A map for big data research in digital humanities

Frédéric Kaplan *

Digital Humanities Laboratory (DHLAB), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

This article is an attempt to represent Big Data research in digital humanities as a structured research field. A division in three concentric areas of study is presented. Challenges in the first circle – focusing on the processing and interpretations of large cultural datasets – can be organized linearly following the data processing pipeline. Challenges in the second circle – concerning digital culture at large – can be structured around the different relations linking massive datasets, large communities, collective discourses, global actors, and the software medium. Challenges in the third circle – dealing with the experience of big data – can be described within a continuous space of possible interfaces organized around three poles: immersion, abstraction, and language. By identifying research challenges in all these domains, the article illustrates how this initial cartography could be helpful to organize the exploration of the various dimensions of Big Data Digital Humanities research.

Introduction: Big Data Digital Humanities vs. Small Data Digital Humanities

Defining the nature and the boundaries of digital humanities is a long-discussed and unsolved issue (Terras et al. 2013), not only because there is no consensus on this question but also because digital humanities are currently undergoing a profound transformation that calls for a reconsideration of its fundamental concepts (Gold 2012). For years, digital humanities have been loosely regrouping computational approaches of humanities research problems and critical reflections of the effects of digital technologies on culture and knowledge (Schreibman et al. 2008). Ten years ago, they emerged as a new label, rebranding and enlarging the idea of "humanities computing" (Svensson 2009). Around this new name and under a "big tent," a progressively larger community of practice thrived (Terras 2011). Each work at the intersection of Computer Science and the Humanities could potentially be part of this welcoming trend. Researchers gathered in national and international meetings, exchanged their views on blogs and mailing lists. If not a well-bounded field, digital humanities were surely a lively conversation.

The welcoming digital humanities label opened doors, connected separated academic silos, built bridges between information sciences and the various disciplines loosely forming what is called the humanities. However, openness was always associated with a need for introspection, self-reflexive writings, tentative boundaries definitions, the "What are digital humanities" articles and monographs became a genre of its own structured around several narratives of exclusion and inclusion (Rockwell 2011). Digital humanities as a research domain define themselves dynamically in the negotiation of these tensions as discussed by several digital humanities scholars (Unsworth 2002; Svensson 2009; Rockwell 2011). **Table 1** gives a non-exhaustive list of these structuring tensions.

The starting point of this article is a relatively new particular structuring tension, opposing Big Data Digital Humanists with Small Data Digital Humanists. Research in Big Data Digital Humanities focuses on large or dense cultural datasets, which call for new processing and interpretation methods.

TABLE 1 | Examples of structuring tensions defining digital humanities.

Structuring tensions	Questions
Humanists vs. digital humanists	When does research in humanities become digital humanities? Can "every medievalist with a website" be part of the digital humanities (Fitzpatrick 2012a)? Does the use of a computer in humanities research make digital humanities research (Unsworth 2002)?
Computer scientists vs. humanists inside digital humanities	Should we still distinguish computer scientists and humanists in digital humanities communities? Is the "two cultures" tension still relevant (Snow 1959)? Are digital humanities a form of "technical upgrade" of the humanities disciplines? Are digital humanities just a particular "application" of the Computer Science fields?
Makers vs. interpreters	Are digital humanities only about "building things"? If you are not a "maker," should you not be considered as "digital humanist" (Ramsey 2011)? Is there room for purely interpretative digital humanities?
Distant readers vs. close readers	Are digital humanities only about "distant reading" (Moretti 2005)? To study literature, should we stop reading books and only focus on quantitative algorithmic measure (Marche 2012)? Can digital humanities also enhance close reading experience? Are "distant reading" approaches a form of radical digital humanities?

The term Big Data itself has disputed origins (Diebold 2012; Lohr 2013). The Oxford English Dictionary defines it as "data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges." In that sense, Big Data are "big" when "manual" analysis becomes cumbersome and new study and interpretation methods must be invented. However, massiveness of Big Data is not tightly linked to a certain number of Terabytes. Boyd and Crawford (2011) note that "Big Data is not notable because of its size, but because of its relationality to other data." Big Data is "fundamentally networked" and challenges in processing it are linked with its interconnected nature. In comparison, the Small Data Digital Humanities regroup more focused works that do not use massive data processing methods and explore other interdisciplinary dimensions linking computer science and humanities research. In comparison with Big Data, Small Data is small in the sense that it is not only smaller-scale but also well-bounded.

