ?organism rdfs:subClassOf taxon:162425 .
    ?protein up:classifiedWith ?go .
    ?go rdfs:subClassOf* <http://purl.obolibrary.org/obo/GO_0006396> .
    bind(replace(str(?protein),
                  "http://purl.uniprot.org/uniprot/", "", "i") as ?id)
}
```

Gene Ontology:
RNA Processing

Create and publish a FAIR Accessor for that query



Container Resource URL

HTTP GET

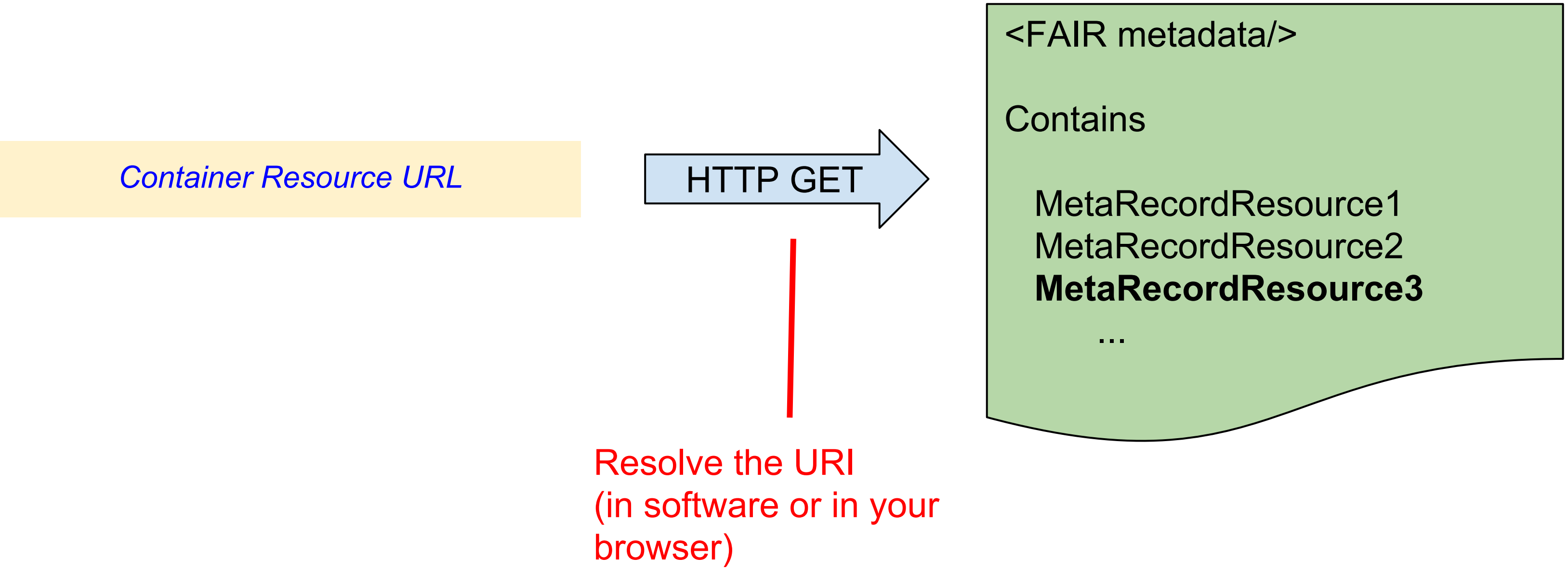
<FAIR metadata/>

Contains

MetaRecordResource1
MetaRecordResource2
MetaRecordResource3
...

`http://linkeddata.systems/Accessors/UniProtAccessor`

Create and publish a FAIR Accessor for that query



Container Resource URL

HTTP GET

<FAIR metadata/>

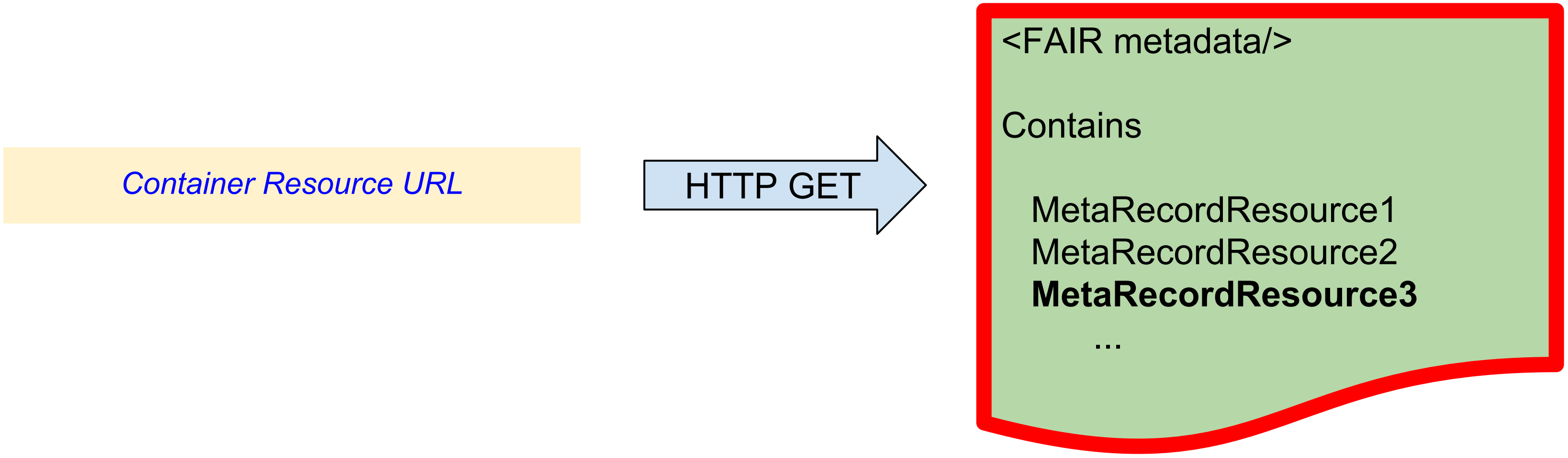
Contains

MetaRecordResource1
MetaRecordResource2
MetaRecordResource3
...

Resolve the URI
(in software or in your
browser)

`http://linkeddata.systems/Accessors/UniProtAccessor`

Create and publish a FAIR Accessor for that query



Container Resource URL

HTTP GET

<FAIR metadata/>

Contains

MetaRecordResource1
MetaRecordResource2
MetaRecordResource3
...

Returns a page of metadata

(in this example, in RDF, rendered into HTML by the Tabulator Browser plugin)

UniProt Slice FAIR Accessor - Aspergillus RNA Processing proteins

creator	wilkinsonlab.info/
language	eng
license	cc by nd4.0
title	UniProt Slice FAIR Accessor - Aspergillus RNA Processing proteins
authored By	0000 0002 9699 485X
entities	412
term has Principal Investigator	Dr. Mark Wilkinson
type	Dataset Basic Container Collection
contact Point	Wilkinson.rdf
description	Takes a SPARQL query of the UniProt database specific to proteins and their GO annotations related to RNA Procssing proteins in Aspergillus and makes it a FAIR Accessor source. The precise query is:

```
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
SELECT distinct ?id

WHERE
{
  ?protein a up:Protein .
  ?protein up:organism ?organism .
  ?organism rdfs:subClassOf taxon:162425 .
  ?protein up:classifiedWith ?go .
  ?go rdfs:subClassOf* <http://purl.obolibrary.org/obo/GO_0006396> .

  bind(replace(str(?protein), "http://purl.uniprot.org/uniprot/", "", "i") as ?id)
}
```

identifier	Uni Prot Accessor
keyword	Aspergillus nidulans Aspergillus Proteins RNA Processing
landing Page	uniprot.org/
language	en
publisher	wilkinsonlab.info/
theme	RNA Processing conceptscheme.rdf
contains	C8UZX9 C8UZY5 C8V0B4 C8V0M2 C8V0U7

UniProt Slice FAIR Accessor - Aspergillus RNA Processing proteins

creator	**wilkinsonlab.info/**
language	**eng**
license	**cc by nd4.0**
title	UniProt Slice FAIR Accessor - Aspergillus RNA Processing proteins
authored By	**0000 0002 9699 485X**
entities	412
term has Principal Investigator	Dr. Mark Wilkinson
type	**Dataset** **Basic Container** **Collection**
contact Point	**Wilkinson.rdf**
description	Takes a SPARQL query of the UniProt database specific to proteins and their GO annotations related to RNA Procssing proteins in Aspergillus and makes it a FAIR Accessor source. The precise query is:

```
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
SELECT distinct ?id

WHERE
{
   ?protein a up:Protein .
   ?protein up:organism ?organism .
   ?organism rdfs:subClassOf taxon:162425 .
   ?protein up:classifiedWith ?go .
   ?go rdfs:subClassOf* <http://purl.obolibrary.org/obo/GO_0006396> .

   bind(replace(str(?protein), "http://purl.uniprot.org/uniprot/", "", "i") as ?id)
}
```

identifier	**Uni Prot Accessor**

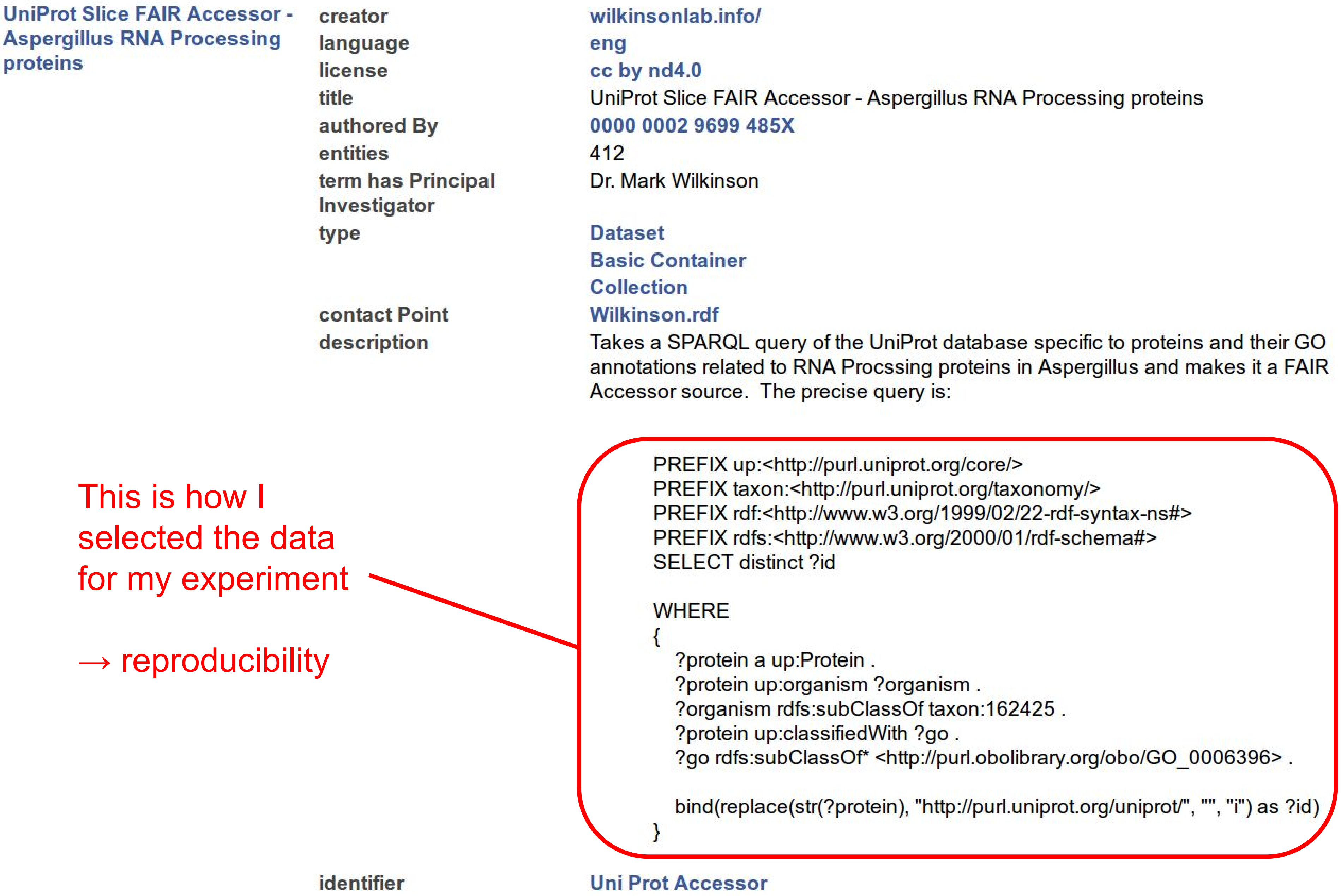
Note that this Metadata is about ME. I am the creator of this dataset, and may be credited for it.

UniProt Slice FAIR Accessor - Aspergillus RNA Processing proteins

creator	wilkinsonlab.info/
language	eng
license	cc by nd4.0
title	UniProt Slice FAIR Accessor - Aspergillus RNA Processing proteins
authored By	0000 0002 9699 485X
entities	412
term has Principal Investigator	Dr. Mark Wilkinson
type	Dataset Basic Container Collection
contact Point	Wilkinson.rdf
description	Takes a SPARQL query of the UniProt database specific to proteins and their GO annotations related to RNA Procssing proteins in Aspergillus and makes it a FAIR Accessor source. The precise query is:

```
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
SELECT distinct ?id

WHERE
{
   ?protein a up:Protein .
   ?protein up:organism ?organism .
   ?organism rdfs:subClassOf taxon:162425 .
   ?protein up:classifiedWith ?go .
   ?go rdfs:subClassOf* <http://purl.obolibrary.org/obo/GO_0006396> .

   bind(replace(str(?protein), "http://purl.uniprot.org/uniprot/", "", "i") as ?id)
}
```

identifier	Uni Prot Accessor



UniProt Slice FAIR Accessor - Aspergillus RNA Processing proteins	creator	**wilkinsonlab.info/**
	language	**eng**
	license	**cc by nd4.0**
	title	UniProt Slice FAIR Accessor - Aspergillus RNA Processing proteins
	authored By	**0000 0002 9699 485X**
	entities	412
	term has Principal Investigator	Dr. Mark Wilkinson
	type	**Dataset** **Basic Container** **Collection**
	contact Point	**Wilkinson.rdf**
	description	Takes a SPARQL query of the UniProt database specific to proteins and their GO annotations related to RNA Procssing proteins in Aspergillus and makes it a FAIR Accessor source. The precise query is:

```
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
SELECT distinct ?id

WHERE
{
   ?protein a up:Protein .
   ?protein up:organism ?organism .
   ?organism rdfs:subClassOf taxon:162425 .
   ?protein up:classifiedWith ?go .
   ?go rdfs:subClassOf* <http://purl.obolibrary.org/obo/GO_0006396> .

   bind(replace(str(?protein), "http://purl.uniprot.org/uniprot/", "", "i") as ?id)
}
```

	identifier	**Uni Prot Accessor**

This is how I
selected the data
for my experiment

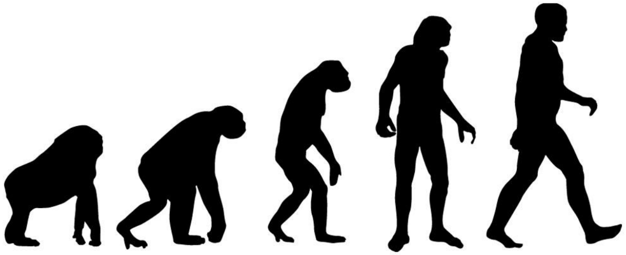
→ reproducibility

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
SELECT distinct ?id

WHERE
{
   ?protein a up:Protein .
   ?protein up:organism ?organism .
   ?organism rdfs:subClassOf taxon:162425 .
   ?protein up:classifiedWith ?go .
   ?go rdfs:subClassOf* <http://purl.obolibrary.org/obo/GO_0006396> .

   bind(replace(str(?protein), "http://purl.uniprot.org/uniprot/", "", "i") as ?id)
}
```

identifier	Uni Prot Accessor
keyword	Aspergillus nidulans Aspergillus Proteins RNA Processing
landing Page	uniprot.org/
language	en
publisher	wilkinsonlab.info/
theme	RNA Processing conceptscheme.rdf
contains	C8UZX9 C8UZY5 C8V0B4 C8V0M2 C8V0U7

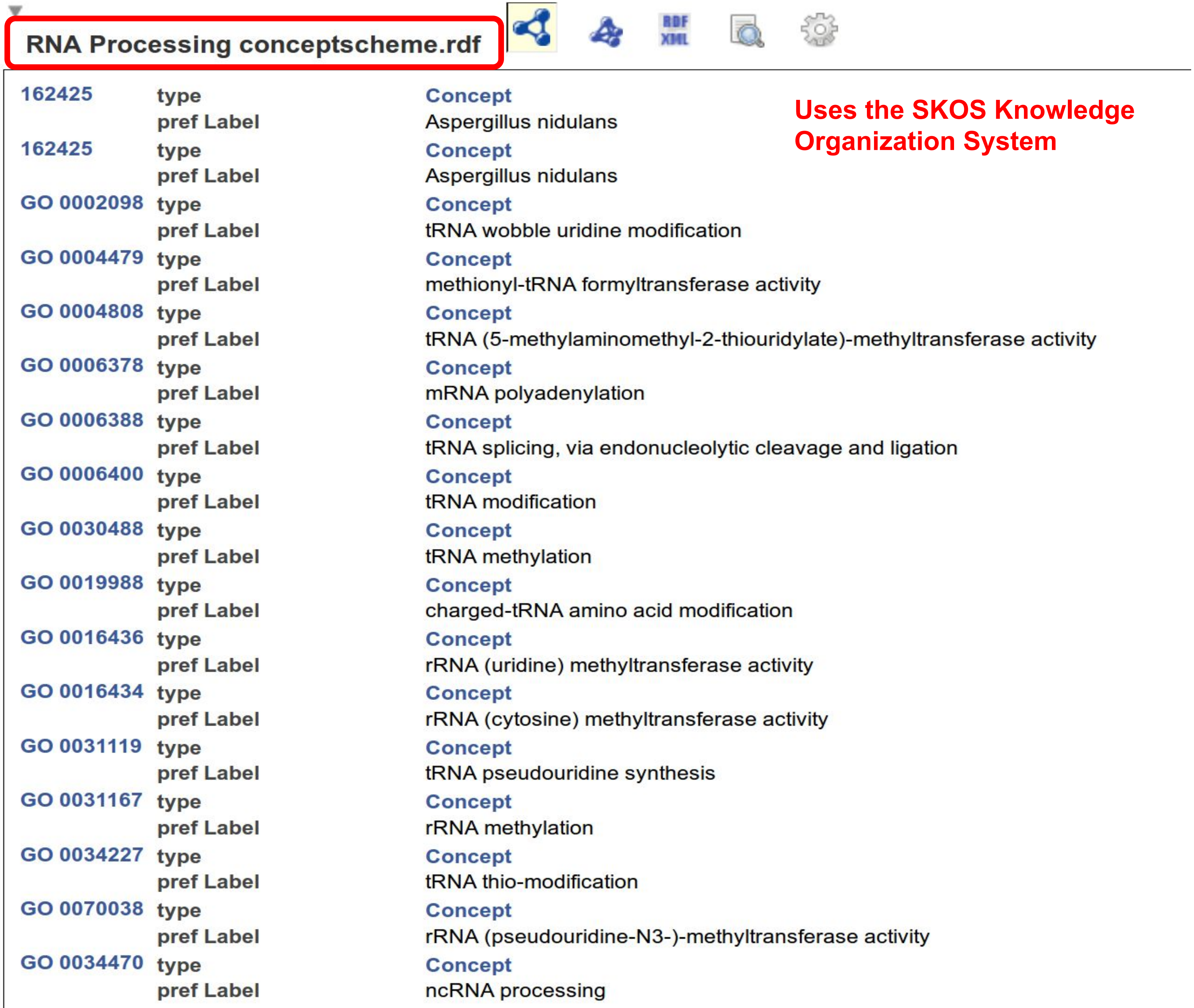
```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
SELECT distinct ?id

WHERE
{
   ?protein a up:Protein .
   ?protein up:organism ?organism .
   ?organism rdfs:subClassOf taxon:162425 .
   ?protein up:classifiedWith ?go .
   ?go rdfs:subClassOf* <http://purl.obolibrary.org/obo/GO_0006396> .

   bind(replace(str(?protein), "http://purl.uniprot.org/uniprot/", "", "i") as ?id)
}
```

identifier	**Uni Prot Accessor**
keyword	Aspergillus nidulans Aspergillus Proteins RNA Processing
landing Page	**uniprot.org/**
language	**en**
publisher	**wilkinsonlab.info/**
theme	**RNA Processing conceptscheme.rdf**
contains	**C8UZX9** **C8UZY5** **C8V0B4** **C8V0M2** **C8V0U7**



```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
SELECT distinct ?id

WHERE
{
   ?protein a up:Protein .
   ?protein up:organism ?organism .
   ?organism rdfs:subClassOf taxon:162425 .
   ?protein up:classifiedWith ?go .
   ?go rdfs:subClassOf* <http://purl.obolibrary.org/obo/GO_0006396> .

   bind(replace(str(?protein), "http://purl.uniprot.org/uniprot/", "", "i") as ?id)
}
```

identifier	Uni Prot Accessor
keyword	Aspergillus nidulans Aspergillus Proteins RNA Processing
landing Page	uniprot.org/
language	en
publisher	wilkinsonlab.info/
theme	RNA Processing conceptscheme.rdf
contains	C8UZX9 C8UZY5 C8V0B4 C8V0M2 C8V0U7

