



Co-funded by the Horizon 2020
Framework Programme of the European Union
Grant Agreement Number 644771



FREME TO MAKE LINKED DATA AVAILABLE TO LOCALIZERS

FEISGILTT 2015 | BERLIN, 3 JUNE 2015

www.freme-project.eu



Felix Sasaki, DFKI / W3C Fellow
On behalf of the FREME Consortium

REMINDER: WHAT IS FREME?

- More info: see presentation from yesterday

<http://slideshare.net/atcfsenzoku/freme-at-feisgilTT-2015-freme-use-cases>

- Design of FREME takes up work from other projects

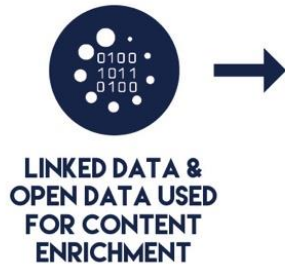
1. LIDER <http://lider-project.eu/>

- In FREME, we deploy best Practices on how to work with linguistic linked data (LLD)
- LLD: Linked data used to represent lexica, corpora, language processing workflows etc.

2. FALCON <http://falcon-project.eu/>

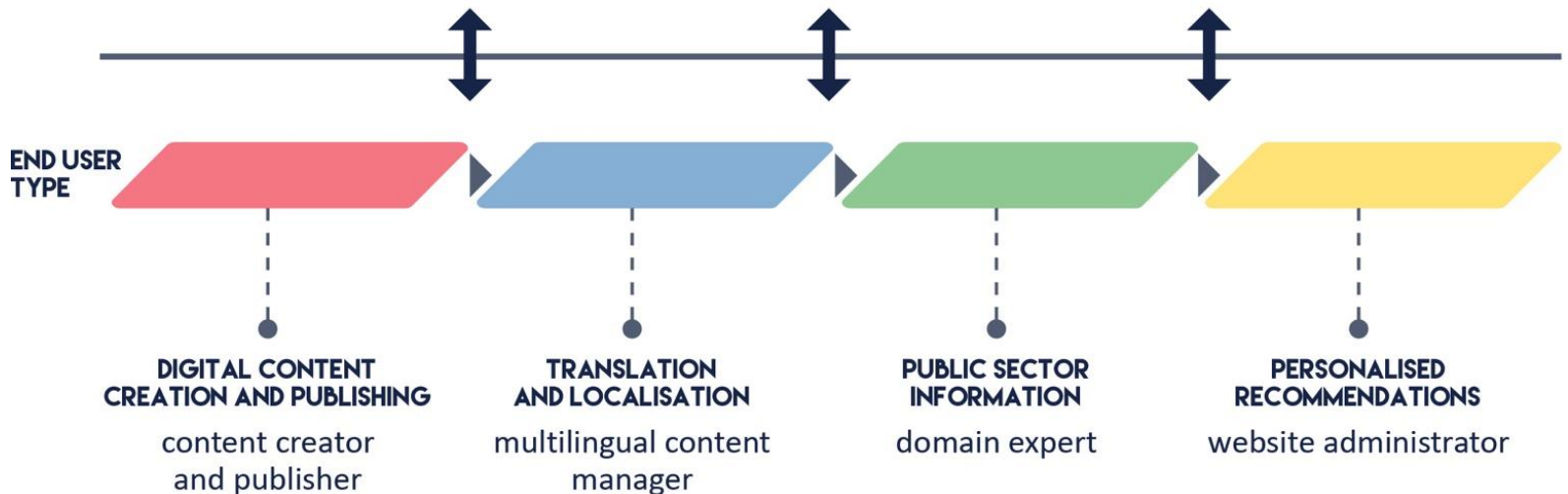
- In FREME, we benefit from experience on working with linked data in localisation scenarios
- One lesson learned: hide linked data in the right way from (localisation) developers
- No need to process linked data always in the native form, see Babelfy <http://babelfy.org/>

FRAME



USER TYPE
interface
developer

USER TYPE
technology
user



FREME E-SERVICES – BIRDS EYE VIEW

- e-Entity
 - Automatic annotation of named entities
- e-Terminology
 - Annotation of terms and linkage to term databases
- e-Link
 - Enrichment with information from (linked) (open) data sources*
- e-Translation
 - Cloud based machine translation
- e-Internationalisation
 - ITS 2.0 metadata to govern the multilingual & semantic content workflow
- e-Publishing
 - Publish enriched content in ePub format

EXAMPLE: E-SERVICE DESIGN

- RESTFul API
- Example

<http://api.freme-project.eu/0.1/e-entity/dbpedia-spolight>

- Under each service endpoint: **tool specific** versions
- Parameters for e-Entity
 - Confidence threshold
 - Informat. Currently text or NIF (explanation see next slides)
 - Outformat.
- Output: NIF in various serializations
 - Currently text/turtle or application/json+ld

WHAT IS NIF?