This article intends to draw a map for Big Data digital humanities showing how it can be organized as a structured field. The ambition of this map is to show that Big Data research in digital humanities can be characterized by common methodologies and objects of studies, therefore transcending some of the tensions that have structured digital humanities so far. As it focuses only on research that deals with these "large body of information" (Katz 2005), this maps does not cover the digital humanities domain as whole. Nevertheless, given the growing importance of massive and networked cultural datasets, it is likely that Big Data digital humanities become a significant part of the whole digital humanities field. In this context, this map may help institutionalize research and education programs with clearer focuses and objectives.

This article presents Big Data research in digital humanities as three concentric circles (**Figure 1**). The first circle corresponds to research focusing on processing and interpretation big and networked cultural data sets, the first object of study of this field. Most of the methods needed to study these datasets need still to be invented, as they are currently not mastered neither by humanists or computer scientists. However, it is important to consider that data processing and interpretation occur in a larger context of the new digital culture characterized by collective discourses, large community, ubiquitous software, and global IT actors. Understanding the relation between these entities could
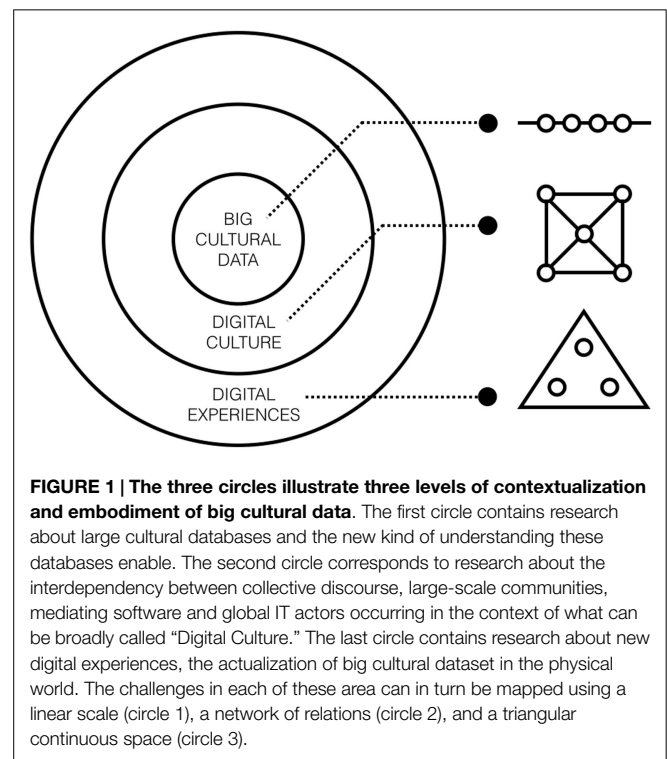

FIGURE 1 | The three circles illustrate three levels of contextualization and embodiment of big cultural data. The first circle contains research about large cultural databases and the new kind of understanding these databases enable. The second circle corresponds to research about the interdependency between collective discourse, large-scale communities, mediating software and global IT actors occurring in the context of what can be broadly called "Digital Culture." The last circle contains research about new digital experiences, the actualization of big cultural dataset in the physical world. The challenges in each of these area can in turn be mapped using a linear scale (circle 1), a network of relations (circle 2), and a triangular continuous space (circle 3).

be considered the second object of study for Big Data Digital Humanities. Eventually, the human experience of such datasets through various kinds of interfaces corresponds to a third family of challenges, differing in scope and methodology from the other two. Therefore, these three areas of studies could be represented as three concentric circles, illustrating three levels of contextualization and embodiment of cultural data. In the next sections, we will briefly discuss each of the circles in more details.

Big Cultural Datasets

Massive cultural digital objects include large-scale corpus like the millions of books scanned by Google and the ones produced by numerous other digitization initiatives (Jacquesson 2010), the millions of photos and micro-message shared on social network services (Thusoo et al. 2010), giant geographical information

systems like Google Earth (Butler 2006), or the ever expanding networks of academic papers citing one another (Shibata et al. 2008). These interconnected objects – either digitally born or reconstructed through digitization pipelines – are too big to be read or watched. The traditional 1:1 ratio of a single scholar confronted with one document cannot cope with such abundance. Moreover, their boundaries are sometimes fuzzy, their content partially unknown and, likely to be in continuous expansion. These characteristics make them profoundly different from corpora traditionally studied by humanities researchers, despite surface resemblances.

The confrontation with these "massive" objects calls for fundamental questions. What can really be extracted from these huge datasets and what interpretations can be drawn based on these extractions? Will we learn more by analyzing 10 millions books that we cannot read individually or by reading five carefully (Moretti 2005)? What is the role of algorithms for mining, shaping, and representing these large digital objects?

