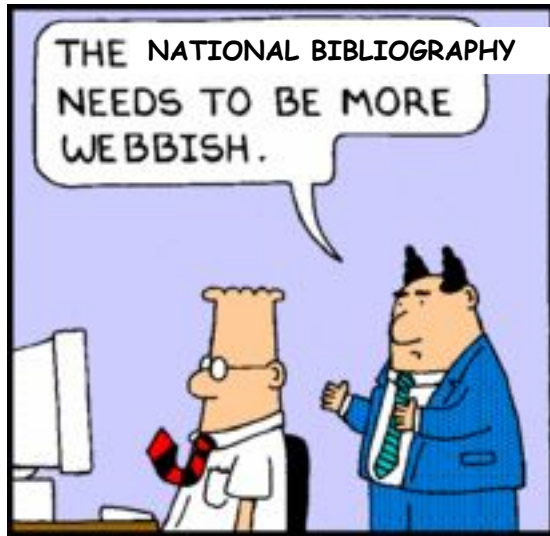


Finnish National Bibliography Fennica as Linked Data

Osma Suominen

SWIB17, Hamburg

6 Dec 2017



with apologies to Scott Adams

Why?

1. Making our data more visible, also internationally
2. Improving the quality and interoperability of our metadata
3. Building competency for the future
4. Why not? :)

1M bib records

bib record
bib record
bib record
bib record
bib record
bib record
bib record
bib record

125k person names

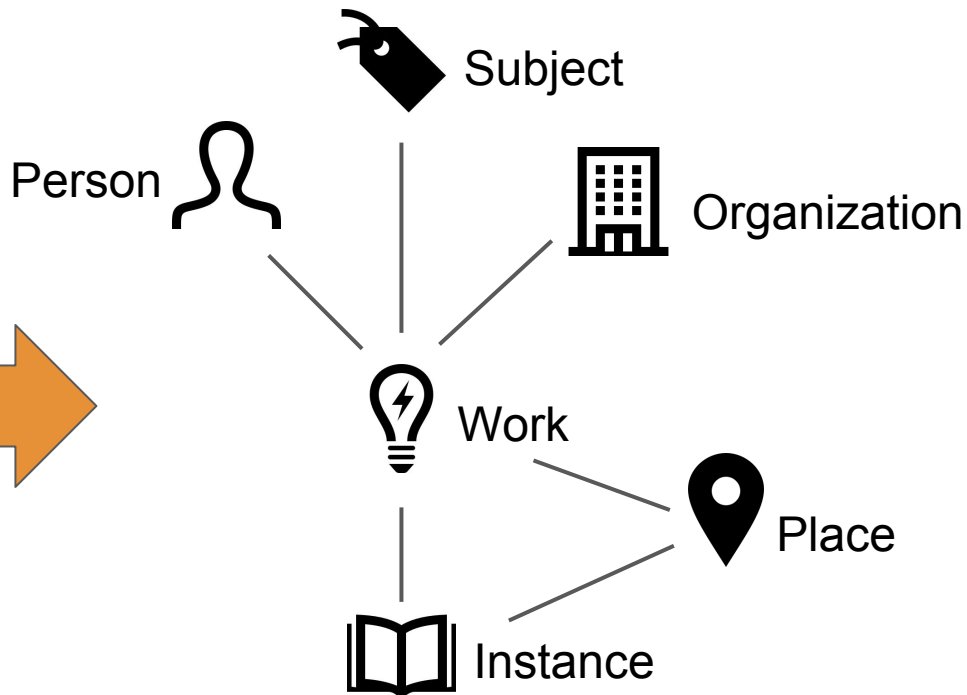
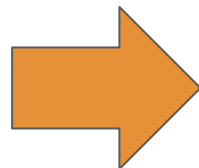
auth record
auth record

