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

FREME WEBINAR

SLIDES CREATED FEBRUARY 2016

FREME
OPEN FRAMEWORK OF E-SERVICES FOR MULTILINGUAL AND SEMANTIC ENRICHMENT OF DIGITAL CONTENT

www.freme-project.eu

Presented
on behalf of the FREME Consortium

Material

- Download of slides

https://www.w3.org/community/ld4lt/wiki/File:Freme-webinar-20160222-slides.pdf

- Download of examples

https://www.w3.org/community/ld4lt/wiki/File:Freme-webinar-20160222-examples.zip

OVERVIEW

- Part I: Outline of FREME
- Part II: Technical aspects of the framework
- Part III: Applications built on top of FREME
- Q&A

FREME

BIRD'S EYE VIEW ON FREME

THE FREME PROJECT

- Two year H2020 Innovation action; started February 2015
- Industry partners leading four business cases around digital content and (linked) data
- Technology development bridging language and data
- Outreach and business modelling demonstrating monetization of the multilingual data value chain

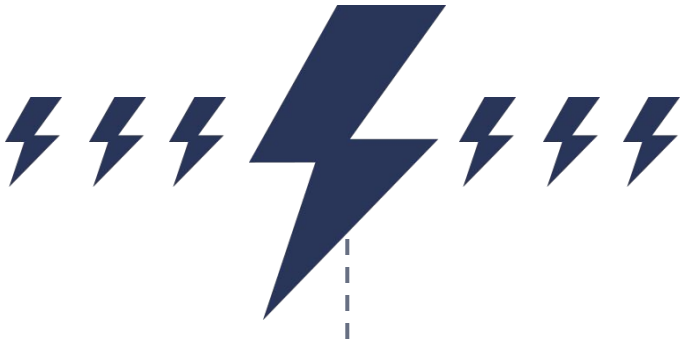
CHALLENGE AND OPPORTUNITY: BIG DATA IS GROWING ACROSS LANGUAGES, SECTORS AND DOMAINS

WHAT LIES AHEAD FOR SEVERAL INDUSTRIES? SEE THE FREME BUSINESS CASES

- BC: Digital publishing
- BC: Translation and localisation
- BC: Agriculture and food domain data
- BC: Web site personalisation

CURRENT STATE OF SOLUTIONS

Machine translation, terminology annotation, ...



Linked data creation & processing

GAPS THAT HINDER BUSINESS:

- Plethora of formats
- Adaptability and platform dependency
- Language coverage
- Usability “The right tool for the right person in given and new enterprises”: technology influences job profiles

FREME TO THE RESCUE: ENRICHING DIGITAL CONTENT

Machine translation, terminology annotation, ...

