Computational Humanities

Computational Humanities Programme Committee, Royal Netherlands Academy of Arts and Science (KNAW)

Computational Humanities Programme Committee

Computational Humanities

Computational Humanities Programme Committee
Royal Netherlands Academy of Arts and Sciences (KNAW)

2010

1. Introduction

In 2025, the field of humanities finds itself in a strong and integrated position among the sciences. Scholars in 2025 looking back 15 years, see a less integrated set of academic disciplines with substantial differences. They also see that by 2010, a collective of unorthodox perspectives on the humanities and strategic initiatives generated a step change in the study of language, music, art, history, religion, and archaeology. The significant breakthrough, that happened both nationally and internationally, was a result of the effective integration of information science and information technology in the humanities. In 2025 traditional research questions are approached differently, and new questions that 15 years ago were beyond imagination are now addressed effectively. During the past 15 years humanities not only benefited from information science and technology but made significant contributions to these fields because of what in recent years became known as the humanities perspective on information. Information is more than bits and patterns. Equally important are the interpretation of information and the meaning and aesthetic value attached by the observer. The proposed Computational Humanities (CH) Programme is a roadmap to this vision of the humanities in 2025.

Today, the humanities make effective use of information technology for data storage and access, and to derive information from data. Observational studies, experiments, imaging and digitalisation are generating vast volumes of data. Data come in various forms: numerical data, images, language (spoken and written), etc. Data represent the real world with different degrees of detail and accuracy. They relate to objects in the real world and relations between objects. They show objects or characterise objects. Data are abundantly available and increasingly accessible. Because of the internet, access is possible anytime and anywhere, even if the data are held at disparate sites. The result is connectivity and interactivity, i.e. the ability to be connected and interact with almost anyone. Virtual worlds (virtual realities) are emerging that are complementary to the physical world. Vast amounts of data are of limited value if data mining technologies are not available or access is restricted, or if the knowledge infrastructure does not yet exist to create new knowledge from data. It takes teams of skilled personnel to compare data, to detect patterns, to describe and comprehend the processes that are revealed in the data, to capture in a model the essential features of a pattern or process and to separate it from secondary features and noise. Skills are also needed to explore and predict how patterns and processes are experienced by target populations and the population at large. Advanced computing technologies are necessary to solve complex computational problems. They are not sufficient, however.

Computational Science is the use of advanced computing capabilities to understand and solve complex problems. It constitutes the “third pillar” of scientific inquiry, together with theory and observation. It challenges researchers to conceptualise problems differently (to think out-of-the-box), to redefine objects and relations between objects, to build models, test hypotheses and simulate the impact of history and context on emergent processes and structures. Computational Humanities is the application of computational science to core issues in the humanities. The humanities differ from the sciences in its concern with expressions of the human mind and their aesthetic value. The expressions include language, literature, music, art, history, religion, and archaeology[1]. Computational Humanities is a new, largely unexplored, field of knowledge. It is situated at the interface between humanities and the (exact) sciences, in particular information science. Developments at that interface are characteristic for modern humanities and are considered important by the Committee National Plan Humanities to the migration of the humanities to a new, sustainable future. Large-scale digitisation and advanced computing capabilities contribute to the sustainable development of the humanities. Today, increasingly, scientific breakthroughs in the sciences are accomplished by teams that have access to massive datasets and are able to process these data intelligently. Data-intensive science consists of three basic activities: capture (e.g. digitalisation), curation[2], and analysis. Discoveries based on data-intensive computing are being referred to as the fourth paradigm for science (Hey et al., 2009; Bell et al., 2009)[3]. In this paradigm, new theories will be developed by combining observational data with data generated by computer models. Computational Humanities is related to several initiatives in the humanities aimed at capturing and curation of data: the development of infrastructures and tool and service provision. Annex 1 briefly describes significant current and related initiatives. Computational Humanities uses the available infrastructure to achieve a change in the humanities by effectively integrating information science and information technology into the humanities.

The Royal Netherlands Academy of Arts and Sciences (KNAW) views the new field of Computational Humanities as a means for enhancing creativity and innovation in the humanities in general and in the humanities institutes of the organization in particular. The KNAW institutes are characterised by extensive and unique collections. The large-scale digitisations of these collections provide opportunities for unorthodox and innovative research aimed at new knowledge about history, literature and music. In April 2009 the KNAW established an advisory committee to identify key areas in the humanities that are likely to fundamentally change the method of research as a result of the information technology revolution. The Computational Humanities Programme Committee (CHPC) consists of scholars from KNAW institutes and Dutch universities[4] [5]. In addition, the KNAW organized two workshops. The first workshop *Bridging the gap between the Humanities and the Computational Sciences* was held at the Trippenhuis in Amsterdam on 4th June 2009. The second workshop brought together scholars from KNAW institutes and universities to discuss the committee's draft report and to identify significant research questions in the humanities that require computational methods to be addressed effectively. The workshop was organised at the Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS) in Wassenaar from 11 to 13 February, 2010.

This report is the outcome of reflections and consultations. Section 2 describes how Computational Science may enhance the study of expressions of the human mind in contemporary societies and history. The section is not about method but about substance. What fundamental questions of scientific inquiry can be addressed given the advanced information technology (including information science) that cannot be addressed otherwise? The questions relate to how expressions emerge, evolve and acquire shared meaning and

---

[1] Commissie Nationaal Plan Geesteswetenschappen (2008) Duurzame geesteswetenschappen. Rapport van de commissie Nationaal Plan Toekomst Geesteswetenschappen (Committee National Plan Humanities, Sustainable humanities. Report of the Committee National Plan Humanities). Amsterdam University Press, Amsterdam. www.aup.nl/geesteswetenschappen. The committee analyses the problems of the humanities in the Netherlands and presents a plan to 'help migrate the humanities to a new, sustainable future.'

[2] Curation of data is the management and long-term care of data. It starts with finding the right data structures and the necessary metadata.

[3] The other three approaches to generating new knowledge are: theoretical research, empirical research (experimental and observational studies) and computer simulation.

[4] For the composition of the CHPC, see Annex 3.

[5] The committee met in 2009 on 14th May , 17th June, 4th September, 25th September and 23rd November.

aesthetic value. Sections 3 to 5 describe particularly pressing issues in the humanities that may be dealt with effectively by Computational Humanities. They are: formalisation (Section 3), representation of meaning and the Frame Problem (Section 4) and types and sources of data in the humanities (Section 5). Section 6 presents three areas of research that are at the forefront of the humanities and require extensive computing capabilities. They are therefore most likely to benefit from information science and information technology. Section 7 proposes a KNAW programme in Computational Humanities. The programme was discussed extensively at the NIAS workshop and the specific projects emerged from that workshop and consultations in and among institutes of the KNAW and university scholars.

