Toward a Model for Digital Tool Criticism: Reflection as Integrative Practice[1]

Authors:

Marijn Koolen (marijn.koolen@huygens.knaw.nl)
Royal Netherlands Academy of Arts and Sciences - Humanities Cluster

Jasmijn van Gorp (J.vanGorp@uu.nl)
Utrecht University - Department of Media and Culture Studies

Jacco van Ossenbruggen (Jacco.van.Ossenbruggen@cwi.nl)
Centrum Wiskunde & Informatica, VU University Amsterdam - Network Institute

1 Abstract

The past decade, an increasing set of digital tools is developed with which digital sources can be selected, analyzed and presented. Many tools go beyond keyword search and perform different types of analysis, aggregation, mapping and linking of data selections, which transforms materials and creates new perspectives, thereby changing the way scholars interact with and perceive their materials. These tools, together with the massive amount of digital and digitized data available for humanities research, put a strain on traditional humanities research methods. Currently, there is no established method of assessing the impact of the digital tools deployed in a specific digital research trajectory. There is no consensus on what questions researchers should ask themselves to evaluate digital sources beyond those of traditional analog source criticism. This article aims to contribute to a better understanding of digital tools and the discussion of how to evaluate and incorporate them in research, based on findings from a Digital Tool Criticism workshop held at the 2017 Digital Humanities Benelux conference. The overall goal of this article is to provide insight in the actual use and practice of Digital Tool Criticism, offer a ready-made format for a workshop on Digital Tool Criticism, give insight in aspects that play a role in Digital Tool Criticism, propose an elaborate model for Digital Tool Criticism that can be used as common ground for further conversations in the field, and finally, provide recommendations for future workshops, researchers, data custodians and tool builders.

[1] This is the author copy of: Marijn Koolen, Jasmijn van Gorp, Jacco van Ossenbruggen; Toward a model for digital tool criticism: Reflection as integrative practice, *Digital Scholarship in the Humanities*, 12 October 2018, https://doi.org/10.1093/llc/fqy048

2 Introduction

The past decade, an increasing set of digital tools is developed with which digital sources can be selected, analysed and presented. Many tools go beyond keyword search and perform different types of analysis, aggregation, mapping and linking of data selections, which transforms materials and creates new perspectives, thereby changing the way scholars interact with and perceive their materials. These tools, together with the massive amount of digital and digitized data available for humanities research, put a strain on traditional humanities research methods. Currently, there is no established method of assessing the impact of the digital tools deployed in a specific digital research trajectory. There is no consensus on what questions researchers should ask themselves to evaluate digital sources beyond those of traditional analog source criticism.

While source criticism is common practice in many academic fields, the awareness for biases inherent in digital tools and their influence on research tasks needs to be increased. When it comes to the criticism of data or sources, source criticism is an established method for historians and humanities scholars. The literature in the humanities on source criticism are primarily aimed at analogue research, but not yet up to date with digital research in the heritage domain. Lara Putnam describes the shift from consulting analogue archives to keyword searching digital archives (Putnam, 2016). Current methods in historical research in physical archives are shaped around leafing through large volumes of materials to identify documents of relevance, with two important consequences. First, the scholar is confronted with the large number of unrelated materials that demonstrates the relative importance of their topic. Second, they are made more aware of what other related and unrelated topics were competing for attention at the time. This prompts the question of how scholars can use digital tools to get a similar understanding of a topic's relative importance and connections with other topics in a digital archive.

Moreover, many digital tools allow scholars to transform, aggregate, count, classify, link and visualize the underlying data. With these modelling steps they further change the materials they are studying. There is as yet little common understanding within and across humanities disciplines of how these steps affect the relation between research questions and materials and how these activities differ from traditional practice in terms of interpreting and contextualizing digital data. Some scholars (e.g. Giuliano, 2017; Underwood, 2014; Gibbs and Owens, 2013) have pointed out the importance of reporting on these parts of the research process to start conversations around how to incorporate them in humanities research. This article aims to contribute to a better understanding of digital tools and the discussion of how to evaluate and incorporate them in research, first by reporting on two experiments held during a workshop at the 2017 DH Benelux conference[2] with participants of different Digital Humanities backgrounds, and, second, by synthesizing the theoretical background of the workshop with a review of

[2] Digital Humanities in the BeNeLux conference: https://dhbenelux2017.eu/

relevant literature and an analysis of the workshop outcomes. We aim to formulate a set of assessment criteria (or building blocks for the conceptualisation) of digital tool criticism. At the workshop we invited the participants to experiment with tools and explicitly asked them to question and criticize the tools at hand. The overall goal of this article is to provide insight in the actual use and practice of Digital Tool Criticism during the workshop and more specifically:

1) Offer a ready-made format for a workshop on Digital Tool Criticism, including assignments, tools and methods for analysis, that can be reused for training and education (cf. section 3)
2) Give insight in all aspects, both reported during the workshop and deriving from our own discussions, that play a role in Digital Tool Criticism (cf. section 4)
3) Propose an elaborate model for Digital Tool Criticism that can be used as common ground for further conversations in the field (cf. section 5)
4) Provide recommendations for future workshops, researchers, data custodians and tool builders (cf. section 6)

Different disciplines may use different methods and may evaluate and reflect on digital tools differently, so there may not be a single common understanding of how digital tools fit in scholarly practice. But we think that a workshop with participants from diverse disciplines, working on the same semi-structured assignments, openly discussing their findings and reflections, and focusing on the exploratory phase in which scholars design their research around questions, materials and methods, is a good starting point for developing meaningful and shareable ways of doing digital tool criticism.

3 Literature on digital tools and their impact on research

In information science, research practices of humanities scholars have been often object of research. The research cycle of social sciences is characterized by Bhattacherjee (2012) and Kendall (2012), while the research cycle of humanities as an iterative process, that continuously revisits all phases (Marshall and Rossman, 2010). Bron et al (2015) distinguish three research phases in media studies research: exploration, contextualisation and presentation. In our conceptualisation of digital tool criticism, it is important to relate the tools and assessment criteria to the phase of research.

If we look at the literature on digital tool criticism, the majority of it can be situated at the *first phase of research: exploration*. Most of the literature that discusses the use of digital tools in humanities scholarship focuses on search interfaces around digital collections. Timothy Burke lists a number of recommendations for scholars to guide their discovery and exploration in digital collections (Burke, 2011). In the exploratory phase, they should exploit the quick

responses of keyword search systems to rapidly iterate through multiple keyword searches, with which they can explore the viability of the collection and the search interface for their research. For this initial phase, simple interfaces should be preferred over advanced interfaces, as the latter require some expertise of the collection, how it is structured and how the search system makes use of that structure to organize search results. Scholars should consciously develop heuristics to evaluate and make sense of search results lists, and develop strategies to gather sets of keywords. We follow this recommendation, by requesting our participants to take notes during their research practice. Another aspect according to Burke (2011) is assessing the quality and authority of found results, which touches on source criticism, but through the lense of digital tools. In our workshop we explicitly asked participants to reflect on this relation between tool and source criticism.