Linked data creation & processing

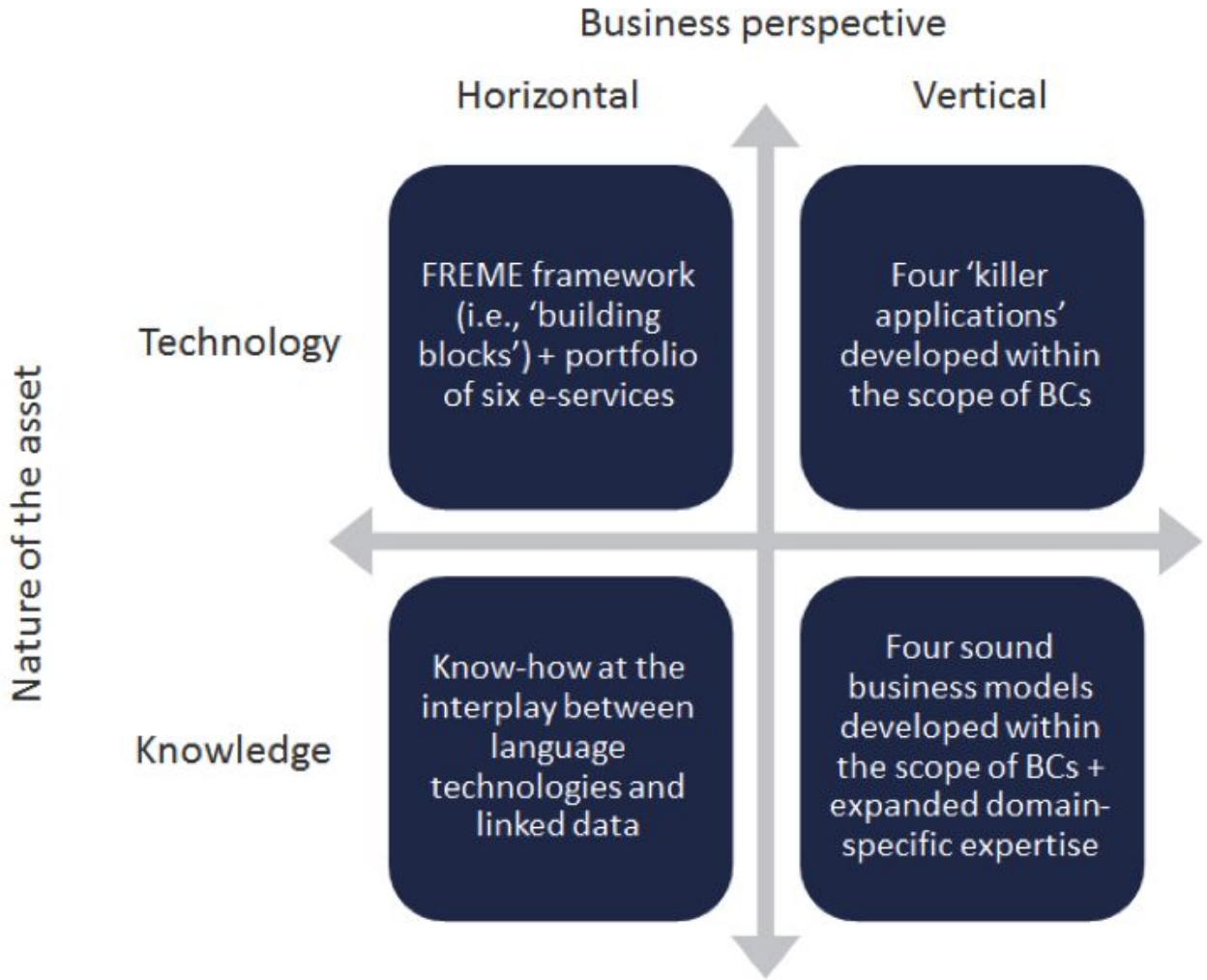
A SET OF INTERFACES* - DESIGN DRIVEN BY BUSINESS CASES

* Graphical interfaces
* Software Interfaces

LT and LD as first class citizens on the Web

LT and LD for various user types: (application) developer, content architect, content author, …

ASSETS AND BUSINESS PERSPECTIVE IN THE FREME PROJECT



OVERVIEW

- Part I: Outline of FREME
- Part II: Technical aspects of the framework
- Part III: Applications built on top of FREME
- Q&A

FREME FROM A TECHNICAL PERSPECTIVE

A framework for multilingual and semantic enrichment of digital content that provides access via a set of APIs and GUIs to six E-services.

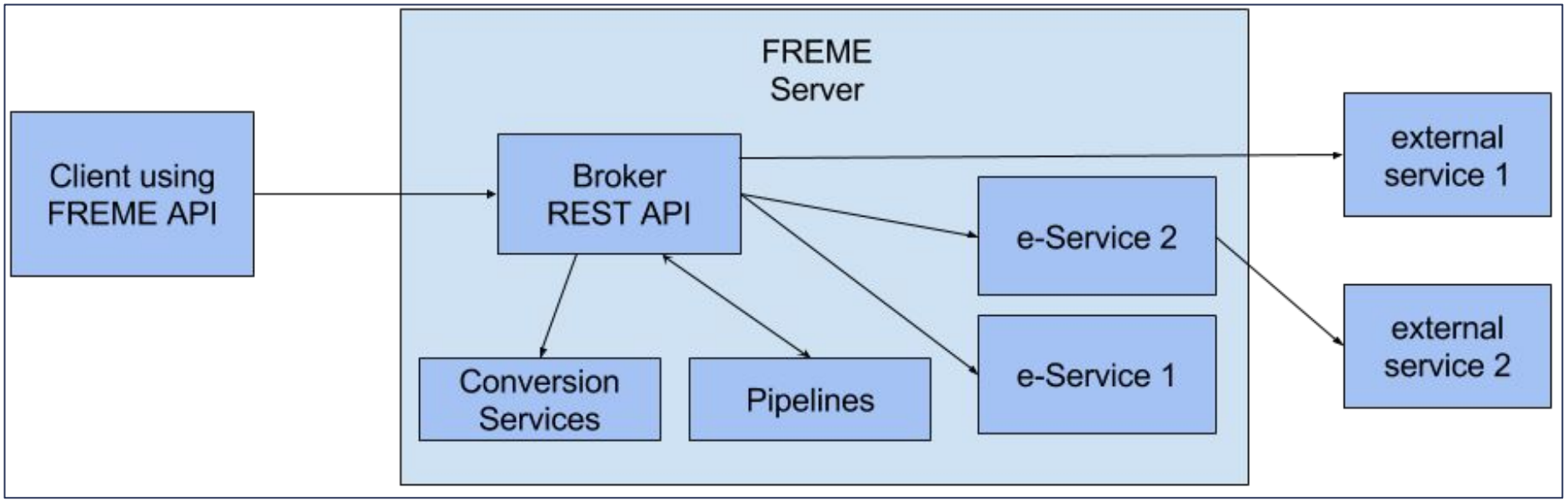
- **e-Entity** for enriching content with information on named entities;
- **e-Link** for enrichment with linked data sources;
- **e-Terminology** for detecting terms and enriching them with term related information;
- **e-Translation** for providing custom machine translation systems;
- **e-Internationalisation** for processing a variety of digital content formats; and
- **e-Publishing** for exporting the outcome of enrichment processes in the ePub format.

