From MARC to Schema.org

-- via BIBFRAME!

Osma Suominen

European BIBFRAME workshop

Frankfurt, September 26, 2017

My assignment

with apologies to Scott Adams

Schema.org forces to think about data from a Web user's point of view

~~"We have these 1M bibliographic records"~~

"The National Library maintains this amazing collection of literary works!
We have these editions of those works in our collection.
They are available free of charge for reading/borrowing
from our library building (Unioninkatu 36, 00170 Helsinki, Finland)
which is open Mon-Fri 10-17, except Wed 10-20.
The electronic versions are available online from these URLs."

From MARC to Schema.org - via BIBFRAME!

To convert to Schema.org, we first need to break down the MARC records into some (any!) kind of RDF data, without losing any important information.

BIBFRAME converters do a fairly good job of this!

1. Zepheira's **pybibframe** was tested briefly. It was rather slow and seems to lose more information than I'd like. Does some internal reconciliation.

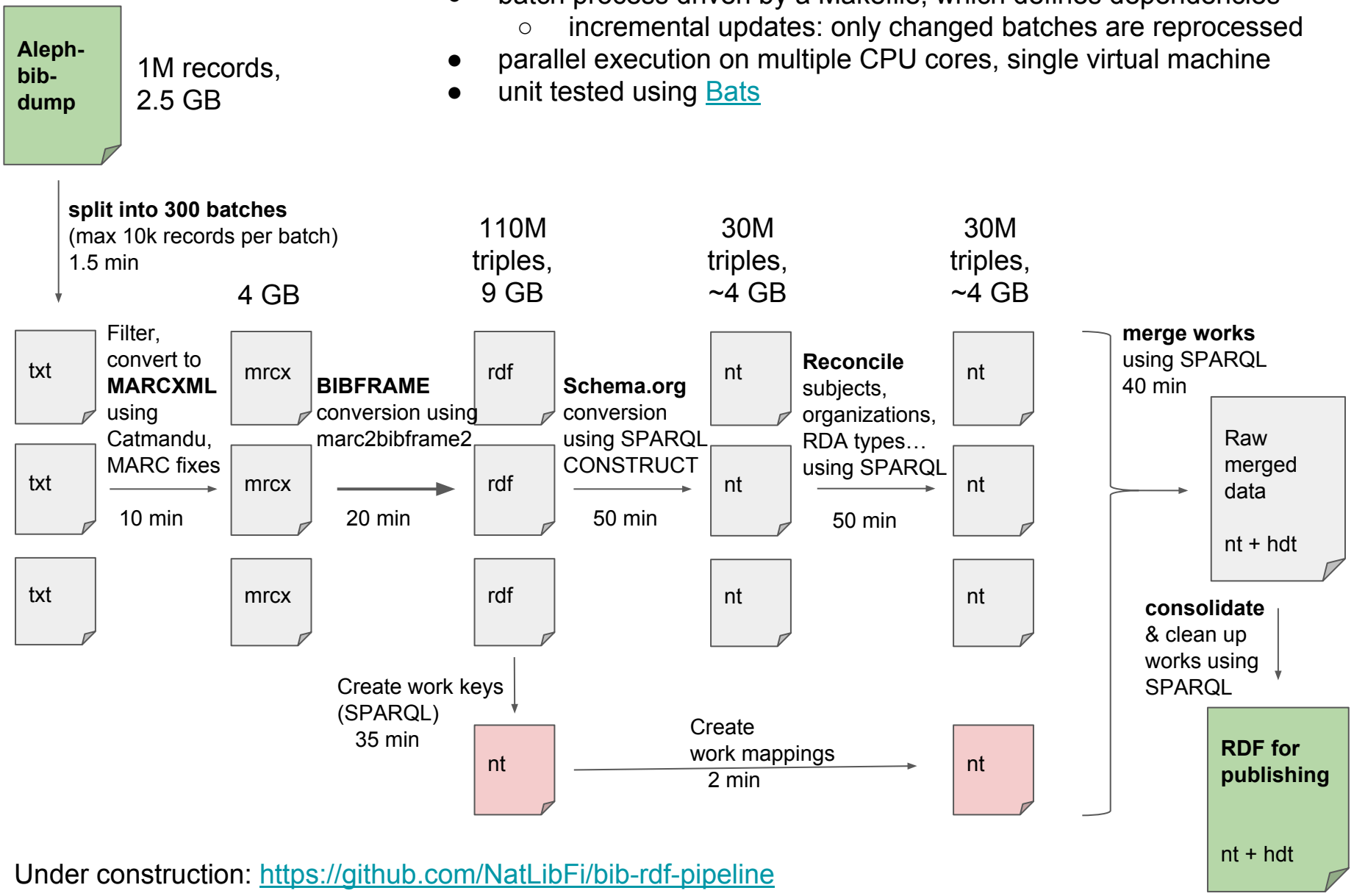
2. We used LoC's **marc2bibframe** for some time. Together with a wrapper, it has relatively good performance and consistent, but quite verbose RDF output. Not maintained anymore!