2. Computational Humanities

Computational Humanities is the application of Computational Science to core issues in the humanities. It includes pattern recognition, sequence analysis in text and historical data, modelling and simulation, development of algorithms, and the presentation of the results in images (visualisation; animation) and sound. It also includes innovative ways of data acquisition, validation, storage, documentation (annotation), processing and dissemination. Computational Science is often associated with vast amounts of data (observations), although that is not a requirement. Computational Science is also used to develop theories from limited data by augmenting the data with information generated by simulation. Computational Humanities has much in common with conventional scientific endeavours. It combines theories, data and methods to address significant research questions. It describes objects and processes and searches for mechanisms (underlying processes or rules) that produce phenomena we are observing in the real world. The mechanisms are often situated at a more abstract level than the details of the real-world observations. For instance, we may explain social movements in terms of the behaviour of actors, interactions between actors, and interactions with the context (time and space). Languages may be characterised in terms of elementary building blocks and rules that govern the combination of the building blocks. Many of the methods in the humanities are generic but adapted to particular research questions and data.

Examples of overarching research questions in the humanities are:

a. How do we express ourselves in language, music and art? How do we convey facts, meaning and emotion and how do systems of shared meaning and mutual understanding develop? How do these expressions contribute to the development of an identity? What factors influence the expressions taking shape and evolving?
b. Do expressions exhibit a structure (patterns)? For instance, does language exhibit a hierarchical structure (tree structure)? What processes generate the structures (how and why do these patterns emerge)? For patterns to emerge, interaction and diffusion mechanisms must be at work. What are these mechanisms? Do similar structures emerge in different contexts (cultures; episodes in history)?
c. What is the probability that a given model describes the structure exhibited by the data? What model maximises that probability (likelihood)? What is the simplest model that gives a plausible (sufficiently accurate) description of the structure exhibited by the data? Measures of simplicity and likelihood guide modelling.
d. What is the relation between a model that describes a structure or pattern and a model that describes a process that generates a structure or pattern?
e. What is the relation between the physical form and other measurable characteristics of a message and its meaning? How does meaning evolve in time and space? How do we encode and decode messages and associate meaning?
f. What factors (events, structure, expressions) in prose, poetry, music or art cause emotional reactions (experiences)?
g. What factors trigger historical events? What is the role of (1) previous events and event sequences, (2) the presence of persons with particular characteristics and

position in society, and (3) the historical context (period)? What insights do we obtain from counterfactual analyses?

h. What do we need to know for accurate predictions of events, conditions and locations?
i. What can Computational Science contribute to the studia humaniora?
j. What is the role of surprising accidental discoveries (*serendipity*) in the identification and explanation of pattern? In what way can a model (or formalised theory) make surprising predictions, taking into account the surprise following unexpected empirical findings, as well as the surprise stemming from unforeseen empirical consequences of the models?

In abstract terms most of these questions also arise in the natural sciences, life sciences and social sciences. They relate to information processing: coding of information (message, meaning, experience), transmission of information, decoding, storage, and retrieval. They also relate to mechanisms of change: events that trigger other events or cause a system to be 'locked in' in a particular state (*absorbing state*), actors who interact and exchange information (and meaning) causing the information to spread in a group (diffusion process). Contextual factors may influence the event sequence and/or the diffusion process, hence acting as a coordination mechanism[6]. To detect the relevant intrinsic and contextual factors is a major challenge in science and humanities. The common approach is to specify a model of the data and to estimate the parameters by maximising the likelihood that the model describes the data (maximum likelihood method) or by minimising the difference between the model data (data generated by the model) and the empirical data (observations) (least square method). Alternatively, novel statistical methods allow for the comparison and evaluation of theories without extensive, or else with only partial empirical support (such as in archaeology or history). In Computational Science, data are viewed as manifestations of an underlying process or a number of plausible underlying processes. Models are specified that describe these processes. They are accurate if they reproduce the empirical observations on (historical) events and structures. Models that describe processes and interactions between processes have several advantages over models that describe structures or patterns. They can integrate path dependencies, feedback effects, heterogeneity and selection, and other factors that influence the emergence of patterns. They can also relatively easily incorporate random mechanisms that capture the effects of chance in pattern formation. Models that have these features are commonly referred to as microsimulation models and agent-based models.

Computational thinking in the humanities (and other sciences) is challenging the humanities (and other sciences) by reformulating old problems and formulating new problems that address core areas and by proposing new ways to solve these problems by perspectives and new methods including reduction, embedding, transformation, and simulation.

3. Formalisation

Formalisation is a description of knowledge about objects and processes in a formal language and using a preconceived framework with the intent to communicate and evaluate. Formalisation is an important dynamic in any kind of knowledge production. It is at the heart of key research activities, because it enables social and technological interaction, as well as standardisation. It is a necessary condition for hypothesis testing and falsification by confronting theory with data (Suppes, 1968). It is also a powerful epistemic strategy that enables processes as diverse as the development of systems of classification, ordering of observations, or the identification of genres.

When the issue of formalisation comes up in relation to the humanities, it is used to refer to

- a necessary step for the use of digital technology;

[6] The peer review coordinating process acts as a filter in a knowledge diffusion process.

- a feature or quality required of data;
- a procedure that is novel for the humanities.

This programme on Computational Humanities proposes to pay due attention to the issue of formalisation as a rich notion that goes beyond the three assumptions listed above. First of all, digital tools do not all require the same type or degree of formalisation. A heuristic visualisation requires a very different kind of formalisation than a statistical programme based on number-crunching algorithms. Similarly, formalisation can take on different shapes, and need not be quantitative. Annotations, such as those used in Alfalab tools (see Annex 2), or textual labels, such as those used in the semantic web, are examples of formalisations that are not intrinsically numerical, yet are eminently amenable to the productive use of digital technologies.

Second, formalisation does not only occur in relation to data. Communication, collaboration or publications are also aspects of research that can be increasingly formalised, when digital tools are introduced in research practices. Consider the following examples: (1) writing in digital media implies different kinds of formalisations than writing with a typewriter; (2) using a database that keeps track of changes in texts, 'contributions' can be measured within a collaboration; (3) web-based interactions with colleagues require formal protocols; (4) social networking software can make relations visible or suggest new ones; (5) alerting services of journals, based on formalised aspects of articles, change our awareness of relevant literature.

