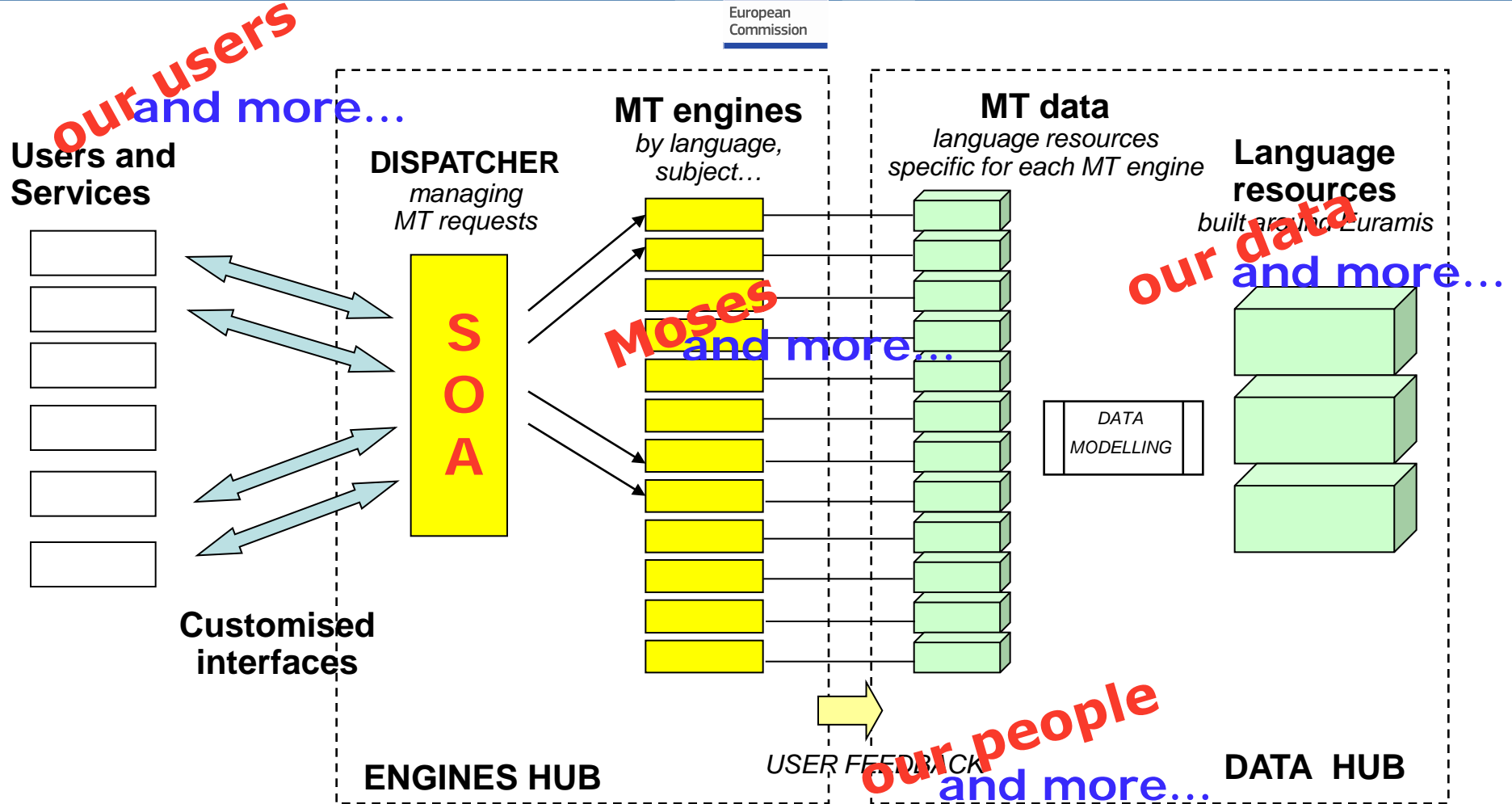# MT@EC

**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

DGT

MT@EC service architecture

European Commission

our users

and more…

Users and Services

DISPATCHER
managing MT requests

S O A

Customised interfaces

MT engines
by language, subject…

Moses

and more…

MT data
language resources specific for each MT engine

Language resources
