

“A picture is worth a thousand words”? - From Project Inception to First Results: Describing Cross-disciplinary Collaboration in the Digital Humanities Project ChIA

Yalemisew Abgaz¹[0000-0002-3887-5342], Amelie Dorn², Gerda Koch³, and Jose Luis Preza Diaz²

¹ Adapt Centre, School of Computing, Dublin City university, Ireland

`Yalemisew.abgaz@adaptcentre.ie`

² Austrian Academy of Sciences, Austria

`{amelie.dorn, JoseLuis.PrezaDiaz }@oeaw.ac.at`

³ Europeana-local Austria, Austria

`kochg@europeana-local.at`

Abstract. Both, historical as well as contemporary images depicting particular aspects of a certain culture, or a certain cultural practice, are widely available in a variety of formats. Typically, historical-cultural images can be found in museums, national archives and libraries both in analogue and digital formats. The Europeana image collection serves as a valuable example of a digital image collection available to the public, that is freely accessible and searchable. A significant amount of the collection, however, does not contain a rich semantic description of the cultural and social aspects represented in the images that would go beyond some metadata like author and title. To enable the enrichment of these images, we propose a project which brings together experts from digital humanities, linguistics, artificial intelligence and semantic web technology. The project aims at analysing the contents of the images with a combination of computer vision, natural language processing and manual curation to represent them with a more descriptive and representative controlled vocabulary. This combination of different types of expertise to address the problem enables us to closely collaborate and learn from each other by taking different roles and perspectives. So far, the collaboration has contributed to the understanding of the detailed requirements from the digital humanities and socio-linguistic perspectives for the representation and processing of cultural images using semantic web technologies, such as multi-disciplinary thesauri, ontologies, computer vision and AI.

Keywords: Cultural image analysis · Semantic enrichment · Computer vision · Ontology · Knowledge design

Copyright 2020 for this paper by its authors. Use permitted under.
Creative Commons License Attribution 4.0 International (CC BY 4.0).

1 Introduction

Fred R. Barnard famously once claimed that “A picture is worth a thousand words”. This assumption resonates well with the aims of the collaborative project discussed in this paper. By cultural image, we understand any image (artwork, photograph, sketch, etc) that depicts cultural artefacts, practices, or social situations, where we particularly focus on food-related content in the context of the project described in more detail below. The historical dimension is just one of several different ones under the umbrella of culture and particularly alludes to places, persons or events. Several books have indeed been written about historical images (pictures, paintings, photographs, sketches etc) such as the Mona Lisa by Leonardo da Vinci. Except for a few Nobel artworks, no matter how much is written about them or captured in the metadata, there remains ample information contained in the images themselves which is left uncaptured in words. In many museums, cultural heritage and public archives, there are images which vividly and meticulously represent the cultural, social, political and other aspects of a society in time. The modern-day digital technology, however, heavily relies on the exchange of information in a form of text to present or represent things. Until recently, often only a few keywords or descriptive sentences were used to annotate, search and retrieve images. This has created a huge gap between the symbolic “thousand words” and the few keywords and descriptors associated with the images.

In this paper, we present a collaborative research endeavour which aims to unpack the number of details contained in cultural images and represent them with carefully selected, semantically interlinked multi-faceted ontologies, taxonomies and thesauri. To this end, a group of researchers from the digital humanities (socio-cultural linguists, archival curators) and computer scientists (computer vision and AI experts and ontology engineers) bring the ChIA⁵ project (accessing & analysing cultural images with new technologies) together, to carry out pilot research in the area [3]. Although the overall objectives of the ChIA project encompass a much wider range of goals, this paper mainly tries to answer the following main questions:

- How can the interaction between Digital Humanities researchers and computer scientists result in a formal collaboration?
- What are the methods of collaboration and technical requirements?
- What are the lessons learned so far, and what are the challenges faced in the collaborative project setting?

We further report the collaboration results so far, and how these results are understood and interpreted by the different experts in the group. The technical solutions that are proposed in the research and the methods we followed to bridge the gap are also discussed in detail.

