European Language Resource Coordination
Connecting Europe Facility

European Commission

Country Profiles

Language Data Matters?

Lilli Smal (DFKI)

OUTLINE:

- What is ELRC all about?
- Country Profiles
    - The reason
    - The process
    - The goal

Language Data Matters?

What is ELRC all about?

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

Vision: True Digital Single Market

Mission: Create sustainable data pipelines

Purpose: Identify and collect language resources

WHAT DATA ARE WE LOOKING FOR?

- Non-personal public sector information
- “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, p.7.)

WHAT DATA ARE WE LOOKING FOR?

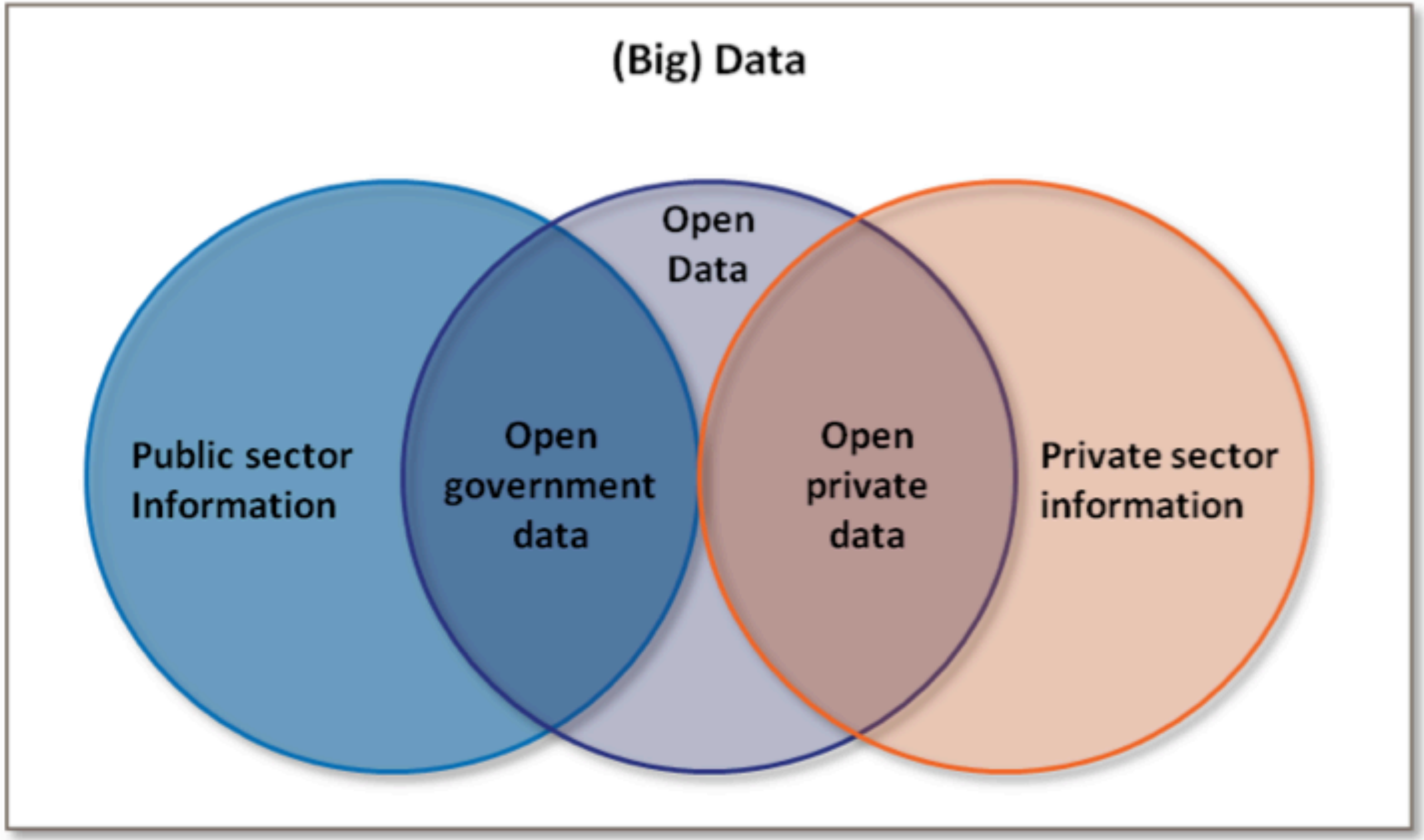
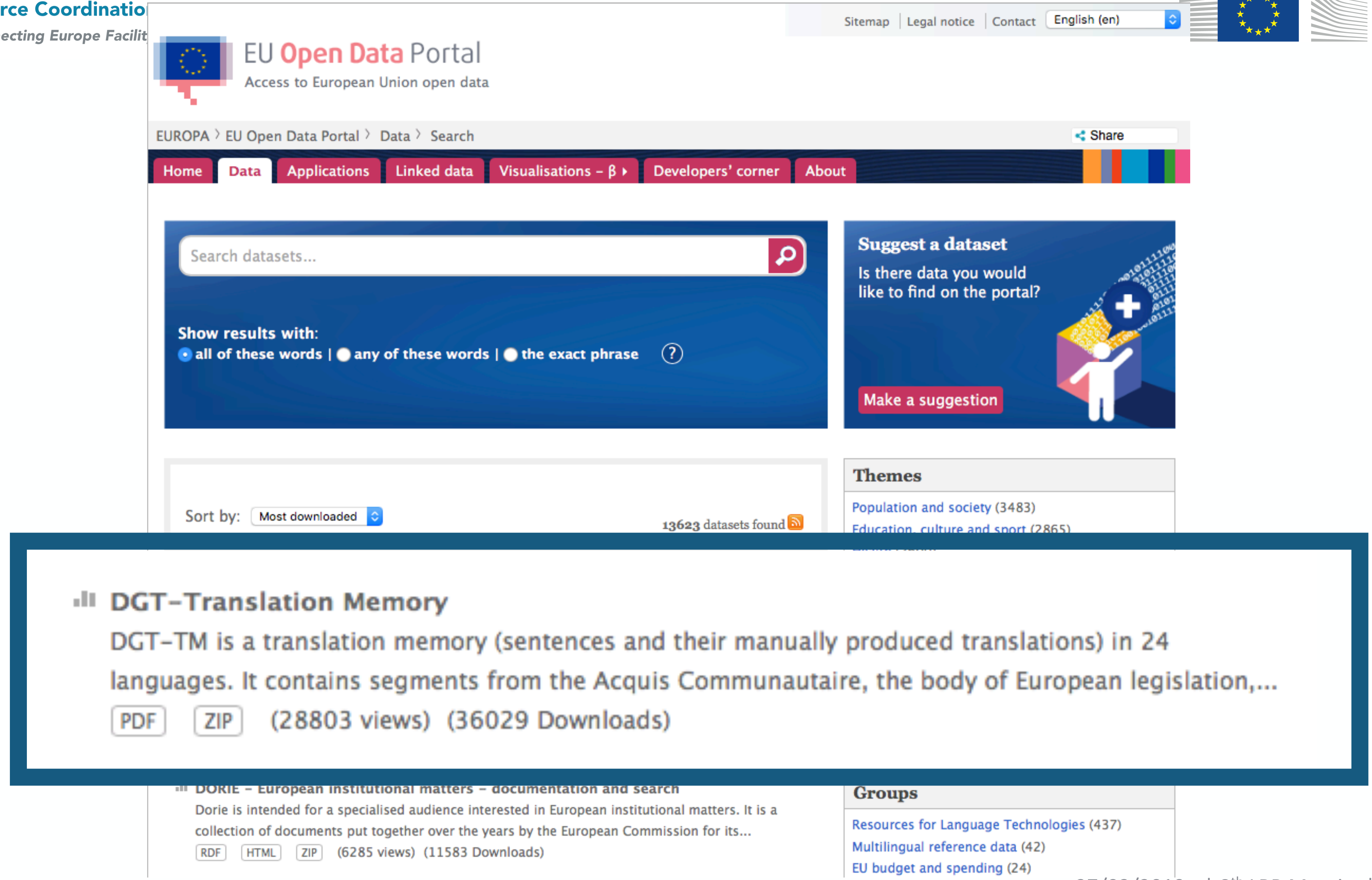


Figure 2 Open Data in the broader data economy

(European Data Portal: Analytical Report 9: The Economic Benefits of Open Data, p.7.)

