

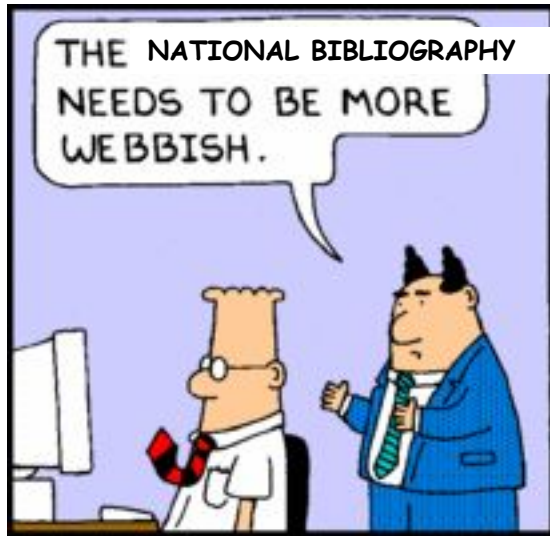
Finnish National Bibliography Fennica as Linked Open Data

Osma Suominen

HELDIG Summit, 23 October 2018



THE NATIONAL LIBRARY OF FINLAND



www.dilbert.com
scottadams@aol.com

12/3/99 © 1999 United Feature Syndicate, Inc.

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

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

125k person names

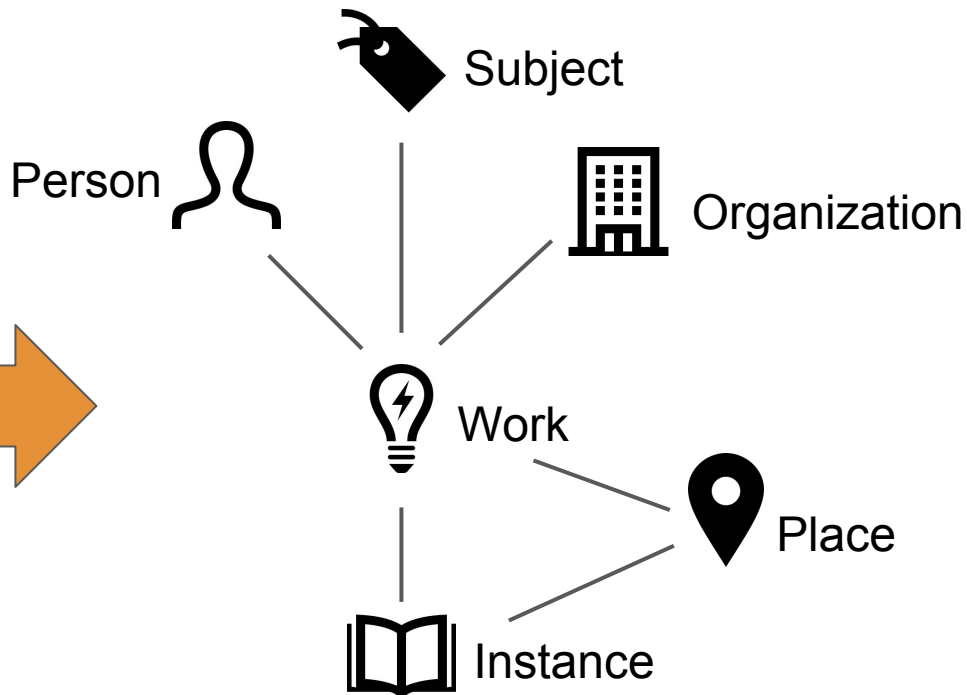
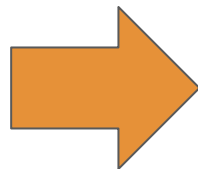
auth record
auth record

