Benefits, challenges and solutions for multilingual digital cultural heritage: a report from the Finnish Presidency Europeana meeting

February 2020

Europeana meeting under Finland's Presidency of EU Council - Espoo, Finland, 24-25 October 2019
Sebastiaan Ter Burg - CC BY

Content

Executive summary

During the Finnish Presidency of the Council of the European Union, Europeana and the Finnish Ministry of Education and Culture jointly held a meeting with high-level policymakers and experts in multilingualism to discuss the challenges and opportunities of multilingualism in digital cultural heritage. The event titled 'Multilingualism in Digital Cultural Heritage - needs, expectations and ways forward' was held on 24-25 October 2019 at the Hanaholmen Culture Centre, Espoo, Finland.

At the meeting, delegates from the DCHE Expert Group, Ministries of Culture across Europe, and multilingualism experts were invited to share their perspectives and experiences with each other through a series of speeches, presentations, case studies, panel discussions and workshops. Their aims were to:

- stimulate reflection on multilingualism in digital cultural heritage at large using Europeana as a case study
- develop a deeper understanding of the multilingualism problem/opportunity space for digital cultural heritage
- consider what options can be pursued to provoke action at the local level, furthering the multilingual capabilities
- provide input and feedback for the Europeana multilingual strategy.

Finland's Ministry of Culture and Education State Secretary Tuomo Puumala emphasised the role of culture and heritage in creating a democratic and inclusive Europe.

The meeting focussed on multilingual metadata, content translation and user interactions. Some of the presentations and speeches included:

- Juliane Stiller - The status quo, the problem space, and historical information on multilingual development
- Rickard Domeij - Multilingualism, technology and policy
- Andrejs Vasiļjevs and Jānis Ziediņš - Learnings from the automatic translation projects and how to apply them for the culture and heritage sector
- Heli Kautonen - UX design
- Dasha Moskalenko - UX design and user testing
- Matias Frosterus - Multilingual linked vocabularies and automatic subject indexing services
- Hugo Manguinhas - Europeana case study
- Anne Bajart - Multilingualism in Digital Cultural Heritage
- Tom Vanallemeersch - eTranslation

Dasha Moskalenko, Hugo Manguinhas and Antoine Isaac presented Europeana case studies on UX design, multilingual knowledge graphs and automatic translation.

Finally, a panel led by Andy Neale discussed conclusions and steps for progress.

See the full programme in the [annexes](#).

The principal opportunities and challenges for various players were identified. Benefits of multilingualism were discussed. And together, delegates developed a broad set of solutions and changes needed to address the identified challenges and advance multilingualism.

This report summarises the key outcomes of the meeting including the benefits, challenges and solutions identified in relation to multilingualism in digital cultural heritage.

Europeana is developing a recommended multilingual strategy and roadmap paper that builds on the output from the meeting in Finland. The Europeana multilingual strategy is scheduled for publication in spring 2020.

