The challenges of multilinguality

Joint EURBICA and APEx meeting

25.11.2013, Brussels

Susanne Waidmann, APEx WP4 Standards & Guidelines

Outline

- Our current users – what are their languages?

- What do we already offer?

- How to manage multilinguality

  – Translation tool

  – Using our standards

  – Semantic assets

- Possibilities for APEx/F
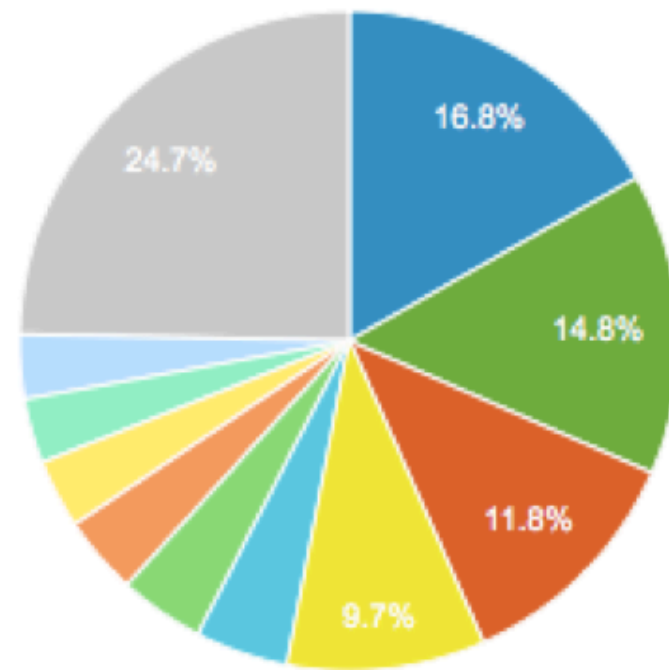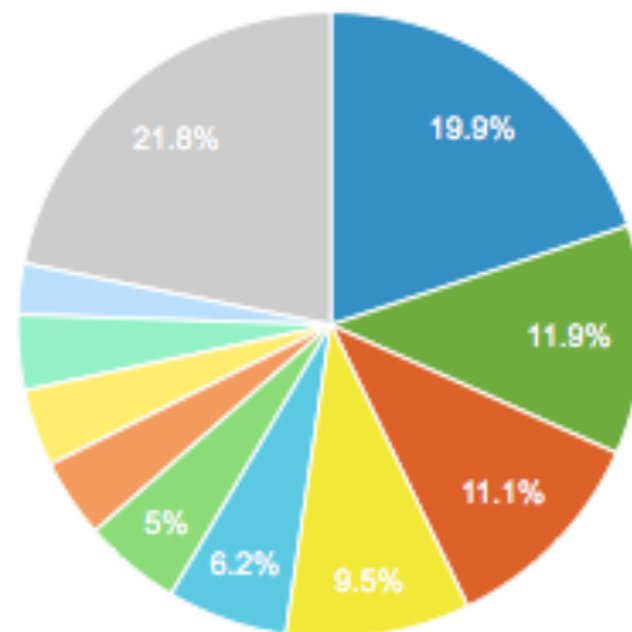
Our current users –
what are their languages?