Huistra and Mellink (2016) provide a critical discussion of full-text searches on historical newspaper archives, specifically the Dutch National Library's newspaper database, and offer three recommendations on how to conduct different types of searches to achieve different types of goals. They formulate as advice a.o. that scholars to keep track of and report the steps they took to select their sources, including which search tools were used, and which queries and filters, to retrieve those sources. Moreover, they write that scholars should discuss these steps with colleagues across disciplines to reach a better understanding both of how these digital technologies influence their research practice and how they can or should adjust their practice when incorporating these tools. This recommendation is incorporated in our workshop format by bringing together researchers with different backgrounds.

Although search may seem a well-understood finding aid, there are many subtleties that scholars should take into account, and introduces experimentation as important element of the research process (Gibbs and Owens, 2013; Underwood, 2014). Gibbs and Owens (2013) argue that scholars should make their data interactions transparent to explain how these interactions contribute to making sense of the historical record. Keyword searches are effective finding aids, but many digital archives and libraries offer additional sense-making tools to get a better understanding of what a digital corpus contains and does not contain and how it is structured, with which scholars can critically evaluate the archive as a whole. These can be indices of topics, persons or periods, faceted classifications based on various metadata fields, timeline visualizations, and documentation that provide details on selection criteria, data formats and search functionalities.

Jennifer Guiliano (2017) argues a move toward recognized methodologies for digital sport history. 'For every affordance the personal computer could offer, as many problems and limitations would be introduced to the practice of research' (p. 147). Similar to Gibbs and Owens (2013), she mentions experimentation with digital tools as an important part of digital scholarship. She illustrates this with an example of using text mining on digital archives of 19th century newspaper. Automatic sentiment analysis using algorithms trained on modern social media data such as tweets, blogs and online user reviews might give unusable results. Adjusting the algorithms by training on 19th century newspaper articles or trying different algorithms that better fit that genre of texts constitutes a form of experimentation that Guiliano considers a core activity (p. 154). We incorporated this recommendation in the workshop by having experimentation as main forma.

In 'Confronting the Digital', Tim Hitchcock argues that the digital makes sources different and there is a need for more than 'being explicit about our use of keyword searching - it is about moving beyond a traditional form of scholarship to data modelling and to what Franco Moretti calls “distant reading”' (Hitchcock, 2013, p. 19). Data modelling is an intellectual activity to determine what elements the data consist of and what these elements represent. When searching through digital collections, scholars should be aware that data modelling has already taken place to make sources searchable, such as indexing of words and phrases for full-text search, or decisions about what to do with metadata that is missing, incomplete or uncertain such as 'circa 1960'.

But scholars also add further layers of data modelling when using digital tools to aggregate, link and visualize data. In ‘Exploring Big Historical Data: The Historian’s Macroscope’, Graham et al (2015) discuss several tools and techniques to analyze large data sets to extract aggregated information that is hard to see by reading and searching. Examples are algorithmic topic modelling to identify what the major topics are in a set of textual documents and which documents cover which topics, or network analysis of how people, places or topics mentioned in metadata records are connected to each other through co-occurrence. To interpret this aggregated information in a meaningful way, scholars need to consider the process by which it was generated, the selection of sources that were included or excluded in the analysis and how the algorithm determines when chunks of data in different documents refer to the same thing. This is regardless of whether they did the aggregation themselves or used information previously aggregated by some tool. Reflecting on the choices that were made for identifying elements of interest in the data (such as topics, keywords or person names) and what alternative choices are possible, can help scholars to consider how the actual choices focus the analysis on certain aspects and pushes others to the background. In our workshop we explicitly asked participants to take these choices into account in assessing their use of tools.

Research by Bron et al (2015) has shown that humanities researchers refine, leave out and change their research questions based on the availability of data and transparency of tools:

> Due to the abundance of material that seems to be available, at first sight a researcher may think that a particular research question can be answered. [...] Another factor are the tools used to gather material. These often lack transparency in terms of how documents are retrieved in response to search terms, which part of a collection is indexed, and which preprocessing steps have been applied, for example, exclusion of a particular field a researcher expected to be present (Bron et al 2015, p. 1553).

This aspect of changing and refining research questions based on tool and data limitations was chosen as a focal point of the workshop assignments, to encourage participants to reflect on this part of the research process.

4 Format of the workshop

4.1 Theoretical working definitions

As first part of the workshop, we provided the participants with a shared theoretical framework. The slides are available online.[3] We are aware that we primed the participants in providing working definitions. We do believe, however, that it is important to start with a common understanding of concepts in order to be able to criticize them and deconstruct them during the experiments and the discussion session.

In the workshop, we focus on the exploratory phase of the research process, in which researchers are determining their goals, shaping their research questions and gathering their materials. To help participants in framing this phase, we let the participants read a text by Trevor Owens (2014) as preparation for the workshop. Trevor Owens argues that researchers can develop their research designs from different starting points, which can be one or more research questions, a collection of research materials, a set of preferred methods, or a specific conceptual framework. The adoption of digital tools affects many aspects of the research, including the research questions, the selection of materials to study and analyse, and the methods employed to study them. Regardless of where the researcher starts, these aspects influence each other, such that making choices to adopt certain methods may prompt the researcher to modify their research questions and materials, and changing the question forces them to reconsider which conceptual frameworks and methods are appropriate. Digital tools mediate between method and materials, such that choosing a specific tool affects what methods are appropriate and what form of materials or data can be used as input for the tool. Indirectly, tool choice thereby affects the research questions and conceptual frameworks. Vice versa, choices in materials, methods and questions affect what tools are appropriate. In practice, the research design and choices are made interactively and iteratively as the researcher explores different ways in which the available materials, methods and tools can be brought together into a coherent and appropriate design.

Owens adopts the research design model from Joe Maxwell (2013) that connects five elements of research design: questions, materials, methods, conceptual framework and validity (see Figure 1). Note that tools are not explicitly mentioned in Maxwell's framework. They are related to, but not the same as, research methods. Methods are modes of inquiry, and tools afford certain modes more than others, so choosing a tool requires reflection on how it affords a method appropriate for a research question. For a certain method there may be multiple tools that are appropriate, to varying extents. Similarly, the data that is used in the inquiry should fit its mode. For the purpose of digital tool criticism, therefore, we provided the participants with a new model (Figure 2). According to us, it is useful to include data and tools as additional aspects of the framework, which are directly connected to methods in an interdependent network. We also

[3] URL: http://bit.ly/2oHsssK

added ‘researcher’ to the model in order to encourage the participants to reflect on their own role and the role of their peers in the research process.

