



MT@EC

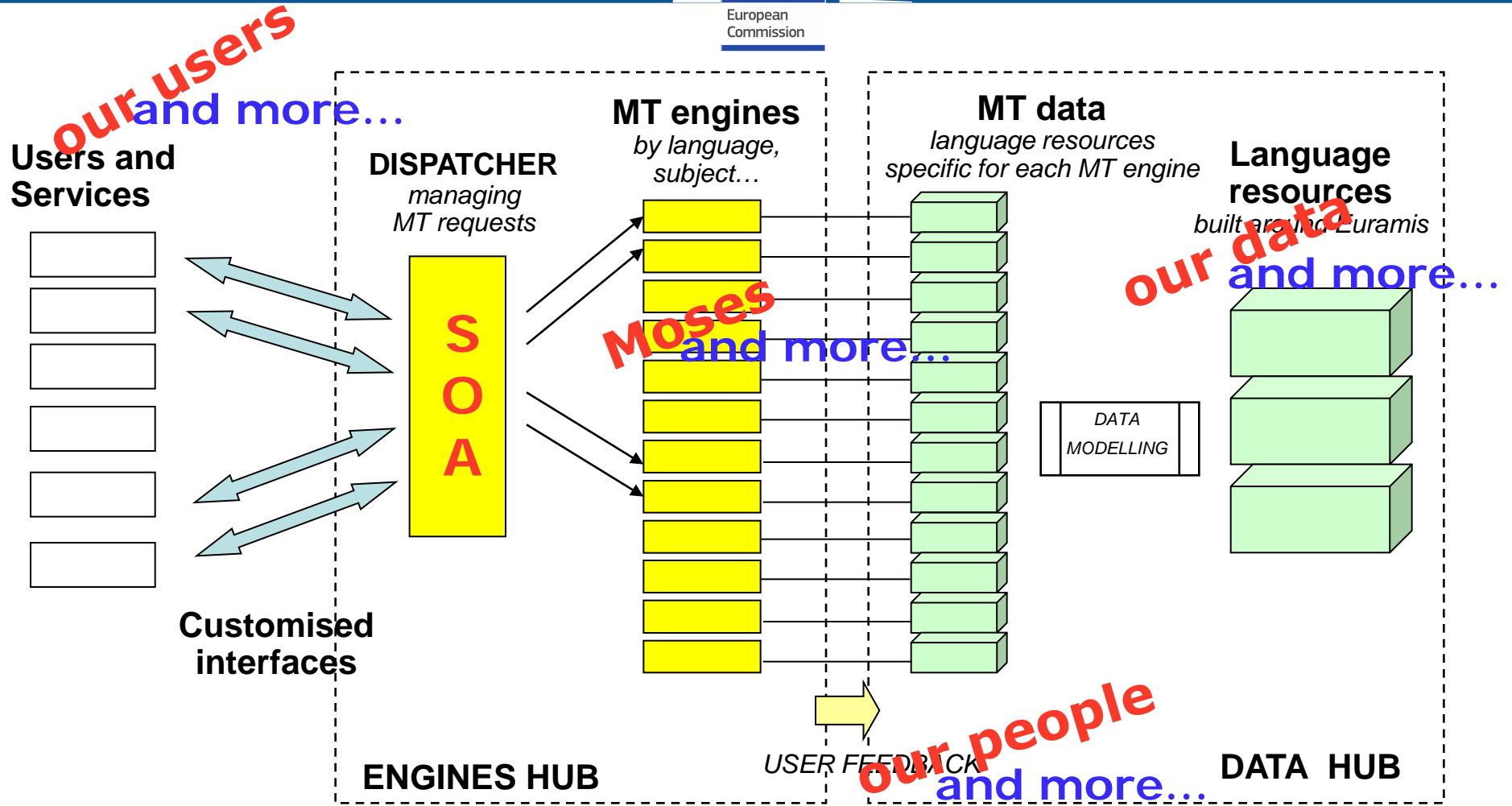
Final Multilingual Web workshop
Luxembourg 15-16 March 2012

