

Translation ~~Quality~~ and Linked Data

Phil Ritchie, CTO, VistaTEC
FEISGILTT, Dublin 2014

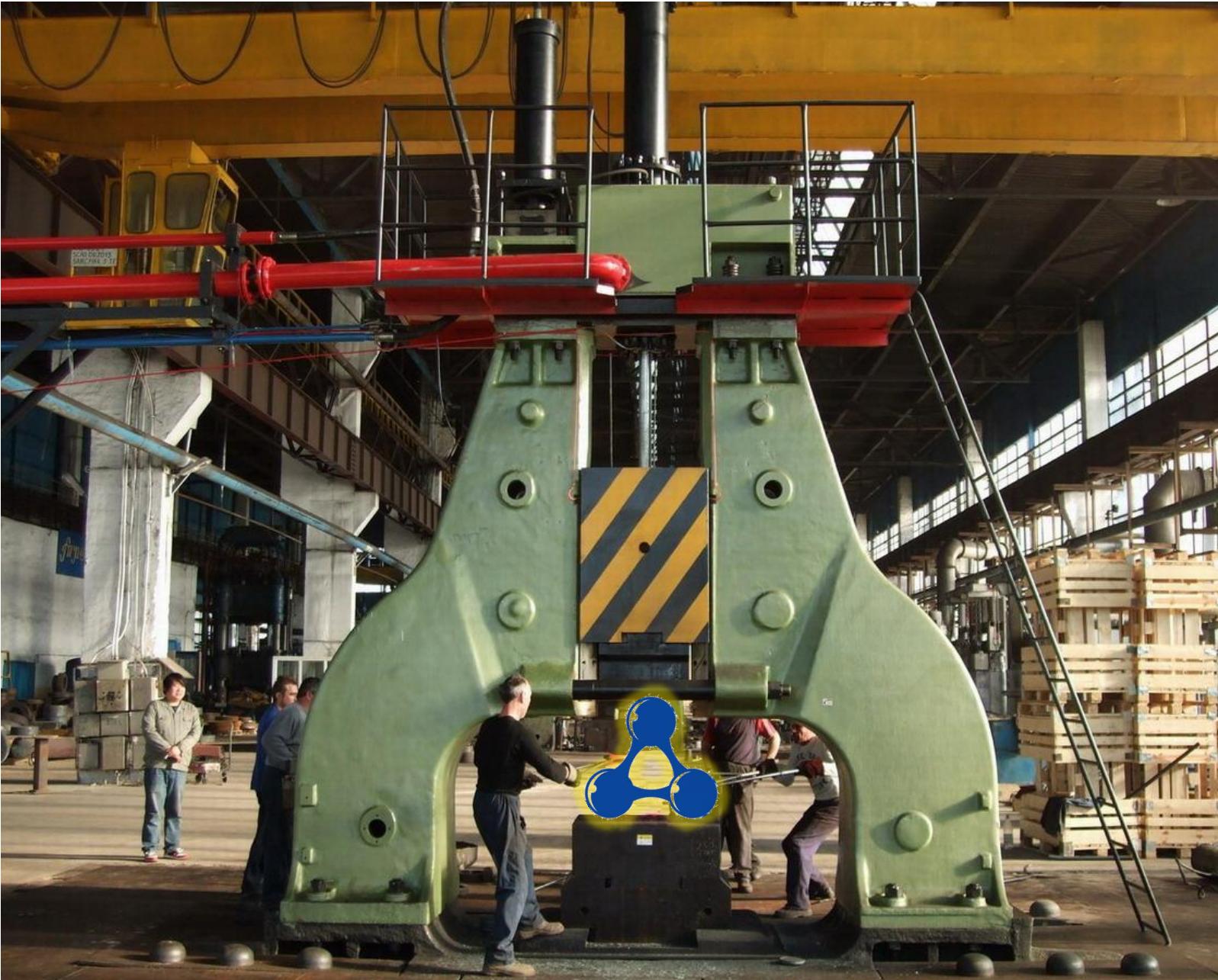