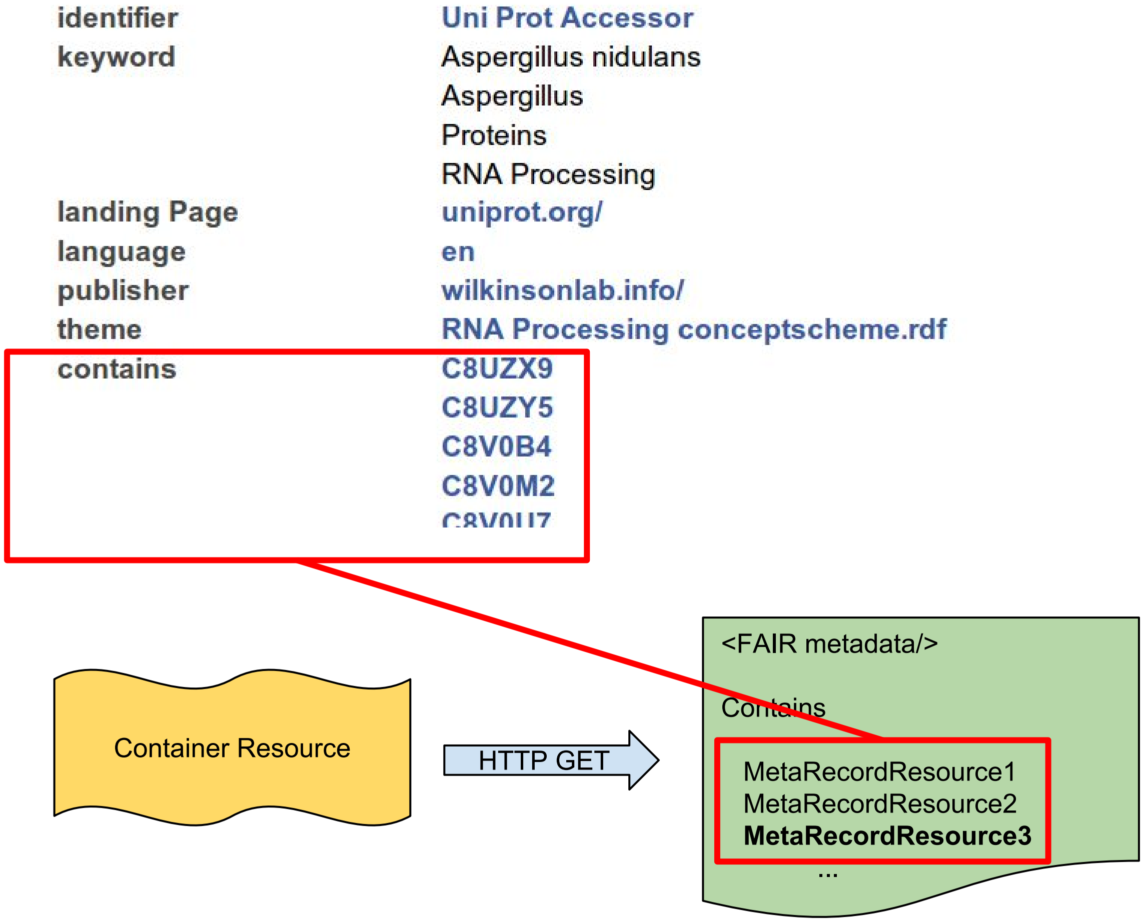


RNA Processing conceptscheme.rdf

Uses the SKOS Knowledge Organization System

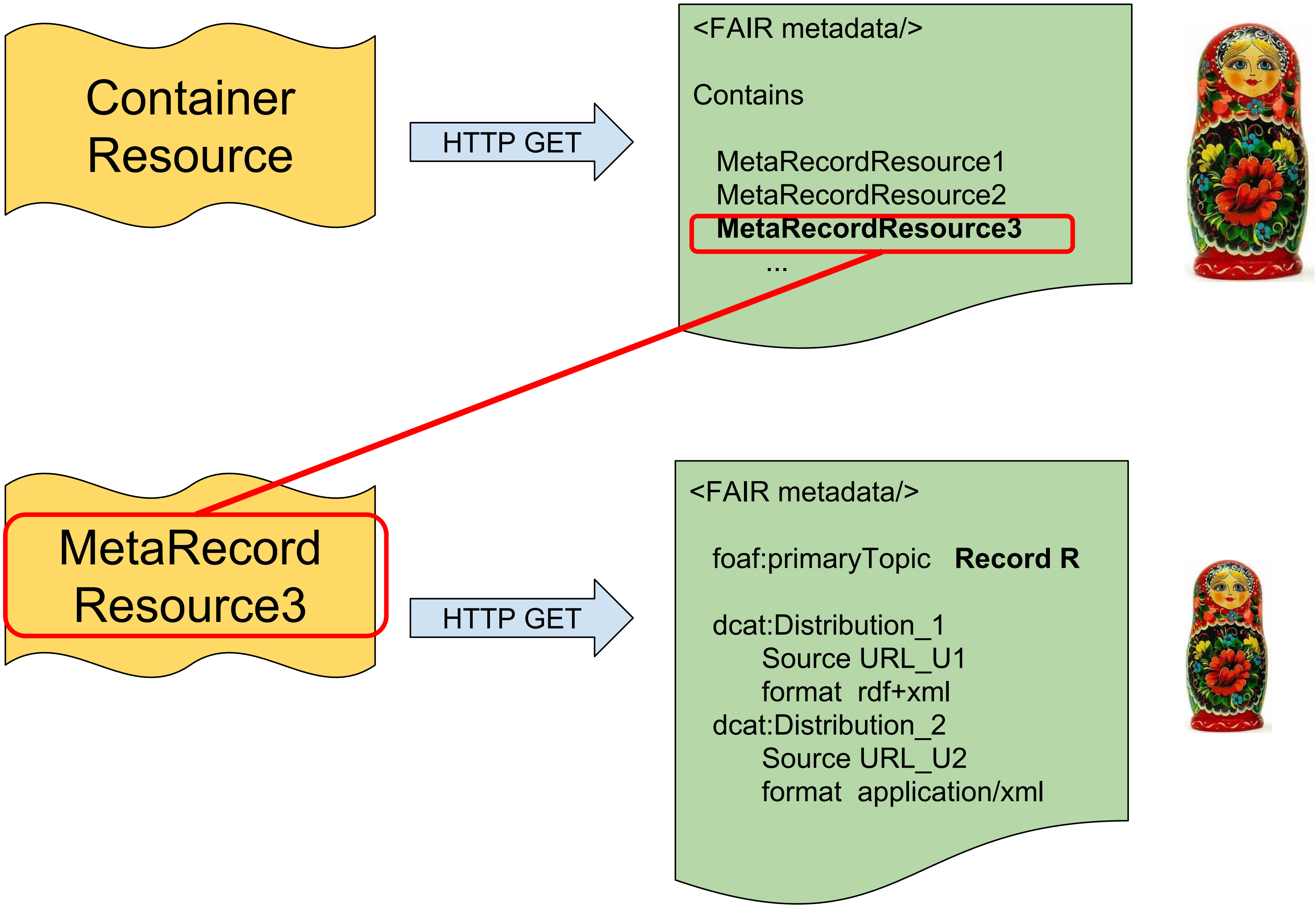
162425	type	Concept
	pref Label	Aspergillus nidulans
162425	type	Concept
	pref Label	Aspergillus nidulans
GO 0002098	type	Concept
	pref Label	tRNA wobble uridine modification
GO 0004479	type	Concept
	pref Label	methionyl-tRNA formyltransferase activity
GO 0004808	type	Concept
	pref Label	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase activity
GO 0006378	type	Concept
	pref Label	mRNA polyadenylation
GO 0006388	type	Concept
	pref Label	tRNA splicing, via endonucleolytic cleavage and ligation
GO 0006400	type	Concept
	pref Label	tRNA modification
GO 0030488	type	Concept
	pref Label	tRNA methylation
GO 0019988	type	Concept
	pref Label	charged-tRNA amino acid modification
GO 0016436	type	Concept
	pref Label	rRNA (uridine) methyltransferase activity
GO 0016434	type	Concept
	pref Label	rRNA (cytosine) methyltransferase activity
GO 0031119	type	Concept
	pref Label	tRNA pseudouridine synthesis
GO 0031167	type	Concept
	pref Label	rRNA methylation
GO 0034227	type	Concept
	pref Label	tRNA thio-modification
GO 0070038	type	Concept
	pref Label	rRNA (pseudouridine-N3-)-methyltransferase activity
GO 0034470	type	Concept
	pref Label	ncRNA processing

```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
SELECT distinct ?id

WHERE
{
   ?protein a up:Protein .
   ?protein up:organism ?organism .
   ?organism rdfs:subClassOf taxon:162425 .
   ?protein up:classifiedWith ?go .
   ?go rdfs:subClassOf* <http://purl.obolibrary.org/obo/GO_0006396> .

   bind(replace(str(?protein), "http://purl.uniprot.org/uniprot/", "", "i") as ?id)
}
```



identifier	Uni Prot Accessor
keyword	Aspergillus nidulans Aspergillus Proteins RNA Processing
landing Page	uniprot.org/
language	en
publisher	wilkinsonlab.info/
theme	RNA Processing conceptscheme.rdf
contains	C8UZX9 C8UZY5 C8V0B4 C8V0M2

Step down to individual Record metadata



Container Resource

HTTP GET

<FAIR metadata/>

Contains

MetaRecordResource1
MetaRecordResource2
MetaRecordResource3
...

MetaRecord Resource3

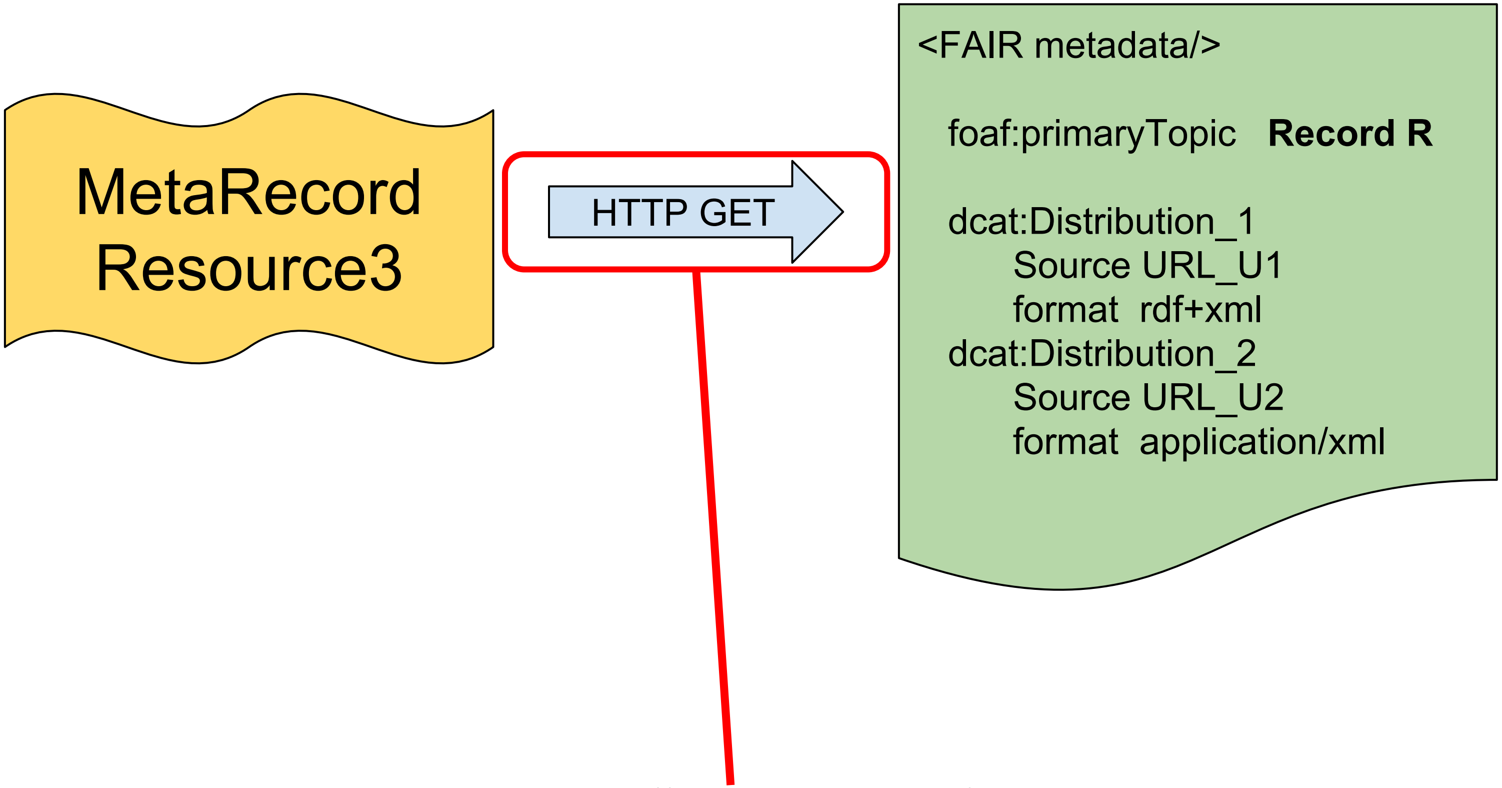
HTTP GET

<FAIR metadata/>

foaf:primaryTopic **Record R**

dcat:Distribution_1
Source URL_U1
format rdf+xml
dcat:Distribution_2
Source URL_U2
format application/xml

Step down to individual Record metadata



MetaRecord Resource3

HTTP GET

<FAIR metadata/>

foaf:primaryTopic **Record R**

dcat:Distribution_1
    Source URL_U1
    format rdf+xml
dcat:Distribution_2
    Source URL_U2
    format application/xml

Call HTTP GET on the URL representing the MetaRecord Resource for the desired record in the Container

The document that is returned

UniProt Protein C8VBH1

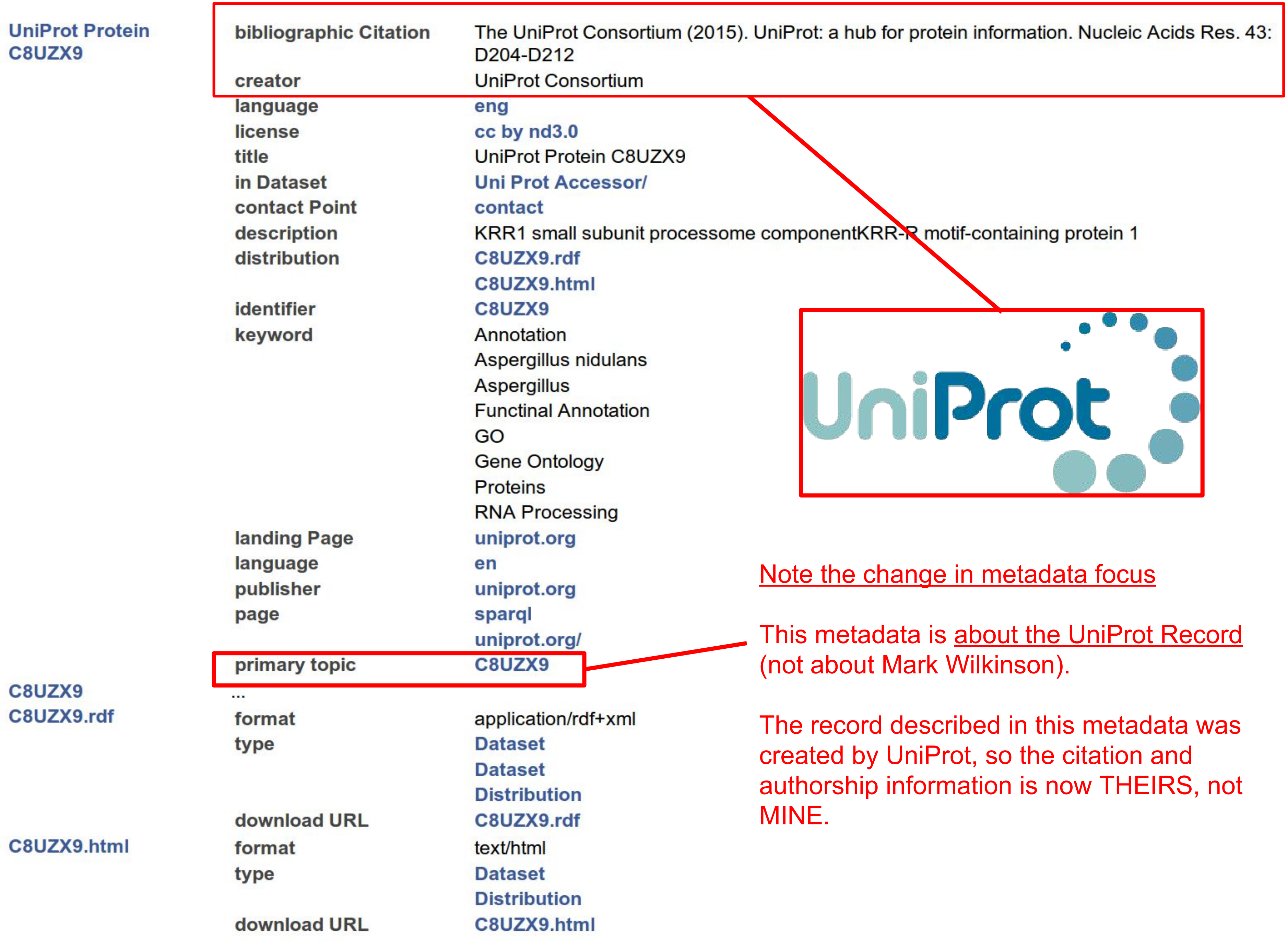
UniProt Protein C8VBH1	bibliographic Citation	The UniProt Consortium (2015). UniProt: a hub for protein information. Nucleic Acids Res. 43: D204-D212
	creator	UniProt Consortium
	language	eng
	license	3.0/
	title	UniProt Protein C8VBH1
	Version	UniProt release 2016_09
	in dataset	Uni Prot Accessor/
	contact point	contact
	description	Exonuclease, putative (AFU_orthologue AFUA_2G05560)
	distribution	fragments?subject=http%3A%2F%2Fidentifiers%2Eorg%2Funiprot%2FC8VBH1&predicate=http%3A%2F%2Fpurl%2Euniprot%2Eorg%2Fcore%2Fclassified With fragments?subject=http%3A%2F%2Fidentifiers%2Eorg%2Funiprot%2FC8VBH1&predicate=http%3A%2F%2Fpurl%2Euniprot%2Eorg%2Fcore%2Forganism C8VBH1.html C8VBH1.rdf
	identifier	C8VBH1
	keyword	Annotation Aspergillus nidulans Aspergillus Functinal Annotation GO Gene Ontology Proteins RNA Processing
	landing page	uniprot.org
	language	en
	publisher	uniprot.org
	page	sparql uniprot.org/
	primary topic	C8VBH1
C8VBH1	...	
Mappings7968453A 9A8F 11E6 A83B A5E65C07C3DD	logical Source	Source7968453A 9A8F 11E6 A83B A5E65C07C3DD
	predicate Object Map	POMap7968453A 9A8F 11E6 A83B A5E65C07C3DD
	subject Map	Subject Map7968453A 9A8F 11E6 A83B A5E65C07C3DD
Mappings7968543A 9A8F 11E6 A83B A5E65C07C3DD	logical Source	Source7968543A 9A8F 11E6 A83B A5E65C07C3DD
	predicate Object Map	POMap7968543A 9A8F 11E6 A83B A5E65C07C3DD
	subject Map	Subject Map7968543A 9A8F 11E6 A83B A5E65C07C3DD
Object Map7968453A 9A8F 11E6 A83B A5E65C07C3DD	parent Triples Map	Subject Map27968453A 9A8F 11E6 A83B A5E65C07C3DD
Object Map7968543A 9A8F 11E6 A83B A5E65C07C3DD	parent Triples Map	Subject Map27968543A 9A8F 11E6 A83B A5E65C07C3DD
POMap7968453A 9A8F 11E6 A83B A5E65C07C3DD	object Map	Object Map7968453A 9A8F 11E6 A83B A5E65C07C3DD
	predicate	organism
POMap7968543A 9A8F 11E6 A83B A5E65C07C3DD	object Map	Object Map7968543A 9A8F 11E6 A83B A5E65C07C3DD
	predicate	classified With
Source7968453A 9A8F 11E6 A83B A5E65C07C3DD	Mapping	Mappings7968453A 9A8F 11E6 A83B A5E65C07C3DD
	reference Formulation	Triple Pattern Fragments
	source	fragments?subject=http%3A%2F%2Fidentifiers%2Eorg%2Funiprot%2FC8VBH1&predicate=http%3A%2F%2Fpurl%2Euniprot%2Eorg%2Fcore%2Forganism
Source7968543A 9A8F 11E6 A83B A5E65C07C3DD	Mapping	Mappings7968543A 9A8F 11E6 A83B A5E65C07C3DD
	reference Formulation	Triple Pattern Fragments
	source	fragments?subject=http%3A%2F%2Fidentifiers%2Eorg%2Funiprot%2FC8VBH1&predicate=http%3A%2F%2Fpurl%2Euniprot%2Eorg%2Fcore%2Fclassified With
Subject Map27968453A 9A8F 11E6 A83B A5E65C07C3DD	class	data 1179
	template	http://identifiers.org/taxon/{TAX}
Subject Map27968543A 9A8F 11E6 A83B A5E65C07C3DD	class	data 1176
	template	http://purl.obolibrary.org/obo/{GO}
Subject Map7968453A 9A8F 11E6 A83B A5E65C07C3DD	class	organism
	template	http://identifiers.org/uniprot/{ID}
Subject Map7968543A 9A8F 11E6 A83B A5E65C07C3DD	class	data 0896
	template	http://identifiers.org/uniprot/{ID}
fragments?subject=http%3A%2F%2Fidentifiers%2Eorg%2Funiprot%2FC8VBH1&predicate=http%3A%2F%2Fpurl%2Euniprot%2Eorg%2Fcore%2Fclassified With	format	application/rdf+xml application/x-turtle text/html
	type	Projector Dataset dataset Distribution
	download URL	fragments?subject=http%3A%2F%2Fidentifiers%2Eorg%2Funiprot%2FC8VBH1&predicate=http%3A%2F

<FAIR metadata/>

foaf:primaryTopic **up:C8UZX9**

dcat:Distribution_1
    Source URL_U1
    format rdf+xml
dcat:Distribution_2
    Source URL_U2
    format application/xml

UniProt Protein C8UZX9	**bibliographic Citation**	The UniProt Consortium (2015). UniProt: a hub for protein information. Nucleic Acids Res. 43: D204-D212
	creator	UniProt Consortium
	language	**eng**
	license	**cc by nd3.0**
	title	UniProt Protein C8UZX9
	in Dataset	**Uni Prot Accessor/**
	contact Point	**contact**
	description	KRR1 small subunit processome componentKRR-R motif-containing protein 1
	distribution	**C8UZX9.rdf**
		C8UZX9.html
	identifier	**C8UZX9**
	keyword	Annotation
		Aspergillus nidulans
		Aspergillus
		Functinal Annotation
		GO
		Gene Ontology
		Proteins
		RNA Processing
	landing Page	**uniprot.org**
	language	**en**
	publisher	**uniprot.org**
	page	**sparql**
		uniprot.org/
	primary topic	**C8UZX9**
C8UZX9	...	
C8UZX9.rdf	**format**	application/rdf+xml
	type	**Dataset**
		Dataset
		Distribution
	download URL	**C8UZX9.rdf**
C8UZX9.html	**format**	text/html
	type	**Dataset**
		Distribution
	download URL	**C8UZX9.html**



UniProt Protein C8UZX9	**bibliographic Citation**	The UniProt Consortium (2015). UniProt: a hub for protein information. Nucleic Acids Res. 43: D204-D212
	creator	UniProt Consortium
	language	eng
	license	cc by nd3.0
	title	UniProt Protein C8UZX9
	in Dataset	Uni Prot Accessor/
	contact Point	contact
	description	KRR1 small subunit processome componentKRR-R motif-containing protein 1
	distribution	C8UZX9.rdf C8UZX9.html
	identifier	C8UZX9
	keyword	Annotation Aspergillus nidulans Aspergillus Functinal Annotation GO Gene Ontology Proteins RNA Processing
	landing Page	uniprot.org
	language	en
	publisher	uniprot.org
	page	sparql uniprot.org/
	primary topic	C8UZX9
C8UZX9	...	
C8UZX9.rdf	**format**	application/rdf+xml
	type	Dataset Dataset Distribution
	download URL	C8UZX9.rdf
C8UZX9.html	**format**	text/html
	type	Dataset Distribution
	download URL	C8UZX9.html

Note the change in metadata focus

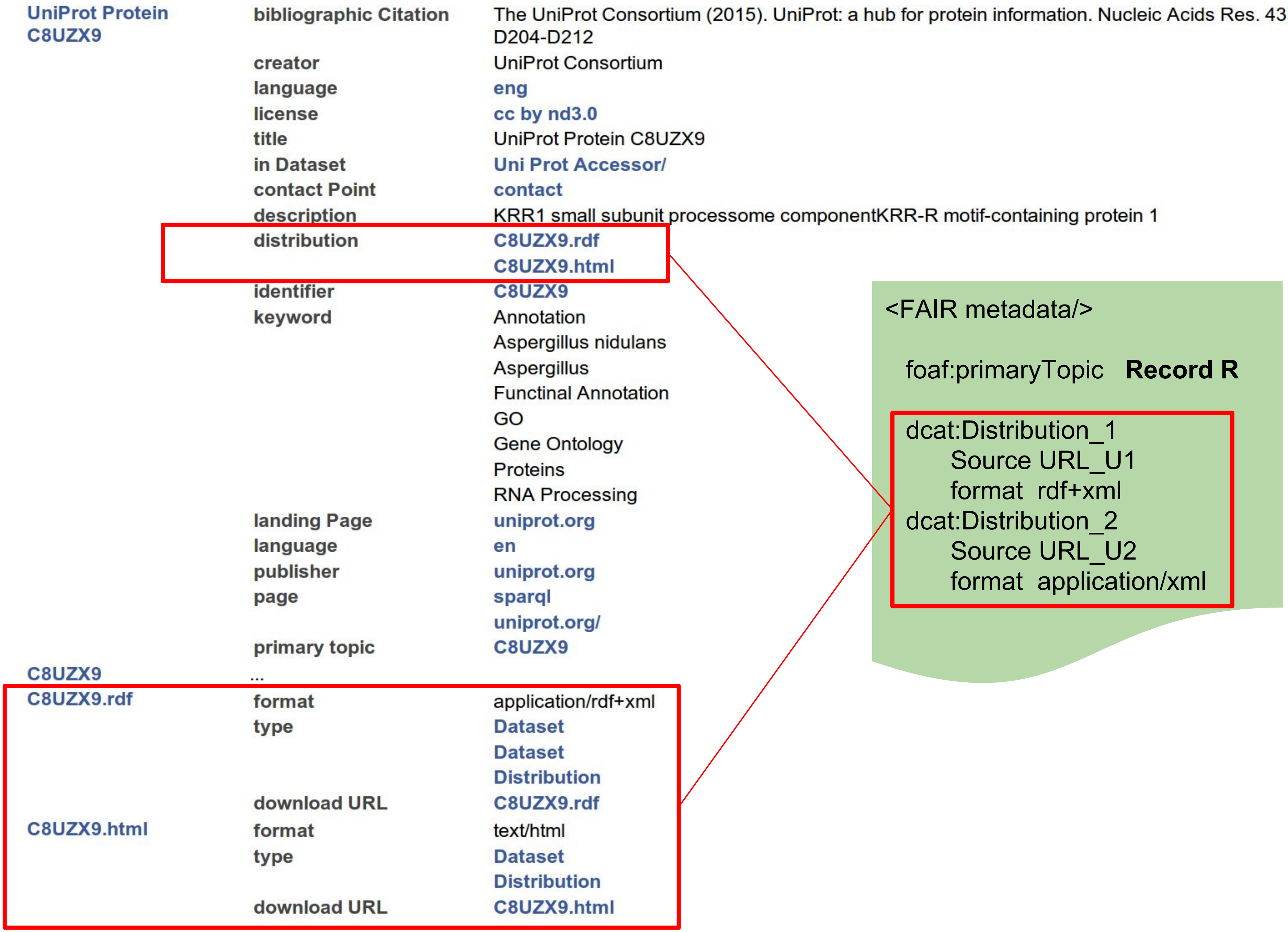
This metadata is about the UniProt Record (not about Mark Wilkinson).

