

# ELRC COUNTRY PROFILES

OVERCOMING OBSTACLES AND CREATING SUSTAINABLE  
LANGUAGE DATA SHARING INFRASTRUCTURES IN EUROPE

Lilli Smal (DFKI)



# THE ELRC COUNTRY PROFILES

---

**The Idea**

Process and Main Findings

Recommendations and White Paper

**Vision:** True Digital Single Market

**Mission:** Creating sustainable language data pipelines

**Purpose:** Identifying and collecting language resources

## A TRUE DIGITAL SINGLE MARKET...

„I think that the overall vision is the true Digital Single Market where all EU citizens can access information relevant and of interest to them, no matter what language(s) they speak. Language should not be a barrier, but is currently one of the most substantial challenges to have a truly integrated European Union. These barriers affect all areas of EU lives – whether it is cross-border services or trade and impacts all levels of society. By providing automated translation we are addressing these challenges, and ELRC’s task of collecting data and establishing pipelines is a concrete action specifically addressing the language barriers.“

Susan Fraser, Project Officer, EC



## A TRUE DIGITAL SINGLE MARKET...

- ...means equal access to information, in all official EU languages,
- ...means having access to cross-border trade using any official EU language,
- ...means having access to digital public services in all official EU languages,
- ...means **language equality** for all official EU languages, Norwegian and Icelandic.



# HOW CAN LANGUAGE DATA HELP TO OVERCOME LANGUAGE BARRIERS AND REALIZE A TRUE DIGITAL SINGLE MARKET?





Language Technologies run on data  
— LANGUAGE DATA





## WHAT DATA ARE WE LOOKING FOR?

## WHAT DATA ARE WE LOOKING FOR?

- Ideally, bi- or multilingual parallel text corpora in TMX format
- Non-personal public sector information (according to the Open Data Directive 2019/1024/EC)
- “Public Sector Information is information generated, created, collected, processed, preserved, maintained, disseminated, or funded by or for the Government or public institution” (European Data Portal: Analytical Report 9: The Economic Benefits of Open Data, [https://www.europeandataportal.eu/sites/default/files/analytical\\_report\\_n9\\_economic\\_benefits\\_of\\_open\\_data.pdf](https://www.europeandataportal.eu/sites/default/files/analytical_report_n9_economic_benefits_of_open_data.pdf), 2017, p.7.)



