

Connecting Collections - Keeping Multiperspectivity

In this paper we reflect on digitally connecting transnational collections for research. The main issue we want to address is the paradox resulting from attempts to datify, model and process sources from diverse origins to a new, digital data source and at the same time to do justice to the myriad of perspectives in the sources.

Our point of reference is mainly our long-standing research project *Migrant, Mobilities and Connection* (MIGRANT). The objective of MIGRANT has been to connect collections and to use digitization as a solution to the loss of migrant heritage (Arthur et al 2018). Our proposed life courses approach leverages serial registration systems for migrants as a central connecting backbone that follow them across borders while the collections stay in place. This makes it possible to connect traces of people and events in different collections in the home and destination countries as well as international archives.

This approach is not much different from the silo breaking agenda (Hotson&Wallnig 2019, Ahnert&Ahnert 2019, p.450, DHBenelux 2024 Call for Papers)

Now we are almost ten years on in the project and we have become increasingly aware that the idea that silos can be broken by digitally connecting was too naive and should be problematized. The crux of the problem is in the conversion and connection of collections, not so much on the technical but on the methodological side. Henceforth we will use 'collections' rather than 'silos', as they better reflect the reality of the heritage world.

The serial life courses approach has proven a useful way to connect different collections, as it will connect traces of (the same) migrant in many different archives and allows us to take into account the myriad of different perspectives on migrants and migration the collections represent. To construct life courses, we need a data pipeline connecting data and digital objects.¹

However, it is important to realize that life courses are (re)constructions that use a formally defined model and abstractions to link event descriptions from collections for migrants. The main pitfall is to assume that collections contain events from migrant lives. Collections do not contain data about migrant lives directly; all heritage collections represent, usually institutional, views on migrants. All collections therefore contain varying perspectives making it necessary to deal with multiperspectivity. Converting such historical sources to a digital representation involves processing

¹ examples at the site <https://migrant.huygens.knaw.nl/home/online-tentoonstelling/>

steps, interpretation and abstractions that lead to possible blind spots for researchers (Hoekstra&Koolen 2019).

Methods of source criticism

The main sources of the serial life courses are the different registration systems that recorded parts of the migrant journey: from the selection process in the Netherlands, through the voyage aboard migrant ships or planes to the arrival in Australia and subsequent housing in migrant camps and support by the Dutch consulates that are kept in four different archives and six different collections in the Netherlands and Australia

These related systems (van Faassen 2020), all partly follow post-World-War II Dutch-Australian migrants, but that does not mean they concern the same people or record the same data about them. The main questions that had to be answered were:

- what exactly do these systems record,
- how uniform or hybrid is the information contained in them
- what is the relation between the different collections,

In the course of the project, we adopted a combined method of analog source criticism extended with digital evaluation procedures. The evaluation of our main migrant registration card system, for example, was only feasible with a combination of sampling and a digital assessment of the information distribution over the collection. Whereas sampling gives a summary of the content, the assessment of the information density gives an overview of the whole collection. (Van Faassen & Hoekstra 2022, 2023). Ideally, such source criticism should be applied to each of the registration systems used for connecting.

There were several interlocking areas of complications, that we outline below.

Datification flattens data

Working with the files from various collections representing different stages of the registration of migrants, showed how much the perspective of administration officials left its mark on the different collections. Obviously, these represented different selections of the migrants, but they also seemed to contradict each other.

- For instance, while usually the emigration date was undisputed, there are cases when the limited time scope of one registration system made it seem that a migrant family did not leave, but data about the same family from another registration system revealed that they did leave a few years later, but under another organized migration program.
- A second example is that our main registration system gives migration dates and ships or planes a migrant traveled on. However, when we compare these dates with the reconstructed list of actual voyages of ships, it is clear that the

variation is much too large. Sometimes this is because of administrative or processing mistakes but they can usually be attributed to a change in travel schedule, alternately registering the scheduled or actual date of departure, or date of arrival in one of the Australian ports.

These examples show how seemingly straightforward data are also constructions. Once they end up in a compound data set they tend to be viewed as fixed, especially if they are combined with other data. The issue is that in order to create a data set by connecting data, we have to link the source data by eliminating the variation in them, but this inevitably also throws away information about different interpretations. The resulting data set is indispensable for analysis that could never be conducted by hand, but every choice to reconcile contradictory data reduces multiperspectivity. Therefore, we argue that serial data analysis should be complemented by critical analysis of the differences between sources (cf Van Lange 2023, 239-240).