The record described in this metadata was created by UniProt, so the citation and authorship information is now THEIRS, not MINE.

UniProt Protein C8UZX9	**bibliographic Citation**	The UniProt Consortium (2015). UniProt: a hub for protein information. Nucleic Acids Res. 43: D204-D212
	creator	UniProt Consortium
	language	**eng**
	license	**cc by nd3.0**
	title	UniProt Protein C8UZX9
	in Dataset	**Uni Prot Accessor/**
	contact Point	**contact**
	description	KRR1 small subunit processome componentKRR-R motif-containing protein 1
	distribution	**C8UZX9.rdf** **C8UZX9.html**
	identifier	**C8UZX9**
	keyword	Annotation Aspergillus nidulans Aspergillus Functinal Annotation GO Gene Ontology Proteins RNA Processing
	landing Page	**uniprot.org**
	language	**en**
	publisher	**uniprot.org**
	page	**sparql** **uniprot.org/**
	primary topic	**C8UZX9**
C8UZX9	...	
C8UZX9.rdf	**format**	application/rdf+xml
	type	**Dataset** **Dataset** **Distribution**
	download URL	**C8UZX9.rdf**
C8UZX9.html	**format**	text/html
	type	**Dataset** **Distribution**
	download URL	**C8UZX9.html**

Container Resource

Symmetrical Link back upward to the Accessor Container, for additional metadata

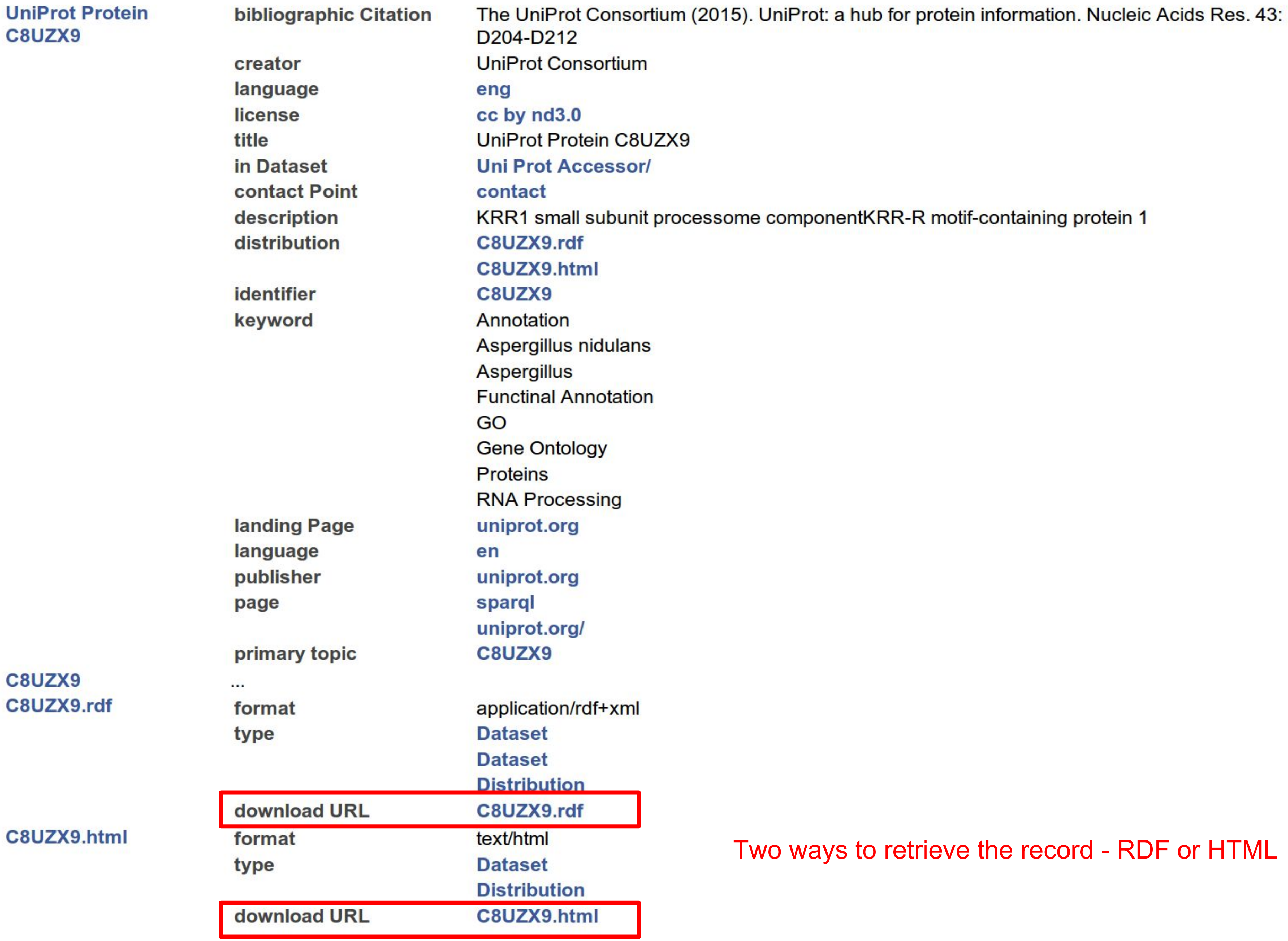


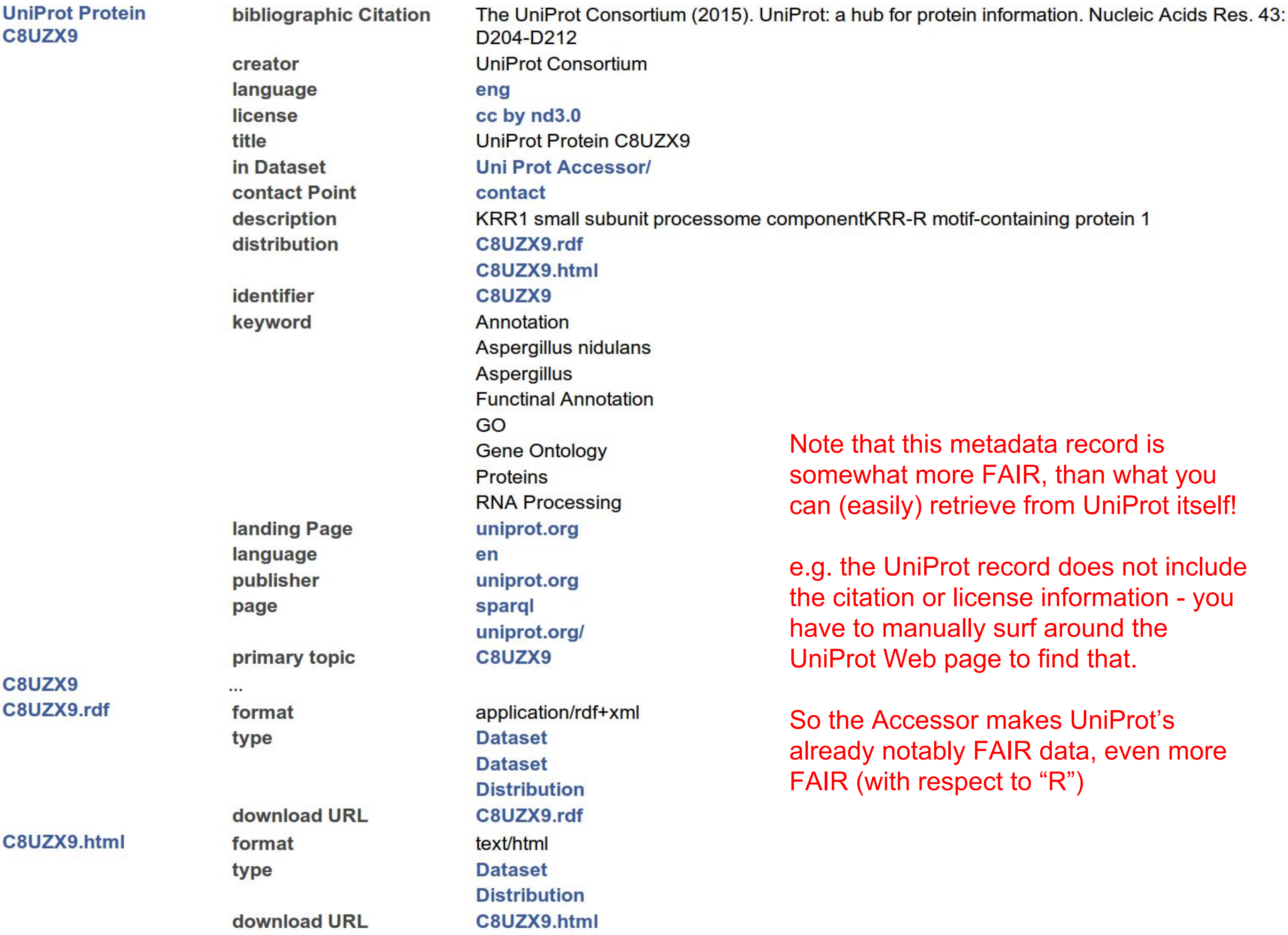
UniProt Protein C8UZX9	bibliographic Citation	The UniProt Consortium (2015). UniProt: a hub for protein information. Nucleic Acids Res. 43: D204-D212
	creator	UniProt Consortium
	language	eng
	license	cc by nd3.0
	title	UniProt Protein C8UZX9
	in Dataset	Uni Prot Accessor/
	contact Point	contact
	description	KRR1 small subunit processome componentKRR-R motif-containing protein 1
	distribution	C8UZX9.rdf
		C8UZX9.html
	identifier	C8UZX9
	keyword	Annotation
		Aspergillus nidulans
		Aspergillus
		Functinal Annotation
		GO
		Gene Ontology
		Proteins
		RNA Processing
	landing Page	uniprot.org
	language	en
	publisher	uniprot.org
	page	sparql
		uniprot.org/
	primary topic	C8UZX9
C8UZX9	...	
C8UZX9.rdf	format	application/rdf+xml
	type	Dataset
		Dataset
		Distribution
	download URL	C8UZX9.rdf
C8UZX9.html	format	text/html
	type	Dataset
		Distribution
	download URL	C8UZX9.html

<FAIR metadata/>

foaf:primaryTopic **Record R**

dcat:Distribution_1
Source URL_U1
format rdf+xml
dcat:Distribution_2
Source URL_U2
format application/xml





UniProt Protein C8UZX9	bibliographic Citation	The UniProt Consortium (2015). UniProt: a hub for protein information. Nucleic Acids Res. 43: D204-D212
	creator	UniProt Consortium
	language	eng
	license	cc by nd3.0
	title	UniProt Protein C8UZX9
	in Dataset	Uni Prot Accessor/
	contact Point	contact
	description	KRR1 small subunit processome componentKRR-R motif-containing protein 1
	distribution	C8UZX9.rdf C8UZX9.html
	identifier	C8UZX9
	keyword	Annotation Aspergillus nidulans Aspergillus Functinal Annotation GO Gene Ontology Proteins RNA Processing
	landing Page	uniprot.org
	language	en
	publisher	uniprot.org
	page	sparql uniprot.org/
	primary topic	C8UZX9
C8UZX9	...	
C8UZX9.rdf	format	application/rdf+xml
	type	Dataset Dataset Distribution
	download URL	C8UZX9.rdf
C8UZX9.html	format	text/html
	type	Dataset Distribution
	download URL	C8UZX9.html

Note that this metadata record is somewhat more FAIR, than what you can (easily) retrieve from UniProt itself!

e.g. the UniProt record does not include the citation or license information - you have to manually surf around the UniProt Web page to find that.

So the Accessor makes UniProt’s already notably FAIR data, even more FAIR (with respect to “R”)

How FAIR are we now?

What does the Accessor give us?

What we have achieved

F

We have created a FAIR record for something - i.e. a slice of a database - that was, historically, un-recordable and un-identifiable in any formal way.

F + R

Accessors are a standard approach to providing human & machine accessible metadata to facilitate appropriate discovery (contextual, biological), proper usage (license) and proper citation for any kind of data.

F + A

The discovery, accessibility, and drill-down/up behaviors do not require any novel API, rather simply rely on global Web standards; this allows them to be indexed by existing Web search engines

What we have achieved

I The metadata itself uses machine-accessible syntaxes, and widely adopted ontologies and vocabularies, thus easily integrates with other metadata

A Accessors provide a lightweight means to protect privacy while still providing the maximum degree of transparency possible

+ Accessors can be static, or dynamic. i.e. we can provide template Accessor file(s) that are edited in Notepad, then published together with the data; or Accessors can dynamically generate their output from code (e.g. layered on a database server)

So far, we have focused on FAIR Metadata

Are there approaches to making the DATA FAIR?

FAIR Projection:

Providing FAIR Data from non-FAIR Data

Dynamically

Imagine the data
we need to integrate
is in a CSV file
in FigShare or Zenodo

How do we discover and integrate <u>that</u> data?

Things we need to do:

We need a way to query “opaque” data blobs (like CSV) about their content

We need a way to retrieve that content in a FAIR format

We need, therefore, to model semantics for that opaque data content

We need to model various semantics for that content (one “size” doesn’t fit all!)

We need to associate those semantic models with a record or record-sets

We need a way to query those semantics determine which “size” fits our req’s

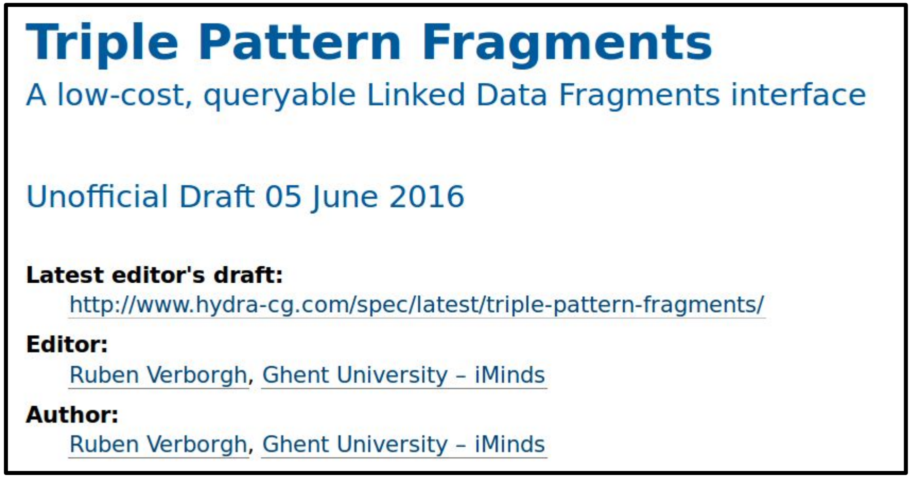
We would like to reuse semantic definitions as much as possible

We need to do all of this without creating a new API :-)

Triple Pattern Fragments

+

RDF Mapping Language



Triple Pattern Fragments

A low-cost, queryable Linked Data Fragments interface

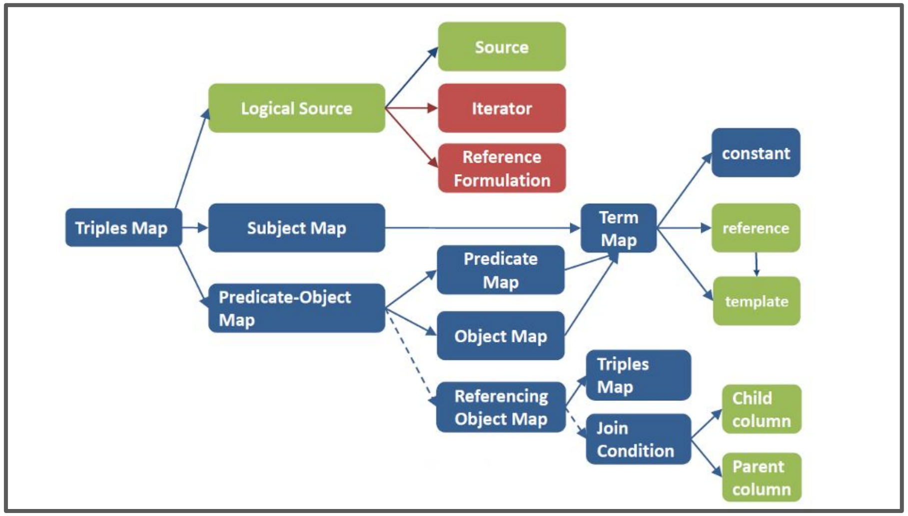
Unofficial Draft 05 June 2016

Latest editor's draft:
http://www.hydra-cg.com/spec/latest/triple-pattern-fragments/

Editor:
Ruben Verborgh, Ghent University – iMinds

Author:
Ruben Verborgh, Ghent University – iMinds

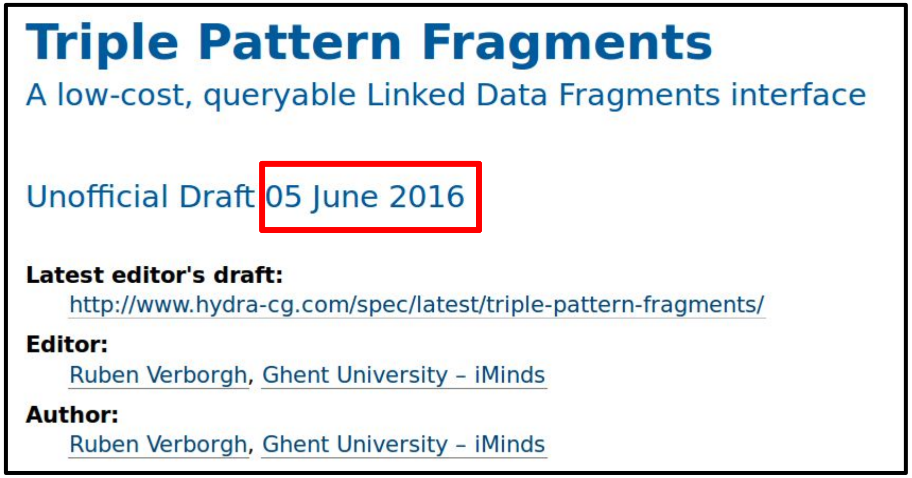
Ruben Verborgh
Ghent University



Anastasia Dimou
Ghent University

Triple Pattern Fragments (TPF)

A REST interface for requesting/retrieving RDF Triples
(from any source)



Triple Pattern Fragments

A low-cost, queryable Linked Data Fragments interface

Unofficial Draft 05 June 2016

Latest editor's draft:
http://www.hydra-cg.com/spec/latest/triple-pattern-fragments/

Editor:
Ruben Verborgh, Ghent University – iMinds

Author:
Ruben Verborgh, Ghent University – iMinds

Ruben Verborgh

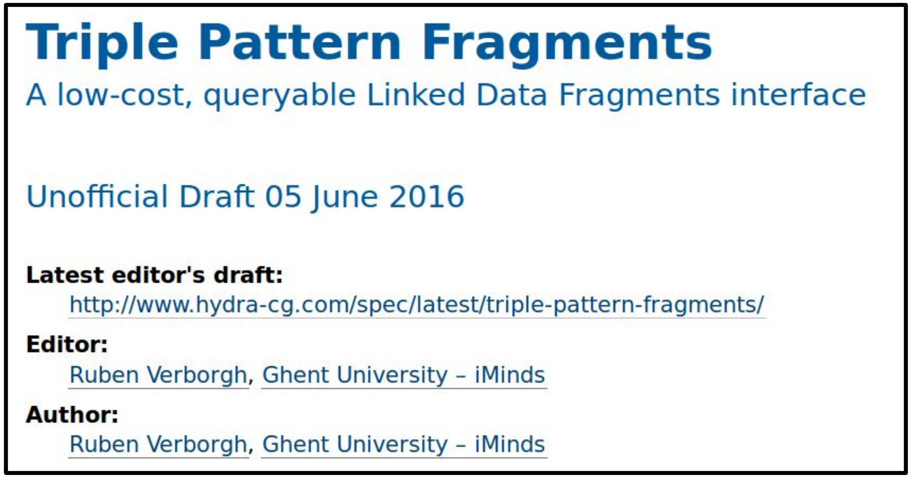
"Slices" of data, from any source, are considered Resources and are therefore represented by a distinct URL:

`http://some.database.org/dataset?s=___;p=___;o=___`

Calling HTTP GET on a TPF URL returns the set of Triples matching {?s, ?p, ?o} PLUS hypermedia instructions and Resource URLs for other relevant slices.