built around Euramis

our data

and more…

DATA MODELLING

ENGINES HUB

USER FEEDBACK

our people

and more…

DATA HUB

MT action lines

| 3. Service | 2. Engines | 1. Data |

Multilingual web = Multilingual content
$(_{ML}\mathbf{W} = {}_{ML}\mathbf{C})$

$_{ML}\mathbf{C} = \mathbf{C}$ available in L_1, L_2, L_3, ..., L_n
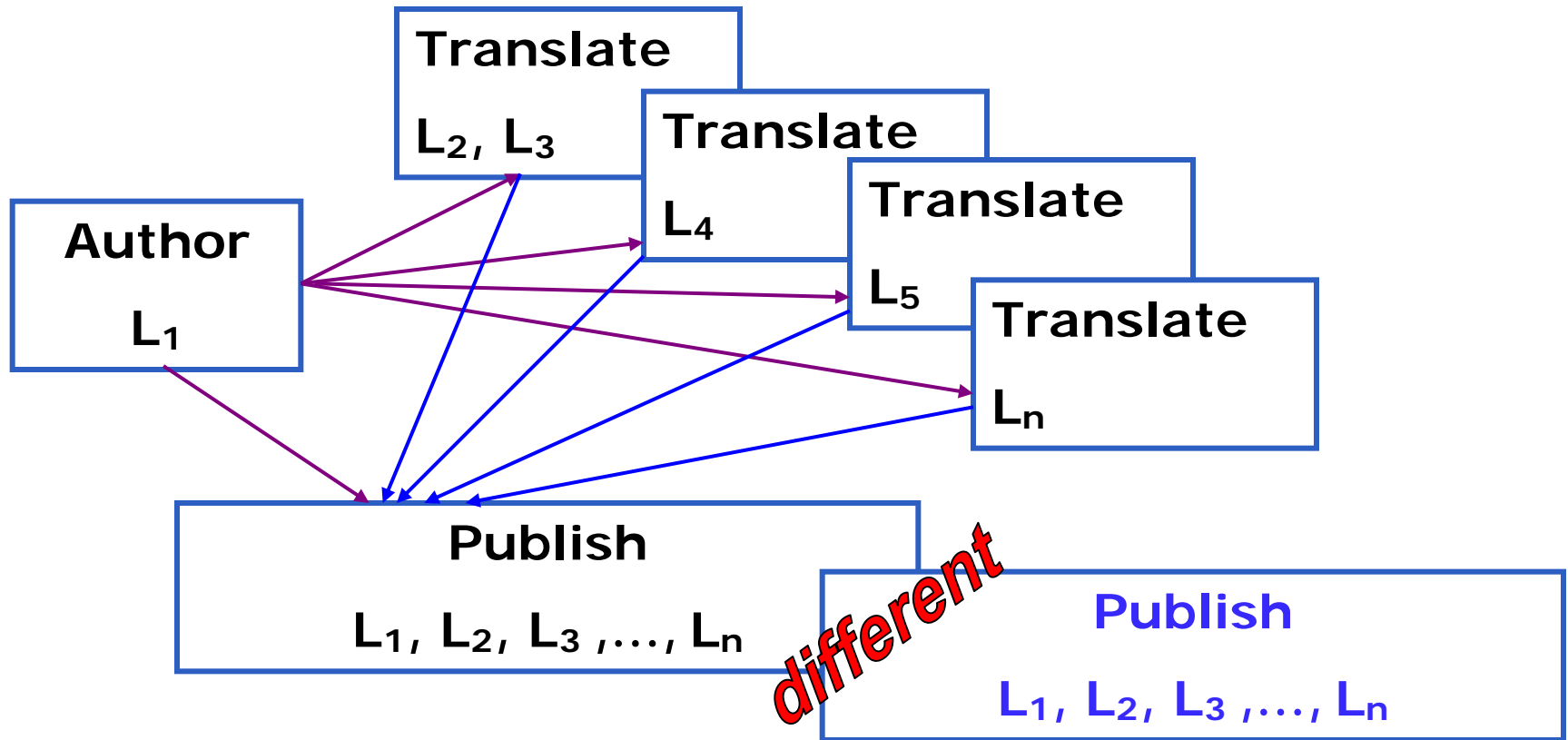
How it is done:

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

DGT

Multilingual web = Multilingual content
$$(_{ML}W = {}_{ML}C)$$
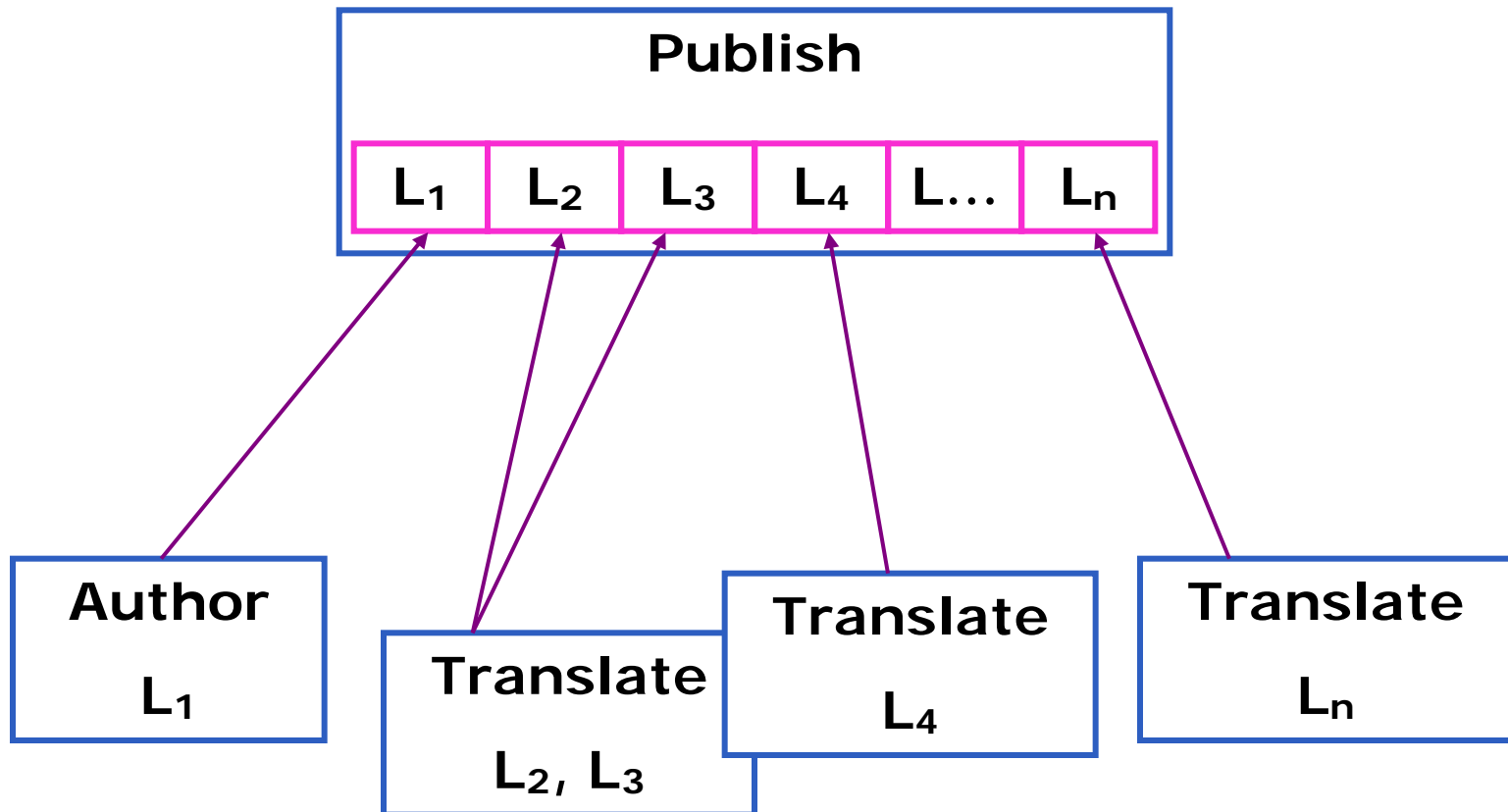
Translate