Scope

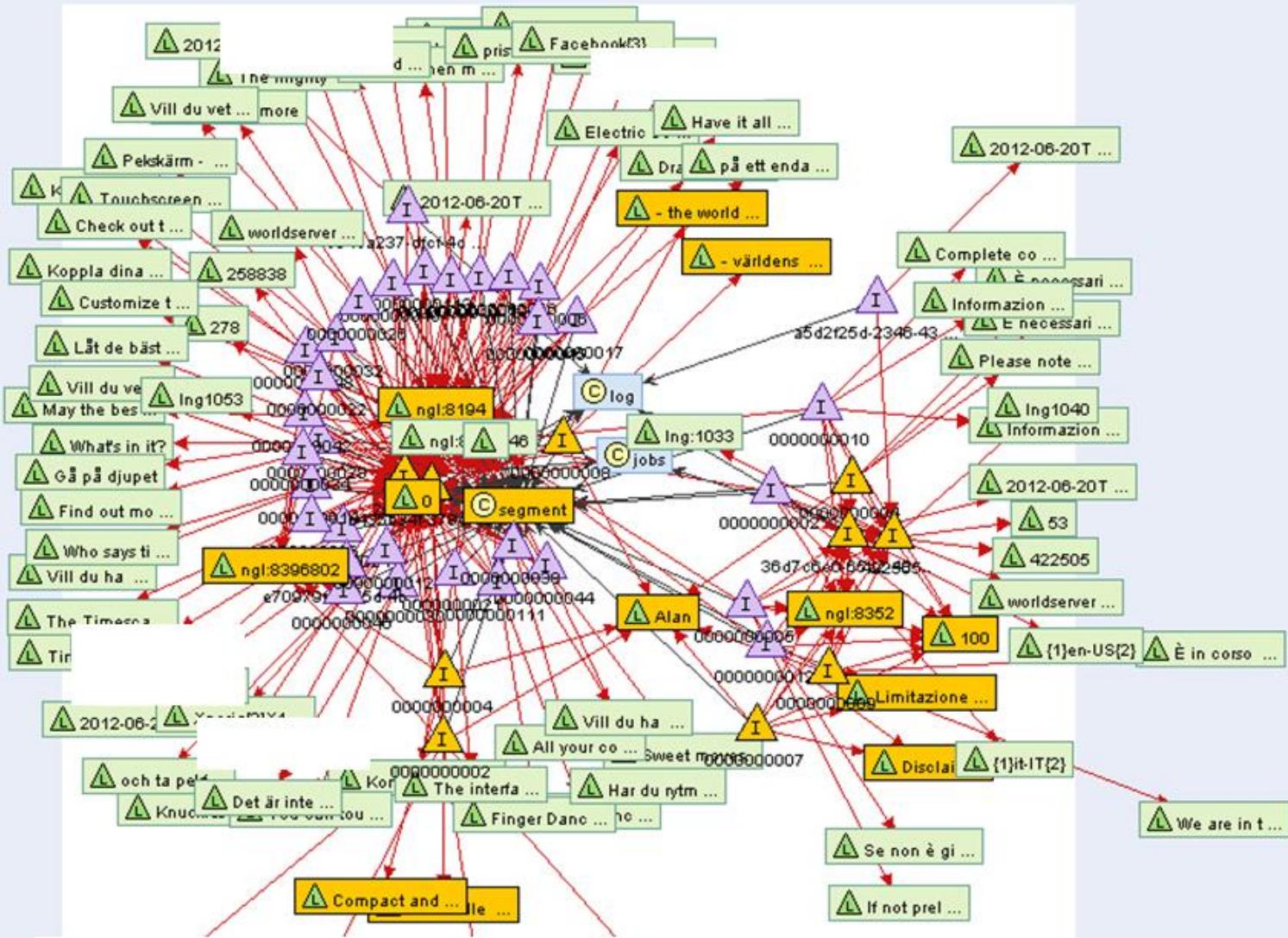
Customer brand adherent, domain accurate, audience engaging translations