Some of these challenges can be structured following the specific parts of data processing: digitization, transcription, pattern recognition, simulation and inferences, preservation, and

FIGURE 2 | Challenges can be structured following the data processing pipeline. At each step, technical challenges are met and choices are made.

curation as show in **Figure 2** and in the **Table 2** below. Each step in the data processing pipeline can be associated with questions that are both technical and epistemological. Consider the processing pipeline of mass book digitization projects. Physical books must be transformed into images (digitization step) that are then transformed into texts (transcription step), on which various pattern can be detected (pattern recognition step like text mining or n-gram approaches) or inferred (simulation step) while being preserved and curated for future research (preservation step). This way of presenting the research challenge insists on the fact that data are never given, but taken and transformed (Gitelman 2013). The technical complexity of pipelines involved clearly demonstrates that, at each step of the data processing, choices are made and biases apply. Understanding these technical choices is crucial to develop new interpretive theories.

Digital Culture

We discussed the relationship between data processing pipelines and large cultural datasets. However, data processing and interpretation happen in a larger context, which we may call *Digital Culture*. The study of this large context can be considered to be the second object of study for digital humanities research. One way to structure this domain is to replace the relation between software and data (the focus of the first circle) in a network of relations between new entities including large-scale communities (MOOCs classrooms, Wikipedia contributors, etc.), collective discourses (Blogs, data journalism, wiki-style collaborative writing), ubiquitous software medium (auto-completion algorithm, search engine), and global actors (Google, Facebook, GLAM, Universities).

Consider the millions of photos shared every hour on Facebook (Huang et al. 2013). In this case, large-scale communities produce both the massive digital objects and the collective discourses
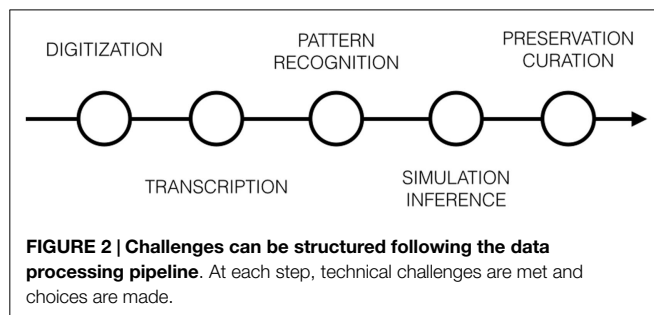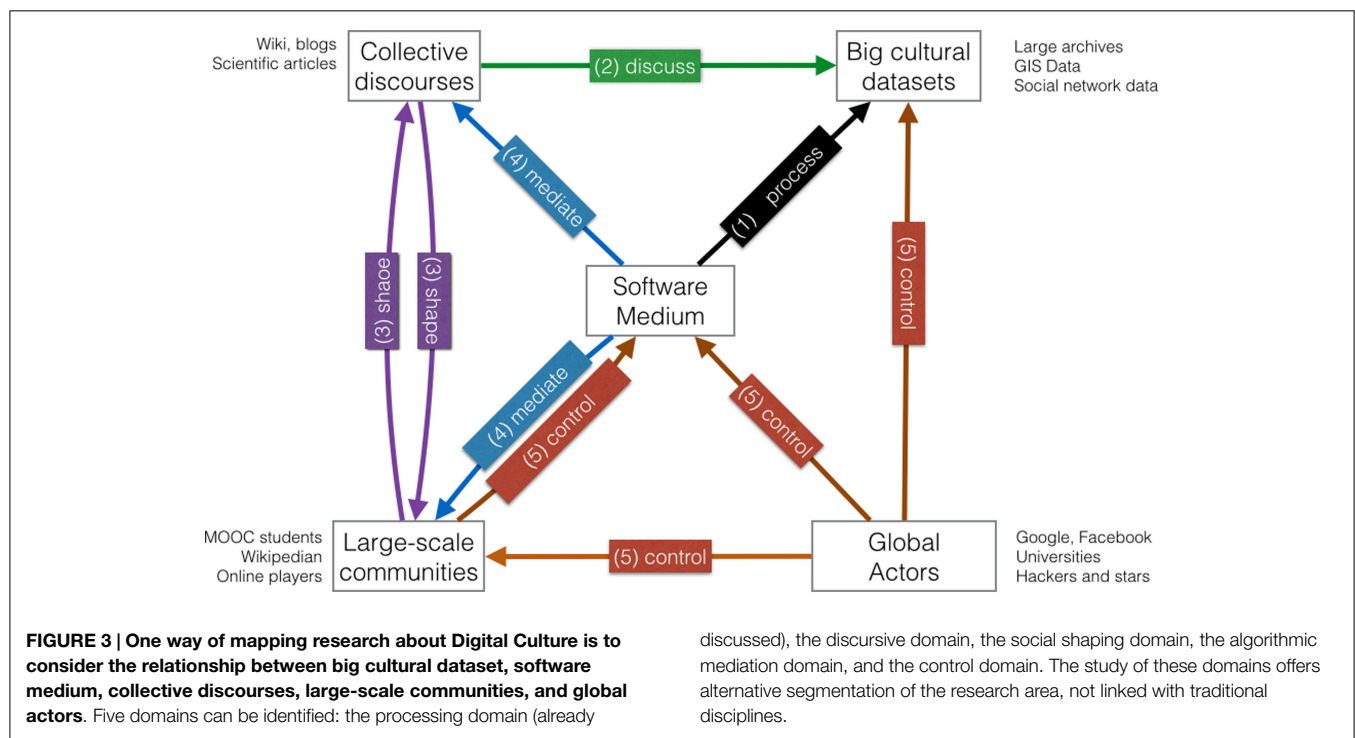
TABLE 2 | Challenges in circle 1.