Triple Pattern Fragments (TPF)

A REST interface for retrieving RDF Triples
(from any source)



Triple Pattern Fragments

A low-cost, queryable Linked Data Fragments interface

Unofficial Draft 05 June 2016

Latest editor's draft:
http://www.hydra-cg.com/spec/latest/triple-pattern-fragments/

Editor:
Ruben Verborgh, Ghent University – iMinds

Author:
Ruben Verborgh, Ghent University – iMinds

Ruben Verborgh

For example, the "BMI" column from a patient registry is a Resource with the URL:

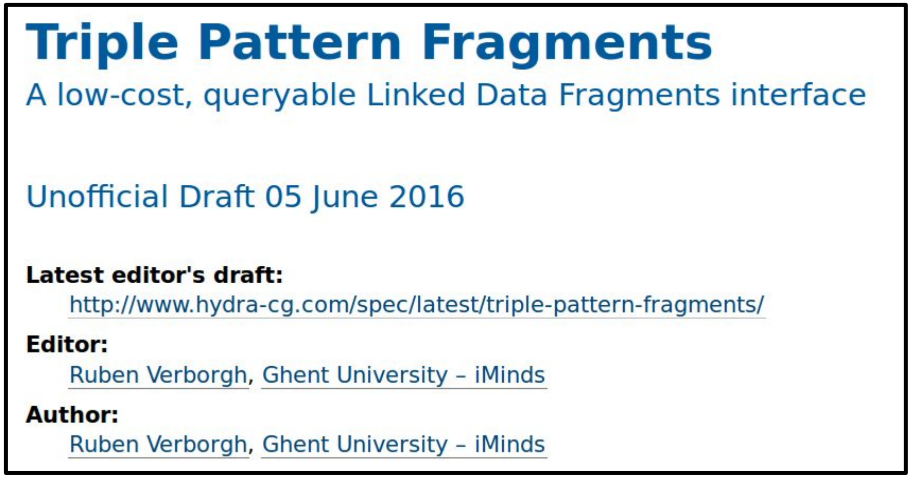
`http://my.registry.org/patients?p=CMO:0000105` **(CMO:0000105 = "body mass index"")**

HTTP GET gives me all BMI triples in the registry, together with other Resource URLs representing other "slices" that might be useful, for example:

`http://my.registry.org/patients?p=CMO:0000004` **(CMO:0000004 = "systolic B.P.")**

Triple Pattern Fragments (TPF)

A REST interface for retrieving RDF Triples
(from any source)



Triple Pattern Fragments

A low-cost, queryable Linked Data Fragments interface

Unofficial Draft 05 June 2016

Latest editor's draft:
http://www.hydra-cg.com/spec/latest/triple-pattern-fragments/

Editor:
Ruben Verborgh, Ghent University – iMinds

Author:
Ruben Verborgh, Ghent University – iMinds

Ruben Verborgh

For example, the "BMI" column from a patient registry is a Resource with the URL:

`http://my.regist` 0105 = "body mass index"")

HTTP GET gives me other Resource URLs
represe r example:

{ REST }

`http://my.registry.org/patients?p=CMO:0000004` (CMO:0000004 = "systolic B.P.")

We have a standard, RESTful way to <u>request</u> triples from any data source

i.e. every slice of every dataset will be considered a distinct Resource

→ simply call HTTP GET on that Resource to get the Triples

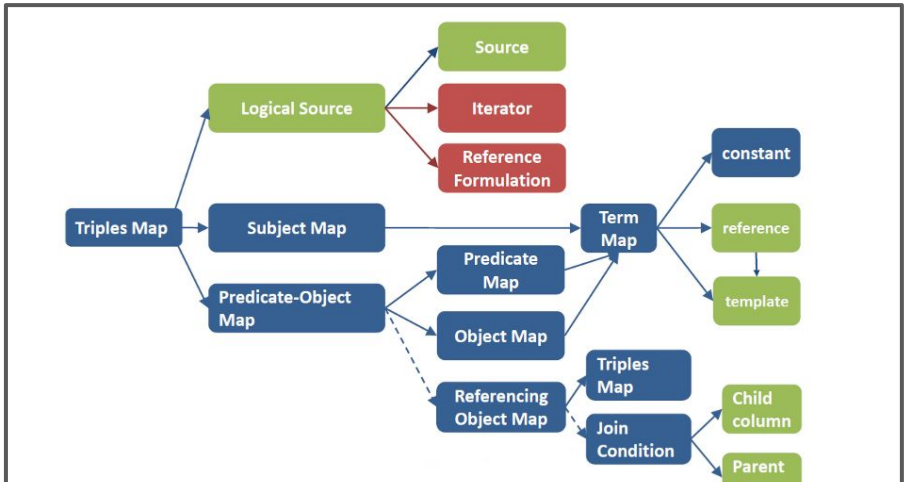
But…

We have no way to know what TPF Resources are available for any given dataset or what those Resources "are"

(proteins? genes? patients? articles?)

RML

A way to describe the structure of an RDF document



http://rml.io/RMLmappingLanguage.html

Anastasia Dimou

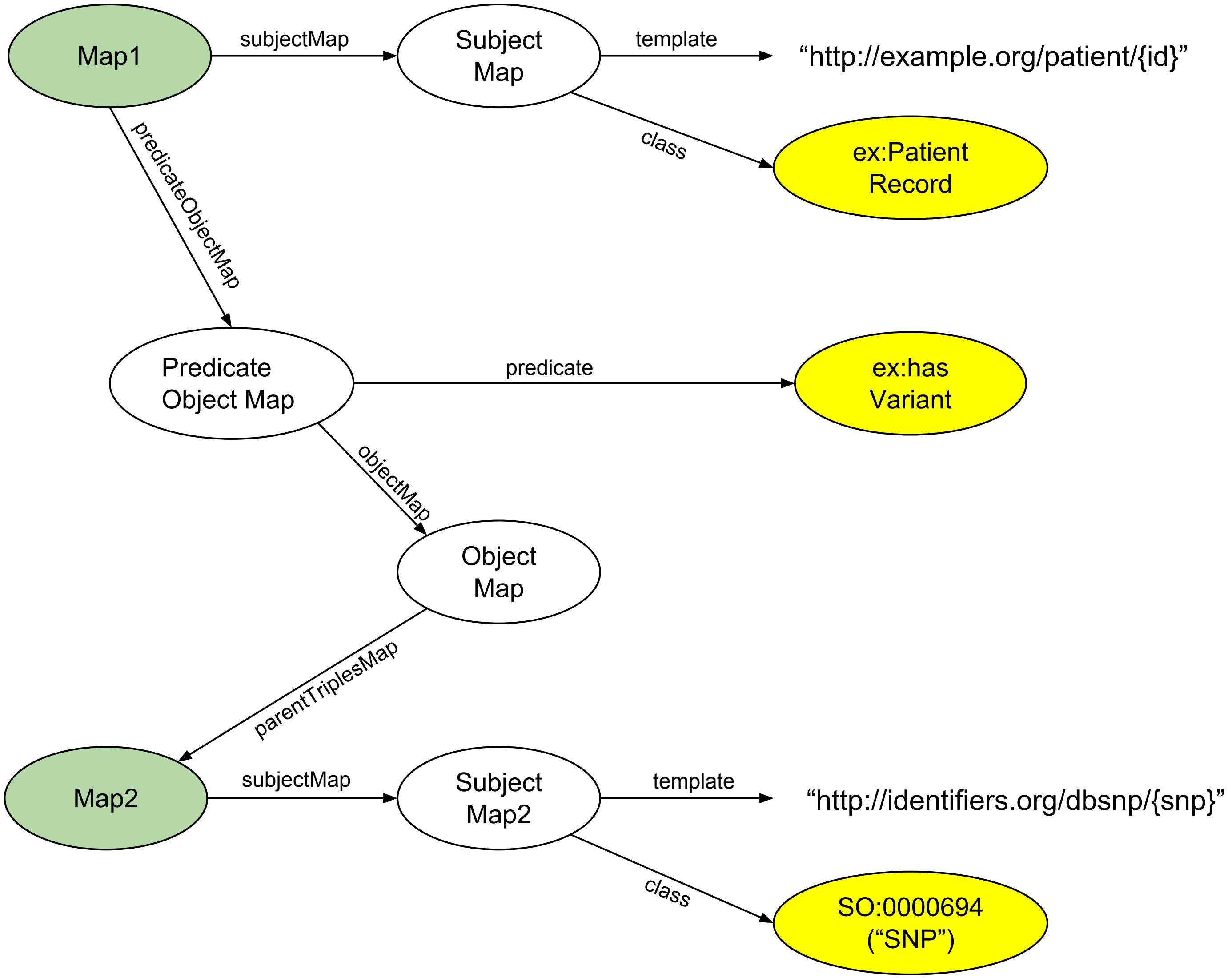
RML allows us to create models of (meta)data structures

“What *could* this data look like, if it were mapped to RDF?”

RML fulfills similar objectives to DCAT Profiles, the Dublin Core Application Profile, and ISO 11179 - Metadata Registries; but has added advantages

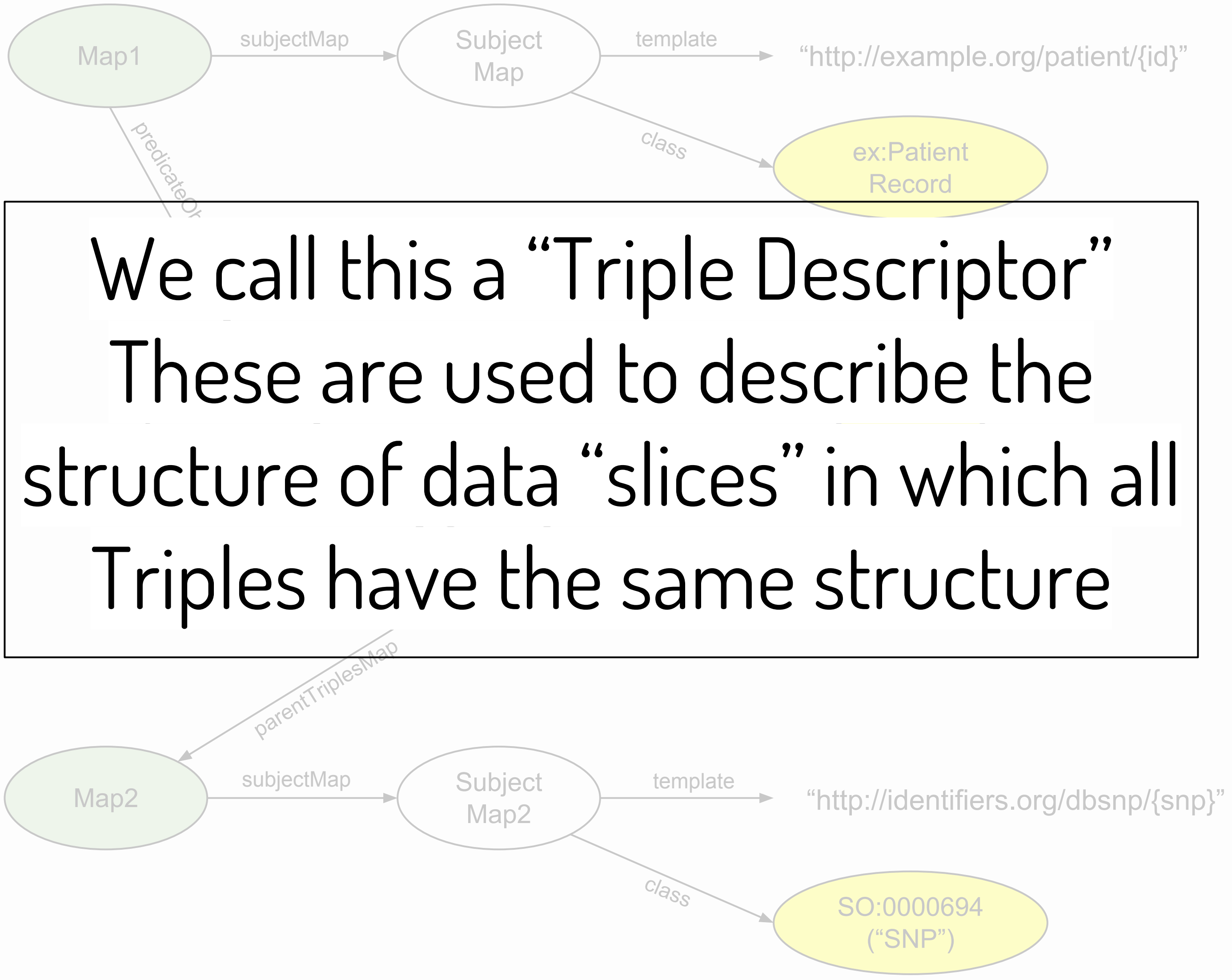
Using RML to describe the structure and semantics of a single Triple

THE MODEL



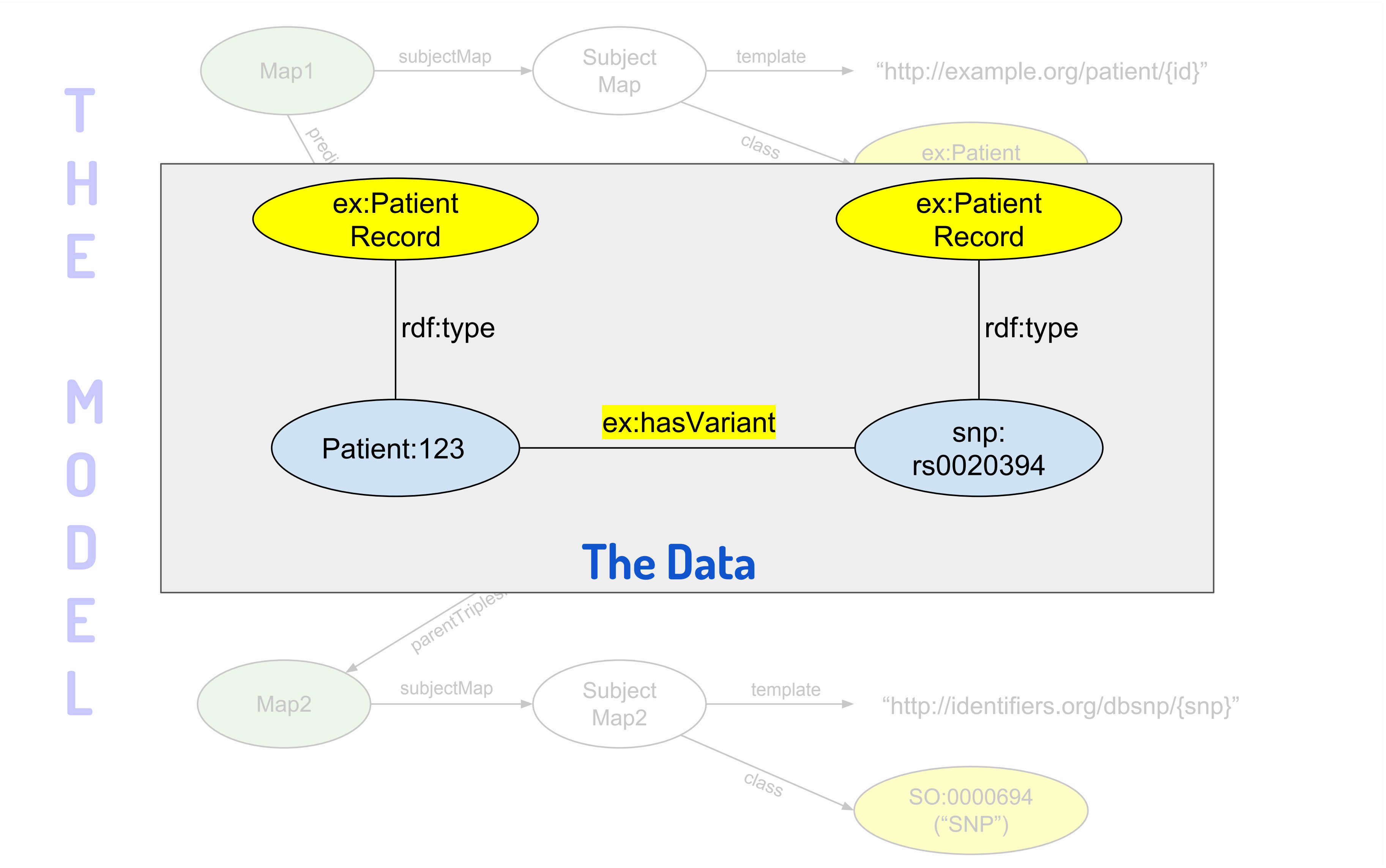
Using RML to describe the structure and semantics of a <u>single Triple</u>

THE MODEL



We call this a "Triple Descriptor"
These are used to describe the structure of data "slices" in which all Triples have the same structure

Using RML to describe the structure and semantics of a single Triple



THE MODEL

Map1

subjectMap

Subject Map

template

"http://example.org/patient/{id}"

predi

class

ex:Patient

ex:Patient Record

rdf:type

ex:Patient Record

rdf:type

Patient:123

ex:hasVariant

snp: rs0020394

The Data

parentTriples

Map2

subjectMap

Subject Map2

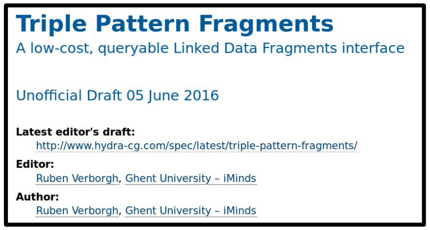
template

"http://identifiers.org/dbsnp/{snp}"

class

SO:0000694 ("SNP")

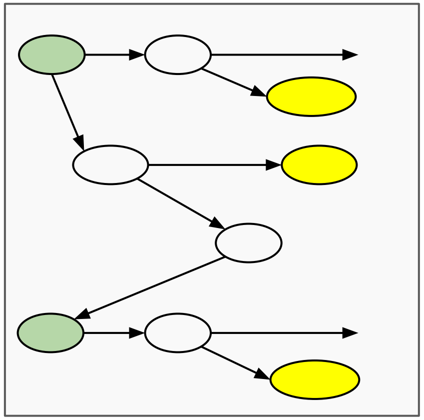
Where are we now?


Triple Pattern Fragments
A low-cost, queryable Linked Data Fragments interface

Unofficial Draft 05 June 2016

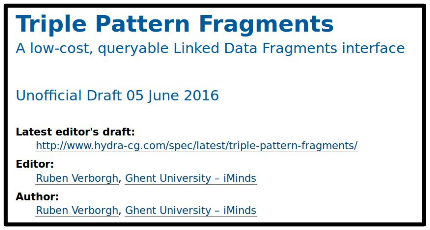
Latest editor's draft:
http://www.hydra-cg.com/spec/latest/triple-pattern-fragments/
Editor:
Ruben Verborgh, Ghent University – iMinds
Author:
Ruben Verborgh, Ghent University – iMinds

TPF – A standard, RESTful way to request Triples



Triple Descriptors – A standard way to describe the structure and meaning of a Triple

Where are we now?



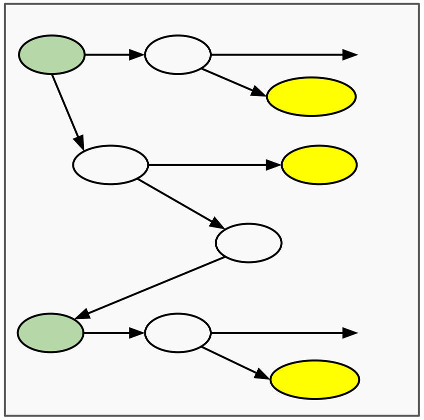
Triple Pattern Fragments

A low-cost, queryable Linked Data Fragments interface

Unofficial Draft 05 June 2016

Latest editor's draft:
http://www.hydra-cg.com/spec/latest/triple-pattern-fragments/
Editor:
Ruben Verborgh, Ghent University – iMinds
Author:
Ruben Verborgh, Ghent University – iMinds

TPF – A standard, RESTful way to request Triples



Triple Descriptors – A standard way to describe the structure and meaning of a Triple

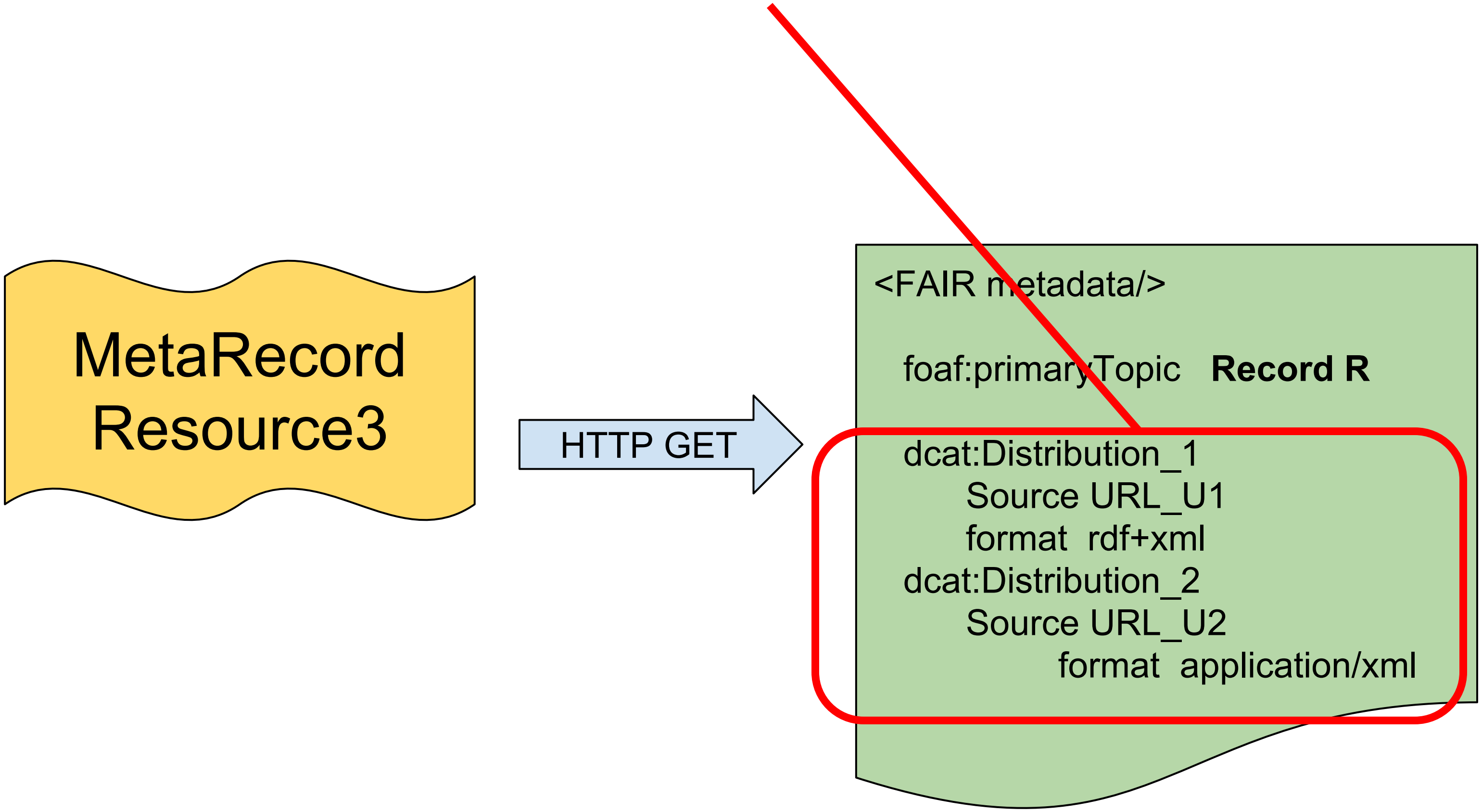
We need a way to associate these with each other

We need a way to associate these with a dataset or record

Luckily, we have already solved this!

The FAIR Accessor can do this

Using the metadata structures defined by DCAT the FAIR Accessor also tells you how to get the content of the record, and what formats are available



MetaRecord
Resource3

HTTP GET

<FAIR metadata/>

foaf:primaryTopic **Record R**

dcat:Distribution_1
Source URL_U1
format rdf+xml
dcat:Distribution_2
Source URL_U2
format application/xml

If we consider the TPF Resource URL to be just another DCAT Distribution, we get...