?

What Data set has the most downloads in the EU open data portal?



Sitemap | Legal notice | Contact | English (en)

EU Open Data Portal
Access to European Union open data

EUROPA › EU Open Data Portal › Data › Search

Share

Home | Data | Applications | Linked data | Visualisations – β | Developers' corner | About

Search datasets...

Show results with:
● all of these words | ○ any of these words | ○ the exact phrase ?

Suggest a dataset
Is there data you would like to find on the portal?
Make a suggestion

Sort by: Most downloaded

13623 datasets found

Themes
Population and society (3483)
Education, culture and sport (2865)

DGT–Translation Memory

DGT–TM is a translation memory (sentences and their manually produced translations) in 24 languages. It contains segments from the Acquis Communautaire, the body of European legislation,...

PDF ZIP (28803 views) (36029 Downloads)

DORIE – European institutional matters – documentation and search

Dorie is intended for a specialised audience interested in European institutional matters. It is a collection of documents put together over the years by the European Commission for its...

RDF HTML ZIP (6285 views) (11583 Downloads)

Groups
Resources for Language Technologies (437)
Multilingual reference data (42)
EU budget and spending (24)

SOME BENEFITS OF OPEN DATA

- Open language data supports the strong language technology research community in Europe
- Governments are one of the main re-users of Open Data themselves (Berners-Lee, Open Data Goldbook for Data Managers and Data Holders, cf. p.15.)
- Allows individuals and businesses to access and engage in cross-border digital public services and trade

COUNTRY PROFILES

The reason

The process

The goal

REASON

- Digital landscape varies a lot in the different countries
- Very difficult to identify decision makers that can authorize data sharing
- There is no “one size fits all” solution for data pipelines and data collection