FREME FROM A TECHNICAL PERSPECTIVE

How to access FREME – several options:

- A life version 0.5 including documentation at http://api.freme-project.eu/doc/0.5/
- A development version at http://api-dev.freme-project.eu/doc/
- A runnable jar file for all versions at http://api-dev.freme-project.eu/freme-distributions/
- Source code in a GitHub project https://github.com/freme-project/
- The framework is available under Apache 2.0 license to ease commercial use
- Underlying services have various licensing conditions

DESIGN OF THE FRAMEWORK



- Client makes a Web service request.
- The broker evokes the actual e-Service.
- The e-Services are part of the server (e.g. e-Entity), or provided externally (e.g. e-Translation).
- Supportive modules provide conversion of digital content formats or pipelining of services (e.g. e-Terminology followed by e-Translation)

FREME = **a framework, not a platform**: modular approach & ease of extensibility

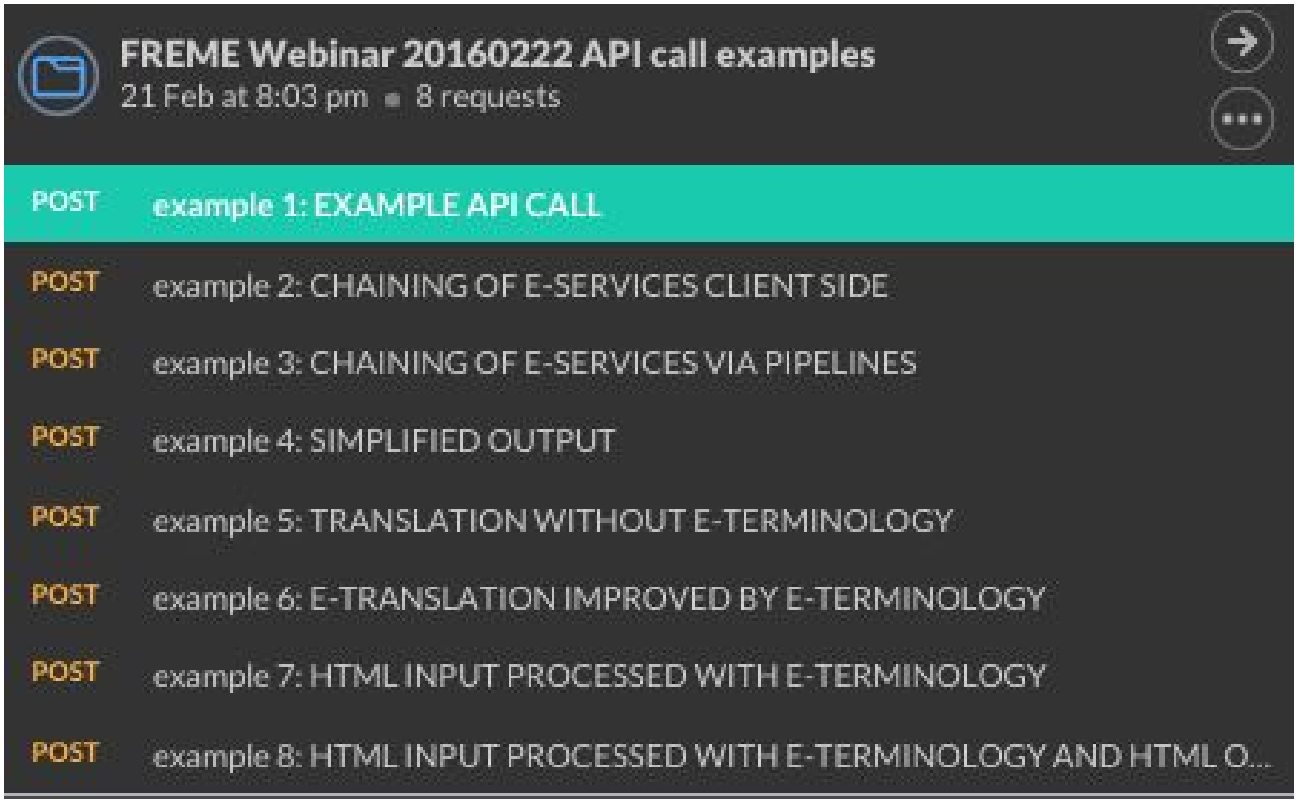
LINGUISTIC LINKED DATA AND OTHER STANDARDS PUT IN ACTION VIA FREME

- **NIF** (Natural Language Processing Interchange Format) for representing digital content and enrichment information in a format agnostic manner, based on the linked data stack;
- **OntoLex lemon** for representing lexical information, to be used e.g. for improving machine translation output;
- **Internationalization Tag Set 2.0** for representing various types of enrichment information in a standardized manner, related e. g. to terminology named entities; and
- The general **linked data technology stack** (RDF, SPARQL etc.)

FREME is built on outcomes of standard driving projects in FP7 in the area of linguistic linked data: LIDER and FALCON

Cf. http://lider-project.eu/ and http://falcon-project.eu/

EXAMPLES



- HTTP requests created with Google Chrome plugin postman
- All postman requests and requests via CURL and jQuery can be downloaded from the LD4LT wiki:

https://www.w3.org/community/ld4lt/wiki/File:Freme-webinar-20160222-examples.zip

FREME

EXAMPLE API CALL

EXAMPLE API CALL

```
http://api.freme-project.eu/current/e-entity/freme-ner/documents/?

        informat=text&outformat=turtle&language=en&dataset=dbpedia&

        input=Welcome to the city of Prague
```

- The request is made to the API for the e-Entity service, a service that enriches content with named entities.
- The input format of content is plain text; the output format is turtle.
- The content to enrich is “Welcome to the city of Prague”.
- The language or the content is English.
- The dataset used for the enrichment is DBpedia.