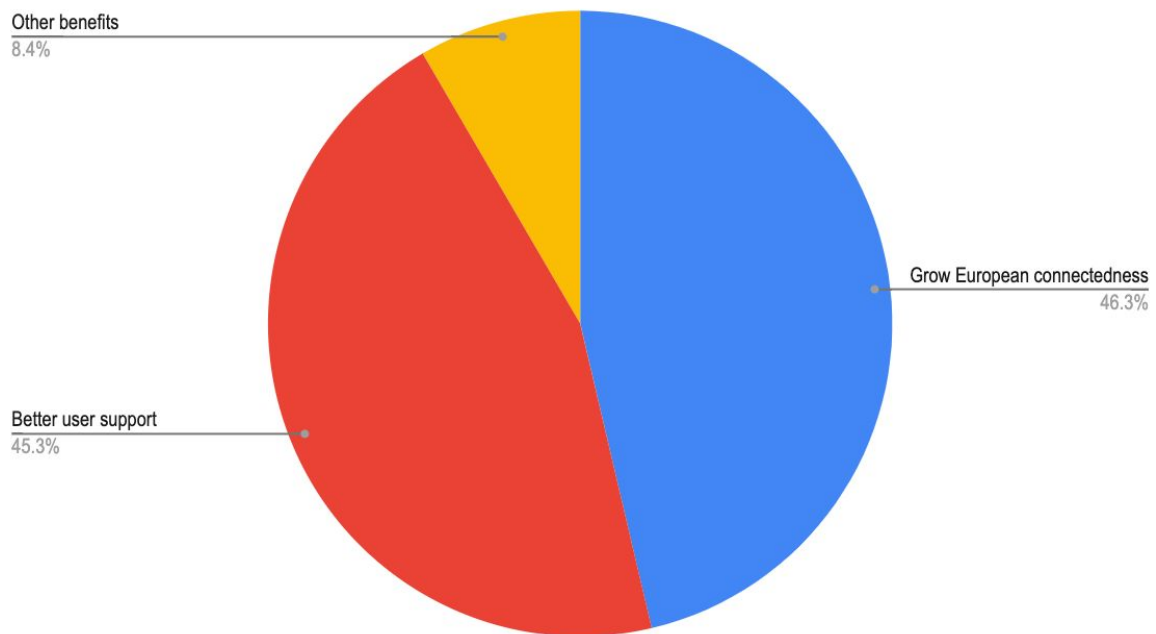# Benefits of multilingual digital cultural heritage

Participants were asked to work in small groups to first identify, and then size, the primary benefits of multilingual digital cultural heritage. In this case, sizing refers to the average proportion of value assigned by the different teams.

The biggest benefit was noted as being that *people can access the information of other cultures and language groups*. Followed by the point that multilingualism *promotes socially inclusive societies and mutual understanding of different cultures*.

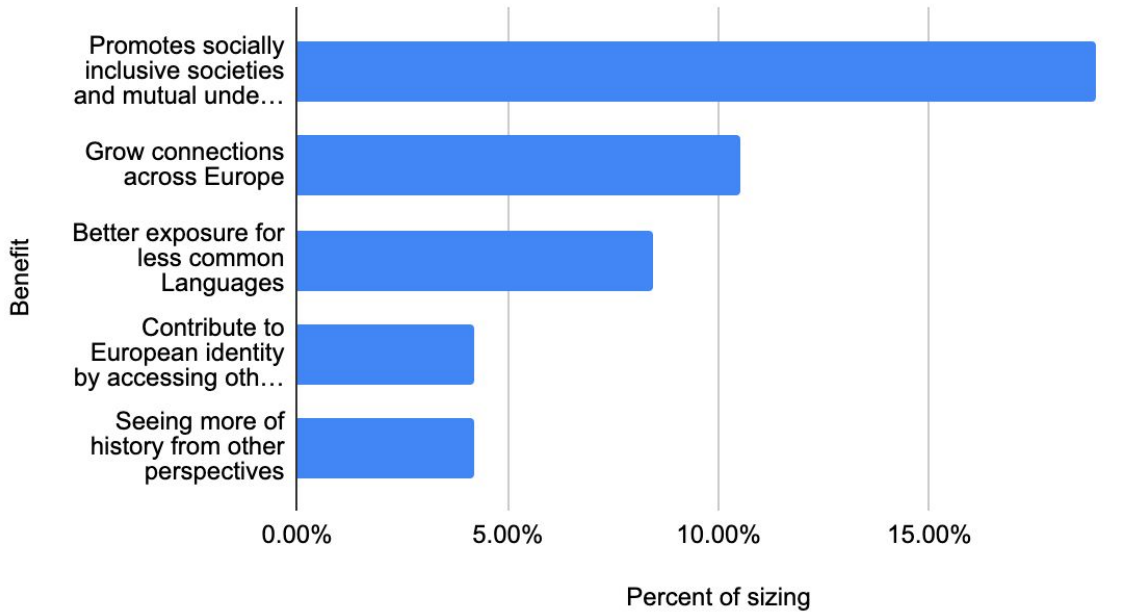| Benefits of multilingual digital cultural heritage	Sizing
People can access the information of other cultures and language groups	28.42%
Promotes socially inclusive societies and mutual understanding of different cultures	18.95%
More sources of material to search across	11.58%
Grows connections across Europe	10.53%
Better exposure for less common languages	8.42%
Increased usability of digital cultural heritage materials in different sectors (e.g. education, research)	5.26%
Fosters and enriches creativity	5.26%
Contributes to European identity by accessing other cultures and languages	4.21%
Seeing more of history from other perspectives	4.21%
Collections attract more visitors	2.11%
Supports learning new languages	1.05%

