IDEAH • Vol. 1, Iss. 1 (DHSI 2017 & 2018)

Building the ARTECHNE Database: How to Develop a Multi-Purpose Database for an Interdisciplinary Project

Marieke M.A. Hendriksen

Published on: May 31, 2020

DOI: 10.21428/f1f23564.e6779c76

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Between 2015 and 2020, the European Research Council-funded project *ARTECHNE: technique in the arts 1500-1950*, was directed by Sven Dupré at Utrecht University and the University of Amsterdam. The project team as envisioned in the grant application consisted of two postdocs, two PhD students, three part-time art conservation specialists (paintings, metals, and glass), and a programme manager.¹ The team is truly interdisciplinary, with backgrounds in philosophy, chemistry, history, art history, history of science, technical art history, and conservation and restoration science. With the help of additional funding, two data entry assistants were added to the team for 2.5 years and six months respectively, and a close collaboration with two scientific programmers from Utrecht University Digital Humanities Lab was developed.

The researchers on the ARTECHNE project study how artistic techniques were transmitted over the centuries, and what the role of texts was in this transmission. It is often assumed that artistic skills such as drawing, painting, silversmithing and glass blowing can only be learned through one-on-one instruction. However, the sheer number of historical artist handbooks, art technical instructions and recipe books suggests texts did play a role in the transmission of artistic knowledge and skills. The ARTECHNE project investigates what that role could have been, through the analysis of primary and secondary sources and through making reconstructions of pigments, grounds, and silver and glass objects, using those written sources as our guides (Dupré; Hagendijk; Boulboulé; Hendriksen, “Casting Life, Casting Death” Pinto).

My research within the project (2016-2019) focused on the meaning of the term “technique” and on the transmission of techniques between the visual arts and other disciplines. The term “technical” is used widely in relation to art and art history today, while we do not have a history of the shifting meaning of the term “technique” in the arts and sciences. Although related forms were occasionally used in European languages before c.1700, the word “technique” was a neologism in the vernacular that started to appear sparsely in treatises on arts and sciences only from the middle of the eighteenth century. Rooted in the Greek *techne*, which was translated routinely as “art” until the mid-eighteenth century, *technique* referred to both processes of making or doing and their products. Yet from c.1750, a distinction of processes of making or doing from the resulting artwork appears to have arisen in German philosophies of art. Exploratory research—consisting mainly of close reading of relevant texts and some corpus analysis using existing digitized corpora such as Kant’s collected work and the ARTFL database—suggested that this distinction may have come about explicitly to develop arguments about judgements of taste, artistic value, and the appreciation of art (Hendriksen, “Art and Technique Always Balance the Scale” 201).

Why a Database?

As the introduction outlines, building a project database was not part of the original plan. Yet, even before the project started, it was clear that relevant textual sources were dispersed and often not digitized, which made accessing and comparing them complicated. Existing large datasets, such as Google Books, were not very useful for our purposes. However valuable Google Books and its analytical tool, Google Books Ngram Viewer, may be for some fields, for those working on sources before 1800, they have very limited value. An n-gram is an instance of a word or phrase within a corpus; n is a variable representing the number of words.² In other words, Google Books Ngram Viewer counts how often a word or a combination of words occurs in the digitized printed sources available in Google Books in any given year between 1500 and 2008 and visualizes that in a chart. Google Books contains over 25 million titles and Ngram Viewer works for those in American English, British English, French, German, Spanish, Russian, Hebrew, and Chinese, so in theory it would be a great way to figure out when the term “technique”, or a combination of words like “art” and “technology” first occurred in European languages, and how it spread. For texts after 1800, this works relatively well, as Figure 1 shows: it at least confirms that “technique” was very much a neologism.

Figure 1: “Technique” and “Art” in Google Ngrams

Yet there are three main issues that make the tool less useful for early modern historians. First, the relatively small number of texts before 1800 in Google Books makes for an eschewed representation, second, the low quality of the OCR of many pre-1800 texts, and thirdly, the high number of false positives for a neologism such as “technique” due to human error in the entry of metadata (Hendriksen, “Google NGram for Early Modern History?”).³ Given the potential difficulties of writing a history of technique in the arts through close reading and the insufficiency of existing resources for the project’s needs, a database was potentially a useful tool to help answer these questions, yet as the Google Books example shows and has been argued by digital humanities (DH) scholars before, there were also many pitfalls to avoid (Hitchcock 9).