Spyridon Pilos
Language applications
DG Translation of the European Commission

A new machine translation service for the European Commission

- Necessary
- Based on data-driven MT technology
- Making best use of Commission language resources
- Making best use of internal linguistic expertise (1700 translators for 23 languages)
- Open and flexible
- Ensuring technological independence
- Being built by DG Translation
- Started: summer 2010
- Deploy: summer 2013

MT@EC service architecture



MT action lines

3. Service

2. Engines

1. Data

Multilingual web = Multilingual content

$$({}_{ML}W = {}_{ML}C)$$

${}_{ML}C = C$ available in $L_1, L_2, L_3, \dots, L_n$

How it is done:

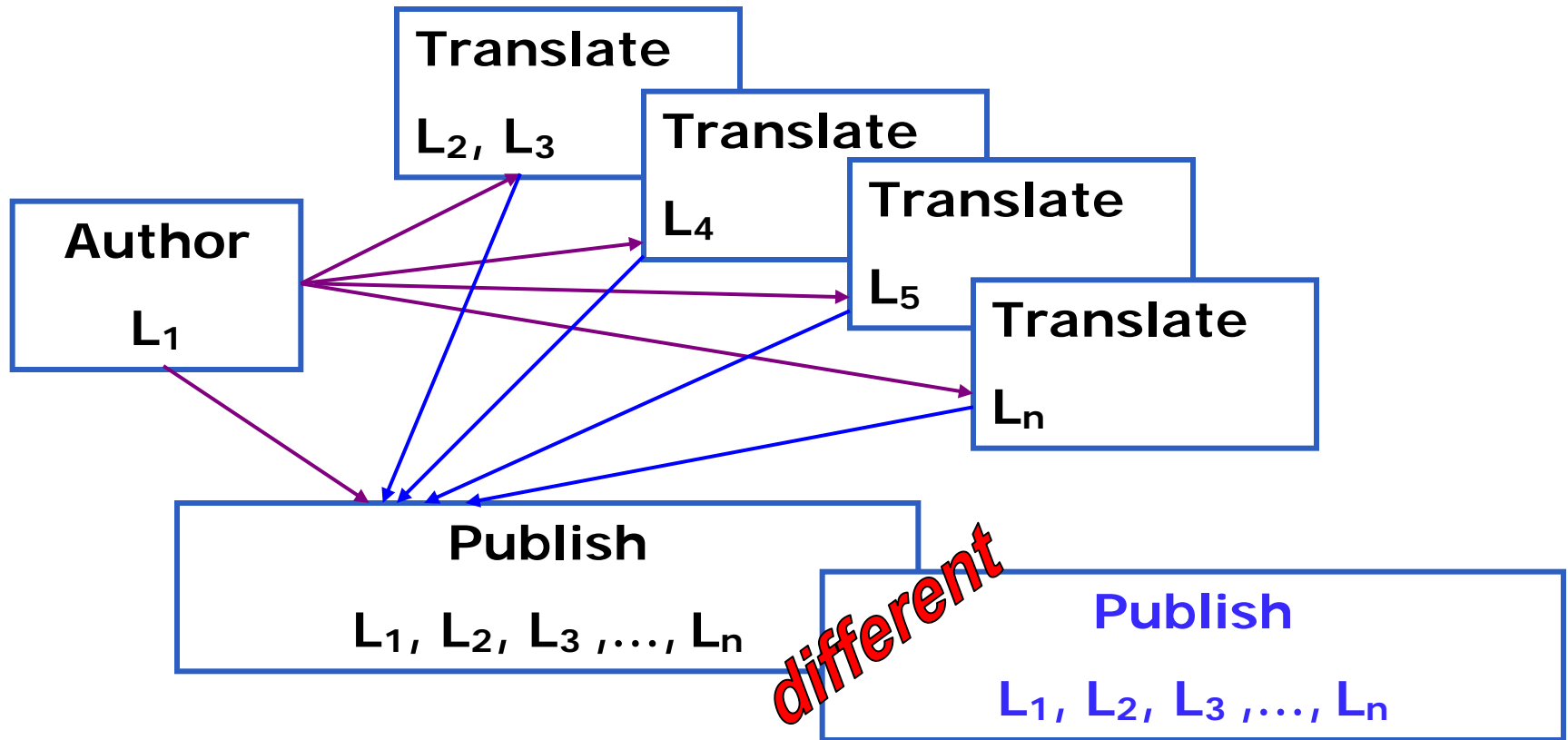
- Author produces **C** in one language
- Translators transfer the **C** to other languages
- Publisher gets **C** in all languages and publishes it