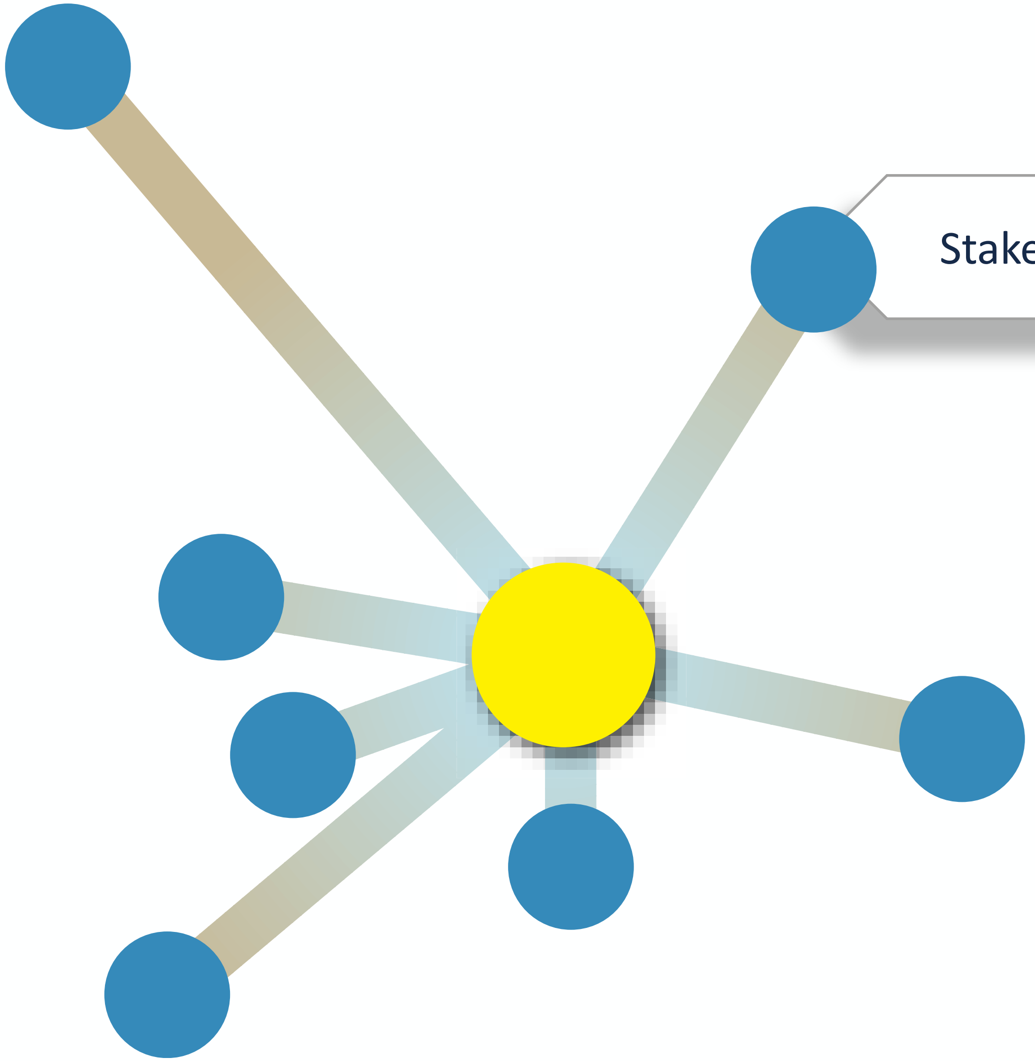
Country Profiles

The reason

The process

The goal

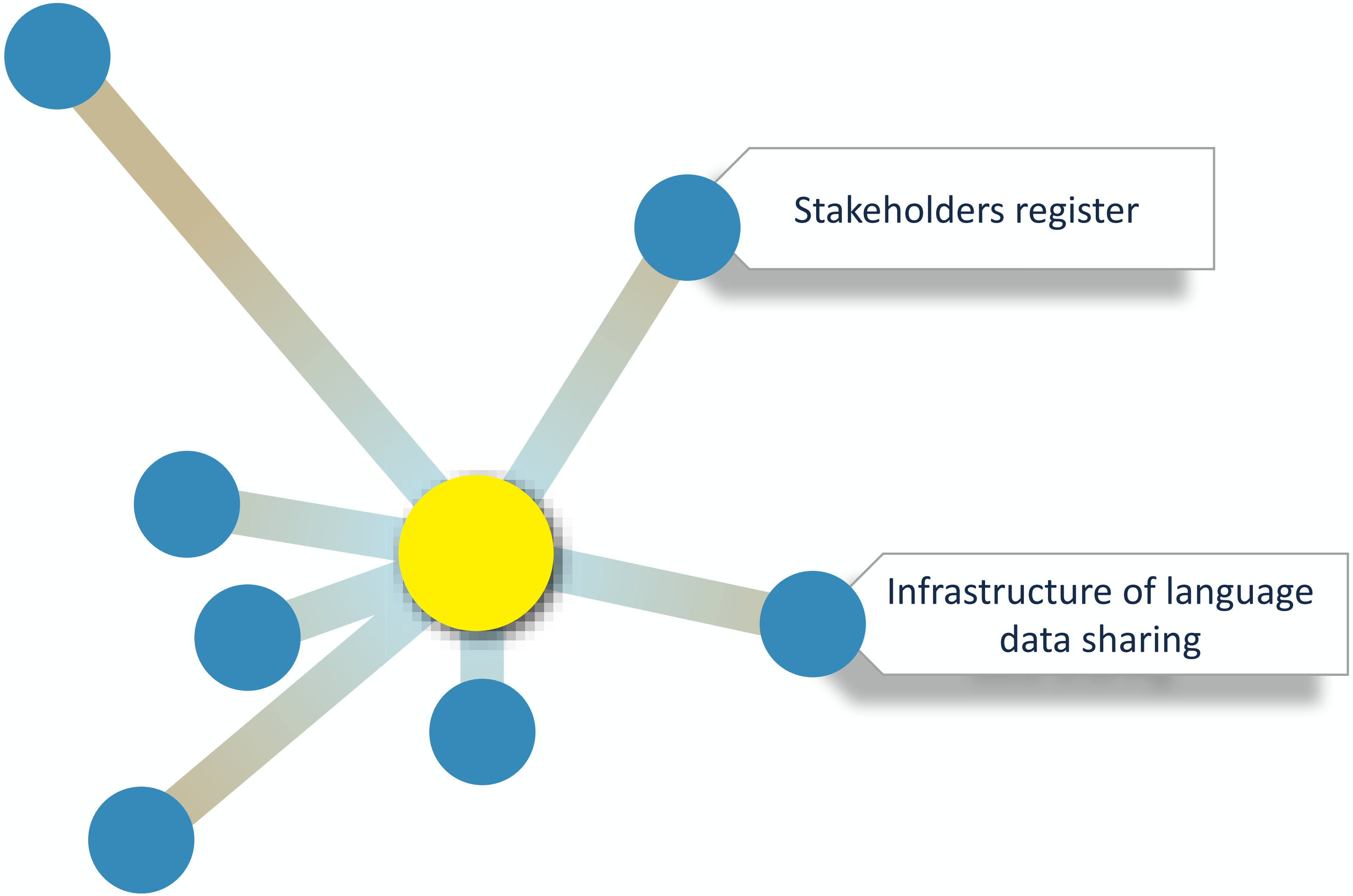
COUNTRY PROFILES



Stakeholders register

STAKEHOLDERS REGISTER

- List of stakeholders including information such as
  - Stakeholder group (e.g. language data creator or holder?)
  - Outsourcing of translations?
  - Stakeholder category (e.g. public body or public online service, research institution, LSP etc.)
  - Stakeholder’s position (who are the key players and decision makers?)
  - etc.



