europeana
network association

Europeana Network Association
Task Forces

Interim Analysis of EuropeanaTech AI in Relation to GLAMs Survey

Task Force Members

Gregory Markus, Netherlands Institute for Sound and Vision (co-chair)
Georgia Angelaki, National Documentation Centre, Greece (co-chair)
Clemens Neudecker, Berlin State Library (co-chair)
Antoine Isaac, Europeana Foundation
Giles Bergel, Oxford University
Werner Bailer, Joanneum Research
Mónica Marrero, Europeana Foundation
Vassilis Tzouvaras, National Technical University of Athens
Johan Oomen, Netherlands Institute for Sound and Vision
Philo van Kemenade, Netherlands Institute for Sound and Vision
Mirjam Cuper, National Library of the Netherlands
Stephan Bartholmei, German Digital Library
David Haskiya, Wikimedia Sweden
José Eduardo Cejudo, Europeana Foundation
Albin Larsson, Europeana Foundation

Introduction

Objectives

This document is an interim report, reporting on the survey by the [EuropeanaTech AI for GLAMs task force](). The purpose of this task force is to do a horizon scanning exercise and to start investigating the expected role and impact of artificial intelligence (AI) in the digital cultural heritage domain especially with regards to collections analysis and improvement.

The aim of this survey was to gain an understanding of who is already working with AI or has plans to do so, the different types of projects being run, the methodologies being used, the challenges faced, the success granted and the resources applied.

The target respondents from this survey were professionals working in museums, libraries, archives, and research institutions as well as the wider industry (technology suppliers, creative industries, etc.) that work with cultural heritage data. A full task force report will be published in early 2021.

For the meantime, this interim report presents our initial takeaways from the survey, organized by survey section. These initial findings will be used to fuel further work during which the task force will more deeply investigate trends and outliers as well as best practices identified through the survey.

Number and provenance of participants

Respondents by country

The survey was made available for institutions from around the world. However, due to the nature of the EuropeanaTech community, the majority of respondents were located in Europe. Most responses came from The Netherlands, which again, is a natural bias due to the Europeana Foundation being located in The Hague and members from the Koninklijke Bibliotheek and The Netherlands Institute for Sound and Vision being involved in the task force. The respondents are primarily from The Netherlands, Sweden, Germany, France, Belgium and the United Kingdom with Spain, Finland, Italy, Czech Republic, Lithuania, Austria and Luxembourg providing 1 to 3 responses. We would have preferred a more geographically diverse group of responses and will certainly investigate why institutes from other EU countries did not respond to the survey.