**SO, WHY IS IT DIFFICULT TO COLLECT  
LANGUAGE DATA?**

# ELRC COUNTRY PROFILES

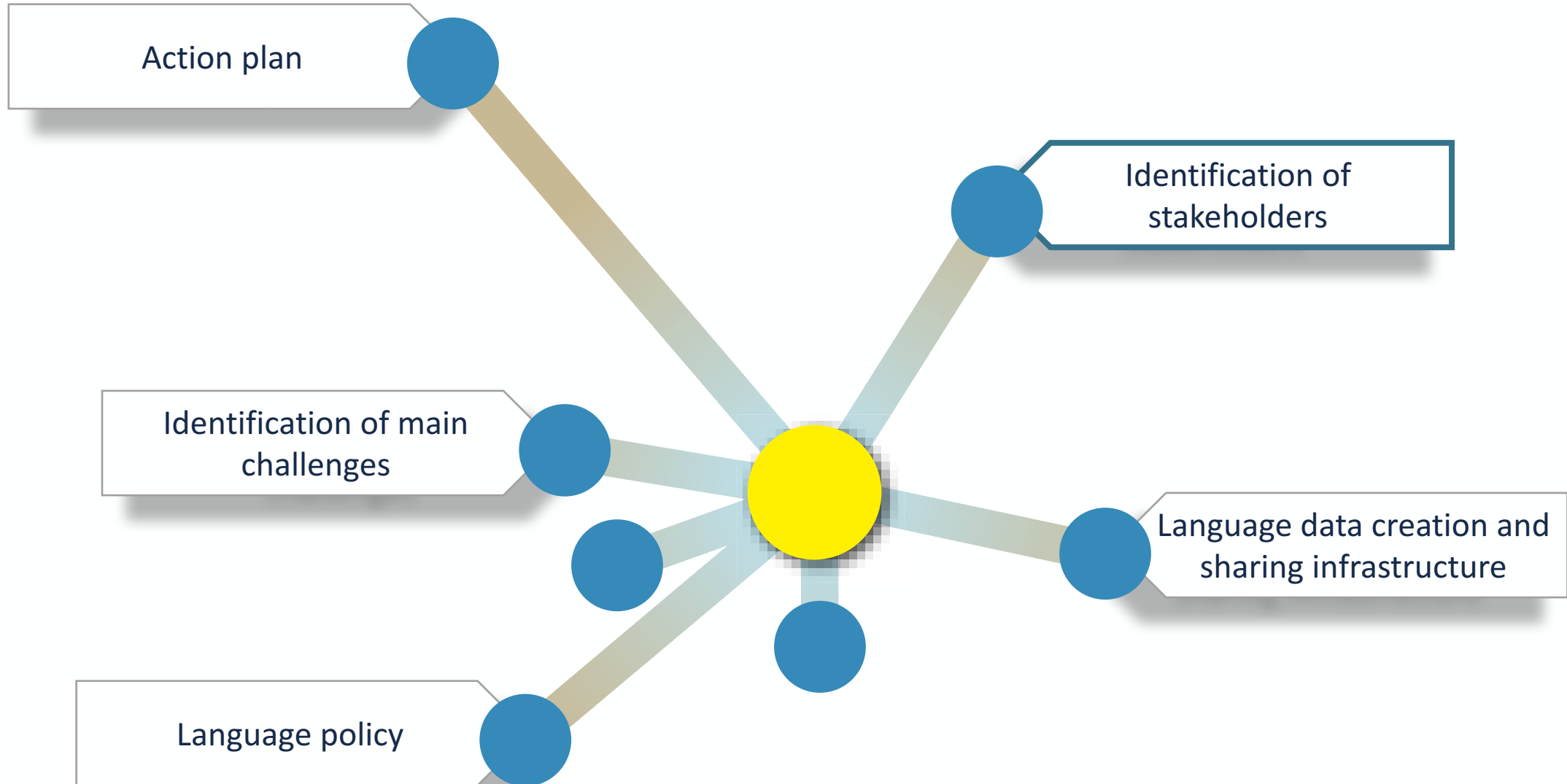
---

The Idea

**Process and Main Findings**

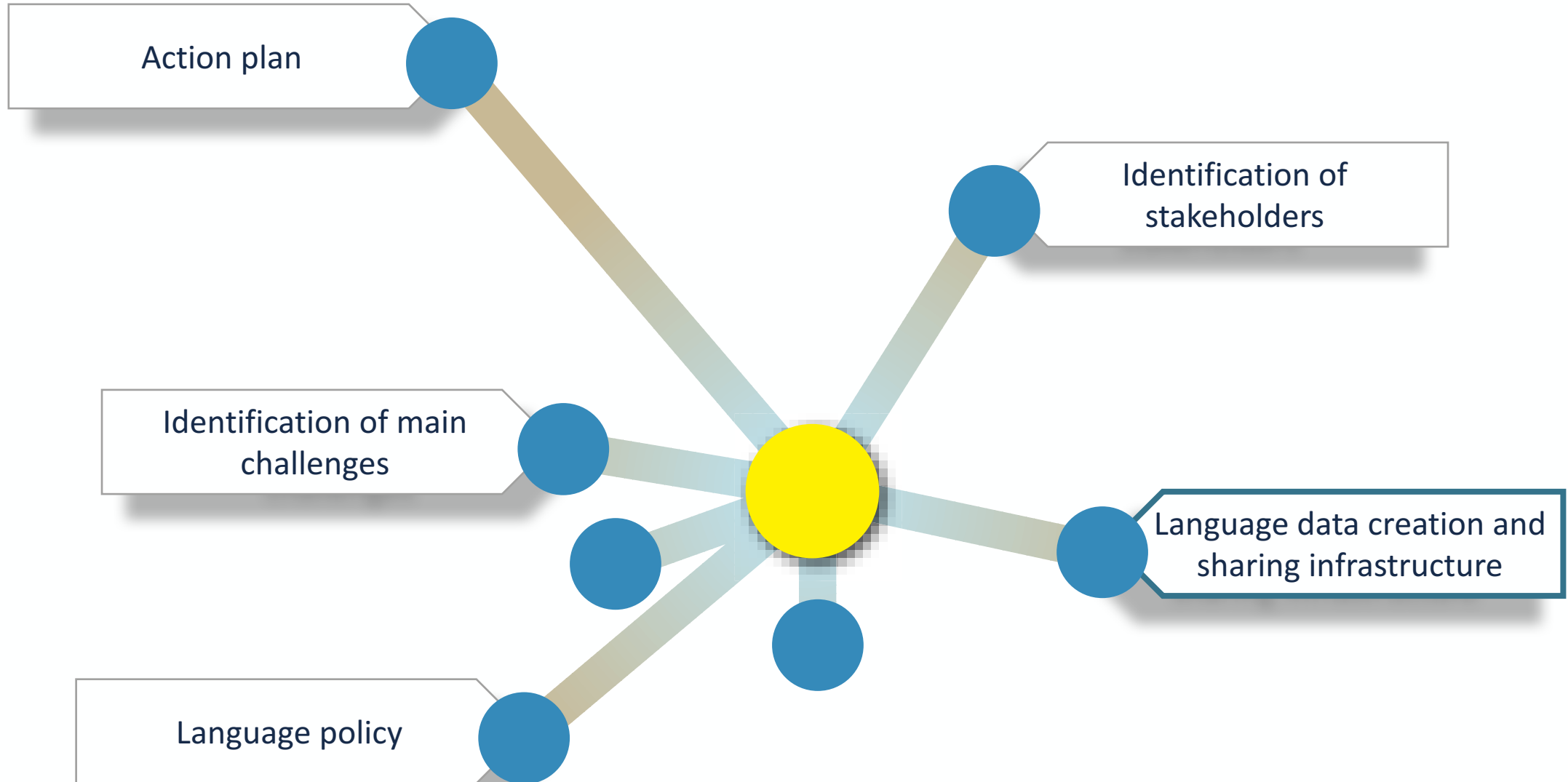
Recommendations and White Paper

# COUNTRY PROFILES

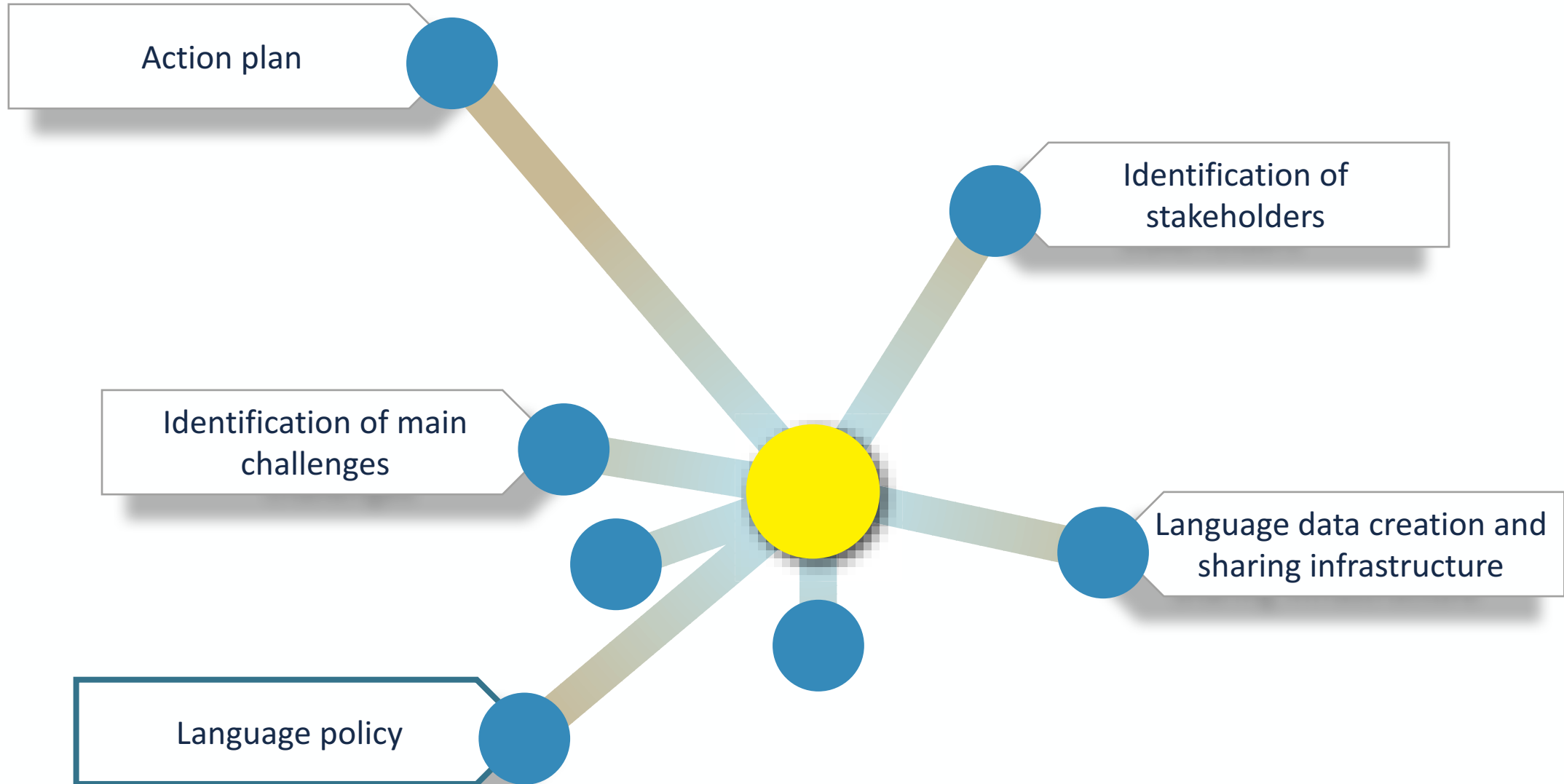