40k corporate names

auth record
auth record

35k subjects (YSA)

auth record
auth record

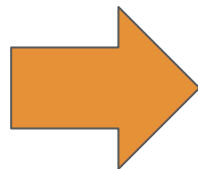


1M bib records

125k person names

bib record

auth record



bib record

auth record

bib record

auth record

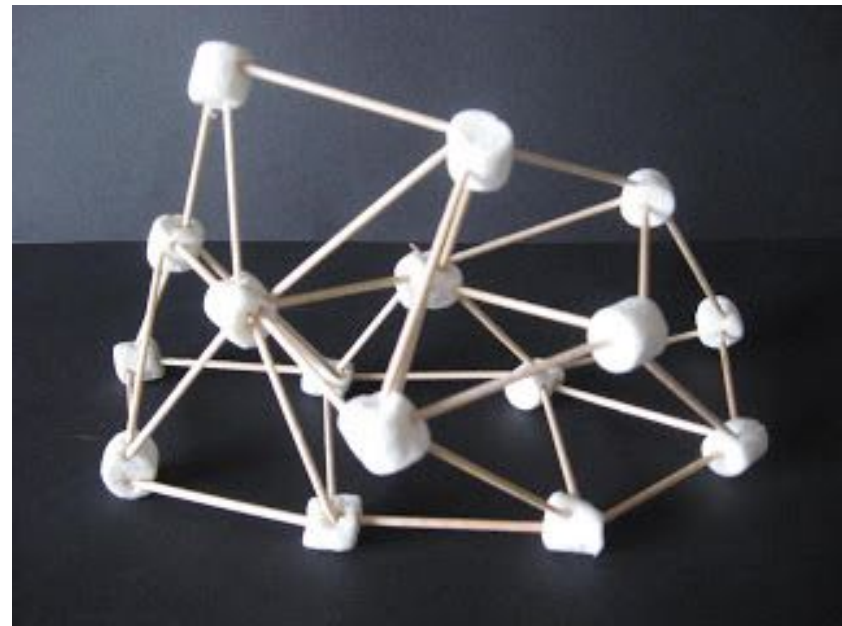
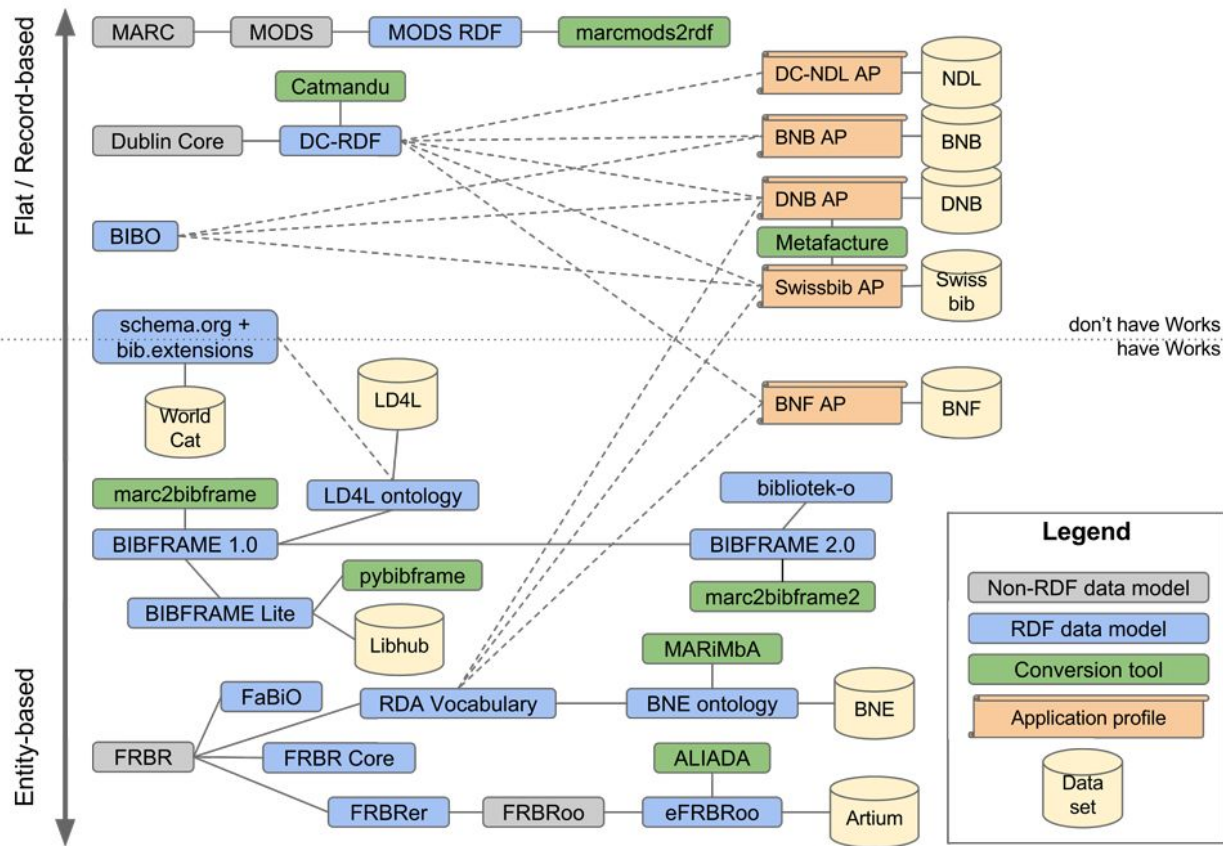