Third, formalisation is far from novel for the humanities. There are specific styles of formalisation that are particular to the humanities, and some that overlap with formalisations familiar to the natural and social sciences, such as those developed in the field of logic. As such, some of the formalisations that will arise in the course of the development of new research practices around digital tools will be unique to the humanities, while others will be similar to those in other areas of research. Furthermore, the introduction of digital tools can make visible the contrast between kinds of formalisations that have a long tradition in humanities and those expected by digital research infrastructures. For example, 'what is a text?' is a question that has long occupied scholars of literature, and a question that is raised again, in both fundamental and urgent ways, in the course of developing digital editions. In this light, formalisation becomes the opportunity to reflect on the way knowledge is produced in different contexts, and to understand more fully the relation between new tools and new forms of knowledge.

Fourth, formalisation makes knowledge explicit. That is important since in an interpretation often a lot of knowledge is 'hidden'. For instance, some aspects of knowledge stay invisible for those who do not know how a particular phenomenon was interpreted. That is why also an interpretation should be made precise and explicit. In that way other scholars and scientists are in a position to trace the conclusions or formulate alternatives leading up to the same results. During the process of formalisation one sometimes becomes aware of the fact that a certain notion is not yet clear enough because an explicit and testable description is simply absent.

To illustrate the concept of formalisation, consider the following example from music theory. It derives from the work of musicologist Fred Lerdahl and linguist Ray Jackendoff (Lerdahl & Jackendoff, 1983). They model music understanding from the perspective of cognitive science. The point of departure is a search for a grammar of music with the aid of generative linguistics[7]. While the impact of their work was only marginal among music theorists and musicologists, the impact in neighbouring fields, such as psychology and computer science, was substantial. The latter appreciated a theory that provided (at least a beginning of) a formalisation of general musicological knowledge, allowing for empirical testable hypotheses. Overall the theory is an illustrative example of the role of formalisation in the

[7] Generative linguistics is a school of thought within linguistics in which the grammar is made fully explicit. A finite set of rules are used to generate those and only those sentences that are grammatical in a given language (Chomsky, 1965).

communication of knowledge within and outside the humanities. The formalisation of the knowledge contained in a certain research domain allows for it to be used more easily and readily *outside* of the discipline in which it was developed. For musicology it might not be so relevant as to what the notions 'meter' or 'tempo' look like in formalised form; they are just axiom's in the study of topics that are, in themselves, far more interesting. For researchers of the neighbouring disciplines, however, that is not necessarily the case. For them these fundamental, music theoretical notions are of great importance.

The programme on Computational Humanities will foster attention to the diversity of forms of formalisation and their relation to the use of digital tools. It will welcome projects that explicitly reflect on formalisation. This close attention to formalisations that are particular to the humanities will enable the development of tools that are more closely aligned to research practices in these fields, and lead to a better understanding the knowledge produced in the humanities.

Two broad types of formalisation of knowledge are commonly distinguished: declarative formalisation and procedural formalisation. The first describes how things are. What are the attributes, events, processes, etc. and how are they related? A mathematical model belongs to this type. The second describes how things work or operate. What are the procedures or rules that guide a process?

4. Representation of meaning and the Frame Problem

Enumerating the phenomena of the world in a formal description has a long history from Parmenides, Plato and Aristotle, to Kant, Husserl and the contemporary advocates of the Semantic Web which promise a heaven of knowledge access *'if only the world has been modelled properly'*. However, there are fundamental stumbling blocks on the path of formalisation of objects and processes. The first is the representation of meaning and the process of meaning-giving (construction of meaning). In computer-based information processing, this issue remains unresolved. The second is the Frame Problem. It is the problem of determining what in the context is relevant to solve a specific problem. The knowledge of what is relevant implies the capacity to anticipate the consequences of an event or action. The consequences can be many and it is unrealistic to determine all the changes that may occur as a result of the event or action.

a. Representation of meaning

The advent of computers heralded new realms in symbolic information processing. Already early in their development, it became apparent that computers were able to execute more interesting forms of processing, beyond numerical computations such as the calculation of ballistic trajectories or the minimum energy needed for initiating nuclear fusion in the atomic bomb. From decryption challenges in the second world war, it was a natural step towards symbolic-information processing using computers. Apart from numerical programming languages such as Fortran (Formula Translation), new languages such as Lisp (List Processing) appeared for the manipulation of symbolic information. The allure of the new methods was that symbolic phenomena such as language and semantics, formerly reserved for human reasoning, might be emulated and implemented, with the hope of constructing an intelligent 'electronic brain', ultimately. However, fifty years later, we may conclude that most of the ambitions that were prevalent during the mid-twentieth century have not yet been realised. It is remarkable that a phenomenon such as language, which is already symbolic and discrete by its own nature, is such a resilient phenomenon to model and emulate using a computer. There are some successes. On the fringe between computing science and humanities, the field of Computational Linguistics has yielded impressive results. However, the focus of this discipline is largely on syntax and statistics. In language and within the humanities at large, a more central topic is that of semantics: the construction and representation of meaning. This is still an unresolved issue in computer-based information processing. Furthermore, as Hofstadter (2001) notes, good models for finding analogies and abstract pattern comparison are still missing. One may conclude that something is wrong in our current views on symbolic information processing. If it is so difficult to model language and semantics by means of the ultimate symbolic engine, i.e., the von Neumann-Turing computer, we may be on the wrong track. Valuable, new knowledge in the humanities is generated by humans, not by machines. It appears probable that for rich information processing, intelligent association between patterns and finding analogies, other mechanisms may be at work that are beyond traditional logic.

A central tool in capturing semantics by means of symbolic representation is the ontology: a formal description of objects in a concrete or abstract world, together with their relations, properties and axioms describing their behaviour:

> "Formal Ontology is the result of combining the intuitive, informal method of classical ontology with the formal, mathematical method of modern symbolic logic, and ultimately of identifying them as different aspects of one and the same science. That is, where the method of ontology is the intuitive study of the fundamental properties, modes, and aspects of being, or of entities in general, and the method of modern symbolic logic is the rigorous construction of formal, axiomatic systems, formal ontology, the result of combining these two methods, is the systematic, formal, axiomatic development of the logic of all forms of being. As such, formal ontology is a science prior to all others in which particular forms, modes, or kinds of being are studied."
>
> Cochiarella (1991)

b. The Frame Problem

The Frame Problem in Artificial Intelligence (see Korb,1998, for an extended overview) refers to the insight that an intelligent actor will by stymied to initiate any action if all consequences and ramifications of that action need to be cognised, beforehand. Traditional logic will fail, and the amended types of logic such as nonmonotonic reasoning suffer from other practical deficiencies. Additionally, such proposed logical solutions still fail to solve the Frame Problem fundamentally. Today, it is realised that the Frame Problem is an instance of the more general class of problems which are concerned with the detection or definition of relevant context to a specific problem.

