15

Cultures of Formalisation: Towards an Encounter between Humanities and Computing

Joris van Zundert, Smiljana Antonijevic, Anne Beaulieu, Karina van Dalen-Oskam, Douwe Zeldenrust, and Tara L. Andrews

Introduction

The past three decades have seen several waves of interest in developing cross-overs between academic research and computing; molecular biology is often cited as the prime exemplar of 'what computation can do for a field'. The humanities and social sciences have also been the terrain of such interactions, at times through bottom-up collaborations, and at times through concerted policy-driven efforts (Wouters and Beaulieu 2006). The main developments vary across national contexts and disciplines. In our local context (in the Netherlands), we can roughly identify the following waves: the 'history and computing' and 'literature and computing' efforts of the 1970s and 1980s; the collaboratory and infrastructure discussions of the last decade; the current efforts at developing computational humanities, and recent emphasis on virtual research environments (VREs) of which Alfalab[1] can be regarded as an example.

Efforts to introduce computational methods typically involve collaborative work between scholars and engineers and the combination of their complementary skills and expertise. Along the lines of Traweek (1988) and Knorr-Cetina (1999), we consider such collaborations as encounters between 'epistemic cultures', that is to say, particular combinations of meanings, material arrangements, and practices that organise an area of scholarly work. In this chapter we focus specifically on formalisation, and we use an analytic metaphor, 'cultures of formalisation', as a means to highlight the epistemic variety underpinning formalisation practices in different epistemic cultures. We argue that critical reflection on formalisation practices is important for any computational program to succeed, and that this is of particular importance in the humanities domain – foremost because experience shows that rigid prescriptive heuristics

and mandatory explicit formalisations, if uncritically imposed from a computational paradigm, generally do not 'land well' in different humanities research communities. While both computational sciences and humanities can be intrinsically characterised by their epistemological and methodological openness to complexity and uncertainty, their sensibilities as to what is acceptable in terms of heterogeneity of method and approaches often do not overlap. By conceptualising and describing cultures of formalisation in the humanities, we can also identify aspects of research that could be better supported if suitable and compatible computing approaches are developed. An approach that stresses cultures of formalisation can therefore also enrich the computing research agenda, and contribute to more symmetrical and constructive interactions between various stakeholders in computational humanities.

From these initial observations, our exploration takes on three forms. First, we propose to look more closely at formalisation and to question whether it is a singular concept. Second, we ask whether formalisation is also an aspect of research in the humanities, even without (necessarily) thinking of it as driven by computation, and we present four case studies that help us explore that question. Finally, we consider how our analysis enriches what can be understood by formalisation, and what kind of light it throws on the encounter between computing and humanities. In the final section we consider how our explorations and findings influenced developments within Alfalab.

What is formalisation in humanities computing?

Formalisation is highly recognisable as a basic principle underpinning the logic of computation. This has led to the perception that any computational approach in the humanities should be rooted in formalisation as well. Repeated failed attempts at deploying computational approaches in the humanities have then been attributed to the lack of formalisation, and, in the worst cases, even the lack of apprehension of formalisation by humanities scholars. Indeed, at first blush it may seem evident that formalisation of humanities research heuristics and hermeneutics is closely tied to the rise in computational efforts. But it is a fallacy to posit that formalisation is a new development in humanities, driven solely by computation. In music theory, for instance, a trend toward formalisation has been present since the 1960s, prompting an increased interest of computer scientists and psychologists in musicology, rather than the other way round (Honing 2006). Such examples support our argument that formalisation is not a newly emerging element in humanities research and, moreover, that methods of formalisation can facilitate interdisciplinary scholarship. Rather than simply a necessary and straightforward condition for computation, formalisation is a rich and productive concept to think about computation, the humanities, and their encounters.

If formalisation is a key component of humanities computing, what form does it take? As noted above, there is a strong tendency in computing to emphasise the importance of formalisation in order to deploy computational approaches – that is to say, to point to the need to explicitly define properties of research objects. Yet we also find formalisation in other aspects of computing in humanities research. Data-sharing also demands formalisation: formalisation of notions of authorship and ownership of data; the formalisation of research methods; and the formalisation of annotations (see Arzberger et al. 2004; Beaulieu 2003). Formalisation is therefore far from a homogeneous standard of quality. It is not an elusive status of epistemic purity to be attained by research objects before they can be 'computed'. So why does formalisation seem to come to the forefront so prominently in the context of interactions between humanities and computing?