Step	Challenges
Digitization	How can we develop more efficient, cheaper, faster digitization techniques allowing to perform mass-digitization programs (Coyle 2006; Lopatin 2006)? How can we develop new sensors and capture systems to obtain more information about the physical artifacts we study (Stanco et al. 2011)? How can we run crowdsourced digitization campaigns (Causer and Melissa 2014)? How can we upgrade datasets digitized with older technical methods (Paradiso and Sparacino 1997)? How can we perform efficient quality controls during digitization processes, anticipating the other steps of the technical pipelines (Liew 2004)? How can we store and compress information as it is being digitized? How can we attach metadata information documenting all these digitization processes?
Transcription	How can we "read" ancient documents (Antonacopoulos and Downton 2007)? How can we recognize specific features in paintings (Smeulders et al. 2000; Saleh et al. 2014)? How can we segment and transcribe audio and video content (He et al. 1999)? What kind of digital preprocessing needs to be performed to facilitate these transcription processes? How can automatic and manual processes be combined? How can we monitor the level of errors and the biases of algorithms in these transcription processes?
Pattern recognition	How can we detect common structural patterns in large collection of paintings, sculptures, and buildings models? How can we find names of people and places in texts (McCallum and Li 2003)? How can we classify the content of messages exchanged, detect events (Das Sarma et al. 2011)? How can we construct semantic graphs of data? How can we reconstruct and analyze networks from these data sets and trace the circulation of patterns?
Simulation and inference	How can we infer new data based on the data sets we study? How can we simulate missing data sets based on patterns detected? How can any uncertainty linked with these reconstructions be assessed (Bentkowska-Kafel et al. 2012)? How can we conduct simulation simultaneously at different scales? How can the inference, extrapolation, and simulation rules be attached to the data they produce in order to document this process (Nuessli and Frédéric 2014)?
Preservation and curation	How should data be stored to ensure both efficient short-term use and long-term preservation? What kind of storage support should be used? How can we assess their longevity? What kind of centralized or decentralized approaches are preferable? How much redundancy is needed? How should data be encoded to ensure traceability despite successive re-encoding? How can privacy, security, and authenticity of data be guaranteed? How can digitally born content be archived (Day 2006)?

TABLE 3 | Challenges in circle 2.

Domain	Examples of challenges
The *processing* domain (1) covers the interaction between software and massive digital objects from a technical and epistemological perspective, studying in particular how to design data-processing algorithms capable of deriving new data out of massive digital objects and how data becomes knowledge through complex processes of interpretation, or hermeneutics. This is a domain we have discussed in the previous section	Challenges of the processing domain have been discussed in the previous section
The *discursive* domain (2) covers the study of the shape of collective discourses in relation with massive digital cultural objects, from Facebook to scientific articles. All the natural categories of "digital linguistic studies" are relevant for this domain: lexical studies, grammatical studies, semantics, pragmatics, and semiotics	How do new technologies redefine scholarly discourses? How is the selective role of recognized academic journals challenged by new forms of open peer review (Shirky 2009; Fitzpatrick 2012b)? Can we imagine new publishing formats of "higher dimensions" allowing to embed videos, visualization interfaces, simulation engines, and source codes (Kaplan 2012)? What is the epistemological status of interactive visualizations? Can simulators be considered as a new kind of representation?
The *social shaping* domain (3) studies how large-scale communities shape and are shaped by the collective discourses they produce. This corresponds to typical sociolinguistic topics, adapted to the context of digital culture	What happens to authorship in crowdsourced projects or wiki-style contributions (Hoffmann 2008)? What is the role of automatic reading machines for plagiarism detection (Sloterdijk 2012) or new form of writing (Goldsmith 2011)? How does mass-digitization projects entail new specific copyright issues (Borghi and Karapapa 2013)?
The *algorithmic mediation* domain (4) covers how software mediates discourses and communities. This is an area traditionally covered by software studies (Manovich 2013; Kitchin and Dodge 2014)	Can the biases of search engines be studied (Rasch and Kanig 2014)? How can we assess the role of taylor-made interface and cultural filters (Pariser 2012)? Could auto-completion algorithms, machine translation, and other text-transforming algorithm have significant long-term effects on natural languages (Somers 1999; Kaplan 2014)? What is the role of algorithm in the structure of collaborative writing (Geiger 2011)?
The *control* domain (5) covers the relationship of communities and global actors with massive digital objects and the software medium. This domain studies how global actors curate both big cultural datasets and software medium to process them or how symmetrically, large-scale communities create or use software infrastructure, for instance, in the context of open source developer communities	Who controls the data? Who controls the software? Who controls the communities? How can control relationships be studied? How can the role of big actors be assessed and monitored this context (Battelle 2005)?