Having some previous experience with GIS and a strong interest in DH, I imagined that a database could help me compare a great number of textual sources efficiently, and that it could also help to visualize how the use and meaning of the term “technique” in Europe spread and changed. Moreover, I realized that many of my colleagues faced similar challenges in researching artistic techniques, comparing sources, and documenting reconstructive research. Hence, I proposed to develop the lacking history of the shifting meanings of the term “technique” in the arts and sciences through an online database containing searchable full-text early modern recipes, artist handbooks, and technical instructions with the help of the UU DH Lab.

We obtained funding to work with the DH lab and started working on the database project in February 2016. The following sections will discuss the aims of the database in more detail, the challenges and problems we encountered, technical choices, and the end result. Subsequently, an attempt is made to answer the question how much digital literacy on the part of historians is required to successfully set up research projects relying on new technologies. I conclude with a critical reflection on the strengths and weaknesses of the resulting dataset and interface and give recommendations for interdisciplinary research project teams considering the development of a project database.

Aims, Challenges, Choices

One of our main aims was to integrate existing, orphaned databases on artistic techniques. A recurrent problem for digital humanities projects is the sustainable preservation and continued accessibility of research data. Although initiatives such as DigitAl Research Infrastructure for the Arts and Humanities (DARIAH, see Kálmán et al. 113) aim to provide structures to resolve these problems, this was too complex to comply with for a small database project at the time (2016) as it was still very much a work in progress, so we had to come up with different solutions that we could implement with the limited time and resources available.

Often when a project ends, no mechanisms are in place to preserve and maintain data sets and interfaces, and accessibility becomes dependent on the goodwill of the former host institution. For

example, several of the project members had previously created other databases at different institutions. Some of them were still accessible through institutional websites, but it was unclear how long that would remain the case. Others solely existed on individual computers, and one was so dated that it required some digital archaeology to make it accessible at all. Therefore, one of our first objectives became moving those databases to Utrecht servers, so we could integrate it into the new ARTECHNE database. Although this sounds relatively straightforward, these moves required planning and time investments from various parties within the project: team members, the DH lab scientific programmers, and data managers at other institutions. In one case, a data manager travelled from Utrecht to Berlin with a hard disk containing the [ColourConText database](#), and spent three days working with an Utrecht-based programmer to successfully migrate a copy of an existing database to local servers. Subsequently we had to manually clean out incomplete and incorrect files from these datasets in order to be able to fully integrate them with our own data.

In the first few months of the project, our scientific programmer, Martijn van der Klis, built a basic relational database structure in Drupal and an interface that allowed us to search the ColourConText files much faster through our own website. Moreover, he applied automated geotagging to the ColourConText files to enable visualizing them on a geographical map. This led to the first problems: the sources in ColourConText had been linked to the location of the archives in which they were kept and not to the location where the sources had been produced or published. Hence, we added two location fields to each record: one for production place, one for current location. Eventually the interface allowed users to search both, either simultaneously or separately. Another problem was that the ColourConText records, which had been added by a group of people, contained quite a number of incomplete records, which could not be used. These had to be cleaned out by hand, as it had to be checked that they really did not contain anything of interest before discarding them.

While completing this process, we simultaneously started working on attaining our primary goal: to create a project database containing digitized, fully searchable historical texts on technique in the arts, which could be linked to other types of data, such as (moving) images, reconstruction and conservation reports, and which offers the user various search and visualization tools. These tools help users to create geographical, interactive semantic maps that situate terms used to describe techniques in the arts before the term “technique” became commonplace. This can for example be words that we call *somatic indicators* – nouns and verbs that instruct the reader on how to use the body or to seek particular sensations or cues, e.g. “rub the pigment until it feels like sand between the teeth”, but also verbs that describe a particular technique, such as “enameling” or “grounding”. Such maps can also provide insight in the changing meanings of terms. In order to make our data linkable to other open data, we first jointly developed an ontology of triples based on the CIDOC Conceptual Reference Model.

Figure 2: The Ontology of the Artechne Database, December 2017

We chose [CIDOC CRM](#) as it is a well-known ISO standard in the cultural heritage sector, which provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation. Moreover, we decided to link our descriptions to the [Getty Vocabularies](#)—ULAN for artist names, AAT for glossary terms, and CONA for artefacts—wherever possible. These vocabularies contain structured terminology for art, architecture, decorative arts, archival materials, visual surrogates, conservation, and bibliographic materials. They are compliant with international standards and provide authoritative information for catalogers,

researchers, and data providers. Another important consideration for us was the fact that the Getty Vocabularies are multilingual and created through user input, as the database was due to contain sources in at least six different languages.