<FAIR metadata/>

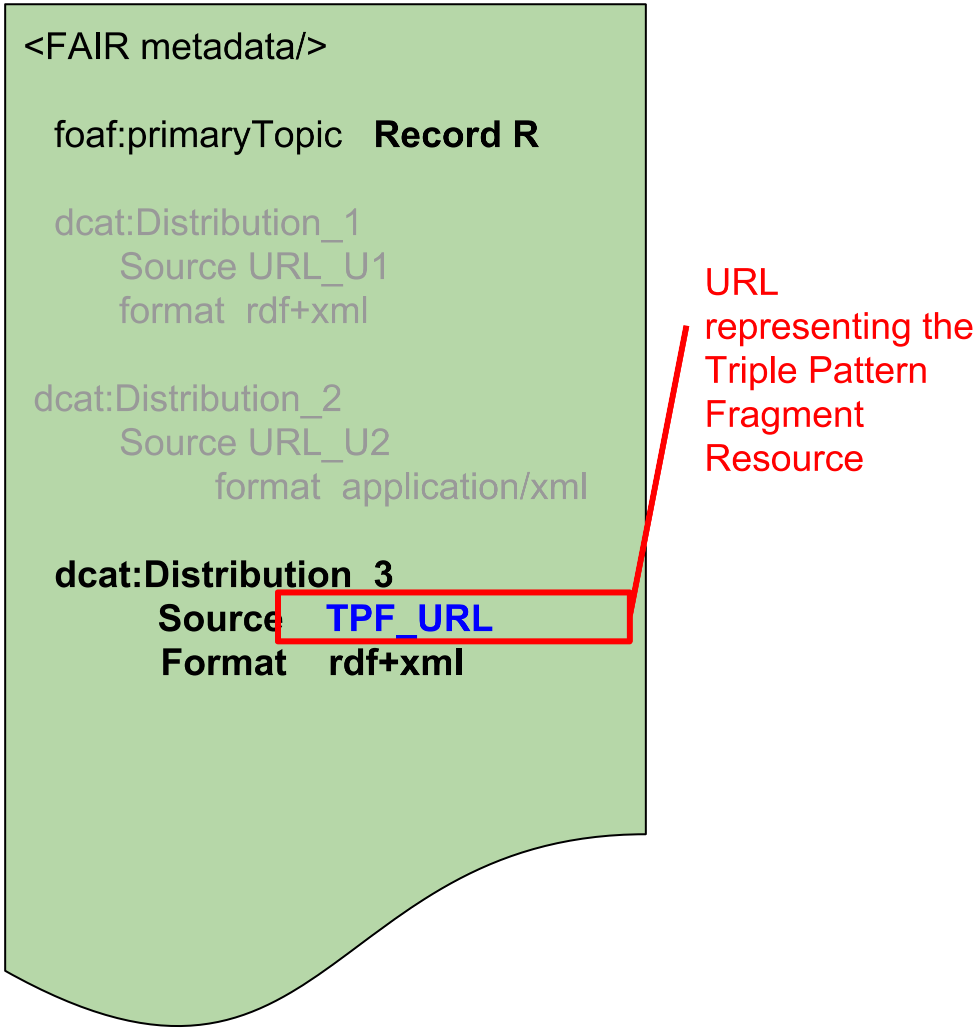
foaf:primaryTopic **Record R**

dcat:Distribution_1
Source URL_U1
format rdf+xml

dcat:Distribution_2
Source URL_U2
format application/xml

dcat:Distribution_3
Source TPF_URL
Format rdf+xml

If we consider the TPF Resource URL to be just another DCAT Distribution, we get...



<FAIR metadata/>

foaf:primaryTopic **Record R**

dcat:Distribution_1
Source URL_U1
format rdf+xml

dcat:Distribution_2
Source URL_U2
format application/xml

dcat:Distribution_3
Source TPF_URL
Format rdf+xml

URL representing the Triple Pattern Fragment Resource

If we consider the TPF Resource URL to be just another DCAT Distribution, we get... now add the Triple Descriptor

<FAIR metadata/>

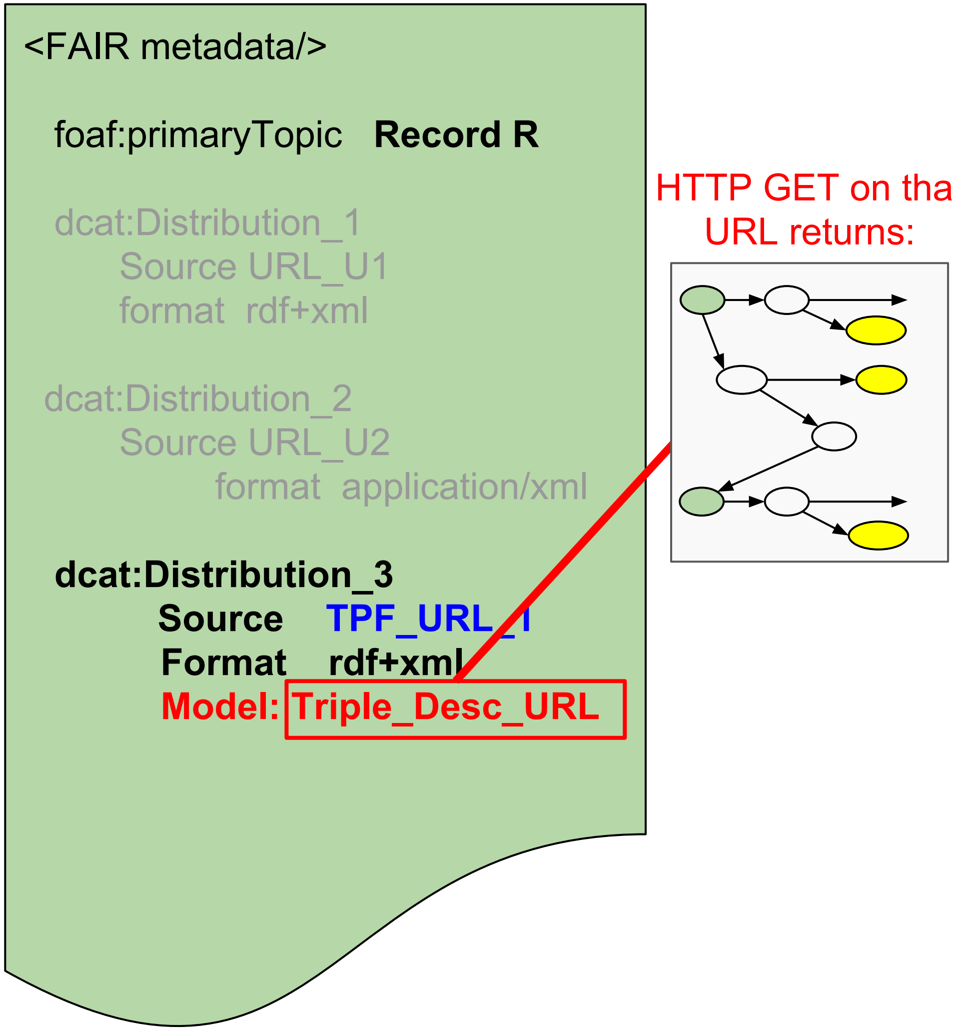
foaf:primaryTopic **Record R**

dcat:Distribution_1
Source URL_U1
format rdf+xml

dcat:Distribution_2
Source URL_U2
format application/xml

dcat:Distribution_3
Source TPF_URL
Format rdf+xml
Model: Triple_Desc_URL

If we consider the TPF Resource URL to be just another DCAT Distribution, we get... now add the Triple Descriptor



<FAIR metadata/>

foaf:primaryTopic **Record R**

dcat:Distribution_1
Source URL_U1
format rdf+xml

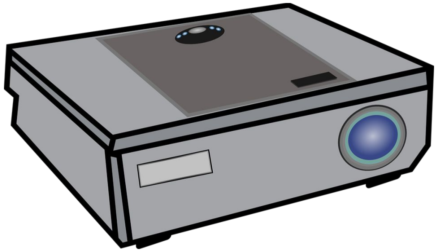
dcat:Distribution_2
Source URL_U2
format application/xml

dcat:Distribution_3
Source TPF_URL_1
Format rdf+xml
Model: Triple_Desc_URL

HTTP GET on that URL returns:

If we consider the TPF Resource URL to be just another DCAT Distribution, we get... now add the Triple Descriptor

Record
+
TPF Server
+
RML Model
=



FAIR Projector

<FAIR metadata/>

foaf:primaryTopic **Record R**

dcat:Distribution_1
Source URL_U1
format rdf+xml

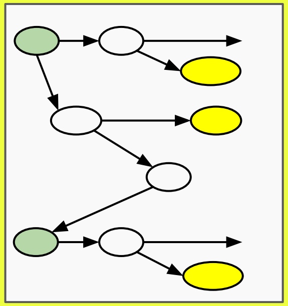
dcat:Distribution_2
Source URL_U2
format application/xml

dcat:Distribution_3
Source TPF_URL
Format rdf+xml
Model: Triple_Desc_URL

If we consider the TPF Resource URL to be just another DCAT Distribution, we get... now add the Triple Descriptor

HTTP GET on **TPF_URL**
returns **rdf+xml** triples from **Record R**

That look like



<FAIR metadata/>

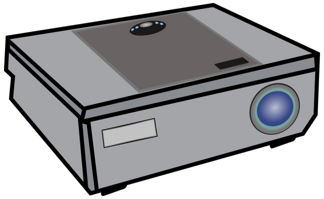
foaf:primaryTopic **Record R**

dcat:Distribution_1
Source URL_U1
format rdf+xml

dcat:Distribution_2
Source URL_U2
format application/xml

dcat:Distribution_3
Source TPF_URL
Format rdf+xml
Model: Triple_Desc_URL

Interoperability without Brute Force

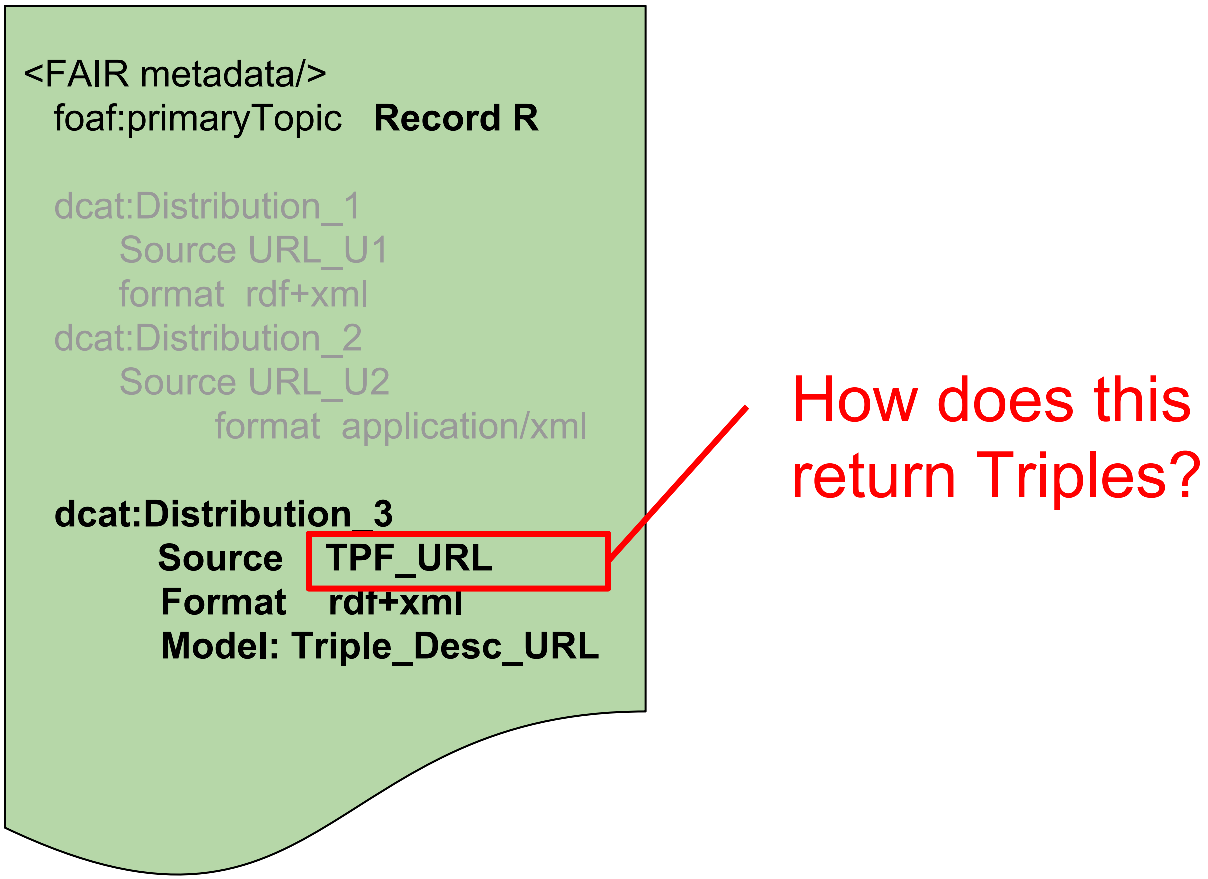


I hear you objecting... I skipped something important!!!

We still have not defined a way to CREATE these triples

I hear you objecting… I skipped something important!!!

We still have not defined a way to CREATE these triples



<FAIR metadata/>
foaf:primaryTopic **Record R**

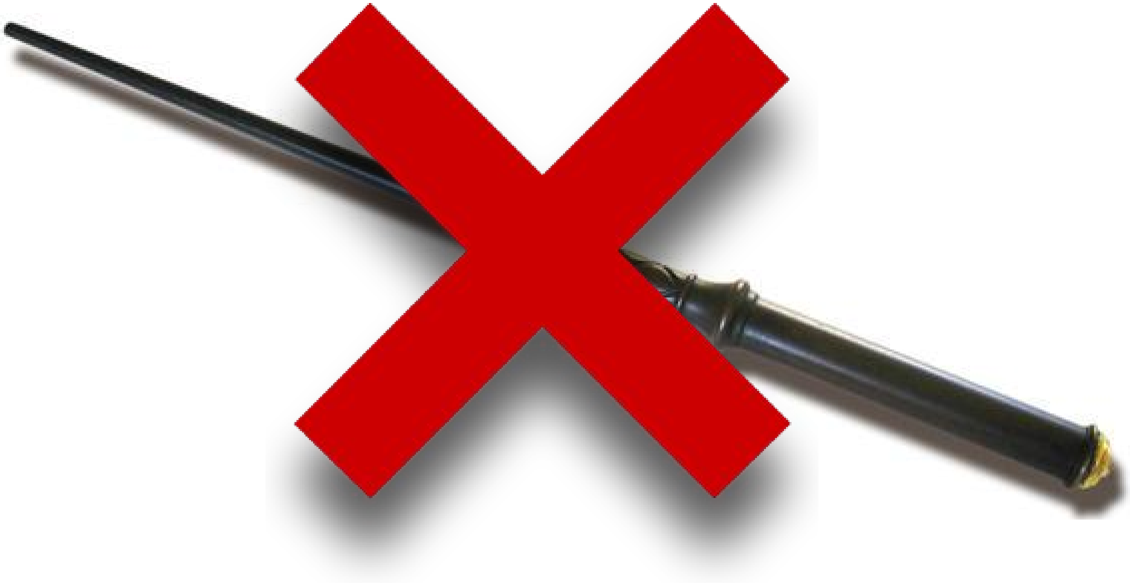
dcat:Distribution_1
Source URL_U1
format rdf+xml
dcat:Distribution_2
Source URL_U2
format application/xml

dcat:Distribution_3
Source TPF_URL
Format rdf+xml
Model: Triple_Desc_URL

How does this return Triples?

Sadly, there is no magic wand to create interoperability

Sadly, there is no magic wand to create interoperability



Someone has to write the TPF server that converts the data

Interoperability will never come “for free”

(because semantics will never come “for free”)

However, there are reasons for optimism!

1. **Researchers transform data <u>anyway</u>**

2. For the most common file formats (e.g. CSV or Excel), there are RML-based tools **to automate the RDF transformation**

3. **Investing time into creating an RML model is more FAIR than *ad hoc* "re-useless" brute-force transformation**. When you create a FAIR Projector for your own data transformation needs, **it is reusable!**

However, there are reasons for optimism!

4. **Citations Citations Citations!**
   FAIR Accessors/Projectors are themselves FAIR objects - You can get credit if other people use your Projected data for their analyses

5. **RML Triple Descriptors are very simple** (one triple!) so we can also **templatize** their construction → creating a FAIR Projector is quite easy in many cases!

How simple?

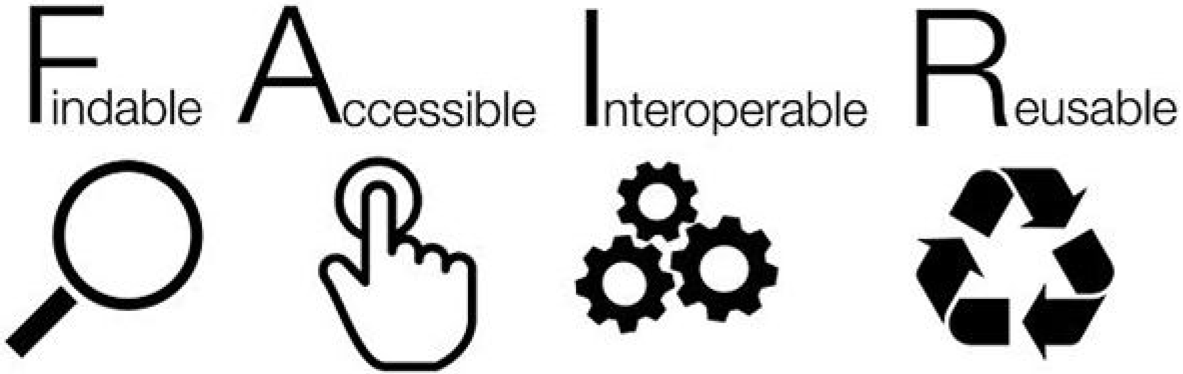
I created this last month
at the BioHackathon in Japan.

It’s the first prototype for
end-user-generated
“FAIR Projection”

FAIR Projection

Goal: Change hard-to-reuse data into reusable data

- In a way that also makes it findable
- In a way that also makes it accessible
- In a way that also makes it interoperable



September 13, 2017

Exemplar Output to demonstrate FAIR Projection from a CSV file

Mark Wilkinson

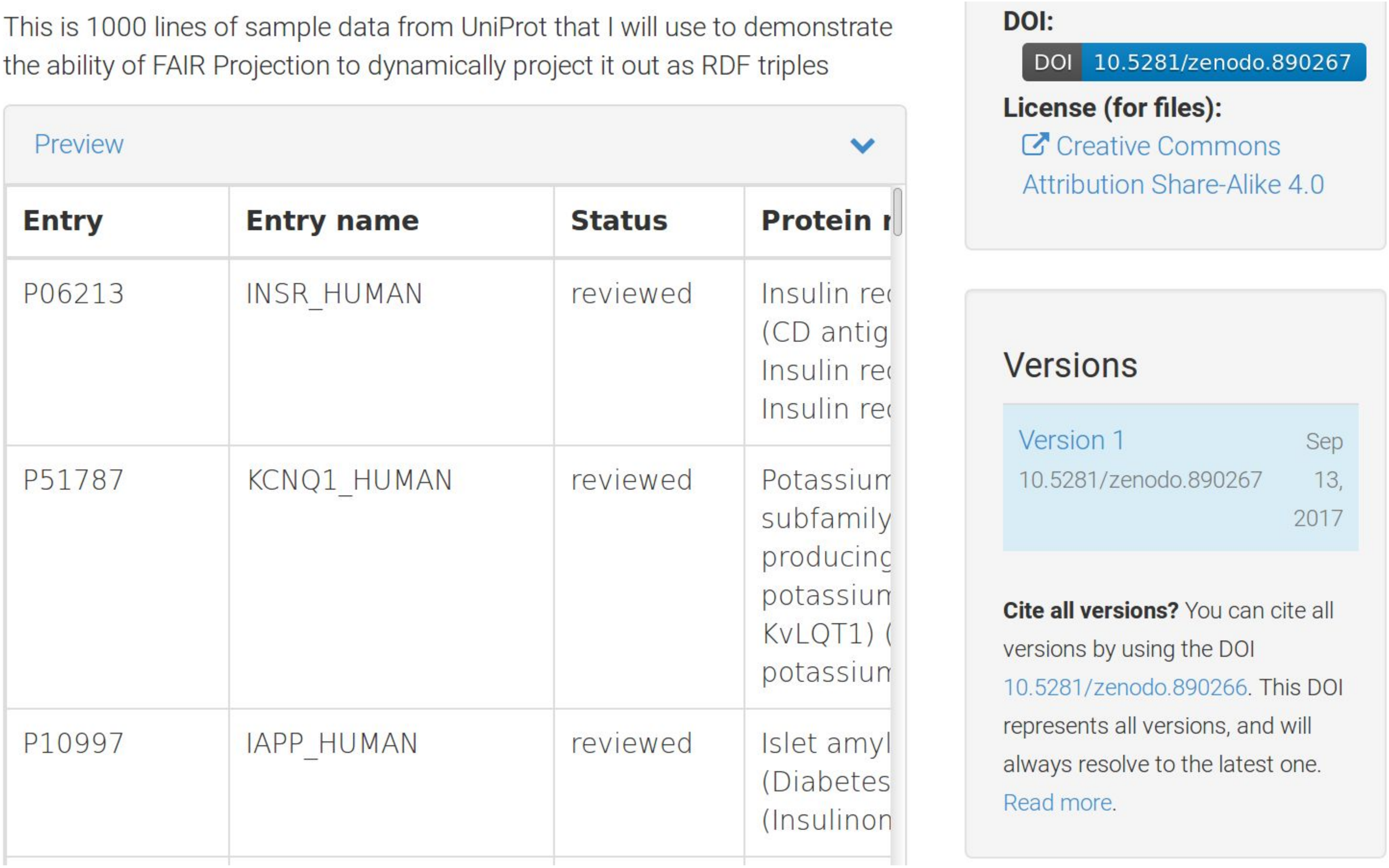
This is 1000 lines of sample data from UniProt that I will use to demonstrate the ability of FAIR Projection to dynamically project it out as RDF triples

Entry	Entry name	Status	Protein r
P06213	INSR_HUMAN	reviewed	Insulin re

Publication date:
September 13, 2017

DOI:
DOI 10.5281/zenodo.890267

License (for files):
Creative Commons Attribution Share-Alike 4.0



This is 1000 lines of sample data from UniProt that I will use to demonstrate the ability of FAIR Projection to dynamically project it out as RDF triples

Entry	Entry name	Status	Protein
P06213	INSR_HUMAN	reviewed	Insulin re (CD antig Insulin re Insulin re
P51787	KCNQ1_HUMAN	reviewed	Potassium subfamily producing potassium KvLQT1) (potassium
P10997	IAPP_HUMAN	reviewed	Islet amyl (Diabetes (Insulinon

DOI:

DOI 10.5281/zenodo.890267

License (for files):

Creative Commons Attribution Share-Alike 4.0

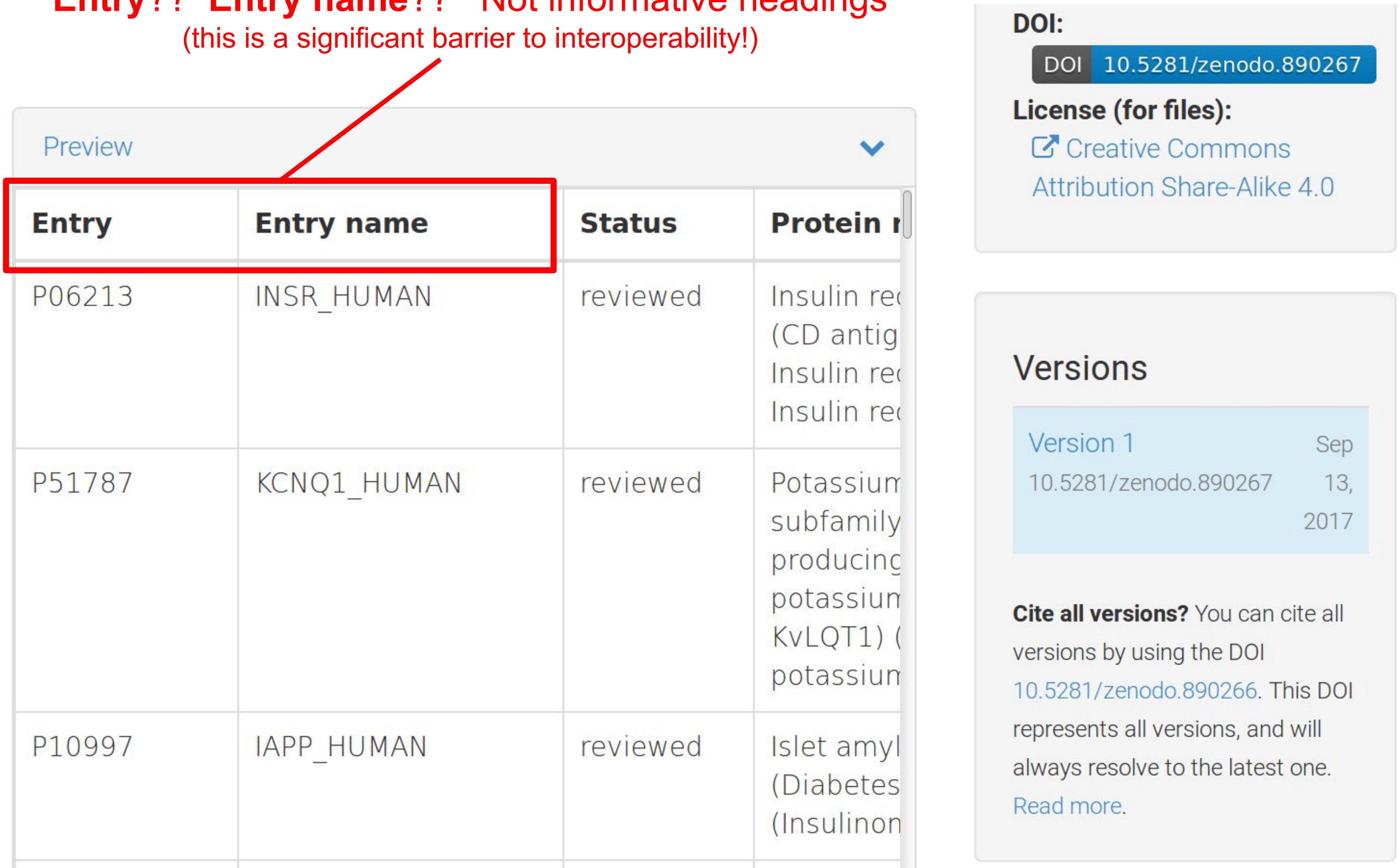
Versions

Version 1 10.5281/zenodo.890267	Sep 13, 2017

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.890266. This DOI represents all versions, and will always resolve to the latest one. Read more.