One of the probable reasons is that encounters between different fields tend to throw difference into relief. The need to explain, to make explicit what one does and how, will tend to highlight processes of articulation, which are related to formalisation. But, while formalisation has been around implicitly in humanities research, making certain kinds of formalisation explicit through the use of computational methods appears as an almost hostile act within some of the humanities domains. We posit that this is because the kind of formalisation that is put forward as a necessary condition for computing is *only one kind of formalisation*, currently dominant in computing but far from universal across research domains. By paying attention to this mismatch in kinds of formalisation, we can see the underlying reasons for resistance or, as is more often the case, indifference, of scholars to computational tools that are proposed.

In order to illustrate these points, we now turn to four examples of formalisation in different humanities domains. These case studies show the distinct approaches, modes, and realisations of formalisations that exist and emerge in various humanities projects taking a computational turn.

Case studies

Our case studies are pursued in the framework of Alfalab and in various institutions inside and outside the Royal Netherlands Academy of Arts and Sciences, including the Virtual Knowledge Studio for the Humanities and Social Sciences and the University of Oxford. Alfalab functions as a meeting point for these various endeavours, enabling encounters that, in turn, can foster critical reflection on our respective projects.

Hypothesising history

In the preparatory phase of research Joris van Zundert and Tara Andrews are carrying through, the objective is to explore the possibilities of computationally

inferring and visualising the hypotheses dependency structure underlying argumentation in historical interpretation.

Historical reconstruction, particularly for medieval periods, rests on scraps of evidence, surmises about its quality and biases, and attempts to fit it into a framework of what we think we know. Elaborate hypotheses must be built to explain our evidence, many of which require intricate argumentation that depends on other hypotheses. These feats of historical analysis can be deeply impressive and thorough (e.g. Dédéyan 2003) but at the same time very dangerous. How can historians assess the full extent of the impact of a new piece of evidence, such as evidence of a market economy found in an area 'known' to have suffered a 'dark age', when it challenges assumptions that have been part of the basis for our understanding of the entire period (cf. Auzépy 2007)? As 'generations' of interpretation build layer upon layer of hypotheses, the complete supportive structure of hypotheses becomes too complex to 'compute' fully, even for the finest minds. Problems of contradictory interpretation due to conflicting structures of hypothesis within one and the same synthesis are a demonstrable result of this complexity.

Our work is therefore a supportive analytical task. It seeks to infer explicitly the relations between hypotheses and to evaluate the internal consistence of the hypothesis dependency trees. This task could be a very valuable addition to the method of aggregative argumentation and interpretation, which is at the core of research heuristics in historical studies. It is also a task that could be highly facilitated through a computational approach. For instance, analysis techniques like topic maps (cf. Bock 2009) and/or rhetorical structure theory (cf. Mann 1986) can readily be applied to describe or visualise argument dependency trees. However, to be able to apply these techniques, we need specific forms of hypothesis and argument formalisation, both of which can be viewed as complex tasks of data curation. We need to be able to capture the hypothetical argumentation in a way that is both simple to apply and unambiguous to a dependency-computing algorithm. Unfortunately, such means of expressing argumentative structure are still a rather abstract form. Instead of taking the form of the original text, the argumentative structure is represented in the form of a series of symbols capturing the argumentative statements and the formalised relations between those statements.

Relevant to the subject of this chapter is the preliminary observation that the process of formalisation and the knowledge of logic involved seem to be non-tangible and actually rather inhospitable to researchers in the field of history. Formalisation of both argumentative structure and the 'surface structure', that is, the very personal idiolect and style of the argumentative text, evokes feelings of 'meddling' with the core capacity, competence, method, and techniques of the researchers. From the computational perspective, however, these formalisations are merely descriptive, and they serve no other purpose than to

compute and visualise the prior hypotheses upon which an argument depends, empowering the researcher to self-evaluate and infer the soundness of argument. At the same time, it is self-evident that the process of transcribing an argument structure from painstakingly stylised idiolect into a computable form is an alienating process for the researcher, not least because idiolect is arguably one of the strongest identifiers of 'self' there is. This is why formalisation in this case creates the risk of resistance and distrust within the targeted research community (historians, in general, and Byzantinists, in the prototype phase of the research, in particular). Furthermore, in order to make formalisation a useful activity, historians need to trust formalisation and not perceive it as an alienating research practice. This will only be achieved if formalisation does not appear as an intrusive process and daunting activity for those researchers. This could be accomplished, for example, by utilising – or even creating – much higher-order computing languages than usually applied. However, this requires computational engineers to recognise form and representation as important factors in establishing trust and enabling unobtrusive formalisation.