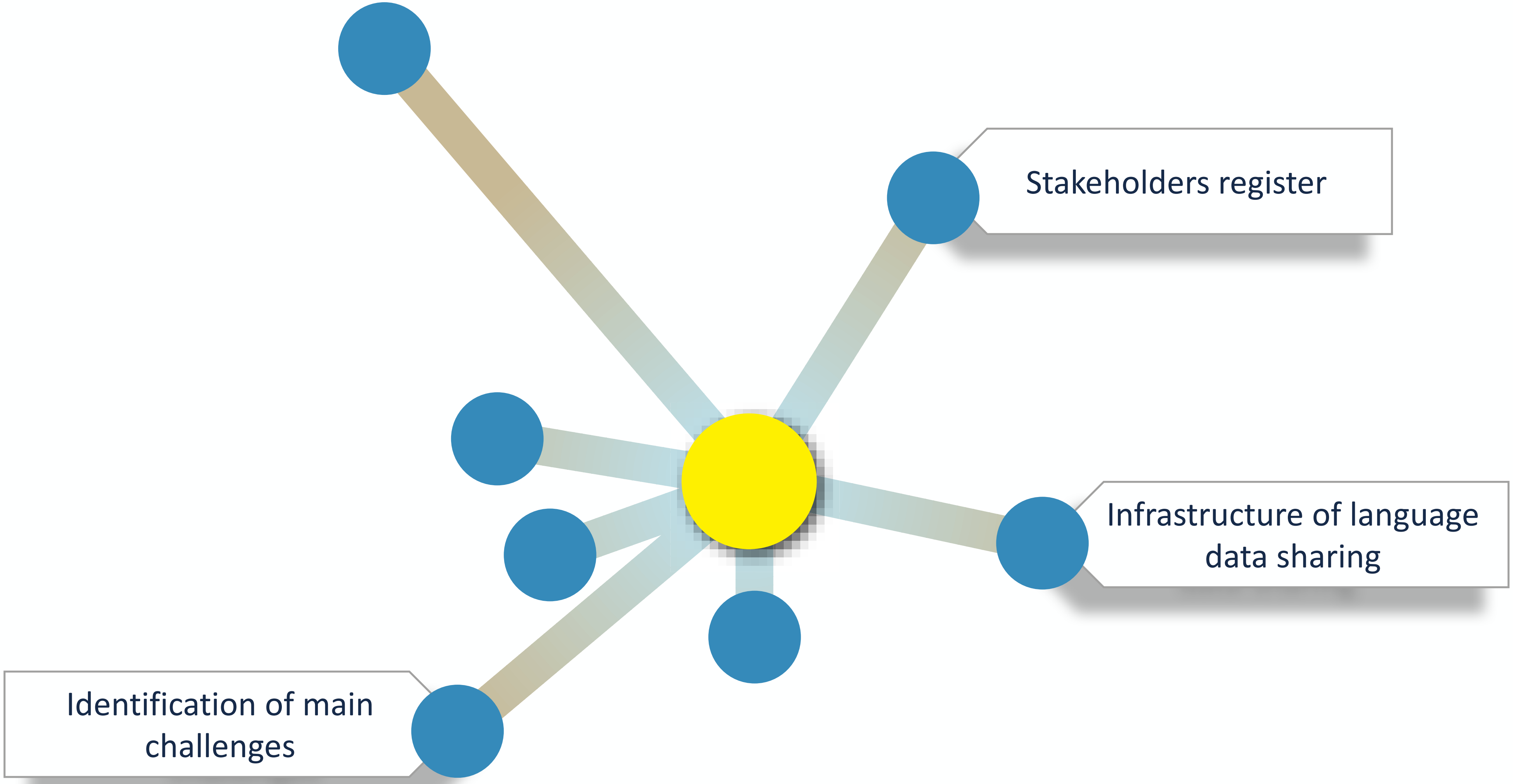
Stakeholders register

Infrastructure of language data sharing

INFRASTRUCTURE OF DATA SHARING

- How are translation needs in the public sector met?
- Who outsources translations and who translates in-house?
- Are these services centralized?
- Do language service providers share their TMs?

