

Dynamic metadata model for B2Share

Invenio User Group Workshop 2017

Nicolas Harraudeau
CERN

How to lose research data 101

“The odds of a data set being reported as extant fell by **17% per year**”

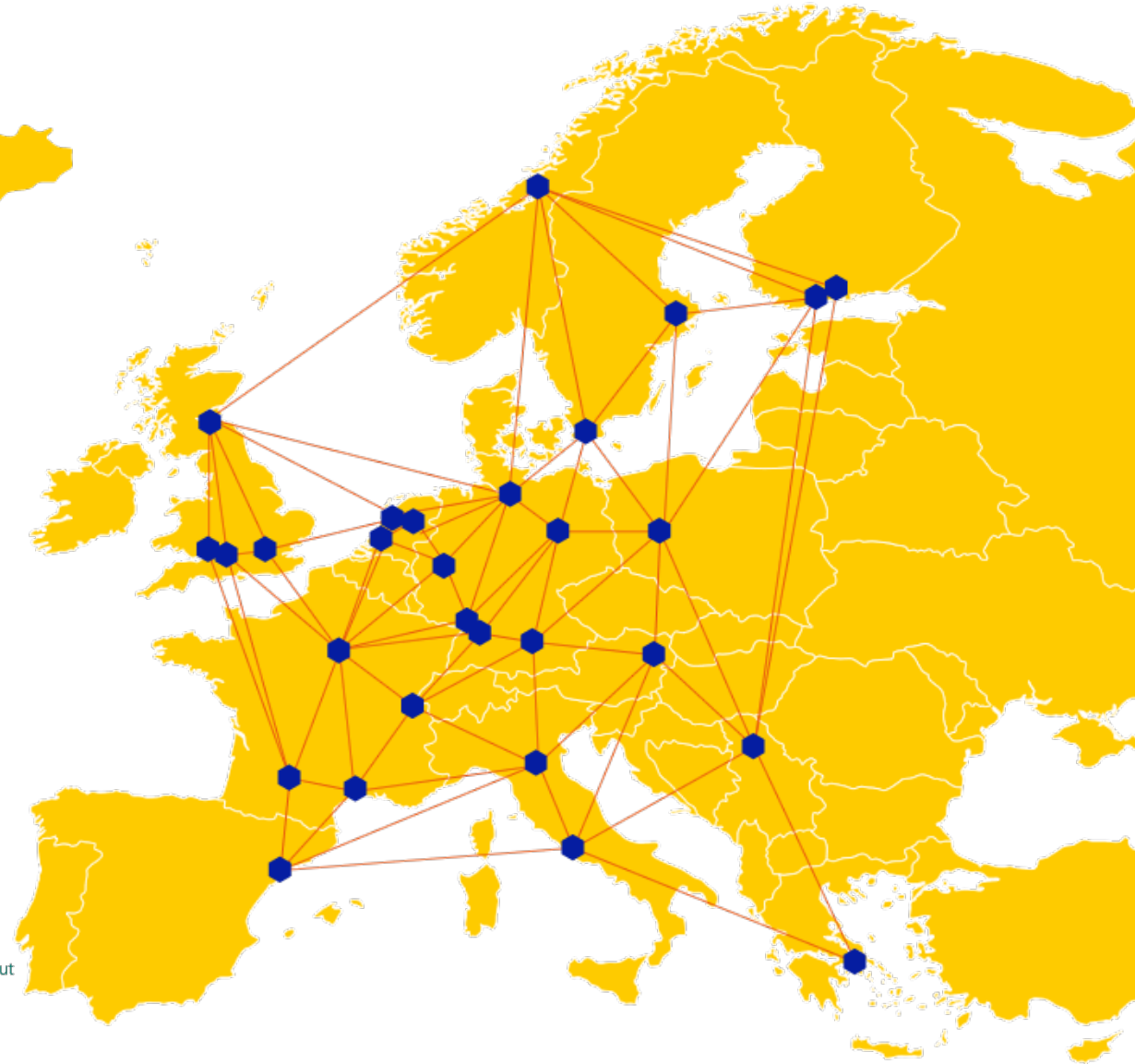
The Availability of Research Data Declines Rapidly with Article Age
October 22, 2013