Research

Industrial Partners







Focus

- Provenance
 - User agents engaged with producing or revising translations
 - Tools
 - People
 - Organizations
- Four interesting characteristics
 - SPARQL – query across repositories
 - Open World Assumption
 - Model Relationships
 - Inferential Queries

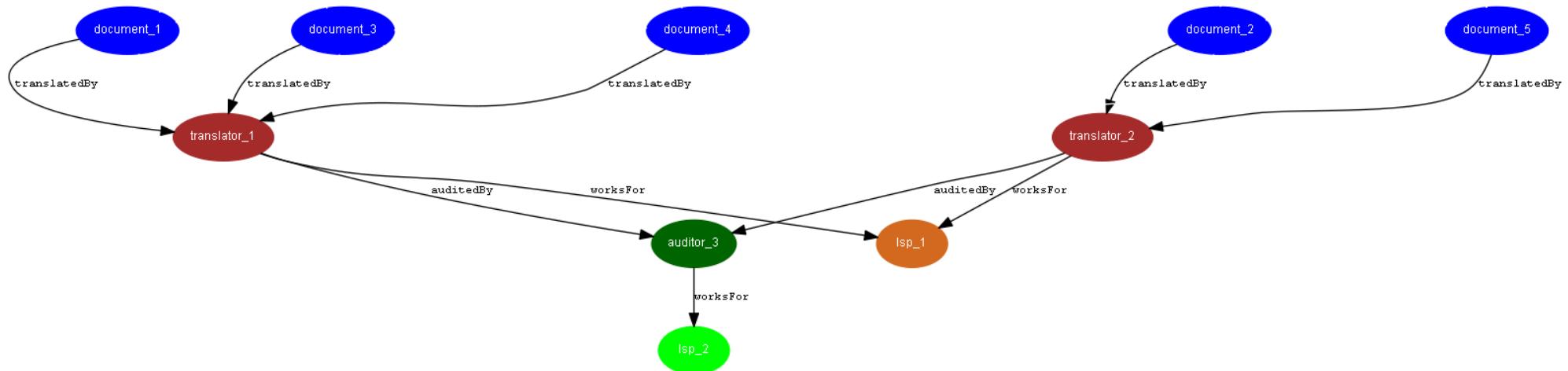
document_1

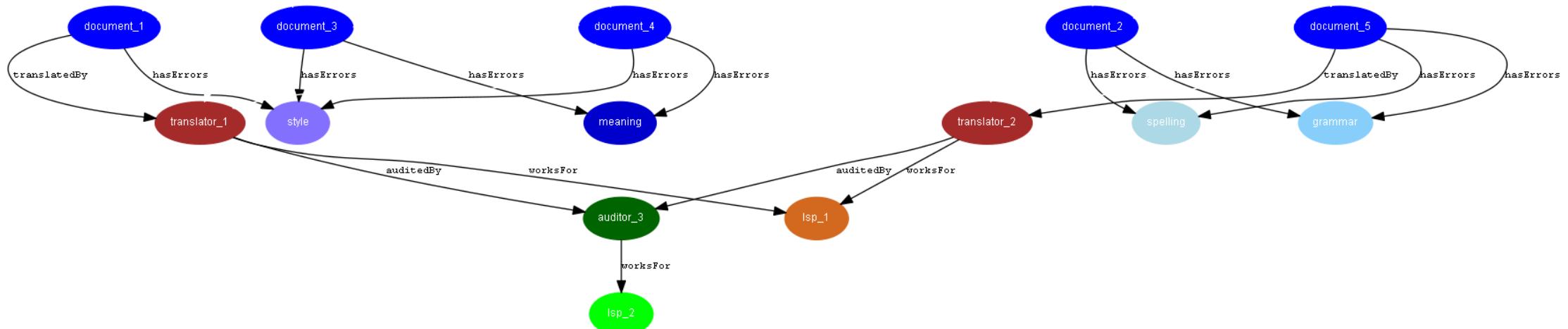
document_3

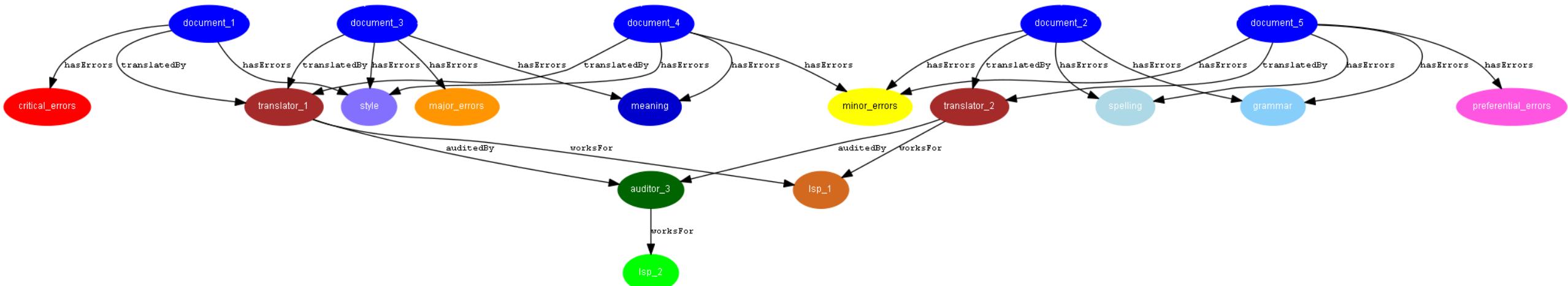
document_4

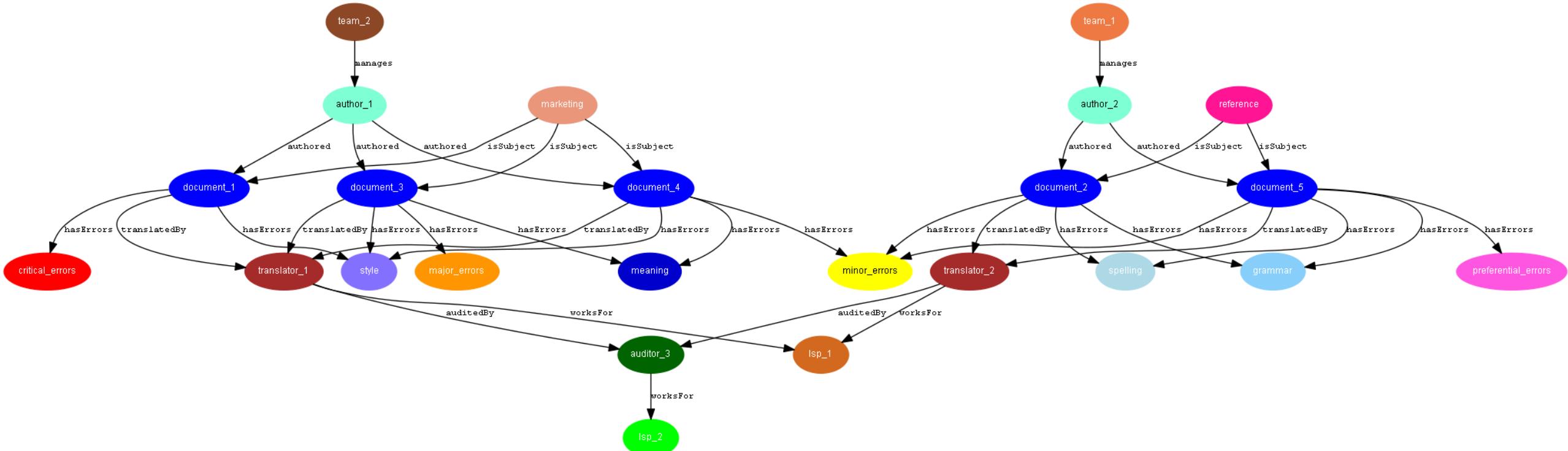
document_2

document_5



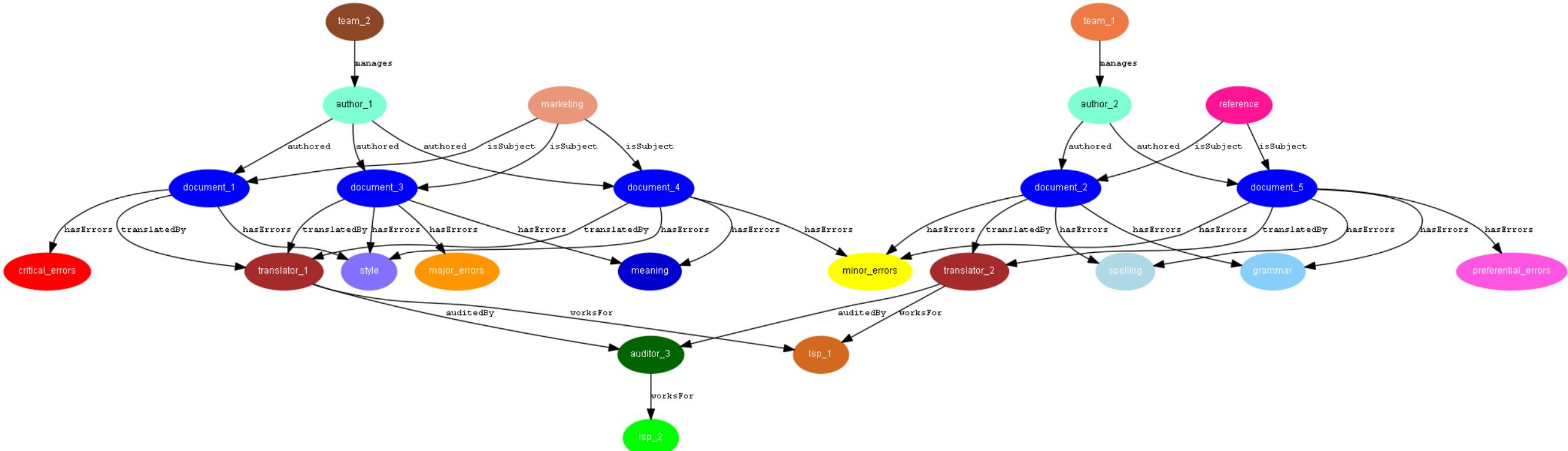


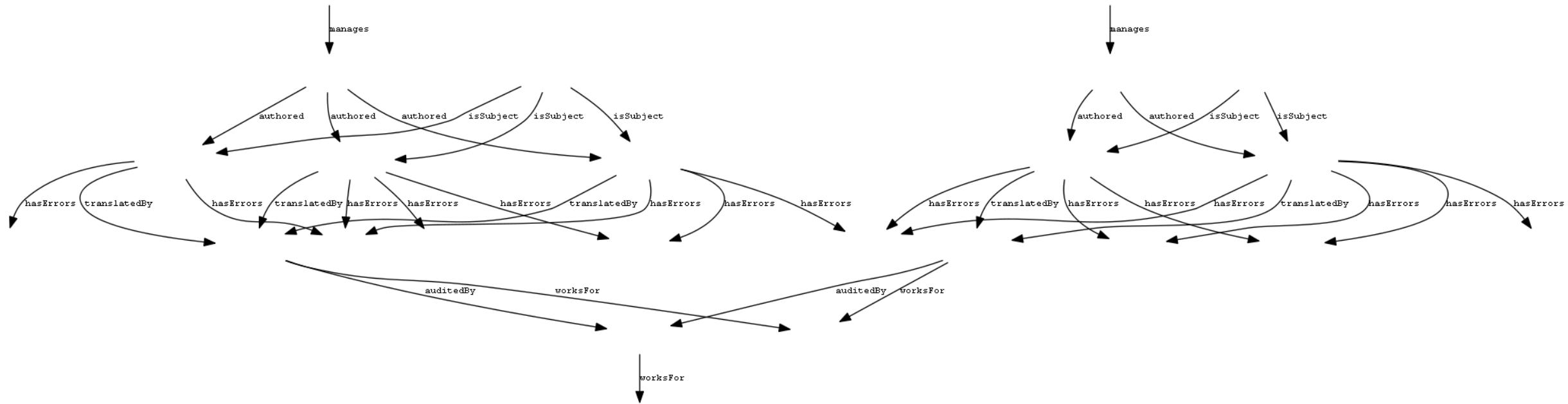




Learning

- Need to aggregate for clarity
- Graphs become complex quickly





Ontologies

Focus

- Provenance
- NLP
- Modelling
- Semantics
- Communities

ITS 2.0, Ocelot and NIF