See example requests: example 1

EXAMPLE OUTPUT: USING NIF TO STORE CONTENT …

```
(1) <http://freme-project.eu/#char=0,29>
(2)  a nif:String , nif:Context , nif:RFC5147String ;
(3)  nif:beginIndex  "0"^^xsd:int ;
(4)  nif:endIndex    "29"^^xsd:int ;
(5)  nif:isString    "Welcome to the city of Prague"^^xsd:string .
```

1) Identifying the content via a URI
2) Adding certain types from NIF*
3) Identifying the start offset of the content
4) Identifying the end offset of the content
5) Providing the string content itself.

* For More on NIF: see a dedicated tutorial http://de.slideshare.net/m1ci/nif-tutorial

… AND ENRICHMENT INFORMATION

(1) <http://freme-project.eu/#char=23,29> …

(2) nif:anchorOf "Prague"^^xsd:string ;

(3) nif:beginIndex "23"^^xsd:int ;

(4) nif:endIndex "29"^^xsd:int ;

(5) nif:referenceContext <http://freme-project.eu/#char=0,29> ;

(6) **itsrdf:taClassRef <http://dbpedia.org/ontology/City>.**

1) Identifying the annotation via a URI
2) Providing the string content of the annotation
3) Identifying the start offset of the content
4) Identifying the end offset of the content
5) Relating the content to annotations
6) Enrichment with ITS 2.0 class information (“Prague” = a city)

FREME

CHAINING OF SERVICES, SIMPLIFICATION OF OUTPUT AND FORMAT COVERAGE

CHAINING OF E-SERVICES

Example: combining e-Entity and e-Link

1) e-Entity input: “Welcome to the city of Prague.”
2) Input to e-Link – goal: find geo position from DBpedia data source

Chaining can be done:

3) On the client side
4) By calling the FREME pipelining service

For more infos on pipelines see

http://api.freme-project.eu/doc/0.5/tutorials/pipeline-entity-link.html

5) The outcome is again NIF, see next slide

See example requests: example 2 and example 3

NIF OUTPUT EXAMPLE (PARTIAL)* FOR CHAINING SERVICES

<http://freme-project.eu/#char=0,30>

nif:anchorOf "Welcome to the city of Prague.".

<http://freme-project.eu/#char=23,29>

nif:anchorOf "Prague".

<http://dbpedia.org/resource/Prague>

<http://www.w3.org/2003/01/geo/wgs84_pos#lat>

"50.0880428938909”;

<http://www.w3.org/2003/01/geo/wgs84_pos#long>

"14.4207572937012" .

*The original output has 110 lines … **linked data can be complex!**

SIMPLIFIED OUTPUT HELPS API DEVELOPERS TO CONSUME LINKED DATA

- FREME provides user specified filter mechanism to simply the output
- Supports CSV, XML or JSON
- Example for previous query as CSV

http://dbpedia.org/resource/Prague, 50.0878367932108,14.4241322001241

For more infos on filtering, see

http://api.freme-project.eu/doc/0.5/knowledge-base/filtering.html

See example requests: example 4

CHAINING OF E-SERVICES TO IMPROVE RESULTS

- e-Services can be improved by taking up the results of previous services
- Example with simplified output: e-Translation service "understands" the output of e-terminology (providing translation suggestions)

e-Translation only

source,target
The EU in brief. The EU is a unique economic and political partnership between 28 European countries that together cover much of the continent.,In de EU. De EU is een uniek economische en politiek partnerschap van 28 Europese landen samen op dat deel van het continent.

e-Translation followed by e-Terminology

source,target
The EU in brief. The EU is a unique economic and political partnership between 28 European countries that together cover much of the continent.,"De voorschriften in DE EU. De EU is een uniek partnerschap tussen politiek en economie in de Europese landen, die gezamenlijk 28 verpakking van het continent.

See example requests: example 5 (e-Translation only) and example 6 (chain of e-Translation and e-Terminology)

FORMAT COVERAGE

- Processing of various content formats
  - NIF, RDF, Text, HTML, OpenOffice, XLIFF 1.2, …
- Many formats are processed via e-Internationalization services
- Format specified in API call as input and (partially supported) output
- More information: http://api.freme-project.eu/doc/0.5/knowledge-base/eInternationalization.html

See example requests: example 7

ROUNDTRIPPING EXAMPLE WITH HTML and E-TERMINOLOGY

```
<!DOCTYPE html> …
<body>
<p>Welcome to the city of Prague.</p>
</body> … </html>
```

Call of e-Terminology

```
<!DOCTYPE html> …
<p>Welcome to the <span its-term="yes">city</span> of Prague.
…</html>
```

See example requests: example 8

COVERING OF XML CONTENT FORMATS

- Currently only as a client-side demo
- Covers DocBook and TEI formats

See

http://api-dev.freme-project.eu/doc/freme-showcase/xml-to-rdf.html

CHAINING OF SERVICES BEYOND FREME - POSSIBLE VIA STANDARDISED ENRICHMENT INFORMATION

- Upcoming: services provided by the project “Digitale Kuratierungstechnologien” http://digitale-kuratierung.de/
    - The services also deploy NIF and ITS 2.0 for storing digital content and enrichment information

The following pipeline will then be possible for you calling the services

1) Call http://digitale-kuratierung.de/ based service to do named entity recognition - output is NIF + ITS 2.0
2) Feed the output to e-Link to gather further information

> Distributed & decentralized language and data technologies, made possible via standardized workflows!