1.		Netherlands	16.79%
2.		Germany	14.76%
3.		France	11.79%
4.		Spain	9.68%
5.		Poland	4.56%
6.		United States	4.17%
7.		Belgium	3.86%
8.		Italy	3.38%
9.		Sweden	3.17%
10.		United Kingdom	3.15%
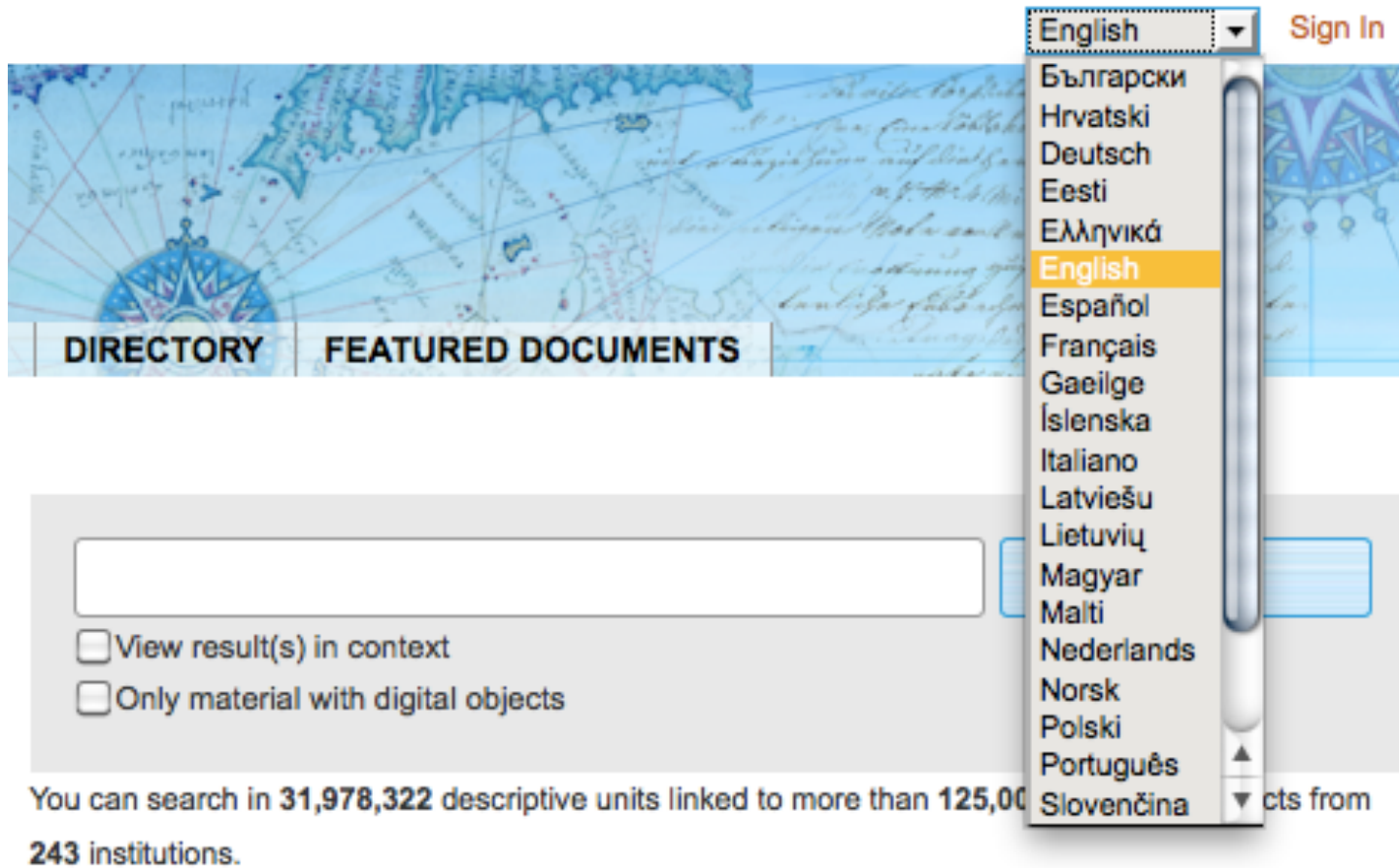
1.	■	en-us	19.87%
2.	■	fr	11.85%
3.	■	nl	11.08%
4.	■	de-de	9.49%
5.	■	es	6.22%
6.	■	de	5.05%
7.	■	nl-nl	4.08%
8.	■	pl	3.99%
9.	■	es-es	3.89%
10.	■	fr-fr	2.63%
11.	■	en	2.07%
12.	■	it	1.72%
13.	■	en-gb	1.67%
14.	■	sv-se	1.52%
15.	■	pt-pt	1.42%
16.	■	it-it	1.39%

What do we already offer?