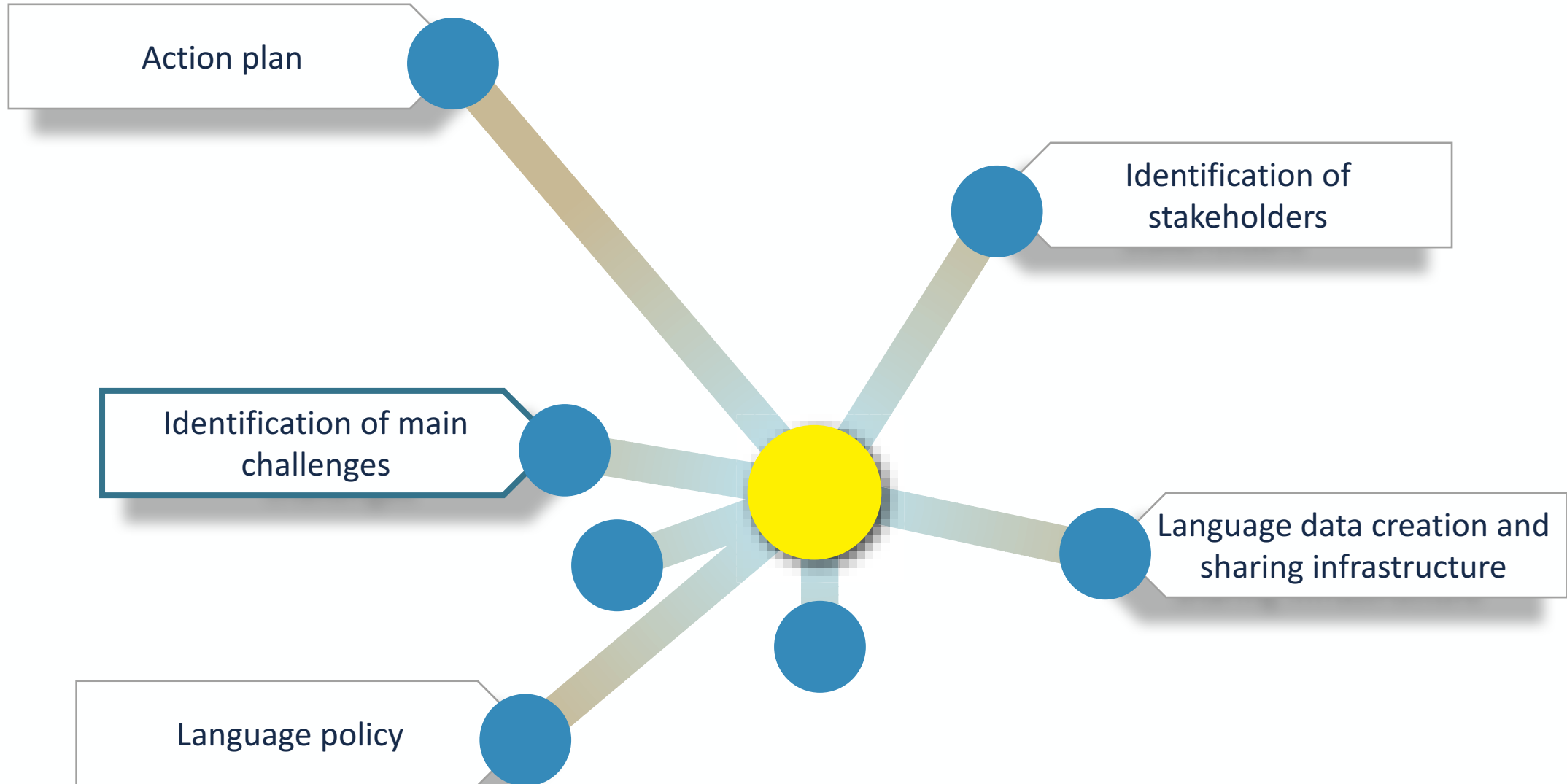
# COUNTRY PROFILES



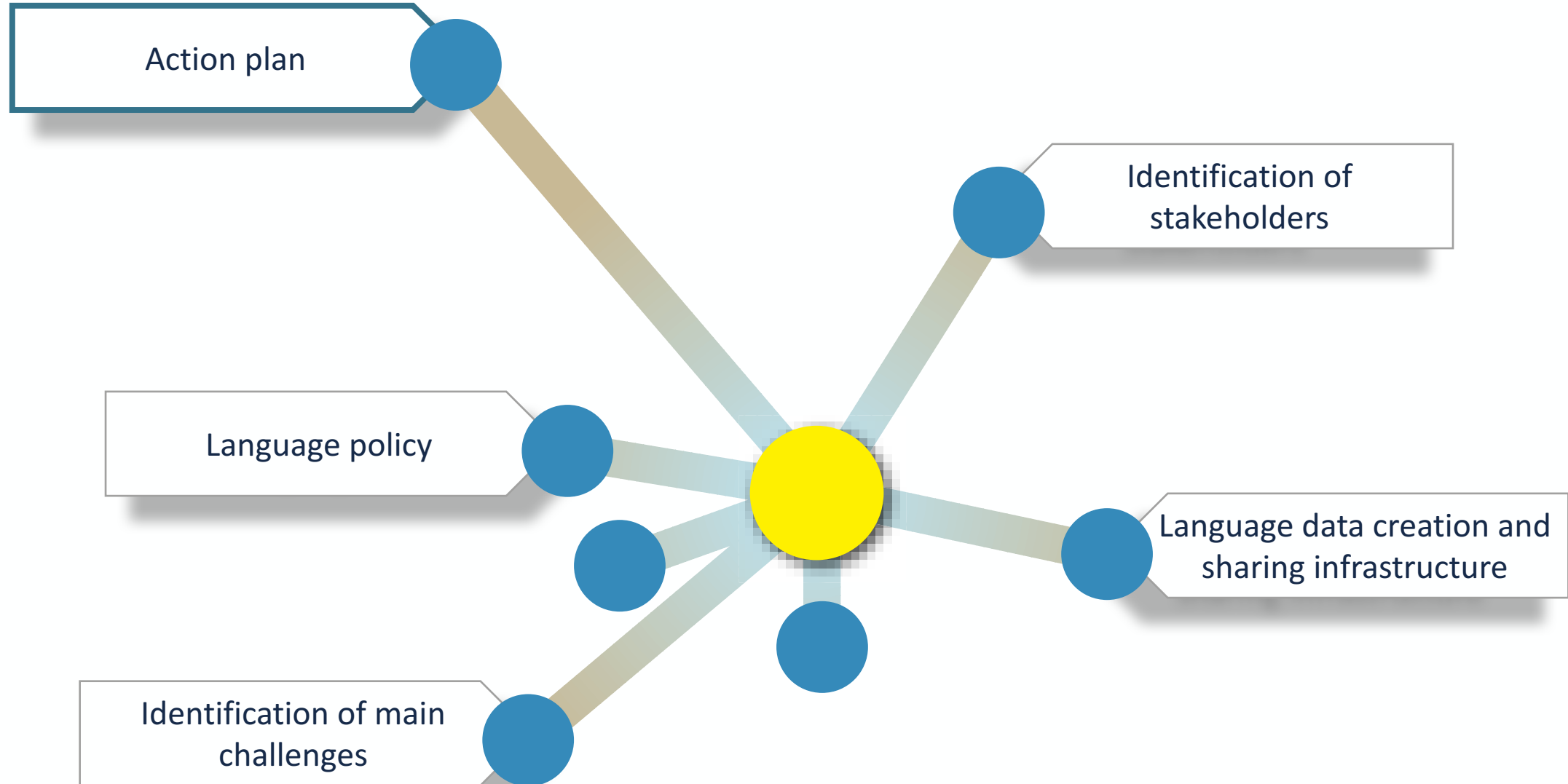
# COUNTRY PROFILES



# COUNTRY PROFILES



# COUNTRY PROFILES

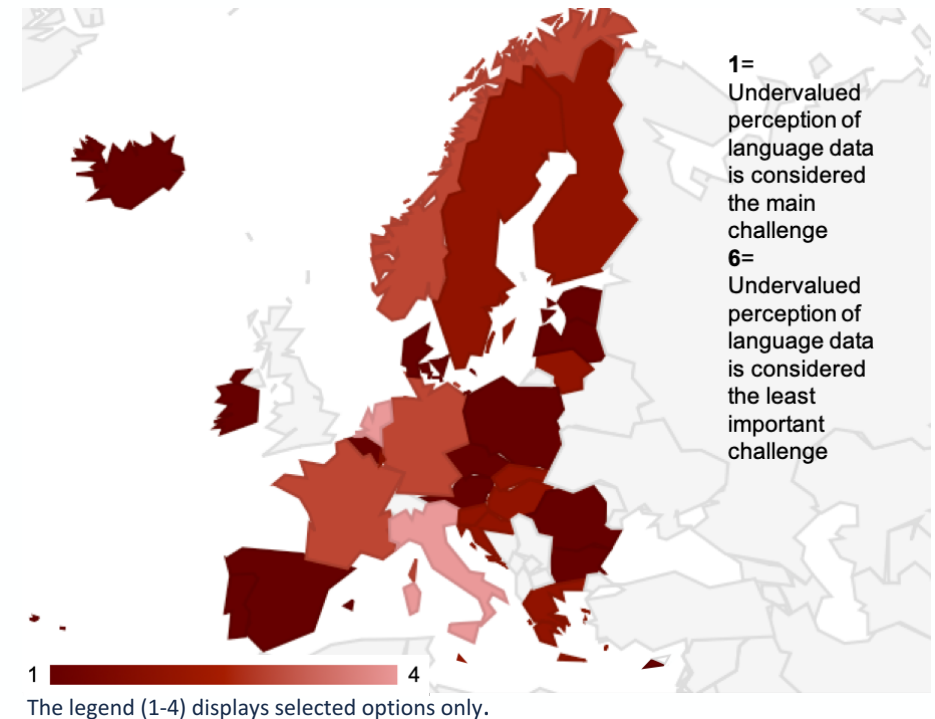


# LANGUAGE DATA IS NOT CONSIDERED VALUABLE



# LANGUAGE DATA IS NOT CONSIDERED VALUABLE

- 14 countries indicated that the most important objective is to change the perception that language data has no added value
- 10 countries ranked this as the second most important objective/challenge
- Since language data is not regarded valuable, little or no resources are allocated towards managing and curating language data