FIGURE 3 | One way of mapping research about Digital Culture is to consider the relationship between big cultural dataset, software medium, collective discourses, large-scale communities, and global actors. Five domains can be identified: the processing domain (already discussed), the discursive domain, the social shaping domain, the algorithmic mediation domain, and the control domain. The study of these domains offers alternative segmentation of the research area, not linked with traditional disciplines.

about massive digital objects. They do so through the mediation of algorithms produced by one giant IT company of the web. Retroactively, collective discourses about the photos have a shaping role on the emergence and structuration of these communities. In addition, as collective discourses reach rapidly a critical mass (e.g., millions of messages or status update) they tend to become themselves massive digital objects, to be archived and studied through specific text and data mining approaches. Understanding photo sharing implies understanding the complexity of this network of interactions.

More generally, research about digital culture can be segmented in subdomains corresponding to groups of relations between some of the entities we have been discussing. This structuration summarized in **Table 3** and **Figure 3**, identifies five domains: the processing domain, the discursive domain, the social shaping domain, the algorithmic mediation domain, and the control domain. This grouping articulates differently the relations of Big Data Digital Humanities with traditional humanities and social sciences disciplines, not considering that digital history, digital sociology, etc., but a new segmentation of domains.

Digital Experiences

Big cultural data, and digital culture at large, are experienced in the real world through physical interfaces, websites and installations. They produce "experiences." This third circle is an area of study on its own.

Some interfaces are essentially immersive, in the sense that they try to project the user into full-fledged environments (e.g., 3d Virtual World). Others provide users with synthetic data representations (e.g., network visualizations). Eventually, some interfaces are essentially linguistic allowing users to browse data via linguistic inputs (e.g., search engine). We can represent the space of possible interfaces with a triangle organized around these three summits (**Figure 4**). Conversational agents (e.g., SIRI) are in between the immersive and linguistics summits. Word clouds are in between abstract and linguistic summits. GIS interfaces can be sorted from the most abstract (Google maps, Open Street Map) to the most immersive (Google Street view). Augmented reality interfaces combine immersive, abstract, and linguistic dimensions. Each dimension of the interface space is associated with specific challenges, some of which are summarized in **Table 4**.
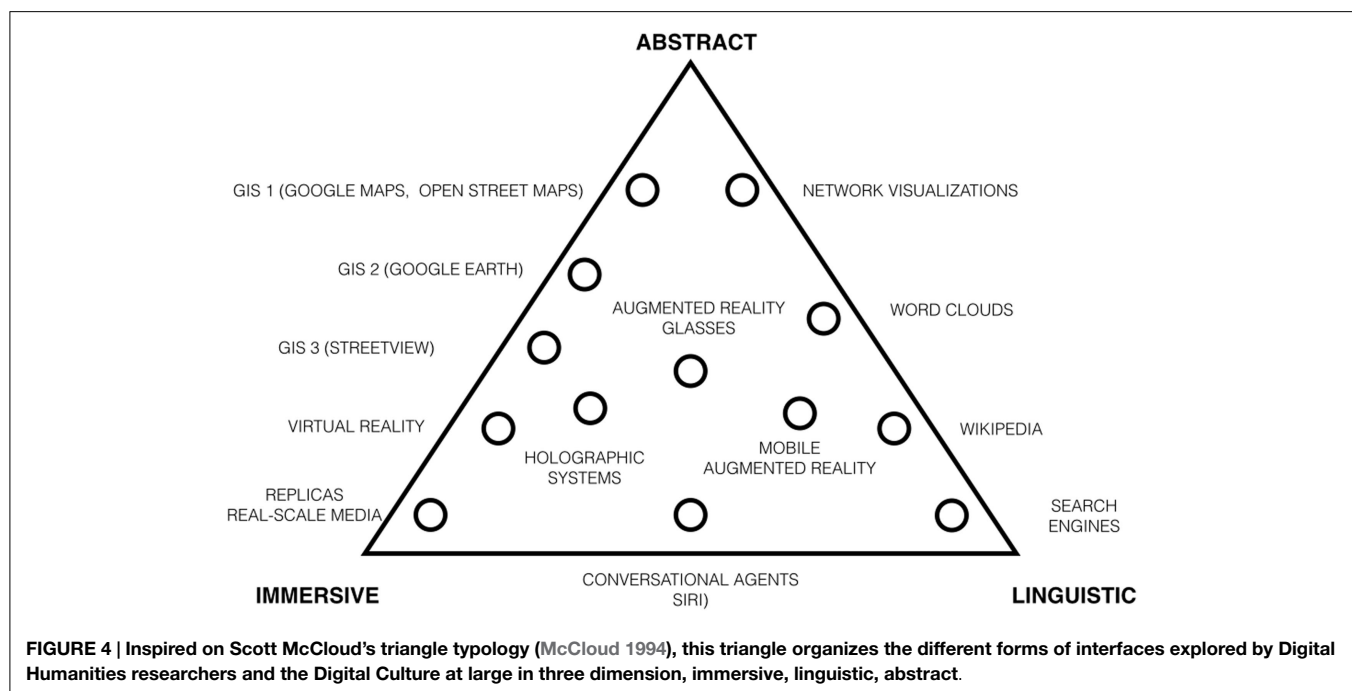