- **multilingual user surfaces**
  - select the language in which you want to see your information

- **multilingual descriptions**
  - first steps within the Archival landscape and archival descriptions
  - constituent examples of multilingual descriptions of several partners (for example Belgium)
  - fitting the "profiles" for multilingualism → xml@lang for all relevant free text fields → fit the tools to show this correctly

- **multilingualism within the search**
  - autocompletion and alternative suggestions are shown in different languages but without translation functionalities behind it (only using search index entries in alphabetical order close to user's search entry

Multilingual surfaces

What do we already offer?

- multilingual user surfaces
  - select the language in which you want to see your information

- multilingual descriptions
  - first steps within the Archival landscape and archival descriptions
  - constituent examples of multiingual descriptions of several partners (for example Belgium)
  - fitting the "profiles" for multilingualism → xml@lang for all relevant free text fields → fit the tools to show this correctly

- multilingualism within the search
  - autocompletion and alternative suggestions are shown in different languages but without translation functionalities behind it (only using search index entries in alphabetical order close to user's search entry)

Multilinguality within the search
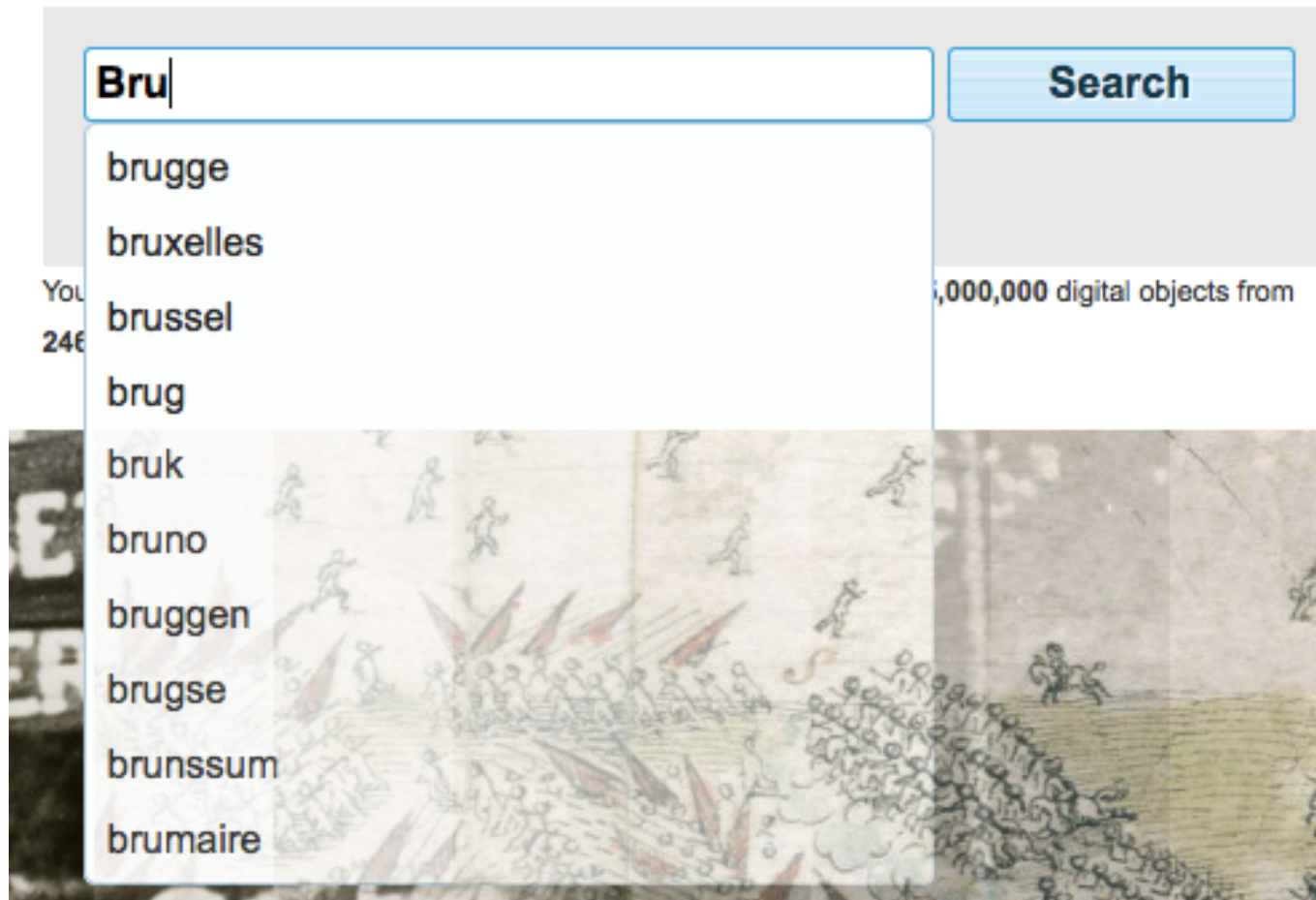
Search