The onymic landscape

Names are so common in our daily lives that we tend to overlook them. Still, names are often a cause for laughter, teasing, and – worse – discrimination. Expecting parents take an enormous amount of time and energy in deciding how to name their child. Individuals sometimes change their own names. And, occasionally, names of cities or streets are changed. Such examples show that names are more than just tools for discriminating or referencing entities. Their very characteristics make names a subtle stylistic tool in literary texts too. Often an author can create an idea or certain expectations of a character in a story just by mentioning a name. Tension can be created by mistakes in names, or by the introduction of aliases, which are only solved at the appropriate time (from the storyteller's point of view). Names can imply a geographical and/or social background of characters. In the case described here, Karina van Dalen-Oskam aims to analyse the usage and functions of names in literary texts from a comparative point of view, that is, between texts, oeuvres, genres, time periods, and even languages (Van Dalen-Oskam 2005).

The first results of the research will become available in 2011.

The study of names in literature is a sub-discipline of onomastics (name studies). Until recently, research in literary onomastics was very eclectic, only pointing out 'significant' names in a literary work and describing their role in the text(s). About a decade ago, scholars started to emphasise the need to look at all the names in a text (or oeuvre, or genre, etc.) and to analyse the so-called onymic landscape. Only with such an analysis can we be sure which names are really significant and have an extra role in the plot of the story, as motive- or theme-bearing stylistic elements.

This, however, is easier said than done. It takes too much time for one researcher to reconnoiter a large onymic landscape. This has to be done by many scholars, and they need to approach their data in a comparable way in order to make their 'mapping' useful for each other. There are no useful software tools yet for this type of work, which is why creating such tools is part of the Alfalab endeavour. It may seem that names are easy to find in texts, and that usage of tools such as those for named entity recognition and classification (NERC) could make such research very easy. But this is not the case. Tools for NERC usually focus on one language only, and for that language on just a few text types, mainly texts from specific topic areas or from newspapers. Even then, the maximum success rate only occasionally exceeds 80 per cent (Sekine and Ranchhod 2009). Literary texts contain many sorts of name types, so several NERC tools would have to be applied before getting a result that still needs a lot of manual cleaning up.

The comparative literary onomastic research Van Dalen-Oskam conducts will look into the frequency and function of names occurring in literary texts. It will also observe the ratio of types of names (e.g. personal first and family names, place names, etc.). To enable comparative analysis, scholars working on different texts will have to follow the same rules. This of course constitutes a kind of formalisation. A deceptively simple challenge, the formalisation of identifiable properties of names and their uses, is in fact a cumbersome task. For example, the most common names in literary texts are personal names. These can be divided into first names, bynames, and surnames. Scholars will usually agree on what category a name belongs to, but what to do with classical names such as Julius Caesar? Julius was not a first name and Caesar not a family name. We can agree about 'Jesus' being a name, but what about 'God'? The second most frequent name type, place names, is also relatively easy to agree on. But is Niagara Falls a location (so a place name) or an object? Several other categories are denoted as being names in theory, for example, currencies and time indications, or as names in one usage while not in another (van Langendonck 2007). But such categories are probably not very relevant to a literary analysis of the texts. And on top of that, scholars would need extensive training before being able to spot these correctly. So these will be excluded from the comparative literary onomastic research. Part of this project therefore involves finding the right degree of formalisation, to do justice to the research goals at hand, to make use of the potential of tools to support comparative analysis, and to maintain the feasibility of the endeavour.

Microtoponymy

Microtoponym is a term used for names of small to very small entities in both natural and human-made landscape. Imagine, for example, a small field called

'the Gallows'. It may very well be that the name is not even formally designated in any land registry but is only known and used by the local population. Such microtoponyms are part of the object of research and formalisation in yet another pilot of the Alfalab project, called GISLab. This specific exploratory research is interested in applications of Geographical Information Systems (GIS) as a suitable platform for humanities research.

Within the broad spectrum of the humanities, the study of onomastic variation, and in particular the study of place names, or toponymy, may come across as a very specialised niche. But microtoponyms are actually of interest to historians, historical geographers, and archaeologists, among others. In the Netherlands, researchers from different disciplines agree that having the Meertens Institute collection, which contains more than 200,000 microtoponyms and their geospatial parameters digitally available, would facilitate and open up new avenues of research in various subjects (Zeldenrust 2005).

Regarding the topic of this chapter, it is useful to point out that in the process of digitising and utilising a legacy consisting of 200,000 physical index cards, systematically recording toponyms and their metadata, formalisation indeed does play a role on several levels: determining functional requirements of the GIS by interviews and studying prior work; development and implementation; and so on. However, the microtoponym case study is less about formalisation in the sense of making heuristics explicit. It is more about formalisation at the level of the objects of research, which, through being formalised and situated in a digital context, can offer new research possibilities.