# STRUCTURAL CHALLENGES



## STRUCTURAL CHALLENGES

- Rated as the second most important challenge
- Less than 30% of the countries have a dedicated (language) policy including language technologies (LT)
- Only few countries have a public body that either coordinates the translation process in public administrations or the language data collection process
- Decision makers are very hard to identify or change frequently

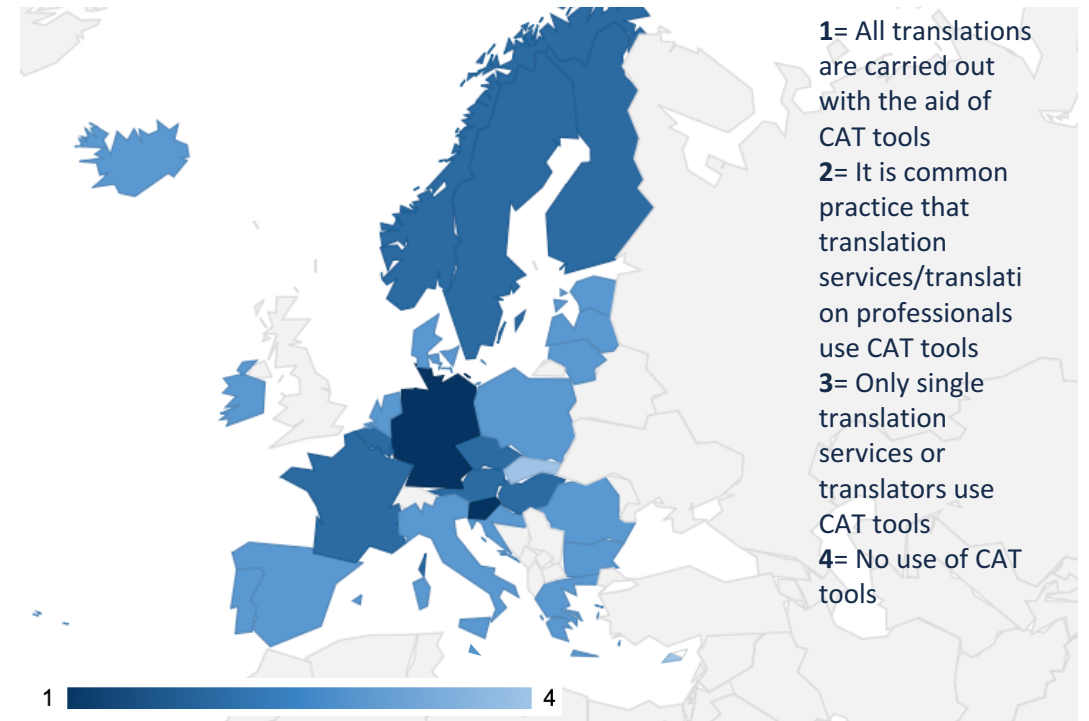


# DISPOSITION TOWARDS LANGUAGE TECHNOLOGY AND LACK OF DIGITAL SKILLS

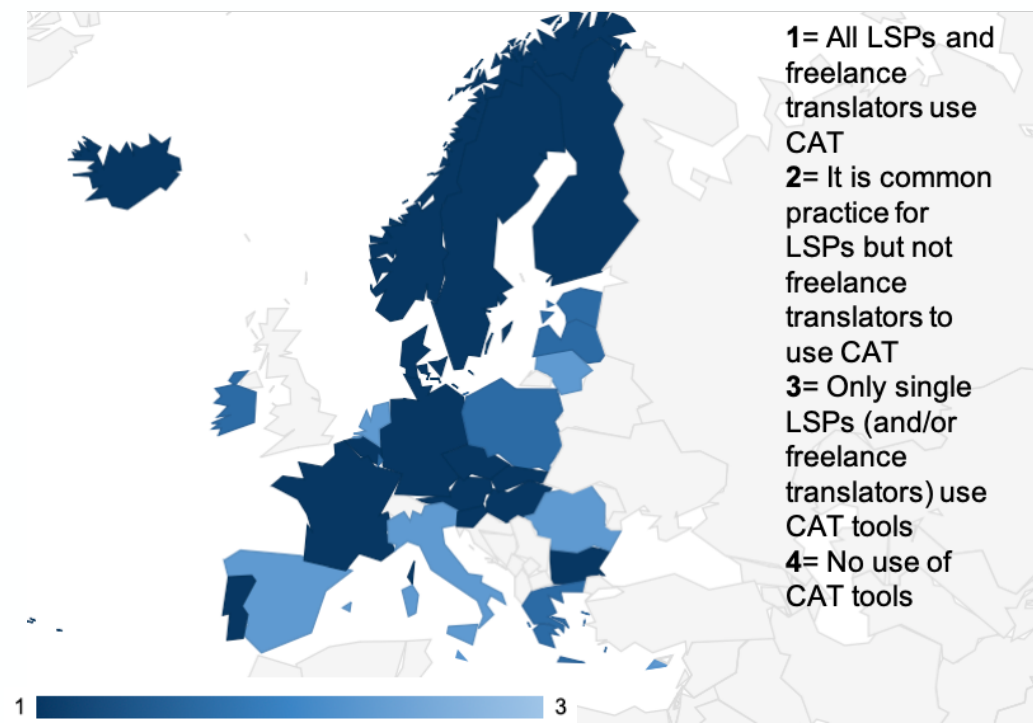
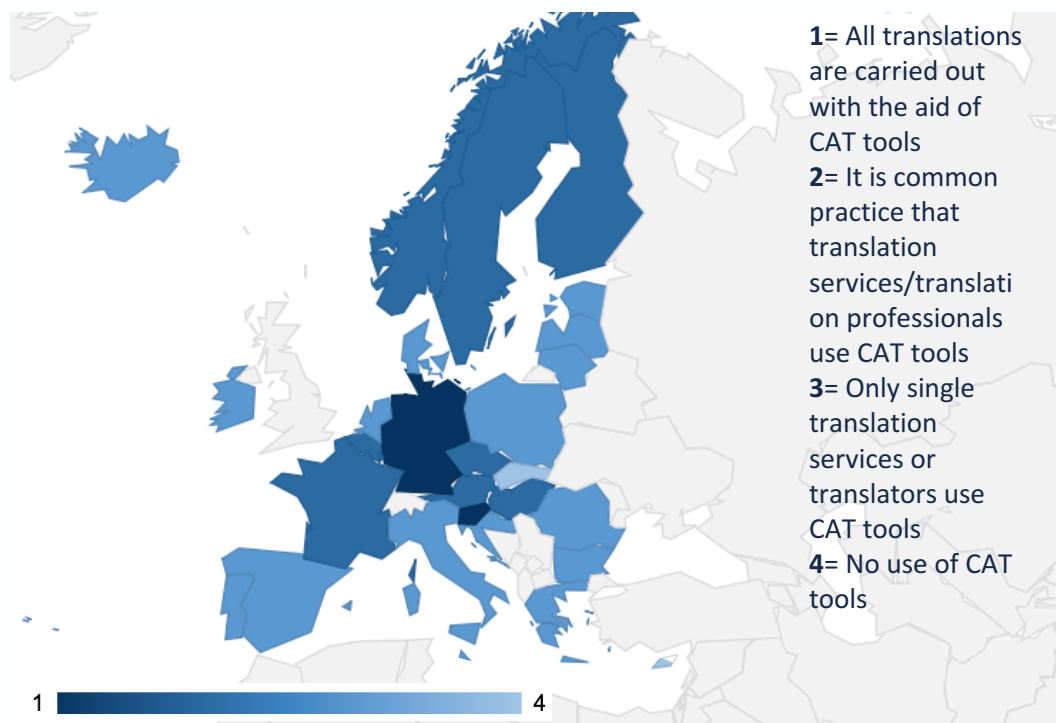


