

cessda eric

Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Multilinguality and Machine Translation Tools & CESSDA

SERISS Seminar “Establishing a roadmap for the use of computational linguistics’ tools and translation technologies in survey research”

Universitat Pompeu Fabra, Barcelona, March 14 & 15, 2018

Mari Kleemola

**Development Manager, Finnish Social Science Data Archive
CESSDA Tools & Services Working Group Leader**



Contents

- What is CESSDA
- Multilinguality within CESSDA SPs
- A User-driven perspective
- CESSDA & Tools

What is CESSDA?

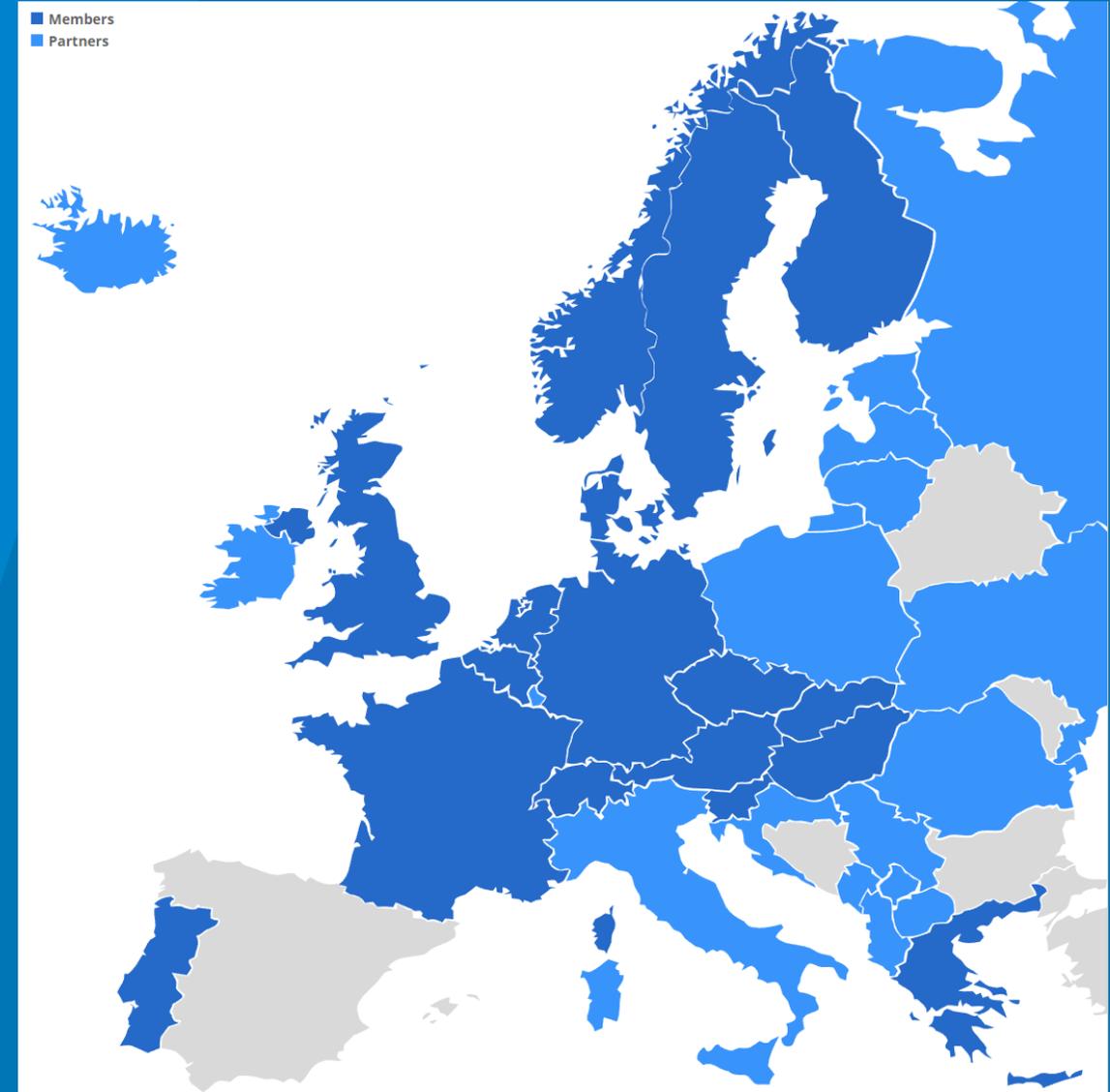
- » Cross-European resource discovery
- » Improved quality of data and metadata
- » A wider selection of comparable data
- » Certification of data archiving organisations
- » Professional training for data archivists and scientific community
- » Improved mechanisms for data dissemination and analysis
- » Strong involvement of organisations outside Europe