Image credit: [MaryMaking blog](#)



“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!)

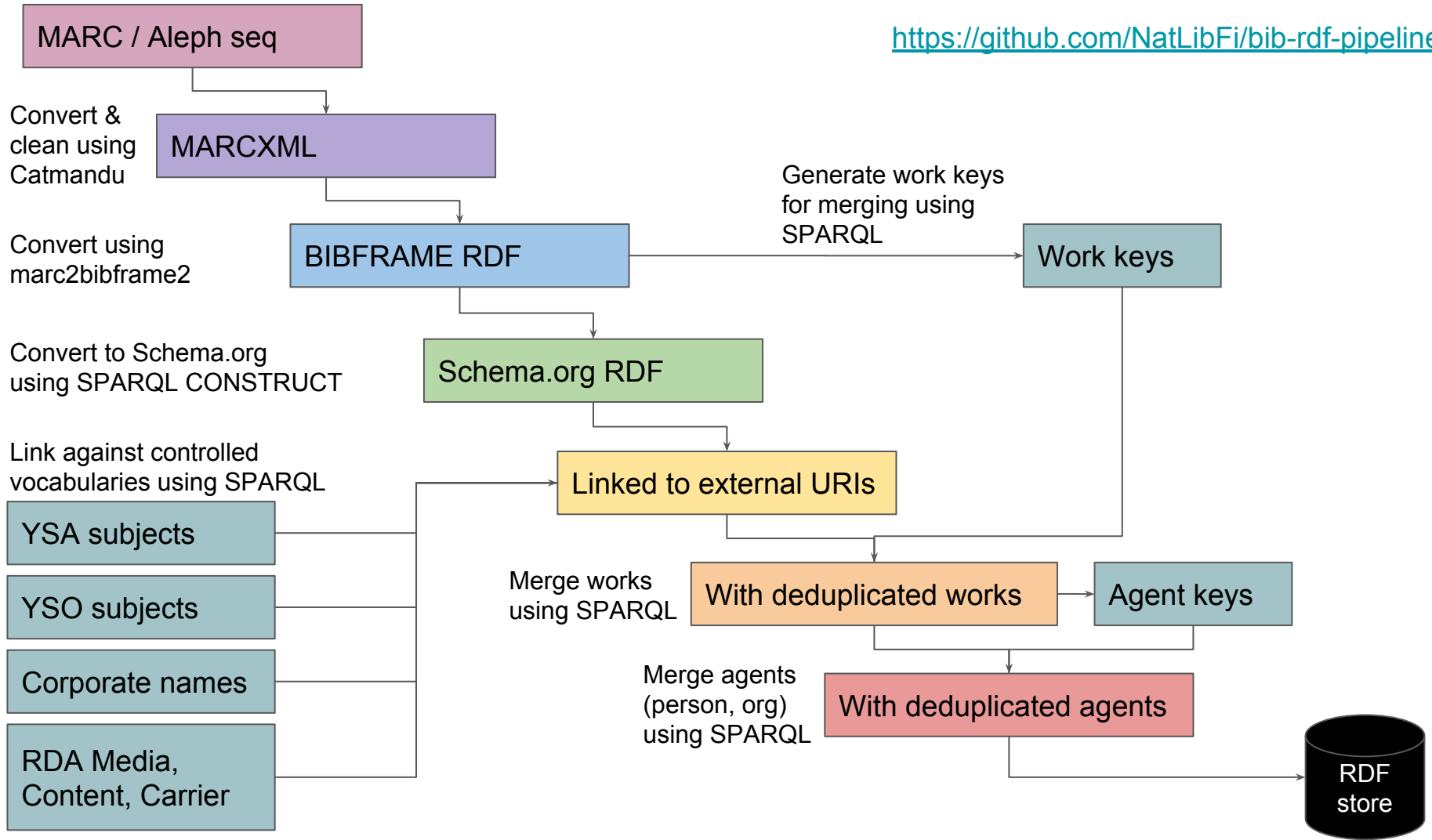


Special thanks to [Richard Wallis](#)
for help with applying schema.org!

