

Country profiles: Status quo and vision

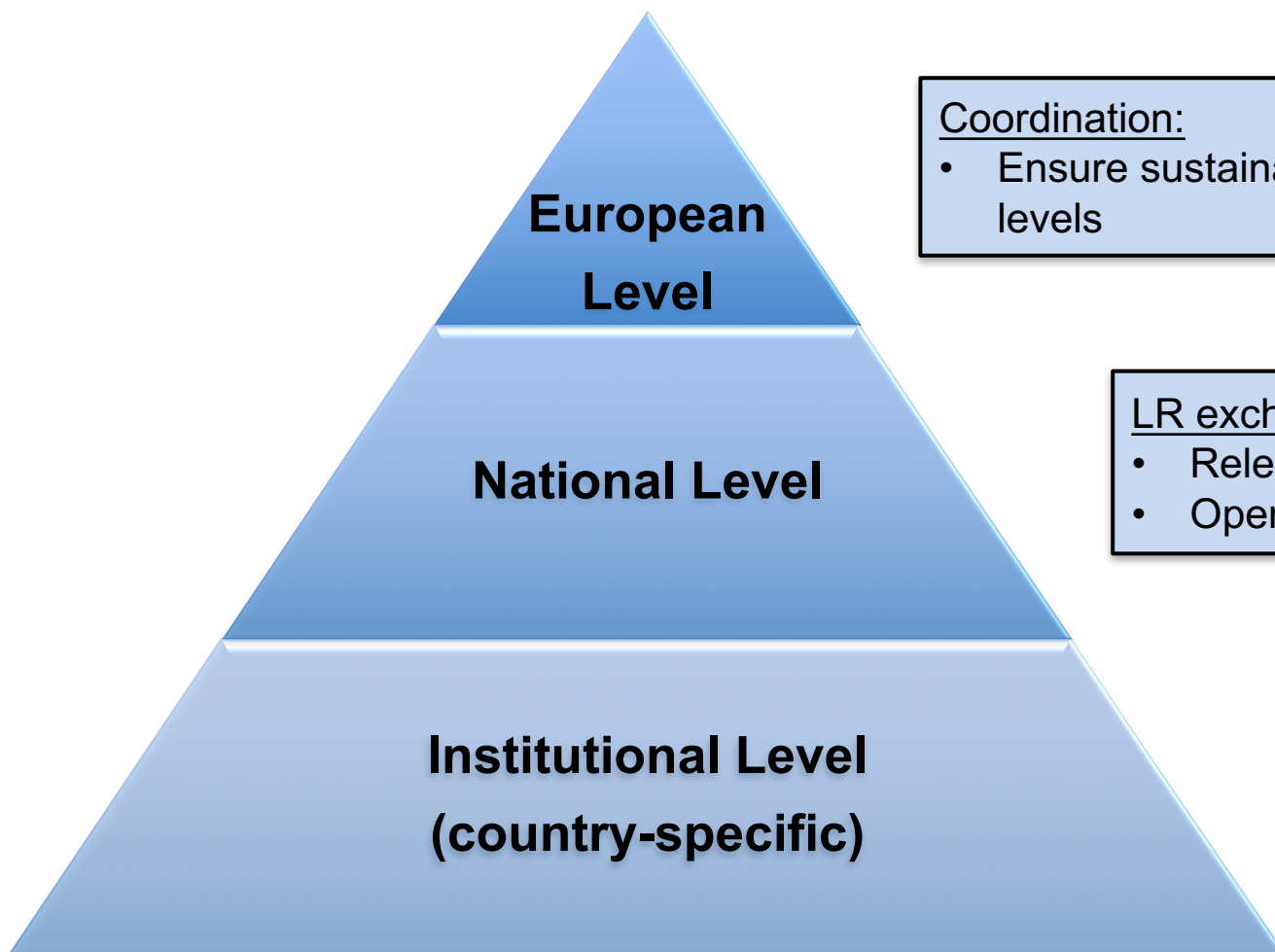
Lilli Smal (DFKI)
Andrea Lösch (DFKI)

... thanks to the ELRC Network

- ELRC National Anchor Points: www.lr-coordination.eu/anchor-points
- Public sector data providers



Key questions to be addressed



Coordination:

- Ensure sustainability of structures on all levels

LR exchange:

- Relevant stakeholders?
- Operational frame?