L_2, L_3

Translate

L_4

Translate

L_5

Author

L_1

Translate

L_n

Publish

L_1, L_2, L_3, ..., L_n

different

Publish

L_1, L_2, L_3, ..., L_n

Multilingual web = Multilingual content
$$(_{ML}W = \ _{ML}C)$$

Multilingual web = Multilingual content
$(_{ML}W = {}_{ML}C)$

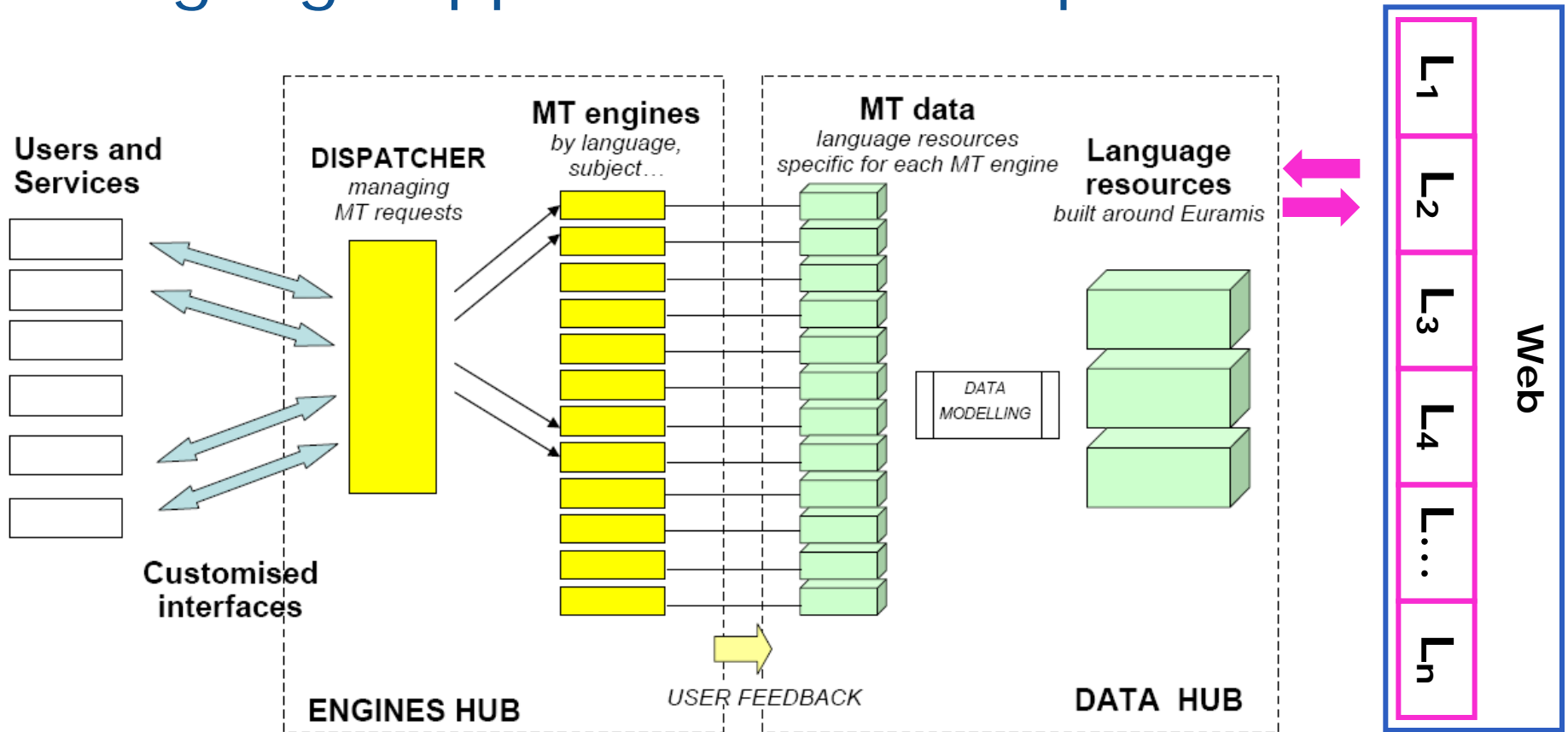$_{ML}C = C$ available in $L_1, L_2, L_3, ..., L_n$