The following example may serve to clarify the previous statement. In the Dutch context, Schönfeld's book, *Veldnamen in Nederland,* or *Microtoponyms in the Netherlands* (Schönfeld 1950), provides a usable starting point for onomasticians concerned with microtoponyms. However, although Schönfeld's work is still a standard in its field, it was written in 1949, long before anyone envisioned a field called computational humanities. The computational approach and interdisciplinary character of the microtoponym virtual research environment (VRE) that will be established through Alfalab could enhance the present onomastic community and create a new interdisciplinary one. Therefore, an a priori focus on descriptive or prescriptive formalisation based on recognised yet superseded theory would potentially hinder the exploration of new and cross-disciplinary possibilities. On the other hand, explicit generic formalisation would also be hard to achieve. For example, since the visual aspects of a GIS specifically allow researchers to make their own interpretations of certain maps *separately* from other researchers' interpretations, their actual method of interpretation can remain fully implicit. The advantages of formalisation in this case have to do with opening up the possibilities for interacting with microtoponyms as digital objects in a VRE.

The empirical image

This case study concerns the use of Flickr as used by researchers who explore graffiti and street art.[2] The case study focuses on the constitution of Flickr as a resource and means of interaction between researchers and empirical material. In every field there is an accepted way of constituting one's object of research, and this aspect of research is a key dimension of epistemic cultures.

Best known as a photo-sharing platform, Flickr can also be used to build a personal archive of photos, to browse material uploaded by (un)known others, or to engage in a wide variety of activities around photos. Flickr has several features of 'ongoing sociability' (Fuller 2003) typically associated with social networking sites. It enables users to represent themselves and to articulate links to other users and the content they upload. Furthermore, Flickr, like other social networking platforms, makes use of traces generated by use of the system and its content, a defining feature of Web 2.0 applications.

The researchers studied in this case are mostly (visual) sociologists or anthropologists, who focus on urban and/or material culture. Among the huge variety of photos on Flickr, urban photography and the documentation of urban life is a prevalent theme (Petersen 2009). All of the researchers use photography as part of their research practices, which they define as 'fieldwork'. Through interviews, email exchanges, analysis of articles and other output, and the researchers' use (or, in one case, vehement non-use) of Flickr, this case study is able to characterise how Flickr is used in relation to empirical material in the researchers' work.

Researchers use Flickr as a source to throw further light on material they have gathered in their fieldwork, by connecting different bits of empirical material. This use resembles searching, browsing, and 'googling' on the Web, but more specifically in relation to visual material and to street culture, for which Flickr is an especially good source. Visual material is also notoriously under-served by search engines, which are oriented to textual, (and even ASCII) material.

This use of Flickr depends on the presence of material from huge numbers of contributors, and, significantly, on the use of recognisable tags or labels. Tagging and labelling subtend formalisation of content, meaning, or significance of aspects of images. While often done without much conscious effort, the seemingly banal gestures of tagging and labelling are important practices that facilitate the constitution of Flickr material as empirical sources. Consider that most of these researchers have very strong feelings about the use of captions for their photos, and condemn these as parasitic textual practices that undermine the narrative power of the visual material. Yet all of them assign titles and tags to their photos on Flickr. These are usually summary, but, nevertheless, they label the photo with a transcription of the 'tag' text (i.e. the 'name' of the writer of a graffiti or artist). Locations are also often used as tags. This labour in turn enables Flickr to function as a searchable

source. Tagging is a recursive practice in these settings: one can deliberately use popular tags for one's photos in order to generate 'views'. This recursive aspect shapes the constitution of categories and modes of organisation of this material.

This case study illustrates how visual material is made usable through formalisations that involve textual labels, which are useful to some extent for researchers – and certainly interesting as an emergent phenomenon. But this reliance on textuality is far from desirable for some researchers (Beaulieu, van Heur, and de Rijcke n.d.). Visual formalisations that do not rely on text would highlight different aspects of this empirical material for researchers. This is a case that suggests how computational approaches might be developed to better serve researchers' needs in relation to their empirical material. Possibilities to formalise image data in ways other than textual labelling, and to make them empirically useful to humanities researchers, would be a valuable contribution of a computational approach.