LR exchange:

- Key persons?
- Institutional frame?

- Goal: Understanding the frame for sustainable sharing of Language Resources
- Information provided (and sought) in country profiles:
 - Review of relevant stakeholders, including:
 - LR holders, LR creators, decision makers, existing networks, CEF representatives, ...
 - Infrastructure of LR exchange, including:
 - meeting of translation needs, services used, exchange of LR/translations/translation-relevant data on national level, coordination, open data (procurement), eGovernment initiatives/legislation, ...

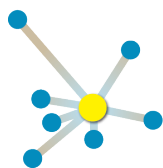
Example: Country Profile GER



	A	B	C	D	E	F
1						
2	Stakeholders register					
3						
4	Stakeholder group ▼	Stakeholder category ▼	Stakeholder category ▼	Stakeholder ▼	Stakeholder's position ▼	Level of influence ▼
5	LR holder and creator	Other	Public body, Terminology network, CEF e	Federal Ministry of the Interior, Building and Community // Bundesministerium des Innern	Resourceful	Federal
6	Language resource holder	Research institution/Academia		Institut für Deutsche Sprache	Resourceful	
7		Existing network		Bavarian State Minister of Finance, for Regional Development and the Homeland IT Commissioner of the Bavarian State Government	Resourceful	State
8				Berlin Senate Department for the Interior and Sport	Resourceful	State
9		Public body		Brandenburg State Ministry of the Interior and of Local Matters	Resourceful	State
10	LR holder and creator	Other	Public body, Terminology network	Bundeskartellamt		
11	LR holder and creator	Other	Public body, Terminology network	Bundesnetzagentur		Federal
12				Bundesverband der Übersetzer und Dolmetscher (BDÜ)		Federal
13				Deutsche Akademie für Sprache und Dichtung	Marginal	Federal
14				Deutsche Bahn		Federal
15				Deutsche Bahn		Federal
16				Deutsche Welle		Federal
17				DIN e.V.		Federal
18	LR holder and creator	Other	Public body, Terminology network	Federal Audit Office // Bundesrechnungshof		Federal
19	LR holder and creator	Other	Public body, Terminology network	Federal Constitutional Court		Federal
20	LR holder and creator	Other	Public body, Terminology network	Federal Criminal Police Office // Bundeskriminalamt		Federal
21	LR holder and creator	Other	Public body, Terminology network	Federal Financial Supervisory Authority // Bundesanstalt für Finanzdienstleistungsaufsicht		Federal
22	LR creator and holder	Other	Public body, Terminology network	Federal Foreign Office // Auswärtiges Amt		Federal
23	LR holder and creator	Other	Public body, Terminology network	Federal Foreign Office // Auswärtiges Amt		Federal
24	LR holder and creator	Other	Public body, Terminology network	Federal Foreign Office // Auswärtiges Amt		Federal
25	LR holder and creator	Other	Public body, Terminology network	Federal Institute for Vocational Education and Training // Bundesinstitut für Berufsbildung	Resourceful	Federal
26	LR holder and creator	Other	Public body, Terminology network	Federal Language Office (Federal Ministry of Defense) // Bundessprachenamt	Resourceful	Federal
27	LR holder and creator	Other	Public body, Terminology network	Federal Language Office (Federal Ministry of Defense) // Bundessprachenamt	Resourceful	Federal

◀ ▶
Stakeholders register
Infos LR exchange
Infrastructure of LR creation
Values for lists
+

