



K O N I N K L I J K E N E D E R L A N D S E
A K A D E M I E V A N W E T E N S C H A P P E N

What Big Data Sounds Like

Zeldenrust, Douwe; 't Hart, Elise; van Zundert, Joris J.

2023

document version

Publisher's PDF, also known as Version of record

document license

CC BY-SA

[Link to publication in KNAW Research Portal](#)

citation for published version (APA)

Zeldenrust, D., 't Hart, E., & van Zundert, J. J. (2023). *What Big Data Sounds Like: The Hermeneutic Significance of the Arts for Big Data in Digital Humanities*.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

pure@knaw.nl

What Big Data Sounds Like – The Hermeneutic Significance of the Arts for Big Data in Digital Humanities

Douwe A. Zeldenrust
Department of Digital Collections
Meertens Institute, Royal Netherlands
Academy of Arts and Sciences
Amsterdam, The Netherlands
douwe.zeldenrust@meertens.knaw.nl

Elise L. 't Hart
Artist in Residence
Meertens Institute, Royal Netherlands
Academy of Arts and Sciences
Amsterdam, The Netherlands
Elisethart@gmail.com

Joris J. van Zundert
Department of Literary Studies
Huygens Institute for the History of the
Netherlands, Royal Netherlands Academy of
Arts and Sciences
Amsterdam, The Netherlands
joris.van.zundert@huygens.knaw.nl

Abstract— In recent scholarly past much attention has been called to the reductive nature of big data analysis regimes and to questions of the perceived but misleading neutral nature of large scale data and algorithms. This paper investigates a particular case of such issues and the frictional process of interfacing big data (analysis) and the hermeneutically geared humanities. We examine in particular the affordances produced by the re-introduction of the artist in a scientific environment.

Keywords— *digital humanities, hermeneutics, arts, audio recordings, scientific process.*

I. INTRODUCTION

The Meertens Institute (Royal Netherlands Academy of Arts and Sciences) specializes in linguistic and ethnological research [1]. This type of research increasingly applies methods and techniques associated with digital humanities and Big Data [2]. Consequently, the institute becomes a site of experimental interdisciplinary research where collaborations between computer scientists, scholars, and digital curators are frequent. Often smaller creative digital humanities experiments turn out to be the grassroots loci of methodological innovation [3].

In October 2018 Elise 't Hart started as the first “Artist in Residence” of the Meertens Institute [4]. The specific aim of appointing an artist in residence was to re-introduce the creative impulse of the artist in a scientific environment. 't Hart initiated the project “The Sound of Time” which added a new perspective to the large scale digitized linguistic audio data that has been collected during decades by scholars of the Meertens Institute. The unique perspective of 't Hart raises a number of important methodological questions pertaining to the significance of the artist's work for scientific investigation and the ways we are used to treating data in a scientific context. In this paper we provide tentative answers to these questions by investigating the affordances produced by the artist's new approach.

II. TURNING TIDES

At the foundation of the Royal Netherlands Academy of Arts and Sciences (KNAW) in 1808 it was called the Royal Institute of Sciences, Literature and Fine Arts [5]. Around 1800 this combination of the different disciplines was self-evident: the ideal for a civilized person was to have knowledge of Science, Literature and of Fine Arts. The king of Holland, Louis Napoléon Bonaparte, installed in this new institute four different ‘Classes’. The First one for Mathematics, Physics and Technique, the Second for Dutch Language, Literature and History, a Third one for Ancient- and Oriental Languages and History and a Forth one for Fine Arts. Artists and scientists were considered equals.

But in the 19th century the disciplines grew apart. Budget cuts and the ideas of prime minister Johan Thorbecke regarding the use and benefits of the Institute resulted in its closure in 1851. It was subsequently replaced by the Royal Academy of Sciences and only the First Class remained. Thorbecke reasoned that most advice for the government came from the First Class and that, contrary to the other Classes, it needed government subsidy to publish the results. Thorbecke encountered persistent resistance and only four years later, after the fall of his cabinet, a Second Class was added. This Class encompassed Language, Literature, History and Philosophy. However, Fine Arts did not return.

Only recently this situation is changing. In 2014 the KNAW founded the Society of Arts with the aim of re-introducing the artist in the field of science. This time the involvement of artists is founded on different ideas. According to the Society there are many similarities between the work of scientists and the work of artists. It states that both are driven by passion in which creativity, curiosity and originality play a crucial role. Moreover, the new Society “[...] will make such similarities more explicit and help restore the relationship between science and art” [6].