Conclusion

If any computational humanities program is to succeed, the policymakers, organisers, and implementers of such programs should take into account how formalisation is put forth and what is understood by formalisation. We have presented four cases that show the highly varied modes and realisations of formalisation in humanities research. The case of van Zundert and Andrews ('Hypothesising history') predominantly focuses on technical and cultural aspects in formalising properties of a research object. The research of van Dalen ('The onymic landscape') draws attention to the formalisation of the heuristics of a specific research domain. Zeldenrust ('Microtoponymy') demonstrates that formalisation can lead to more freedom, not less. Finally, the case of Beaulieu ('The empirical image') calls attention to emergent formalisation as a driver for the development of a computational approach, rather than the other way round, thereby tapping into the creative potential created by reversing the dominant dynamic. These different modes of formalisation are connected to, but not singularly driven by, current computational practices. Formalisation manifests itself as a multi-faceted, multi-directional and multi-motivated complex of activities, not as a simple, unitary principle underlying computational approach.

The case studies presented in this chapter also illustrate that formalisation can be supported by computation, if we recognise formalisation as an integral part of humanities practice and not as a feature driven only by computation. Such an understanding can be used to align technology and tool development efforts more usefully to the needs and ambitions of researchers. Furthermore, by recognising and articulating different modes of formalisation, computational

science can enrich its own research agenda, further expanding its ambitions in terms of what computation can mean.

Furthermore, if we identify and describe 'cultures of formalisation', researchers will have a more explicit means to recognise practices of formalisation in their own and other humanities domains. In other words, researchers will be able to 'look over the walls' and identify both implicit and explicit formalisation practices in different humanities domains. Such a recognition of different modes of formalisation would enable researchers to interact with each other's modes, allowing them to cross-fertilise different knowledge domains. From there, a community or network of researchers could develop to enhance and to foster awareness of formalisation practices as a value-added means for the humanities.

Finally, in examining promises and challenges of computational methods in general and formalisation in particular, two points should be taken into account. First of all, the ongoing computational 'waves' and 'turns' should not steer the research community away from maintaining and promoting the traditions of humanities in contemporary scholarship. Computational humanities should be unequivocally recognised as only one stream of contemporary humanities research. Perpetuating claims about potency and ubiquity of computational methods, while regarding non-computational scholarship as conservative, creates resistance towards methodological and epistemological innovation. Such claims also obscure the fact that not all questions in humanities research can and should be approached by way of some unified computational analysis. The variety of cultures of formalisation illustrated in this chapter highlights that there is no single golden road to computation.

Secondly, an interplay among computation, formalisation, and humanities should not be light-heartedly considered as yet another way of doing humanities research. Such an interplay is rather more about cognition than about method (cf. Brey 2005): 'when the computer functions as an enhancement of human cognition ... human and computer are best regarded as a single cognitive unit, a hybrid cognitive system that is part human, part artificial, in which two semi-autonomous information-processing systems cooperate in performing cognitive tasks' (Brey 2005: 392). Understanding the cognitive interplay of computational systems and human users is important for analysis of formalisation in humanities research. The 'computational turn' does not involve 'just' a specific formalisation of research hermeneutics; it possibly also involves a specific formalisation of the research thought process. However, the community of computational humanists seems to shy away from such a view, and instead seems to specifically highlight methodological and epistemological aspects of formalisation. Yet recognising the cognitive and affective aspects of scholarship (Antonijevic, Dormans, and Wyatt n.d.) could help us understand some of the reasons for the resistance towards computational methods that

still prevails. Such an understanding would also help acknowledge scholars' right to not compute, and to decide which turn to take.

Epilogues: connecting cultures of formalisation

The main conclusion of our case studies is that any computational humanities program that is to succeed should take into account how formalisation is put forth, and what is understood by formalisation in the various subfields of the humanities. The cases studies demonstrated wide epistemic variety in humanities research, and, subsequently, the co-existence of diverse formalisation practices in the humanities disciplines. In this final section, we document how we have endeavoured to put the insights presented in the chapter into practice.

The case studies presented in this chapter are all linked to the Alfalab project (http://alfalablog.knaw.nl/), and this is not coincidental. One of the objectives of Alfalab has been to foster interdisciplinary collaboration, foremost by developing and applying computational means suited to research interests of the research settings involved. However, fostering interdisciplinary collaboration might be a goal more easily set than achieved. An implicit consequence of such interdisciplinary efforts is that various epistemic cultures will meet and will need to interact within a project. This has been the case in Alfalab. There we have become aware of methodological and epistemological differences among researchers involved in the project, and of the need for explicit attention for such differences and for flexibility in shaping the project. As a result, we adapted the implementation of Alfalab. It was initially envisioned as a generic Web-based humanities laboratory – a computational infrastructure project, within which a variety of generic tools and databases would be available (Van Zundert 2009). In this sense, the initial plans for Alfalab could be compared to those underlying other humanities-oriented infrastructural initiatives, such as Project Bamboo (Bamboo 2011), DARIAH (Dariah 2011), and CLARIN (Clarin 2011). These projects primarily aim for shared digital infrastructures, as an explicit pre-condition for the successful application of computational approaches in the humanities. However, the research partners in Alfalab realised that developing a shared infrastructure could not, by itself, bring about interdisciplinary or collaborative research. Alfalab needed a way to connect more closely the common digital infrastructure and particular research questions posed by involved research communities. In order to achieve this, we adopted the concept of tailored virtual research environments (VREs).