Bereit



Example: Country Profile GER



	A	B	C	D	E
1	Infrastructure of language resource exchange				
2					
3	CEF Coordinators		How are translation needs in the public sector met? What services are used?	Is there any exchange of data on a national level?	Does anyone coordinate data exchange/translations?
4	CEF Telecom		All federal ministries have their own language service, only few ministries on the state level have their own language services; all members of the terminology network have translation services and use MultiTerm Online;	No exchange of information or data between public services unless they are part of the terminology network (Terminologiedatenbankverbund) All members use the collaborative platform: <i>MultTerm Online</i>	Federal CEF coordinator subordinated to Ministry of Transport and Digital Infrastructure (Email Alexandra Soska, 8.7.16)
5	eTranslation			Terminology network - members:	
6	eDelivery			AA= Federal Foreign Office // Auswärtiges Amt BaFin= Federal Financial Supervisory Authority // Bundesanstalt für Finanzdienstleistungsaufsicht	
7	eID			BfJ= Federal Office of Justice // Bundesamt für Justiz	
8	eInvoicing			BIBB= Federal Institute for Vocational Education and Training // Bundesinstitut für Berufsbildung	
9	eSignature			BKA= Bundeskriminalamt BKartA= Bundeskartellamt	
10				BMAS= Federal Ministry of Labour and Social Affairs // Bundesministerium für Arbeit und Soziales	
11				BMBF= Federal Ministry of Education and Research // Bundesministerium für Bildung und Forschung	
12				BMEL= Bundesministerium für Ernährung und Landwirtschaft // Federal Ministry of Food and Agriculture	
13				BMF= Federal Ministry of Finance // Bundesministerium	
14					

◀ ▶

Stakeholders register

Infos LR exchange

Infrastructure of LR creation

Values for lists

+

Bereit

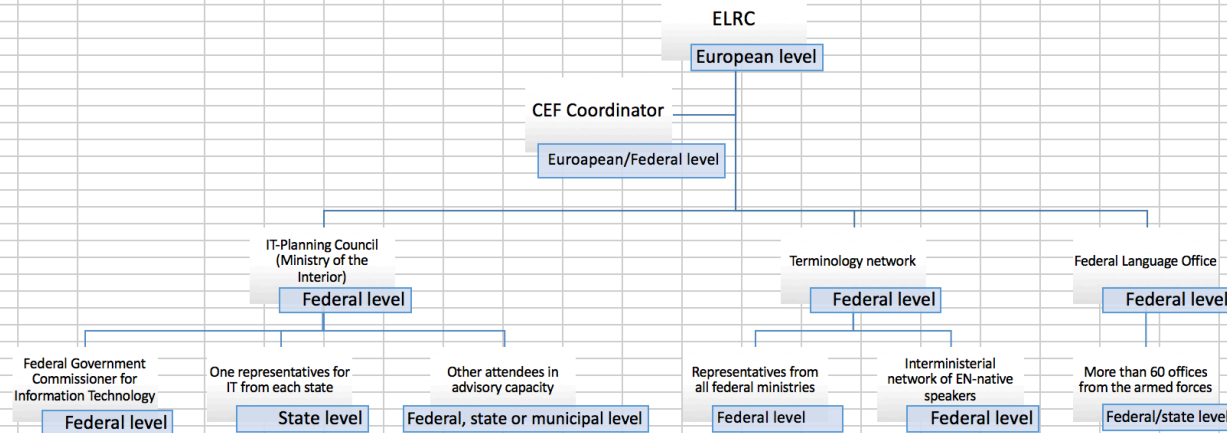
90 %



Example: Country Profile GER



Decision making - Strategic level of eGovernment and IT



Stakeholders register

Infos LR exchange

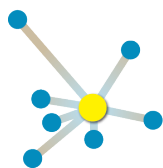
Infrastructure of LR creation

Values for lists



Bereit





Example: Country Profile GER



	A	B	C	D	E	F	G	H
1								
2								
3	Stakeholder group		Stakeholder categories		Stakeholder's position (help and harm potential)		Level of influence	To be contacted by:
4	Language resource holder	I.e. institutions that subcontract translations	Public body		Resourceful	I.e. high help potential, low harm potential	Federal	ELRC consortium
5	Language resource creator	I.e. institutions that create translations	Public online service		Key Player	I.e. high help and high harm potential	State	T NAP
6	LR holder and creator		Existing network		Show Stopper	I.e. high harm, low help potential	Municipal	P NAP
7			Language service provider		Marginal	I.e. low help and low harm potential		N/A
8				This includes the CEF telecom coordinator and the coordinators for each building block				
9			Research institution/Academia		N/A			
10			CEF Coordinator					
11			Decision maker					
12			Other					
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								
25								
26								
27								
28								
29								
30								
31								

