Multilingual Legislative Process

M.T. Carrasco Benitez
Luxembourg BabelNet Workshop
2-3 March 2016
Multilingual Legislative Process

A bird’s eye view
Zoom-in: five multilingual experiments
Restate the obvious
Bird’s eye view

EU institutions: making laws
NLP: a mean to an end
Here: address form, not content
Simplistic toddler's view

Daddy makes cars
Daddy makes multilingual publications
Cars : factories
Laws: institutions
Process

Authoring, Translation, Publishing
Partial idealised view
Choreography among EU institutions
Interoperability

Standards
Email: format + protocol
Internet Message Format
Data formats usable as interfaces
Zoom - five experiments

Multilingual Electronic Dossier (MED)
Generation of multilingual parallel docs
Multilingual Toolbox (mux)
Data recycling
Legal Acts Format (LAF)
Multilingual Electronic Dossier (MED)

Container format, multimedia
Shipping container, only the box
Web friendly
Linked data
Container format

Zipped directory structure
Common trick: EPUB, ODF, DOCX
Package or perish
Structure the data, how is secondary
Authoring, TRANSLATION, Publishing

Computer-aided translation (CAT)
Vendor independent translation package
foo.med - MED in OmegaT
Self-contained + server: like email
Generation of multilingual parallel docs

Skeleton + segs 24 langs = 24 docs
Neutral segments, such as numbers
Quality + speed = cost
Camera ready, final format
XML, entities
Multilingual Toolbox (mux)

Aligning and cleaning multilingual data
44 billions: 1952-2016
Large: terabyte, programmatic
No wastage
Clean aligned data: basic ingredient
Flour: bread, pastry
N-languages alignments

Bilingual considered harmful
Translation mostly bilingual
XLIFF, SMT
EU 24 languages
Comparison among all the languages
Good field to look for Ph.D. subjects
Data recycling

Reuse considered during production
Ready for reuse: clean and aligned
Result from generation: docs + segments
BabelNet: hooks during production
More extended process

Preparation
Authoring, Translation, Publishing
Archival (recycled data)
Legal Acts Format (LAF)

Legal acts of the EU
No formal specs, such as a DTD
Prose + tradition
LAF: MED, generation, recycled
Conclusion

Consider the whole process
Impossible design in one go
Specify basic components
Generation of multilingual parallel docs

dragoman.org/lbw