In Alfalab, a virtual research environment came to be defined as a set of tools tailored to facilitate a specific research workflow over a distributed digital infrastructure that transcends institutional borders (Zeldenrust 2010). In current digital humanities practice, a VRE typically combines three types of digital

tools. The first type comprises digital communication facilities such as mailing lists, wikis, RSS feeds, video, and chat functions. The second type, used alongside the first, enables researchers to access and/or create data within a VRE. Third, a VRE will likely offer tools to analyse digital research data, and possibly to visualise the results of analyses (Early Modern 2011).

Although some of the VREs are quite generic, others opt to adapt to the research of the research communities involved. Such specific tailoring of a virtual research environment usually derives from the recognition that different epistemic cultures apply different methodologies, heuristics, and epistemological approaches. In Alfalab, we found that offering the same, generic infrastructure to all of the researchers included in the project did not facilitate successful integration and implementation of computational methods. Again, infrastructure was a prerequisite, but in itself it could not sustain a vibrant research community focused on using digital resources to engage with a certain research topic. Such engagement could only result from digital resources that served the specific purpose of the researchers involved, and that corresponded to the epistemological and methodological approaches of those researchers.

For that reason, Alfalab has chosen to opt for more differentiated, rather than generic, computational approaches. We have created three demonstrator virtual research environments targeted at three specific research approaches pursued by networks of researchers. The resulting three VREs are TextLab, GISLab, and LifeLab. TextLab is built around text-oriented research. It comprises a text transcription tool to capture digitally the text of physical documents that cannot be digitised by automated means such as optical character recognition. The tool has elaborate facilities to structure, comment, index, and publish digital editions of text sources. GISLab is a VRE geared towards curation and analysis of geospatial data. It involves, among other features, a Web-based facility to pinpoint locations in old maps onto a modern geospatial grid, hence relating historical geospatial information to current reference points. LifeLab opens up census and life course data from a variety of databases to the use of, for example, scholars interested in historical life course analysis and historical economics. Currently, LifeLab does this by making it possible to select datasets for specific variables out of the whole collection. These specific datasets can then be analysed, for example, with statistical means.

Around all three demonstrator projects, Alfalab has organised workshops and is providing various forms of documentation. These serve to familiarise targeted researchers with tasks and workflows that can be achieved by combining Web-based and other digital tools. Workshops also serve to tailor the tools and workflows more closely to the needs of researchers involved.

Last but not least, an expertise group called InterfaceLab has been developed in the course of Alfalab. The work of the InterfaceLab has been to ensure that, as much as possible, VREs were developed through effective

collaboration between humanities researchers, computer scientists, and science and technology studies experts. Collaboration across these epistemic cultures did not occur spontaneously or effortlessly. The InterfaceLab has developed and deployed a variety of strategies to stimulate and enhance this collaborative work among researchers in Alfalab and to support the interactions with the digital humanities community and potential users. Among these are the development of shared understandings of research agendas, of data/tool coupling, and of researchers' needs when working in digital settings. These are detailed on the Alfalab portal, and in various publications – indeed, the analysis in this chapter is an illustration of the kind of work stimulated and supported by this part of the project. Through these sustained interactions in Alfalab, mutual learning and sharing of experiences has taken place. The success of Alfalab can therefore be traced to the expertise of participants in their respective field, to the feedback cycles implemented by the InterfaceLab, and to the ongoing stimulation of critical reflection on the processes of translations going on within the project. As such, the outcomes of Alfalab are equally the mechanisms for interaction set up in and around the project as well as the specific VREs developed.

Alfalab has endeavoured to develop digital resources that not only acknowledged specific research needs, but also recognised different epistemic cultures and types of formalisation applied by those research communities. The project thus incorporated lessons learned from the case studies presented in this chapter, and was able to put forward digital resources useful for specific research groups. However, this did not mean that the goal of fostering interdisciplinary collaboration was also realised. Acknowledging the variety of methods and epistemologies involved in Alfalab helped us to recognise that more generic computational approaches would not suit our purposes. But did it also enable us to identify successful properties for interdisciplinary digital collaboration?