Summary of all benefits

These benefits can be grouped into two primary types of benefits, they can *grow European connectedness* (referring to the feeling of connection between European citizens) and *provide better user support*.

When looking at the size of these groups, a more balanced view emerges, showing that the benefits are split evenly between individuals and the state.

Benefits that contribute to growing European connectness

Benefit

- Promotes socially inclusive societies and mutual unde…
- Grow connections across Europe
- Better exposure for less common Languages
- Contribute to European identity by accessing oth…
- Seeing more of history from other perspectives

Percent of sizing (0.00%, 5.00%, 10.00%, 15.00%)

Benefits that contribute to better user support

Benefit

- People can access the information of other cultures and language groups
- More sources of material to search across
- Increased usability of digital cultural heritage materials in different sectors (e.g. education, research)

Percent of sizing (0.00%, 10.00%, 20.00%)
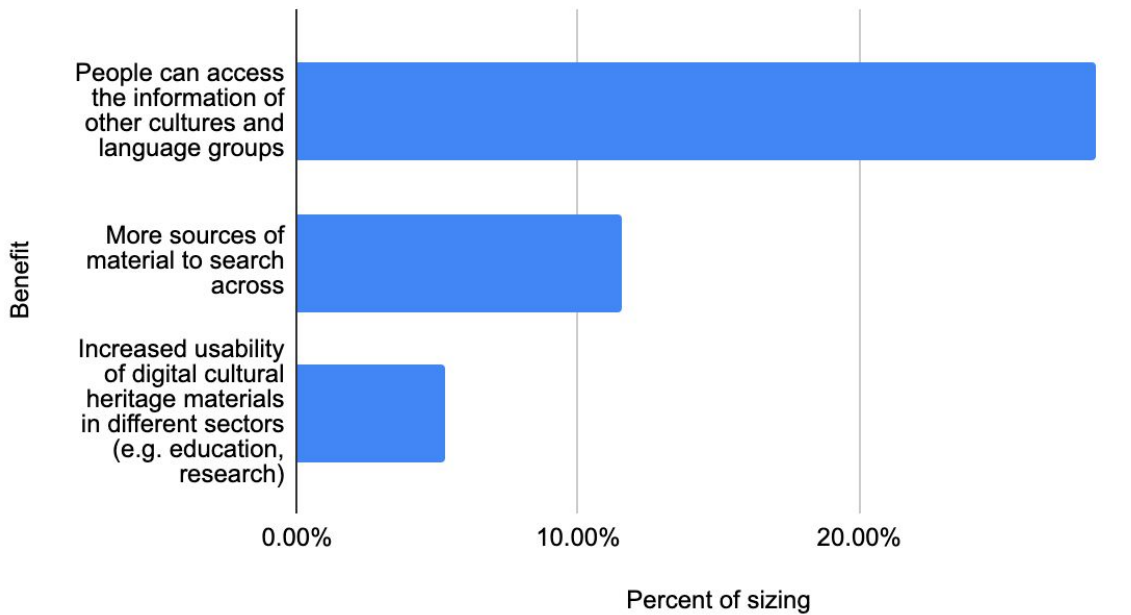
Other benefits
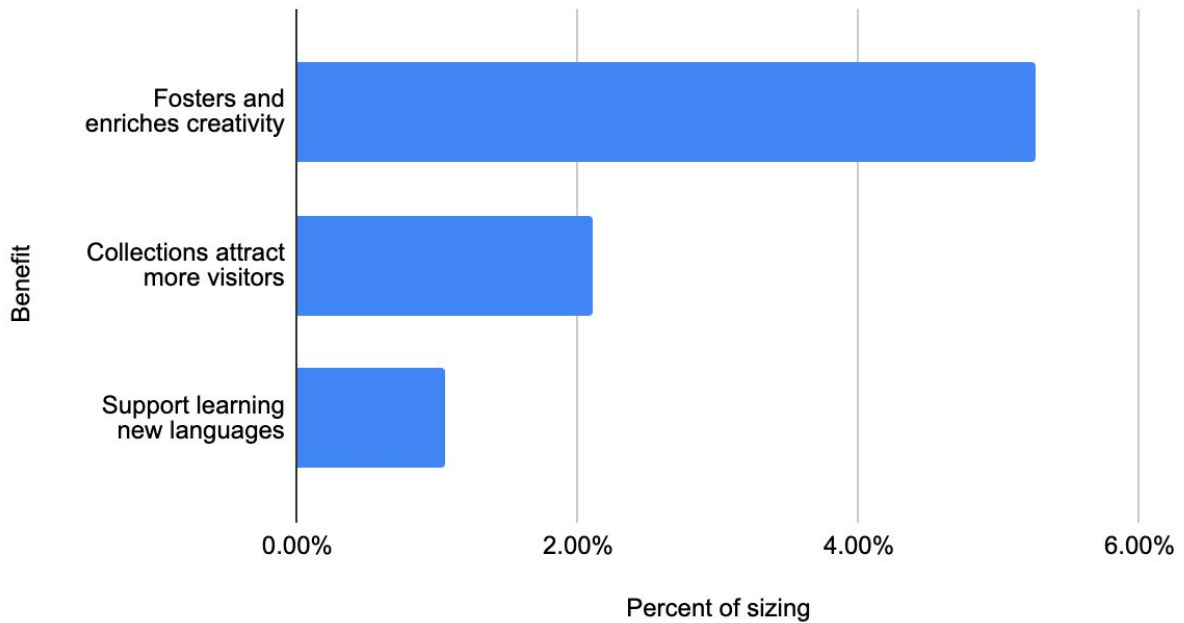
Issues that limit the delivery of multilingual digital cultural heritage