> "The Frame Problem is one concerning internal representations about the world. .. [it] concerns our *descriptive knowledge* of events and the capacity (human or machine) to infer the consequences. The heart of the Frame Problem concerns where, in a model of the world, to search for consequences and when to stop. In this sense, the problem can be reduced to designing [an] optimum search strategy …"
>
> Brewster (2003)

A range of related problems such as the 'ramifications problem' and 'selective attention' in cognitive-science research can be identified. Humanities researchers are continuously faced with such problems: on the one hand, the abundance of information in libraries, whether in analogue or digital form, provides a limitless source of arguments for constructing or supporting a narrative. On the other hand, search and selection of relevant information become extremely time consuming and costly. What is more, an oversight of the (dialectic) propagation of knowledge over networks of researchers is logically and practically unmanageable, even and perhaps just in case all such knowledge is maintained neatly in an ontology. Choosing a topic, finding the relevant material and anticipating the effect of the new interpretations and narratives under construction present a daunting task.

The interdisciplinary field of Computational Humanities needs to be developed to deal effectively with the problems of semantic representation and framing. The essence of association and analogy needs to be uncovered in greater detail. Within the discipline of Artificial Intelligence it has become clear that the modelling of intelligence is futile when it is detached from physically embodied agents in a social and cultural context. Knowledge cannot be completely detached from the physical and social reality if it needs to be used intelligently. Language is a vehicle, carrying 'aboutness', i.e., reference to an external world. In order to understand phenomena in this world, a rich representation of propositions (facts), predicates, imagined images, remembered sounds and voices, spatial and geographical patterns need to be activated in the brain of a reader of text or a listener to speech. The Frame Problem is concerned with internal representations or mental models about the world. The symbolic and ontological representations as we know them today represent a fundamental problem in humanities as in other sciences.

The good news is that new disciplines in computing science, artificial intelligence, pattern recognition and machine learning are beginning to show algorithms which display capabilities of geometric-pattern association, intuitions maybe, derived from large data repositories and text corpora. Within The Netherlands, a research programme funded by the Dutch Science Foundation NWO since 2005, i.e., the programme 'Continuous Access to the Cultural Heritage (CATCH)', has shown that it is possible, today, to search in image and sound collections and provide meaningful associations between diverse sources of information, using these new methods. Interestingly, the most interesting results appear in those areas where humanities researchers, volunteers and computers work together on the enrichment and refinement of raw information carriers such as song and speech recordings, archaeological image material and handwritten books. Time is ripe for new paradigms involving the intelligent processing of large amounts of data.

5. Types and sources of data

The observation of objects of study results in data. In the humanities, four types of data may be distinguished: numerical data, text, images and sound. Data may represent one-time observations or sequences of observations (repeated measurements, detected changes in series of images, in language use over time, etc.). Data may come together with information on the data (metadata; annotations; references; keywords; textual labels), such as when (date) and where (location) the data were produced, the meaning of the data, the object producing the data, the reliability and validity of the data, the degree of belief in the data, etc. That information is needed to interpret the data. The processing of these seemingly unrelated data types shares key features (Bod, 2008).

Data are symbols that carry information and data analysis has much in common with information processing. The interpretation of symbols, i.e. the decoding of the information, requires shared meaning-giving. The latter is culturally determined (cultural meaning systems) with a strong path dependence (earlier interpretations have a strong influence on current interpretations; learning). Path dependence underlies mental models, cognitive structures or schemata. These models or schemes filter the information we use to 'construct' knowledge. They also result in the selective attention that is central to the Frame Problem. According to some cognitive theories, at the individual level, these models can evolve through processes of assimilation (fitting observations into cognitive structures) or accommodation (adjustment of cognitive structures as a result of new observations) (Piaget and Inhelder, 1968).

The chances for finding analogies and meaningful associations between diverse sources of information on objects and events, and to identify path dependencies typical for the humanities are strongly influenced by the transformation of content collections into digital libraries. Digital humanities libraries may contribute to the increased understanding of knowledge production in the humanities due to the more facetted role of metadata and annotation processing:

(i) access tools link distributed collections at a semantic level via the traditional annotation layers known as metadata; metadata fields can cover the tomb stone details that are beyond dispute (author, date of excavation, location of discovery, format conversion, etc.), but also annotations that are the result of scholarly analysis and interpretation;
(ii) enrichment of the primary content descriptions can more easily be extended with annotations from secondary users; this may involve elements reflecting the aforementioned path-dependencies, such as the introduction of error corrections, as well as changes due to the dynamics in analysis frameworks, or to the dynamics in the availability of contextual details, including as geospatial and timing information (see below); this enrichment process can be linked to concepts such as web 2.0, Delphi and crowd sourcing, dependent on the profile of the annotators involved;
(iii) human-generated metadata can be enriched with automatically extracted content features, e.g. resulting from al kinds of content mining, including text stylistics, genre classification, date detection, etc. Partly due to the probabilistic models on which machine-generated annotations are likely to be based, this type of metadata enrichment may introduce new forms of uncertainty and fuzziness in addition to the annotation layers indicating reliability and validity.

The successful exploitation of the potential offered by digital libraries will be crucial for the success of the proposed programme. The future may show that the Fourth Paradigm in the humanities was the effective exploitation of digital libraries following large digitalisation projects.

Many information processing mechanisms are not deterministic but stochastic (they involve random elements). The outcome depends on the belief in the data (degrees of uncertainty associated with the data measured by subjective probabilities), ignorance and random factors. Such a model of scientific reasoning can be translated into a Bayesian framework. The framework enables modelling data analysis, information processing and knowledge acquisition (Howson and Urbach, 1989). The approach is a mode of formalisation that can be used to translate some aspects of scientific reasoning into computational applications.

Sources and observations in the humanities are positioned in time and space. Determining the date of a source or an observation is often a complex task. Data in sources or the source itself are often not dated or only dated approximately, i.e. the precise date is not known. Dating problems arise in various areas of the humanities: dates in time series, the date of a specific event, the dating of maps or parts of maps, etc. Historical research handles dating problems in several ways, ranging from leaving the data underspecified to making sophisticated guesses with the goal of approximating the underlying exact date.

Several research questions to the dating problem can be distinguished, such as
- How do ontologies/thesauri handle time where boundaries are uncertain and overlapping? For instance, the period of the Enlightenment differs regionally, and when ontologies are connected this should not create a mismatch, but rather a fuzzy match.
- How are maps affected by the time periods in which they are constructed? The dependence of maps on the time when they are constructed and on the purpose of the maps leads to different maps for the same area. Problems can arise when different maps are merged or connected.