- Natural Language Processing Interchange Format
- “The XLIFF of natural language processing workflows” (Phil Ritchie, FEISGILTT 2015)
- NIF: Linked data based representation of digital content and NLP related annotations
- Anchoring in source format possible -> basis for roundtripping
- More info: see <http://site.nlp2rdf.org/>

NIF EXAMPLE: DESCRIBING DOCUMENTS

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#> .
<http://example.org/document/1#char=0,11>
  a nif:String , nif:Context , nif:RFC5147String ;
  nif:isString "the content"^^xsd:string;
  nif:beginIndex "0"^^xsd:nonNegativeInteger;
  nif:endIndex "11"^^xsd:nonNegativeInteger;
  nif:sourceUrl <http://differentday.blogspot.com/2007_01_01_archive.html> .
```

DESCRIBING STRINGS

<http://example.org/document/1#char=0,21>

a nif:String , nif:Context , nif:RFC5147String ;

nif:isString "We talk about Xiamen."^^xsd:string;

nif:beginIndex "0"

nif:endIndex "21"

nif:sourceUrl <http://differentday.blogspot.com/2007_01_01_archive.html> .

<http://example.org/document/1#char=14,20> a nif:String , nif:RFC5147String ,
nif:Word, nif:Phrase;

nif:referenceContext <http://example.org/document/1#char=0,21> ;

nif:anchorOf "Xiamen" ;

nif:beginIndex "14" ;

nif:endIndex "20";

nif:wasConvertedFrom

<http://example.org?t=url&f=html&i=http://somewebpage.com#char=0,2820> ;

STORING E-ENTITY ENRICHMENT

```
<http://example.org/document/1#char=14,20> a nif:String , nif:RFC5147String ,  
nif:Word, nif:Phrase;
```

```
itsrdf:taIdentRef <http://dbpedia.org/resource/Xiamen> ;
```

```
itsrdf:taClassRef <http://dbpedia.org/ontology/City> ;
```

```
itsrdf:taClassRef <http://dbpedia.org/ontology/Settlement> ;
```

```
itsrdf:taClassRef <http://dbpedia.org/ontology/PopulatedPlace> ;
```

```
itsrdf:taClassRef <http://dbpedia.org/ontology/Place> .
```

- NIF allows to add multiple annotations to content
- No constraints on the structure of annotations

KEEPING PROVENANCE

```
<http://example.org/document/1#char=0,21> ...
```

```
nif:wasConvertedFrom
```

```
<http://example.com/?informat=html&intype=url&
```

```
input=
```

```
http://differentday.blogspot.com/2007_01_01_archive.html/
```

```
&xpath=/html/body[1]/h2[1]/span[1]/text()[1]>.
```

- XPath only an example
- nif:wasConvertedFrom can hold source format specific information
- Can be the basis for round tripping

BENEFIT AND DRAWBACKS OF NIF

Benefits

- NIF can store all information of enrichment services
 - e-Entity, e-Link, e-Terminology, e-Translation
- Via NIF we can chain services easily
 - No constraints on structures: NIF format constitutes general annotation structure

Drawbacks

- No tool support of heterogeneous input formats in current tooling
 - Working on that 😊 -> integration of Okapi and NIF tooling
- Size of NIF annotations may be an issue
 - State: currently gathering implementation experience

DEMO: COMBINING E-SERVICES VIA NIF

- Try things yourself at <http://api.freme-project.eu/doc/0.1/>

Demo workflow:

1. Input: text
2. Processing via e-Entity
3. Output: NIF, input to step 4
4. Processing via e-Link
5. Output: NIF

HIDING COMPLEXITY (1/2): NIF AND E-SERVICE USER

FREME version 0.1: service endpoints understand text only or NIF content

<http://api.freme-project.eu/0.1/e-entity/dbpedia-spolight>

- Future version: support additional formats via integrating Okapi into NIF
 - Informat: HTML, XML, Word, PDF; ...
 - Outformat: NIF, in some cases (HTML, XML, ...) roundtripping
- API user sets input and output e.g. via Accept header
- NIF is processed internally, “hidden from the user”

HIDING COMPLEXITY (2/2): THE CASE OF E-LINK

- Many users don't know linked data sources:
 - What type of data is available?
 - What linked data vocabularies are used: NIF, LEMON, ...
 - What queries do I need to get information of type X
- FREME e-Link allows them to query linked data without looking at it
 - Input: content plus a query template : “Find my all events close to a given entity” , “Find me all museums close to a given entity”, ...
 - Output: content enriched with information relevant to the query, also as JSON-LD
- Concept of query templates: similar to “Schematron for information architects” approach, cf. George Bina at XML Prague 2015

<http://archive.xmlprague.cz/2015/files/xmlprague-2015-proceedings.pdf#page=199>

E-LINK EXAMPLE: TEMPLATE “PROVIDE GEO-INFORMATION FOR A GIVEN ENTITY”

```
<http://example.org/document/1#char=0,6> ...  
nif:anchorOf "Berlin"@en;  
nif:beginIndex "0";  
nif:endIndex "6"; itsrdf:taldentRef  
<http://dbpedia.org/resource/Berlin> .
```

<http://api.freme-project.eu/0.1/e-link/?outformat=turtle&templateid=1>



```
...  
<http://dbpedia.org/resource/Berlin>;  
itsrdf:taldentRef <http://dbpedia.org/resource/Berlin>;  
  geo:lat "52.516666";  
  geo:long "13.383333" .
```

LINKED DATA AND LOCALISATION: LESSONS LEARNED

- Integration of linked data and tooling: loose coupling wins
 - Localisation tools talking to linked data enabled web services
- Hide complexity in the right manner
 - Cf. e-Link template approach
- Give people “their output format” – probably json
 - json-ld to the rescue
- Linked data world can benefit from localisation tooling
 - Cf. work on OKAPI – NIF integration

CONTACTS

FELIX SASAKI

Senior Researcher DFKI / W3C Fellow

On behalf of the FREME consortium

E-mail: felix.sasaki@dfki.de



CONSORTIUM