COUNTRY PROFILES



Stakeholders register

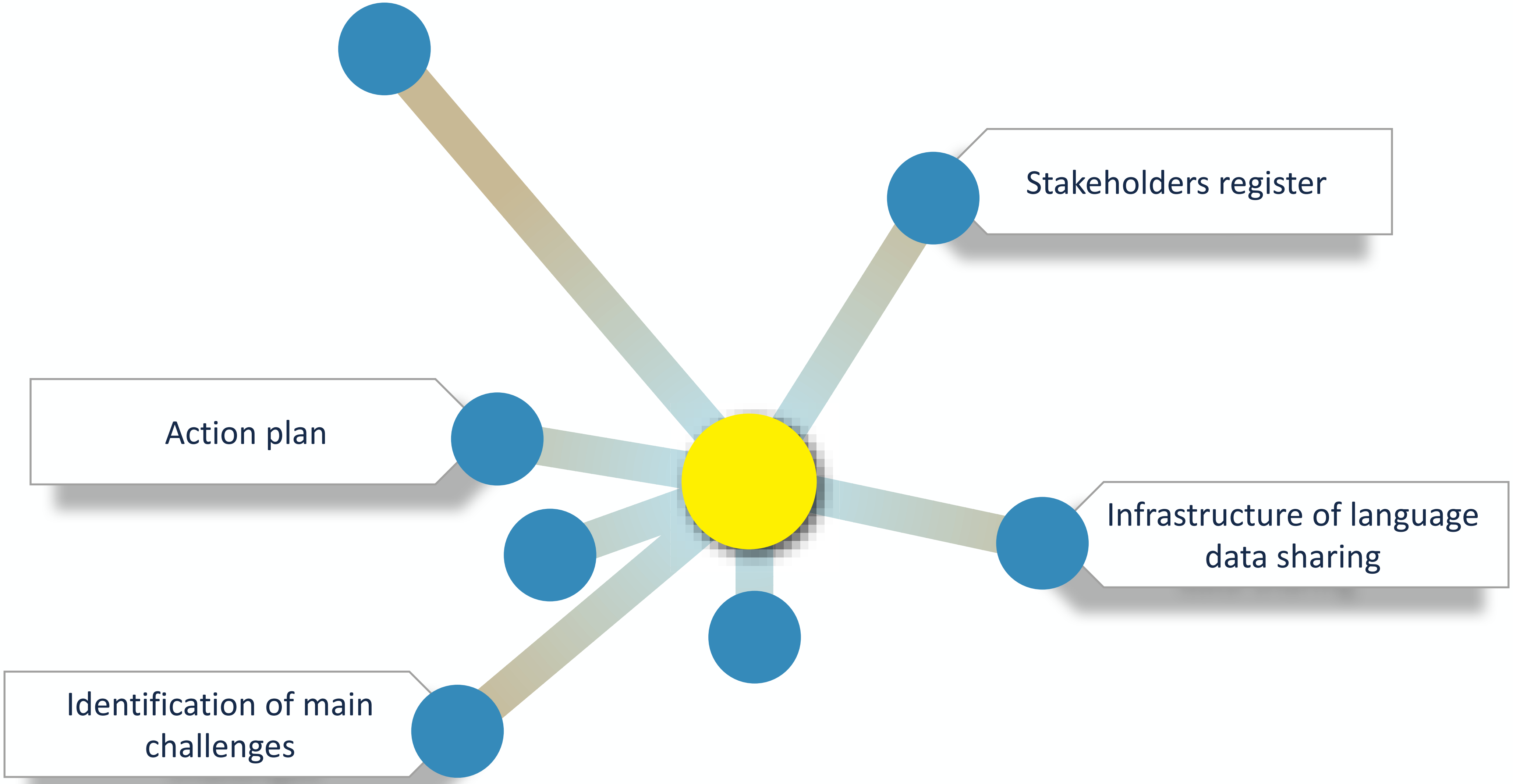
Infrastructure of language data sharing

Identification of main challenges

IDENTIFICATION OF MAIN CHALLENGES

- Language data is not considered valuable
- Little or no coordination of information/data exchange between public services
- Legal concerns (especially personal data)
- TMs are not requested back
- No person in charge of (language) data management