⁵ <https://chia.acdh.oeaw.ac.at/>

2 The Humanities Research Problem

The cradle of this collaborative project is an inspiring, informal brainstorming session held between a digital humanist, who is an expert in the area of lexicographic collections, a computer science expert and an ontology and knowledge engineer. The discussion identified several pertinent, to-date unresolved challenges related to historical collections including cultural images, pictures, photographs, sketches etc., collected from several centuries and stored in libraries, archives and museums. The main challenge identified during the discussion was that the majority of these resources lack any rich semantics (description) except for some available metadata, for example, date, author and title. Despite this fact, these collections, particularly historical images, bear rather rich aspects describing the detailed social, cultural, political, economic, etc. interaction of different societies. These aspects are not represented in a written form except for images well studied by experts. This led to the under-exploitation of the resources as a whole for research, educational and business purposes. Considering the richness of digital collections and the impact they have for understanding the culture and interaction between societies at different times, it became clear that digital humanities was concerned with the long-standing question of how to make these resources better accessible and available by drawing on the expertise from different disciplines. The challenging factor until now, however, is concerned with the questions of how to organise cultural image resources automatically, how to analyze their content and the aspects and represent them in a more detailed, interlinked and interrelated manner. It also focuses on how to support systematic search and retrieval of these resources using rich semantics which can go beyond searches based on facets such as author and title.

Since the inception of this research question, we conducted studies [4] to consolidate our research question and to understand the magnitude of the problem [10]. During the proposal preparation phases, we began to understand that it was necessary to include additional experts from digital humanities who were directly involved in providing platforms and access to historical and cultural image collections for their users. After inviting such experts from Europeana local Austria, we advanced our understanding of the day-to-day challenges of the experts [8]. The challenges are non-trivial ones which still require a systematic approach and a deeper collaboration between linguists, computer scientists, knowledge organisation/ontology experts and digital humanity researchers. Thus, the main research questions from the humanities point of view are:

- How can the rich information contained in historical images be explicitly represented, semantically enriched and interlinked by these resources?
- Which intelligent and interactive tools can be provided to make such resources searchable, analyzable and exploitable for both humans and machine agents?

Within this set of specific research questions, we further included the following technical questions:

- What is the best method to analyze and represent the contents of historical images with or without written detailed description about the images?
- What would be the best standards to follow to semantically enrich these images once their content was analysed.

This opens up the path to the investigation of best knowledge organisation methods and tools that support semantic annotation/enrichment, and semantic search of historical images.

With the three major research questions emanating from the first question, we developed a collaboration between four major disciplines: digital humanities, socio-cultural linguistics, computer science (computer vision and AI in particular) and knowledge organisation and semantic web technologies. This unique collaboration between the experts (Section 3) enabled us to deeply understand the problem at each corner of the interaction and helped us to tackle the technical details outlined in Section 4.

3 The Collaboration Experience

ChIA is an ongoing collaborative project and it is very important to give a detailed description of the composition of the team and how the team collaborates.

3.1 The ChIA-team Details

There are four core members in the ChIA collaboration, broadly categorised as digital humanists (two) and computer scientists (two). The digital humanist group mainly focuses on the analysis and representation of cultural images to enhance better access for users of the systems. Their main research interests include the information is contained in cultural images, the aspects of a society the images represent, how such aspects are represented and how the information contained in the images is represented using different languages and formats. The two digital humanists have further specialisation in their interest. The first member is a linguist by profession who is interested in conducting a socio-cultural and socio-linguistic analysis of cultural images and design thinking methods on how societies represent complex historical and cultural events using digital images. The second member is a manager and content-coordinator and also focuses on providing metadata and platforms for collecting, organising and presenting cultural images on the Europeana-local-Austria platform. The first focuses on the research aspect, the second focuses on the implementation and provision of the actual service.

The second group, computer scientists, also has two members specializing in different domains in computer science. One of them is an AI and computer vision expert with the objective of looking at cultural images as opposed to contemporary image collections to find out how details contained in an image can be extracted by computer vision tools. Finally, we have the fourth member who specialises in knowledge representation, focusing on how complex social and cultural aspects can be represented across disciplines using ontologies, taxonomies

and thesauri. The main objective of this team from the technical point of view is to analyse cultural images and extract as much detail as possible and represent such detail with semantically rich, interlinked and interoperable system and to support its exploitation via chatbots.

There are two principal investigators in this project, one from the digital humanities and the other from the technical computer science. The most important element next to the research question is enabling a thorough understanding among the team members. According to research on a diversified multidisciplinary team [7], our team is diversified in several ways including task-related diversity (e.g. education, profession etc) and relations-oriented diversity (e.g. sex, geography). Having a diversified background often is a problem unless it is managed and directed systematically. For this purpose, we have built strong interpersonal and project management skills to lead this project as PIs. Since the very interaction between these two domains to address the outlined problem is important, we have also set up different communication channels to keep each other up-to-date on our day-to-day interactions. Besides, the project also draws on advice and guidance from an international advisory board, of again four members, who are also experts in the fields of semantic technologies, knowledge design, the GLAM sector and AI.