Comparing different heuristics and epistemic approaches among partners involved in the Alfalab project provided an insight that it would be very hard to come up with an easy solution, or natural 'fit', for facilitating interdisciplinary collaboration within a VRE. But reflecting upon those heuristics explicitly and in a comparative manner also revealed certain commonalities in concepts and workflows. It became obvious that, if the project was to offer a successful route to interdisciplinary digital resources and collaboration, we should target those commonalities and try to create shared digital functionalities on the basis of these.

One such commonality appears to be the concept of annotation. Researchers involved across a range of different disciplines represented in the project seem to value annotation. The specific forms of annotation may be different (recall the tension between textual and visual material in the Flickr case study above) but the basic concept is not; for the scholars involved, annotation is the act

of enriching or interpreting research data. In TextLab, this might result in textual notes explicitly related to certain text fragments. In GISLab, annotation could imply identifying certain locations with registered microtoponyms. In other words, the 'act of annotation' can be regarded as a common interface among researchers involved in the project. Although formalisation in these different research settings may differ, the concept of annotation travels across epistemic communities. From the information science viewpoint, such concepts can be useful handles for leveraging computational approaches. For Alfalab annotation was one of the axes along which its interdisciplinary digital infrastructure could be built. We have therefore cross-referenced annotations across the three demonstrator virtual research environments. For this purpose a fourth demonstrator was conceptualised as a cross-demonstrator annotation discovery tool. The implementation of this demonstrator will consist of a repository that harvests annotations from the 'back ends' of the other three demonstrators. Annotations will then be categorised, and a user interface for exploration of these annotations will be provided. The implementation is modeled after the principles of the Open Annotation Collaboration (OAC 2011) and Linked Data initiatives (Linked Data 2011). Currently, this annotation exploration demonstrator is a centralised and snapshot-based tool for data discovery. Ideally this will become a real-time active annotation alerting system. In future developments, we hope that, for instance, a researcher who annotates 'Amsterdam' in a text in TextLab might be alerted that GISLab contains annotations on 'Amsterdam' from the same period as the text's publication year.

The current results achieved in Alfalab show how recognition and careful consideration of differences in cultures of formalisation can facilitate development of useful and applicable digital resources for humanities research. Analysis and reflection on such cultures pre-empts the risk of naive strategies for pushing technology to unreceptive researchers. Such reflections can also lead to the identification of conceptual commonalities across different sets of research heuristics that appear to be good leveraging points for computational approaches.

Although this is a promising beginning, Alfalab is a time-limited project, and it has explored the variety of formalisation practices in humanities research only within this limited project scope. In order to network virtual research environments that support various heuristics and formalisation practices in humanities research successfully, we need far more explicit knowledge of epistemic cultures in the humanities (Wouters et al. n.d.). As the Alfalab example shows, the active coupling of this knowledge to the development of new research practices and new tools would provide the greatest benefits for projects and initiatives developing digital infrastructures for humanities research.

Notes

1. Several institutes of the Royal Netherlands Academy of Arts and Science (KNAW) have joined forces in a cross-institute project named Alfalab (see http://alfalablog.knaw.nl/). All authors of this chapter are members of the Alfalab team. The KNAW has also set up a committee to develop a program of research on Computational Humanities; Beaulieu and van Zundert are members of the program committee.
2. The material presented here is part of an ongoing ethnographic project, Network Realism, pursued at the Virtual Knowledge Studio for the Humanities and Social Sciences, Amsterdam by Sarah de Rijcke and Anne Beaulieu. See the project blog: http://networkrealism.wordpress.com/.

References

Antonijevic, S., Dormans, S., and Wyatt, S. (Under Review), 'Working in Virtual Knowledge: Affective Labor in Scholarly Collaboration', in P. Wouters et al. (eds), *Virtual Knowledge* (Cambridge, MA: MIT Press).

Arzberger, P., et al., (2004), 'Science and Government: An International Framework to Promote Access to Data', *Science* 303(5665): 1777–8.

Auzépy M., (2007), 'State of Emergency (700–850)', in J. Shepard (ed.), *The Cambridge History of the Byzantine Empire c.500–1492* (Cambridge: Cambridge University Press, 251–91).

Bamboo (2011), http://projectbamboo.org/. Date accessed 7 February 2011.