Within the field of digital humanities this initiative is now followed by the Meertens Institute [4]. At this humanities institute collaborations between experts of different disciplines

such as computer scientists, scholars, digital curators are already frequent and now a new perspective has been added: that of the artist. Elise 't Hart, artist and founder of the "Institute for Domestic Sounds" [7], started in October 2018 for six months as the first Artist in Residence of the Meertens Institute [4]. Her project was called "The Sound of Time". The aim was, beside making works of art, to explore experimental ways of questioning and researching Big Data.

III. 250 DAYS OF NON-STOP LISTENING

The Meertens Institute has a strong tradition in studying Dutch language and culture [1]. It possesses a large library and numerous (audio) collections. The institute has more than six thousand hours, or 250 days of non-stop listening, of audio recordings of dialects, interviews, folk songs and stories from all parts of the Netherlands and of Dutch speaking people abroad. The oldest recording dates from 1927 and the collection has been augmented over the decades following. The Meertens Institute still adds new (digital) recordings to the collection [8].

The entire collection has been digitized according to the standards of the Library of Congress. Currently over 1,000 hours (in total 2,216 recordings) are available via the website "Nederlandse Dialectenbank" (transl. Dutch Dialect Database) of the Meertens Institute [9][10][11]. Typescripts of 660 of the available recordings have been digitized and added to the database, comprising in total more than 11,000 scans. The entire dataset has additionally been made available via the Common Language and Resources Infrastructure (CLARIN), a European research infrastructure to share scientific linguistic resources [12]. Each resource is described using CMDI (Component Metadata Infrastructure) [13] and PIDs (Persistent Identifiers) have been assigned. To allow others to harvest the metadata records an OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) [14] provider serves the DCMI (Dublin Core Metadata Initiative) [15] and the CMDI documents.

The Meertens Institute is in the process of publicizing all the audio recordings and additional data via the Dutch Dialect Database. This effort is costly and time consuming. A significant, if justified, complication is presented by the General Data Protection Regulation, GDPR (Dutch: Autoriteit Persoonsgegevens). This privacy oriented legal measure demands that the rights of the interviewees are respected [16]. This proves to be an issue as it is often not exactly known who is mentioned and in what context. Moreover, most recordings are in various dialects for which speech recognition is not yet an option. Therefore, only manual labor-intensive solutions are currently available to digitize the recordings. Consequently, the Dutch Dialect Database is a work in progress. Nevertheless, it generates more than 200,000 page views annually, showing that it is well known to researchers and the general public [9].

IV. TICKING CLOCKS

Traditionally the focus of the scholars of the Meertens Institute is on the dialects, the stories, and the songs audible in the recordings. Other sounds, such as the squeaking of a staircase, the tinkling of coffee cups, and the ticking of clocks

are seemingly of no value to them. However, for Elise 't Hart the audio collection contains a wealth of hidden historic domestic audio treasures. The very tick of a clock may take a person back to, for instance, the living room of one's youth or that of one's grandparents [7]. These everyday sounds are therefore at the center of her interest. In 2013 she founded the Institute for Domestic Sounds. Originally this was part of a graduation assignment at the Utrecht University of Fine Arts. The Institute for Domestic Sounds proved to be an instant public success. For instance, the artist offers a subscription to a "Sound of the Month" on SoundCloud which shows an ever increasing number of subscriptions [7] [17]. The public interest in these domestic sounds has resulted in the continuation of systematic registering of everyday domestic sounds.

By recording the domestic sounds and adding context information, 't Hart makes what usually goes unnoticed take center stage. The audio collection of the Meertens Institute proved to be a rich resource for this. The focus of the Meertens Institute collection is in line with the interest of the Institute for Domestic Sounds – both are interested in everyday culture. Moreover, because the collection also includes the dimension of time as the recordings date back more than 80 year, Elise 't Hart is able to track historic changes in domestic sounds. To give but one example: the ringing of a telephone in the hallway makes place for the soft buzzing of a mobile on the table [4].

As part of her Artist in Residence fellowship Elise 't Hart adds annotations to the audio collection of the Meertens Institute. Thus, in addition to linguistic detail now also domestic sounds, such as the striking of a match for a cigarette, or the whistling of a tea kettle, become available as annotations no the typescripts (cf. figure 1). She also includes these annotations in the dictionary of sounds that she is currently working on. The work of annotating later serves to transform recorded sound into digital works of art that are streamed as podcasts [4].