3.2 The CHIA Cross-cultural Team Communication

The team is not only professionally diversified but also physically dispersed with only two of the team members working in the same building while the third team member works in the same country but in a different city. The fourth team member is in Dublin, Ireland. This physical distance introduced challenges and opportunities. The main challenge is that the team members are not able to meet and update their day-to-day activities. However, different channels are used to plan tasks and update progress. The most important one is the weekly project meeting which is held via Skype. This meeting allows us to update our progress, to plan our next tasks and serves as a podium to ask a question about topics we do not understand. This gives us a great opportunity to have live interactions among the team members. We have understood that there are at least three types of members in the team: those who provide up-to-date information daily, sending updates and any relevant information about the project, those who keep the work going by experimenting and building prototypes to test the ideas that are circulated among the team and those who prepare, plan and evaluate the overall direction and progress of the project. They keep the meetings running smoothly, supervise the adherence of the work to relevant standards etc. The four members usually play one or more of these roles interchangeably during our meetings or throughout the week. Our official communication channel is via email and our document sharing platform is google drive. We prefer to use google as it supports sharing and collaborative editing of project-related documents and supports easy and quick interaction among the project members wherever they are. Also, a project space on slack has been set up to communicate important information to team members and the advisory board.

One of the major drawbacks of the online communication system is that it is very easy to lose track of the agenda while members try to explain complex concepts. To reduce this, we further introduced internal review meetings every six months to come together physically to review our progress and discuss complex matters in person. The other limitation is the lack of sufficient time to discuss and understand complex issues within a single meeting. This is a difficult scenario and when such complex issue arises, it is discussed offline in one-to-one settings. This makes it easier to understand the issues in detail and further allow for an unhastened and friendly discussion. In general, the physical distance between the members has contributed to slightly reduce the progress of our understanding of other members perspectives and interests.

4 Description of the Technical Collaboration: The Case of Building Experimental Dataset

One of the main deriving research questions is bridging the gap between the information packed in the images and the explicit annotation of the content of the images using ontologies. The interaction between the team members to understand the problem and to work toward the solution required the following interaction (Fig. 1) to be held all the time among the team members. These interactions represent various levels of research collaboration within the project. For the sake of brevity, the interaction is explained by taking the dataset preparation phase of the project.

4.1 The DH Quadrant

The Digital Humanities (DH) quadrant is dedicated to identifying cultural images that are widely used by users of the platform. The digital humanist uses the collection to search and retrieve cultural images. The total image base consists of data originating from many different repositories created by a considerable number of cultural heritage professionals during several decades [11]. As a result, the data is very diverse and the collection is huge. For the experiment, we, therefore, limit ourselves to cultural images related to staple food edible for humans. To carry out the selection of the images, the DH directly interacts with the Socio-Cultural Linguistic (SL) to determine how food-related culture is represented in the language of the search (which is German and English at this stage). This collaboration yields a list of terminologies that are widely used to represent the cultural aspects contained in the images. For working with the images in all the other quadrants, we needed access to the Europeana-Local-Austria system. The DH set up a web-based interface that makes use of a customised Europeana Search API to enable search, selection and management of selected subsets within the image collection. The interface allows a (Boolean) search for images on metadata element level and the retrieval of a considerable image base for CV analysis. The DH gave us a short introductory training on how to use the system. This very useful collaboration improves the understanding of the rest of