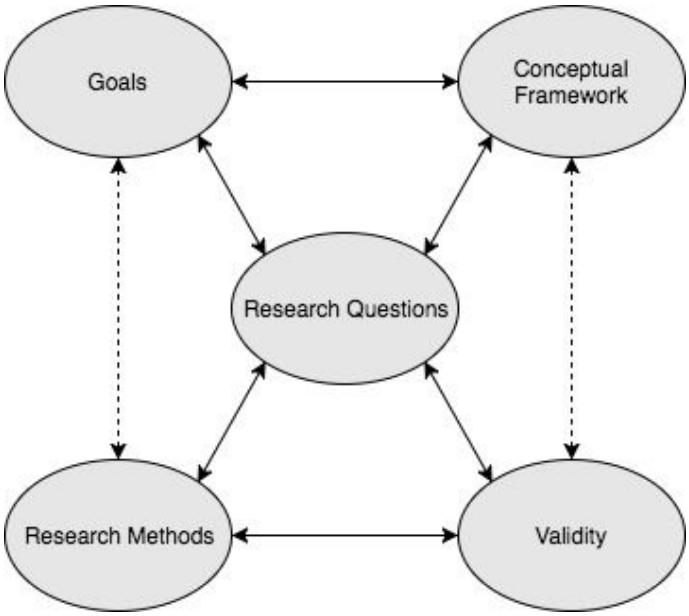

Fig. 1. An interactive Model of Research Design, as developed by Maxwell (2013)

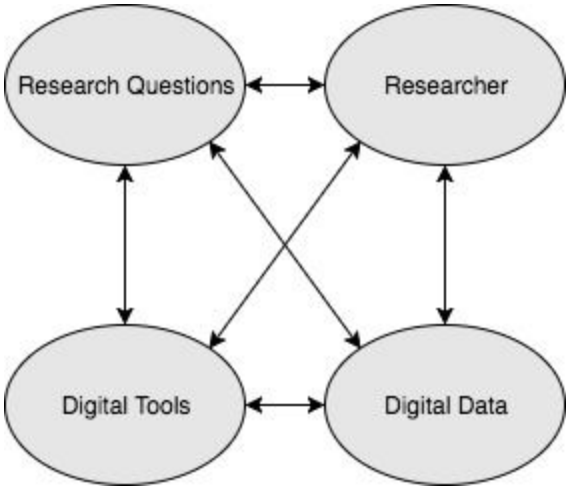

Fig. 2. A model of interdependent concepts of Digital Tool Criticism as made by us and presented to the workshop participants

Besides the theory of Owens and the two models, we also provided the participants with a working definition on 'source criticism' as hook-up for the demarcation of tool criticism.[4] Source criticism is a method or approach common in the humanities and specifically in historical research for evaluating information sources (cf. Fickers, 2012). Internal source criticism focuses solely on the content of the text itself and excludes external aspects. External source criticism, on the other hand, focuses on the metadata of the text, i.e. contextual aspects. Fickers posits five basic questions that are essential for historical source criticism:

- Who created the text?
- What kind of document is it?
- Where was it made and distributed?
- When was it made?
- Why was it made?

We argue it is important to also address the open question whether 'digital' source criticism is different from 'analogue' source criticism and in what way. The same basic questions can be asked of digital sources, whether these sources were born digital or were digitized versions of analogue sources. Tool criticism adds a question for source criticism to the list of five, namely: how was a (version of a) source made? This question can be translated into questions about the tool itself:

- Who made the tool?
- What kind of tool is it?
- When was it made?
- Why was it made?
- How does the tool function?

This prompts further questions, such as: What makes digital tool criticism different from digital source criticism? And to what extent are digital tool criticism and digital source criticism entangled? We added that when thinking about why a tool was made and what it was developed to do, it is important to take into account that it can be and often is used for other things than it is intended purpose.

Before discussing the methodology of the workshop, we also provided a working definition for digital tools. Tools can be studied and evaluated from different perspectives: as research instruments, as methods and as platforms. In the workshop we equated the concept of digital tool with that of computational tool. This can be a tool which is available and used online, that is, the computations are performed remotely on a server that hosts the tool, not locally on the researcher's own computer. A tool can also be software installed locally (such as excel, gephi,...). More specifically, we used the working definition by Van Eijnatten et al. (2013):

[4] This part on digital source criticism is derived from the following book chapter that co-author Van Gorp was writing at the time of the workshop: Van Gorp, J. & de Leeuw, J.S. (2018) Methods of Data Collection with/in Digital Television Archives: Digital Television Historiography. In Van den Bulck H. et al (eds.) *Palgrave Handbook for Media Policy Methods*. Forthcoming

> Digital tools are used in opening up, presenting and curating textual and multi-media sources, in heuristic techniques of retrieval and accumulation of digitised data, in data analysis, in various forms of visualisation and in enhanced and multi-media publications of research results.'

This working definition proved to be a fruitful one as it fits our perspective to link tool criticism to stages of humanities research in the heritage domain.

4.2 Experimental Setup

We based our experimental Setup on recommendations in the existing literature, as elaborated in Section 2. To investigate how tools affect the exploratory phase of research (Bron et al, 2015), we chose a flexible experimental setup in which participants could start from any of the aspects mentioned in Figure 1 and work out a research design that has a research question, a method of investigation and a set of digital sources and tools to investigate.

We wanted the participants to investigate and reflect on the role and impact of digital tools during the exploratory phase, both in establishing a research question and in the selection of digital sources to be used in addressing that research question. Therefore, we ran two short experiments covering different steps in the exploratory research phase, in which participants worked in small groups, and wrote down the steps, choices taken and their findings. In the first experiment they explored data sets and tools to establish a research question, in the second to select appropriate digital data and tools for their research questions. We also decided to have a single research theme that participants were encouraged to adopt to give direction to their exploratory research steps, so that they could compare their findings relatively easily. The topic/theme was ‘migration in Europe’, although they were allowed to ignore this theme and choose their own.

In each part of the experiment, we asked to participants to keep a logbook of their research process, in which to keep track of the chosen goals, framework, questions, methods and their validity during the experiments. We provided them with post-it notes and a Google Document per group to write down any questions they had about the tools and datasets they used, as well as any reflections and insights. We advised them to appoint one person to log considerations, choices, questions, observations. Participants could take screenshots and photos to document their research process. We also encouraged participants to talk out loud and discuss with each other during this process.

We asked participants to think during the experiments about the following questions, related to Fickers’ five Ws:

- Which tools do you use, and why? When do you switch, and why?
- What type of use was the tool intended for?
- Who is the intended audience or user group of a tool?
- What should you know about a tool w.r.t. the access, presentation and transformation of data?

- Do digital tools change our research, and if so, how? in shaping research questions, in selecting or analysing materials?
- To what extent can digital source criticism and digital tool criticism be separated?

After both experiments, participants were asked to analyse their written notes and post-it notes and to create a simple poster to present to the other groups. Specifically, we asked each group to address the following questions. What are most important questions on specific tools and tool use? What are important considerations, reflections and insights? How did the tools you used influence or steer your exploration and analysis?

4.2.1 Data and Tools

We introduced a limited number of digital tools to give participants an idea of what is available and to ensure that there was some overlap in the tools used by multiple groups of participants so we could compare experiences. Again, participants could choose other tools as well so as not to constrain their explorations.

In the workshop we focused on online digital heritage collections, which are many and diverse, and for which different types of tools are available, both tools that are specific to individual collections and tools that are generic and can be used on many different collections. We provided a list of current tools, both generic tools in which data can be imported, and tools that are tied to and built around specific datasets.