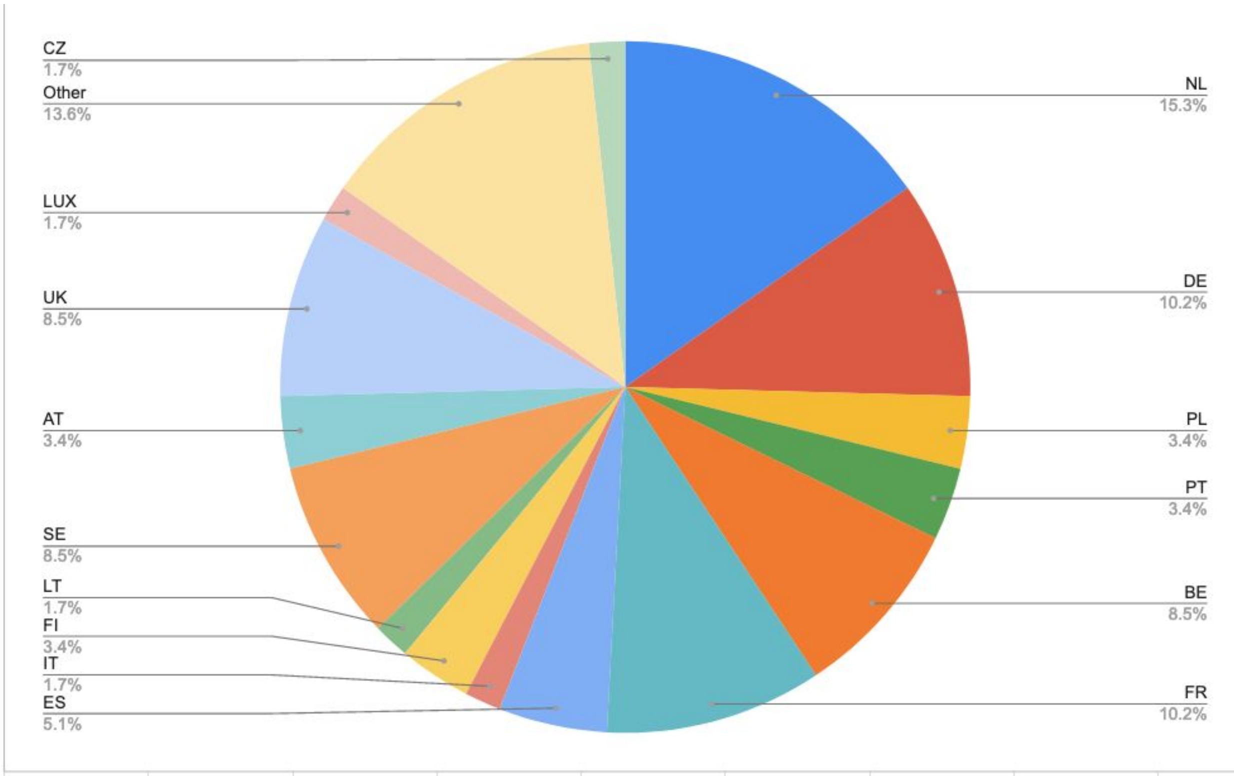

Figure 1: Distribution of respondents by country

Respondents by institution type

Overall the survey respondents represented either a library, archive, museum (LAM), research institution, some combination of two, or more. Many of the institutes identify themselves as a research institution and a LAM. This makes sense as such R&D intensive work like machine learning would usually require resources that accompany having additional research or technology departments. The assumption is that this institution type would have an impact on the diversity of content that is used for AI projects. However, it would appear that this did not have a huge impact on diversity of content types as respondents worked with a wide range of content types from medieval texts to still images, 3D images, audio, moving image and plain text. This could be because of the still experimental and high threshold for AI work.

Provided use cases

Nearly two thirds of all respondents had a use case(s) that they wished to share. Several respondents noted that they wanted to share use cases but then failed to fill in any meaningful data. Only two institutions shared two or more use cases. Three use cases was the max amount of use cases shared by a respondent.

Findings

Level of expertise or interest in AI

Almost all the respondents (91.8%) are interested in at least one AI topic (provided for or added by them), and more than half of them (54%) show expertise in one of the provided topics.

In a more detailed analysis of the interest by topic (see Table 1), we can see that *(Meta) Data quality* is the topic for which people have more practical experience with AI (26,67%), followed by *Collections Management* (18.64%), *Discovery and Search* (16.67%), and *Knowledge Extraction* (18.64%). In terms of interest, *Knowledge Extraction* is also the topic for which people show more interest (76.27%, most of them 'very interested'). The least interesting topics for our respondents are *Machine Translation* (25.42% are ‘not interested’ on this), followed by *Audience Analysis* (18.33%), *Crowdsourcing* (16.95%), and *Creative Engagement* (16.95%). *Machine Translation* is also the topic that was considered less useful among the ones who applied it, followed by *Audience Analysi*s (25%, and 20% of people who applied it considered it ‘not useful’ respectively).

	Not interested	Somewhat interested	Very interested	applied this, not useful	applied this, useful
Knowledge Extraction	5.08% 3	16.95% 10	59.32% 35	0.00% 0	18.64% 11
(Meta) Data Quality	6.67% 4	5.00% 3	60.00% 36	1.67% 1	26.67% 16
Audience Analysis	18.33% 11	36.67% 22	36.67% 22	1.67% 1	6.67% 4
Crowdsourcing and Human in the Loop	16.95% 10	28.81% 17	44.07% 26	0.00% 0	10.17% 6
Visualizing GLAM collections	13.33% 8	20.00% 12	53.33% 32	0.00% 0	13.33% 8
Collections Management	8.47% 5	10.17% 6	61.02% 36	1.69% 1	18.64% 11
Discovery and Search	8.33% 5	16.67% 10	55.00% 33	3.33% 2	16.67% 10
Creative or Engagement projects and initiatives	6.95% 10	27.12% 16	45.76% 27	1.69% 1	8.47% 5

Machine Translation	25.42% 15	28.81% 17	32.20% 19	3.39% 2	10.17% 6

Table 1. Results for the question “What is your level of interest or expertise with the following (AI) techniques/ topics?”

Additional areas suggested by our respondents that can be considered interesting for them or for which they have experience with are: Layout Recognition, Photogrammetry Automation, Production of 3D Content, Data Extraction (e.g. Optical Character Recognition (OCR) and Handwritten Text Recognition (HTR)), Music Information Retrieval, Collection Content Analysis, Semantics (Linked Data, Knowledge Representation), and Visual Recognition (object, subject, color of image/video).