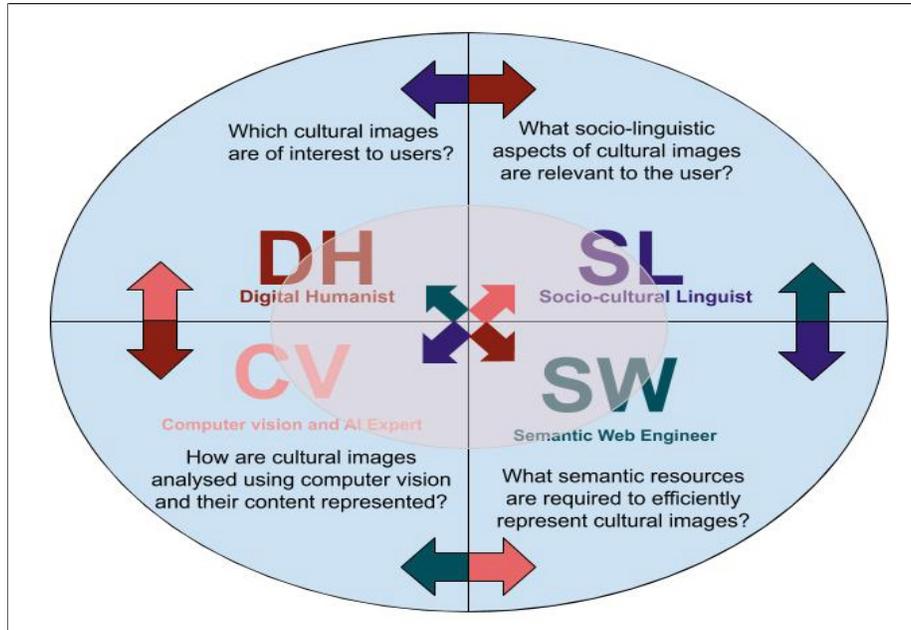


Fig. 1: Interaction diagram among project members.

the team and interaction with the system and further helps us to reflect on the problem from the DH side.

4.2 The SL Quadrant

The SL quadrant, in turn, interacts with the DH to understand what kind of cultural images are available in the collection and how they are represented in the metadata so far. This includes understanding the conceptualisation of the cultural aspects of societies [6]. The SL directly collaborates with the Semantic Web (SW) engineer to identify existing ontologies, thesaurus and dictionaries that represent the socio-cultural aspects of the images. The collaboration further looks into the novel combination of existing tools or creation of new ontologies to represent concepts that were not represented. This collaboration serves as an input to the SW in utilising existing and new ontologies to semantically annotate and represent cultural images in a more robust manner.

4.3 The SW Quadrant

The main question in the Semantic Web (SW) quadrant is to identify useful ontologies to represent the rich information contained in the images to support semantic annotation, reasoning and semantic search. This will enable the project

to provide a rich, semantically interlinked open data of the image collection which will be consumed by the AI system to build an interactive chatbot. The SW interacts with the CV quadrant to represent the output of the computer vision using ontologies. The collaboration between the SW and SL quadrant also benefits from previous collaborative research that has been carried out in [5, 1, 2, 4]. Furthermore, the DH interacts with the SW in listing and selecting useful ontologies and vocabularies that have been in use in the platform.

4.4 The CV Quadrant

The Computer Vision (CV) quadrant focuses on extracting as much detailed information from the digital cultural image collections identified by the DH. This quadrant consumes existing computer vision tools [9] to analyse and represent the contents of cultural images and evaluate the accuracy. Although the metadata is not always available for all the images, the results from the CV will be compared with the existing metadata of the images. The CV quadrant further focuses on building a chatbot by consuming the outputs of the SW quadrant. The CV quadrant also provides several outputs of a computer vision and enables the other team members to understand what a computer vision can output when given historical images of several years old.

All the four quadrants have a common goal which is to understand the requirements of the other quadrants and making their research question and requirements clear to the other members. The circle in the middle of the collaboration diagram (Fig. 1) represents the information we all need to have in common about the project. We believe that as the more interaction we have among ourselves and advance the project, we will be able to push the radius of the circle to learn and collaborate with more overlapping interest from the other team members. Our objective is to push the inner circle to the edges of the outer circle in all directions. Until that happens every member of each team will have a blind spot which requires the consultation of the responsible member of the quadrant. For example, the SL may not fully understand how the SW quadrant works. In such cases, depending on the problem, the SL may rely on the SW expert to carry out the task. Such tasks may remain the blind spot of the SL.

Much of the CV work is technical and requires time and effort to fully understand it. This could be another blindspot for the team members. We look forward to having training on computer vision and its use of algorithms to familiarise ourselves with the concept. We also look forward to having more technical training on how to build a chatbot system and on how to evaluate its outputs. This will also help us to bridge the gap between the technical part of the project and the research (theoretical) aspects of the project.

5 Conclusions and Recommendations

In this paper, we explored the collaboration among four experts from different academic backgrounds working together to address long-standing digital humanities questions. The collaboration from the inception of the project to its current

implementation stage is demonstrated. We looked into the major research questions and how the team collaborated to address each of these research questions. Collaboration between technical computer scientists and digital humanities experts is not usually smooth. In this project, we demonstrate how we managed to maintain the collaboration at its best by enabling smooth interaction between the team members. The lessons we have learned from this collaboration and which we continue to learn is that it is very important to start with a sound digital humanities research question and understand the problem from many different sides before rushing to a technical solution. While working with the digital humanists to understand the problem with an example, it is also important to introduce the technical requirements early to ensure that the digital humanists capture all the necessary information which is required to complete the technical solution and most importantly all the requirements to evaluate the system at the end. Finally, we conclude that a picture can indeed say more than a thousand words, however, with the right set of digital tools, controlled vocabularies, thesauri and ontologies we should be able to successfully access implicit knowledge via human and machine analysis.