Participants were then asked to consider the issues that prevent them from fully delivering the benefits of multilingual digital cultural heritage. In small groups they also considered whether these issues were technology or capability constraints. Capability in this case encompasses skills, knowledge, or funding.

Capabilities available

Translation option not available to user
(Europeana interface)

Lack of attractiveness

Awareness/ promotion/ dissemination

All metadata
understandable
(translation & enrichment)

Lack of unified language/
translation policy

Lack of understanding inside the
instituuion about the necessity of
creating/providing a
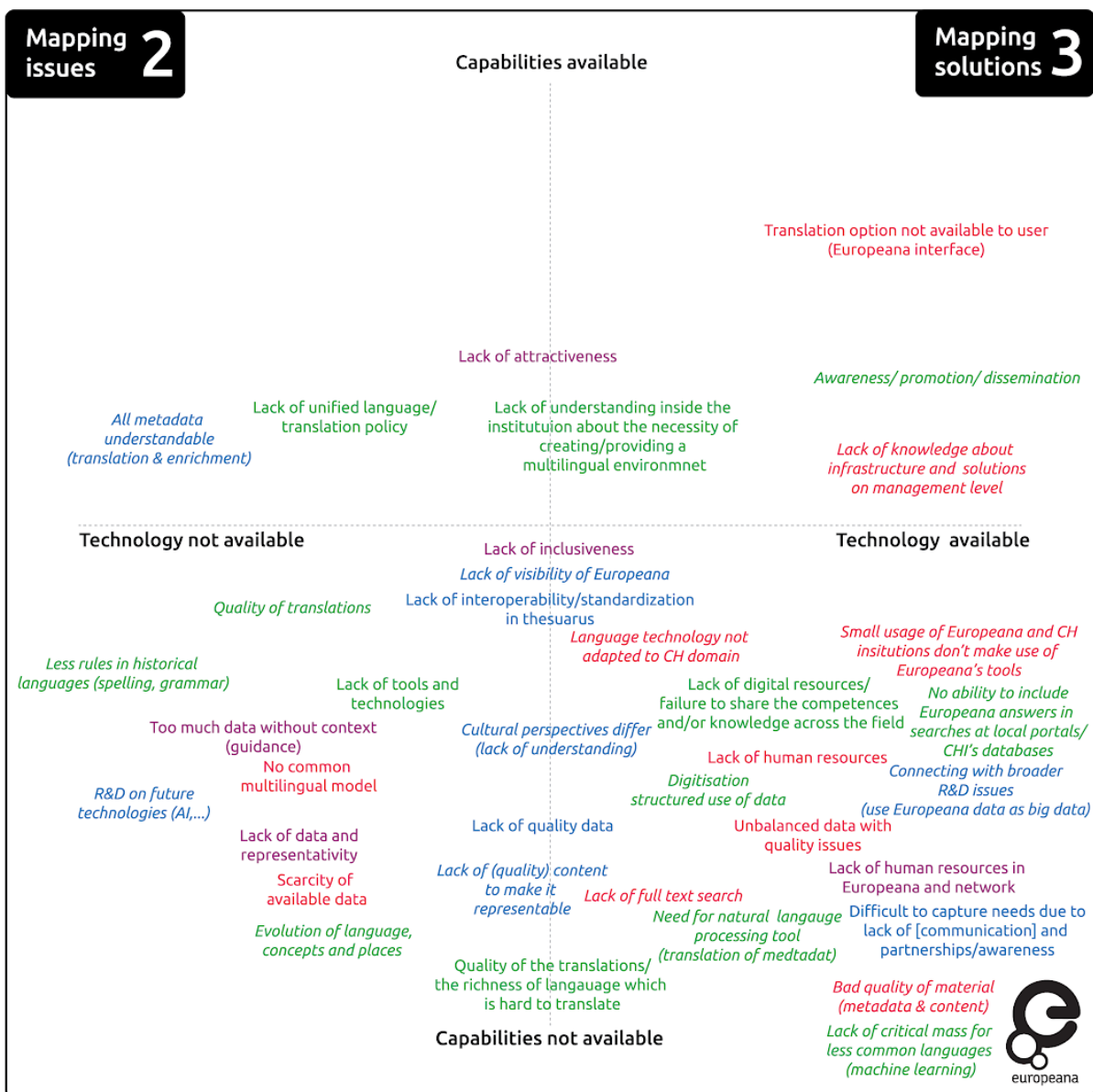multilingual environmnet

*Lack of knowledge about
infrastructure and solutions
on management level*

Technology not available

Lack of inclusiveness

Lack of visibility of Europeana

Quality of translations

Lack of interoperability/standardization
in thesuarus

Technology available

*Language technology not
adapted to CH domain*

*Small usage of Europeana and CH
insitutions don't make use of
Europeana's tools*

*Less rules in historical
languages (spelling, grammar)*

Lack of tools and
technologies

Lack of digital resources/
failure to share the competences
and/or knowledge across the field

*No ability to include
Europeana answers in
searches at local portals/
CHI's databases*

Too much data without context
(guidance)

*Cultural perspectives differ
(lack of understanding)*

Lack of human resources

*Connecting with broader
R&D issues
(use Europeana data as big data)*

*R&D on future
technologies (AI,...)*

No common
multilingual model

*Digitisation
structured use of data*

Lack of data and
representativity

Lack of quality data

Unbalanced data with
quality issues

*Lack of (quality) content
to make it
representable*

Scarcity of
available data

Lack of full text search

Lack of human resources in
Europeana and network

*Evolution of language,
concepts and places*

*Need for natural langauge
processing tool
(translation of medtadat)*

Difficult to capture needs due to
lack of [communication] and
partnerships/awareness

Quality of the translations/
the richness of langauage which
is hard to translate

*Bad quality of material
(metadata & content)*

*Lack of critical mass for
less common languages
(machine learning)*

Capabilities not available

europeana

Capability gaps for issues that have technology solutions (lower-right quadrant)

In summary, we can see a concentration of issues in the lower-right quadrant. This equates to issues for which participants believe the technology is available, but for which the capability in cultural heritage institutions and Member States is the constraining factor.