◀ ▶

Stakeholders register

Infos LR exchange

Infrastructure of LR creation

Values for lists

+

Bereit

📄

📄

⏮

⏭

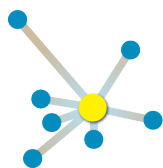
⏪

⏩

+

- Please provide your input by **Monday, 14 May**
- Individual follow-up
- Presentation of results during next round of Q&A online sessions early June

- What's the purpose: eTranslation is not just its MT engines. It will integrate additional NLP tools to enhance its performance -> Need to compile a meaningful list of relevant tools
 - Except for collecting language resources, ELRC acts also as an observatory of existing tools and services
- A preliminary list has been shared with you (based on tools already catalogued in LT community repositories and databases). It is as simplified as possible:
 - Name of tool/service
 - Short description
 - Language(s)
 - Function(s)
 - Some admin info (e.g. owner, source repository etc,)



Contributor	Source	Name	Short description	Language	Free text & Language
Kanella Poulakaki	MetaShare	Coral Corpus Aligner	Coral (CORpus ALigner) is a tool for easy bilingual parallel corpora alignment	language-independent	
	MetaShare	Lingua-Align	A toolbox for automatic syntactic tree alignment and for word alignment	language-independent	
	Robert C. Moore	Moore's aligner	NO INFO		
	Johns-Hopkins University	GIZA++	GIZA++ is an extension of the program GIZA (part of the SMT toolkit EGYPT) which was developed by the Statistical Machine Learning Group		
	MetaShare	Yet Another Word Alignment Algorithm (YAWA) [Romanian to English]	YAWA is a Romanian to English word alignment algorithm that will generate word alignments	multilingual	EN, RO
	PRESEMT	Present Phrase Aligner Module (PAM)	The Phrase Aligner Module processes the bilingual corpora by performing token-to-token alignment	language-independent	DE, EL, EN, ZH
	MetaShare	hunalign	hunalign is a sentence aligner. It can use bilingual lexicons as a resource, but it does not require them	language-independent	
	MetaShare	Collocation Extractor	A stand-alone application which identifies and extracts collocations along with their frequency	language-independent	
	MetaShare	Hydra - tool for developing wordnets	Hydra is a tool for editing, viewing, searching and validating wordnet. The Hydra API for wordnet processing uses abstract syntax trees		
	MetaShare	PELCRA EN Lemmatizer	PELCRA EN Lemmatizer is a British National Corpus-derived lemma dictionary	EN	
MetaShare	MetaShare	Lemmatizer for Portuguese	Based on the MXPOST part of speech tagger and UNITEK dictionaries for Portuguese	PT	
	SoapLAB	ILSP Lemmatizer	Lemmatizer for Greek texts	EL	
	UPF/ SoapLAB	anonymizer	The anonymizer web service can be used to substitute proper nouns with tokens	multilingual	Asturian-Bable
	MetaShare	Croatian Lemmatisation Server	The Croatian Lemmatisation Server (CLS) is a web-based service for lemmatization	HR	
	MetaShare	Morphological analyzer for Estonian ESTMORF	ESTMORF is a computer program for analysing unrestricted Estonian text. It can be used as a morphological analyzer, POS-tagger and dependency/constituent parser	ET	
	MetaShare	SweGram Constraint Grammar Parsers for Swedish	morphological analyzer, POS-tagger and dependency/constituent parser for Swedish	SV	
	MetaShare	NERanka: Named Entity Recognition and Annotation Tool	This tool is a web application for named entity (NE) recognition and tagging	language-independent	SR
	MetaShare	U-Compare E-txt2DB: Giving structure to unstructured data	Extxt2DB is a framework for specifying and executing Entity Recognition (ER) queries	EN	



Northern region – Room “Président”

South-Western region – Room “Président”

South-Eastern region – Room “Belle Epoque”

Generic services – Room Meyerbeer

Thank you for your attention!

Email: info@lr-coordination.eu
Website: www.lr-coordination.eu