3. LoC and Index Data released marc2bibframe2, a BIBFRAME 2.0 converter in March and we started using it the following week. Fast, works well, some issues were reported and quickly fixed!

4. LD4L-Labs is working on the bib2lod converter, from MARC to their flavor of BIBFRAME 2.0. Following closely!

Fennica RDF conversion pipeline (draft)

- batch process driven by a Makefile, which defines dependencies
  - incremental updates: only changed batches are reprocessed
- parallel execution on multiple CPU cores, single virtual machine
- unit tested using Bats



Under construction: https://github.com/NatLibFi/bib-rdf-pipeline

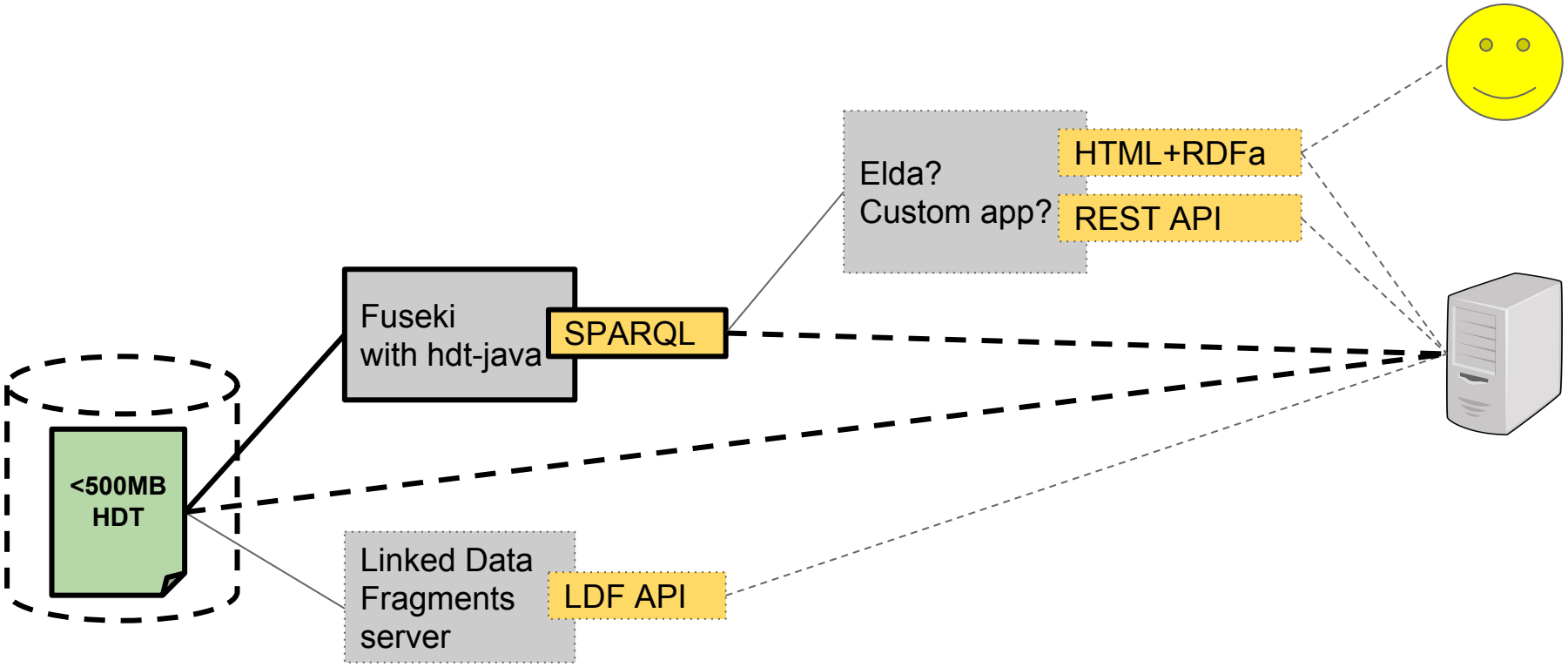
Issues in BIBFRAME

The marc2bibframe2 converter produces quite verbose output, including some redundancy. Information from one MARC field may be repeated up to 8 times. Works are not extracted and merged by the converter. Not good for further maintenance!

Relationship between BIBFRAME and RDA is unclear. Is BIBFRAME the best way to represent RDA-conformant metadata as RDF?

We have strong collaboration with museums and archives. BIBFRAME is very focused on library materials.

Publishing as LOD



Bolded parts already exist

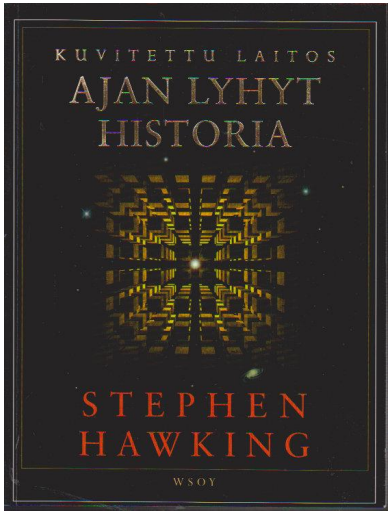
http://linkeddata-kk.lib.helsinki.fi (temporary site)

Thank you!
Questions?

osma.suominen@helsinki.fi

code: https://github.com/NatLibFi/bib-rdf-pipeline
these slides: http://tinyurl.com/marc-bf-schema

Fennica using Schema.org




```
# The original English language work
fennica:000215259work9 a schema:CreativeWork ;
  schema:about ysa:Y94527, ysa:Y96623, ysa:Y97136,
               ysa:Y97137, ysa:Y97575, ysa:Y99040,
               yso:p18360, yso:p19627, yso:p21034,