Medical science experiments and research lost in University of Leicester blaze

By [Tom Mack](#) | Posted: April 23, 2016

-  CSC
-  UNINETT Jigma
-  BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación
-  Science & Technology Facilities Council
-  CEPIS
-  UCL
-  rzg
RECHENZENTRUM GARCHING
-  CINECA
-  Trust-IT Trust-IT Services Ltd
Communicating ICT to markets
-  Jisc
-  grnet
-  INES
Institut National de l'Enseignement Supérieur
-  KIT
Karlsruher Institut für Technologie
-  DKRZ
DEUTSCHES KLIMARECHENZENTRUM
-  LIBER
-  CERN
-  umweltbundesamt
ENVIRONMENT AGENCY AUSTRIA
-  Data Archiving and Networked Services
DANS
-  HELSINKI UNIVERSITY OF TECHNOLOGY
HELSINKI UNIVERSITY OF TECHNOLOGY
-  epcc
-  e-Science Data Factory
-  EMBL
-  INGV
-  SNIC
-  JÜLICH
FORSCHUNGSZENTRUM
-  Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment
-  KTH
KTH VETENSKAP OCH KONST
-  SURF SARA
-  GFZ
Helmholtz Centre
POTSDAM
-  Max-Planck-Institut
für Meteorologie
-  CLARIN
Common Language Resources and
Technology Infrastructure



Different metadata for each scientific Domain

Common Metadata

(Title, Authors, License, Open Access...)



Domain specific Metadata

Language,
Country/Region...

CLARIN



Linguistic

Geographic Location,
Spatial Resolution...

DRIHM



Meteorology and Climate

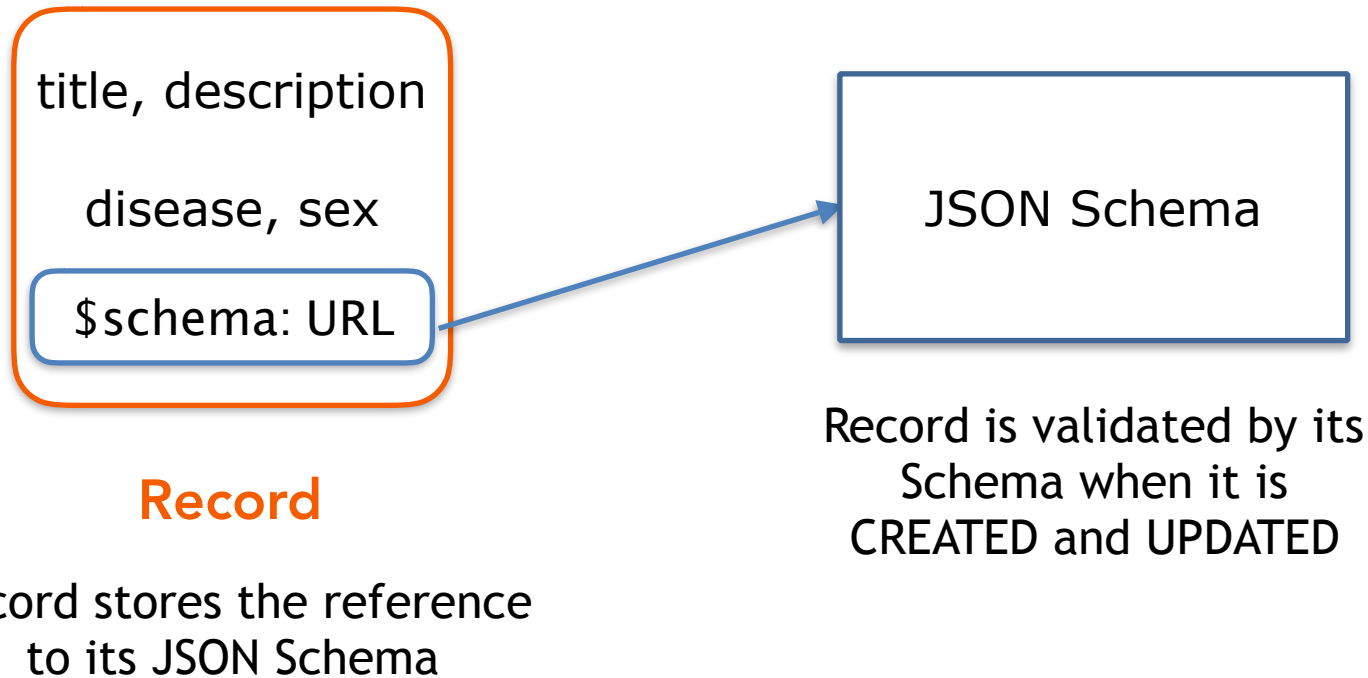
Disease, Sex, Age,
Study design...