Tools for specific datasets:

- Cultural heritage
    - Europeana (https://www.europeana.eu/): a digital platform giving access to heritage collections from more than 3000 European heritage and memory institutions.
    - European Library (http://www.theeuropeanlibrary.org/): gives access to the digital collections of 49 national libraries in Europe. Users can search through 200 million metadata records and over 24 million pages of full-text content.
- Broadcast media
    - EuscreenXL (EU) (http://euscreen.eu/): gives access to European audiovisual heritage, with over 1 million metadata records and over 60,000 media items.
    - Delpher newspaper collection (NL) (https://www.delpher.nl/): a faceted search interface for a range of collections of the National Library of the Netherlands, including 88 million newspaper articles of the Dutch historical newspaper archive, digitized books and journals and radio bulletins.
    - AVResearcherXL (NL) (http://avresearcher.clariah.beeldengeluid.nl/): a comparative search tool that gives access to the Dutch television and radio archive of the Netherlands Institute for Sound and Vision and the Dutch historical newspaper archive offered by the Delpher tool described above. The tool offers two search boxes so users can compare queries. Each search box is connected

to its own search results list and to a combined timeline view that shows the number of search results per year for the two queries.

- Politics:
    - Parliamentary debate search (http://search.politicalmashup.nl/): a faceted and structured querying interface on top of archives of parliamentary debates from seven European countries. Users can narrow the search by political party, party member, and analyse search results through a number of visualisations and aggregations, such as word clouds and timelines.
    - Talk of Europe (http://www.talkofeurope.eu/data/): a platform for querying a Linked Data representation of the same parliamentary debates described above. Users can search the collection using SPARQL queries and download result sets for further analysis in other tools.
    - Migration Flows - Europe (http://migration.iom.int/europe/): a platform that visualizes European data on migration on a geographical map, including migrant registrations, transit routes and relocations, and a map of migrations offices. The site also gives access to the statistical reports on which the visualizations are based.

Generic tools:

- Voyant Tools (https://voyant-tools.org/): an online text analysis tool in which users can create a text corpus by uploading documents or providing lists of URLs. The tool parses the text of documents and offers a range of statistical tables and visualizations for analysis.
- OpenRefine (http://openrefine.org/): a desktop application in which users can upload tabular data and perform data cleaning and aggregation. The tool keeps track of the steps taken, so users can see how a particular view on the data was reached and repeat those steps as a recipe on similar data.
- Digital Methods Initiative tools (https://wiki.digitalmethods.net/Dmi/ToolDatabase)
- Digital Research Tools Directory (https://dirtdirectory.org/): a directory of digital tools that organizes a long list of research tools by type of access and use.

In addition, we encouraged participants to use any tools they know well, such as MS Excel and Google Spreadsheet.

4.2.2 Participants

Participants worked collaboratively in small groups, so that they could share their experiences, ideas and questions regarding data and tools. The workshop was attended by 19 participants. After a short introduction about the workshop, each participant introduced themselves and described their background, experience and expectations. The group was very heterogeneous, representing many humanities disciplines (historical sciences, media studies, literary studies, linguistics, (digital) heritage) and library and information science. Some had little experience with digital tools and digital research, others had years of experience with many different tools

and methodologies. The 19 participants split up into six groups, five groups of three participants each and one of four participants.

4.2.3 Method of analysis

Each group kept notes of their explorations in a Google Document, so it is possible to compare how different groups develop their research questions, how they choose their methods of analysis and make data and tools selections. To analyse the research process in terms of these activities, we categorized phrases in the participants collaborative notes for five aspects, and color-coded the phrases with different colors for the aspects: *Research Question (blue), Method (red), Data (green), Tools (pink),* and *Reflection (yellow).* To visually analyze how groups shift between these aspects, we created versions where we removed white space to collect the notes of a group on a single page. This offers a form of distant reading of these notes that reveals patterns that might otherwise go unnoticed. We call these visualisations 'research-process-visualisations'.

5 Results

5.1 General trends in research processes

We observed that groups interpreted the note taking process differently, with some groups writing down each step in exploring and reflecting in chronological order, while others summarised at the end of each experiment. Even when taken these procedural differences into account, the notes show some interesting patterns. Figure 3 shows the color-coded notes of the 6 groups.

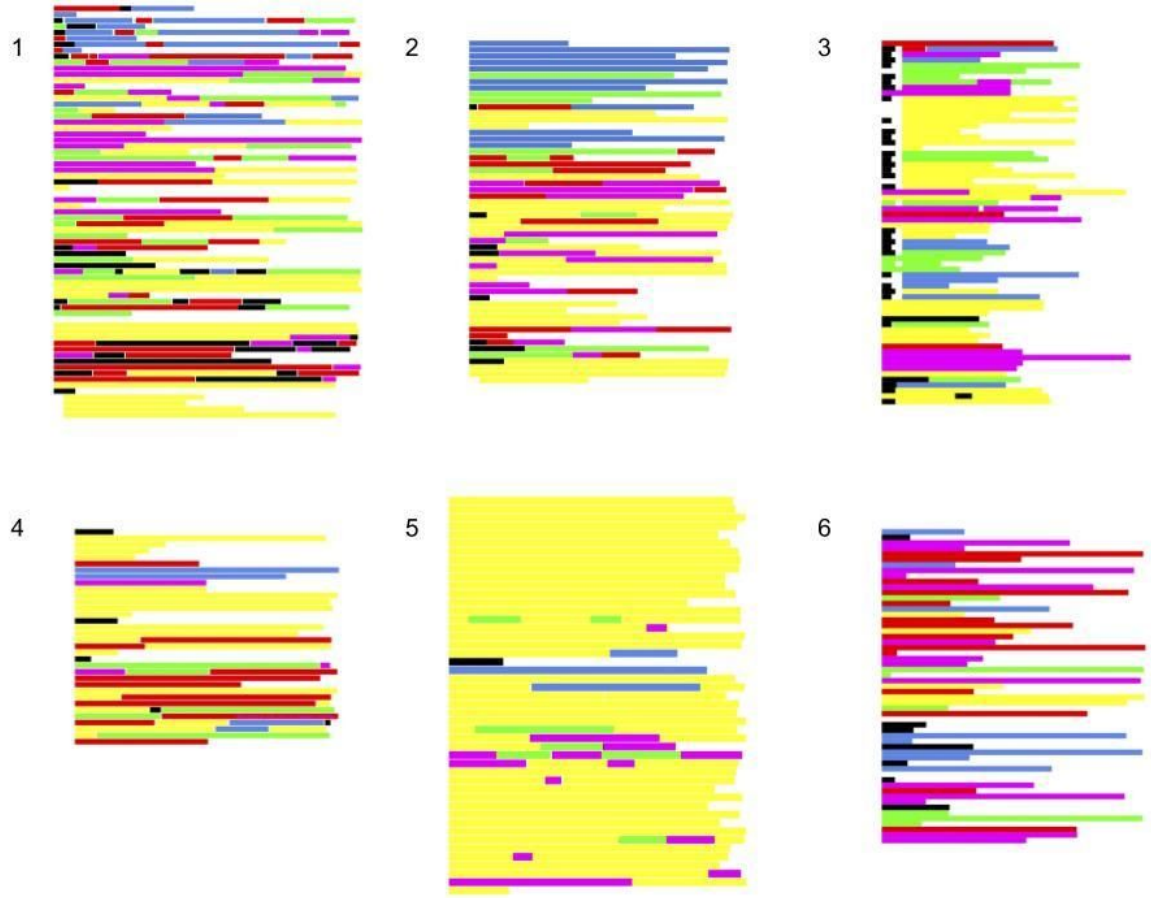

Fig. 3. Research-process-visualisations: research process notes of the six groups, color coded by research aspect. The numbers on the left of the images correspond to the numbers of the groups.