Use cases analysis

Most of the projects fall under the span of digitization and discoverability. Figure 2 contains a histogram where each bar corresponds to a different goal, where the category “Others” refers to projects about bias detection, machine translation, quality assessment and duplicate detection. Each bar is divided into different sections corresponding to the different media types in each category. We will summarize the main goals and media types of the different projects in the following paragraphs.

It is clear from the different categories considered that the various goals stated are actually quite aligned, and that GLAM institutions are most interested in using AI for facilitating the exploitation (and to some extent, the production) of their digitized collections. By digitizing their cultural heritage objects they can improve the accessibility to these objects by the public, who might be able to access them via online portals. Once the objects are digitized, their metadata needs to be enriched for improving findability and searchability.

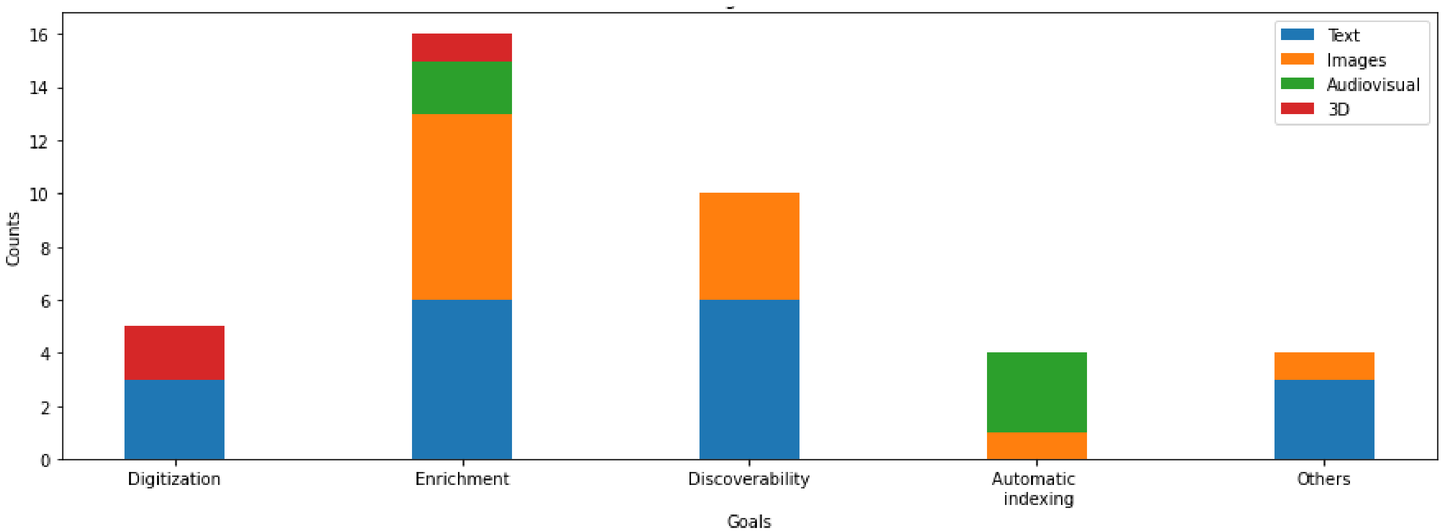

Figure 2

In total eleven projects are working on image classification based on style, technique or painter for automatic enrichment and indexing. Another topic of interest is image retrieval based on style and color, where the goal would be to find similar images to a source (query) image. A recurrent topic of interest is object detection in images, where the goal is automatically improving the metadata, so images are findable based on the elements they contain. Therefore, the main focuses of image analysis are enrichment and discoverability. Another topic with certain presence is the 3D reconstruction of 2D images, being applied mainly to buildings and historical objects for digitization and enrichment.

Regarding text, it is an ubiquitous format being present in most of the goals presented in Figure 2. Several projects are using Optical Character Recognition and Handwritten Text Recognition for the digitization of text contained in documents. This is used for obtaining the full text of the document and therefore improving the metadata and discoverability of documents based on their content. Due to the multilingual nature of some of the data sources, four projects are considering multilingual approaches, although only two of them are planning the use of machine translation. Some of the technologies used for enriching text are Named-Entity Recognition and Linked Open Data.

Five projects mention working with video and audio data, two of them explicitly mentioning audio processing. The applications are diverse, ranging from video segmentation to speech and music recognition for enrichment and discoverability.

Regarding the teams, most of them are of small size, typically 1-3 people. In most of the projects there is one or more GLAM specialists, which indicates the need for expert domain knowledge from the cultural heritage sector. In most projects there is also at least one software developer or data scientist due to the technical nature of the work.

