Bernhard Rieder, Assistant Professor, Laboratoire Paragraphe, Université de Paris VIII
Theo Röhle, Postdoc Researcher, Graduiertenkolleg Automatismen, Universität Paderborn

Digital Methods: Five Challenges

"*There is no safety in numbers, or in anything else.*" - James Thurber

With the immense success of the Web, the richness and complexity of the objects and actors that populate the Internet has become quite astounding. The intensification of cultural production, distribution, and consumption online poses new challenges for the humanities and the social sciences. Faced with an abundance of empirical material, researchers are increasingly turning towards automated methods of analysis in order to explore the new wealth of datasets. These computational tools hold a lot of promise: they are able to process much more data than would ever be possible manually; they promise extended zoomability between micro and macro and the ability to reconcile breadth and depth of analysis; they can help to reveal patterns and structures which are impossible to discern with the naked eye; they are even probing into semantic relations and meaning.

So, something is definitely happening in the humanities. A discipline that has held numbers, calculations, and computers at a safe distance for a long time is suddenly warming up to the new computational possibilities. "Digital Humanities" and "Cultural Analytics" are the new buzzwords, data visualisation projects abound, and, if we believe the title of the conference, we may even be in the middle of a "Computational Turn". But what are we to make of all of this? Are these new digital methods just another set of tools in our arsenal or do they represent a deeper transformation that challenges established epistemology? Could there be a danger of losing something in the process? Do we overlook something important with our eyes fixed on the digital frontier?

Our starting point for investigating these questions is the problem of method, loosely defined as a systematic way of pursuing knowledge. The term "Digital Humanities" designates a wide range of software tools, not all of which actually pretend to the status of method. Some of them target auxiliary functions such as communication, organization and archiving, or pedagogy. What interests us here, however, are tools that explicitly function as methods: they process data systematically and associate certain "truth claims" with their results. Two types of software can be distinguished here.

The first consists of automated versions of existing manual methods, often related to deduction and hypothesis testing, such as regression analysis and other concepts from statistics, game theory, or linguistics. The central question here is how these methods are affected when they move into the digital realm and are implemented as software. Despite ongoing debates about the validity and epistemological standing of these established methods, there are traditions of reflexivity and established knowledge on pitfalls, best practises, quality standards, etc. Still, the change in technology may have important consequence for how these methods are used, how they evolve, and how they produce knowledge. New visualisation techniques, for example, may

open up new possibilities for communicating the results of established statistics. Interactive toolkits could supplant the teaching of basic formulae.

The second type of tools, which can be broadly subsumed under the term "data exploration", represents a more inductive tendency. While they do not pretend to verify or falsify hypotheses, they still try to generate knowledge about the data they analyse. By rendering certain aspects, properties, and relations visible, they offer us particular perspectives on the phenomena we are interested in. While their results may be visually impressive and intuitively convincing, the methodological and epistemological status of their output seems unclear at best. Nevertheless, it is these very tools that provoke the most enthusiastic reactions. What is rarely reflected by advocates of an "end of theory" (Chris Anderson) though, is that theory is already at work on the most basic level of methodology, i.e. when it comes to defining units of analysis, algorithms, and visualisation procedures.

In both cases, digital technology is set to change the way scholars work with their material, how they "see" it and interact with it. The question is, now, how well the humanities are prepared for these transformations. If there truly is a paradigm shift on the horizon, we will have to dig deeper into the methodological assumptions that are folded into the new tools. We will need to uncover the concepts and models that have carried over from different disciplines into the programs we employ today (and tomorrow). This would seem like a task well suited to the humanities. But only if we don't get bogged down, neither by naive technological enthusiasm, nor by a paralysing fear of technology. We will also have to watch out for our (sometimes not so) secret longing to compete with the natural sciences on their own terms, by being as "objective", as "rigorous" with the help of machines.

What follows is a non-exhaustive list of issues that we believe will have to be addressed it we want to productively integrate the new methods, without surrendering control over the conceptual infrastructure of our work.

1. Objectivity

Despite frequent assertions that the humanities produce interpretive rather than verifiable knowledge, there is a latent doubt, a certain kind of "envy" of the cold rationality, objectivity, "down to the essentials" of the natural sciences. The scientific ideal still follows the idea of mathematics. The roots of this ideal can be traced back to the historical proponents of objectivity like Russell, Poincaré, and Carnap, who explicitly excluded subjective elements from the realm of science. Scientific knowledge for them encompassed only structural features that could be established intersubjectively. Empirical methods that rely on strong formalisation epitomise this ideal, especially if they are implemented in an automated way. Automatic collection and processing appears to remove the data one step further from the perils of human error and subjective judgement.

Two important aspects are often ignored when the ideal of objectivity is invoked: Firstly, there is much in the natural sciences that is open to debate, especially in mathematics. In the wake of the strong programme, STS has provided a range of examples showing that these fields are just as connected to societal developments as the humanities. Secondly, even the most rigid meth-

odological process involves elements of subjective judgement. Since units of analysis are never given, decisions at the level of selection, operationalisation, modelling etc. are taken on the basis of rote knowledge in the field. Thus, even a simple quantitative transformation is a proto-interpretation.

Exposing ones own empirical research to this kind of scrutiny is hardly attractive and the methodological questions can easily be buried under layers of impressive data output. Since the underlying problems cannot be avoided, though, it seems wise to confront them and integrate them into the research agenda. This calls for a highly reflexive attitude towards our methodological and epistemological premises. A pressing question is whether we have the terminology to discuss these matters and, even more important, to convey them to colleagues in the humanities.

2. Visual evidence

The visual output of the new tools poses similar kinds of questions, but with a special urgency. Visualisations have a special explanatory role since a) they often have to reduce informational dimensions to fit graphical requirements, and b) they emphasise certain aspects at the expense of others by choosing and specifying modes or representation.