First, the amount of text devoted to critical reflection (yellow) differs from group to group. It dominates the notes of group 5 and is almost absent in group 6. Also, the focus of the discussions around the research topic (blue) is remarkable. Given the two parts of the workshop, with establishing a research question as the explicit task in the first part, one would expect the most blue-coded phrases in the top half of the notes. This is indeed the case for group 1, 2 and 4, but it is clearly in the middle of group 5 and in the second half of the notes of group 3 and 6. This observation is in line with Maxwell's claim that the formation of a research question is an iterative process influenced by multiple aspects of the research design (see Figure 1). We also observed that since many datasets are only available through a specific tool, discussion about data (green) are often mixed with discussions around the associated tool (pink). This observation also supports the idea that digital source criticism and digital tool criticism are hard to separate. Likewise, the functionality of a tool (pink) is often discussed in terms of the research method (red), to the extent that the two become hard to distinguish. This corroborates our earlier claim that tools mediate between data and methods. In some cases it is clear that participants are discussing specific aspects of a tool, such as what features it has or does not have or to what extent they are configurable. In other cases that they are talking about

a method of analysis in general without considering specific types of tools. But in many parts of the notes these aspects blend into each other. This demonstrates that tool and method clearly are interdependent, but should be considered separate aspects in a model of digital tool criticism, as we will elaborate on in Section 5.

5.2 Impact of Data & Tools on Research Question Refinement

The first task given to the six groups was to refine the given theme 'migration in Europe' into a more specific research question. As first steps in this exploratory phase, all participant groups use rapid searches to establish whether a given data set or tool is suitable for a certain line of inquiry and iteratively adjusted questions, tools and data selections until they are aligned enough to warrant further exploration in a specific direction. Once they had established a fruitful direction, they use the same strategy 'to rapidly test and refine questions and hypotheses' (Solberg 2012, p. 64).

Group 1 started with questions around a chosen topic of interest, then looked for About pages of tools to see which ones give access to the data required for these questions. Having found that the Parliamentary Debate Search (PDS) and Talk of Europe (ToE) tools give access to recent materials and promising results based on an initial keyword search, they try several related keywords to get a feel for the extent of the relevant data. Their overall research goal, 'Compare discussion of migration in broadcast media and in European parliamentary debate speeches 1990-2014' was formulated relatively early in the process, and formulated in terms of the corpora of the investigated tools.

Group 2 investigated perception and stereotyping of immigrants and refugees by different political parties, and used keyword searches initially to establish which historical periods best fit this investigation. Once they focused on a specific period, from the Geneva convention in 1951 until 1994 (as more recent newspapers were not available due to copyright), they used ‘pearl growing’ (Drabenstott, 2000; Yakel, 2010) or what Burke (2011) calls ‘keyword harvesting’ as a manual form of topic modeling, to investigate the evolution of terminology around the main topic.

This is also reflected in central role terms play in the research question formulated by this group: ‘In which ways do the terms that are used in newspapers and parliamentary debates to describe immigrants and refugees from distinct nationalities evolve between 1970 and 1990?’.

Group 3 started with the tool AVResearcherXL, which gives access to two collections, a Dutch radio and television archive; and a Dutch newspaper archive. It allows users to run two keyword queries side-by-side, either on the same collection or on different collections. The group quickly realised that what at first seemed to be an affordance of the tool, comparative analysis, is in fact difficult because the two collections do not fully overlap in the periods covered (for copyright reasons) and the newspaper archive includes full-text search whereas the radio and television archive only uses metadata. This group’s research question is somewhat similar to that of previous group: ‘How did word usage of migration changed over time?’. The comparative nature of the tool is, however, clearly reflected in the research method formulated

by this group: ‘Using the parliamentary debates via the Parliamentary Debate Search system as a baseline to trace the development of word usage, how can other data sets be used to characterize the developments?’.

Group 4 is relatively brief in their notes. They explicitly address the question whether their research question may or may not depend on available data and tools:

> We struggled with the scope of the question: should we adapt it to the sources we have at hand right away? Or do we want to make up a question that we are not sure we can answer, because we might not be able to extrapolate from the materials that we have available (because of limitation of the sources)? It is likely that when we do the latter, we end up more with tool criticism than with actual answers to questions.

Their reflections on their own research question: ‘How is the topic migration present in cultural expression? Comparing end of 60s with 90s’ follows a similar pattern. They noted: ‘We started with ambitious research questions. Through bumping into limitations, research question slowly disappeared from view’.

Group 5’s notes are hardly about tools, data, method and research questions directly, but mainly reflections on these topics, indicated by the yellow color. For example:

> The type of questions we think of is already influenced by what we expect to be possible with the tools (‘how did people think about’ became ‘what terms were used’, so this is based on available metadata/presentation of the material).

The resulting research question is indeed term-centric: ‘What were the terms used for migrants around the time of Suriname’s independence in 1975? Taking a five year window from 1975 to 1980’.

Group 6 quickly starts with keyword searches related to 'migrants' and 'integration' to identify which specific topics are viable for inquiry. Once they have established that 'integration' is more fruitful, they use explorations around this topic to address questions about how the tool constraints and steers them towards specific questions and analyses. Their lab notes suggest that part of the time during the workshop is used to try to carry out the actual research with the goal to find the answers to the research question discussed. For some queries it is unclear to what extent these are still intended to help in refining the research question. They formulated their research question as: 'In what way can we use word frequencies in parliamentary speeches as an indicator for political viewpoints on integration?'

The main point in the process *when* questions changed was when scholars identified the boundaries of the available corpus and the properties of the (meta)data. In all cases, questions around the discussion of migration and refugees were refined by zooming in on either specific organizations (e.g. Dutch political specific parties PVV and VVD), specific regions (Surinam), specific periods (1990-2014, 1970-1990, late 60s and 90s) or specific topics (assimilation).

5.3 Meta-Discussion about the workshop

The workshop closed with a general discussion in which participants were asked to reflect on the value of the format and outcomes of the workshop.

One of the main points raised is that, in using digital tools, scholars are not always reflectively questioning what they are doing. Participants who had worked on the same datasets in the workshop as in previous projects realized that back then they did not reflect in the same way and ask the questions they asked themselves in this workshop. The participants agreed that the explicit reflection on tool use in the format of a workshop, where they work together and can discuss findings on the same or similar assignments, tools and datasets, is an effective way to critically assess the use of digital tools. Here -interestingly- analogue tools such as post-its and pen-and-paper can help to stimulate this reflection as they pull scholars out of the environment of digital research.