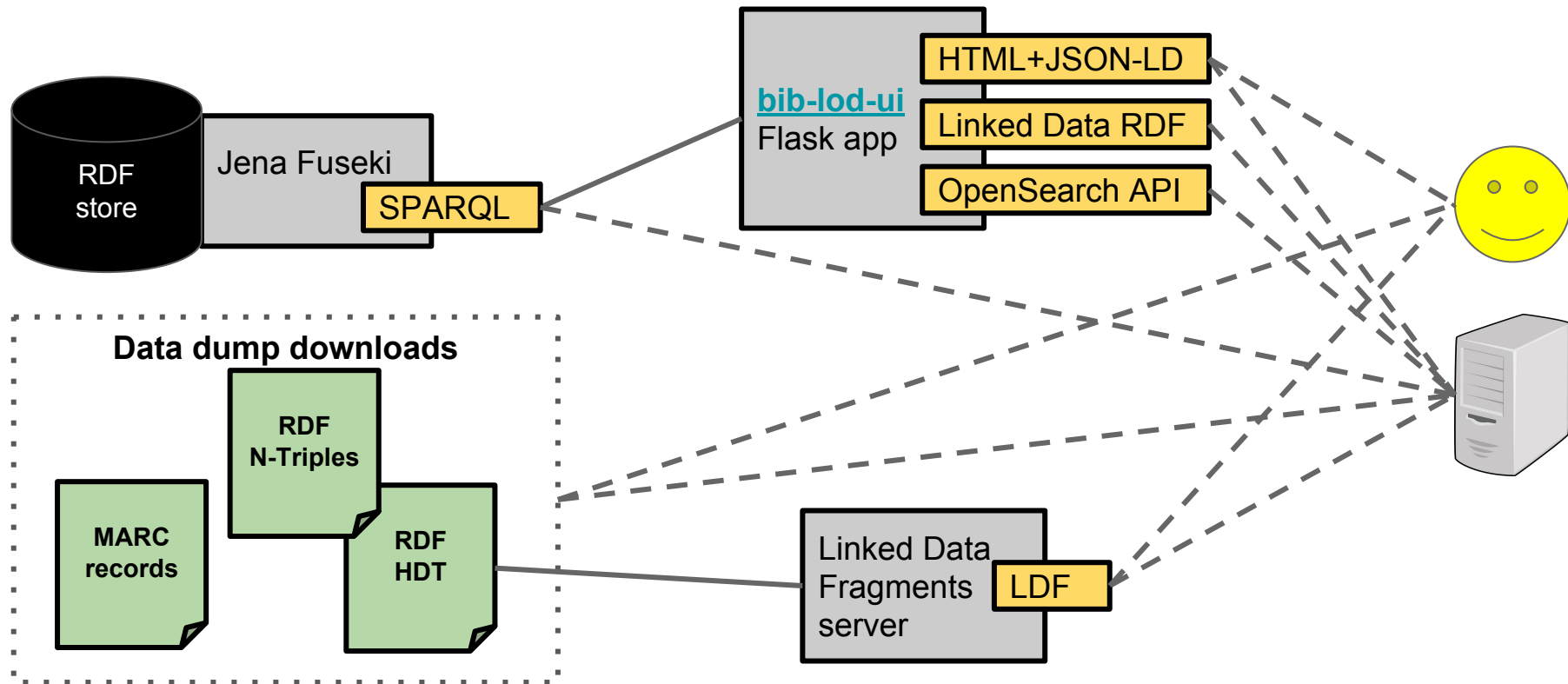
MARC →



→ **Linked
Data**



Publishing as Linked Open Data for human & machine access



Demo

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

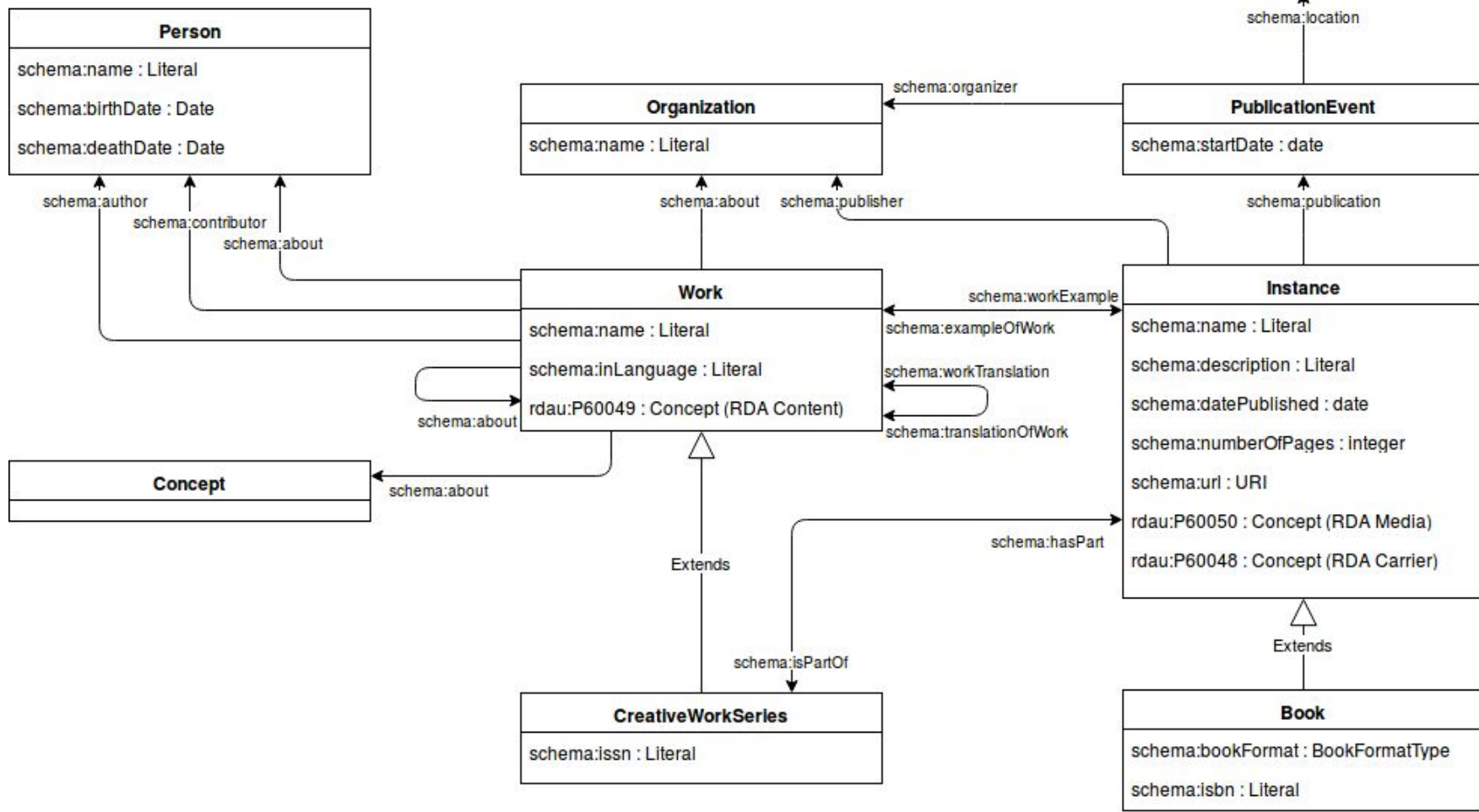
Downloadable dumps

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

Data model documentation