               yso:p2872, yso:p4403, yso:p9145 ;
  schema:author fennica:000215259person10 ;
  schema:inLanguage "en" ;
  schema:name "The illustrated A brief history of time" ;
  schema:workTranslation fennica:000215259 .

# The Finnish translation (~expression in FRBR/RDA)
fennica:000215259 a schema:CreativeWork ;
  schema:about ysa:Y94527, ysa:Y96623, ysa:Y97136,
               ysa:Y97137, ysa:Y97575, ysa:Y99040,
               yso:p18360, yso:p19627, yso:p21034,
               yso:p2872, yso:p4403, yso:p9145 ;
  schema:author fennica:000215259person10 ;
  schema:contributor fennica:000215259person11 ;
  schema:inLanguage "fi" ;
  schema:name "Ajan lyhyt historia" ;
  schema:translationOfWork fennica:000215259work9 ;
  schema:workExample fennica:000215259instance26 ;
  rdau:P60049 rdacontent:1020 .

# The manifestation (FRBR/RDA) / instance (BIBFRAME)
fennica:000215259instance26 a schema:Book, schema:CreativeWork ;
  schema:author fennica:000215259person10 ;
  schema:contributor fennica:000215259person11 ;
  schema:datePublished "2000" ;
  schema:description "Lisäpainokset: 4. p. 2002. - 5. p. 2005." ;
  schema:exampleOfWork fennica:000215259 ;
  schema:isbn "9510248215", "9789510248218" ;
  schema:name "Ajan lyhyt historia" ;
  schema:numberOfPages "248, 6 s. :" ;
  rdau:P60048 rdacarrier:1007 ;
  schema:publisher [
        schema:name "WSOY" ;
        a schema:Organization
  ] .

# The original author
fennica:000215259person10 a schema:Person ;
  schema:name "Hawking, Stephen" .

# The translator
fennica:000215259person11 a schema:Person ;
  schema:name "Varteva, Risto" .
```

Special thanks to Richard Wallis
for help with applying schema.org!