COUNTRY PROFILES



Stakeholders register

Action plan

Infrastructure of language data sharing

Identification of main challenges

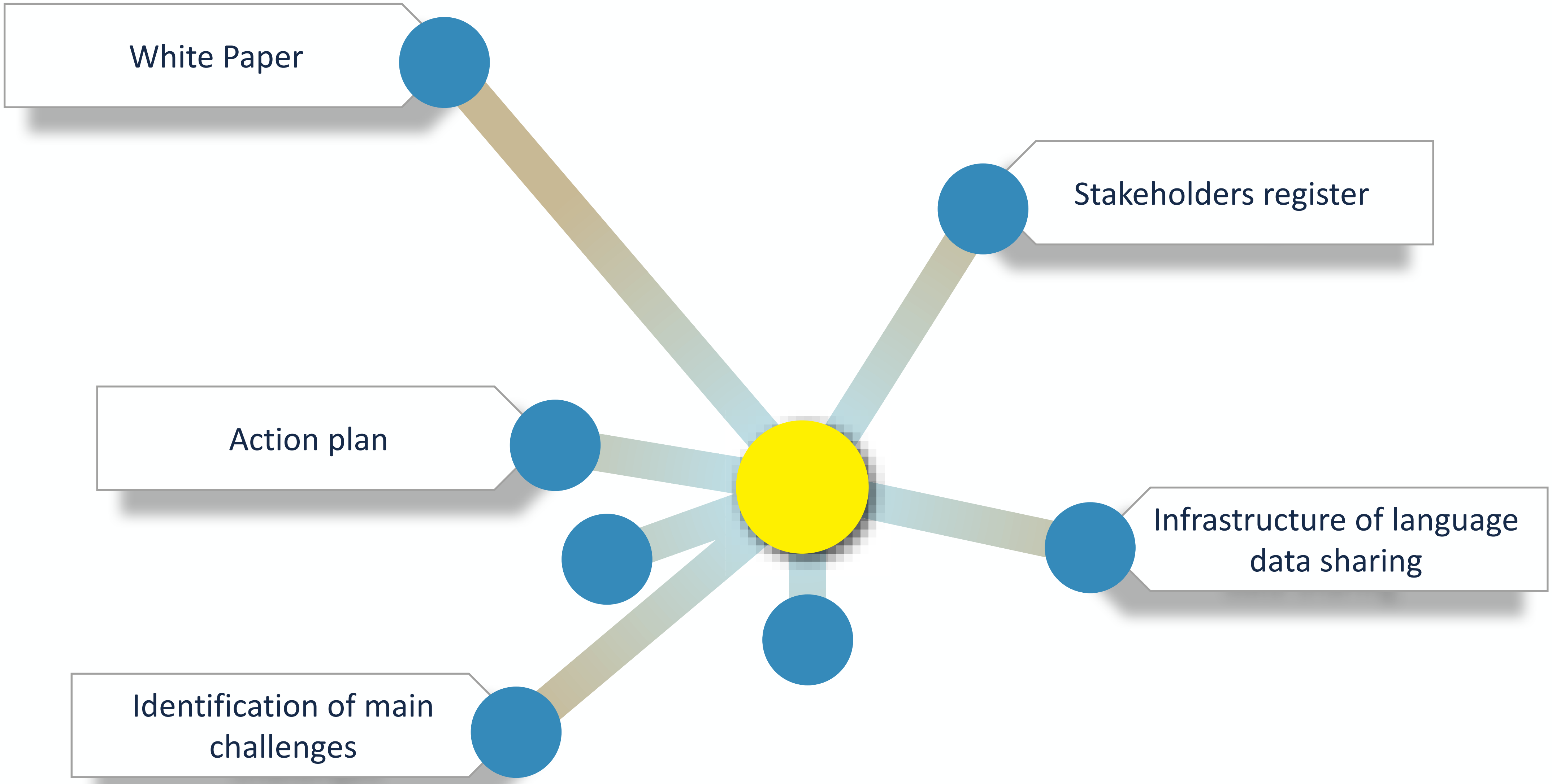
ACTION PLANS

- Objective 1: Raising awareness of language data as open data and a valuable asset
- Objective 2: Increasing interest in MT/LT in public services as part of the national digital policy
- Objective 3: Tackle legal/ethical issues and concerns
- Objective 4: Identify and gain access to outsourced translations
- Objective 5: Establish good data management practices in public services

ACTION PLANS

- The ELRC consortium answered as many questions as possible and published the responses on the ELRC website: http://helpdesk.lr-coordination.eu/cef_dashboard/
- Discussion of three main challenges in working groups

COUNTRY PROFILES



White Paper

Stakeholders register

Action plan

Infrastructure of language data sharing

Identification of main challenges

WHITE PAPER

- Publication of results stressing the importance and value of language date
- All National Anchor Points as co-authors
- To be presented at ELRC Conference in Helsinki in November
- Laying the groundwork for further collaboration and application for CEF funded projects

Country Profiles

The reason

The process

The goal

Language Data Matters!

Languages are the **HEART** of Europe

THANK YOU FOR YOUR ATTENTION!

Website: www.lr-coordination.eu

Twitter: @LR_Coordination

Email: info@lr-coordination.eu/
lilli.smal@dfki.de

REFERENCES:

- **Halevy, Alon et al.**: *The Unreasonable Effectiveness of Data*, https://static.googleusercontent.com/media/research.google.com/de//pubs/archive/35179.pdf, 2009.
- **European Data Portal**: *Open Data Goldbook for Data Managers and Data Holders*, https://www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf, 2018.
- **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, 2017.
- **European Parliament**: *Vote on Language Equality in the Digital Age*, http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2018-0228+0+DOC+XML+V0//EN&language=en, 2018.
- **van der Meer, Jaap et al.**: *Capitalizing on Translation Data*, https://www.taus.net/think-tank/reports/translate-reports/taus-data-market-white-paper, 2017.
- **Massardo, Isabella; van der Meer, Jaap**: *The Translation Industry in 2022*. https://info.taus.net/translation-industry-2022-report-download, 2017.