Meanwhile, the ARTECHNE team made a selection of sources to be added to the database. The first batch were circa 100 art technical texts from western Europe in six different languages, from between 1500 and 1900. This time frame not only covered most of the period studied in the overall project but, focusing on texts from before 1900 had the added benefit that we avoided copyright issues. The texts were selected for relevance to our individual research projects and the overarching goals of the project. Many texts were already digitized in some way or another, with PDFs available in online repositories, but often there was no reliable OCR nor useful metadata. In January 2017, after funding was acquired for a part time data entry employee, we started adding sources to the database. If necessary, we applied OCR using ABBYY Fine Reader or we transcribed digitized sources. Subsequently they were added to the database in chunks, better known as records, to enhance searchability. For the same reason, glosses of current and historical names were added to each record. A manual was developed for all project members, so anyone who came across anything they felt needed to be added immediately could also do this themselves.

Subsequently, our programmer developed visualization tools to allow the search results to be viewed in list, geographical map, timeline, and as word cloud and collocations. Finally, a text comparison feature was added to enable easy one-on-one comparison of various works within the database, as many early modern art technical recipes and instructions circulated in various forms. These features were completed in the fall of 2017. During this main development stage of the database, the team was confronted with a number of practical problems. The first one was that OCR correction and all data entry had to be done manually to ensure data quality, which was very time-consuming, and severely limited the amount of text we could add. Special challenges in this respect were that many of our sources are early prints (pre-1800) in various languages, and a significant portion of the German texts was printed in Gothic font. These sources were nearly impossible to read through OCR at the time, so they had to be transcribed word for word.

Moreover, the process of enriching texts and linking objects and textual sources proved to be labour intensive too. Given our limited resources, we chose not to apply standard text encoding, but to use a lighter form of annotation, creating our own glossary. To ensure that the data would remain usable for external parties and after the project as well, we not only committed to the CIDOC-CRM-based ontology and the linking to the Getty Vocabularies, the database was also indexed using Apache Solr, allowing for export of data from the application to .csv-format, and was given stable URIs. The database thus adhered to the 5-star open data plan formulated by [Berners-Lee](#), which should make it possible to link it to more open data, thus enriching our own data.

Although initially aimed at aiding and documenting the ARTECHNE research project, we believed from the beginning that the database should and could serve a much broader community of art historians, historians of science, conservators, restoration specialists, and other researchers. Our final, but certainly not the least important, aim was therefore the sustainable preservation of the project database for future use, and compliance with the 5-star open data plan alone was necessary but not sufficient to meet this goal. We did not want to end up in the situation that when the ARTECHNE project ends in 2020, the database would be left on university servers without further maintenance or guarantees for continued upkeep of the interface. Johanna Drucker has argued, given the importance of the performativity of interfaces, it is extremely important to preserve not only the raw dataset, but also the interface, and to make sure that the data can also be easily accessed through a user-friendly interface after the project ends, either the original interface or another one (16, 38). This is especially important for the humanities, where many researchers, regardless age, have very limited skills in retrieving data from repositories.

Hence we decided on a dual approach: on one hand, the complete dataset and the code for the interface will be deposited in two open access online repositories, and on the other, the dataset is migrated to and integrated into the [digital resources of the RKD Netherlands Institute for Art History](#) in The Hague. The former ensures the ongoing preservation of the original dataset and interface code. By choosing to deposit in both [Zenodo](#) and [DANS](#), we have a copy in both an international and a national government-funded open access repository. The latter ensures immediate, easy and ongoing access to the data (but not the analytical tools) for both researchers and a general audience. This is a pragmatic and defensible choice, as the analytical tools were developed for the specific purposes of the ARTECHNE project and will remain available on university servers for at least two years after the project ends. Moreover, even though the analytical tools were not migrated to RKD, the data can be analyzed with the tools offered on the RKD website and are linked to the data already available at this institution.

Results

As the ARTECHNE project draws to an end, the database continues to function as a repository and research resource for the team members. For example, when several ARTECHNE team members became involved in the development of the exhibition [Back to Black. From historic colour recipes to our contemporary experience of black](#) at Museum Hof van Busleyden in Maline, Belgium, they added the dye recipes and other resources used for the exhibition and the catalogue to the database, where they could easily be linked to other relevant materials, and be consulted and analyzed by all collaborators.