40k corporate names

auth record
auth record

35k subjects (YSA)

auth record
auth record

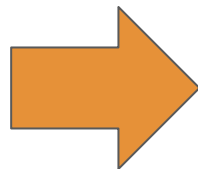


1M bib records

125k person names

bib record

auth record



bib record

auth record

bib record

auth record

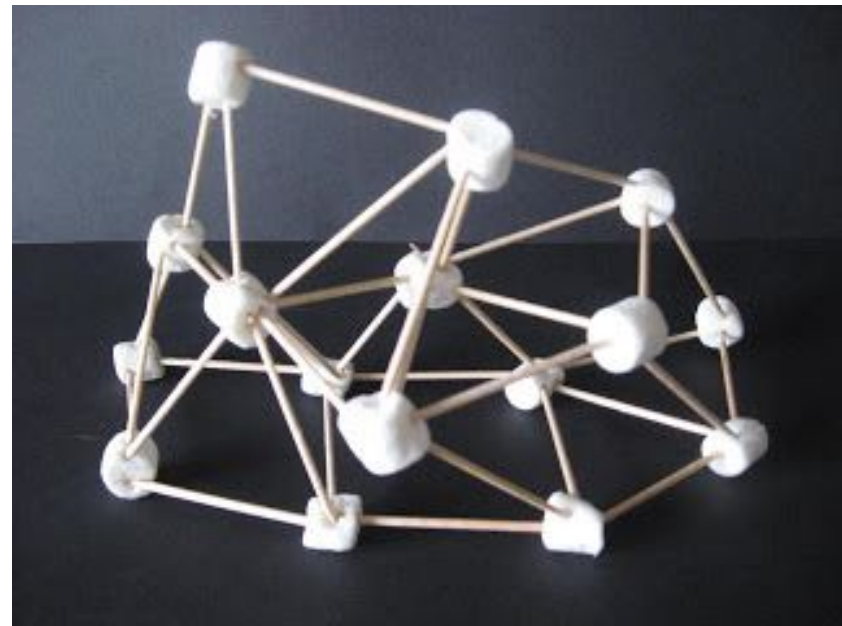
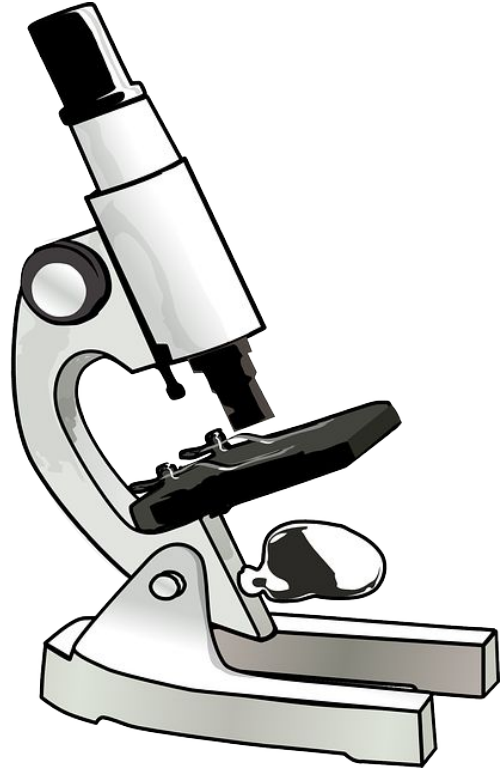
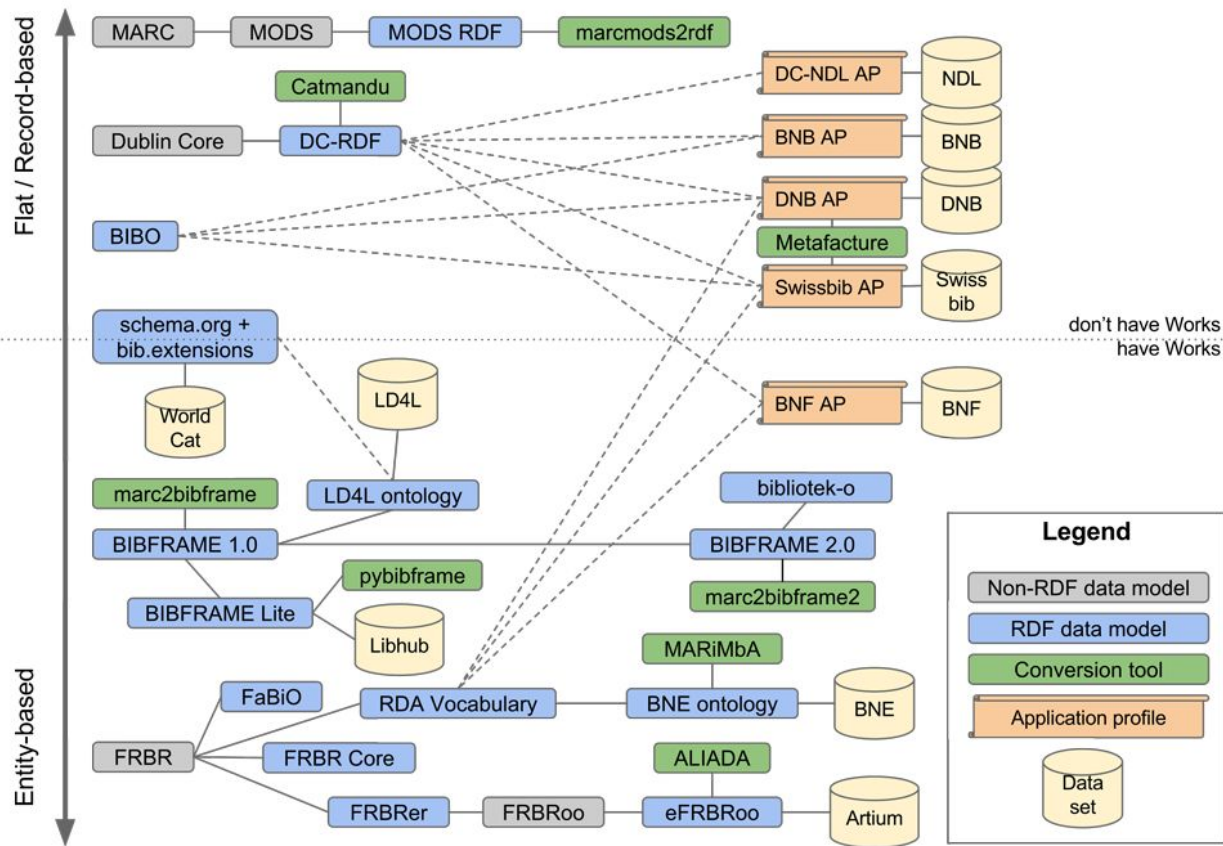