The importance of documentation was another important topic in the discussion. One group mentioned they explicitly looked for documentation on the digital tools they considered, to find out how these tools work, what data they give access to or what formats they accept, how they transform data and for what purposes these tools were made. Such documentation is often limited or not present at all but is crucial in understanding whether a tool does what a user thinks it does. Digital tools are boxes that can be opened up to a certain extent by tool builders, either by providing source code or documentation, or working directly with (other) scholars and discuss how they work. Another group noted that scholars often attempt to use a tool for a specific part of the research but upon hitting the limitations of that tool, come up with workarounds. These are often very useful but rarely documented. One participant said he would like to know what workarounds others have developed, so he can possibly reuse them.

The third main topic that was discussed is data literacy and the complex interactions between digitals tools and data. Some participants argued that the opacity of tools means they only get in the way of getting to grips with the data: 'We don’t want a tool, we want the raw data.' They felt that researchers should have a basic understanding of data and how it is structured. They noticed that in using digital tools for research, they keep going back to the data and metadata, and the underlying structures and schemes used. 'Being able to look at a SPARQL[5] query and maybe not being able to write it yourself but at least to understand what it‘s doing … That is the literacy that we certainly should have.' 'The more directly you are able to query data, the more confident you are about what you get out.'

This points to the difficulty of separating tools and data. Once you separate the digital tool from the digital data, whatever you do with the data will involve some other tool, as interacting with digital data always requires some tool, however rudimentary, to mediate. 'Tools are intimately related to the data.' Before choosing a tool to perform data transformation or analysis, a researcher has to critically evaluate the data they use as input to the tool. Although the question remains to what extent one can separate data criticism from tool criticism, because one of the aspects of digital data criticism is to assess how it was created and shaped by previous digital technologies in the first place. This prompted the question: 'What actually is the raw data?'

[5] SPARQL is a structured query language for Linked Data. See https://www.w3.org/TR/rdf-sparql-query/

There is a long process of tools, even for digitization only. When confronted with a digitized data set, there are already many questions regarding the digitization process. Especially around OCR and text interpretation. Did the OCR process use language-specific models and parameters in deciding between candidate characters or words? How did the digitization process deal with aspects like image noise, marginalia, tilted scans, missing fragments, cuts and holes in the page, etc?

Furthermore, critiquing the chain of tools that are involved to create an online keyword search interface of a large digitized archive blends naturally with critiquing of analogue processes of constructing that archive. One question is how the metadata formats, institutional cataloging policies, selection criteria for materials to include and the cataloging choices and behaviors of individual cataloguers and documentalists have changed over the decades or centuries of an institute's history.

This led to the suggestion that we also primed in our workshop set-up: work out a method of digital data and tool criticism in phases that follow the phases of the research process, e.g. exploration, analysis, presentation (Bron et al, 2015). In each phase, criticism should focus on tool use as a chain of steps or interactions. In analyzing data that is presented in a particular tool at a particular step, it is important to understand what previous data interactions and transformations led to that view on the data and how that process shapes what a user sees.

6 Discussion: reflection as integrative practice

Digital tool criticism forces us to step back and assess how tools fit in our research methodologies. We chose to focus on the exploratory phase to draw out the questions around digital tools in the initial steps.

The most important lesson learned in this workshop is that the choice to have participants work in groups and write down their steps, encouraged them to reflect on their own research process and the role of tools in it. By introducing the model of Maxwell (2013) and Owens' discussion of its role in digital humanities research, participants could easily separate tools, data and methods and question and reflect on each aspect individually and in interaction with each other. Digital tool criticism requires scholars to relate the choice and use of tools to the phase of their research. Scholarly publications should not only focus on what we have learned about e.g. migration through using digital tools, but also reflect on the process by which we learn and generate new knowledge and insights.

Therefore, we consider reflection is the central concept in digital tool criticism. Reflection as practice integrates all elements of research to critically assess and use digital tools: research questions, methods, tools and data are interdependent and choices regarding them are shaped in an interactive and reflective research process. Why are particular data, tools and functionalities chosen? Why are certain directions discarded in favour of different directions? What insights led to a change in direction, and what new insights does that give? Our analysis

of the notes and posters made by the participants suggests that research method should be included as a separate concepts in a model for digital tool criticism. At the same time, the role of the researcher is not mentioned in the notes and posters, but only came up in the closing discussion of the workshop when participants were reflecting on the workshop and on digital tool criticism is a method, so we argue that *researcher* makes less sense as an explicit concept in the model. These considerations lead to a different model, shown in Figure 4, in which *research method* is brought back into the model and *reflection* is added to replace 'researcher' and is considered as integrative practice encompassing all other concepts.

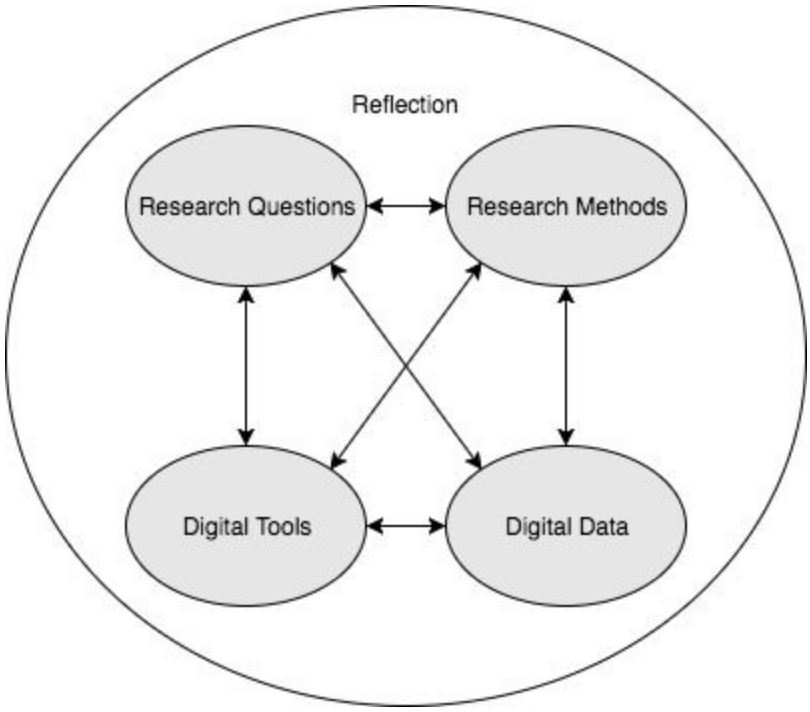

Fig. 4. An interactive model of digital tool criticism, where reflection integrates the four concepts of research questions, methods, data and tools as interactive and interdependent parts of the research process.

Adopting this type of reflection in research practice has consequences for how we conduct and organise our work. In other words, it affects our methodologies. Much like research in the late 19th and early 20th century, we have to reflect on how tools organize, access and analyse our materials before we can apply them in researching the materials. As Scheinfeldt (2008) argues,

> Late 19th and early 20th century scholarship was dominated not by big ideas, but by methodological refinement and disciplinary consolidation. Denigrated in the later 20th century as unworthy of serious attention by scholars, the 19th and early 20th century, by contrast, took activities like philology, lexicology, and especially bibliography very seriously. Serious scholarship was concerned as much with organizing knowledge as it was with framing knowledge in an ideological construct.