ABeaulieu, A. (2003), 'Annex 1, Case Study of Data Sharing at the fMRI Data Center, Dartmouth College, USA', in *Promoting Access to Public Research Data for Scientific, Economic, and Social Development, Final Report,* OECD Follow Up Group on Issues of Access to Publicly Funded Research Data.

Beaulieu, A., van Heur, B., and de Rijcke, S. (Under Review), 'Authority and Expertise in New Sites of Knowledge Production', in P. Wouters, et al. (eds) *Virtual Knowledge* (Campbridge, MA: MIT Press).

Bock, B., et al. (2009), *Automatic Extraction of Topic Maps based Argumentation Trails* (Leipzig: Topic Maps Lab). http://www.topicmapslab.de/publications/automatic_extraction_of_topic_maps_based_argumentation_trails?locale=en. Date accessed 7 February 2011.

Brey P. (2005), 'The Epistemology and Ontology of Human-Computer Interaction', in *Minds and Machines* 15(3–4) (Hingam: Kluwer Academic Publishers, 383–98).

Clarin (2011), http://www.clarin.eu. Date accessed 7 February 2011.

Dariah (2011), http://www.dariah.eu. Date accessed 7 February 2011.

Dédéyan, G. (2003), *Les Arméniens entre Grecs, Musulmans et Croisés: étude sur les pouvoirs arméniens dans le Proche-Orient méditerranéen (1068–1150)* (Lisbon: Fundação Calouste Gulbenkian).

Early Modern (2011), http://www.earlymoderntexts.org/VRE_What/index.html. Date accessed 7 February 2011.

Fuller, M. (2003), *Behind the Blip: Essays on the Culture of Software* (Brooklyn, NY: Autonomedia).

Honing, H. (2006), 'On the Growing Role of Observation, Formalisation and Experimental Method in Musicology', in *Empirical Musicology Review* 1(1), 2–6.

Knorr-Cetina, K. (1999) *Epistemic Cultures: How the Sciences Make Knowledge* (Cambridge, MA: Harvard University Press).

Linked Data (2011), http://linkeddata.org/. Date accessed 7 February 2011.

Mann, W. C., and Thompson, S. A. (1986), *Rhetorical Structure Theory: Description and Construction of Text Structures* (Nijmegen, NL: Information Sciences Institute).

Petersen S. M., (2009), *Common Banality: The Affective Character of Photo Sharing, Everyday Life and Produsage Cultures*, PhD thesis (Copenhagen: IT University of Copenhagen).

Schönfeld, M. (1950), *Veldnamen in Nederland* (Amsterdam: Noord Hollandsche Uitg. Mij.).

Sekine, S., and Ranchhod, E. (2009), *Named Entities: Recognition, Classification and Use* (Amsterdam/Philadelphia: John Benjamins).

Traweek, S. (1988), *Beamtimes and Lifetimes: The World of High Energy Physics* (London: Harvard University Press).

van Dalen-Oskam, K. H. (2005), 'Vergleichende literarische Onomastik', in A. Brendler and S. Brendler (eds), *Namenforschung morgen: Ideen, Perspektiven, Visionen* (Hamburg: Baar, 183–91). English translation: K. H. van Dalen-Oskam (2011), 'Comparative Literary Onomastics'. http://www.huygensinstituut.knaw.nl/vandalen. Date accessed 7 February 2011.

van Langendonck, W. (2007), 'Theory and Typology of Proper Names', W. Bisang (ed.), *Trends in Linguistics: Studies and Monographs* 168 (Berlin/New York: Mouton de Gruyter).

oac (2011), http://www.openannotation.org. Date accessed 7 February 2011.

Wouters, P., and Beaulieu, A. (2006), 'Imagining e-Science beyond Computation', in C. Hine (ed.), *New Infrastructures for Knowledge Production: Understanding e-Science* (London: Idea Group, 48–70).

Wouters, P., et al. (eds) (Under Review), *Virtual Knowledge* (Cambridge, MA: MIT Press).

Zeldenrust, D. A. (2005), 'DIMITO: Digitization of Rural Microtoponyms at the Meertens Instituut', in *Humanities, Computers and Cultural Heritage* (Amsterdam: Royal Netherlands Academy of Arts and Sciences, 301–7).

—— (2010), 'Spheres of Alfalab, Virtual Research Environments between Research Questions and Infrastructure' (Vienna: SDH 2010).

van Zundert, J. J., Zeldenrust, D. A., and Beaulieu, A. (2009), 'Alfalab: Construction and Deconstruction of a Digital Humanities Experiment', in *e-Science 2009, Fifth IEEE International Conference on e-Science* (Oxford, 1–5).