FIGURE 4 | Inspired on Scott McCloud's triangle typology (McCloud 1994), this triangle organizes the different forms of interfaces explored by Digital Humanities researchers and the Digital Culture at large in three dimension, immersive, linguistic, abstract.

TABLE 4 | Challenges in circle 3.

Dimension	Challenges
Immersive	How can effective immersion be designed? How can full-fledged environment be created based on big cultural datasets (Greengrass and Hughes 2008)? How can collective experiences occur in immersive situations? How can uncertainty in 3d world be conveyed (Bentkowska-Kafel et al. 2012)? How can the effectiveness of immersive environment be evaluated in various contexts (museum, schools, etc.)?
Abstract	How can dense representations be created out of large amount of data (Tufte 2001)? How can users navigate within abstract representations? How can multi-scale navigation be realized? How can users use data visualization to detect new patterns?
Linguistic	How can large quantities of text be visualized and sorted (Rockwell et al. 2010)? How can users navigate within different text layers? How can distant and close reading be combined?

Conclusion

Research in Big Data in digital humanities is becoming a well-structured field with specific objects of study. In this article, we identified three concentric areas of study and discussed how challenges in each area could be mapped. We illustrated how challenges focusing on the processing and interpretations of large cultural datasets can be organized linearly following the data processing pipeline, how challenges concerning digital culture at large could be structured around a network of relations between the new entities that emerged with the digital revolution and eventually, how challenges dealing with the experience of digital data can be described using the continuous space of possible interfaces. There are surely other ways of mapping this emerging field and the suggested structuration could be certainly refined and amended. However, we hope that this initial cartography will help paving the road ahead, acting as an invitation for exploring further the idea of Big Data Digital Humanities as a structured field.

References

Antonacopoulos, Apostolos., and Downton, Andy C. 2007. Special issue on the analysis of historical documents. *International Journal of Document Analysis and Recognition (IJDAR)* 9: 75–7. doi:10.1007/s10032-007-0045-1

Battelle, John. 2005. *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. New York, NY: Portfolio.

Bentkowska-Kafel, Anna., Denard, Hugh., and Baker, Drew. 2012. *Paradata and Transparency in Virtual Heritage*. Farnham: Ashgate.

Borghi, Maurizio., and Karapapa, Stavroula. 2013. *Copyright and Mass Digitization*. Oxford: Oxford University Press.

Boyd, Danah and Crawford, Kate. 2011. "Six Provocations for Big Data." A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, September 21, 2011. http://ssrn.com/abstract=1926431; http://dx.doi.org/10.2139/ssrn.1926431.

Butler, Declan. 2006. Virtual globes: the web-wide world. *Nature* 439: 776–8. doi:10.1038/439776a

Causer, Tim., and Melissa, Terras. 2014. Many hands make light work. Many hands together make merry work: transcribe Bentham and crowdsourcing manuscript collections. In *Crowdsourcing Our Cultural Heritage*, Edited by M. Ridge. 57–88. Surey: Ashgate.

Coyle, Karen. 2006. Mass digitization of books. *The Journal of Academic Librarianship* 32: 641–5. doi:10.1016/j.acalib.2006.08.002

Das Sarma, A., Jain, A., and Yu, C. 2011. Dynamic relationship and event discovery. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 207–216. Hong Kong: ACM.