Selection biases

Selecting the collections required previous selection by us, the project Migrant research team. With a background as publishers of historical sources and the traditions of external and internal source criticism, we realized we needed a collection overview (Van Faassen 2014a-b) as archives grew out of the administrative processes of the institutions involved. Policies and implementation change over time and influence each other. Institutional hierarchy, policy and personal relations and communications connect these different collections, but these relations are not recorded within all collections even if they bear their stamp. When governmental document collections are transferred to archive institutions, archive laws and policies further complicate the situation. In MIGRANT, part of the policy files were digitized in our collaboration with archival partners. However, it is difficult to turn them into data that integrate with the registration systems. But it is also indispensable as interpretative aids for datasets that cannot say anything about themselves.

Selections influence data creation at many levels. Some examples:

- Collaborative endeavors between the Dutch and Australian governments to craft a cohesive narrative on migration are manifest in the propaganda photo collection centered around the 100,000th migrant, Adriana Zevenbergen. This comprehensive collection, extending to the minutiae of photo captions, resides in both National Archives. There are several ways in which the paradox in which digitization makes the archive material usable outside their original context but leads to the loss of perspectives:
 - Only the Australian collection has been digitized making the shared policy effectively invisible to researchers
 - Collections featuring high-quality images, such as those portraying Zevenbergen, undergo digitization without contextual information. Consequently, these digitized materials find their way into a general search engine, now the primary access point. Selecting materials using

the search engine revives old propaganda, presenting a constructed image instead of the diverse and multiform historical reality

- Shifts in policy often cause changes in registration that can easily mislead researchers. For instance, quantitative analysis showed sudden increases in housing issues in the early 1960s and the remarkable spike of financial data for youths. These were not indicative of increasing problems, but rather reflections of changes in administrative priorities (Van Faassen & Hoekstra 2022).

Expanding upon these instances would undoubtedly reinforce our core argument that selection of sources at any level should be informed by knowledge of the context influencing the formation of the collection. Otherwise, even seemingly straightforward results may merely be artifacts shaped by selection biases on the part of the researcher, or rooted in the sources and their organization. It is usually difficult to account for them in a data set. Digitizing and connecting collections, however, remains both inevitable and desirable, but comes with all sorts of unforeseen side effects. Especially the pretension that the separation between collections can be removed or broken without question should be viewed with great reserve.

We do not pretend to be able to solve the paradox, as we ourselves continue to be confronted with it many times in project Migrant. But we do believe that there are major methodological challenges in balancing datafication and multiperspectivity by trying to better account for context and selection. This starts with a more general awareness of the paradox.

References

- Ahnert, Ruth and Sebastian E. Ahnert (2019), 'Networking the Republic of Letters', in Hotson & Wallnig, *Reassembling*, pp. 399–410
- Arthur, Paul, Jason Ensor, Marijke van Faassen, Rik Hoekstra, and Nonja Peters, 'Migrating people, migrating data: Digital approaches to migrant heritage', *Journal of the Japanese Association for Digital Humanities*, 3 (October 2018), pp. 98-113
- Hotson, Howard and Thomas Wallnig (2019), "Introduction", in: Hotson & Wallnig, *Reassembling*, pp. 7-22
- Hotson, Howard and Thomas Wallnig, eds. (2019), *Reassembling the Republic of Letters in the Digital Age: Standards, Systems, Scholarship*, Göttingen
- Faassen, M. van (2014a) *Polder en Emigratie. Het Nederlandse emigratiebestel in internationaal perspectief 1945-1967* (Den Haag)

Faassen, M. van (2014b), *Research Guide Emigratie 1945-1967*
<https://resources.huygens.knaw.nl/emigratie>.

Faassen, M. van, Oprel, M. (2020), 'Papertrails to private Lives. The performative power of card indexes through time and space'. in: Ida Nijenhuis et al. (eds) ,*Information and Power in History. Towards a Global Approach* (Routledge), pp. 254-274

Faassen, M. van, Hoekstra, R. (2022), 'Migrant Visibility: Digitization and Heritage policies', <https://www.frontiersin.org/articles/10.3389/fhumd.2022.908456/full>

Hoekstra, Rik and Marijn Koolen (2018), 'Data scopes for digital history research', *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, pp. 1–16.

Lange, M. van (2023), *Emotional Imprints of War: A Computer-Assisted Analysis of Emotions in Dutch Parliamentary Debates, 1945-1989*, (Bielefeld)
<https://doi.org/10.14361/9783839464854> .