Some of the original intentions could not be realized with the database, but it did provide researchers with insights that would have been extremely difficult to develop based on a close reading of sources alone. For example, it proved impossible to gather enough data that allowed for a reliable collocations

analyses that could tell us more about the words used to describe “technique” in the arts before the term became common in the vernacular. Yet the database did provide other insights, e.g. it gave us a much clearer idea of the historical spread and circulation of written sources on technique in the arts, particularly when it came to specific techniques that turned out to be central to more than one form or arts, such as the use of wax.

Figure 3: List View of the Results for a Simple Search for “Wax”, August 2019

For some project members however, the database never became an integral part of their research practices, either because they made a conscious choice not to invest their scarce time in meticulously adding their research data to the database, or because there was no clear benefit to their research in doing so. This never became a point of contention, as we agreed from the start that the database should serve project members and the larger community wherever possible but creating and maintaining it was a means to do so, not a purpose in itself.

With the ARTECHNE database project drawing to a close, the database is still used frequently by a variety of users. Former and current project members consult it on a regular basis to retrieve recipes, information on the nature and use of specific materials and techniques, and to compare textual sources. One project partner included links to her personal dataset, which was integrated in the ARTECHNE database, in her book, which means that readers will continue to consult the data. Other users with very specific questions about artistic materials and techniques show an interest too. For

example, an art historian who is interested in the role of smelling and scents in artistic practices, will find that the database contains 26 references to smell in various languages. Finally, the ARTECHNE database is used in education, to discuss the opportunities and challenges of databases in art history with research master students at Utrecht University.

1 November 2019, the ARTECHNE database fully migrated to the servers of the RKD, where it will become accessible in the near future, and both dataset and code are now available at DANS and [Zenodo](#). For the foreseeable future, the [database](#) will also remain accessible via its original homepage on the servers of Utrecht University. The future will tell if it will remain a useful resource for those interested in technique in the arts in (early) modern Europe.

Reflections

Leading this collaborative database project for nearly four years has given me a chance to reflect critically on how much digital literacy on the part of (art) historians and other humanists is required to successfully set up research projects relying on new technologies, and what I would advise interdisciplinary research project teams considering the development of a project database.

Regarding the first point, as a historian you do not need to become a data scientist, but it is essential that you have a basic grasp of concepts, methods and techniques. Vice versa, a programmer who wants to work with humanities scholars needs a basic grasp of *their* concepts, methods and techniques. An open mind and a willingness to learn from each other, to grasp each others' way of "seeing the world" and the possibilities and limitations of the others' discipline will help you to work constructively together. Compare it to two people with different native languages trying to cook together: you do not need to speak each others' language fluently, but a basic (or preferably even an intermediate) grasp definitely makes things easier, will prevent big misunderstandings, and will equip both of you and the project better for the future.

Using digital methods in the humanities will not work in the long run if humanists see it as some sort of "programming on demand" model, because this will most likely lead to narrowly applicable tools which will not survive the project they have been designed for. Similarly, scientific programmers willing to collaborate with humanists will have to delve into the questions they ask and how they try to answer them in order to come up with approaches that are beneficial for all parties. If time and budget allow for it, the role-based collaboration model with a designated fulltime data steward suggested by Berg-Fulton et al. can be a fruitful approach (152). However, as the ARTECHNE project database shows, even with limited resources, it is possible to create a sustainable project database that is a useful resource within the project and beyond.

To those considering a similar undertaking, my first advice would be: read up and talk to data specialists. Ask questions, educate yourself. There are plenty of free online resources that can help you

to quickly develop a rudimentary understanding of data science methods and questions. The trilingual [Programming Historian](#) website is a great starting point for humanities scholars from all historical disciplines. To get an idea of what is out there in terms of DH projects, check websites such as that of the *European Association for Digital Humanities* and that of the [Office of Digital Humanities](#) at the National Endowment for the Humanities. Consider taking an introductory massive open online course (MOOC) on programming or developing databases. Consult your local Digital Humanities Lab. This should be sufficient to get you started when you engage in conversation with data scientists.

In addition to those first steps, remember that the quality of your data and the flexibility of your ontology are everything, both in terms of usability within your envisioned project and in terms of sustainability. From the start of your project and throughout the process, think and talk about the standards your data have to adhere to in order for you to be able to search and analyze them optimally, how you put them in your ontology, and how you do all of this in such a way that they can be reused and integrated with other data, either right away or in the future. Once you have a clear idea of where you are heading with your DH project, continue to educate yourself. The [Digital Humanities Summer Institute](#) is obviously a great place to do so, but even if you are not willing or able to attend, as I argued in the previous paragraph, there are enough opportunities.