| bruxelles brussel brüssel brus | **Search** |

☐ View result(s) in context

☑ Match at least one word

☐ Only material with digital objects

brussel

brusselse

brustem

brussels

bruselas

brus

brusselsestraat

brusse

brust

bruschy

Clear search

ICTPSP

How to manage multilinguality

The translation tool

Using our standards – EAG2012

Bundesarchiv

▷ **Bundesarchiv Koblenz (Headquarter)**

▷ **Bundesarchiv Berlin-Lichterfelde (Branch)**

▷ **Bundesarchiv Bayreuth (Lastenausgleichsarchiv) (Branch)**

▷ **Bundesarchiv Berlin-Wilmersdorf (Filmarchiv) (Branch)**

▷ **Bundesarchiv Freiburg im Breisgau (Militärarchiv) (Branch)**

▷ **Bundesarchiv Hoppegarten (Zwischenarchiv) (Interim archive)**

▷ **Bundesarchiv Ludwigsburg (Branch)**

▷ **Bundesarchiv Rastatt (Branch)**

▷ **Bundesarchiv Sankt Augustin-Hangelar (Zwischenarchiv) (Interim archive)**

OTHER INFORMATION (SEE MORE)

Related resource:	Beständeübersicht des Bundesarchivs Edition "Die Kabinettsprotokolle der Bundesregierung" Edition "Akten der Reichskanzlei. Weimarer Republik" Portal Zwangsarbeit im NS-Staat
Type of Archive:	National archives Specialised governmental archives
Last update:	24.05.2013

Bundesarchiv

- Bundesarchiv Koblenz (Headquarter)
- Bundesarchiv Berlin-Lichterfelde (Branch)
- Bundesarchiv Bayreuth (Lastenausgleichsarchiv) (Branch)
- Bundesarchiv Berlin-Wilmersdorf (Filmarchiv) (Branch)
- Bundesarchiv Freiburg im Breisgau (Militärarchiv) (Branch)
- Bundesarchiv Hoppegarten (Zwischenarchiv) (Interim archive)
- Bundesarchiv Ludwigsburg (Branch)
- Bundesarchiv Rastatt (Branch)
- Bundesarchiv Sankt Augustin-Hangelar (Zwischenarchiv) (Interim archive)

OTHER INFORMATION (SEE MORE)	
Related resource:	Beständeübersicht des Bundesarchivs Edition "Die Kabinettsprotokolle der Bundesregierung" Edition "Akten der Reichskanzlei. Weimarer Republik" Portal Zwangsarbeit im NS-Staat
Type of Archive:	National archives Specialised governmental archives
Last update:	24.05.2013

Information in blue bloxes is shown automatically in the user selected language with help of the translation tool (resp. the Country Managers!): labels, headers, fixed vocubulary, automatically created content (i.e. update date)

Information in orange boxes needs to be managed by institution managers, if it is their wish to offer information in various languages.