BBMRI



Biomedical research

Invenio 3 metadata validation



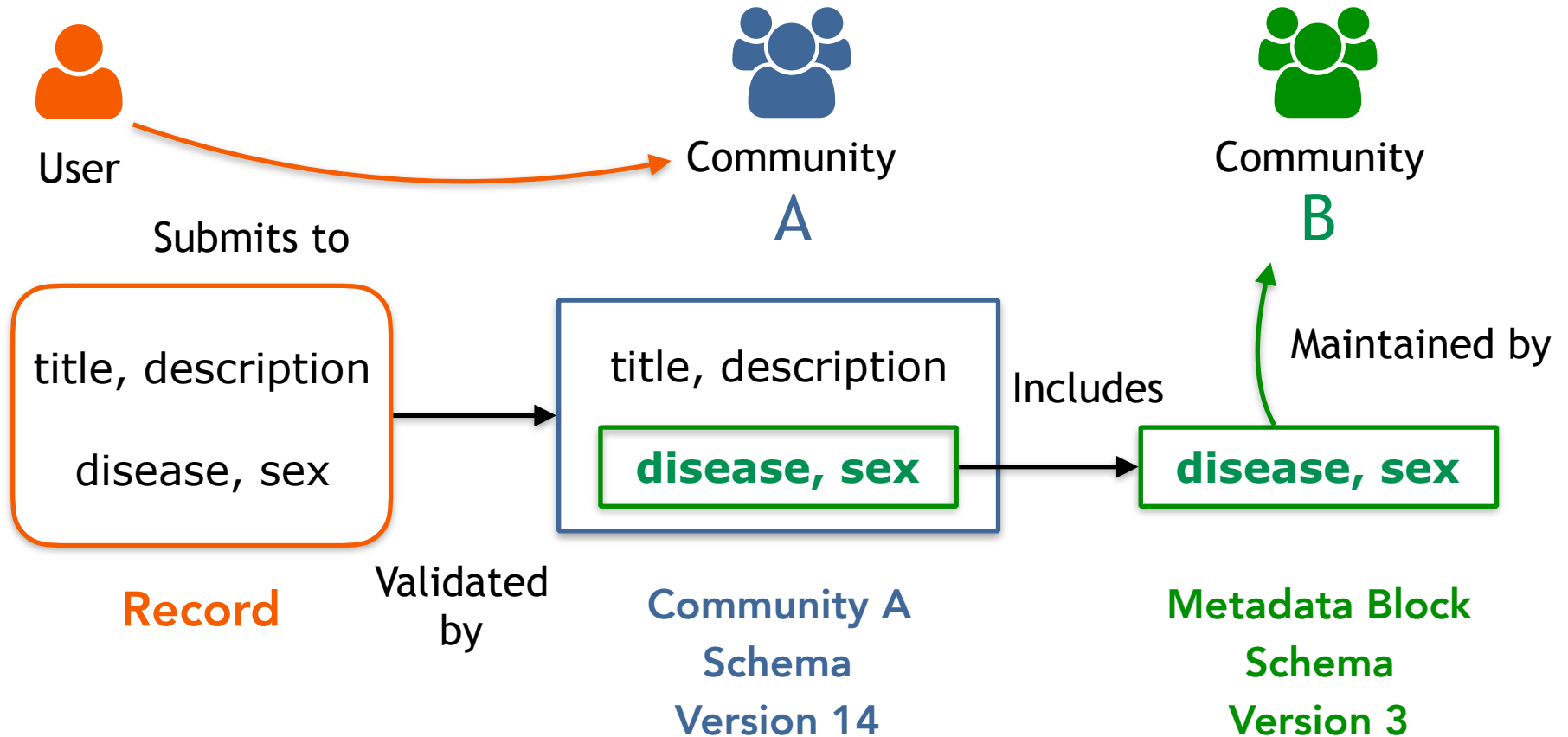


DEMO

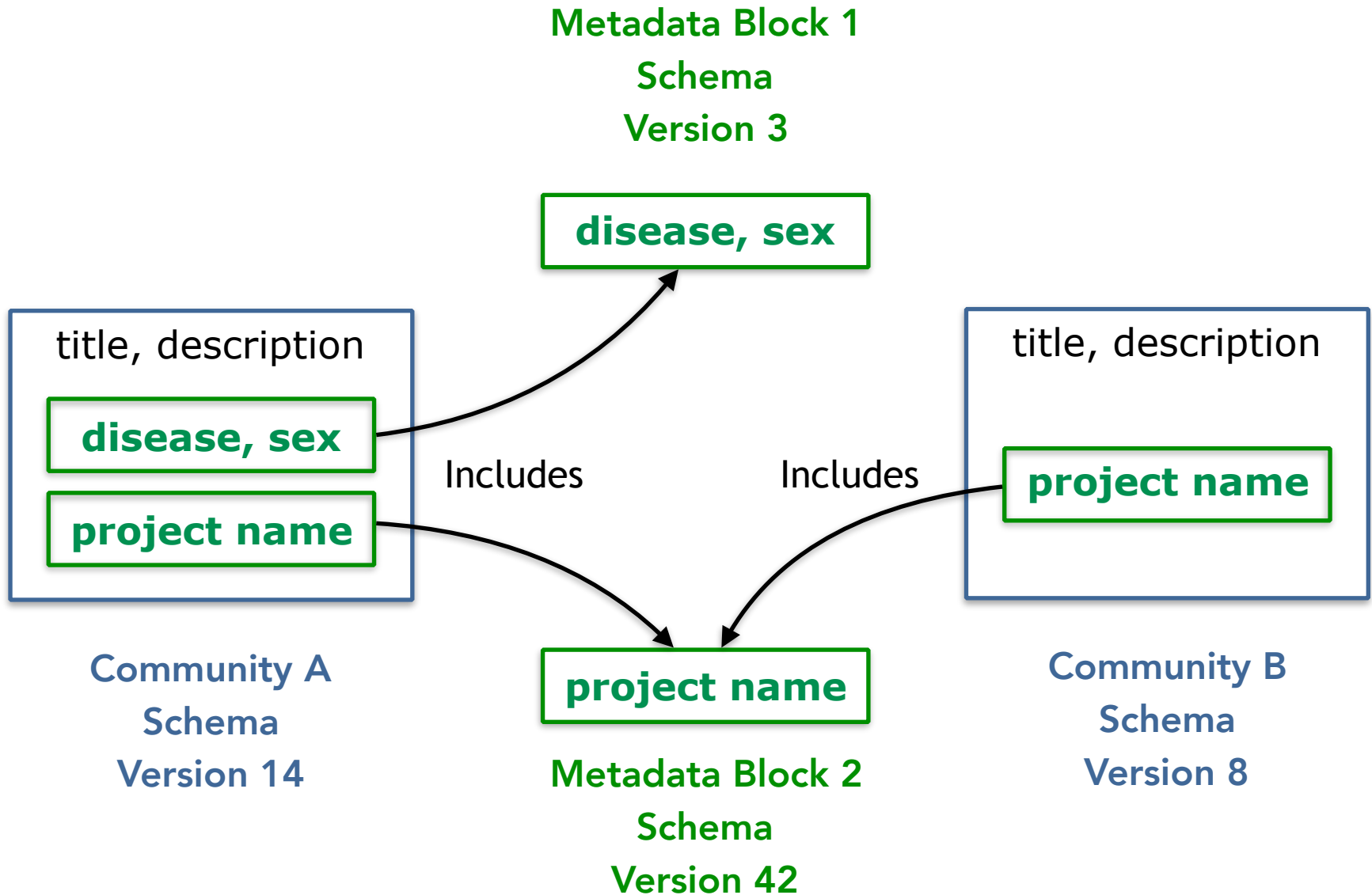
Record Creation

B2Share's composite schemas

Every Community has one versioned schema. It is composed of a **common fields** and a set of **metadata block schemas** which are maintained by other (or the same) communities.