Image credit: [MaryMaking blog](#)



“From MARC silos to Linked Data silos”

“Family forest” of bibliographic data models, conversion tools, application profiles and data sets



Original image by Doc Searls. CC BY 2.0
<https://www.flickr.com/photos/docsearls/5500714140>

As seen in:

[SWIB16 talk](#)

[DCMI webinar](#)

[o-bib journal article](#)

schema.org

with separate Works and Instances like BIBFRAME,
as enabled by the bibliographic extensions

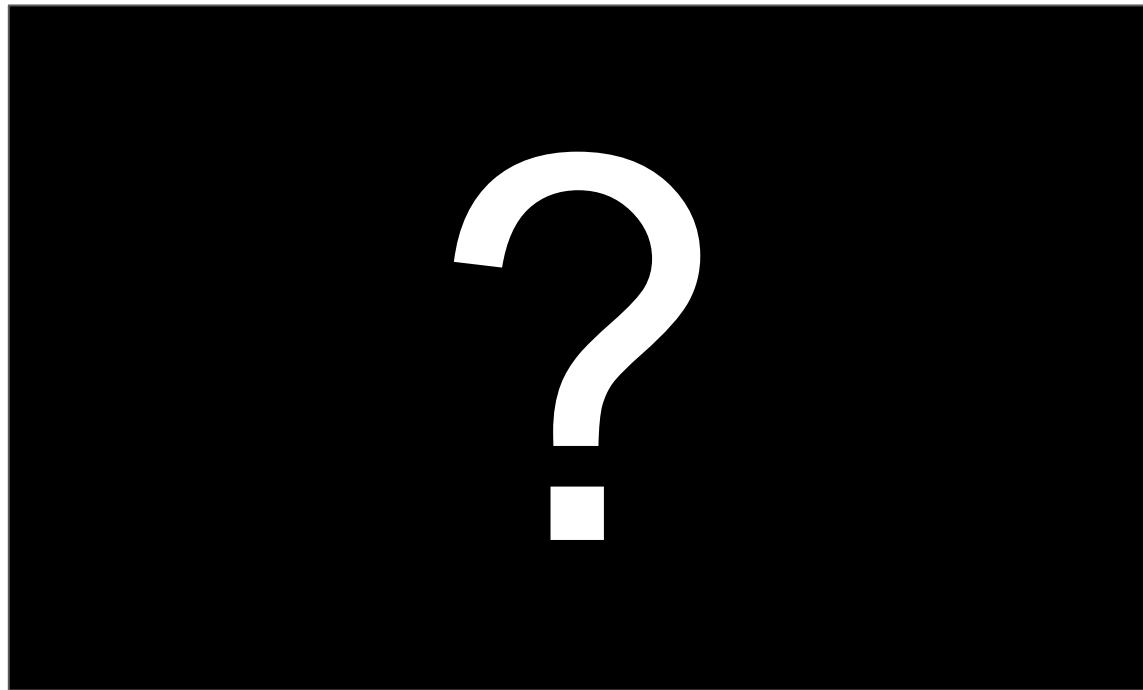
**because it allows us to describe our resources
from a common-sense, Web user perspective**

(and we get a metadata haircut for free!)