Entry?? **Entry name**?? Not informative headings

(this is a significant barrier to interoperability!)



Preview

Entry	Entry name	Status	Protein
P06213	INSR_HUMAN	reviewed	Insulin re (CD antig Insulin re Insulin re
P51787	KCNQ1_HUMAN	reviewed	Potassiu subfamily producin potassiu KvLQT1) (potassiu
P10997	IAPP_HUMAN	reviewed	Islet amyl (Diabetes (Insulino

DOI:

DOI 10.5281/zenodo.890267

License (for files):

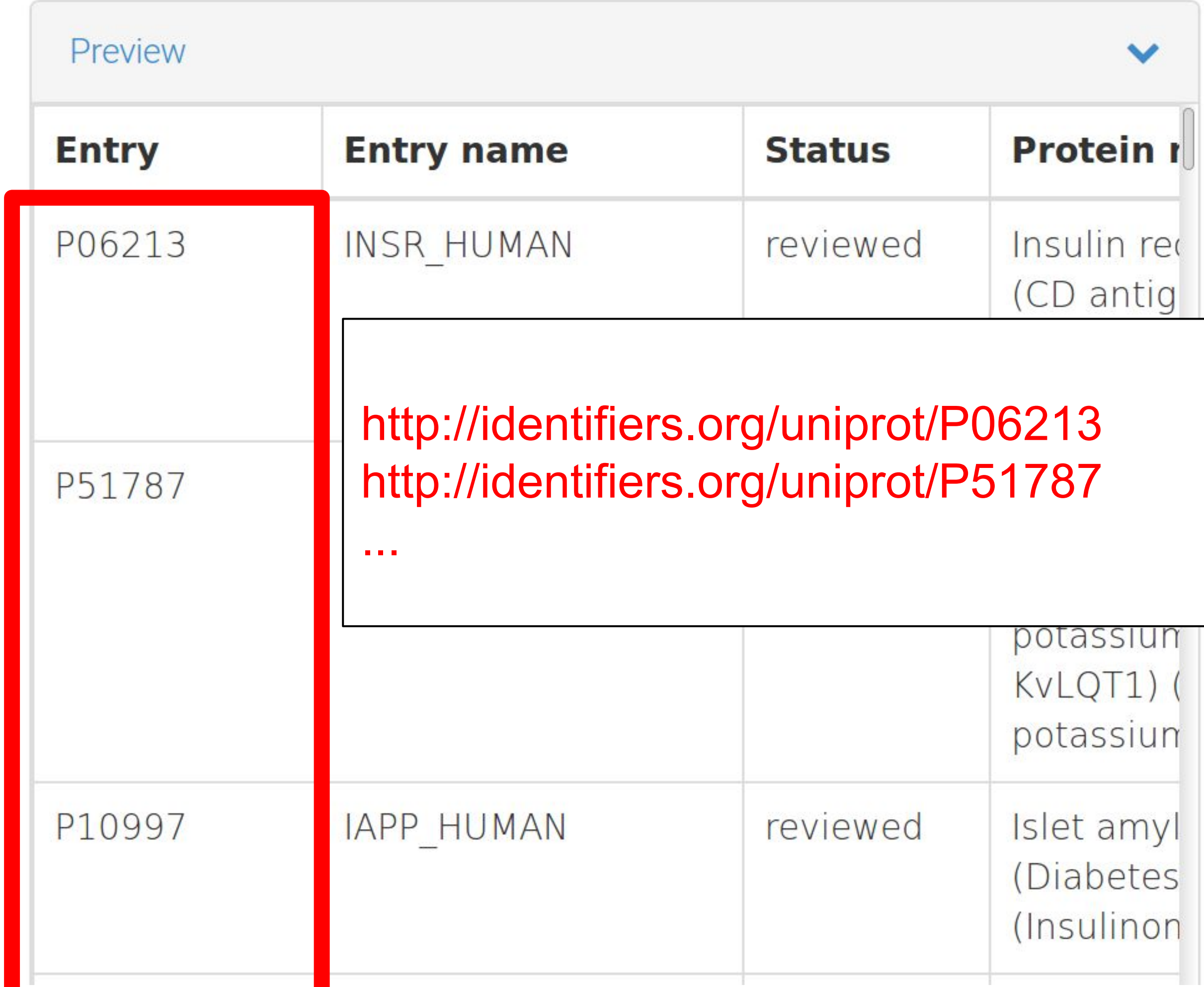
Creative Commons Attribution Share-Alike 4.0

Versions

Version 1 10.5281/zenodo.890267 Sep 13, 2017

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.890266. This DOI represents all versions, and will always resolve to the latest one. Read more.

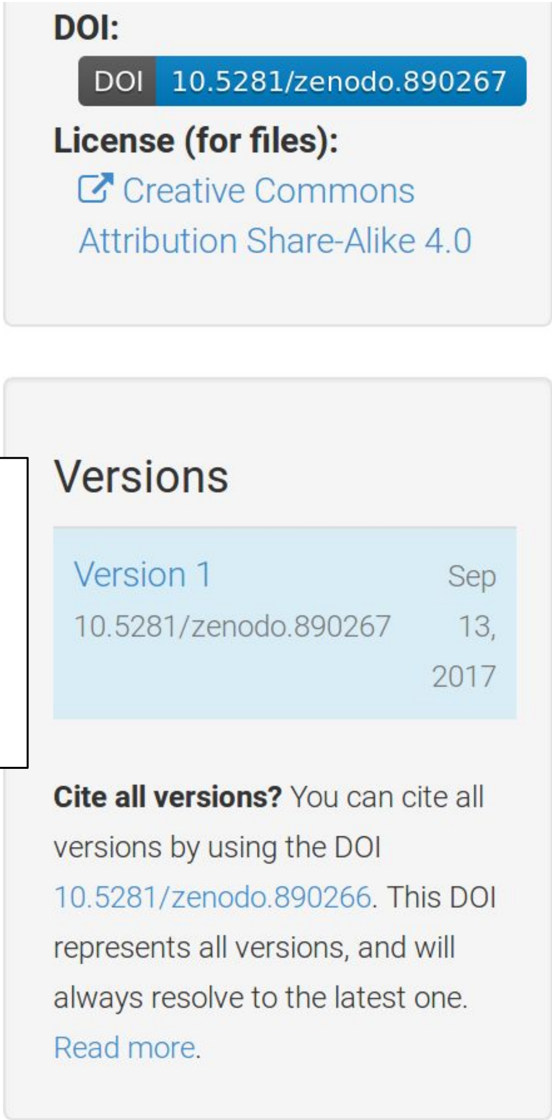

That's what I do. I drink, and I know things.



Preview

Entry	Entry name	Status	Protein r
P06213	INSR_HUMAN	reviewed	Insulin re (CD antig
P51787			potassiun KvLQT1) (potassiun
P10997	IAPP_HUMAN	reviewed	Islet amyl (Diabetes (Insulinon

http://identifiers.org/uniprot/P06213
http://identifiers.org/uniprot/P51787
...



DOI:

DOI 10.5281/zenodo.890267

License (for files):

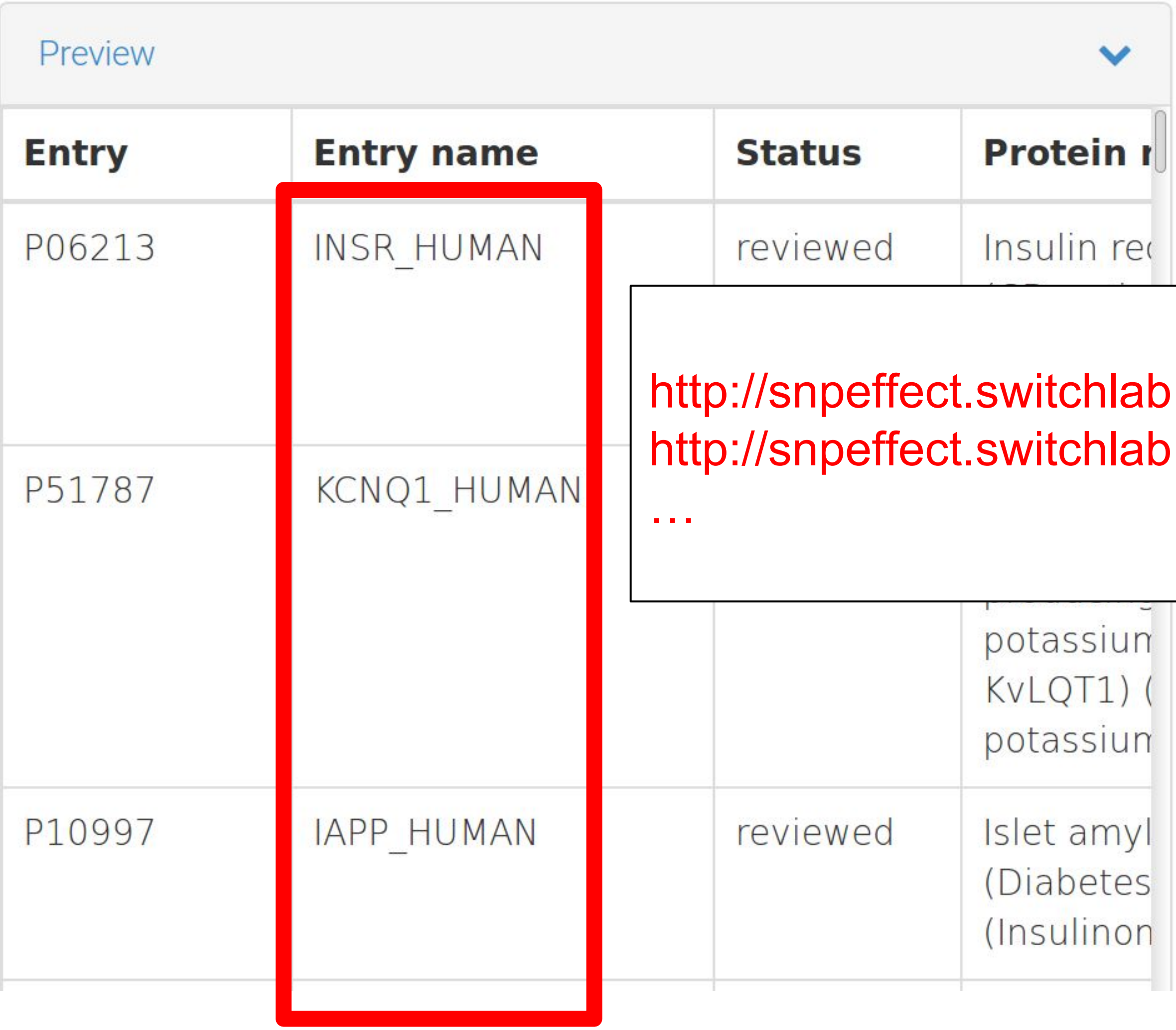
Creative Commons Attribution Share-Alike 4.0

Versions

Version 1 — 10.5281/zenodo.890267 — Sep 13, 2017

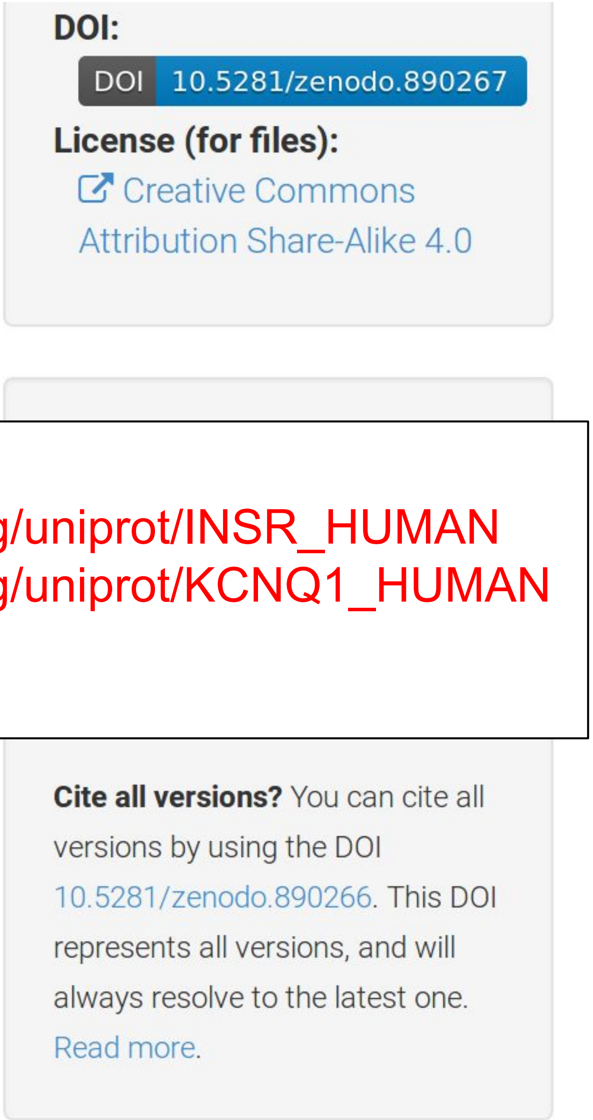
Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.890266. This DOI represents all versions, and will always resolve to the latest one. Read more.

SNPeffect 4.0 PHENOTYPING HUMAN MUTATIONS



Entry	Entry name	Status	Protein r
P06213	INSR_HUMAN	reviewed	Insulin re
P51787	KCNQ1_HUMAN		potassium KvLQT1) (potassium
P10997	IAPP_HUMAN	reviewed	Islet amyl (Diabetes (Insulinon

http://snpeffect.switchlab.org/uniprot/INSR_HUMAN
http://snpeffect.switchlab.org/uniprot/KCNQ1_HUMAN
…



DOI:

DOI 10.5281/zenodo.890267

License (for files):

Creative Commons Attribution Share-Alike 4.0

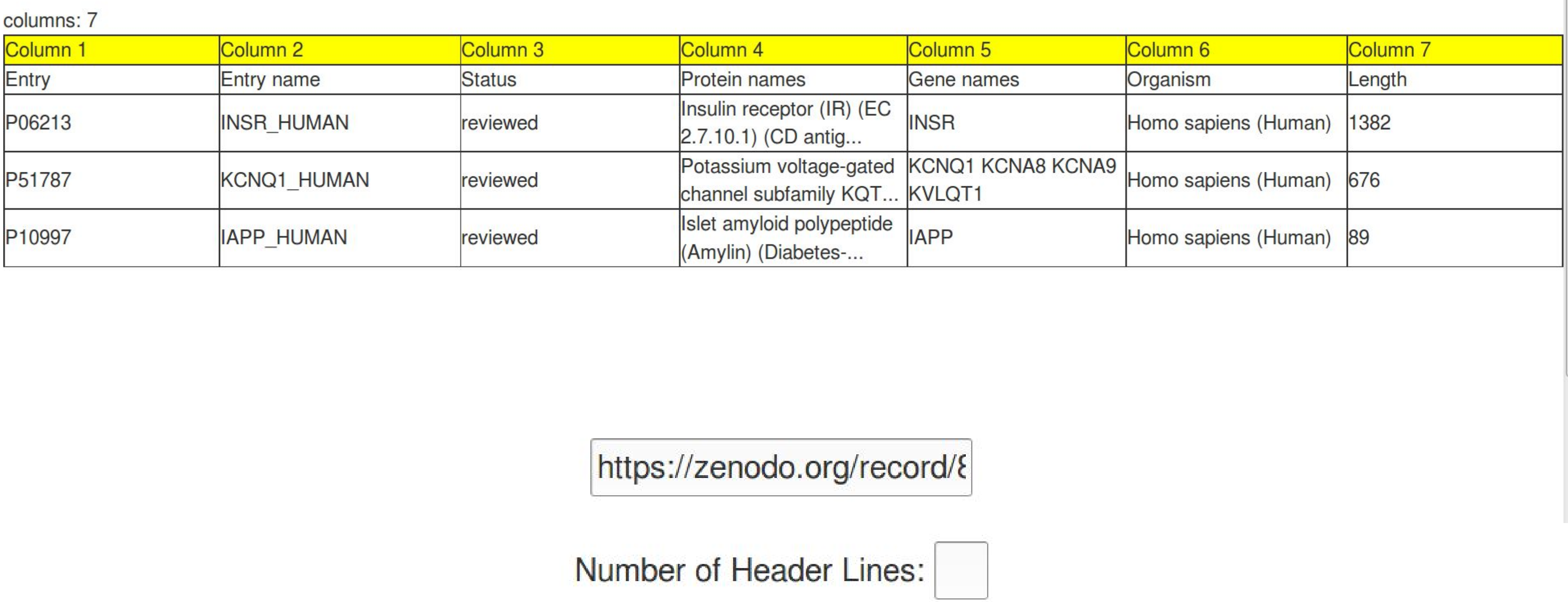
Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.890266. This DOI represents all versions, and will always resolve to the latest one. Read more.