How it could be done:

- Publisher prepared for **C** in <u>all</u> languages ("placeholders")
- Author produces **C** in <u>one</u> language "ready for publication"
- Translators produce **C** in <u>other</u> languages "ready for publication"
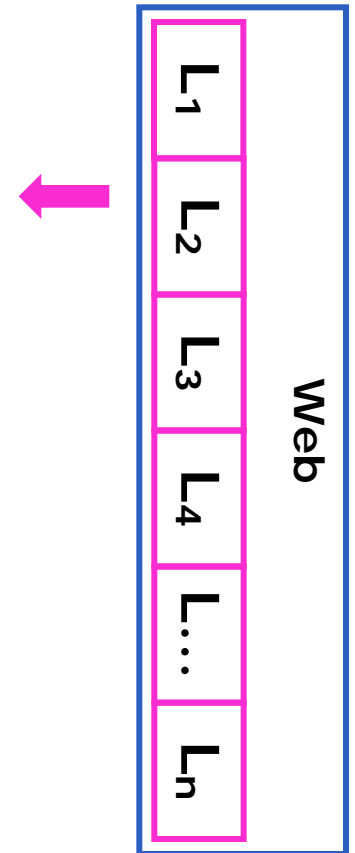
DGT

Language applications: example MT

Getting multilingual data from the web

- europa.eu has lots of $_{ML}C$

- We tried with *web site 1*: not that difficult

  BUT

  ...when going to *web site 2*, we need to analyse the way <u>that site</u> manages linguistic versions

  ...

  ...web site n has yet another way of managing linguistic versions

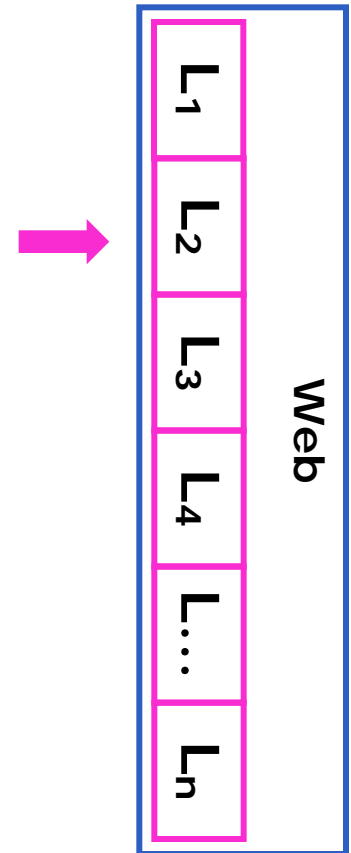- ... and why not, since there is no standard to follow?

L₁	
L₂	
L₃	Web
L₄	
L...	
Lₙ	

DGT

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"

L_1

L_2

L_3

L_4

$L_{...}$

L_n

Web

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 <u>feasible</u> approach
  - demonstrate the benefit <u>for all</u>
- Commission DG for Translation: founding member of "Language interoperability portfolio" (Linport)

DGT

Thank you
for your attention

DGT