6. Promising areas of research

Three promising areas of research have been identified. They relate to the formalisation of knowledge; the representation of meaning and the Frame Problem; and the modelling of content and metadata.

a. Formalisation

Both descriptive formalisation using mathematical models and procedural formalisation using rules-based models (e.g. agent-based models) are needed in the humanities to enable effective use of information science and information technology. Although in the humanities different types of formalisations exist, they have several aspects in common and some exists also in other sciences, e.g. the cognitive sciences. Scientific breakthroughs reveal the significance of formalisation and the need for a formalisation of knowledge that is not hindered by frame problems that result from narrow disciplinary perspectives on what should be formalised and how formalisation should proceed. The unorthodox perspectives referred to in the introduction are by nature not constrained by any type of barrier.

b. Representation of meaning and frames

To enumerate the phenomena of the world in formal descriptions of interdependent objects and processes that make sense, two central issues in the humanities must be resolved: semantics and the Frame Problem. The formalisation and representation of meaning and meaning-giving processes (construction of meaning) on the computer as part of an information system remain a major challenge. State-of-the-art information science and information technology is not yet able to add meaning to objects and processes that humans can understand and share. Humans interpret symbols within a context. In other words, meaning is contextualised. The context is essential for a correct interpretation of symbols. Hence, the problem of semantics is interwoven with the Frame Problem. Selective attention is too often the cause of not grasping the true meaning of an object, a gesture, a symbol or a process. Since a person’s mental model of the world determines what he or she considers relevant in a given context, the solution to the Frame Problem requires the formalisation and representation of world views, i.e. to build representations in the computer of the mental models humans use when they process signals that originate in their environment.

The context is physical and social. One way of including the social context in the description and interpretation of objects and processes is crowd sourcing. In general, groups of people have more accurate information about an object or event than individuals, or even experts. With the arrival of Internet some of this information became readily available. In addition user-generated content ('Web 2.0') has become a new and rich source of information that allows for access to knowledge that has not been tractable in such detail before. Large groups of people contribute information that is a result of all contributions, and as such reveals knowledge that remained implicit when not studied such a large scale (Honing & Reips, 2008). Computational Humanities may be furthered significantly by incorporating data from crowd sourced intelligence, especially as a way to enrich data through web 2.0 social tagging and networking facilities. Another way of using social and cultural factors in the description and interpretation of objects and processes is to incorporate serious gaming as a means to gather research data. We have seen very successful applications of serious gaming in the realm of civil engineering, social studies and the military (mainly based on simulation and role playing games). Games as serious simulations offer interesting new possibilities to gather data and conduct experimental research. Lastly the research dynamics may go both ways: the internet is a growing playing ground for the cultural industry, and Computational Humanities is in a position to provide novel ways of interpreting and accessing the information implicitly and explicitly carried by these social and cultural networks.

c. *Modelling*

A model is an abstract and simplified representation of reality. Examples in the field of (computational) humanities are a language production model, a music perception model, a visual interpretation model, or an archeological excavation model. Models are especially useful when experimentation or direct access and measurement are not possible. This is often the case in the humanities where objects of study, such as archeological sites or complex systems such as the human brain, are not easily accessible.

At the core of a model is a modelling language, which can range from a mathematical language or computer language to natural language. Modelling descriptions in a formal language brings the advantage of unambiguous interpretability, facilitating implementation in computers.

Implemented models can generate simulations of the real-world object or process they model. In the humanities, models typically simulate a human process (from music performance to text production) or a humanities artefact (from damaged scrolls to architectural constructions). This allows for a model to be empirically evaluated in terms of accuracy, efficiency and simplicity. There are a large number of methodological and empirical issues related to models, ranging from assessment through human judging or gold standards, to the ontological status of models. The identification of a key number of generic models and good practices (promising applications) is likely to be a driving force guiding innovation in the humanities.

In the humanities, an important issue is the modelling of content and metadata. The myriad of metadata types potentially evolving with the emergence of digital libraries (cf. Section 5) is capturing the process of scholarly research at multiple layers and offers a complex and rich new starting point for capturing the essence of association and analogy described in Section 3. It requires renewed attention for metadata model design (e.g., how to capture fuzziness and uncertainty), annotation interoperability issues (e.g., metadata integration for content annotated with coding tools such as thesauri and ontologies), and the coding of change over time (e.g., metadata for the dynamics of temporal and geospatial details), as well as attention for metadata standards, including formatting issues. For each class of problems specific metamodels may be needed, potentially including more abstract and widely applicable generic models (e.g. models of sequences of events; systems of differential equations), as well as algorithms, modelling languages, and modelling procedures telling how to apply the modelling language to create results.

7. The Computational Humanities Programme

The aim of the programme is to stimulate research in Computational Humanities. The development of new approaches to research, combining information and computational science on the one hand and humanities on the other is expected to lead to important innovations that could result in a breakthrough in the humanities. The formalisation of objects, events, contexts and procedures, and the advances in modelling and simulation, are expected to open up interesting avenues for research. To guarantee success, the Computational Humanities Programme focuses on a single overarching issue, creating an intersection between the humanities and computational sciences and integrating issues of formalisation, representation of meaning, framing and modelling. The proposed programme builds on the uniqueness and strengths of the KNAW institutes that result from their dual mission of maintaining collections and research. The overarching issue is: ***Ties that bind. Discovering high-level concepts, patterns and motifs across media*** (numerical data, text, images and sound).

The committee and the participants at the NIAS workshop believe that computational methods may significantly enhance the search for high-level concepts, patterns and motifs in humanities data and, as a result, enable researchers to leverage sources (whether history, language and music) differently in the course of research. From a computational point of view, the recognition of low-level patterns (e.g. shapes, linguistic dependencies, beat in music, brush strokes) is well-established. Through the work to be pursued in the programme, there is an opportunity to pursue substantial innovation in the discovery of high-level patterns, such as global structures of texts and music, emotions, semantics, and social, economic, demographic and geographical patterns and structures in history. High-level concepts, patterns and motifs can contribute to the construction and explanation of meaning.

The overarching theme is made specific in a limited number of research projects. The projects are carried out by teams of scholars from KNAW institutes and universities in the Netherlands[8]. Each project addresses a specific research question. The question will be addressed through methods that seek to identify concepts, patterns and motifs across media. At the NIAS meeting scholars from the KNAW institutes presented research questions that are central to their institute's dual mission of maintaining collections and research. The following research questions are among the questions to be addressed:

- How to identify and understand expressions of nationalism in diaries, folksongs, posters and other expressions of political activism?
- How to identify and interpret continuities and discontinuities in expressions of regional identity?
- How to identify patterns of political activism and passivism in 2nd world-war diaries?
- How to identify feelings of belonging and other expressions of emotion in sailors' letters and songs?