Figure 3 contains a histogram of the number of projects per team size in people. Only those projects that clearly stated their team size were considered.

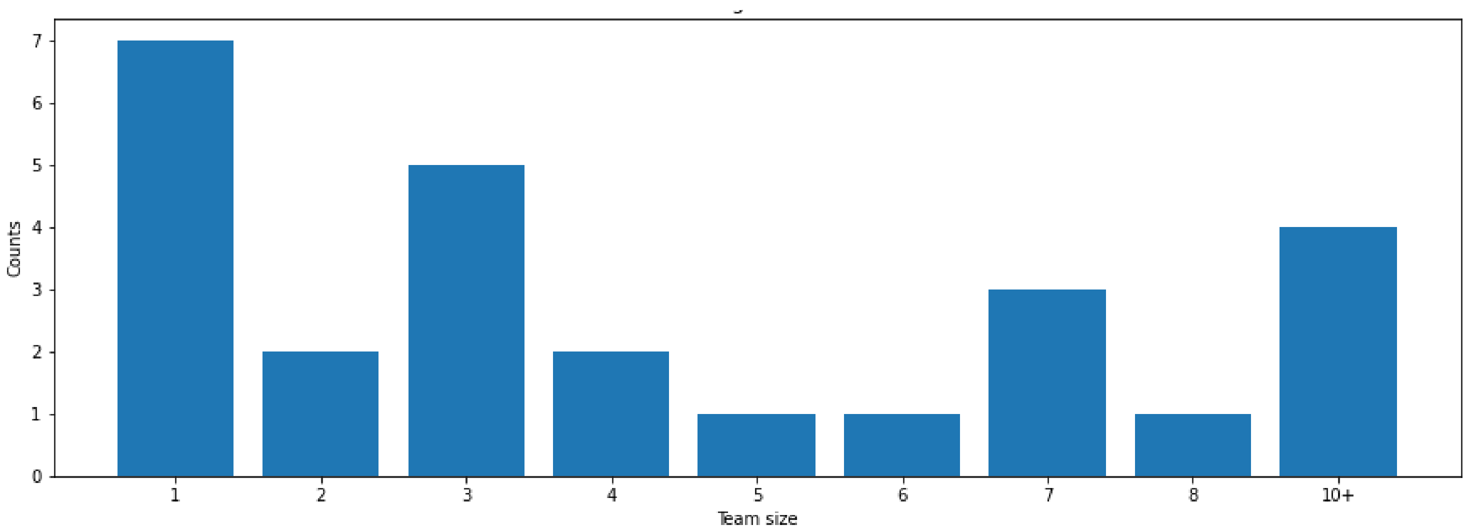

Figure 3

Technology analysis

The questions about the use of technical aspects of the use cases, such as content types, tools, frameworks, infrastructure and resources used were answered for slightly more than half of the responses received (33-35). The content types processed in the reported use cases (cf. Figure 2) focus on text (including scanned/OCRed and handwritten documents, 16 mentions) and images/photos (12 mentions). Other types of content are used less frequently, with 5 use cases processing audio/video and 6 various types of metadata and occasional mentions of other content like 3D and maps. Overall, this result is not surprising, as these content types are most common in GLAM's collections and also by far the largest amount of AI algorithms target still images and text, while e.g. video and 3D are less covered.

The responses related to the tools used indicate that about two thirds of the respondents used their own tool chain, but making in many cases use of open source algorithms. About 60% used (also) pretrained models, and about 75% trained models using their own data, often via transfer learning. Some comments suggest that a common workflow is to use pretrained models at first before taking steps with more custom tooling. The results indicate that the majority of organisations who explored the use of AI on their data moved quite deeply into the field by building their own tool chain and training on their own data (ed note. We will seek to exemplify these in the final report). Similarly, the most frequent response to the question about the computing infrastructure was using local infrastructure (16 mentions, note that this might range from developers' laptops to powerful servers), followed by research infrastructures on a consortium/regional level or service providers (6 mentions) and public cloud services (5 mentions).

In terms of AI frameworks, TensorFlow/Keras was most frequently mentioned (14). Other more commonly used frameworks are PyTorch (5) and Scikit-learn (3), while the rest of the responses are quite scattered. One interesting aspect are the external data resources used in the work: 2 mention Europeana, 1 mentions Wikidata, while all others rely on institutional, local or regional resources. The reason for this should be further studied, which may lead to identifying further needs for supporting CH organisations with shared data resources.

