Digital Resources and Tools in Historical Research

Lars Wieneke and Gerben Zaagsma
Luxembourg Centre for Contemporary and Digital History
https://www.c2dh.uni.lu

Abstract.

This proposal will discuss the use of digital picture archives and associated tools in historical research from the perspective of digital history with a focus on resources for the history of education. Our starting point will be threefold:

- digital picture archives need to be seen as part of a wide range of digital resources that are currently available for historical research; while certain methodological, epistemological and technical questions are specific for digital picture archives, many pertain to digital resources in general;
- similarly: discussing the prerequisites for a picture archive on educational history implies, first, to address the broader question of what prerequisites should be considered for digital archives more generally; and addressing the question of whether or not such general prerequisites can be formulated at all given the wide range of research questions and use cases researchers bring to the table

With this broader contextualisation in mind we will focus on the possibilities and limitations of digital picture archives for the history of education through a brief discussion of the following points:

- what are the characteristics of digital picture archives; technical and otherwise, and within that context, digital archives for the history of education; can specific characteristics be observed with regard to the latter or not?
- what layers of information are currently embedded in digital picture archives for the history of education (taking Gerhard Paul's differentiation as a starting point); how can we improve the design, annotation and classification of images through metadata etc, to build corpuses that meet researcher's requirements?

Introduction

Our paper will have to start with a confession: neither of us are historians of education. Both of us, however, have a keen interest and experience in what is called "digital history" and hence in the use and potential of digital resources in historical research. This workshop presents us with an excellent opportunity to consider a particular case study, the visual history of education, and think through how, and in what ways, digital resources are already used and could advance the field further.

In our paper we will discuss the use of digital picture archives and associated tools in historical research with a focus on the possibilities and limitations of digital picture archives for the history of education. Before doing so some contextual parameters need to be established:

First of all: digital picture archives need to be seen as part of a wide range of digital resources that are currently available for historical research; while certain methodological, epistemological and technical questions are specific for digital picture archives, many pertain to digital resources in general;

Secondly: discussing the prerequisites for a picture archive on educational history implies, first, to address the broader question of what prerequisites should be considered for digital archives more generally; and addressing the question of whether or not such general prerequisites can be formulated at all given the wide range of research questions and use cases researchers bring to the table.

The call for papers for this Pre-Conference workshop announced a focus on "the impact of the discipline on developing and maintaining of a picture archive" and listed a number of pertinent questions. For the purposes of our talk we have reframed these questions as three major topics to be addressed, and added some of our own concerns:

Prerequisites:
what prerequisites, if any, are there for a picture archive on educational history? Are there common basic requirements?

Existing data archives: in how far can existing data archives meet the needs? What do they offer and how are they used? What is their strength and weakness in regard to the analytical possibilities they offer?

Potential: Can existing data archives meet the demands of the visual history of education? Is there a need for another solution? And what potential do new technological approaches, for example from the field of Computer Vision, offer?

The visual turn in the digital humanities

Before addressing the issues above, a very brief word on the visual turn in the digital. The Digital Humanities are traditionally text-based and the engagement with images, as more than digitised artefacts, is only recent. As Patrik Svensson wrote as recently as 2009: "The so-called "visual turn" or research on multimodal representation does not seem to have had a large impact on humanities computing."[1] Many digitisation projects of the 1990s focused on textual materials and text editions. Large scale digitisation of images is a development of roughly the past 15 years. Gerhard Paul talks about the "the technological quantum leap of the world wide web" as a result of which "historians have had completely new possibilities of image research at their disposal for merely the last ten years."[2]

The question is of course how that potential has been used and to what extent it has been realised. For our purposes a further question is whether digital picture archives are merely used as repositories of visual material and images, now easily accessible and in much larger quantities than before, or if new digital methods are used to actually analyse them. As to the latter: recent developments in Computer Vision are now beginning to offer exciting new possibilities. Using neural networks and other techniques rapid advances are being made in visual pattern discovery, image recognition and classification[3] As a result, it is now possible to analyse large image data sets and categorise them[4] to a certain extend. If all that sounds somewhat deterring to many historians, a creative use of metadata can already yield interesting research results for those willing to invest in some technical expertise.

Prerequisites

Let's now turn our attention to the impact of the discipline on developing and maintaining of a picture archive. A first question is what prerequisites, if any, exist for a digital picture archive on educational history? This is a problematic that relates to the more fundamental question what the demands of the visual history of education

1 Patrik Svensson, Humanities Computing As Digital Humanities. DHQ: Digital Humanities Quarterly 3/3 (2009).

2 Gerhard Paul, 'Visual History', Version: 1.0, in: Docupedia-Zeitgeschichte, 11.02.2010. URL: http://docupedia.de/zg/paul_visual_history_v1_de_2010. DOI: http://dx.doi.org/10.14765/zzf.dok.2.557.v1

3 See for a succint overview of recent developments the DH2017 workshop proposal on Computer Vision in Digital Humanities: https://dh2017.adho.org/abstracts/639/639.pdf.