These questions were proposed by scholars from the KNAW institutes, and are indicative of the type of questions that the programme wants to address. Additional questions will be proposed before 15th March 2010. The questions are being elaborated in cooperation with scholars from outside the organization. A template has been prepared to guide this activity (Annex 4).

Projects will be carried out by teams of scholars from a number of KNAW institutes and universities. The projects are designed to make optimal usage of the collections maintained by the KNAW institutes and other accessible data sets and digital libraries (e.g. Europeana), and the existing infrastructure (e.g. infrastructure and expertise produced by the Alfalab project, CLARIN, DANS, etc.). Conventional methods of detection and understanding of high-level concepts, patterns and motifs will be augmented by computational methods, such as automatic data-enrichment and metadata techniques[9], visualisation techniques, high-level concept detection algorithms for document retrieval, machine-learning algorithms, models for the study of events and event sequences[10], spatial analysis using geographic information systems, and the expectation-maximization (EM) algorithm[11]. The latter is a widely used generic method for inferring missing data and latent characteristics of objects. In addition,

---

[8] The partnerships may evolve into networks of a more permanent nature, provided they generate synergy and are successful.
[9] A metadata register is being developed as part of the CLARIN project.
[10] Including Markov models that are important in the study of life histories and population dynamics in historical and contemporary societies, and Hidden Markov Models that are important in text categorization, document classification and speech recognition technology.
[11] The "shape from shading" technique used by David Stork to estimate the direction of illumination in Vermeer's *Girl with a pearl earring* is an example of the EM algorithm (see http://www.diatrope.com/stork/VermeerGirl.html). Professor Stork presented the research findings at the NIAS workshop. A few years ago it was also established that the gravity model, presented by John Nerbonne at the NIAS meeting to describe linguistic distance as a function of geographical distance, can be written as a variant of the EM algorithm. It shows that the method that involves an iterative procedure involving using a model to predict an outcome (expectation) and estimating the parameters of the model using the maximum likelihood method (maximisation) is truly generic.

experiments in collaborative annotation, user-generated content and other forms that can be labelled ‘Web 2.0 approaches’ will be supported.

To guarantee the success of the Computational Humanities programme, a number of conditions should be satisfied:

- A plan of action should be elaborated to assure that creative ideas and innovation are implemented. Driving innovation on the one hand and consolidation and anchoring of new knowledge and operational methods within established settings on the other, are two different things that need to be very carefully balanced. Sustaining innovation requires support among leaders and followers. Creating support is paramount to the success of the programme. Excellence is a necessary condition for support but it is not sufficient. An appropriate involvement of stakeholders and supportive allies are significant as well.
- The work programme needs to be peer-reviewed. Although a relatively large number of scholars are involved in the design of the programme, the committee recommends international peer review because it is the established mechanism of quality control by independent experts.
- A steering committee should be established. The steering committee supervises the overall programme, arranges evaluations and other quality control measures, assures that the programme makes optimal usage of collections and infrastructures in the Netherlands and in Europe, and proposes mechanisms that enhance effective cooperation. A selection of members of the Programme Committee could serve on the steering committee.
- The programme should be implemented in two steps to add a build-in learning process. The Computational Humanities programme is a new initiative involving scholars with different disciplinary and institutional background. It is recommended to start with three pilot projects to give momentum to the unorthodox combination of humanities and computational methods / approaches in order to find answers to established research questions and new questions. The steering committee, as part of their mandate to supervise the overall programme, should monitor the projects from the point of view of programme strategy. After one year, stakeholders meet to take stock, determine whether the programme remains on the road to success, and identify the factors that enhance/inhibit the development of computational humanities as a new approach.
- Participants in the programme should receive training to develop a shared understanding of research problems and approaches. Both substantive and interactional aspects of research should receive attention. A humanities perspective on the need to formulate conceptually coherent research questions in relation to a contextualised set of sources, and the formalisation of research problems and research objects from a computational science perspective should be part of the training. The training should also address the factors that make interdisciplinary teamwork a success. Examples of good practice are likely to be illuminating.
- Participants in the Computational Humanities programme should be concentrated at one location, at least during part (e.g. 50%) of the working time. A variety of support is needed, including both physical co-presence as well as computer-supported collaborative work practices.
- To foster the sharing of computational resources, the programme makes sure that institutes collaborate in setting up a connected computational infrastructure.
- The programme will yield several results in the form of components, libraries, applications and (web) services. An apt development process methodology and quality control process must be put in place to monitor and warrant seamless integration of such results in existing infrastructure of the Academy and beyond, and to guarantee the sustainability of IT results.

- Outreach activities should be considered to demonstrate the potential of computational methods for the humanities and to highlight the opportunities the humanities offer to Computational Science. Such activities are designed to maintain the engagement of Computational Humanities with emerging research in the 'mainstream' humanities and Computational Science. A specific format for this activity could be a 'road-show for graduate schools'. Other formats include seminars and workshops in cooperation with groups at universities.

The programme is divided into four work packages. They are:

- History
- Literature
- Music
- Integration and dissemination, including training.

It is recommended that working groups are organized on history, literature and music to promote discussions, cooperation and the development of shared interests and goals. Scholars may participate in more than one working group as long as a shared vision prevails.

Expected results include:

- New computational methods, algorithms and software to detect high-level concepts, patterns and motifs in text, images and sound that serve scientific understanding in history, literature and music. The new methods are expected to meet the challenge of capturing concepts, patterns and motifs from a humanities and computational point of view. They should be open-source to promote sustainability, dissemination and enhancement.
- Libraries, applications and (web) services.
- Research networks involving scholars from KNAW institutes and universities. The cooperation may evolve into a national network for computational humanities.
- Increased visibility of Computational Humanities in the research programmes of KNAW institutes.
- Renewed insight in the origin, structure and digital future of the collections maintained by the KNAW institutes. In addition, an improved insight is expected in the relative position of the collections and digital libraries among other collections in the Netherlands and abroad.
- Publications in academic journals, conference proceedings and book chapters. Digital scholarship (digital publications, open institutional repositories, websites, enriched publications) is considered important.
- PhD dissertations. Several PhD dissertations are expected to result from the programme.

The funds made available by the KNAW amount to € 3.5 million for a period of five years, i.e. € 700,000 per year. The KNAW requests partners in the Computational Humanities programme to contribute own funds. As a result, a total of five projects may be funded. Each project has either one post-doc researcher and two Ph.D. students, or one post-doc researcher, one Ph.D. student and one software engineer. The KNAW funds represent seed money to develop the field of Computational Humanities in a sustainable way and to effectively integrate information science and information technology in the humanities to obtain the breakthrough improvements envisaged in the introduction.

References

Bell, G., T. Hey and A. Szalay (2009). Beyond the data deluge. *Science*, 323 (6 March 2009), 1297-1298.