OVERVIEW

- Part I: Outline of FREME
- Part II: Technical aspects of the framework
- Part III: Applications built on top of FREME
- Q&A

OVERVIEW – GUIS BUILT BY FREME BC PARTNERS

GUI / BC	Description	E-Services
eBook authoring environment / BC1	Authoring of eBook content	e-Entity, e-Link, e-Terminology, e-Translation, e-Internationalisation
CKEditor plugin / BC1	FREME enabled browser CMS plugin (e.g. for WordPress, Drupal)	e-Entity, e-Link, e-Translation, e-Internationalisation
Ocelot editor / BC2	Vistatec GUI for quality assurance in localisation processes	e-Entity, e-Link, e-Terminology, e-Translation, e-Internationalisation
Browser interfaces	Interactive access via graphical interfaces in the browser	All e-Services

BUSINESS CASE "LINKED DATA IN PUBLISHING WORKFLOWS"

- Wolters Kluwer, Agroknow
- Enrichment of academic publication metadata

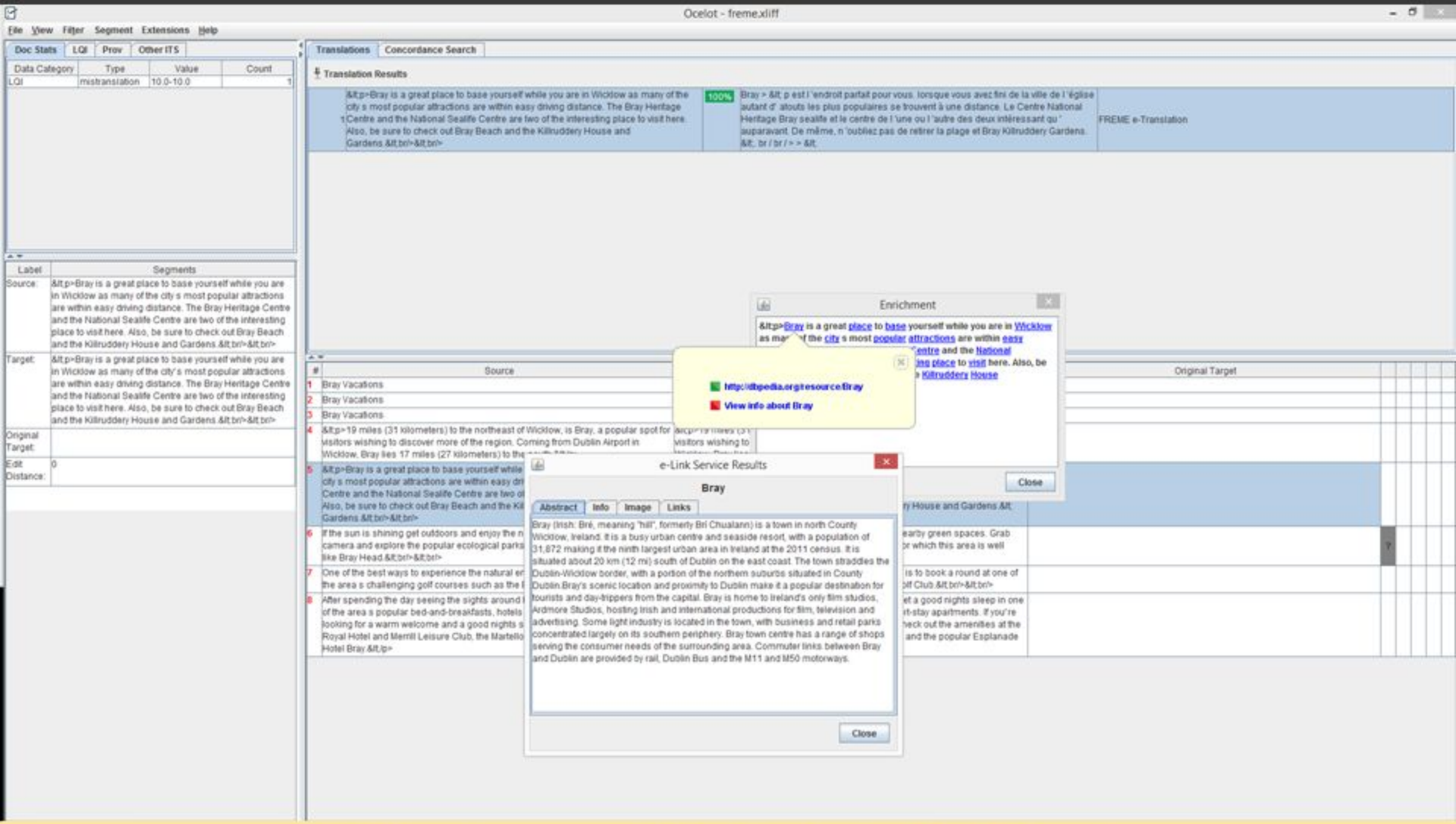
Before FREME	Result of deploying FREME
<dc:creator> <ags:creatorPersonal> Stoitsis, Giannis, Agroknow </ags:creatorPersonal> </dc:creator>	<dc:creator> <ags:creatorPersonal>Stoitsis, Giannis</ags:creatorPersonal> **<nameIdentifier schemeURI=** **"http://orcid.org/"** **nameIdentifierScheme=** **"ORCID">0000-0003-3347-8265** **</nameIdentifier>** **<affiliation>Agroknow</affiliation>** </dc:creator>
<dc:subject> <ags:subjectClassification scheme="ags:ASC"> <![CDATA[J10]]> </ags:subjectClassification> </dc:subject>	<dc:subject **freme-enrichment=** **"http://aims.fao.org/aos/agrovoc/c_426** **http://aims.fao.org/aos/agrovoc/c_24135** **http://aims.fao.org/aos/agrovoc/c_4644** **http://aims.fao.org/aos/agrovoc/c_7178">** <ags:subjectClassification scheme= "ags:ASC"><![CDATA[J10]]> </ags:subjectClassification> </dc:subject>