As the Artechne database project shows, even if all partners have the sustainability of the dataset and interface in mind throughout the project, you will encounter unexpected challenges. This makes developing a multi-purpose database for an interdisciplinary project difficult, but also fun and rewarding.

Works Cited

Berg-Fulton, Tracey, Alison Langmead, Thomas Lombardi, David Newbury, and Christopher Nygren. "A Role-Based Model for Successful Collaboration in Digital Art History." *International Journal for Digital Art History*, vol. 3, 2018, pp. 152-179. <https://doi.org/10.11588/dah.2018.3.34297>.

Boulboulé, Jenny. "Drawn up by a Learned Physician from the Mouths of Artisans. The Mayerne Manuscript Revisited." *Netherlands Yearbook for History of Art / Nederlands Kunsthistorisch Jaarboek*, vol. 68, 2019, pp. 204-249.

Dupré, Sven. "The Role of Judgment in the Making of Glass Colors in the Seventeenth Century." *Ferrum*, vol. 90, 2018, pp. 26-35.

Drucker, Johanna. "Performative Materiality and Theoretical Approaches to Interface." *Digital Humanities Quarterly*, vol. 7, no. 1, 2013, pp. 1-43.

<http://www.digitalhumanities.org/dhq/vol/7/1/000143/000143.html>.

Hagendijk, Thijs. “Learning a Craft from Books. Historical Re-enactment of Functional Reading in Gold- and Silversmithing.” *Nuncius*, vol. 33:2, 2018, pp. 198-235.

Hendriksen, Marieke M.A. “Casting Life, Casting Death: Connections Between Early Modern Anatomical Corrosive Preparations and Artistic Materials and Techniques.” *Notes and Records, The Royal Society Journal of the History of Science*, vol. 73:3, 2019, pp. 369-397.
<https://doi.org/10.1098/rsnr.2018.0068>.

---. “Art and Technique Always Balance the Scale”: German Philosophies of Sensory Perception, Taste, and Art Criticism, and the Rise of the Term Technik, ca. 1735–ca. 1835.” *History of Humanities*, vol. 2, no. 1, 2017, pp. 201-219.

---. “Google NGram for Early Modern History?” *The Medicine Chest*, 2016
<https://mariekehendriksen.nl/2016/03/13/google-ngram-for-early-modern-history/>.

Hitchcock, Tim. “Confronting the Digital. Or how academic history writing lost the plot.” *Cultural and Social History*, vol. 10, no. 1, 2015, pp. 9-23. <https://doi.org/10.2752/147800413X13515292098070>.

Kálmán, Tibor, Matej Ďurčo, Frank Fischer et al. “A landscape of data – working with digital resources within and beyond DARIAH.” *Int J Digit Humanities*, vol. 1, no. 1, 2019, pp. 113-131.
<https://doi.org/10.1007/s42803-019-00008-6>.

Pinto, Mariana. “Taking Paint Samples for Pigment Analysis in Nineteenth-Century England.” *Studies in Conservation*, 2018, <https://doi.org/10.1080/00393630.2018.1550612>.

Acknowledgements

The research for this paper was funded by the European Research Council under the European Union’s Horizon 2020 Research and Innovation Programme (Grant Agreement No. 648718), an Utrecht University Digital Humanities Grant, and a Renaissance Society of America travel grant for the Digital Humanities Summer Institute 2018.

Footnotes

1. Many thanks to the entire project team and their collaborators for their contributions to the ARTECHNE database over the years: Sven Dupré, Jenny Boulboulé, Thijs Hagendijk, Mariana Pinto, Jill Briggeman, Roland van Blokland, Jurjen Munk, Deborah van den Herik, Maartje Stols-Witlox,

Martijn van der Klis, Alex Hebing, José de Kruijf, Anita Hopmans, Reinier van 't Zelfde, Erik Samson, Anna Carlgren, and Suzanne Bul. [↵](#)

2. In computational linguistics, n-grams are also used for probability predictions. [↵](#)

3. Google has used the data it accrued through the digitization of books to develop an excellent OCR tool, Google Vision, which has an impressively high accuracy in reading early modern print. However, unlike the scanned images and low-quality OCR freely available online in Google Books, Google Cloud Vision AI is a commercial product. See <https://cloud.google.com/vision/pricing> (last consulted 31 July 2019). [↵](#)