Representations of network topology filter out individual aspects in order to visualise structure. This way, they tap both into the ideal of objectivity mentioned above and into the power of visual evidence. Visualisation has a long history as a rhetoric device in the sciences; it is one of the prime vehicles for reducing complexity and conveying a certain perspective on the material.

The fact that a visualisation is not given, but always a specific projection of the data is often forgotten in the process, especially when visual competence is lacking. This opens up for manipulative uses of visualisation (a standard example being the use of different scales in a coordinate system). But even without manipulative intent, it is important to stress that there is no self-evident transformation from data to output. Due to the iterative nature of the algorithms in many of the new tools, changing small parameters can have substantial consequences for the visual output.

The central question that has to be tackled is the epistemological status of the visual results our new tools produce. It seems clear that they have to be disconnected from notions of objectivity and from notions of a direct transfer from world to image. Rather, visualisations are specific kinds of representations that involve specific kinds of reductions. But is it therefore feasible to treat them *exclusively* as a rhetoric device? The visualisations seem to carry some kind of valid proposition about the world, but how can their range be properly delineated, what are kosher ways to integrate them into a scientific argument?

3. Black-boxing

Another element is related to the technological underpinnings of the new methods. In creating truly digital methods, we mechanise a (variable) part of the scientific process. This process of

formalisation into data structures, algorithms, modes of representation, and possibilities for interaction paradoxically reduces the transparency of the methodological procedures. Transparency, in this case, simply means our ability to understand the method, to see how it works, to reproduce it, and to criticise it. An open process of scrutiny is, however, a pillar of science's claim to social legitimacy. Technological black-boxing may prove to be a major issue if digital methods become more widespread. A solution may consist of publication of the source code combined with detailed technical specifications that help others understand how the software works. But even with technical transparency guaranteed, at least two problems remain:

a) The first concerns the classic *two cultures* problem: even if specifications and source code are accessible, who can actually make sense of them? Algorithms for automatic classification based on full-text analysis or techniques for producing readable graph layouts are by no means trivial. Methodological coherence will in part depend on our ability to transpose our epistemological enterprise into the language of computer science and back.

b) The second problem is even more diffuse. Some of the tools computer science uses are positively *experimental*, in the sense that the results they produce cannot be easily mapped back to the algorithms and the data they process. Many of the techniques issued from the field of machine learning show a capacity to produce outputs that are not only unanticipated but also very difficult to backtrack to the inputs. Even in purely deterministic systems, small variations in the data or in system parameters may have far-reaching consequences, especially in techniques with high iteration count (most of them). What we are trying to say is that techniques imported from the computer sciences may never be understood in the same way we understand statistical concepts like variance or regression.

These two caveats apply specifically to deductive epistemologies with strong claims to objectivity and scientific validity. In contexts where tools are used rather to generate ideas than to "prove" them, epistemological concerns may matter less. But still, without the full consciousness of what it means to mechanise methodology, we may find ourselves in a situation where large parts of knowledge production are delegated to software tools that we do no longer understand.

4. Institutions

A fourth point for discussion can be derived from a sociological viewpoint on scholarly work. The speed and efficiency of the new methods, their capacity to analyse large chunks of data, may render more time-consuming and labour-intensive approaches unattractive and too costly. In settings where even humanities' research is increasingly financed on a project basis - which implies very particular pragmatics based on clear time-frames, planned expectations, and identifiable "deliverables" - digital methods have a clear advantage.

This can have consequences for recruiting (think computer science PhDs appointed to humanities professorships) and, more importantly, affect our idea of what "good work" looks like. To put it in provocative terms: will the scholarly memoir hold up to the interactive full-screen animation? There is a true danger of catering too much to short attention spans while at the same time cruising on technology's aura of objectivity. To counteract these potential dangers, we will

have to understand not only the potential but also the limits of the new methods. Obviously, building research tools is not an end in itself an there is an argument to be made for the confident defence of methods that are built on other principles than mechanical reasoning.

5. Total Science

A fifth concern is the larger philosophical and theoretical outlook that is incorporated in our methodological apparatus. By producing a particular representation of the world, our tools may act as agents of a specific ontology, with far-reaching effects on our research agenda. A prime example is the representation of network topologies. Many of these projections foster a trend towards a reification of structure, simultaneously extending the explanatory reach of these structures.

In this sense, network theory appears as a predecessor of earlier attempts at creating a universal science, such as structuralism, cybernetics, and systems theory. These attempts are fuelled by the hope to be able to reconcile widely diverging disciplines around a common explanatory hub. The failure of these attempts should make us particularly wary about new aspirations and pretentions at creating a new total science. Instead we should focus upon explicitly limiting the scope of the explanations produced by the new tools.

Conclusions

These five areas of potential difficulty are not meant to discourage from the exploration of new digital methods; they rather represent a list of issues that will have to be addressed if we want these new approaches to be epistemologically sound. While uncritical enthusiasm is out of place and the challenges we face are indeed considerable, there is one aspect in the whole story that we see as a positive development, independently from the actual results software tools might deliver. The omnipresence of digital technologies has already forced the humanities to pay considerable attention to the dominant technological paradigm of our time, the computer. The rise of digital research methodologies may encourage us to take the next step and develop a better understanding of the inner workings of software. The delegation of certain aspects of knowledge production - and consequently of decision-making - to computers is a phenomenon that is not confined to scholarly research but a larger process that touches most areas that depend heavily on information processing. The confrontation with advanced software tools in our own fields can help us in formulating the theories and forging the disciplinary alliances we need to build a critical understanding of the infiltration of software into every pore of contemporary society.