Challenges

Identity management

Libraries have traditionally managed identities (e.g. persons, works, places, subjects) by using authorized names and headings - i.e. **strings**.

This is a fragile way to assert identity. It would be better to represent **things** and give them persistent **identifiers**. This is not yet standard practice in MARC.

We have a relatively large number of **duplicate** persons and works in the data set:

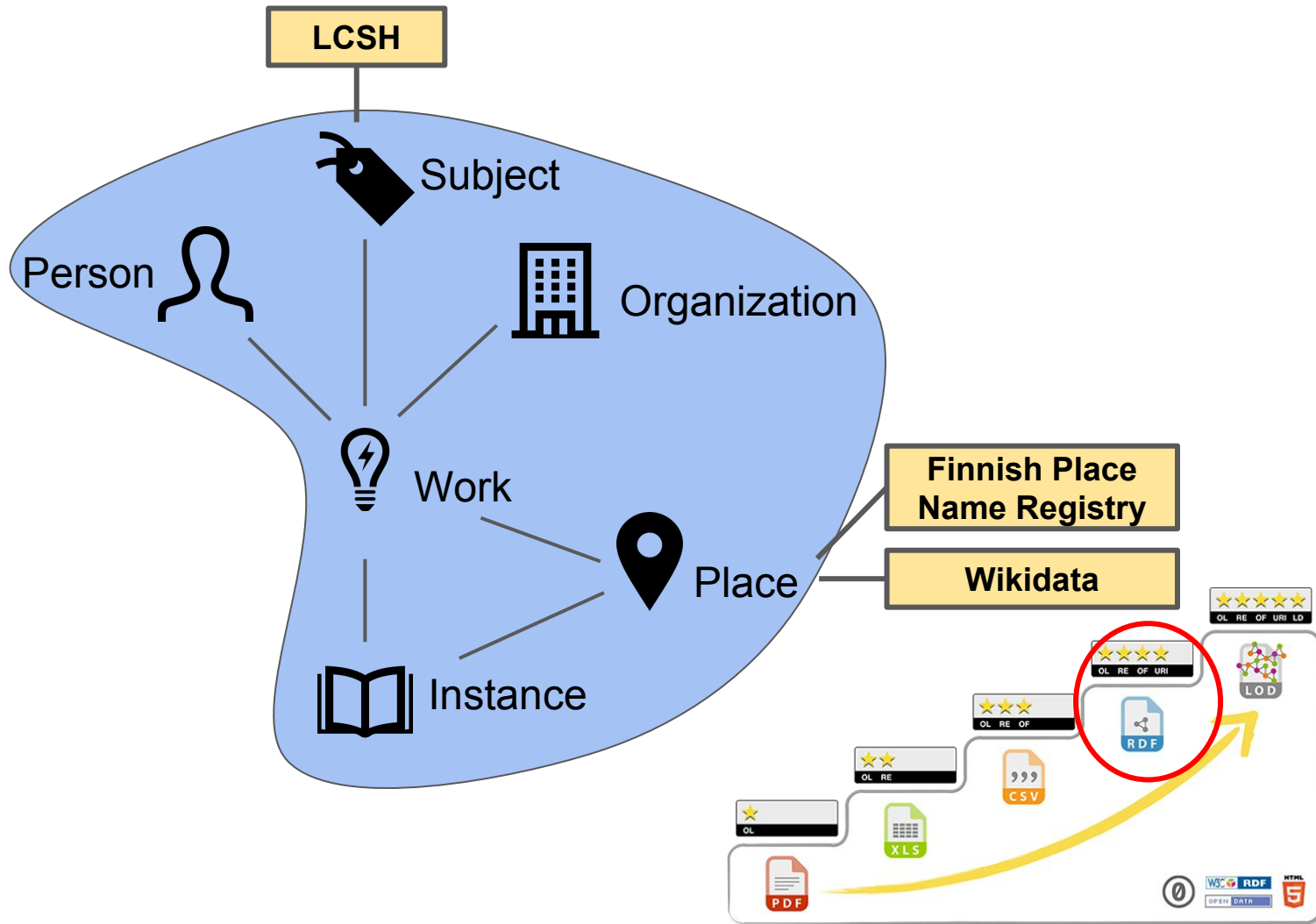
- cannot know for certain if **persons** with the same name are really the same
- extracting **works** from traditional bibliographic records is a hard problem