## USE OF CAT TOOLS IN PUBLIC SERVICES

- Only two countries (Germany and Slovenia) indicated that all translations are carried out with the help of CAT tools
- Nine countries said that it is common practice to use CAT tools in the translation process
- And 16 countries said that only single translation services or translators use CAT tools
- Two countries indicated that CAT tools are not used at all (Cyprus and Slovakia)



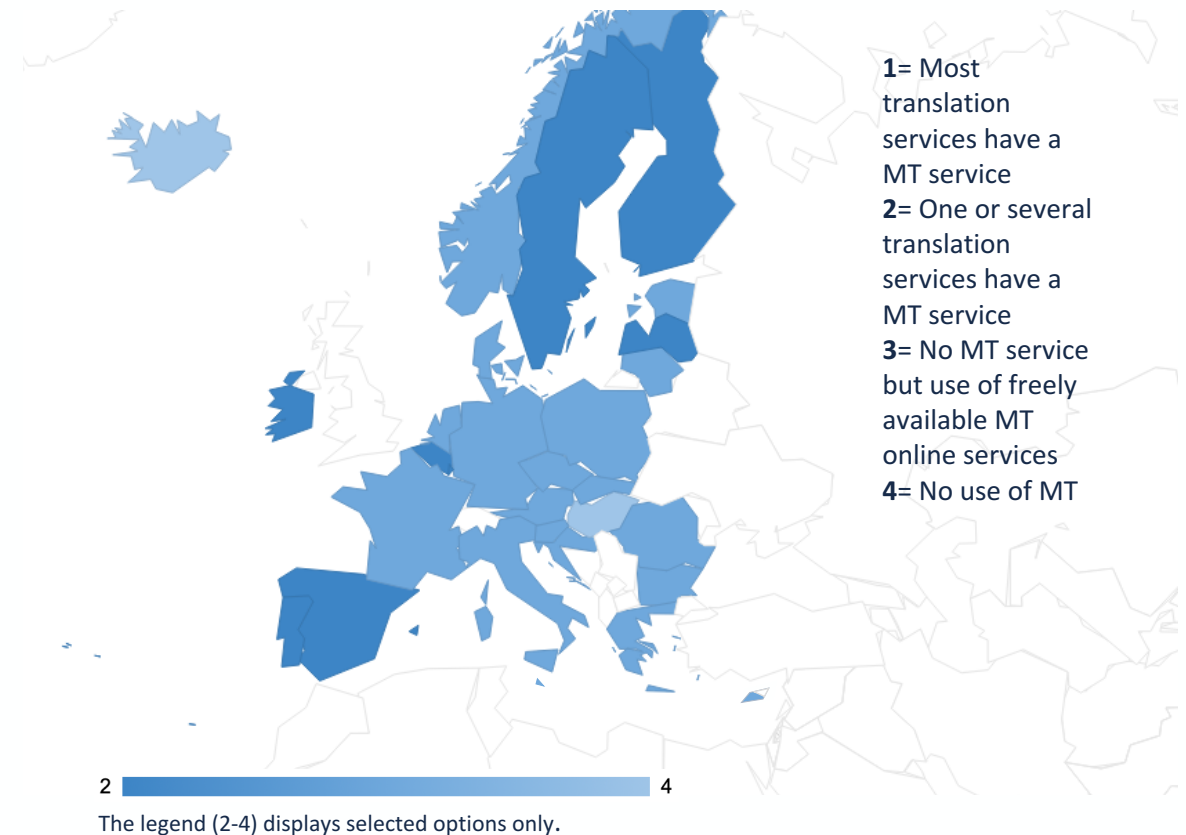
## USE OF CAT TOOLS IN PUBLIC SERVICES COMPARED TO LANGUAGE SERVICE PROVIDERS



The legend (1-3) displays selected options only.