FAIR Projector Builder

Please enter the URL of the tab-delimited file

https://zenodo.org/record/

Put the URL for the Zenodo data here



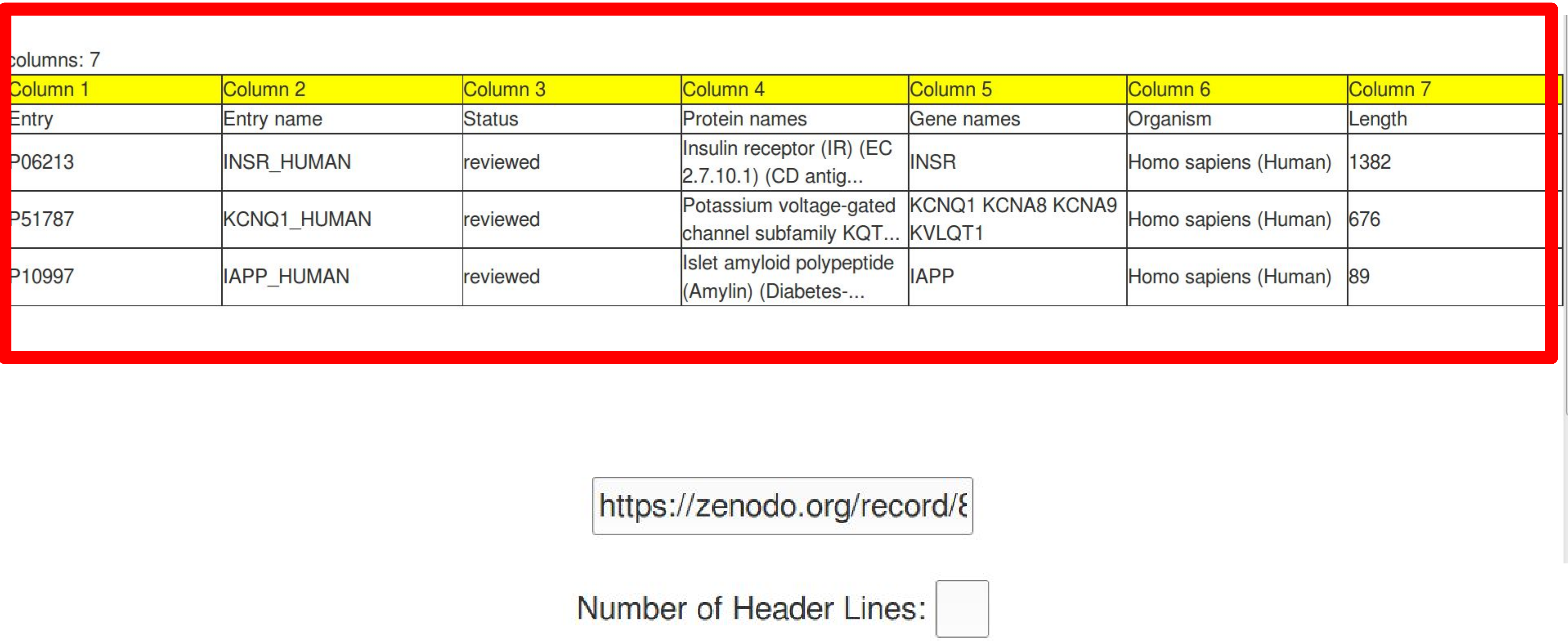
columns: 7

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7
Entry	Entry name	Status	Protein names	Gene names	Organism	Length
P06213	INSR_HUMAN	reviewed	Insulin receptor (IR) (EC 2.7.10.1) (CD antig...	INSR	Homo sapiens (Human)	1382
P51787	KCNQ1_HUMAN	reviewed	Potassium voltage-gated channel subfamily KQT...	KCNQ1 KCNA8 KCNA9 KVLQT1	Homo sapiens (Human)	676
P10997	IAPP_HUMAN	reviewed	Islet amyloid polypeptide (Amylin) (Diabetes-...	IAPP	Homo sapiens (Human)	89

https://zenodo.org/record/

Number of Header Lines:

The first few lines of the file, with columns numbered

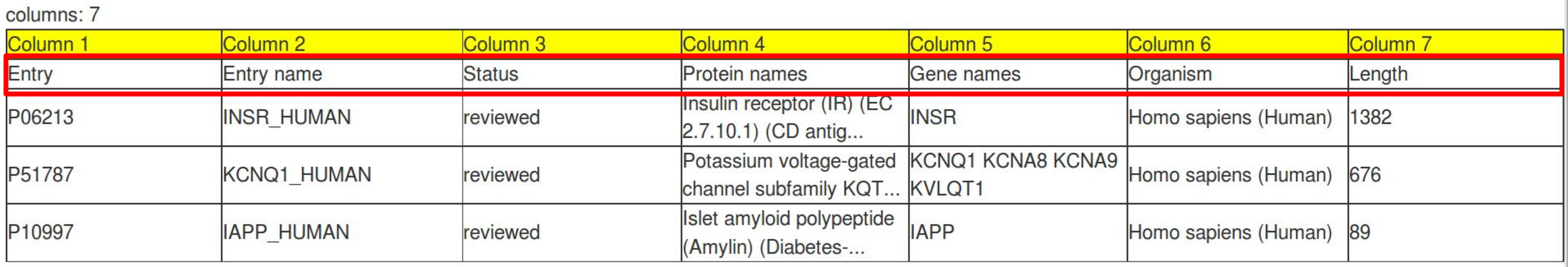


columns: 7

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7
Entry	Entry name	Status	Protein names	Gene names	Organism	Length
P06213	INSR_HUMAN	reviewed	Insulin receptor (IR) (EC 2.7.10.1) (CD antig...	INSR	Homo sapiens (Human)	1382
P51787	KCNQ1_HUMAN	reviewed	Potassium voltage-gated channel subfamily KQT...	KCNQ1 KCNA8 KCNA9 KVLQT1	Homo sapiens (Human)	676
P10997	IAPP_HUMAN	reviewed	Islet amyloid polypeptide (Amylin) (Diabetes-...	IAPP	Homo sapiens (Human)	89

https://zenodo.org/record/

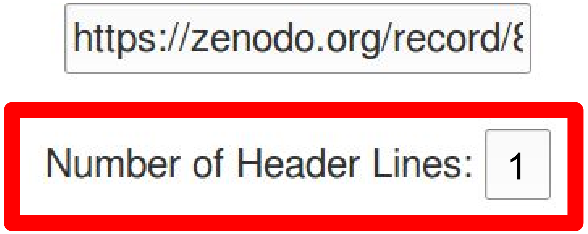
Number of Header Lines: ☐



columns: 7

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7
Entry	Entry name	Status	Protein names	Gene names	Organism	Length
P06213	INSR_HUMAN	reviewed	Insulin receptor (IR) (EC 2.7.10.1) (CD antig...	INSR	Homo sapiens (Human)	1382
P51787	KCNQ1_HUMAN	reviewed	Potassium voltage-gated channel subfamily KQT...	KCNQ1 KCNA8 KCNA9 KVLQT1	Homo sapiens (Human)	676
P10997	IAPP_HUMAN	reviewed	Islet amyloid polypeptide (Amylin) (Diabetes-...	IAPP	Homo sapiens (Human)	89

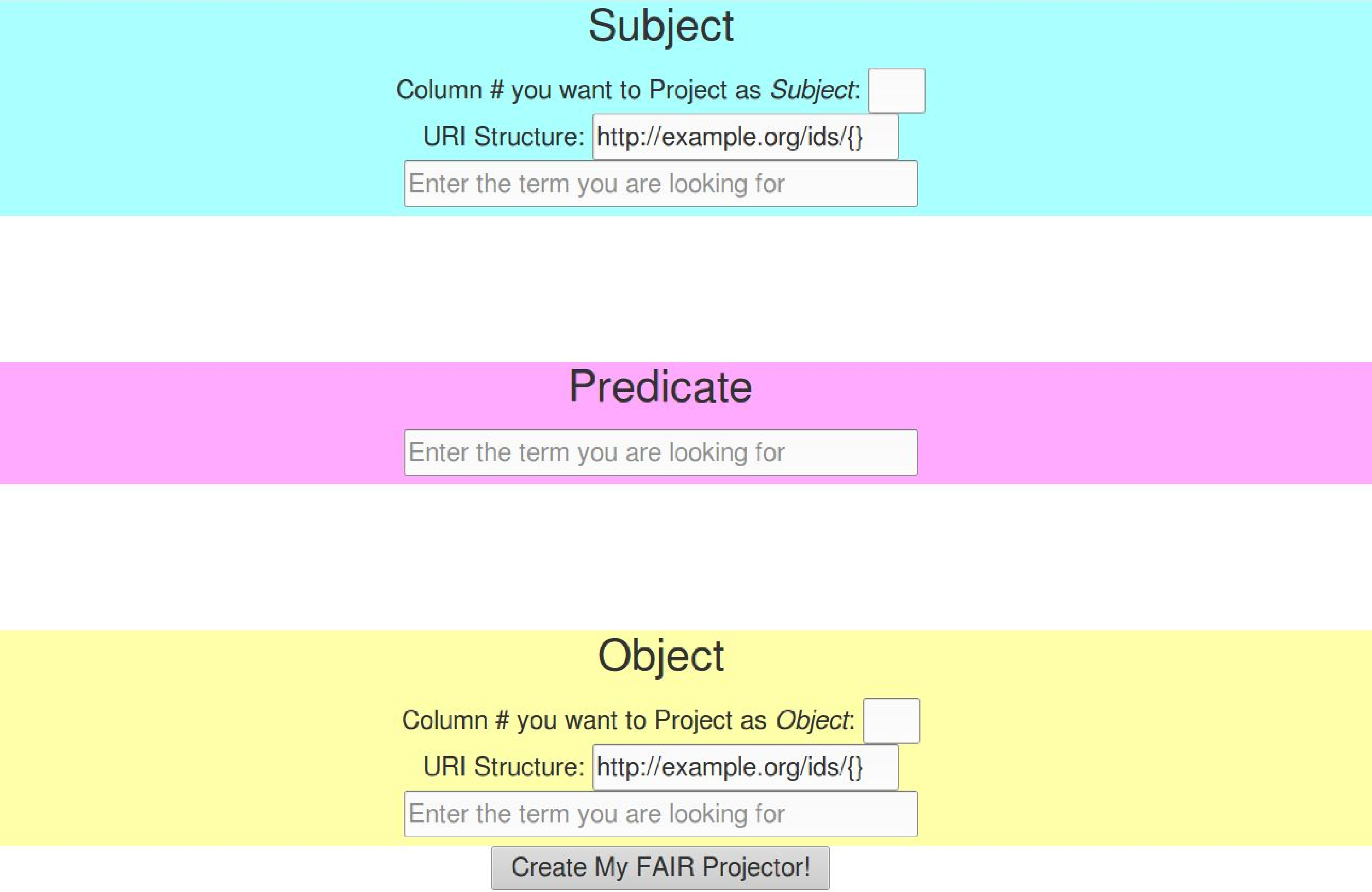
How many lines are headers? (in this case, one)



https://zenodo.org/record/

Number of Header Lines: 1

The next three questions tell the computer
how to interpret the two data columns of interest



Subject

Column # you want to Project as *Subject*:

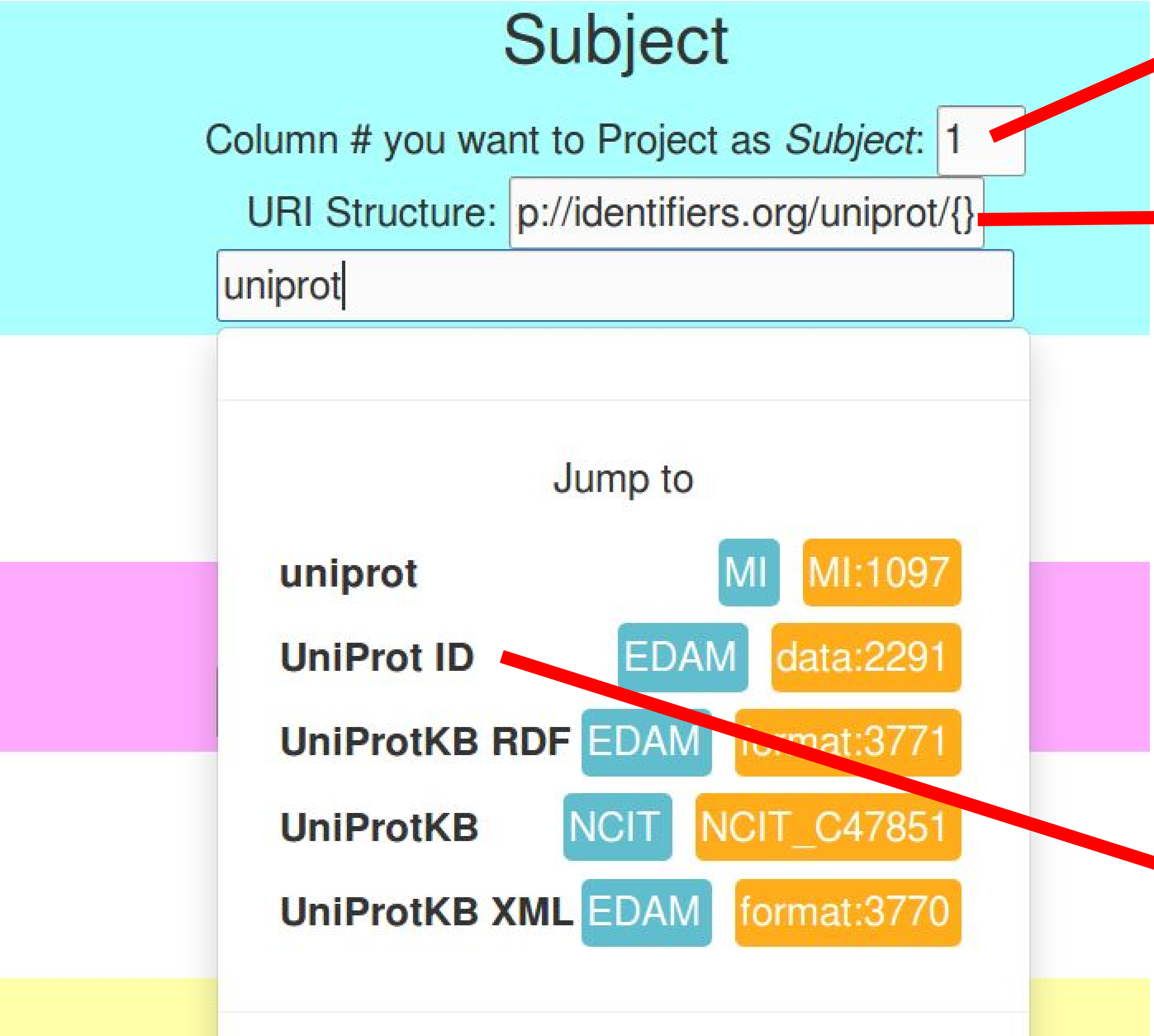
URI Structure: http://example.org/ids/{}

Predicate

Object

Column # you want to Project as *Object*:

URI Structure: http://example.org/ids/{}



Subject

Column # you want to Project as *Subject*: 1

URI Structure: p://identifiers.org/uniprot/{}

uniprot

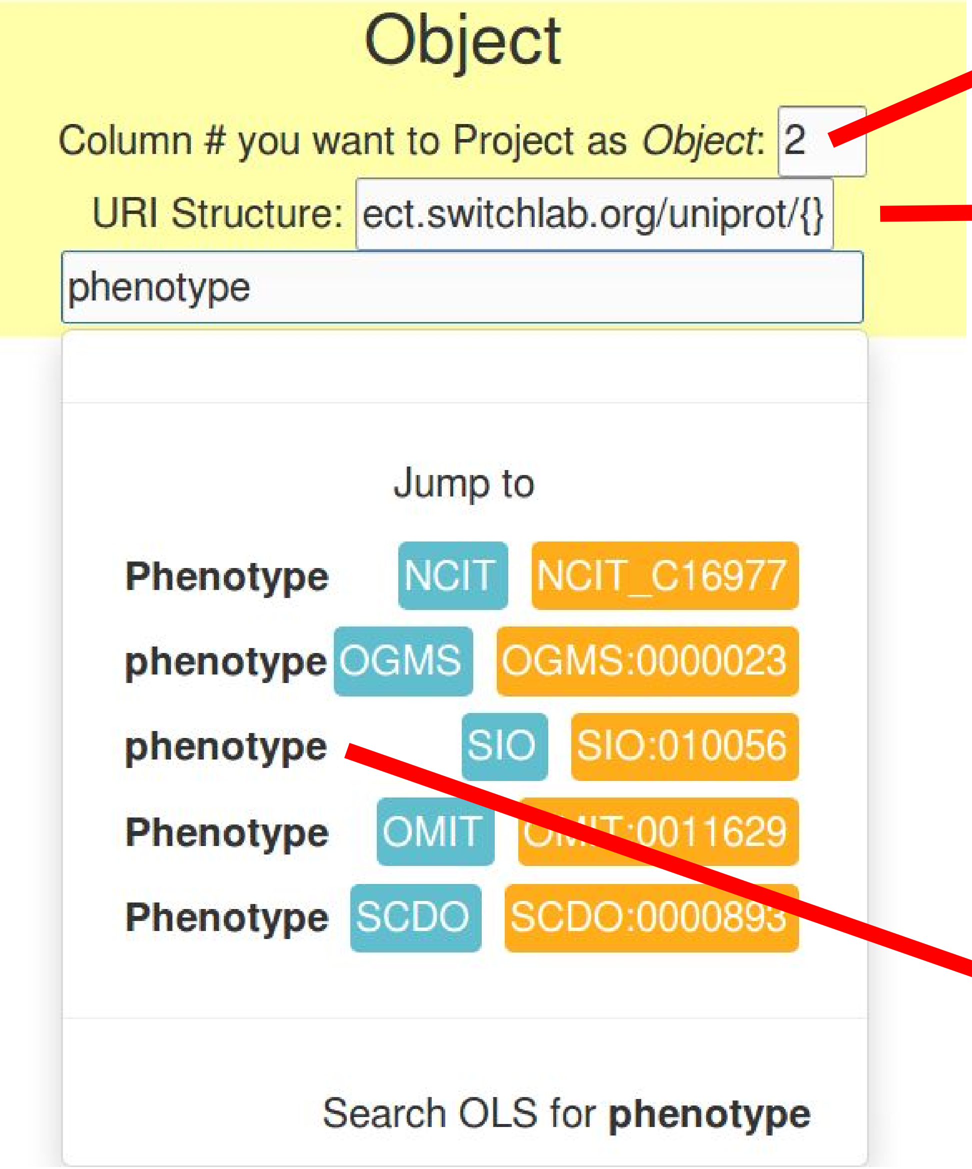
Jump to

uniprot	MI	MI:1097
UniProt ID	EDAM	data:2291
UniProtKB RDF	EDAM	format:3771
UniProtKB	NCIT	NCIT_C47851
UniProtKB XML	EDAM	format:3770

Column 1

http://identifiers.org/uniprot/{ }

The type of data is a **UniProt ID**



Object

Column # you want to Project as *Object*: 2

URI Structure: ect.switchlab.org/uniprot/{}

phenotype

Jump to

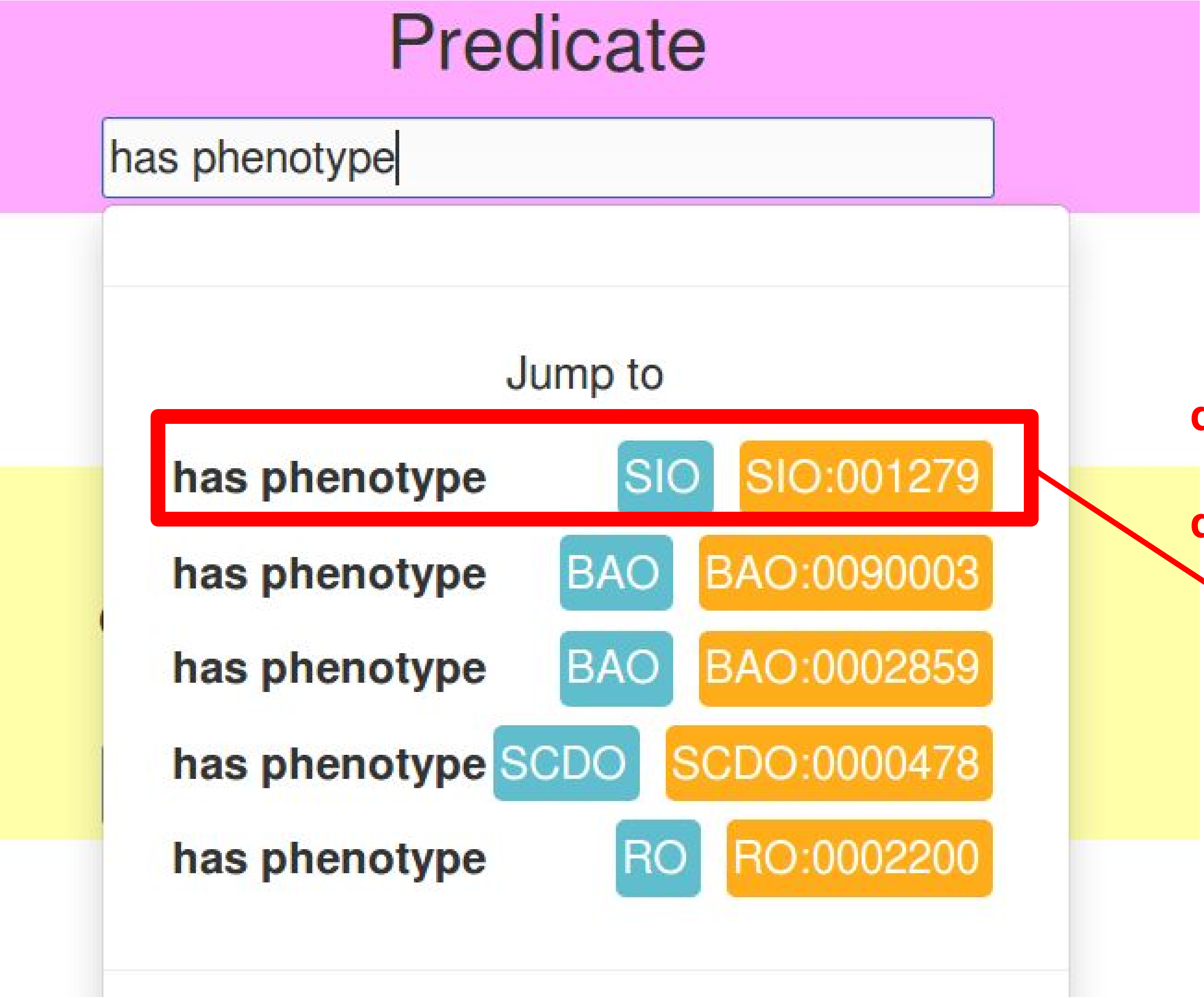
Phenotype	NCIT	NCIT_C16977
phenotype	OGMS	OGMS:0000023
phenotype	SIO	SIO:010056
Phenotype	OMIT	OMIT:0011629
Phenotype	SCDO	SCDO:0000893

Search OLS for **phenotype**

Column 2

http://snpeffect.switchlab.org/uniprot/{ }

The type of data is <u>**Phenotype**</u>



Predicate

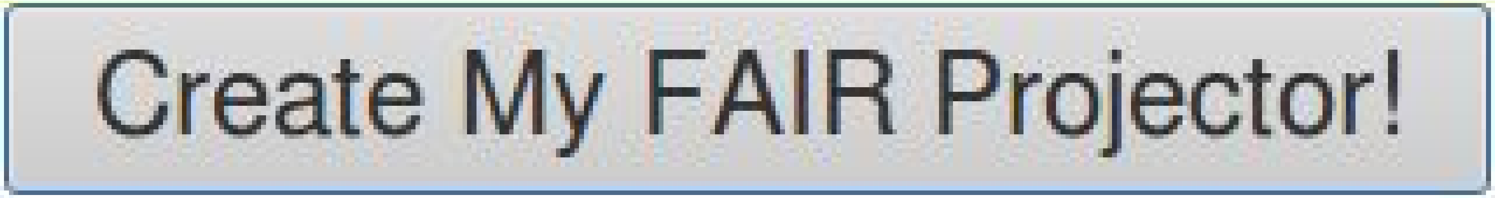
has phenotype

Jump to

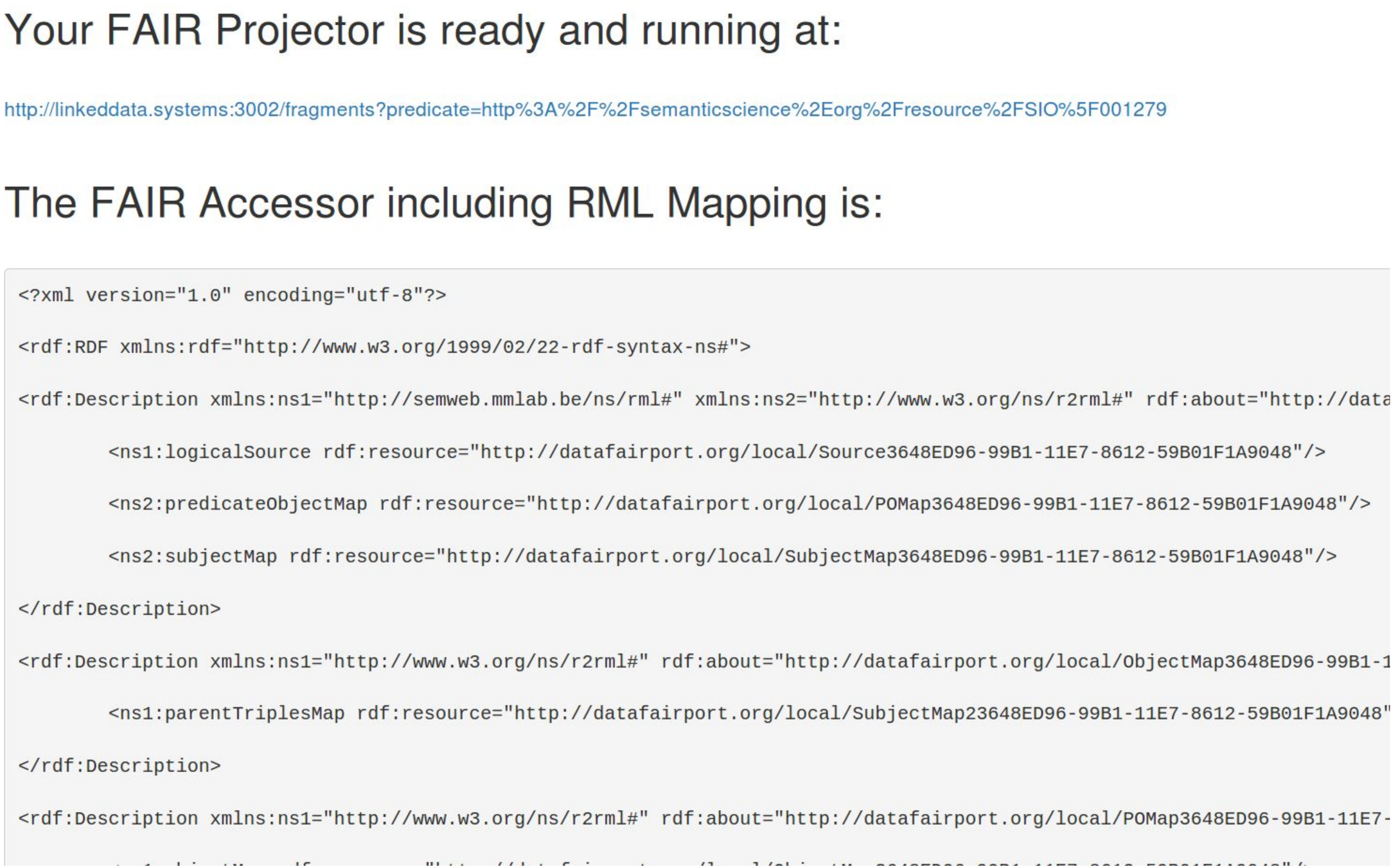
has phenotype	SIO	SIO:001279
has phenotype	BAO	BAO:0090003
has phenotype	BAO	BAO:0002859
has phenotype	SCDO	SCDO:0000478
has phenotype	RO	RO:0002200