Day, Michael. 2006. The long-term preservation of web content. In *Web Archiving*, 177–199. Berlin: Springer. Available at: http://link.springer.com/chapter/10.1007/978-3-540-46332-0_8

Diebold, Francis X. 2012. "A Personal Perspective on the Origin(s) and Development of 'Big Data': The Phenomenon, the Term, and the Discipline, Second Version." PIER Working Paper No. 13-003, November 26, 2012. http://ssrn.com/abstract=2202843; http://dx.doi.org/10.2139/ssrn.2202843.

Fitzpatrick, Kathleen. (2012a). The humanities, done digitally, debates in the digital humanities. In *Debates in the Digital Humanities*, Edited by M.K. Gold. 12–15. Minneapolis, MN: University of Minnesota Press.

Fitzpatrick, Kathleen. (2012b). Beyond metrics: community authorization and open peer review. In *Debates in the Digital Humanities*, Edited by M.K. Gold, 452–459. Minneapolis, MN: University of Minnesota Press.

Geiger, R. Stuart. 2011. The lives of bots. In *Critical Point of View: A Wikipedia Reader*, Edited by G. Lovink and N. Tkacz, 78–93. Amsterdam. Available at: http://www.networkcultures.org/_uploads/%237reader_Wikipedia.pdf

Gitelman, Lisa. 2013. *"Raw Data" Is an Oxymoron*. Cambridge, MA: MIT Press.

Gold, Matthew K. 2012. *Debates in the Digital Humanities*. Minneapolis, MN: University of Minnesota Press.

Goldsmith, Kenneth. 2011. *Uncreative Writing: Managing Language in the Digital Age*. New York, NY: Columbia University Press.

Greengrass, M., and Hughes, L.M. 2008. The virtual representation of the past. In *Digital Research in the Arts and Humanities Series*, Edited by M. Greengrass and L. Hughes. Ashgate. Available at: http://books.google.ch/books?id=ZZn3JnHW868C

He, Liwei., Sanocki, Elizabeth., Gupta, Anoop., and Grudin, Jonathan. 1999. Auto-summarization of audio-video presentations. In *Proceedings of the Seventh ACM International Conference on Multimedia (Part 1), MULTIMEDIA'99*, 489–498. New York, NY: ACM.

Hoffmann, Robert. 2008. A wiki for the life sciences where authorship matters. *Nature Genetics* 40: 1047–51. doi:10.1038/ng.f.217

Huang, Qi., Birman, Ken., van Renesse, Robbert., Lloyd, Robbert., Kumar, Sanjeev., and Li, Harry C. 2013. An analysis of Facebook photo caching. In *Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles, SOSP'13*, 167–181. New York, NY: ACM.

Jacquesson, Alain. 2010. *Google Livres et le futur des bibliothèques numériques*. Paris: Editions du Cercle de La Librairie.

Kaplan, Frédéric. 2012. How books will become machines. In *Lire Demain. Des Manuscrits Antiques à L'ère Digitale*, Edited by C. Clivaz, J. Meizos, F. Vallotton, and J. Verheyden, 25–41. Lausanne: PPUR.

Kaplan, Frederic. 2014. Linguistic capitalism and algorithmic mediation. *Representations* 127: 57–63. doi:10.1525/rep.2014.127.1.57

Katz, S.N. 2005. Why technology matters: the humanities in the twenty-first century. *Interdisciplinary Science Reviews* 30. 105–118. doi:10.1179/030801805X25909

Kitchin, Rob., and Dodge, Martin. 2014. *Code/Space: Software and Everyday Life*. Cambridge: MIT Press.

Liew, C.L. 2004. Digitizing collections – strategic issues for the information manager. *Library Collections, Acquisitions, and Technical Services* 28: 349–51. doi:10.1016/j.lcats.2004.05.008

Lohr, Steve. 2013. *The Origins of 'Big Data': An Etymological Detective Story*. Bits Blog. Available at: http://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/

Lopatin, Laurie. 2006. Library digitization projects, issues and guidelines. *Library Hi Tech* 24: 273–89. doi:10.1108/07378830610669637

Manovich, Lev. 2013. *Software Takes Command*. New York, NY: Bloomsbury Academic.

Marche, Stephen. 2012. *Literature Is Not Data: Against Digital Humanities*. Available at: https://lareviewofbooks.org/essay/literature-is-not-data-against-digital-humanities/

McCallum, Andrew., and Li, Wei. 2003. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003, CONLL '03*, Vol. 4, 188–191. Stroudsburg, PA: Association for Computational Linguistics.

McCloud, Scott. 1994. *Understanding Comics: The Invisible Art*. Reprint ed. New York, NY: William Morrow Paperbacks.