Outcomes and Impact

General overview of the impact and outcomes from the use cases

Almost all respondents whose projects were in progress reported some outcomes to date. These included working code and trained models; infrastructure and APIs; enhanced collections descriptions; and reports or presentations. Impacts were typically at an early stage, but ranged from production workflows to experience gained in deploying or learning more about AI technologies. Few respondents were able to report specific impacts on user experiences (with two cases having reached production level).

Challenges

Not many respondents of the survey provided detailed accounts of challenges faced while implementing AI related methodologies. From those that did denote challenges, issues ranged from practical matters like technical knowledge and personnel, costs of purchasing machines with enough processing power, training and hiring in external consultants to provide assessment of processes. Technical issues included having accurate and appropriate data, clear definitions of goals, scalability and being able to accurately evaluate the outcomes.

Concerns

Initial concerns for respondents varied. Overall concerns were more practical questions such as communication, timeframes, staffing, IPR and GDPR related issues, data quality, data quantity, experience, having clear use cases and presentation possibilities for the results.

Ethics

The overall ethical questions that institutions faced related primarily to copyright both for the data they were using and the open source tools they implemented. Some institutions only used items that they were the copyright holder of or where the material was openly licensed / public domain. Others were using mixed datasets or copyrighted content that they could use for only specific purposes proved challenging as it limited possibilities for exploitation. Tools for AI methodologies also proved difficult for a few as some off-the-shelf tools were not open source but perhaps easier to implement while open source tools would require more developer resources. Questions about facial recognition and personal data protection were also encountered as well as the ethics behind the material itself, especially materials related to colonialism.

Evaluation and metrics

One part of the survey asked about how the organisation using AI tools evaluated them, and which data sets and metrics have been used. Nearly two thirds of the respondents reported having carried out some form of evaluation, however, only about one third of those reporting evaluations provide details about the metrics. Four responses mention information retrieval metrics (precision, recall, F-measure, mean average precision), three mention classification accuracy and two mention metrics for speech/character recognition (word/character error rate). Concerning benchmarks, only CLEF-HIPE-2020 and Labelled Faces in the Wild (LFW) were mentioned. We see that many answers in this section are very general, and benchmarks are rarely used. Providing evaluation frameworks, guidance and best practices for evaluating AI technology seems to be needed by the community. It should also be further explored why existing benchmarking datasets are rarely used. If it turns out that the reason is their weak overlap with relevant tasks for GLAM institutions, then efforts on preparing and sharing datasets could benefit the community.

Feedback on role of EuropeanaTech

For several years now, EuropeanaTech's role as a facilitator and enabler for R&D work within cultural heritage has grown. Within the survey, the need and desire for such a facilitator and catalyzing body was made more apparent. Most respondents viewed EuropeanaTech as a body that could provide open access to knowledge from different institutes working with AI. This could be done either through white papers, blogs, journal articles in EuropeanaTech Insight or through events such as the past EuropeanaTech conferences (2015, 2018). While the sharing and transmission of knowledge is welcome, EuropeanaTech is also seen as a space and community where feedback and discussion can be held and validated. This could be in relation to less technical questions such as ethics, but also for technical feedback and validation of methods, tools and standards. With such a large member base primarily across Europe, EuropeanaTech offers a community that has the potential to move towards cultural heritage to a more synergized and interoperable body as opposed to siloed workings.

In the words of one respondent EuropeanaTech's role can be:

Collecting and spreading the word and knowledge of cases - especially cases from non-English speaking countries and GLAMs. Advocating for all GLAMs that create machine learning models, suitable datasets, and source code to publish them under open licenses. Help connect people and organisations to find partners for shared projects

However, and rightly so, a few respondents have casted doubts on whether or not EuropeanaTech could sufficiently carry out this role. There are still many issues with regards to metadata quality that are constantly in need of improvement within Europeana and AI work could be one means of bettering these issues.

Yet, with relation to the Europeana Innovation Agenda that was published in 2019, “Cultural policies should stimulate research into methods that can enhance the quality, usability and retrieval of complex digital objects. Machine learning and artificial intelligence will play a crucial role here, offering innovative solutions for automatic extraction of metadata and optimisation of content searchability. the implementation of appropriate preservation methods. This will be a crucial step towards assuring the durability of Europe’s cultural legacy.” This section was labeled by the community as an “Act Now!” priority. EuropeanaTech can certainly help move the Europeana Network Association forward with regards to AI.

Future work

Following this interim-analysis, the task force will select multiple institutions to engage with and conduct more in-depth interviews in order to gain more insights into their experiences working AI technologies. These interviews and the survey will go on to inform the final task force report that will be delivered in early 2021.

Should you wish to provide feedback or share insights please contact gmarkus@beeldengeluid.nl