BUSINESS CASE
"LINKED DATA IN XML LOCALIZATION WORKFLOWS"

- Vistatec – workflows integrating localization XML formats XLIFF, ITS 2.0 and linked data, in the Ocelot editor for translation editing and review – see GUI screenshot next slide

Process Step	FREME e-service
Conversion of native document to Extensible Localization Interchange File Format	e-Internationalization
Translation	e-Terminology and e-Entity
Semantic enrichment	e-Link
Content publication	e-Pub

vistatec



SPARQL queries are executed to retrieve desired related information

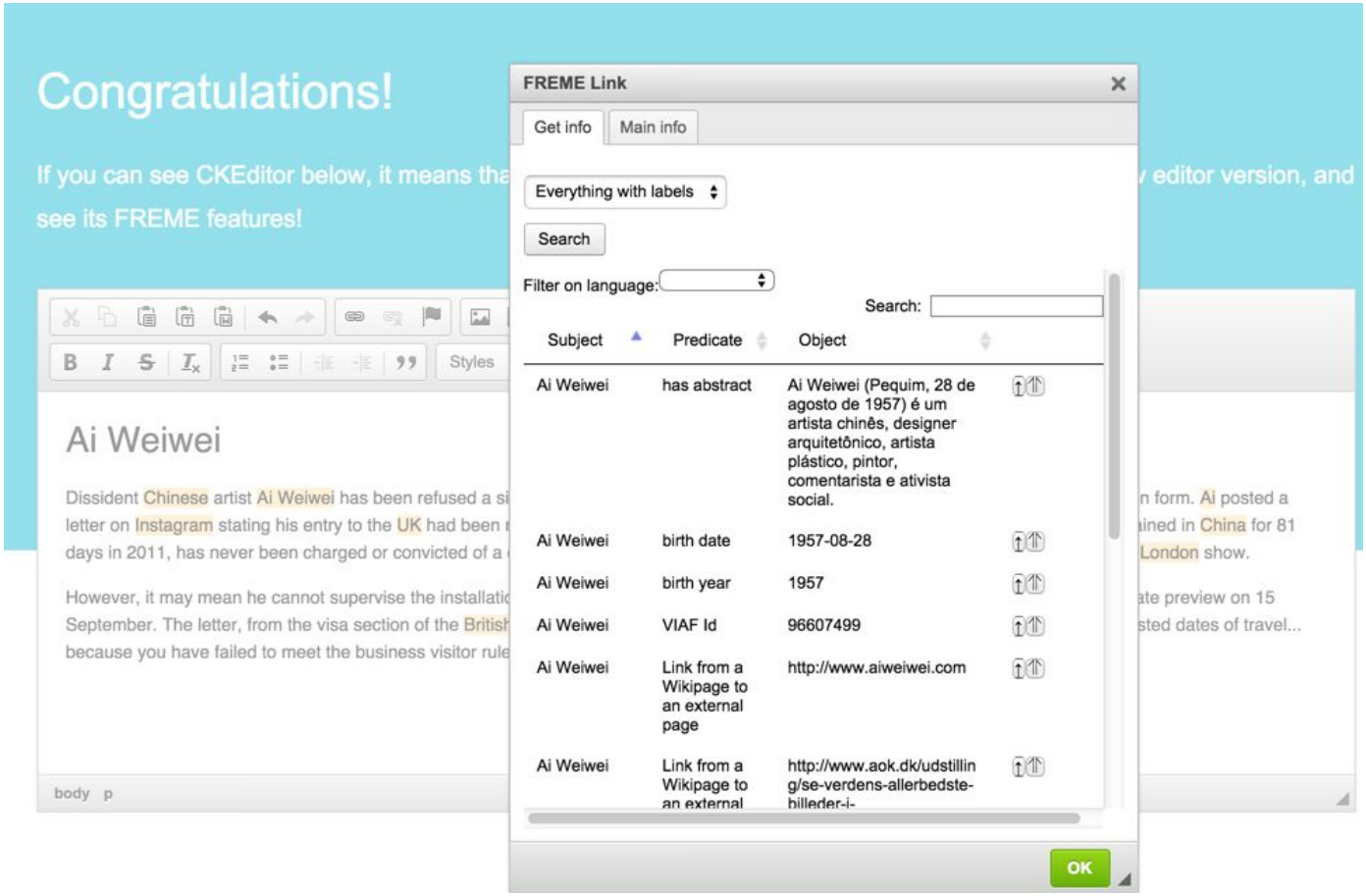
BUSINESS CASE "LINKED DATA IN BOOK METADATA"

- iMinds – linked data in book metadata
- A potential approach for embedding linked data in ONIX

```
<Contributor>
 <NameIdentifier>
  <NameIDType>
   <IDTypeName>Meta4Books ContributorID</IDT
   <IDValue>65097</IDValue>
  </NameIDType>
 </NameIdentifier>
 <ContributorRole>A01</ContributorRole>
 <SequenceNumber>1</SequenceNumber>
 <NamesBeforeKey>Jonathan</NamesBeforeKey>
 <KeyNames>Franzen</KeyNames>
 <Entity>
  <URI>http://viaf.org/viaf/84489381/</URI>
 </Entity>
</Contributor>
```