## USE OF MACHINE TRANSLATION IN PUBLIC SERVICES

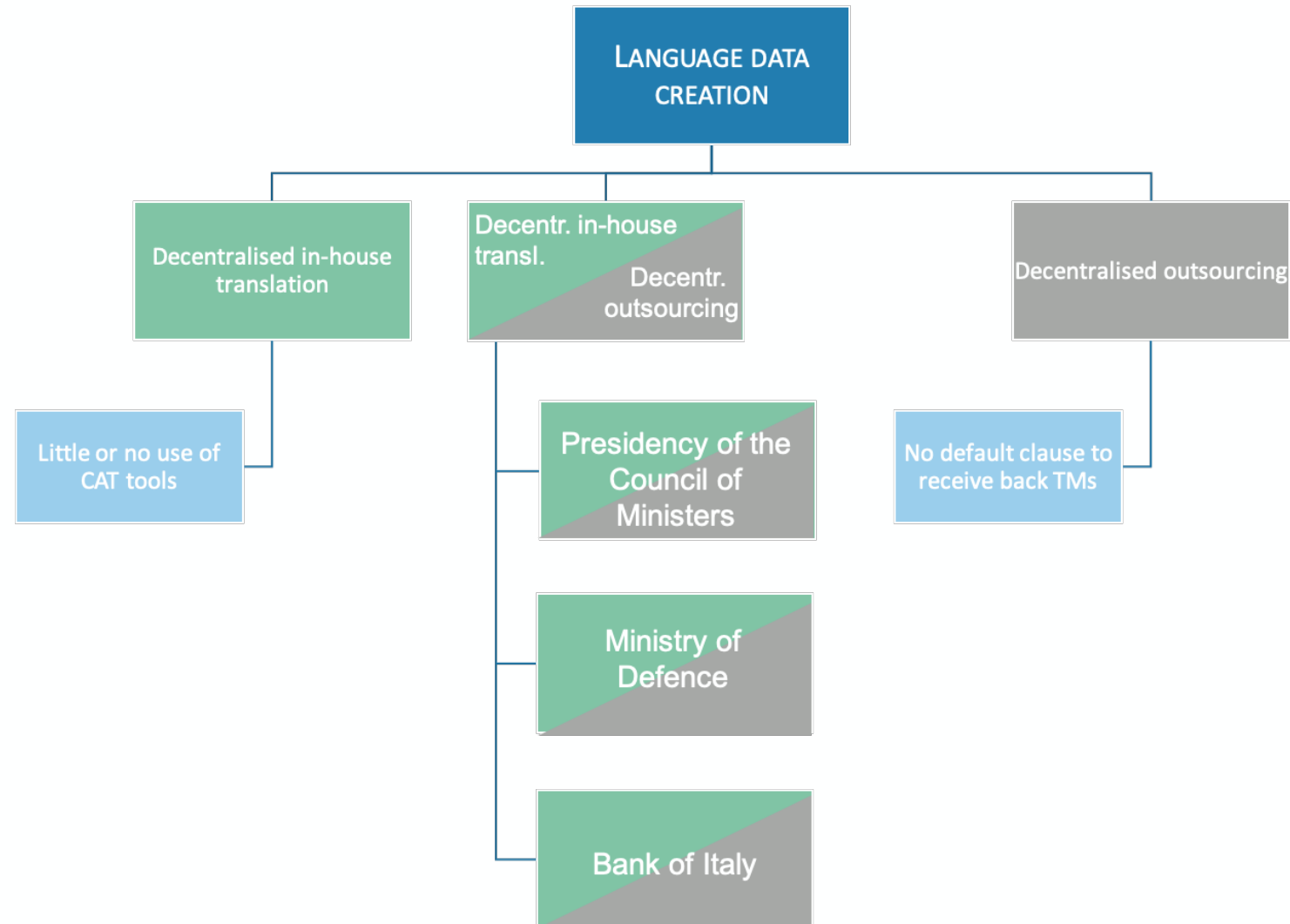
- Only eight countries stated that one or several PA have an MT system integrated into the translation process
- In no country, is it a standard practice to use machine translation in the translation process apart from free online MT services