@prefix...

```
<http://example.com/myitsservice?informat=html?intype=url&input=http://example.com/doc.html&char=0,179>
  a nif:Context ;
  a nif:RFC5147String ;
  nif:beginIndex "0" ;
  nif:endIndex "179" ;
  nif:isString " transport or transportation is the movement of people, animals and goods from one
location to another. Modes of tranport inc. air, rail, road, water, cable, pipeline, and space. " .
```



```
<http://example.com/myitsservice?informat=html?intype=url&input=http://example.com/doc.html&char=1,10>
  nif:beginIndex "1" ;
  nif:endIndex "10" ;
  a nif:RFC5147String ;
  itsrdf:hasLocQualityIssue [
    a itsrdf:LocQualityIssue ;
    itsrdf:locQualityIssueEnabled "no" ;
    itsrdf:locQualityIssueProfileRef <http://example.org/qaMovel/v1> ;
    itsrdf:locQualityIssueSeverity "50" ;
    itsrdf:locQualityIssueType "typographical" ;
  ];
```

- Exchange mechanism with NLP tools
 - Synergy with our interests and development of predictive text analysis and classification tools

CNGL and GLOBIC

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns>.
@prefix its: <http://www.cngl.ie/ontologies/#>.
@prefix lang: <http://findSomeIsoUri/>.
@prefix ngldatatypes: <http://www.cngl.ie/ontologies/datatypes/>.
@prefix nglcategories: <http://www.cngl.ie/ontologies/tagcontent/>.
@prefix ngltools: <http://www.cngl.ie/ontologies/tools/>.
@prefix nglorigins: <http://www.cngl.ie/ontologies/origins/>.
@prefix nglphases: <http://www.cngl.ie/ontologies/phases/>.
@prefix nglalttranstype: <http://www.cngl.ie/ontologies/alttranstype/>.
@prefix nglCountedContent: <http://www.cngl.ie/ontologies/#CountedContent>.
@prefix gic: <http://www.cngl.ie/ontologies/#>.
@prefix jobs: <http://www.cngl.ie/jobs/>.
@prefix logs: <http://www.cngl.ie/logger/logs/>.
@prefix prv: <http://replace-with-provenance/>.
@prefix prov:<http://www.w3.org/ns/prov#>.
@prefix dcterms:<http://purl.org/dc/terms/>.
@prefix str: <http://nlp2rdf.lod2.eu/schema/string/> .
@prefix its: <http://www.w3.org/2005/11/its/rdf#>.

<http://www.cngl.ie/jobs/> a gic:Project;
<http://www.cngl.ie/jobs/> a prov:Activity;
    gic:fileFormat "XLIFF 1.2";
    gic:activityId "";
```

CNGL and GLOBIC

```
<http://www.cngl.ie/jobs//fr-fr> a prov:Entity;
  gic:xliffSourceLang lang:en-us;
  gic:xliffTargetLang lang:fr-fr;
  rdf:XMLLiteral ngldatatypes:;;
  prov:wasMember <http://www.cngl.ie/jobs//fr-fr#collection>;
.