Moretti, Franco. 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History*. New York: Verso.

Nuessli, Marc-Antoine., and Frédéric, Kaplan. 2014. *Encoding Metaknowledge for Historical Databases*. Lausanne: Digital Humanities.

Paradiso, J., and Sparacino F. 1997. "Optical Tracking for Music and Dance Performance." It's conference paper presented at the Fourth Conference on Optical 3D Measurement Techniques, ETH, Zurich, September, 1997.

Pariser, Eli. 2012. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Reprint ed. New York, NY: Penguin Books.

Ramsey, Stephen. 2011. Who's in and who's out. In *Terras, Nyhan and Vanhoutte 2013, Defining Digital Humanities: A Reader, Édition: New edition*. Farnham: Ashgate Publishing Limited. Available at: http://stephenramsay.us/text/2011/01/08/whos-in-and-whos-out/reprinted

Rasch, Miriam. and Kanig, Rene. 2014. *Society of the Query Reader: Reflections on Web Search*. Amsterdam: Instituut voor Netwerkcultuur.

Rockwell, G. 2011. *Inclusion in the Digital Humanities*. http://www.philosophi.ca/pmwiki.php/Main/InclusionInTheDigitalHumanities

Rockwell, Geoffrey., Wong, Garry., Ruecker, Stan., Meredith-Lobay, Megan., and Sinclair, St. 2010. The big see: large scale visualization. *Journal of the Chicago Colloquium on Digital Humanities and Computer Science* 1. https://letterpress. uchicago.edu/index.php/jdhcs/article/view/65

Saleh, Kaplan., Abe, Kaplan., Arora, Ravneet Singh., and Elgammal, Ahmed. 2014. Toward automated discovery of artistic influence. *Multimedia Tools and Applications* 1–27. doi:10.1007/s11042-014-2193-x

Schreibman, Susan., Siemens, Ray., and Unsworth, John. 2008. *A Companion to Digital Humanities*. Malden, MA: Wiley-Blackwell.

Shibata, John., Kajikawa, Yuya., Takeda, Yoshiyuki., and Matsushima, Katsumori. 2008. Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation* 28: 758–75. doi:10. 1016/j.technovation.2008.03.009

Shirky, Clay. 2009. *Here Comes Everybody: The Power of Organizing Without Organizations*. Reprint ed. New York, NY: Penguin Books.

Sloterdijk, Peter. 2012. Plagiat Universitaire: Le Pacte de Non-lecture. *Le Monde*. http://www.lemonde.fr/idees/article/2012/01/28/le-pacte-de-non-lecture_ 1635887_3232.html

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., and Jain, R. 2000. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22: 1349–80. doi:10.1109/34.895972

Snow, C.P. 1959. Introduction. In *The Two Cultures and the Scientific Revolution*, Edited by S. Collini, 1993. Cambridge: Cambridge University Press.

Somers, John. 1999. Review article: example-based machine translation. *Machine Translation* 14: 113–57. doi:10.1023/A:1008109312730

Stanco, Filippo., Battiato, Sebastiano., and Gallo, Giovanni. 2011. *Digital Imaging for Cultural Heritage Preservation: Analysis, Restoration, and Reconstruction of Ancient Artworks*. CRC Press.

Svensson, P. 2009. Humanities computing as digital huminites. *Digital Humanities Quaterly* 3: 3. http://www.digitalhumanities.org/dhq/vol/3/3/000065/000065. html

Terras, Melissa. 2011. *Peering Inside the Big Tent*. reprinted in Terras, Nyhan and Vanhoutte 2013, *Defining Digital Humanities: A Reader, Édition: New edition*. Farnham, Surrey, England: Burlington, VT: Ashgate Publishing Limited. Available at: http://melissaterras.blogspot.ch/2011/07/ peering-inside-big-tent-digital.html

Terras, Melissa., Nyhan, Julianne., and Vanhoutte, Julianne. 2013. *Defining Digital Humanities: A Reader. Édition: New edition*. Farnham: Ashgate Publishing Limited.

Thusoo, Ashish., Shao, Zheng., Anthony, Suresh., Borthakur, Dhruba., Jain, Namit., Sen Sarma, Joydeep., et al., 2010. Data warehousing and analytics infrastructure at Facebook. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data, SIGMOD'10*, 1013–1020. New York, NY: ACM.

Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. 2nd ed. Cheshire, CT: Graphics Press.

Unsworth, J. 2002. What is humanities computing and what is it not? In *Jahrbuch für Computerphilologie*, Vol. 4, Edited by G. Braungart, K. Eibl, and F. Jannidis, 71–84. Paderborn: Menis Verlag.