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

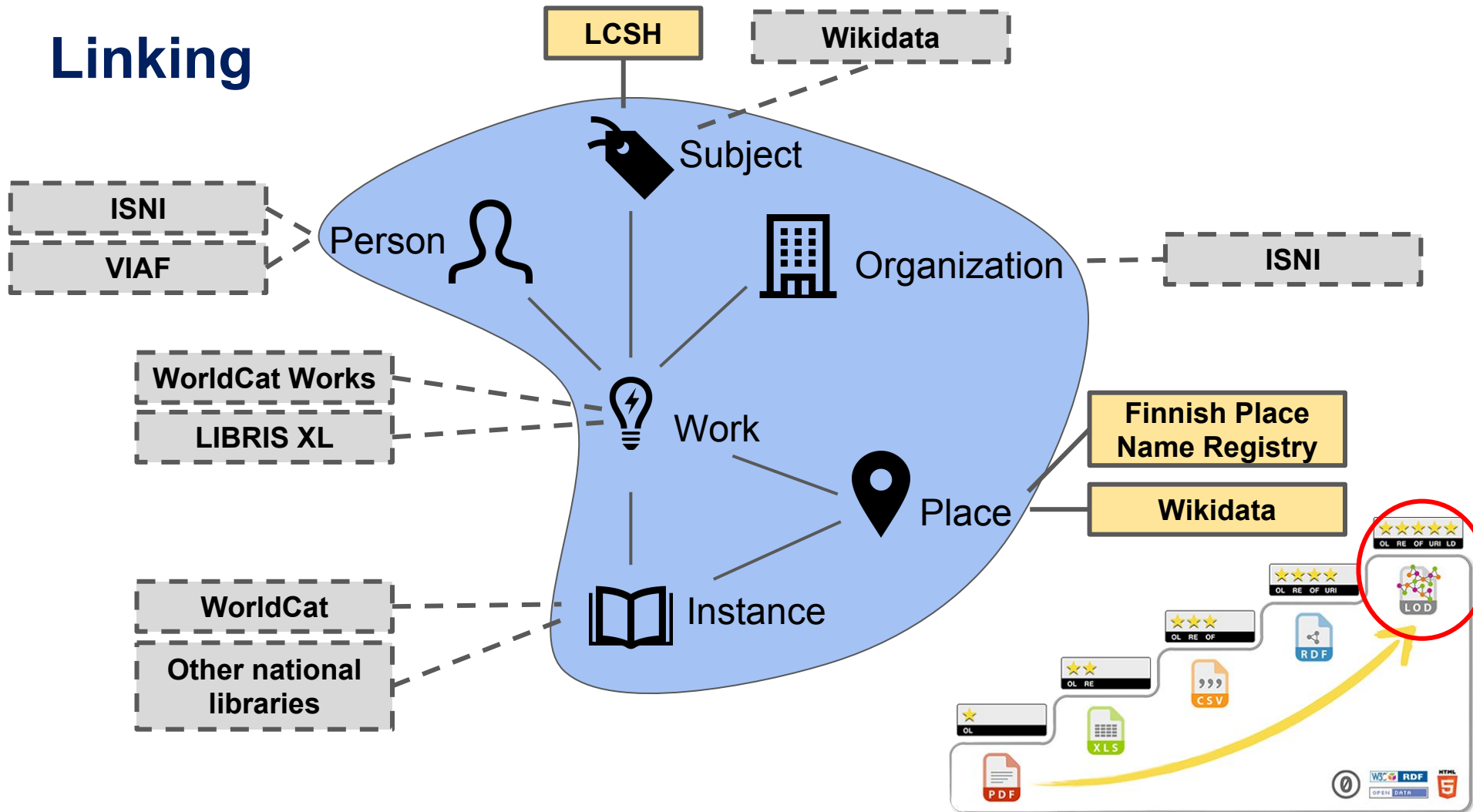


...but we rely on conversion of MARC records that change all the time!

Linking



Linking



Opportunities

SPARQL endpoint


<http://data.nationallibrary.fi/bib/sparql>

SPARQL query form




Use the form below to write a query. Some example queries to get you started:



- Authors of works with the same subjects as works by Esko Valtaoja
- Authors with more than 500 works
- EBooks about cosmology
- Persons featuring in more than 100 works
- Subjects with more than 5000 works
- Works by Aleksis Kivi
- Works by authors born before 1550

Query × +



```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 SELECT * WHERE {
4   ?sub ?pred ?obj .
5 }
6 LIMIT 10
```

 **Table** Response Pivot Table Google Chart Geo 

Persons featured in >100 works

	person	personName	works
1	http://urn.fi/URN:NBN:fi:au:pn:000105186	Jeesus Kristus	"396" ^{xsd:integer}
2	http://urn.fi/URN:NBN:fi:au:pn:000067771	Mannerheim, Carl Gustaf Emil	"365" ^{xsd:integer}
3	http://urn.fi/URN:NBN:fi:au:pn:000068760	Aalto, Alvar	"355" ^{xsd:integer}
4	http://urn.fi/URN:NBN:fi:au:pn:000058877	Sibelius, Jean	"303" ^{xsd:integer}