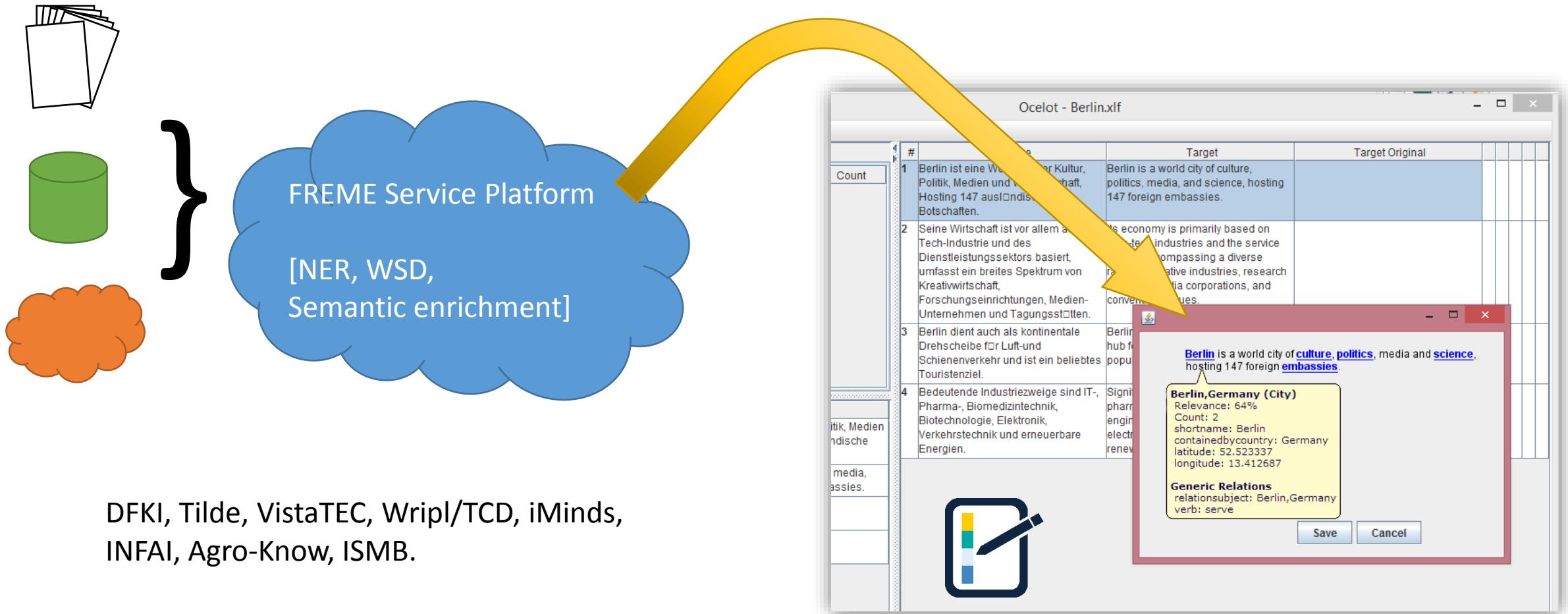
<http://www.cngl.ie/jobs//fr-fr> a gic:InformationResource;.