The explicitness of digital tools prompts scholars to ask questions about them that may not always have been obvious when working with analogue tools. Questions regarding the selection, normalization and organization of data in indexes has correspondences with questions about traditional access tools for archives, libraries and heritage collections. This goes beyond recognizing the politics and rhetorical construction of archives (Finnegan, 2006, p. 118), to understanding the history of collection creation, organization and management. An institution's history of gathering and organizing materials into collections, and changes in institutional policy regarding these activities are rarely documented in great detail, but are also rarely considered or reported in research that makes use of these collections. E.g. how selection criteria and topical or subject indexing of archival materials has changed over time, how indexers applied the chosen controlled vocabularies and conducted their document analysis, how different indexers made different interpretive choices regarding the relevance of index terms, etc. All these affect accessibility of archival materials. Yet with digital tools and data, these types of questions are posed frequently. Perhaps the disconnect between distant reading perspectives and established close reading methods prompts scholars to question how to make sense of such reductive views on the data and how these views relate to a scholar's expectations derived from background knowledge. For instance, seeing search results represented as a frequency graph on a timeline, a scholar might see a peak or a dip in a certain period and wonder how it relates to what they know about that period, but also how it relates to the history of the collection being searched.

The main questions center around complex relationship between tools and data in a digital environment. The first aspect is how tools select, filter and give access to data. Tool limitations may form a barrier to having full access to a set of data because a tool may be the only way to access them, as with web-based tools that gives access to digital archives and heritage collections. Access to digital sources is often mediated through digital tools, which suggests an integrated criticism of tools and sources. Another issue with many digital tools working on integrated data sets is that they lack information about what data is accessible through the tool, how that data has been selected and how tool features include or exclude certain parts of the data. This makes it hard for scholars to judge whether what they see is all there is, or that other data has been filtered out or is simply not available in the tool.

The second aspect is how tools transforms the data they operate on and thereby can change the nature of the data and how they can be interpreted. In order to critically evaluate the suitability of digital tools for a particular research scope and approach, a scholar needs to have a basic understanding of how they work and what they do and don't do. We agree with Benjamin Schmidt (2016) that this need not necessarily be at the level of algorithmic detail, but at the level of data transformations. Some tools are extremely complex with hundreds of algorithms, and some require advanced mathematical knowledge to fully comprehend but which is not necessary to meaningfully use the tool in research. However, at the level of data transformations, the workings of tools represent data interpretations and directly affect methodology. In this sense, the selection and filtering of data discussed above are also transformative. Keyword search not only selects or filters, but also reorganizes data sources, taking them out of their individual contexts and placing them together in a list of search results, often ordered by algorithmically determined relevance. This also makes it clear that the choices

made by the researcher to use certain keywords or to use certain tools in a particular order should be included in the critical assessment of a tool and that this an important reflective step in the research process.

Another aspect of tools are interfaces. Interfaces are often introduced with comments about how easy to use they are. Incorporating digital tools in research is never easy and always requires critical reflection on how they mediate between researchers and their materials of study. Attractive and intuitive interfaces make it easy to forget that under the hood, many choices are made based on implicit or explicit assumptions of the creators of the tools, that may or may not align with the assumptions of their users.

This has lead to the following definition or demarcation of the concept of digital tool criticism:

> With digital tool criticism we mean the reflection on the role of digital tools in the research methodology and the evaluation of the suitability of a given digital tool for a specific research goal. The aim is to understand the impact of any limitation of the tool on the specific goal, not to improve a tool's performance. That is, ensuring as a scholar to be aware of the impact of a tool on research design, methods, interpretations and outcomes.

This requires researchers, data custodians and tool providers to understand issues from different perspectives. Researchers need to be trained to anticipate and recognize tool bias and its impact on their research results. Data custodians and tool providers, on the other hand, have to make information about the potential biases of the underlying processes more transparent. This includes processes such as collection policies, digitization procedures, data enrichment and linking, quality assessment, error correction and search technologies (Traub and Van Ossenbruggen, 2015).

Reflection on tool use in a research process suggests an element of experimentation, the latter being widely considered as important element in digital tool use (cf. Section 2). One way to critically evaluate a tool for a given purpose is to experiment with different ways of applying the tool. This allows evaluation from multiple experiences and perspectives. A concrete example is a simple heuristic of testing alternative keyword queries and compare the number of results or analyse the overlap in results, which can reveal the inner workings of tools. Experimentation is a skill in the sense that there are good and bad ways to experiment with a tool to assess its impact on data and interpretation. Experimentation also helps scholars to reflect on and challenge their own assumptions regarding tools and data.

Reflection on procedure and method does not come naturally while doing research, especially when interfaces resemble those we use everyday. This is where collaborative sessions are useful, each person bringing their own experiences and skills. For digital tool criticism it helps to have both scholars and tool developers involved in the discussion. Collaboration also affords brainstorming ideas and coming up with experiments to quickly test hypotheses. At same time, collaborative research raises the issue of being less involved in the entire process, especially in presenting parts of scholarly work that were done by others. In the case of humanities scholars and computer scientists, it may be difficult to establish to what extent they understand each others' contributions.

7 List of recommendations

Based on the discussion points above, we provide a list of recommendations for conducting digital tool criticism for (1) tool creators and maintainers and (2) humanities scholars.

First, *creators and maintainers* that give access to data sets, stand alone tools and tools built around datasets should provide documentation describing a range of details of these data sets and tools:

- For data sets it is important to describe the selection criteria and any data processing and transformations performed on the selected data before it is made available. Selections, normalizations, aggregations and other steps that affect the input data need to be described, at least at a high level, so that researchers can reason about what is in the data sets and what is not, and how the transformations affect the ways they can validly interpret the data.
- For tools it is important to describe what functionalities are available and how each of these selects, filters and transforms data, so scholars can reason how they change the nature and scope of the data from input to output. From the workshop discussion came the recommendation for tool builders to have an “about” page with each digital tool that covers these aspects.

Second, *humanities scholars* using digital tools in their research should reflect and report on their choices for those tools. We make the following recommendations:

- Digital tool criticism should analyse and discuss tools at the level of data transformations. Reflect on how inputs and outputs differ and what this means for interpreting the transformed data.
- Source criticism, tool criticism and data criticism (as output of the tools they used) should be integrated and incorporated in the research process. Scholars should reflect and report on how these three aspects contribute to the scope of the data and how that aligns with the scope of the research questions.
- Scholars should document and share the workarounds they develop in dealing with limitations of tools. Aspects to document are the types of activities that a tool does not support well and what alternative steps with the same or other tools have been taken.
- The research process should include experimentation to find out how digital tools work in terms of modelling and transforming data, and to bring out and refine scholars’ own assumptions about tools.