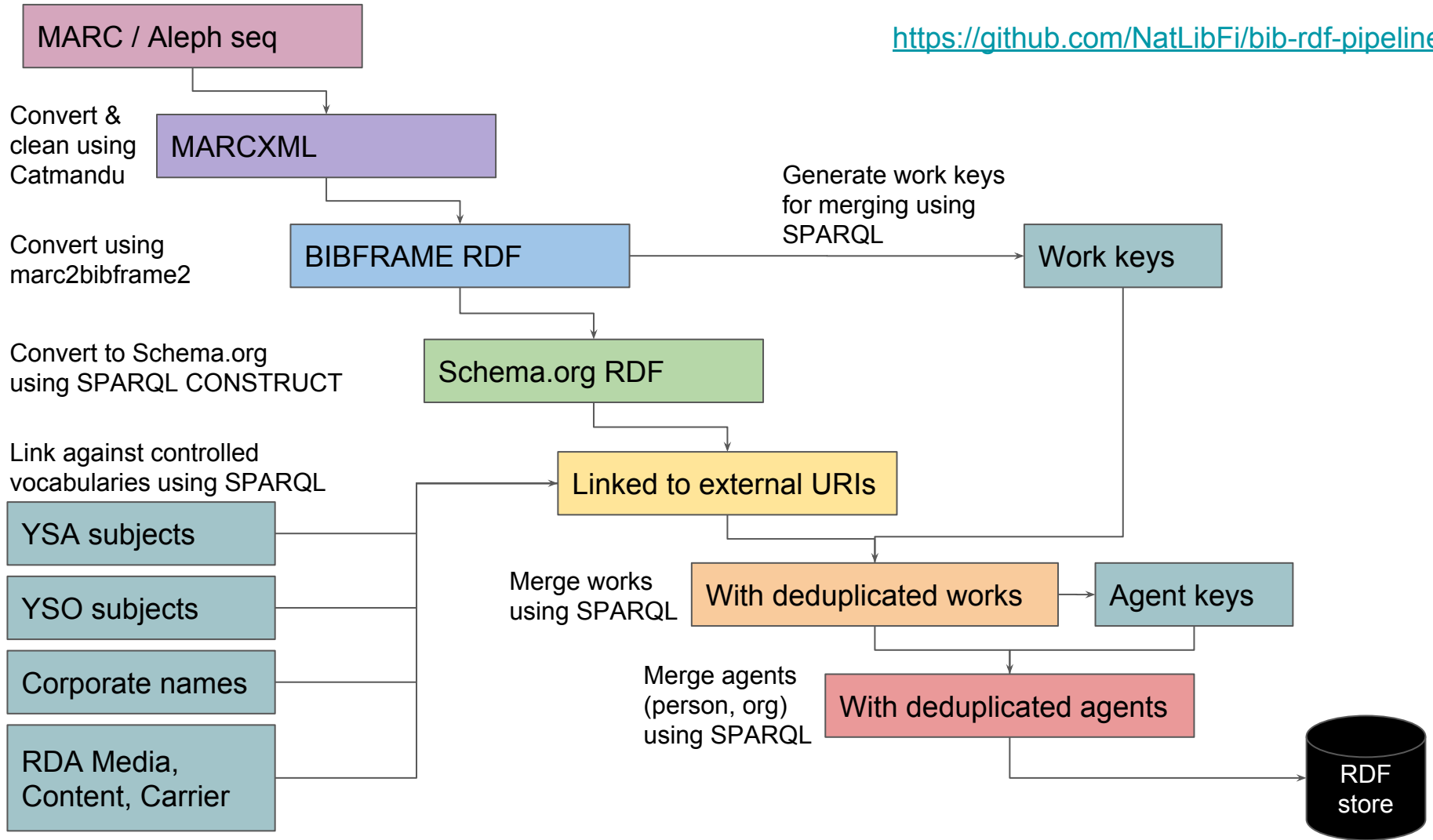


Special thanks to [Richard Wallis](#)
for help with applying schema.org!