B2Share's composite schemas



B2Share's composite schemas

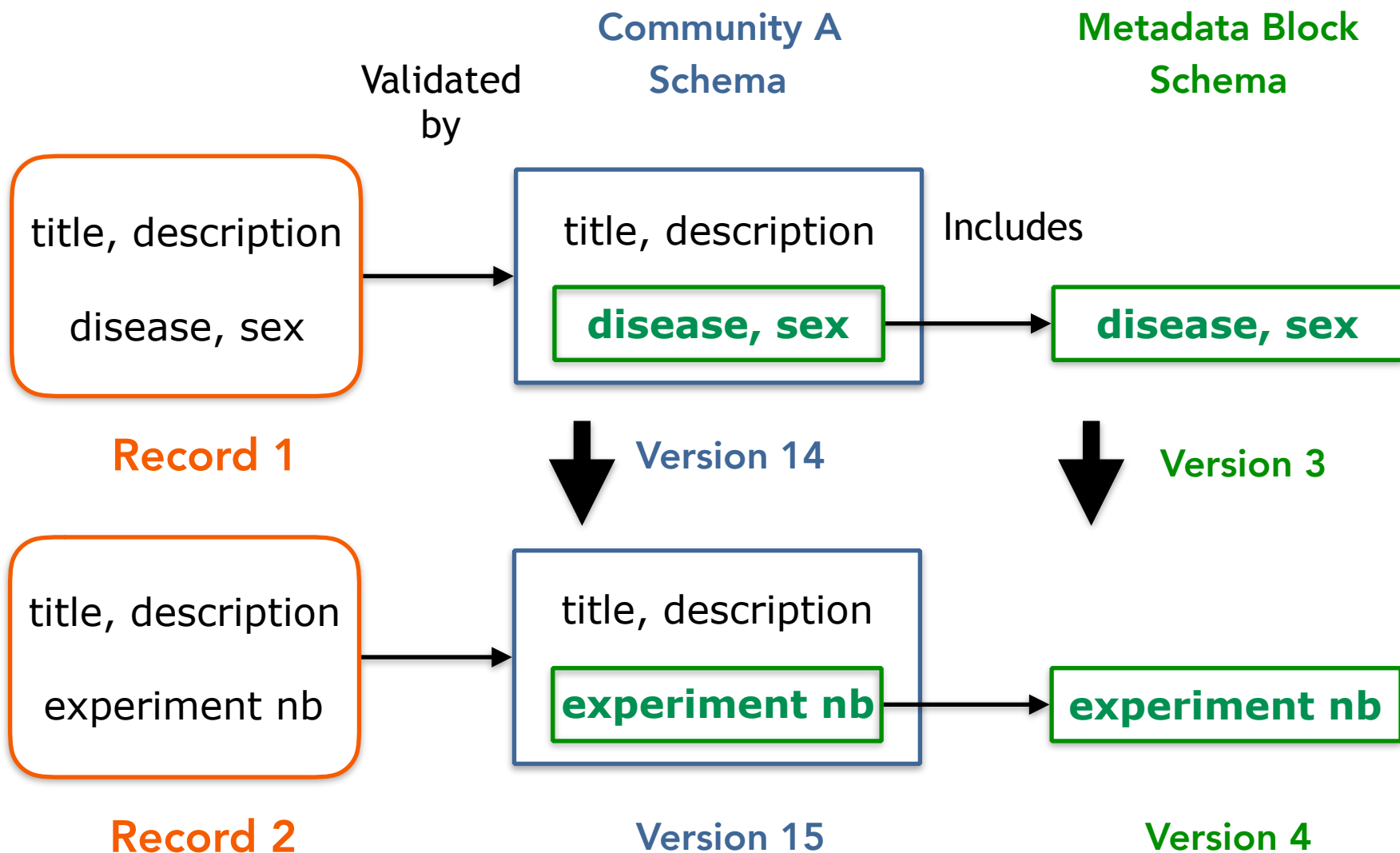
GET “http://.../api/communities/<community_ID>/<version_nb>”

```
1 {
2   "properties": {
3     "title": "...",
4     "description": "...",
5     ...
6   }
7   "community_specific": {
8     "<schema_ID1>": {
9       "$ref": "http://.../api/schemas/<schema_ID1>/versions/5#/json_schema"
10    },
11    "<schema_ID2>": {
12      "$ref": "http://.../api/schemas/<schema_ID2>/versions/3#/json_schema"
13    },
14    ...
15  }
16 }
17 }
18 }
19 }
20 }
21 }
```

Fields common to every dataset

Metadata Blocks

Schemas versioning



Schema backward compatibility

“http://.../api/schemas/<schema_ID1>/versions/1#/json_schema”

```
1 {
2   "properties": {
3
4     "disease": "...",
5     "sex": "...",
6     ...
7   }
8 }
```

type: “string”

“http://.../api/schemas/<schema_ID1>/versions/2#/json_schema”

```
1 {
2   "properties": {
3
4     "disease": {
5       "name": "...",
6     },
7     "sex": "...",
8     ...
9   }
10 }
```

type: “object” => INCOMPATIBLE

Schema backward compatibility




- Community Schemas and Metadata Block schemas have to be **backward compatible**.

i.e: Any **JSON pointer** (path) should always resolve the **same field type** whatever the schema version.

Consequence: If one wants to **change the structure**, he has to either **rename** the field and **deprecate** the old field name, or **create a new schema**.

-  [doschema](#) library enables to check schemas backward compatibility

Next steps

-  Enable to create schemas via REST API (only CLI right now)
-  Create a nice user interface enabling to create/update schemas
-  Improve search

Links and bibliography

- EUDAT eudat.org
- B2Share instance at CSC b2share.eudat.eu
- Invenio <http://invenio-software.org/>
- JSON Schema <http://json-schema.org/>

- *Medical science experiments and research lost in University of Leicester blaze* (Leicester Mercury)
<http://www.leicestermercury.co.uk/Medical-science-experiments-research-lost/story-29163384-detail/story.html>