't Hart's efforts allow listeners to become more aware of domestic everyday sounds. Seemingly trivial audio information thus makes for a more comprehensive understanding of the soundscapes that are part of the social context [18]. Elise 't Hart: "It's about learning to listen, becoming aware and gaining complete experience of our environment" [7].



Fig. 1. An example of a transcription with annotations in Dutch. This example is from a "one minute" podcast made by Elise 't Hart using the audio recording from the village of Fijnaart in the Netherlands [19]. In gray are the annotations of Elise 't Hart: [stilte] (transl. "silence"), [krakende stoel] (transl. "creaking chair") and [tikkende klok] (transl. "ticking clock").

V. BIG DATA AND HERMENEUTICS

What make big data exactly big is a challenging question in itself [20], but constraining a working definition to the sheer amount of data, big data in the humanities are often still comparatively small with regard to big data in, for instance, gene sequencing or particle physics. However, the digitizing of vast collections of text, and especially the digitization of moving image and sound has created a rich ground for the application of big data analysis techniques [21, 22]. On a methodological level the work from 't Hart can be seen as both being afforded by big data as well as being a reaction to the dominant discourse on big data. "The Sound of Time" project was only made possible through the arduous digital and manual curation work that resulted in the massive scale digitized sound collections of the Meertens Institute. At the same time the work of "The Sound of Time" confronts us with a number of myopic effects that hide as caveats in big data and associated analysis patterns.

In the recent scholarly past much attention has been called to the reductive nature of big data analysis regimes [23] [24], and to questions of the perceived but misleading neutral nature of large scale data [25] and algorithms [26]. Conventional approaches in the humanities are hermeneutic, that is: they are based on theories of interpretation and not on theories of numerical evidence. This creates an uneasy friction between said conventional methods and "the unreasonable effectiveness of data" [27]. The reductive nature of, for instance, machine learning in big data analysis, and the risks of biases in big data have genuinely and justifiably deterred a lot of scholars from venturing into big data analysis. However, a lot of the defense of hermeneutics on the scholarly side also rests in sheer unwillingness to explore the possible methodological interfaces between big data analysis and hermeneutics. Thus, an intellectual inertia leads to the suboptimal or even non-exploration of what big data hermeneutics entail. It is in this respect that 't Hart's work serves as an eye-opener for many researchers involved. Her work adds evidence to our idea that the myopic effects of big data do reside in the researcher more than in the data. Generations of linguistic scholars have perceived the Meertens Institute's collection of audio recordings as linguistic data first and only. It took an artist's perspective to see what other information was hidden by the linguistic perspective. At the same time 't Hart's approach shows what riches may be gained from engaging with big data in a purely – and too often called "old fashioned" – hermeneutic fashion.

VI. IMPACT

The introduction of the artist 't Hart's perspectives into the field of digital humanities results not only in digital works of art, it also performs as a catalyst of method, research, and dissemination in various ways. On a methodological level it opened up new perspectives on existing research, adding to analysis design. At the Meertens Institute the project "Migration and Material Culture" researched the domestic interiors of migrants and their descendants in the Netherlands. As part of the project interviews were held and recorded. It resulted in several publications and a PhD-theses [28]. Domestic sounds were not considered in this study. The ideas

of 't Hart have now resulted in a dialogue with one of the researchers involved on how a new dimension of domestic sounds could be integrated in the data analysis. In addition, the work of 't Hart results in a recursive form of archiving, which could be called a methodological side-effect. At the end of the project – as it is the policy of the Meertens Institute to document project results – the digital works of art have again become part of the (open access and online available) collections of the Meertens Institute [4], having themselves resulted from the same archive [19].

The work of 't Hart generates renewed research interest in the use of (digitized) audio collections among scholars. The presentation of the "The Sound of Time" at an interdisciplinary meetup for artists, scholars and archivists of the Amsterdam Research Institute of the Arts and Sciences (ARIAS), sparked a new interest in the collection of Louise Kaiser. Kaiser (1891 – 1973) was director of the "Experimental Phonetics Laboratory" and the first female lecturer of the University of Amsterdam. She built a substantive audio collection of over 1,800 mostly experimental recordings. Elise 't Hart used this collection – that is now archived by the Meertens Institute – for her podcasts. The meeting resulted in the formation of a consortium of researchers and collections specialists from the Meertens Institute, the University of Amsterdam and the Institute for Domestic Sounds. Plans are being designed to develop a digital research platform for the Kaiser collection.