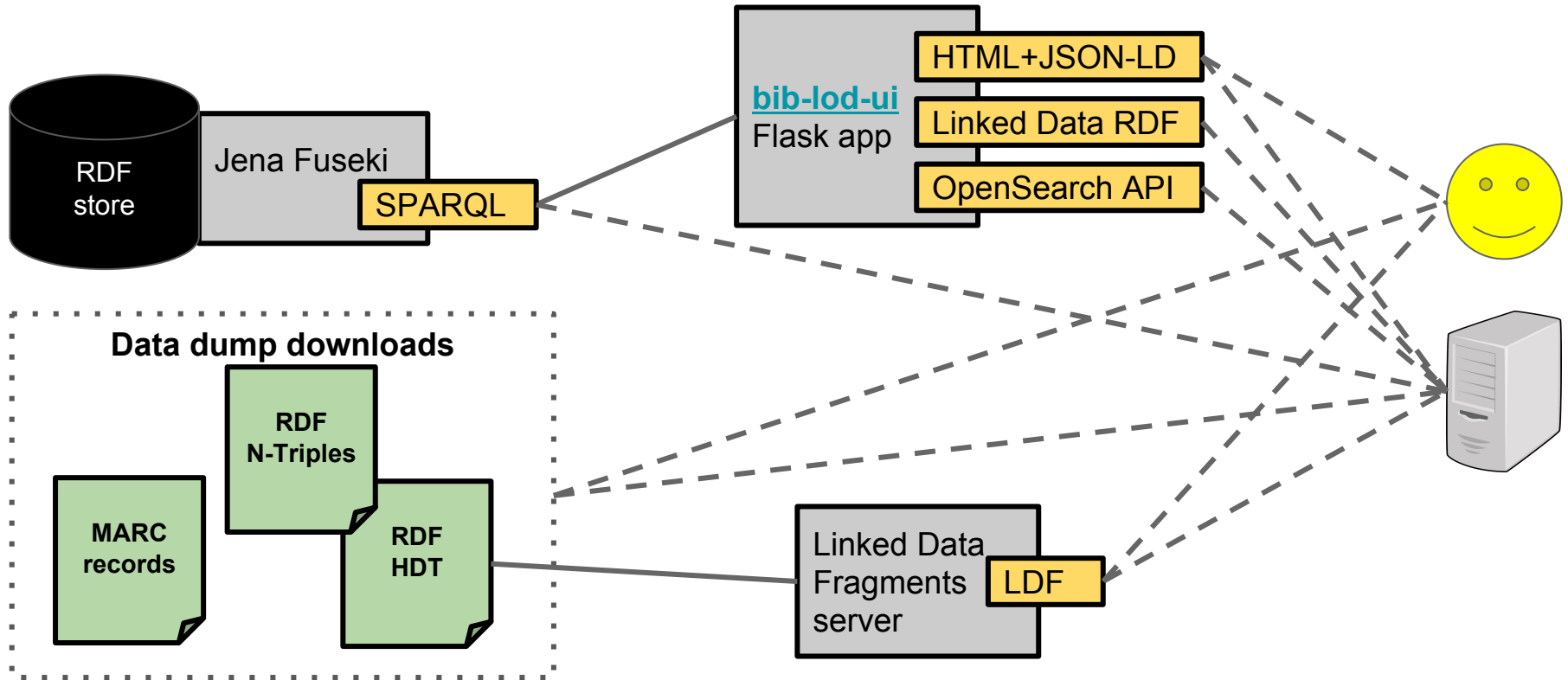
MARC →



→ **Linked
Data**



Publishing as Linked Open Data for human & machine access



Demo

<http://data.nationallibrary.fi/bib/me/W00009584100>

Spelunking UI...maybe?

Our hairballs

Index of /download

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
 Parent Directory		-	
 fennica-marc.zip	2017-12-05 00:23	254M	
 fennica.hdt	2017-12-05 00:23	492M	
 fennica.hdt.index	2017-12-05 00:23	293M	
 fennica.nt.gz	2017-12-05 00:22	739M	

Apache/2.4.18 (Ubuntu) Server at data.nationallibrary.fi Port 80

NatLibFi Linked Data Fragments server



Available datasets

Browse the following datasets as **Triple Pattern Fragments**:

Fennica

Fennica with an HDT back-end

The current dataset *index* contains metadata about these datasets.