5	http://urn.fi/URN:NBN:fi:au:pn:000066777	Kekkonen, Urho	"290" ^{xsd:integer}
6	http://urn.fi/URN:NBN:fi:au:pn:000051426	Luther, Martti	"200" ^{xsd:integer}
7	http://urn.fi/URN:NBN:fi:au:pn:000057305	Runeberg, Johan Ludvig	"198" ^{xsd:integer}
8	http://urn.fi/URN:NBN:fi:au:pn:000059409	Snellman, Johan Vilhelm	"194" ^{xsd:integer}
9	http://urn.fi/URN:NBN:fi:au:pn:000067546	Gallen-Kallela, Akseli	"166" ^{xsd:integer}
10	http://urn.fi/URN:NBN:fi:au:pn:000051570	Lönnrot, Elias	"160" ^{xsd:integer}
11	http://urn.fi/URN:NBN:fi:au:pn:000061174	Topelius, Zacharias	"136" ^{xsd:integer}
12	http://urn.fi/URN:NBN:fi:au:pn:000068031	Kivi, Aleksis	"136" ^{xsd:integer}
13	http://urn.fi/URN:NBN:fi:au:pn:000095240	Kustaa III, Ruotsin kuningas	"133" ^{xsd:integer}
14	http://urn.fi/URN:NBN:fi:au:pn:000073620	Heidegger, Martin	"117" ^{xsd:integer}
15	http://urn.fi/URN:NBN:fi:au:pn:000045590	Jansson, Tove	"105" ^{xsd:integer}
16	http://urn.fi/URN:NBN:fi:au:pn:000047948	Koivisto, Mauno	"105" ^{xsd:integer}
17	http://urn.fi/URN:NBN:fi:au:pn:000047948	Koivisto, Mauno, 1923-2017	"105" ^{xsd:integer}

Showing 1 to 17 of 17 entries (in 6.051 seconds)

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

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-heldig>