APEx

ARCHIVES PORTAL EUROPE

ICTPSP

| Your institution | Identity | Contact | **Access and Services** | Description | Control | Relations |

Opening hours*:

Monday to Friday: 9am to 6pm

Select a language: English ▲▼

Opening hours:

Montag bis Freitag: 9:00 bis 18:00 Uhr

Select a language: German ▲▼

Opening hours:

Lunes a viernes: 9 – 18

Select a language: Spanish ▲▼

Sindhi
Sinhalese
Slovak
Slovenian
Somali
South Ndebele
Southern Sotho
Spanish
Sundanese
Swahili
Swati
Swedish
Tagalog
Tahitian
Tajik
Tamil
Tatar

(Add opening times)

Closing dates:

Select a language:

(Add closing dates)

Description of directions:

Select a language:

APEx

ARCHIVES PORTAL EUROPE

ICTPSP

Using our standards – EAG2012

```
<relations>
 <resourceRelation resourceRelationType="creatorOf" href="http://
   startext.net-build.de:8080/barch/MidosaSEARCH/Bestaendeuebe
   rsicht/index.htm">
   <relationEntry xml:lang="ger">Beständeübersicht des Bundesarch
     ivs</relationEntry>
   <relationEntry xml:lang="eng">Holdings guide of the Federal
     Archives of Germany</relationEntry>
 </resourceRelation>
 [...]
</relations>
```

Semantic assets

Semantic assets contain all kind of controlled vocabularies:

- encoded lists, eg ISO standards

- simple list of words / terms

- dictionary - list of words with definitions

- thesaurus – list of words with relations / synonyms / hierarchies

- ontology – knowledge concept for terms / entities

Semantic assets

→ we will integrate controlled vocabularies in the archival standards we are using

→ we will not create totally new ones, but find solutions to reuse and enhance the existing ones

→ joint WP2, WP4 and WP6 work, which was initiated at a joint work package meeting in Berlin in early November

Possibilities for APE(x/F)

- What we do anyway...
  - use possibilities the standards offer to us
  - investigate further how to use ontologies and controlled vocabularies within the Archives Portal Europe
- Additional possibilities...
  - add possibility to use e.g. Google Translate, but only intentionally (not as default option) → needs to be evaluated
  - language specific searches
  - add a virtual keyboard to permit typing special characters in search box, such as diacritics
  - translations (and/or transcriptions) as user generated content (UGC)

APEx

APEnet ARCHIVES PORTAL EUROPE

| HOME | ABOUT APEX | CONTRIBUTE | NEWS | EVENTS | OUTCOMES | CONTACT |

Welcome to the APEx website

APEx – The project to boost the Archives Portal Europe and let us zoom in on our shared European archival heritage!

The Archives Portal Europe network of excellence - APEx - is a European-based Best Practice Network project aiming at **expanding, enriching, enhancing and sustaining** the Archives Portal Europe.

Spotlight on!

Possibilities for APE(x/F)

- What we do anyway...
  - use possibilities the standards offer to us
  - investigate further how to use ontologies and controlled vocabularies within the Archives Portal Europe
- Additional possibilities...
  - add possibility to use e.g. Google Translate, but only intentionally (not as default option) → needs to be evaluated
  - language specific searches
  - add a virtual keyboard to permit typing special characters in search box, such as diacritics
  - translations (and/or transcriptions) as user generated content (UGC)

Questions and comments?

s.waidmann.apex@bundesarchiv.de

APEx project co-funded by the European Commission within the framework of the ICT Policy Support Programme