Dataset index

Query dataset index by triple pattern

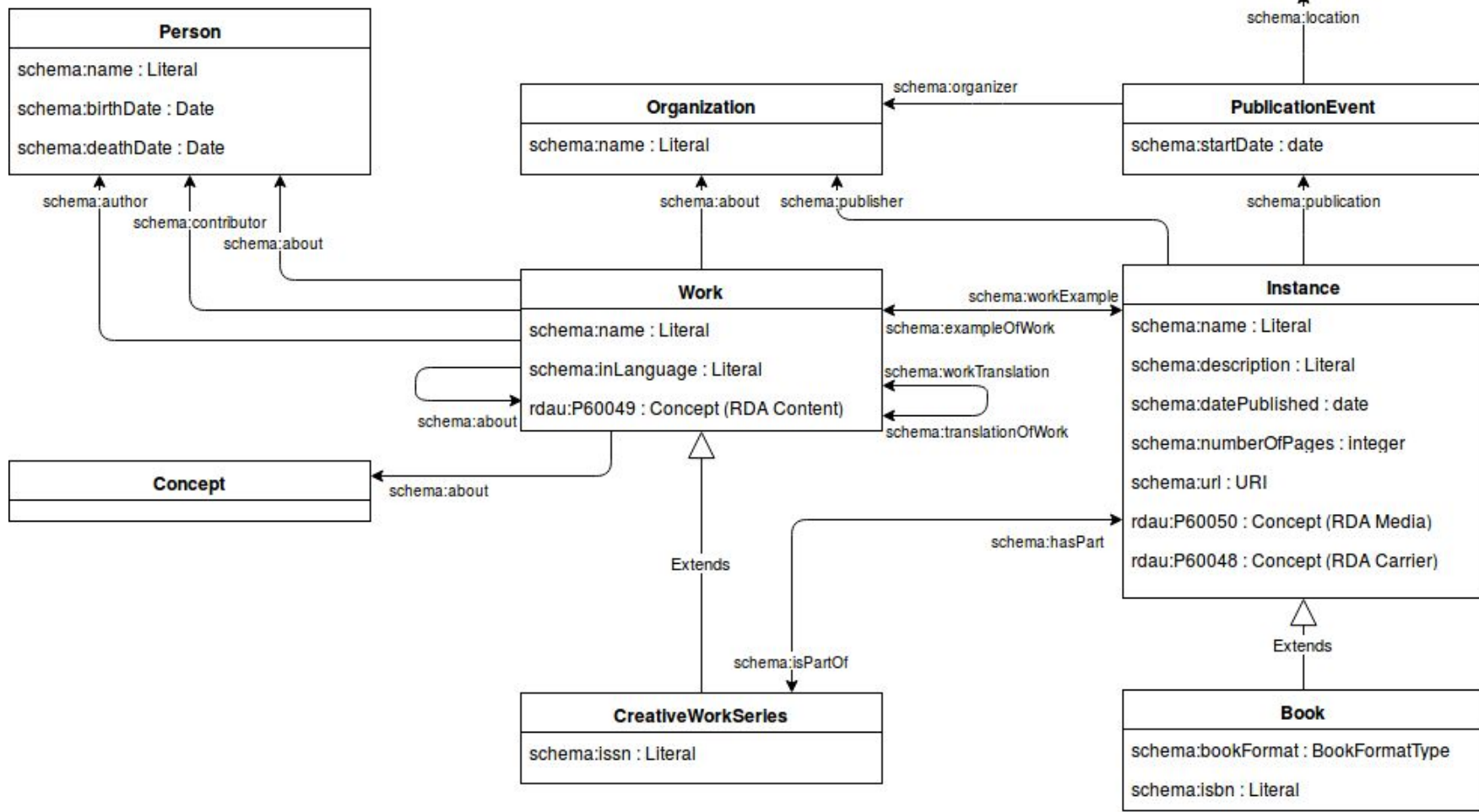
subject: _____

predicate: _____

object: _____

Find matching triples

Data model documentation



Challenges

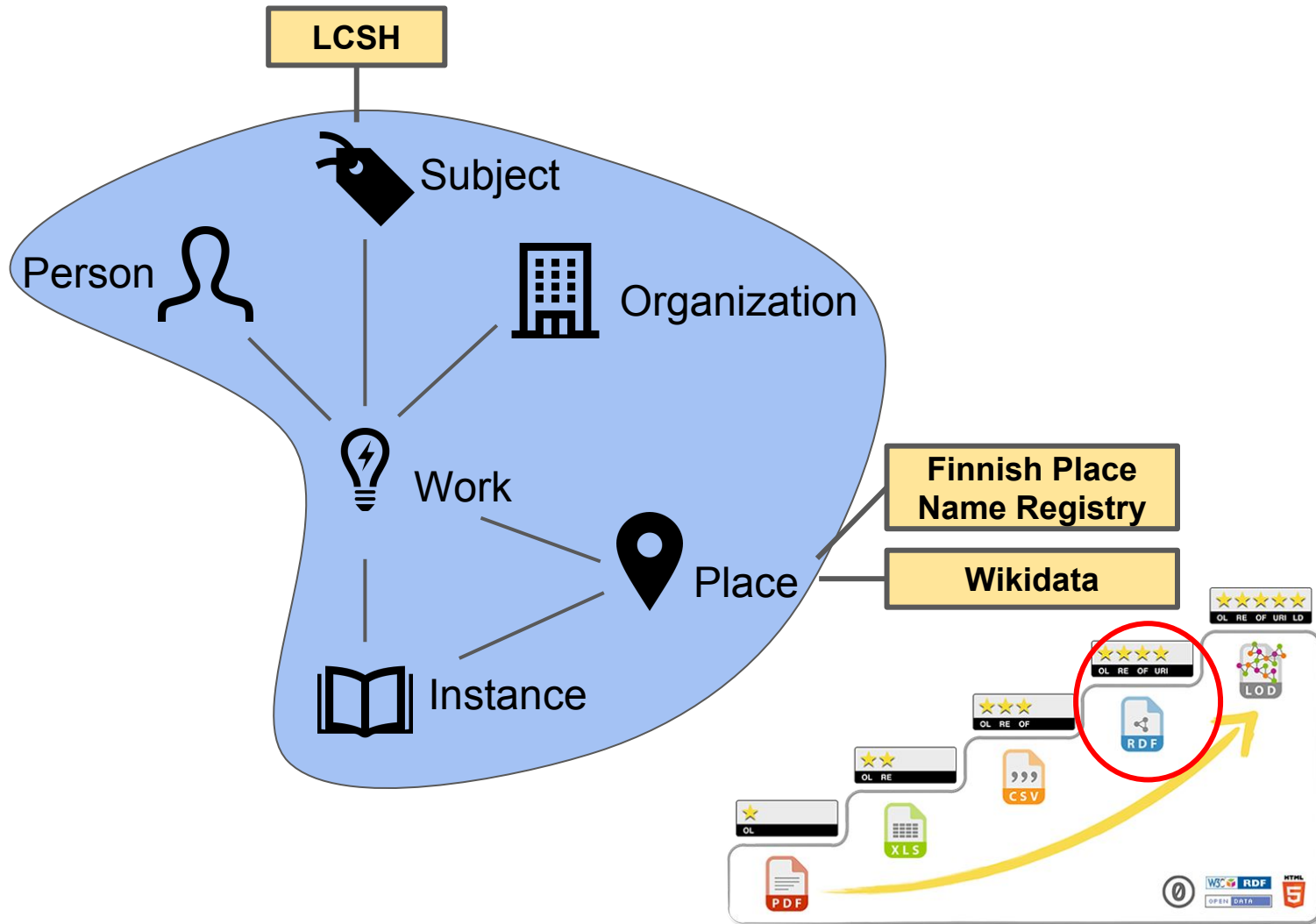
Work extraction

1. Extract works from MARC records
2. Create a work authority
3. Use and maintain it for cataloging
4. ???
5. Profit!

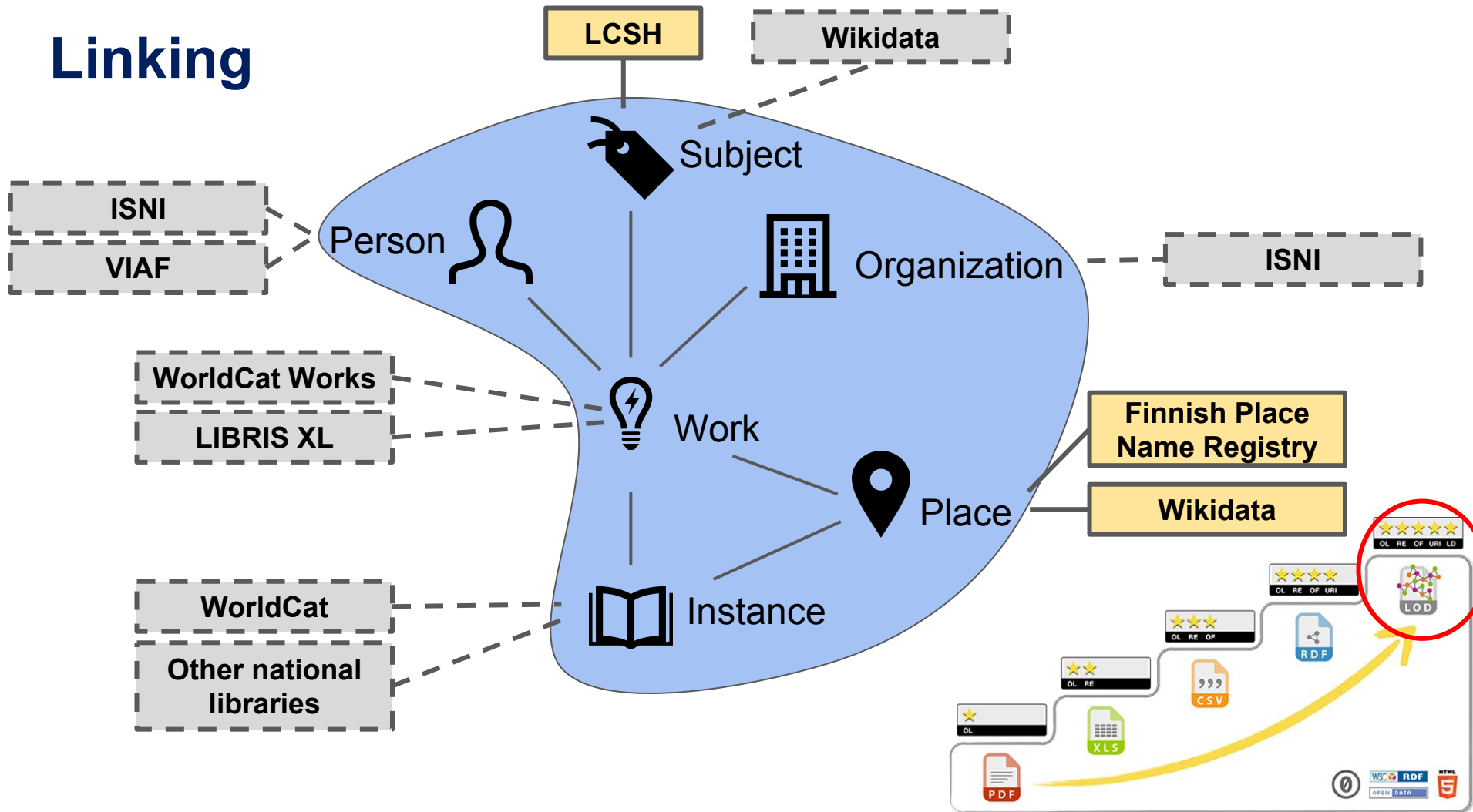
Not so easy in practice.

Lots of problems in the metadata that cause inconsistencies in the output.

Linking



Linking



“Cool URIs don’t change” -- Tim Berners-Lee



but we rely on conversion of MARC records that change all the time!

It's open data.

Is it FAIR?

1. Findable:

URIs as identifiers, with rich metadata

2. Accessible:

URI lookup, SPARQL and LDF endpoints, downloadable data dumps

3. Interoperable:

RDF representation using Schema.org and a little bit of RDAu

4. Reusable:

CC0 license. Entities that are references also from other metadata

What next?

1. Enriching and cleaning the RDF data, e.g. using subclasses like Map
2. More links to other Linked Data sets
3. Expanding to new data sets: Viola discography, Arto article database

My birthday present

Suomi
Finland
100

The Finnish Declaration of Independence was adopted by the Parliament of Finland on 6 December 1917

Thank you!

Questions?

osma.suominen@helsinki.fi - [@OsmaSuominen](https://twitter.com/OsmaSuominen)

<http://data.nationallibrary.fi> - [@NatLibFiData](https://twitter.com/NatLibFiData)

Code:

<https://github.com/NatLibFi/bib-rdf-pipeline>

<https://github.com/NatLibFi/bib-lod-ui>

These slides: <http://tinyurl.com/fennica-ld>