European
Commission

Multilingual web = Multilingual content

($MLW = MLC$)

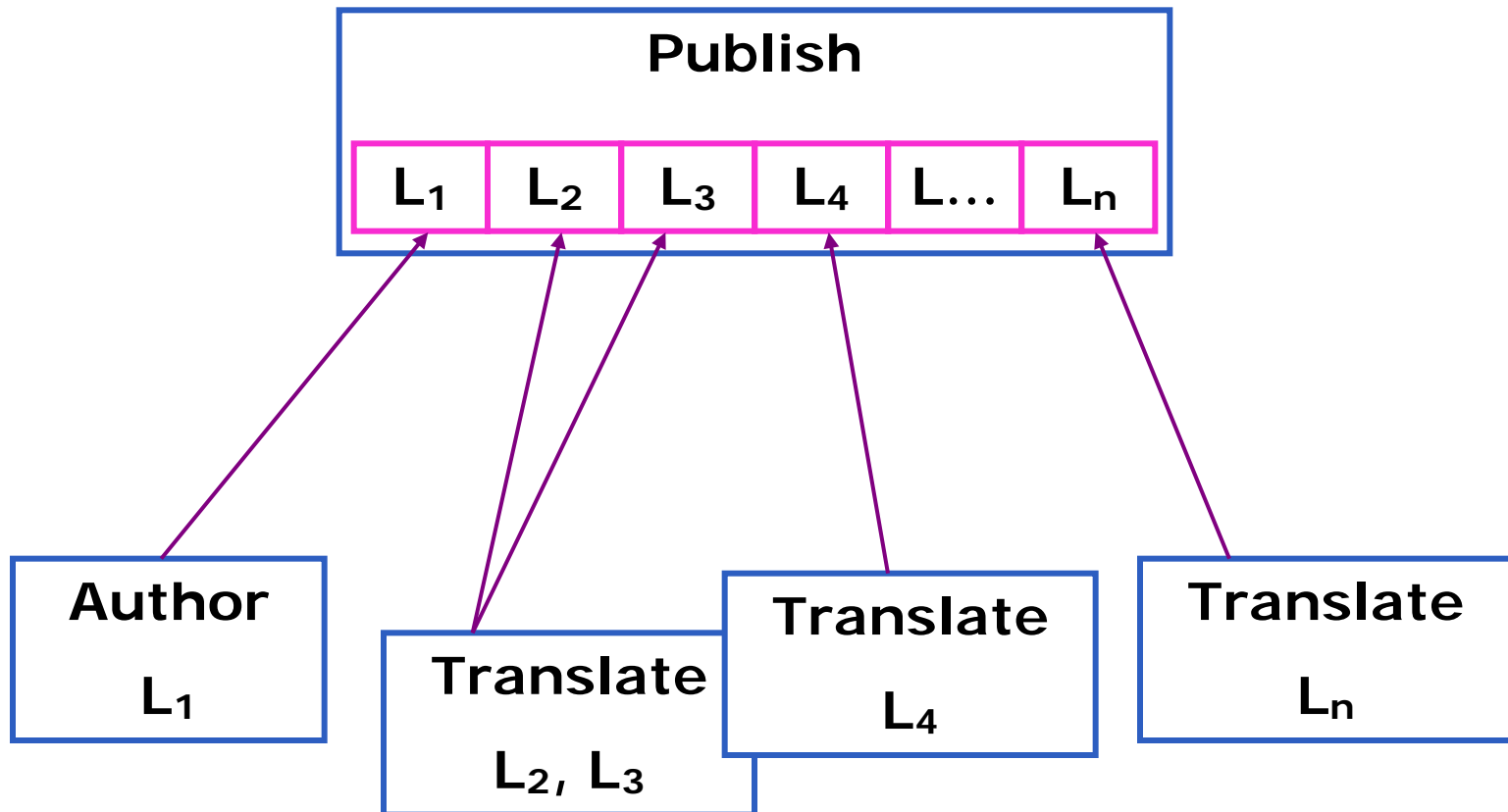




European
Commission

Multilingual web = Multilingual content