Bod, R. (2008). De unificatie van menselijke cognitie. *Algemeen Nederlands Tijdschrift voor Wijsbegeerte*, 100(2), 129-137.

Brewster, C. (2003). Knowledge maintenance and the Frame Problem, *Proceedings of the 6th Annual Colloquium for Computational Linguistics in the United Kingdom.* Edinburgh: AKT ePrint Archive id 130.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press.

Cocchiarella, N. (1991). Formal Ontology, In: Barry Smith, Hans Burkhardt (eds.), *Handbook of Metaphysics and Ontology*. Munich: Philosophia Verlag, p. 640.

Hey, T., S. Tansley and K. Tolle eds. (2009). *The fourth paradigm. Data-intensive scientific discovery*. Redmond, Washington: Microsoft Research. (http://research.microsoft.com/en-us/collaboration/fourthparadigm/).

Hofstadter, D. (2001). Analogy as the Core of Cognition, In: Dedre Gentner, Keith Holyoak, and Boicho Kokinov (eds.), *The Analogical Mind: Perspectives from Cognitive Science*. Cambridge, Mass.: MIT Press/Bradford Book, 499-538.

Honing, H. (2006). On the growing role of observation, formalisation and experimental method in musicology. *Experimental Musicology Review*, 1(1), 2-6.

Honing, H., & U.-D. Reips (2008). Web-based versus lab-based studies: a response to Kendall (2008). *Empirical Musicology Review*, 3(2), 73-77.

Howson, C. and P. Urbach (1989). *Scientific reasoning. The Bayesian approach*. La Salle, IL: Open Court Publishing.

Korb, K.B. (1998). The Frame Problem: An AI fairy tale, *Minds and Machines*, 8(3), 317-351.

Lerdahl, F. & R. Jackendoff (1983). *A Generative Theory of Tonal Music*. Cambridge, Mass.: MIT Press.

Mathijsen, M. (2003). *Naar de letter. Handboek editiewetenschap*, 3rd edition. The Hague.

Piaget, J. and B. Inhelder (1968). *Memory and intelligence*. London: Routledge and Kegan Paul.

Suppes, P. (1968). The desirability of formalisation in science. *Journal of Philosophy*, 65(20), 651-664.

Annex 1 – Related programmes and initiatives

The Computational Humanities Programme will take a leadership role in the development of new approaches to research in the humanities. This will be accomplished by supporting a number of innovative research projects and by contributing to capacity-building through training and user-engagement activities. In fulfilling this mission, the Programme distinguishes itself from other initiatives that are aimed at infrastructural development, and tool and service provision, such as CLARIN, DARIAH, HERA, DANS, SURF, TARGET, and the Arts and Humanities e-Science Support Centre (AHeSSC at King's College ,UK)[12]. These initiatives develop ICT infrastructure for the humanities by collecting and combining content and tools of different types and origins, by developing standards and interoperability devices that are needed for combining these elements, or by providing infrastructural standard and services, in contrast to the Programme which focuses on research. While the emphasis is different between these initiatives, it is nevertheless crucial that coordination and dialogue takes places so that 'supply and demand' can be aligned, whether with existing or developing efforts, such as the e-Science Research Center, the Science Park Watergraafsmeer, Amsterdam.

Other initiatives that focus on federating research in digital humanities (e.g. HASTAC) or in consolidating new approaches to research in the humanities (e.g. Alfalab) play an important role in community building. The Programme will not duplicate these efforts, and will seek to use these initiatives as platforms to contribute and learn from these communities. Existing links between these bodies and KNAW institutes will be useful here (six KNAW institutes are members of Alfalab; VKS is a member of the HASTAC).

Constructive interaction with these initiatives is essential, however, and will constitute one of the tasks of the Steering Committee. At the level of the funded research projects, researchers will be expected to make maximum use of the expertise within these initiatives through collaborations and partnerships. Where appropriate, researchers in the Programme will also ensure that any tools, protocols or platforms developed become part of the research infrastructure.

Other initiatives also fund research at the intersection of new technologies and humanities. In the Netherlands, the NWO programme CATCH is a related initiative, in which the main objective is to develop, in collaboration with scientific programmers and other ICT experts, new ICT tools that can be used in humanities research. Initially, the main focus of CATCH was on cultural heritage institutions and the data that concerns them (ICT). More recently, a third dimension was added: humanities research. A significant difference exists between the CATCH programme and the Computational Humanities Programme. In CATCH, research into technologies for the disclosure of cultural heritage data goes hand in hand with the development of access tool , whereas research into discovery techniques and training are central to the Computational Humanities Programme.

[12] Information on each centre or programme is given in the glossary (Annex 2).

Annex 2 – Glossary

AHeSSC (Arts and Humanities e-Science Support Centre at King's College ,UK
www.kcl.ac.uk/iss/cerch/advice/national/ahessc.html
AHeSSC supports, coordinates and promotes e-Science in all arts and humanities disciplines. It also liaises with e-Science and e-Social Science communities, computing, and information sciences.

Alfalab
http://alfalablog.huygensinstituut.nl/?page_id=68
A joint pilot project by six institutes of the Royal Netherlands Academy of Arts and Sciences (KNAW): DANS, the Frisian Academy, the Huygens Institute, the IISH, the Meertens Institute, and the Virtual Knowledge Studio (VKS). Alfalab is part of the KNAW's strategy of supporting humanities research in general and digital and computational humanities in particular. Alfalab applies and promotes the use of digital tools and methods in humanities research, and fosters cooperation of humanities researchers at national and international levels. The current Alfalab pilot project runs from 1 January 2009 – 1 March 2011. It is financed by the KNAW.

CATCH (Continuous Access to Cultural Heritage)
www.nwo.nl/catch
An NWO programme that develops generic methods and techniques cutting across the areas of the humanities and computer science, aiming to facilitate an interaction with cultural heritage institutions. Innovation, multidisciplinary collaboration and transferability are essential. Since 2005 CATCH finances teams which focus on improving the cross-fertilisation between computer science, humanities and cultural heritage. The teams consist of a Ph.D. student, a post-doc researcher and an IT programmer. In the light of transferability and interoperability, the research teams execute their research at the heritage institutions, according to the *laboratorium extra muros* formula. Currently CATCH is financing 14 research projects conducted in 12 cultural heritage institutions. CATCH is financed by NWO and is a coordinated effort from the Dutch cultural heritage institutions together with the two NWO divisions, the Physical Sciences (EW) and the Humanities (GW).