This quadrant cites issues with:

- Data quality and content to support translations
- Standardisation (and adoption by cultural heritage domain)
- Use of tools that already exist
- Human resources

Gaps where we have both the capability and the technology (upper-right quadrant)

In the upper-right quadrant we can see a small number of issues for which it is proposed that we have both the capability and the technology.

This quadrant cites issues with:

- Understanding within institutions and Member States as to the necessity of multilingual environments (and their attractiveness)
- Awareness and visibility of the tools and options available
- Managers' knowledge of solutions
- User interface options not being available

Technology gaps (left-side quadrant)

When looking at the issues for which technology is seen as the limiting factor, it's the quality and nature of cultural heritage data that are identified as the main blockers.

These quadrants cite issues with:

- Scarcity and representativeness of data
- Historical language without rules and context
- Evolution through time
- Lack of unified multilingual policies and models

Solutions to explore in the resolution of multilingual issues

As a final exercise, participants were asked to explore different solutions to the issues they had identified. This report groups these solution ideas into six areas:

1. Data issues
2. Translation quality
3. Human resource
4. Awareness and visibility of tools and options
5. Multilingual policies and models
6. User experience

Solutions: Data issues

- Better ingestion of metadata from aggregators to Europeana
- Better curation and crowdsourcing of data for training, including the involvement of minority language speakers
- Better use of thesauri
- Investigate automated data cleaning tools

Solutions: Translation quality

- Launch a mass translation effort in the cultural domain (that will also deliver better machine translation outcomes)
- Experiment with training on other non-cultural heritage data, including other public domain source texts
- Use proximity (related languages) to generate new translations
- Provide tools for speech-to-text conversion of audio recordings
- Test existing translation tools further
- Better post-correction OCR for less common languages

Solutions: Human resource

- Develop programmes to build more multilingual expertise at the local level
- Support aggregators in their collection of more multilingual content
- Advocate for more Member State funding

Solutions: Awareness and visibility of tools and options

- Raise awareness about the benefits of good quality content for translation
- Take action to encourage cooperation at national and local levels, including across institution types such as libraries, museums and language institutions
- Use communication and dissemination tools to increase visibility of Europeana and its projects
- Better strategic thinking in the multilingual space
- Cooperate with technology industry to disseminate cheaper digitisation solutions

Solutions: Multilingual policies and models

- Identify minimum standards for working in multilingual space
- Provide frameworks for cataloguing practices (that aim to grow multilingual coverage)

Solutions: User experience
- Conduct surveys and market research to better understand expectations
- Provide alternatives and indicators about correctness of translations to users
- Keep simplicity as a design principle

Authors

Andy Neale, Technical Director, Europeana Foundation
Antoine Isaac, R&D Manager, Europeana Foundation
Dasha Moskalenko, Service Design Manager, Europeana Foundation
Hugo Manguinhas, API Product Manager

Annexes

- **Other documents and communications**

  Post event update on Europeana Pro '*Next steps for multilingualism in digital cultural heritage*'
  https://pro.europeana.eu/post/next-steps-for-multilingualism-in-digital-cultural-heritage

  About the event on Europeana Pro
  https://pro.europeana.eu/event/multilingualism-in-digital-cultural-heritage-and-europeana

  Presentations from the meeting
  https://www.slideshare.net/Europeana/tag/finnish-presidency

- **Technical discussion paper for extending multilingual support in Europeana**
  https://docs.google.com/document/d/1frFjKGW-IXVxuFXlCCgLWCSCgPual0WcPsFiHA67978/edit#heading=h.80i9y9ppewgq

- **Event programme and list of participants**
  https://pro.europeana.eu/files/Europeana_Professional/Event_documentation/Events/Presidency/Finnish-Presidency-Programme.pdf