Acknowledgements: This research is funded by the Austrian Academy of Sciences under the funding scheme: go!digital Next Generation (GDNG 2018-051). The ChIA project is carried out in collaboration with the Adapt Centre, DCU. The ADAPT SFI Centre for Digital Media Technology is funded by Science Foundation Ireland through the SFI Research Centres Programme and is co-funded under the European Regional Development Fund (ERDF) through Grant # 13/RC/2106.

References

1. Abgaz, Y., Dorn, A., Piringer, B., Wandl-Vogt, E., Way, A.: Semantic modelling and publishing of traditional data collection questionnaires and answers. *Information* **9**(12) (2018). <https://doi.org/10.3390/info9120297>
2. Abgaz, Y.M., Dorn, A., Piringer, B., Wandl-Vogt, E., Way, A.: A semantic model for traditional data collection questionnaires enabling cultural analysis. In: Proceedings of the LREC 2018 Workshop "6th Workshop on Linked Data in Linguistics (LDL-2018)". Japan, Miyazaki (2018)
3. Amelie Dorn, Yalemisew Abgaz, G.K., Diaz, J.L.P.: Harvesting knowledge from cultural images with assorted technologies: the example of the ChIA project. In: The 16th International ISKO conference. Aalborg, Denmark (2020)
4. Dorn, A., Abgaz, Y., Wandl-Vogt, E.: Opening up cultural content in non-standard language data through cross-disciplinary collaboration: insights on methods, processes and learnings on the example of exploreat! In: Krauwer, S., Fišer, D. (eds.) *TwinTalks-DHN 2019. Twin Talks Workshop at DHN 2019. Proceedings of the Twin Talks Workshop at DHN 2019, co-located with Digital Humanities in the Nordic Countries (DHN 2019)*. Copenhagen, Denmark, March 5, 2019. vol. 2365, pp. 82–89 (2019)

5. Dorn, A., Wandl-Vogt, E., Abgaz, Y., Benito Santos, A., Therón, R.: Unlocking cultural conceptualisation in indigenous language resources: Collaborative computing methodologies. In: Soria, C., Besacier, L., Pretorius, L. (eds.) Proceedings of the LREC 2018 Workshop "CCURL 2018 – Sustaining Knowledge Diversity in the Digital Age". pp. 19–22 (2018)
6. Goikhman, A., Therón, R., Wandl-Vogt, E.: Designing collaborations: Could design probes contribute to better communication between collaborators? In: Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality. p. 1219–1222. TEEM '16, Association for Computing Machinery, New York, NY, USA (2016). <https://doi.org/10.1145/3012430.3012431>, <https://doi.org/10.1145/3012430.3012431>
7. Jackson, S.E.: The consequences of diversity in multidisciplinary work teams. *Handbook of work group psychology* pp. 53–75 (1996)
8. Koch, G., Koch, W.: Aggregation und management von metadaten im kontext von europeana. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare* **70**(2), 170–178 (Sep 2017). <https://doi.org/10.31263/voebm.v70i2.1776>, <https://journals.univie.ac.at/index.php/voebm/article/view/2071>
9. Kubany, A., Ishay, S.B., sacha Ohayon, R., Shmilovici, A., Rokach, L., Doitshman, T.: Semantic comparison of state-of-the-art deep learning apis for image multi-label classification (2019)
10. Longhurst, B., Smith, G., Bagnall, G., Crawford, G., Ogborn, M., Baldwin, E., McCracken, S.: *Introducing cultural studies*, second edition. Taylor and Francis (2008). <https://doi.org/10.4324/9781315834344>, <http://usir.salford.ac.uk/id/eprint/40855/>
11. Wasner-Peter, I.: *Handbuch kulturportale. online-angebote aus kultur und wissenschaft*, hrsg. v. euler, ellen/hagedorn-saupe, monika/maier, gerard/schweibenz, werner/sieglerschmidt, jörn. berlin/boston: De gruyter saur 2015. *Communications of the Association of Austrian Librarians* **71**(3-4), 558–560 (Dec 2018). <https://doi.org/10.31263/voebm.v71i3-4.2176>, <https://journals.univie.ac.at/index.php/voebm/article/view/2159>