INTEGRATING FREME IN GUI

- FREME plugin for open source WYSIWYG HTML editor CKEditor
- Semantic enrichment & machine translation service

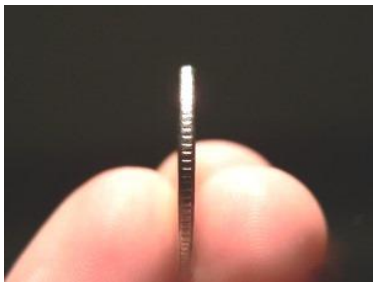


BUSINESS CASE "Empowering user engagement via Personalisation and Analytics"

- wripl – content recommendations and analytics
- Increase the value of your online content by addressing both sides of the coin:

Personalisation (FrontEnd) & **Analytics** (BackEnd)

Recommend the right content to the right user segment

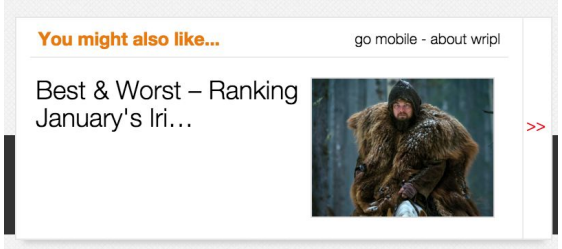


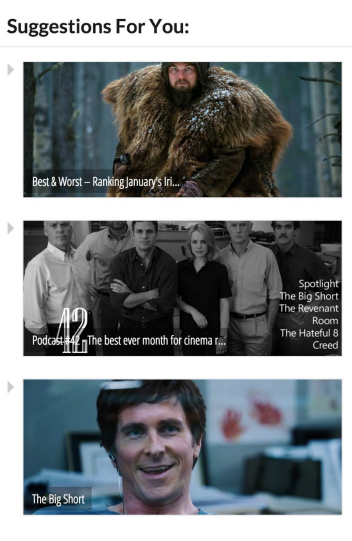
Create the right content based on user interest engagement analytics

Increase Content Value

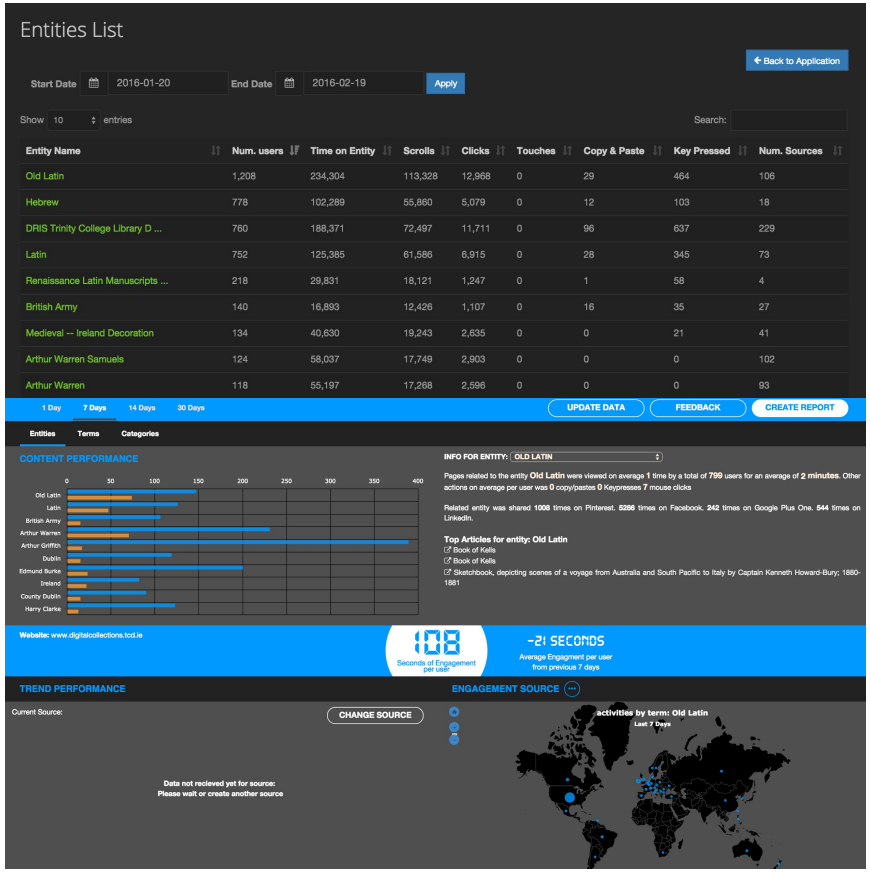
BUSINESS CASE “Empowering user engagement via Personalisation and Analytics”