- » <https://www.cessda.eu/>

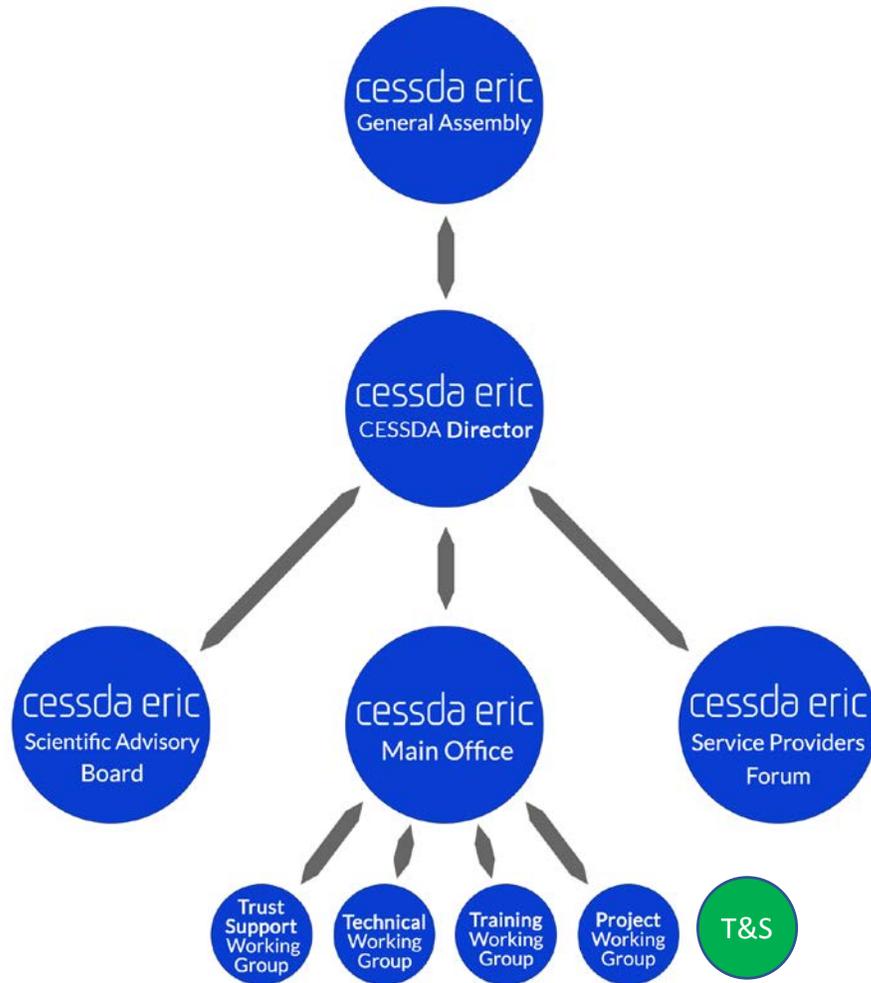
The mission of CESSDA is to provide a full scale sustainable research infrastructure that enables the research community to conduct high-quality research in the social sciences which can contribute to effective solutions to the major challenges facing society today.

MEMBERS

- » Austria
- » Belgium
- » Czech Republic
- » Denmark
- » France
- » Finland
- » Germany
- » Greece
- » Hungary
- » Netherlands
- » Norway
- » Portugal
- » Slovakia
- » Slovenia
- » Sweden
- » Switzerland
- » UK



CESSDA ERIC Governance



Bodies of CESSDA ERIC

- » General Assembly
- » Service Providers' Forum
- » Scientific Advisory Board
- » Working Groups

Multilinguality within CESSDA SPs

- » Focus on re-use and long-term preservation
- » National and international studies (data): surveys, qualitative textual data,...
- » 17 member countries > data and other material in 17+ languages
- » Questionnaires in several languages (pdf files)
- » Metadata on study level usually in English and in national language(s), variable level metadata in national language and/or English (DDI C, DDI L)
- » Data usually available in national language and/or English