# INEFFICIENT TRANSLATION AND PROCUREMENT PRACTICES

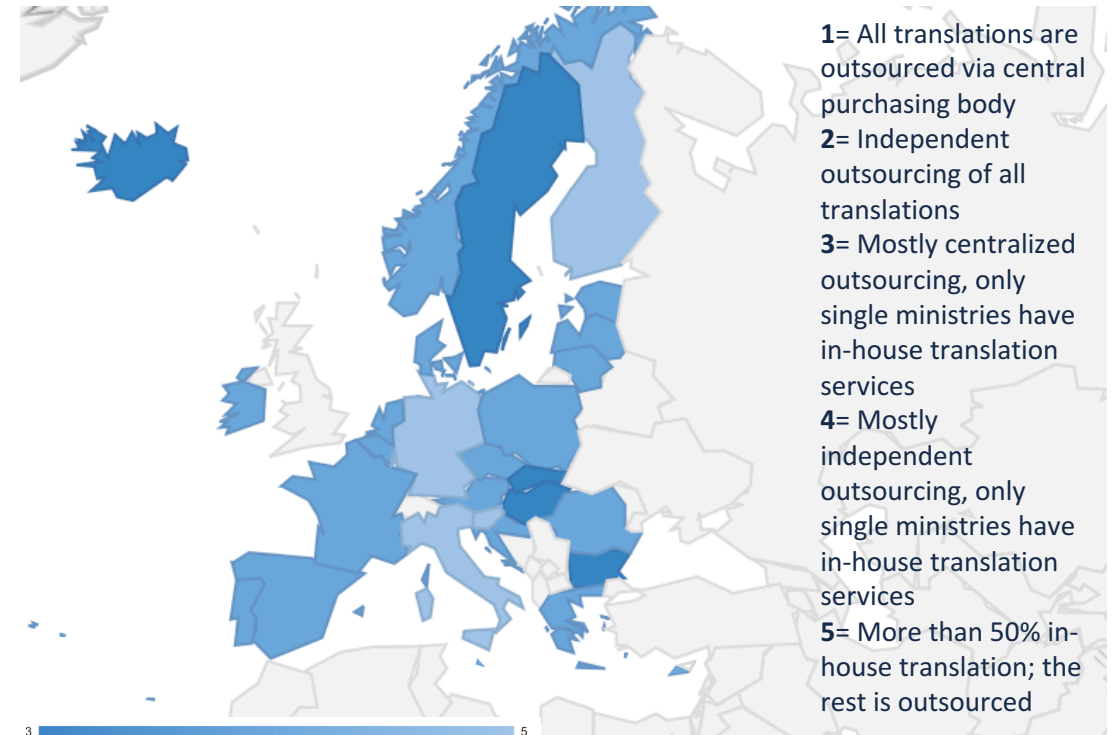


# EXAMPLE: ITALY



# IN-HOUSE TRANSLATION VS. OUTSOURCING

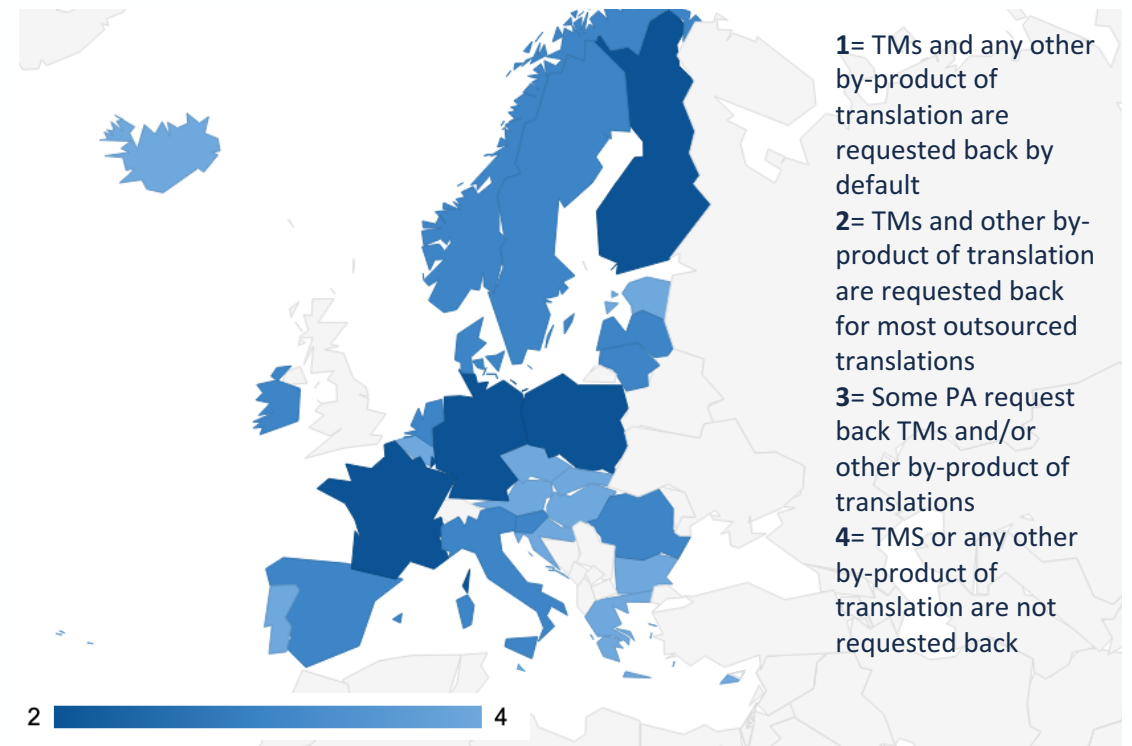
- Only 5 countries translate more than 50% in-house (Finland, Germany, Luxembourg, Italy and Slovenia)
- 24 countries indicated that single ministries have in-house translation services, the rest of the translation services is outsourced independently (19) or through a central purchasing body (5)



3  5  
 The legend (3-5) displays selected options only.

## REQUESTING BACK TRANSLATION MEMORIES FOR OUTSOURCED TRANSLATIONS

- 13 countries indicated that translation memories are not requested back at all
- 11 countries indicated that some public administrations request back translation memories
- Only 5 countries said that translation memories were requested for most translations (Finland, Poland, Germany, Luxembourg and France)
- Not a single country indicated that on the national/federal level Translation Memories are requested back by default



The legend (2-4) displays selected options only.



# LEGAL CONCERNS



## LEGAL CONCERNS

- Little expertise about copyright and Intellectual Property Rights of translations within translation services lead to reluctance to share data
- Concerns about GDPR infringement are used as an argument against sharing language data

# THE ELRC COUNTRY PROFILES

---

The Idea

Process and Main Findings

**Recommendations and White Paper**

# MAIN RECOMMENDATIONS

## At the European and national policy level:

- The Open Data Directive should reference language data as a valuable data category.
- A study on the value of language data identifying and quantifying the value of language data for citizens, public administrations and businesses should be commissioned by the European Commission.
- National policies (e.g. national Open Data policy, digital agenda or strategy for Artificial Intelligence) should be updated to explicitly support the sharing of language data and language technologies.
- National surveys assessing the translation practices in public administrations on all administrative levels should be conducted.
- Inclusion of language data management plans in all relevant national funding policies and calls for proposals should be obligatory (if not yet included).

# MAIN RECOMMENDATIONS

## At the organisational/institutional level:

- Translation and data management practices need to be adapted and improved in order to allow for easy/easier language data sharing in the future, including appropriate licensing of translations.
- Investments in human capital should be made, in particular the provision of technical and legal training for translators and translation managers.
- Investments in IT infrastructures, equipment and tools are necessary, in particular the provision of computer-assisted translation (CAT) tools, machine translation, data anonymization methods and tools, as well as language data management tools.

# ELRC WHITE PAPER

## Sustainable Language Data Sharing to Support Language Equality in Multilingual Europe

### WHY LANGUAGE DATA MATTERS

- Further information about the challenges for language data sharing
- Recommendations addressing each of the challenges
- Country Profiles for each country including information about
  - Translation and language data sharing practices
  - Language and digital policy
  - Main challenges
  - Action plan addressing the identified challenges





# Language Data Matters!



# THANK YOU FOR YOUR ATTENTION!

Website: [www.lr-coordination.eu](http://www.lr-coordination.eu)

Twitter: @LR\_Coordination

Email: [info@lr-coordination.eu](mailto:info@lr-coordination.eu)