On the level of research dissemination the work of 't Hart has proven tremendously productive. The podcasts created substantial additional audience among the general public for the collections of the Meertens Institute, which serves considerably to valorize both the institute and its collections [29]. Moreover, because of the success of "The Sound of Time", the Royal Netherlands Academy of Arts and Sciences has made additional funds available for a permanent and annually rotating Artist in Residence. A call has been issued in June 2019 for an artist to work on a project that makes a meaningful contribution to promoting cross-fertilization between the Arts and Science [30].

VII. CONCLUSION

We have found that the re-introduction of the artist in the scientific process as it pertains to big data has resulted in several beneficial effects both for the researchers and artist involved. The unique perspective that comes with the eye and attention of the artist serves to alleviate forms of myopia that exist in researchers with regard to big data and with regard to the potential scientific uses of curated big data. Artists joining in the scientific process by way of fellowships may play a pivotal role in the difficult and frictional process of interfacing big data, big data analysis, and hermeneutically geared humanities. At the same time engaging an artist in the scientific process proved to be a major factor in enlarging the public dissemination effect of research.

In the recent scholarly past much attention has been called to the reductive nature of big data analysis regimes [21] [22], and to questions of the perceived but misleading neutral nature of large scale data [23] and algorithms [24]. Conventional approaches in the humanities are hermeneutic, that is: they are

based on theories of interpretation and not on theories of numerical evidence. This creates an uneasy friction between said conventional methods and “the unreasonable effectiveness of data” [25]. The reductive nature of, for instance, machine learning in big data analysis, and the risks of biases in big data have genuinely and justifiably deterred a lot of scholars from venturing into big data analysis. However, a lot of the defense of hermeneutics on the scholarly side also rests in sheer unwillingness to explore the possible methodological interfaces between big data analysis and hermeneutics. Thus, an intellectual inertia leads to the under exploration of what big data hermeneutics entail. It is in this respect that ’t Hart’s work serves as an eye-opener for many researchers involved. Her work adds evidence to our idea that the myopic effects of big data do reside in the researcher more than in the data. Generations of linguistic scholars have perceived the Meertens Institute’s collection of audio recordings as linguistic data first and only. It took an artist’s perspective to appreciate what other information was hidden by the linguistic perspective. At the same time ’t Hart’s approach shows what riches may be gained from engaging with big data in a purely – and too often called “old fashioned” – hermeneutic fashion.