Personalisation





Interest Analytics



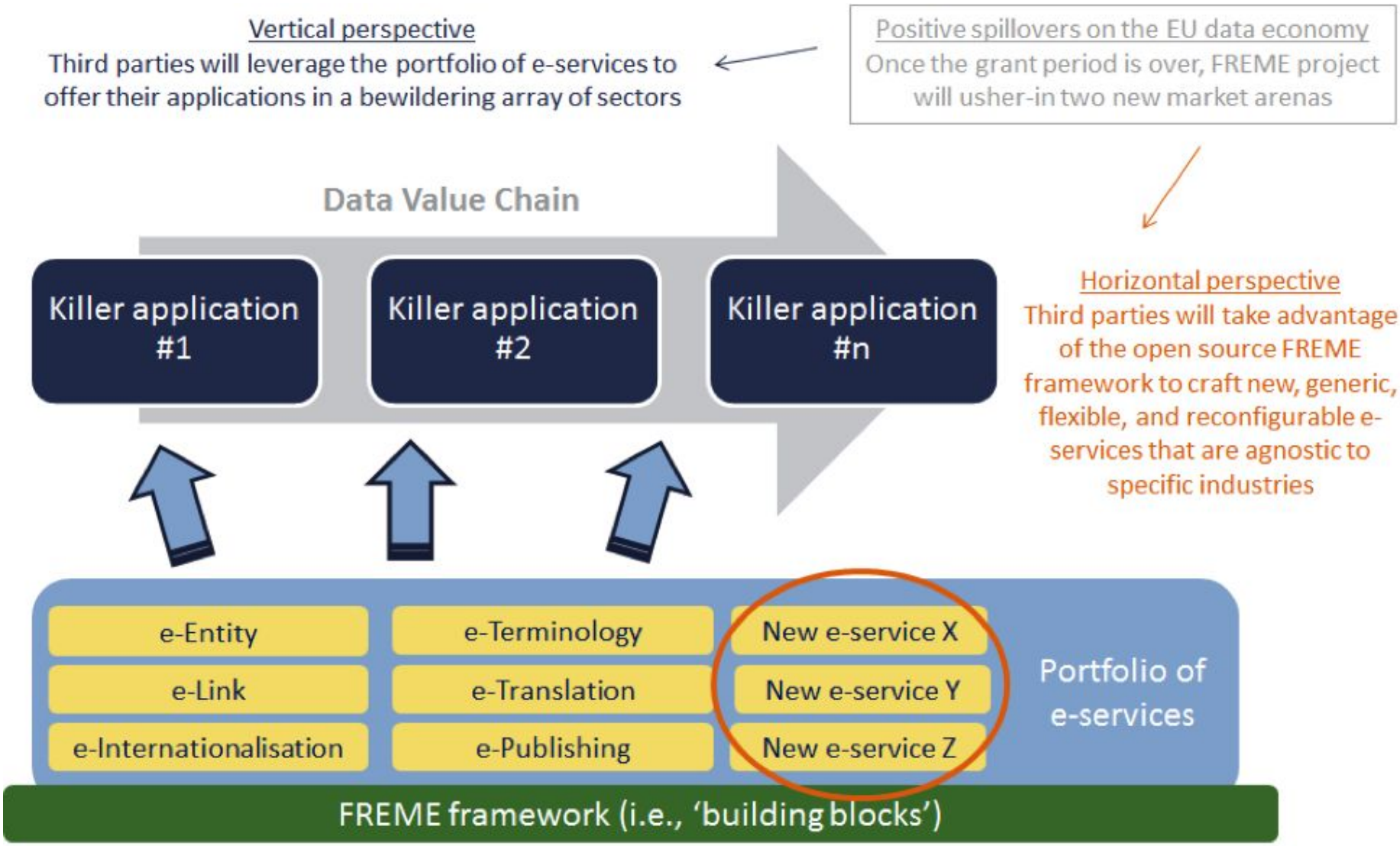
OVERVIEW

- Part I: Outline of FREME
- Part II: Technical aspects of the framework
- Part III: Applications built on top of FREME
- Q&A

WE ARE LOOKING FOR FEEDBACK! ON:

- Technical design of the framework
  - Your analysis of shortcomings and strong points
- Characteristics of existing e-Services
- Relation to other frameworks and platform – let's build synergies!
- Applications you want to build
- New services you want to have – or contribute to the framework - See our vision in the next slide

A VISION FOR THE FUTURE OF THE FREME FRAMEWORK



KEY PEOPLE AROUND THE FRAMEWORK

FREME

FREME-PROJECT.EU

Felix Sasaki
Tatjana Gornostaja
felix.sasaki@dfki.de
tatjana.gornostaja@tilde.lv

DOMAIN AND TECHNOLOGY SPECIFIC

Business case “Publishing” and e-Publishing
Frank Salliau frank.salliau@ugent.be

Business case “Localisation” and e-Internationalisation
Phil Ritchie phil.ritchie@vistatec.com

Business case “Agriculture”
Giannis Stoitsis stoitsis@agroknow.com

Business case “Web site personalisation”
Kevin Koidl kevin@wripl.com

Language technologies *(e-Translation, e-Terminology)*
Tatjana Gornostaja tatjana.gornostaja@Tilde.lv

Data technologies and data sets
Milan Dojchinovski (e-Entity, e-Link)
milan.dojchinovski@fit.cvut.cz

Business opportunities
Michele Osella osella@ismb.it

FRAMEWORK GENERAL

Jan Nehring <jan.nehring@dfki.de>

DKT Project

Georg Rehm <georg.rehm@dfki.de>

STRATEGIC DEVELOPMENTS DIRECTIONS

More business cases

More technologies

More (multilingual) data set integration

CONTACTS

E-mail: info@freme-project.eu

CONSORTIUM