4 See for a description of a very recent example this DHBenelux 2018 abstract: Seeing History: Analyzing Large-Scale Historical Visual Datasets Using Deep Neural Networks. http://2018.dhbenelux.org/wp-content/uploads/sites/8/2018/05/Wevers_Smits_Seeing_History_DHBenelux2018.pdf.

should look like and can be approached from either a content or technical perspective. As to content, researchers and heritage institutions should jointly decide digitisation priorities, ensuring that what is being digitised represents a broad spectrum of relevant topics for potential future research (avoiding the pitfall that what is being digitised only reflects existing master narratives).

From a technical perspective other factors come into play, related to content: First of all the size of the collection affords different approaches when dealing with hundreds, thousands or millions of images. Furthermore the technical provenance of the images (raster images, photographs, etc.) comes into play while resolution and file size of the scanned images require different strategies depending on the size of the collection. And finally possible research questions and topics require certain technical possibilities, for example, certain types of metadata to allow a researcher to find relevant materials. In the case of images, high-quality and consistent metadata are of crucial importance as they provide, at least so far, the only way to find relevant non-textual materials (as compared to being able to perform a full-text search in OCR'ed textual materials). Conversely, new technological possibilities can help to open up new avenues of research and generate new research questions. Here one can think about interlinking materials from various repositories, for example through the so-called International Image Interoperability Framework[5].

State of the Art

Let's consider a couple of example of relevant digital picture archives for the history of education. Pictura Paedagogica Online is the BBF's digital picture archive, Historywallcharts is a collaborative project offering history wallcharts from Germany, The Netherlands and Denmark; and DigiPorta is a digital portrait archive. These archives differ considerably when it comes to search and browsing options, extent and quality of metadata, possibilities to save and/or export found objects and their metadata, etc. As image repositories they function well but there is much room for improvement, especially when search options and quality of metadata are concerned. One factor to keep in mind here is that the migration of data from legacy websites to newer more state of the art content and or asset management systems is costly. In many cases, the question then is how existing databases can be improved until funding is secured for entirely new solutions.

5 See https://iiif.io

Potential

Can existing digital picture archives meet the demands of the visual history of education? This, of course, all depends on how one formulates these demands. To provide an example: suppose we wanted to conduct a comparative wallchart analysis of the depiction of World War II in Germany, Denmark and The Netherlands based upon the collection in http://historywallcharts.eu/. This can certainly be done, yet it requires quite some time as there is no advanced keyword search that would allow us to retrieve all relevant images at once and/or per country; moreover, one needs to search using multiple languages to obtain all possibly relevant results. In this particular example the main point to address would be the quality and consistency of metadata. A different approach would be to use IIIF to interlink the original databases the wallcharts come from, obviating the need for a new application that brings them together in a new database.
Websites like DigiPorta allow users to export metadata, but only for individual records. If this could be done for all relevant records that a search yields the options for research could be significant. Consider the following example, a description by Dutch historian Martijn Kleppe of his research into iconic images used in Dutch history textbooks:

> "This presentation will focus on the methods applied to establish which photos can be called iconic. One of the characteristics of iconic photos is the repetitive publication of the same image. We therefore made an inventory of all photos that were published in Dutch High School History textbooks during 1970 – 2000. A total of 412 books have been analysed and 5.395 photos were identified. All the photos were digitised and added into a database, using software package Fotostation Pro. A total of 42 variables containing information about the photo and the textbook were written down and saved in the Exif file of each photo. This enabled the researcher not only to 'read' the information in different types of Photo- editing and viewing software but we could also export the data into statistical software like SPSS, enabling us to calculate which photos were used most often, resulting in a list of most used photos."[6]

If we move to the realm of computer vision technologies we have other options. Apart from using, for example, deep learning approaches to determine the type of image we are looking at (a drawing, photo, engraving, etc) we could look for all kinds of categories of interest, such as depictions of war, cities, cars, animals etc.

6 Martijn Kleppe, 'Photographic Icons – Building and researching large-scale photo collections', Brainstorm Meeting – e-Humanities: Innovating Scholarship (29 March 2011, NIAS Wassenaar). URL: https://www.ehumanities.nl/v02/beheer/wp-content/uploads/2011/04/Booklet-e-Humanities-Meeting1.pdf.

Concluding remarks

The above was only a very short exploration of digital picture archives in historical research and the visual history of education. What is clear to us is that a multilayered strategy is necessary to realise the potential of digital picture archives for the visual history of education more fully. Enriching and improving the quality and consistency of metadata of existing repositories is one important approach as is exploring what improvements in browsing and advanced search options could be implemented. As to new solutions, migration to more modern systems is costly but of course preferred. Interlinking repositories, through IIIF and Linked Open Data should be part of such an effort.

For research, a way to export metadata of search results is crucial to open up more possibilities for digital historical analysis with some of the existing repositories. Nonetheless, whereas the design and technical possibilities embedded in digital archives obviously shape and constrain what researchers can do with the materials located within them, a researcher's creativity, imagination and willingness to experiment are equally important. In the end, though, we have to return to the question posed in the beginning: what are the specific requirements in the history of education from the perspective of its researchers? Only by formulating these can we hope to build corpuses that meet researcher's requirements.