What is the **semantic relationship** between the **data in column 1** and the **data in column 2**

"Changes in this protein result in this phenotype"



(If you drink, and you know things,
you can answer these questions
in about 60 seconds)



Your FAIR Projector is ready and running at:

http://linkeddata.systems:3002/fragments?predicate=http%3A%2F%2Fsemanticscience%2Eorg%2Fresource%2FSIO%5F001279

The FAIR Accessor including RML Mapping is:

```
<?xml version="1.0" encoding="utf-8"?>

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

<rdf:Description xmlns:ns1="http://semweb.mmlab.be/ns/rml#" xmlns:ns2="http://www.w3.org/ns/r2rml#" rdf:about="http://data

        <ns1:logicalSource rdf:resource="http://datafairport.org/local/Source3648ED96-99B1-11E7-8612-59B01F1A9048"/>

        <ns2:predicateObjectMap rdf:resource="http://datafairport.org/local/POMap3648ED96-99B1-11E7-8612-59B01F1A9048"/>

        <ns2:subjectMap rdf:resource="http://datafairport.org/local/SubjectMap3648ED96-99B1-11E7-8612-59B01F1A9048"/>

</rdf:Description>

<rdf:Description xmlns:ns1="http://www.w3.org/ns/r2rml#" rdf:about="http://datafairport.org/local/ObjectMap3648ED96-99B1-1

        <ns1:parentTriplesMap rdf:resource="http://datafairport.org/local/SubjectMap23648ED96-99B1-11E7-8612-59B01F1A9048"

</rdf:Description>

<rdf:Description xmlns:ns1="http://www.w3.org/ns/r2rml#" rdf:about="http://datafairport.org/local/POMap3648ED96-99B1-11E7-
```

Your FAIR Projector is ready and running at:

http://linkeddata.systems:3002/fragments?predicate=http%3A%2F%2Fsemanticscience%2Eorg%2Fresource%2FSIO%5F001279

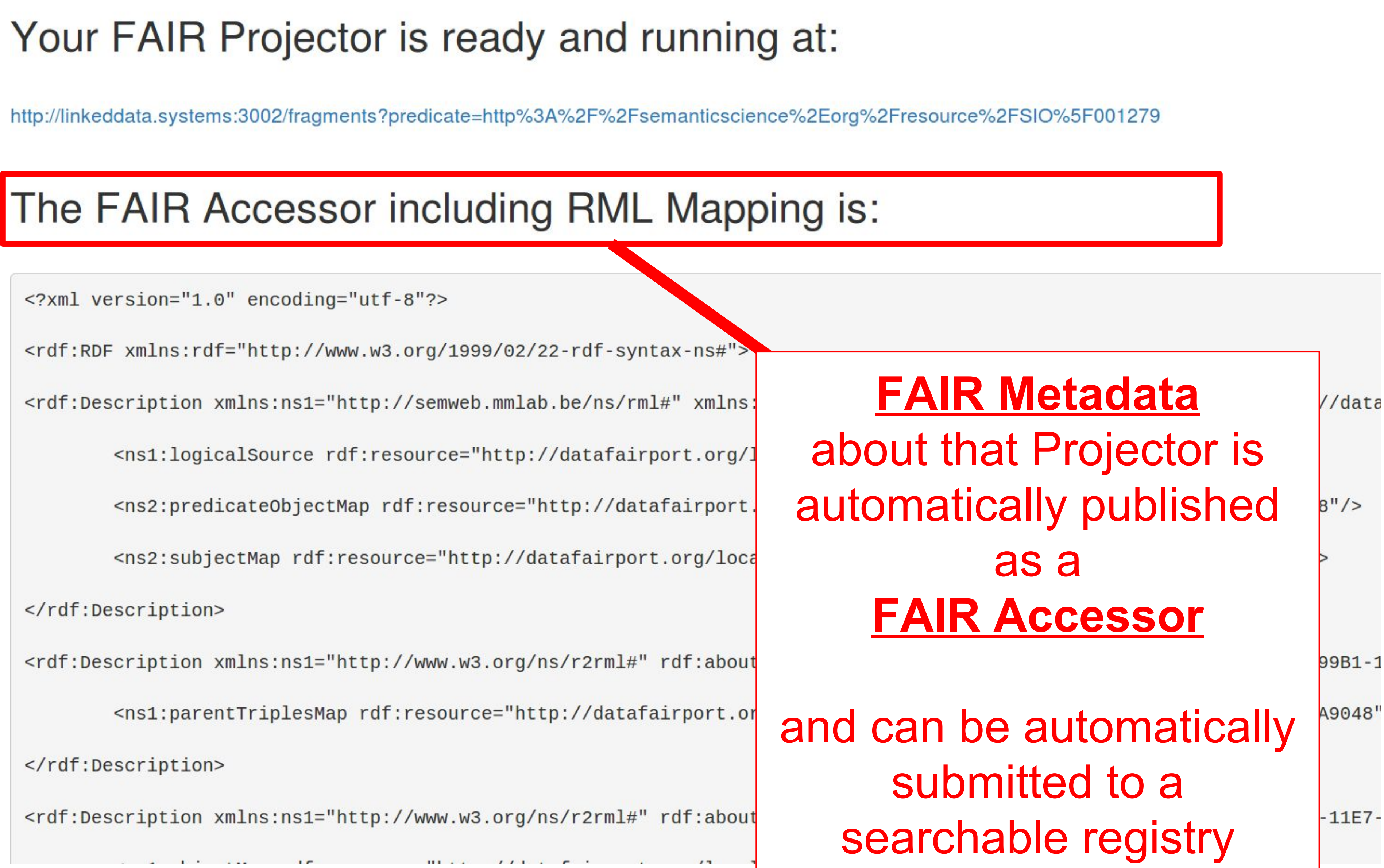
The FAIR Accessor including RML Mapping is:

```
<?xml version="1.0" encoding="utf-8"?>

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

<rdf:Description xmlns:ns1="http://semweb.mmlab.be/ns/rml#" xmlns:                                  //data

        <ns1:logicalSource rdf:resource="http://datafairport.org/l

        <ns2:predicateObjectMap rdf:resource="http://datafairport.                                  8"/>

        <ns2:subjectMap rdf:resource="http://datafairport.org/loca                                  >

</rdf:Description>

<rdf:Description xmlns:ns1="http://www.w3.org/ns/r2rml#" rdf:about                                  99B1-1

        <ns1:parentTriplesMap rdf:resource="http://datafairport.or                                  A9048"

</rdf:Description>

<rdf:Description xmlns:ns1="http://www.w3.org/ns/r2rml#" rdf:about                                  -11E7-
```

The URL to retrieve your “Projected” machine-readable

FAIR Data

...dynamically generated from the CSV file



Your FAIR Projector is ready and running at:

http://linkeddata.systems:3002/fragments?predicate=http%3A%2F%2Fsemanticscience%2Eorg%2Fresource%2FSIO%5F001279

The FAIR Accessor including RML Mapping is:

```
<?xml version="1.0" encoding="utf-8"?>

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

<rdf:Description xmlns:ns1="http://semweb.mmlab.be/ns/rml#" xmlns:                                              //data

        <ns1:logicalSource rdf:resource="http://datafairport.org/l

        <ns2:predicateObjectMap rdf:resource="http://datafairport.                                         8"/>

        <ns2:subjectMap rdf:resource="http://datafairport.org/loca                                         >

</rdf:Description>

<rdf:Description xmlns:ns1="http://www.w3.org/ns/r2rml#" rdf:about                                              99B1-1

        <ns1:parentTriplesMap rdf:resource="http://datafairport.or                                         A9048"

</rdf:Description>

<rdf:Description xmlns:ns1="http://www.w3.org/ns/r2rml#" rdf:about                                              -11E7-
```

<u>FAIR Metadata</u>
about that Projector is automatically published as a
<u>FAIR Accessor</u>

and can be automatically submitted to a searchable registry

Your FAIR Projector is ready and running a

http://linkeddata.systems:3002/fragments?predicate=http%3A%2F%2Fsemanticscience%

The FAIR Accessor including RML Mappin

```
<?xml version="1.0" encoding="utf-8"?>

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-s#">

<rdf:Description xmlns:ns1="http://semweb.mmlab.be/ns/rml#" xmlns:ns2="http://www.w3.org/ns/r2rml#" rdf:about="http://data

        <ns1:logicalSource rdf:resource="http://datafairport.org/local/Source3648ED96-99B1-11E7-8612-59B01F1A9048"/>

        <ns2:predicateObjectMap rdf:resource="http://datafairport.org/local/POMap3648ED96-99B1-11E7-8612-59B01F1A9048"/>

        <ns2:subjectMap rdf:resource="http://datafairport.org/local/SubjectMap3648ED96-99B1-11E7-8612-59B01F1A9048"/>

</rdf:Description>

<rdf:Description xmlns:ns1="http://www.w3.org/ns/r2rml#" rdf:about="http://datafairport.org/local/ObjectMap3648ED96-99B1-1

        <ns1:parentTriplesMap rdf:resource="http://datafairport.org/local/SubjectMap23648ED96-99B1-11E7-8612-59B01F1A9048"

</rdf:Description>

<rdf:Description xmlns:ns1="http://www.w3.org/ns/r2rml#" rdf:about="http://datafairport.org/local/POMap3648ED96-99B1-11E7-
```

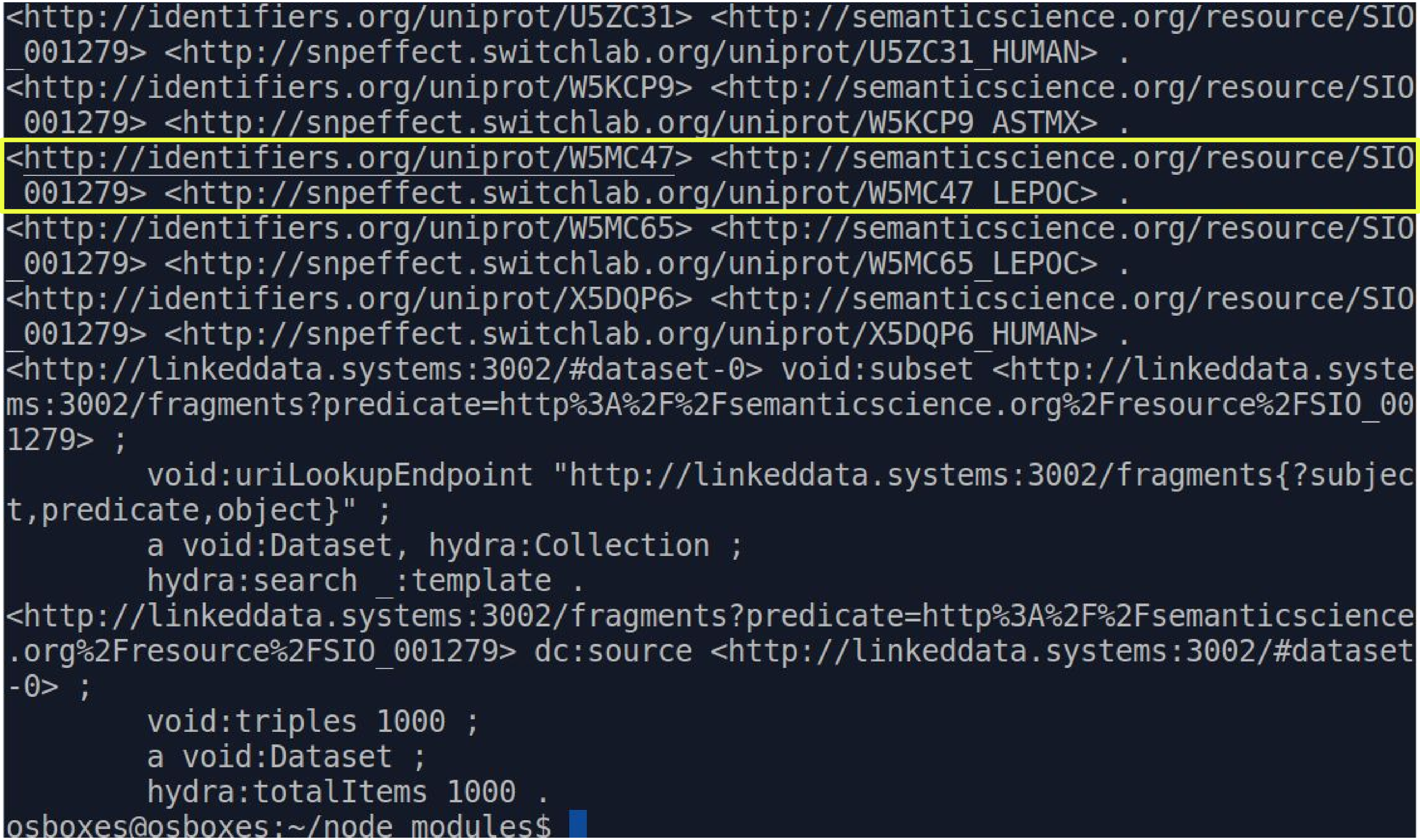
(RML enables **discovery** of the CSV data by its “type”)

```
<rdf:Description xmlns:ns1="http://purl.org/dc/elements/1.1/" xmlns:ns2="http://www.w3.org/ns/dcat#"

        <ns1:format>application/rdf+xml</ns1:format>

        <ns1:format>application/x-turtle</ns1:format>

        <ns1:format>text/html</ns1:format>

        <rdf:type rdf:resource="http://datafairport.org/ontology/FAIR-schema.owl#Projector"/>

        <rdf:type rdf:resource="http://purl.org/dc/elements/1.1/Dataset"/>

        <rdf:type rdf:resource="http://rdfs.org/ns/void#Dataset"/>

        <rdf:type rdf:resource="http://www.w3.org/ns/dcat#Distribution"/>

        <ns2:downloadURL rdf:resource="http://linkeddata.systems:3002/fragments?predicate=http%3A%2F

</rdf:Description>
```

Your **FAIR Data** are registered as a new DCAT Distribution of *https://zenodo.org/record/890267/files/dataoutput.csv*
In three formats (flexible reusability)

Call the URL:

http://linkeddata.systems:3002/fragments?predicate=http%3A%2F%2Fsemanticscience.org%2Fresource%2FSIO_001279



```
<http://identifiers.org/uniprot/U5ZC31> <http://semanticscience.org/resource/SIO
_001279> <http://snpeffect.switchlab.org/uniprot/U5ZC31_HUMAN> .
<http://identifiers.org/uniprot/W5KCP9> <http://semanticscience.org/resource/SIO
_001279> <http://snpeffect.switchlab.org/uniprot/W5KCP9_ASTMX> .
<http://identifiers.org/uniprot/W5MC47> <http://semanticscience.org/resource/SIO
_001279> <http://snpeffect.switchlab.org/uniprot/W5MC47_LEPOC> .
<http://identifiers.org/uniprot/W5MC65> <http://semanticscience.org/resource/SIO
_001279> <http://snpeffect.switchlab.org/uniprot/W5MC65_LEPOC> .
<http://identifiers.org/uniprot/X5DQP6> <http://semanticscience.org/resource/SIO
_001279> <http://snpeffect.switchlab.org/uniprot/X5DQP6_HUMAN> .
<http://linkeddata.systems:3002/#dataset-0> void:subset <http://linkeddata.syste
ms:3002/fragments?predicate=http%3A%2F%2Fsemanticscience.org%2Fresource%2FSIO_00
1279> ;
        void:uriLookupEndpoint "http://linkeddata.systems:3002/fragments{?subjec
t,predicate,object}" ;
        a void:Dataset, hydra:Collection ;
        hydra:search _:template .
<http://linkeddata.systems:3002/fragments?predicate=http%3A%2F%2Fsemanticscience
.org%2Fresource%2FSIO_001279> dc:source <http://linkeddata.systems:3002/#dataset
-0> ;
        void:triples 1000 ;
        a void:Dataset ;
        hydra:totalItems 1000 .
osboxes@osboxes:~/node_modules$
```

UniProt:W5MC47 relates to phenotype **SNPEffect**:W5MC47_LEPOC

I can immediately use a Web query language (SPARQL) to automatically combine this new FAIR Data with other data sources like EBI, NCBI, PubMed, ChEMBL,CHEBI, etc.

(This entire process took about 2 minutes…)

Summary of FAIR Projectors

F+I+R

FAIR Projectors provide a discoverable and standardized REST interface to retrieve interoperable data, and its interoperable metadata

I

FAIR Projectors can convert non-FAIR data into FAIR data, or can change the structure, URL format, or semantics of existing FAIR data sources

A + I

FAIR Projectors can be deployed over, and provide a common interface to:

- Static Data Deposits, in any format, anywhere
- Databases
- Triplestores
- Certain (common) types of Web Services

R+++

Triple Descriptors are independent FAIR entities, intended for reuse,

None of this required a new API

FAIR Metrics

FAIRness is quantifiable!

A small but stakeholder-representative working group has designed *quantitative* metrics to evaluate the “FAIRness” of a resource

This evaluation can be (semi-)automated

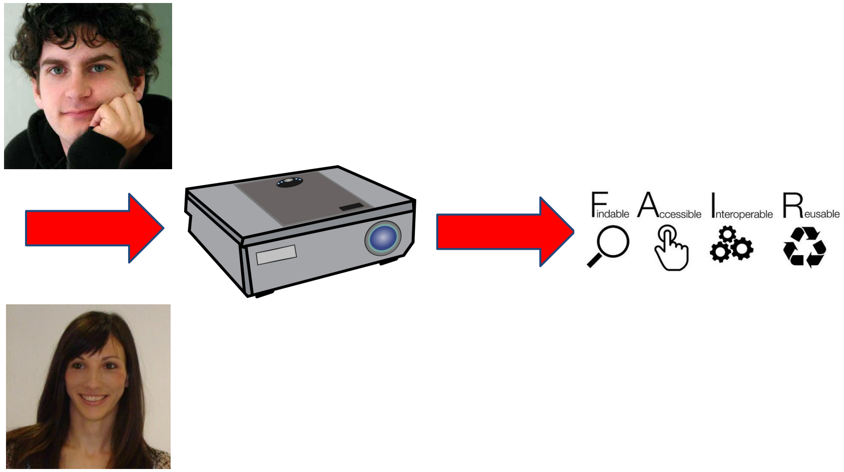
Publication being written, but the
current draft of the Metrics should be available in a few weeks
for public comment

Findable Accessible Interoperable Reusable

Findable Accessible Interoperable Reusable

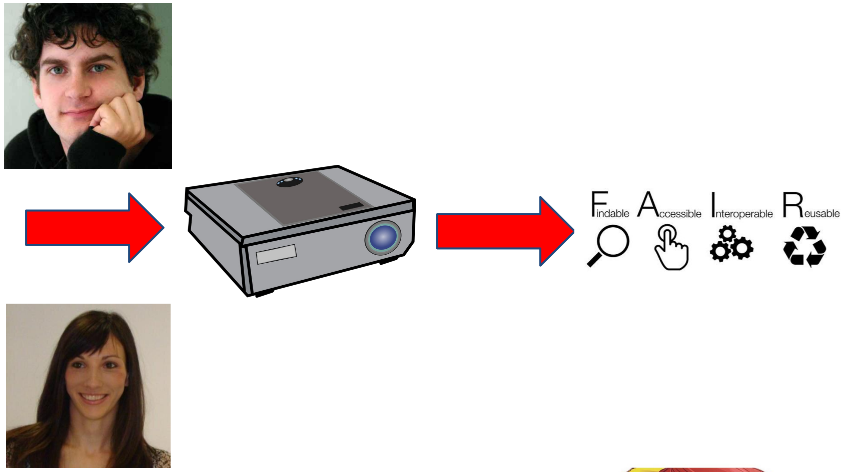
Findable Accessible Interoperable Reusable

	N	O	T		
		F	A	I	R



Findable Accessible Interoperable Reusable

	N	O	T		
		F	A	I	R



Findable Accessible Interoperable Reusable

The forecast is cloud-y
…but increasingly FAIR

Special Thanks to:

Michel Dumontier - Stanford Center for Biomedical Informatics Research, Stanford, California.

Ruben Verborgh – Ghent University – imec, Ghent, Belgium

Luiz Olavo Bonino da Silva Santos - Dutch Techcentre for Life Sciences, Utrecht, The Netherlands - Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.

Tim Clark - Department of Neurology, Massachusetts General Hospital Boston MA and Harvard Medical School, Boston, MA, USA

Morris A. Swertz - Genomics Coordination Center and Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands

Fleur D.L. Kelpin - Genomics Coordination Center and Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands

Alasdair J. G. Gray - Department of Computer Science, School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, UK

Erik A. Schultes - Department of Human Genetics, Leiden University Medical Center, The Netherlands

Erik M. van Mulligen - Department of Medical Informatics, Erasmus University Medical Center Rotterdam, The Netherlands

Paolo Ciccarese - Perkin Elmer Innovation Lab, Cambridge MA and Harvard Medical School, Boston MA, USA

Mark Thompson - Leiden University Medical Center, Leiden, The Netherlands

Jerven T. Bolleman - Swiss-Prot group, SIB Swiss Institute of Bioinformatics, Centre Medical Universitaire, Geneva, Switzerland

CBGP
CENTRO DE BIOTECNOLOGÍA
Y GENÓMICA DE PLANTAS
UPM-INIA

Fundación BBVA

MARIE CURIE ACTIONS

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

http://wilkinsonlab.info

Funding for the Wilkinson Laboratory from:

UPM Isaac Peral Programme
Spanish Ministerio de Economía y Competitividad grant number TIN2014-55993-R.

BioHackathon 2017
in Tokyo/Iwate

BioHackathon Funding from DBCLS and NBDC.

Additional support for FAIR “Skunkworks” members comes from:

- European Union funded projects ELIXIR-EXCELERATE (H2020 no. 676559), ADOPT BBMRI-ERIC (H2020 no. 676550), CORBEL (H2020 no. 654248)
- Netherlands Organisation for Scientific Research (Odex4all project)
- Stichting Topconsortium voor Kennis en Innovatie High Tech Systemen en Materialen (FAIRdICT project)
- BBMRI-NL
- RD-Connect and ELIXIR (Rare disease implementation study FP7 no. 305444).

Mark D Wilkinson, CBGP UPM-INIA, 2017
ORCiD:0000-0001-6960-357X

You are not only ***allowed*** to use this presentation for your own purposes, you are ***encouraged*** to!

http://tinyurl.com/FAIRDataPortugal

Some graphical elements were taken from slide templates provided by:

FGST

Free Google Slides Templates