($MLW = MLC$)



Multilingual web = Multilingual content

($_{ML}W = {}_{ML}C$)

$_{ML}C = C$ available in $L_1, L_2, L_3, \dots, L_n$

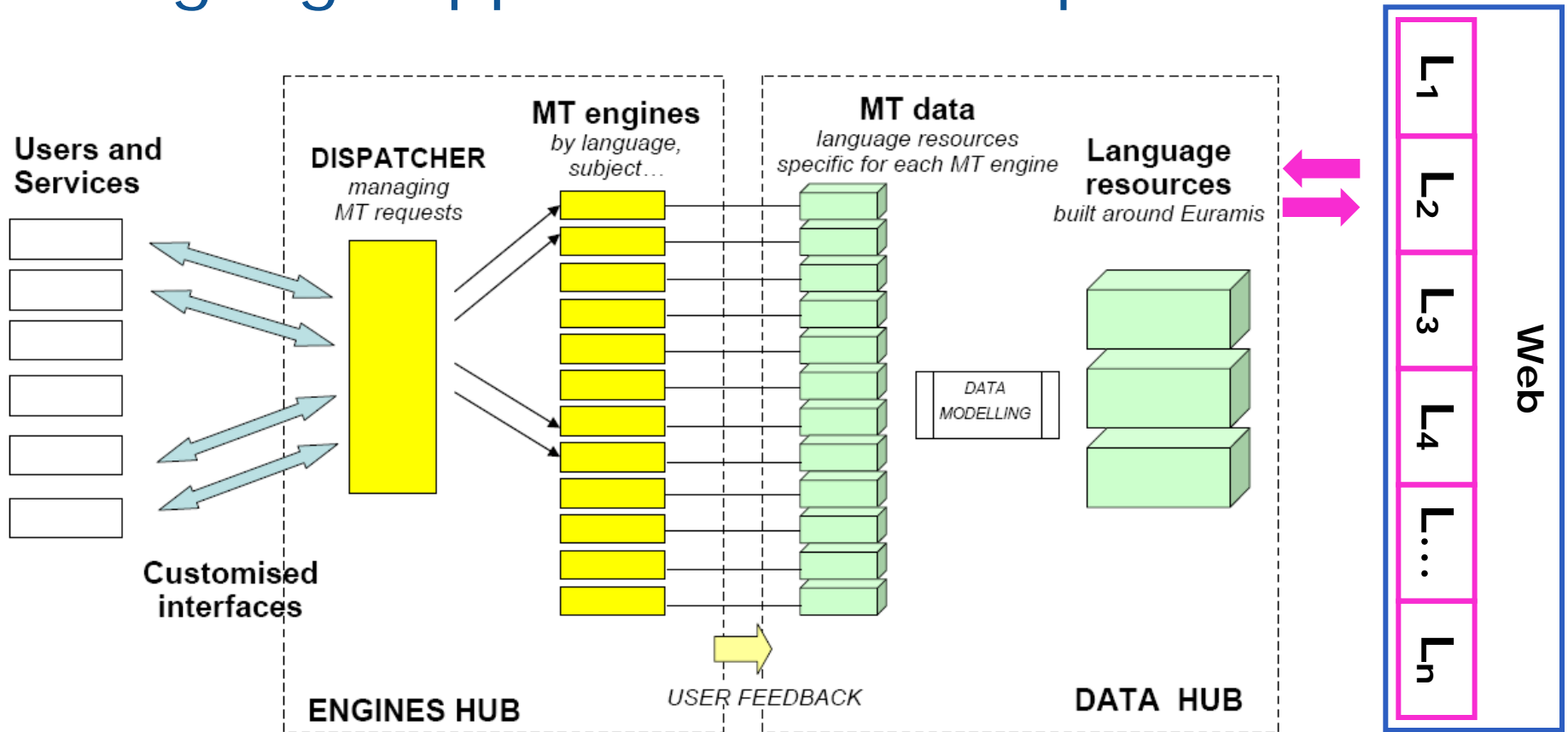
How it could be done:

- Publisher prepared for **C** in all languages (“placeholders”)
- Author produces **C** in one language “ready for publication”
- Translators produce **C** in other languages “ready for publication”



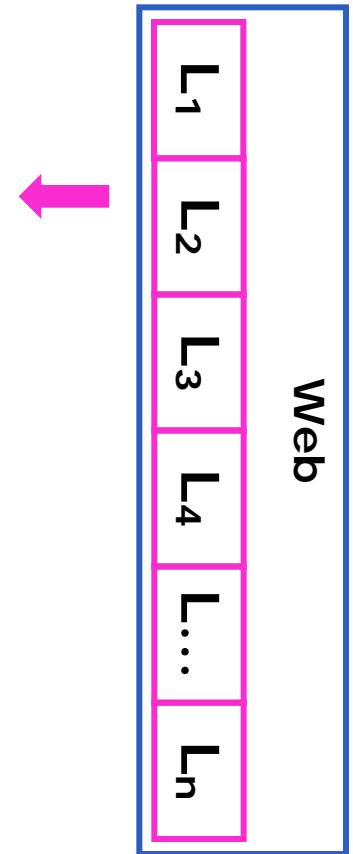
European
Commission

Language applications: example MT



Getting multilingual data from the web

- europa.eu has lots of **MLC**
- We tried with *web site 1*: not that difficult
BUT
...when going to *web site 2*, we need to analyse the way that site manages linguistic versions
...
...*web site n* has yet another way of managing linguistic versions
- ... and why not, since there is no standard to follow?



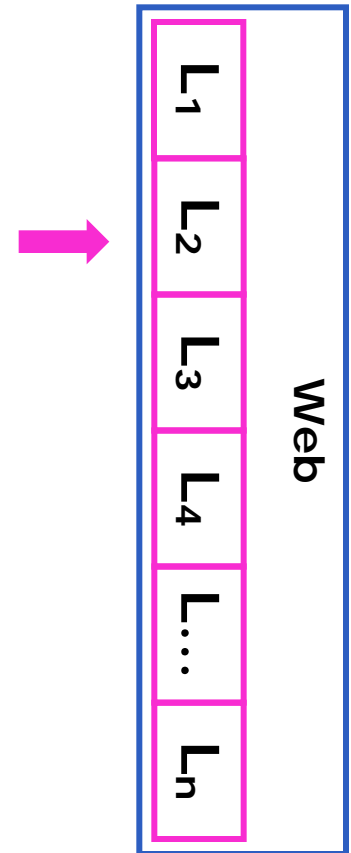
Giving multilingual data to the web

DGT-TM

- Translation memories from the Official Journal of the EU
- uses tmx
- yearly updates (to come): “do not change, so that there is continuity”

DGT-Acquis

- New resource (to come)
- Parallel text from the Official Journal of the EU
- First version: “do not change too much compared to JRC-Acquis”
- establish “our” standard way of working (which we make public)
- yearly updates (to come): “do not change, so that there is continuity”



Conclusion

- Need some order for multilingual content on the web
 - “Standard” data models?
 - “Standard” structure for data “storage” or “publication”?
 - Standard?
- MW consortium (and MW-LT) are meant to :
 - propose a feasible approach
 - demonstrate the benefit for all
- Commission DG for Translation: founding member of “Language interoperability portfolio” (Linport)