A good way to perform digital tool and data criticism is to use a checklist of questions to ask about the tools and data:

- Questions to ask about digital data: Where does the data come from? Who made the data? Who made the data available? What selection criteria were used? How is it organised? What preprocessing steps were used to make the data available? If digitized from analogue sources, how does the digitized data differ from the analogue sources? Are all sources digitized or only selected materials? What are known omissions/gaps in the data?
- Questions about digital tools: Which tools are available and relevant for your research? Which tool best fits the method you want to use? How does the tool fit the method you want to use? For which phase of your research is this tool suitable? What kind of tool is it? Who made the tool, when, why and what for? How does the tool transform the data that it works upon? What are the potential consequences of this?
    - *Questions about digital search tools*: What search strategies does the tool allow? What feedback about matching and non-matching documents does the tool provide? What ways does the tool offer for sense-making and getting an overview of the data it gives access to?
    - *Questions about digital analysis tools:* What elements of the data does the tool allow you to analyze qualitatively or quantitatively? What ways of analyzing does the tool offer, and what ways to contextualize your analysis?

Although there are also digital publication tools, we did not yet look into this within the confines of the workshop. The workshop focused on tools for exploration and also on tools for analysis, as exploration often incorporates different forms of analysis.

8 Conclusion and future steps

In this article we argued that reflection can be seen as an integrative practice. Our research is based on the outcomes of a workshop in which we brought together people with an interest in Digital Humanities research. One of the findings was that collaborative note taking and reflection is an effective way to make scholars more aware of limitations of data and tools but more importantly of their own research process and the questions, considerations and choices they have. In that sense, the format of the workshop was a success. Therefore, we are planning further iterations of this workshop where we tighten the protocol for tracking the research process.t. For instance, we will try to let our future participants make their own ‘research-process-visualisations’ since we expect these visualisations to be a great help in their reflection process. We also plan to include logging of system interactions in future workshops, so that participants can connect the steps in their research process to specific interactions with tools and also see when they switch between tools.

A challenge of any workshop is to find a balance between priming of participants in providing working definitions, tools and assignments and enabling to draw conclusions on the outcomes of workshop in a collaborative fashion. We believe it is important to build on existing knowledge and experiences, and therefore we plan to share this article with all future

participants, so we can build an even more broadly shared framework for Digital Tool Criticism. In a follow-up workshop to the one discussed in this paper, we have to think about a way to let participants also co-author guidelines, perhaps by let them write and test guidelines during the workshop and/or create a voting system by which guidelines can be ranked according to their perceived importance. Moreover, our workshop focused on the first phase of research - exploration - and related tools. It would be valuable to retake the workshop for all other phases as well to test our model of reflection as integrative practice.

9 Funding

This work was supported by the VRE4EIC project, a project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 676247 and by the CLARIAH-CORE project financed by NWO (www.clariah.nl).

10 Notes

11 References

Bhattacherjee, A. (2012). Social science research: Principles, methods, and practices. Tampa, FL: Global Text Project.

Bhavnani, S., K. Drabenstott, and D. Radev. (2001). Towards a unified framework of IR tasks and strategies. In: Proceedings of the American Society for Information Science Annual Meeting 38, pp. 340-354.

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T. and Schnapp, J., (2012). Digital_Humanities. Mit Press.

Burke, T. (2011). How I Talk About Searching, Discovery and Research in Courses. [Blog] Easily Distracted. Available at: https://blogs.swarthmore.edu/burke/blog/2011/05/09/how-i-talk-about-searching-discovery-and-research-in-courses/ [Accessed 1 Mar. 2018]

Drabenstott, K.M. (2001). Web Search Strategy Development. Online, Vol. 25, No. 4.

Eijnatten, J. van, Pieters, T. and Verheul, J., (2013). Big Data for Global History: The Transformative Promise of Digital Humanities. BMGN - Low Countries Historical Review. 128(4), pp.55–77. DOI: http://doi.org/10.18352/bmgn-lchr.9350

Fickers, A. (2012). Towards a New Digital Historicism? Doing History in the Age of Abundance. View journal, volume 1 (1). http://orbilu.uni.lu/bitstream/10993/7615/1/4-4-1-PB.pdf

Finnegan, C.A., (2006). What is this a picture of?: Some thoughts on images and archives. Rhetoric & Public Affairs, 9(1), pp.116-123.

Gibbs, F. and Owens, T. (2013). The Hermeneutics of Data and Historical Writing. In Kristen Nawrotzki; Jack Dougherty. Writing History in the Digital Age. University of Michigan Press, 2013. DOI: http://dx.doi.org/10.3998/dh.12230987.0001.001.

Graham, S., Milligan, I. and Weingart, S. (2015). Exploring big historical data: The historian's macroscope. London, Imperial College Press.

Guiliano, J. (2017). Toward a Praxis of Critical Digital Sport History. Journal of Sport History, Volume 44, Number 2, Summer 2017, pp. 146-159.

Hitchcock, T. (2013). Confronting the Digital - Or How Academic History Writing Lost the Plot. Cultural and Social History, Volume 10, Issue 1, pp. 9-23. https://doi.org/10.2752/147800413X13515292098070

Huistra, H., and Melink, B. (2016). Phrasing history: Selecting sources in digital repositories. Historical Methods: A Journal of Quantitative and Interdisciplinary History, Vol. 49, NO. 4, 220-229. DOI: 10.1080/01615440.2016.1205964

Kendall, D. (2012). Sociology in our times. Belmont, Canada: Cengage Learning.

Marshall, C. and Rossman, G.B. (2010). Designing qualitative research. London: Sage.

Maxwell, J. (2013). Qualitative Research Design: An Interactive Approach, 3rd edition. SAGE publications.

Owens, T. (2014). Where to start? On research questions in the digital humanities. [Blog] Trevor Owens. http://www.trevorowens.org/2014/08/where-to-start-on-research-questions-in-the-digital-humanities/ [Accessed 1 Mar. 2018]

Putnam L. (2016). The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast. American Historical Review, Volume 121, Number 2, pp. 377-402.

Scheinfeldt, T. (2008). Sunset for Ideology, Sunrise for Methodology? [Blog] Found History. http://foundhistory.org/2008/03/sunset-for-ideology-sunrise-for-methodology/ [Accessed 1 Mar. 2018]

Schmidt, B. (2016). Do Digital Humanists Need to Understand Algorithms? Debates in the Digital Humanities, 2016 Edition.

Solberg, J. (2012). Googling the Archive: Digital Tools and the Practice of History. Advances in the History of Rhetoric, Volume 15, pp. 53-76.

Traub, M.C, and Van Ossenbruggen, J.R. (2015). *Workshop on Tool Criticism in the Digital Humanities*. CWI Techreport https://ir.cwi.nl/pub/23500

Underwood, T. (2014). Theorizing Research Practices We Forgot to Theorize Twenty Years Ago. Representations, Vol. 127 No. 1, Summer 2014; (pp. 64-72) DOI: 10.1525/rep.2014.127.1.64.

Van Gorp, J. and de Leeuw, J.S. (2018) Methods of Data Collection with/in Digital Television Archives: Digital Television Historiography. In Van den Bulck H. et al. (2018) *Palgrave Handbook for Media Policy Methods*. Forthcoming

Yakel, E. (2010). Searching and Seeking in the Deep Web: Primary Sources on the Internet. In Working in the Archives: Practical Research Methods for Rhetoric and Composition, pp. 102-118. Southern Illinois University Press.