CLARIN (Common Language Resources and Technology Infrastructure)
www.clarin.eu
A pan-European programme to create, coordinate and make language resources and technology available and readily usable. CLARIN currently has 33 consortium partners, 168 member institutions and 184 members sites (working units) in 32 countries. The Dutch part of CLARIN is called CLARIN-NL; members are Meertens Institute-KNAW, DANS-KNAW, Huygens Institute-KNAW, UvA, VU, UT, RUG, UL, MPI, RUN, UvT, UU, INL, and the Digital Library for Dutch Literature (Leiden).

DANS (Data Archiving and Networked Services)
www.dans.knaw.nl/nl
DANS is an institute under the auspices of Royal Netherlands Academy of Arts and Sciences (KNAW) which is also supported by the Netherlands Organisation for Scientific Research (NWO). DANS archives data in the arts and humanities and social sciences and makes the data permanently accessible.

DARIAH (Digital Research Infrastructure for the Arts and Humanities)
www.dariah.eu
A pan-European programme to facilitate long-term access to, and use of all European arts and humanities data for the purposes of research. DARIAH is the digital research infrastructure that will connect scholarly data archives and repositories with cultural heritage for the arts

and humanities across Europe. DARIAH currently has 14 consortium partners, among which DANS-KNAW, where the project management of the programme is located.

Digital Humanities
Wikipedia has an interesting, well-articulated and multi-layered entry on digital humanities
http://en.wikipedia.org/wiki/Digital_humanities

E-Science Research Center (eSRC)
A new network organisation that should develop e-science concepts for all scientific disciplines. The concepts are subsequently to be produced and supported by SURF. eSRC is financed by, among other parties. NWO, SURF, and several companies.

Expertise center Target
www.rug.nl/target
A sustainable economic cluster of intelligent sensor network information systems in the Northern part of the Netherlands, aimed at data management for very large amounts of data. Prominent scientific research groups and innovative businesses jointly develop and improve complex and scalable data systems. The starting point here is the Target paradigm: full integration of large-scale data processing, archiving and analysis. In these experimental surroundings, the Target model is developed into actual market applications, and participants in follow-up projects will develop further products and services. Target has 10 partners, among which ASTRON, the UMCG, the AI department of the RUG, and companies such as Oracle and IBM.

Europeana
http://www.europeana.eu/
Multi-lingual online collection of millions of digitized items from European museums, libraries, archives and multi-media collections. Currently it links to 6 million digital items in several categories:

Images - paintings, drawings, maps, photos and pictures of museum objects
Texts - books, newspapers, letters, diaries and archival papers
Sounds - music and spoken word from cylinders, tapes, discs and ratio broadcasts
Videos - films, newsreels and TV broadcasts

Europeana.eu is funded by the European Commission and the member states.

GIS (Geographical Information System)
A software system developed in order to present and analyse geographical information (e.g. the geographical distribution of data concerning specific linguistic or cultural phenomena).

HASTAC (Humanities, Arts, Science, and Technology Advanced Collaboratory)
www.hastac.org
HASTAC ("haystack") is a network of individuals and institutions committed to new forms of collaboration across communities and disciplines fostered by creative use of technology. HASTAC is open to anyone. Participants share their work and ideas with the HASTAC community. HASTAC was founded by Cathy N. Davidson, former Vice Provost for Interdisciplinary Studies at Duke University, and David Theo Goldberg, Director of the University of California's state-wide Humanities Research Institute (UCHRI). For additional information, see the HASTAC website and http://en.wikipedia.org/wiki/HASTAC

HERA (Humanities in the European Research Area)
www.heranet.info
HERA is a partnership between fifteen Humanities Research Councils across Europe and the European Science Foundation, with the objective of firmly establishing the humanities in the

European Research Area and in the 6/7th Framework Programmes. The HERA ERA-Net was financed by the EU Framework programme 6's ERA-NET scheme and was established from the ERA-NET ERCH (European Network for research Councils in the Humanities) formulated by the Danish, Dutch and Irish Research Councils. Recently the HERA Joint Research Programme partners have launched a joint call for trans-national Collaborative Research Projects (CPRs) in two humanities research areas: “Cultural Dynamics: Inheritance and Identity” and “Humanities as a Source of Creativity and Innovation”.

Annex 3 – Computational Humanities Programme Committee (CHPC)

Chair
Prof.dr.ir. Frans Willekens
Director, Netherlands Interdisciplinary Demographic Institute (NIDI-KNAW), The Hague, and
Professor of Population Studies, University of Groningen
http://www.nidi.knaw.nl/en/staff/willekens/

Members
Dr. Anne Baulieu
Senior Research Fellow, The Virtual Knowledge Studio for the Humanities and Social Sciences (VKS-KNAW), Amsterdam
http://www.virtualknowledgestudio.nl/staff/anne-beaulieu/

Prof.dr. Rens Bod
Professor of Artificial Intelligence
University of St Andrews (UK) and
Institute for Logic, Language and Computation
University of Amsterdam
http://staff.science.uva.nl/~rens/

Prof.dr. Antal van den Bosch
Professor of Memory, Language and Meaning
Induction of Linguistic Knowledge Research Group and Centre for Creative Computing
Faculty of Humanities
Tilburg University
http://ilk.uvt.nl/~antalb/

Dr. Henkjan Honing
Associate Professor of Music Cognition
Institute of Logic, Language and Computation (ILLC)
University of Amsterdam
http://hum.uva.nl/mmm/hh

Prof.dr. Franciska de Jong
Professor of Language Technology
University of Twente
http://wwwhome.cs.utwente.nl/~fdejong/

Prof.dr. Kees Mandemakers
International Institute of Social History (IISH-KNAW), Amsterdam, and
Professor of Large Historical Databases
Erasmus University, Rotterdam
http://www.iisg.nl/staff/kma.php

Prof.dr. Lambert Schomaker
Professor of Artificial Intelligence
Artificial Intelligence and Cognitive Engineering Institute (ALICE)
University of Groningen
http://www.rug.nl/staff/l.r.b.schomaker/index

Dr. Joris van Zundert
Senior Researcher
Huygens Institute-KNAW, The Hague
http://www.huygensinstituut.knaw.nl/__eng/?p=440

Professional support
Dr. Corrien Blom
Senior Policy Advisor
KNAW, Amsterdam
http://www.knaw.nl/cfdata/telefoon/afdeling.cfm?id__afdeling=107

Annex 4 - Template for contributions to be prepared by the KNAW institutes
(prepared by Programme Committee)

i. Elaboration of questions with specification of how research questions can be answered via an approach that uses ‘patterns and motif detection’
ii. Why are the questions and results of this research valuable to the humanities?
iii. Why does the proposed research require a computational-humanities approach?
iv. Specification of expertise (in-house and needed)
v. Specification of collections needed for the project
vi. Connections of this work to other efforts (in-house and existing collaborations)
vii. What resources does the institute commit?