CESSDA Multilinguality Tools

- » Metadata portfolio CMM1 supports multilinguality, based on DDI L
- » CMM1 includes multilingual Controlled Vocabularies (CESSDA Topic Classification and selected DDI CVs)
- » The European Language Social Science Thesaurus (ELSST)
- » In development: CV Manager (Sep 2018), EuroQuestionBank
- » CESSDA SPs use a variety of tools for ingest and management of data collections, machine translation has not been in focus
- » For example, in FSD we use WordFast translation memory software to translate survey data labels (questions) from Finnish to English
- » Tools developed in projects like DASISH and SERISS are of interest to CESSDA

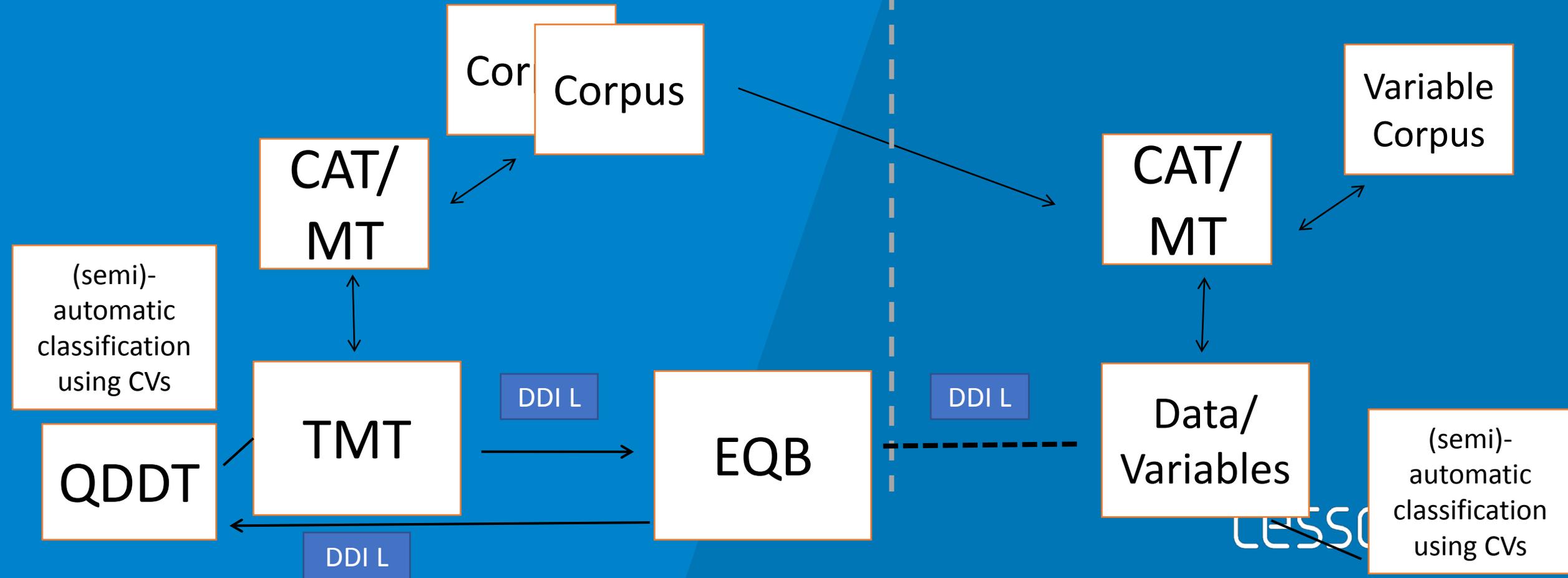
A User-driven Perspective (ideas)

cross-national comparative surveys by design

English > national languages

archived national surveys

> English



CESSDA Software Adoption Policy

- » Work in progress
- » How 3rd party software could make its way on to the CESSDA Technical Research Infrastructure
- » CESSDA is willing in principle to adopt, maintain, and make tools available for the public good within the CESSDA Research Infrastructure, provided that they conform to certain conditions and requirements

For example, must support one or more of the CESSDA ERIC Objectives, with a preference for facilitating the sharing and re-use of data and metadata; and should be usable by different kinds of projects and user types, and not be specific to one particular project.

Thank you!

cessda eric