REFERENCES

- [1] “The Netherlands in the World: Language and culture in everyday life in the Netherlands,” Meertens Instituut. [Online]. Available: <https://www.meertens.knaw.nl/cms/en/onderzoek>. [Accessed: 14-Aug-2019].
- [2] M. Terras, J. Nyhan, and E. Vanhoutte, Eds., *Defining Digital Humanities*. Ashgate, 2013.
- [3] B. Nowvskie, “On the Origin of ‘Hack’ and ‘Yack,’” in *Debates in the Digital Humanities 2016*, M. K. Gold and L. F. Klein, Eds. Minneapolis, London: University of Minnesota Press, 2016, p. 7.
- [4] “Hoe klinkt de tijd?,” Meertens Instituut. [Online]. Available: <https://www.meertens.knaw.nl/cms/nl/nieuws-agenda/artist-in-residence>. [Accessed: 14-Aug-2019].
- [5] K. Van Berkel, *De stem van de wetenschap, Geschiedenis van de Koninklijke Nederlandse Akademie van Wetenschappen, deel 1: 1808 – 1914*. Amsterdam: Bert Bakker, 2008.
- [6] “Society of Arts,” Royal Netherlands Academy of Arts and Sciences. [Online]. Available: <https://www.knaw.nl/en/about-us/society-of-arts>. [Accessed: 14-Aug-2019].
- [7] E. L. ’t Hart, “Instituut voor Huisgeluid,” Elise ’t Hart. [Online]. Available: <https://www.Elisethart.com/instituut-voor-huisgeluid>. [Accessed: 14-Aug-2019].
- [8] “Audio Collections,” Meertens Instituut. [Online]. Available: <http://www.meertens.knaw.nl/cms/en/collections/audio-collections>. [Accessed: 14-Aug-2019].
- [9] “Nederlandse Dialectenbank,” Meertens Instituut. [Online]. Available: <https://www.meertens.knaw.nl/ndb>. [Accessed: 14-Aug-2019].
- [10] M. Oostendorp, “Phonological and phonetic databases at the Meertens Institute,” in *The Oxford Handbook of Corpus Phonology*, J. Durand, U. Gut, and G. Kristoffersen, Eds. Oxford: Oxford University Press, 2014, pp. 546–551.
- [11] D. Zeldenrust, “The creation, curation and dissemination of humanities digital resources: The Dutch Dialect Database,” presented at the European Social Science History Conference 2016, Valencia, 2016.
- [12] “CLARIN - European Research Infrastructure for Language Resources and Technology,” CLARIN. [Online]. Available: <https://www.clarin.eu/>. [Accessed: 14-Aug-2019].
- [13] “Component Metadata,” CLARIN. [Online]. Available: <https://www.clarin.eu/content/component-metadata>. [Accessed: 14-Aug-2019].
- [14] “Open Archives Initiative Protocol for Metadata Harvesting.” [Online]. Available: <https://www.openarchives.org/pmh/>. [Accessed: 14-Aug-2019].
- [15] “Dublin Core Metadata Initiative: innovation in metadata design, implementation & best practice,” 2019. [Online]. Available: <https://www.dublincore.org/>. [Accessed: 14-Aug-2019].
- [16] H. Tjalsma, “Gevolgen nieuwe AVG voor data-archieven,” *Edata & Research*, vol. 12, no. 3, p. 4, 2018.
- [17] E.L. ’t Hart, “Instituut voor Huisgeluid,” SoundCloud, 2019. [Online]. Available: <https://soundcloud.com/Elisethart/sets/geluidenabonnement>. [Accessed: 14-Aug-2019].
- [18] S. J. Smith, “Soundscape,” *Area*, vol. 26, no. 3, pp. 232–240, 1994.
- [19] “Collectie behorende bij het project: ‘Hoe klinkt de tijd’,” collection nr. 1116, Meertens Instituut (KNAW), 2018-2019 (2019).
- [20] F. Kaplan and I. di Lenardo, “Big Data of the Past,” *Frontiers in Digital Humanities*, vol. 4, p. 12, 2017.
- [21] C. Martinez-Ortiz et al., “From Tools to ‘Recipes’: Building a Media Suite within the Dutch Digital Humanities Infrastructure CLARIAH,” in *Abstracts DHBelux 2017 conference: Wednesday 5 July 2017, Utrecht, 2017*, pp. 78–81.
- [22] A. Piper, “There Will Be Numbers,” *Journal of Cultural Analytics*, vol. 1, no. 1, May 2016.
- [23] J. J. van Zundert, “Screwmenetics and Hermenumericals: the Computationality of Hermeneutics,” in *A New Companion to Digital Humanities*, S. Scheibman, R. Siemens, and J. Unsworth, Eds. Malden (US), Oxford (UK), etc.: John Wiley & Sons, Ltd, 2016, pp. 331–347.
- [24] J. Drucker, “Graphesis: Visual Knowledge Production and Representation,” *Poetess Archive Journal*, vol. 2, no. 1, Dec. 2010.
- [25] L. Gitelman, Ed., “Raw Data” Is an Oxymoron. Cambridge (MA), USA: The MIT Press, 2013.
- [26] T. McPherson, “Why Are the Digital Humanities So White? or Thinking the Histories of Race and Computation,” in *Debates in the Digital Humanities*, M. K. Gold, Ed. Minneapolis: University of Minnesota Press, 2012, pp. 139–60.
- [27] A. Halevy, P. Norvig, and F. Pereira, “The Unreasonable Effectiveness of Data,” *IEEE Intelligent Systems*, vol. 24, no. 2, pp. 8–12, Mar. 2009.
- [28] H. Van der Horst, “Materiality of belonging: The domestic interiors of Turkish migrants and their descendants in the Netherlands,” PhD, University of Amsterdam, Amsterdam, 2008.
- [29] E. Dijksterhuis, “Het is geluid van nostalgie,” *Het Parool*, p. 16, 28-Nov-2018.
- [30] “Artist in Residence at the Meertens Institute: nominations now being accepted,” Akademie van Kunsten, 2019. [Online]. Available: <https://www.akademievankunsten.nl/nl/english/artist-in-residence-at-the-meertens-institute-nominations-now-being-accepted>. [Accessed: 14-Aug-2019].