<http://www.cngl.ie/jobs//fr-fr/0> a gic:Segment;.
<http://www.cngl.ie/jobs//fr-fr/0> a prov:Entity;
  gic:sourceSegment "Microsoft Office 2010, as revealed by the just-released Technical Preview, brings a
set of important if incremental improvements to the market-leading office suite.";
  gic:sourceLang lang:en-us;
  gic:targetSegment "Microsoft Office 2010, tel que révélé par la Technical Preview
vient de paraître, apporte un ensemble d'améliorations importantes si supplémentaires à la suite office de
leader sur le marché.";
  gic:targetLang lang:fr-fr;
  its:hasProvRecord <http://www.cngl.ie/logger/logs/prov0>;.
```

Community Engagement

- Best Practices for Multilingual Linked Open Data
 - <http://www.w3.org/community/bpmlod/>
- Linked Data for Language Technologies (LD4LT)
 - <http://www.w3.org/community/ld4lt/>
- LIDER
 - <http://www.lider-project.eu/>

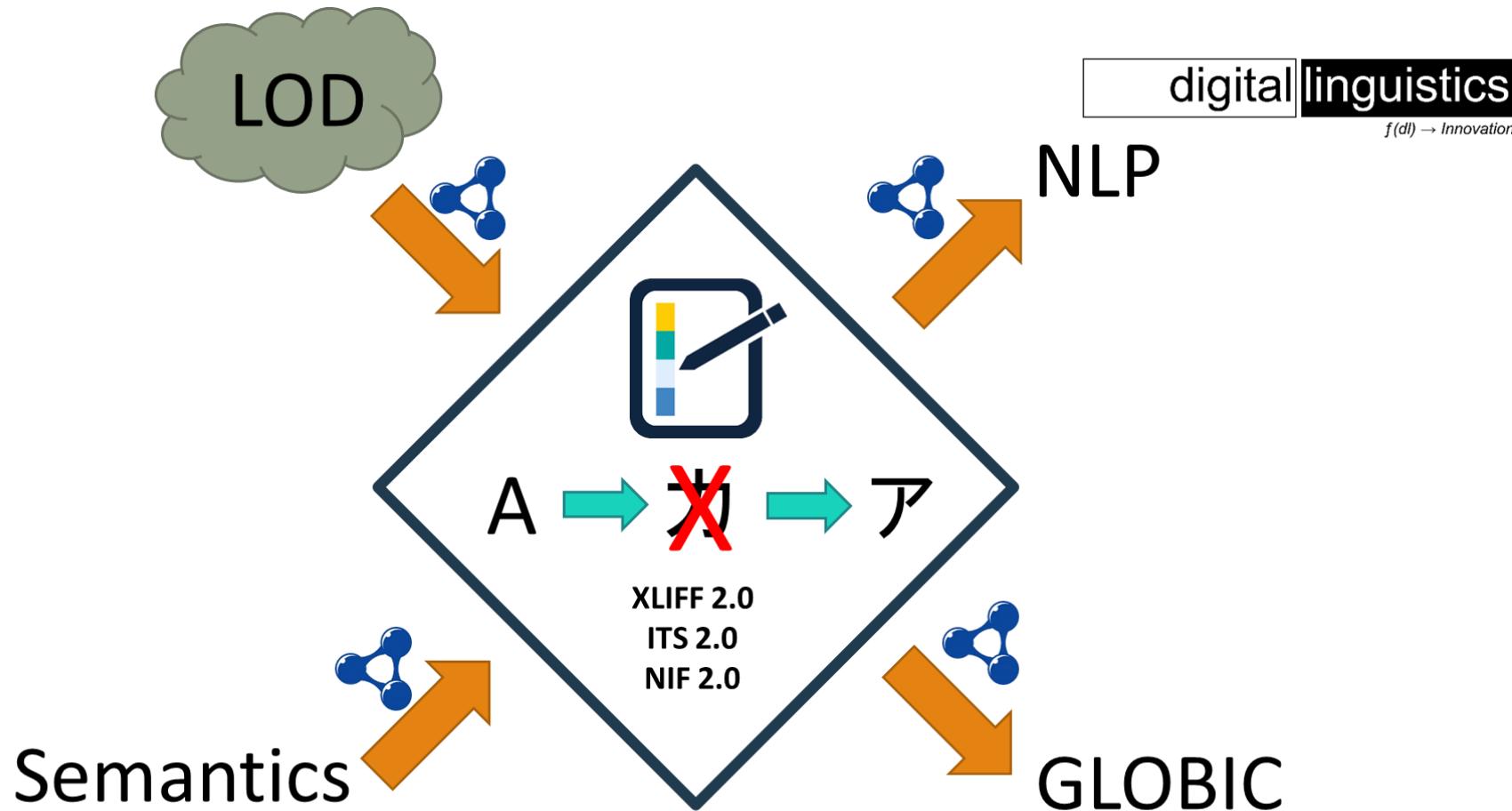
FREME



Conclusions

- Facilitates concept of Intelligent Content
 - Semantic models and enrichment
- Socialises information
 - Potential value-add to customers?
- Up-skills linguists
 - Linguists can judge automatic suggestions and curate
- Deep linguistic structure and process representations
 - Reason about language and meaning
 - Allow automation to move to the next level

Vision



Thank You!

phil.ritchie@vistatec.com

@philinthecloud

blog.phil-ritchie.net