cultural science

$ sciendo

Cultural Science Meets Cultural Data Analytics

INDREK IBRUS

Tallinn University, Estonia; e-mail: indrek.ibrus@tlu.ee

MAXIMILIAN SCHICH

Tallinn University, Estonia; e-mail: maximilian.schich@tlu.ee

MAREK TAMM

Tallinn University, Estonia; e-mail:marek.tamm@tlu.ee

Keywords: cultural data analytics, cultural analytics, digital humanities, complexity, semiosphere, cultural semiotics, cultural science

For developing Cultural Science as a research field and practice it is worthwhile reconsidering the ways to approach the study of large corpora of digital content and data. In this context, Digital Humanities (DH) has been a success story in the academic world. However, we argue that it is better to consider DH as a transitory phenomenon that needs to be developed into more specific research fields, while at the same time it could benefit from being extended towards an even more multidisciplinary science. To achieve this, it is vital to first transcend the artificial division of cultural inquiry into the qualitative analysis of idiographic phenomena and the quantification of nomothetic phenomena. It is furthermore important to surpass the dichotomy of specific versus general as research objects; for example replacing this with the notion of the semiosphere as a research object, defined as the 'smallest' functioning element of culture by Juri Lotman. In this perspective, the singular cultural unit is always conditioned by the whole of the semiosphere, while the whole can be always changed by the singular, both in line with classic hermeneutic inquiry and recent notions of complexity science. Further, the label of 'humanities' in DH is at the same time both too large and too restrictive. We instead argue for a study of meaning-making practices in human society, but without confining ourselves to traditional humanities scholarship, but rather, learning from new developments in systems biology, evolutionary economics, complexity science and many more. We think that this new transdisciplinary field of study can help define the scope of the *Cultural Science Journal*. Indeed, it has already found practical application in a variety of 'post-DH' collaborations in 'Cultural Data Analytics', often with the aim to explore the dynamics of meaning-making practices by computational means and by looking at a spectrum of materials (textual, sonic, visual, multimodal, etc.) both regarding the *longue durée* and in real-time applications, if not anticipating the future.

Introduction

This new issue of the *Cultural Science Journal* marks the handover of the journal's editorship – it is moving from Curtin University in Australia to be hosted and published by Tallinn University in Estonia. The journal builds on 13 years of groundbreaking work linking the study of culture with novel approaches

within Internet studies, systems theory, complexity science, evolutionary biology, evolutionary and institutional economics, media economics and other approaches within the sciences and economics fields. The movement, which started at the Queensland University of Technology in Australia and has been driven by John Hartley, Jason Potts, Lucy Montgomery, Terry Flew, Stuart Cunningham, Carsten Hermmann-Pillath and many others, has produced a colossal body of work clearing the pathways between cultural and media studies, cultural theory, economics and the sciences. 'Cultural Science 2.0', as framed by John Hartley, was initially developed to build on the work of Raymond Williams and his interest not only in the micro-contexts of culture, but also in the systematic and general processes and the mutually conditioning relationships between the micro, meso and macro contexts. Hartley and colleagues efforts were aimed at conceptualising and studying such relationships systematically, working towards transdisciplinary frameworks linking the study of culture with evolutionary and insti-tutional economics, complexity science and other evolutionary approaches to change. This work has been pathbreaking, feeding into new approaches and further studies into evolutionary processes in media markets and cultural industries internationally (including the emergent fields of media inno-vation studies, cross-innovation systems and open publishing).

Yet, parallel approaches have also emerged, linking the study of culture with other disciplines, especially biology, but also mathematics, physics, data science, network science, etc., resulting in novel and epoch-making approaches, such as cultural analytics, cultural evolution studies, computational humanities and computational social science. In sum, 'Cultural Science', it can be argued, is evolving into an intensely dialogic, multidisciplinary and 'explosively' (to use a Juri Lotman term) evolving domain.

In this dynamically evolving context, in its next phase, the *Cultural Science Journal* is well placed to become the arena where the relevant dialogues are held, where centuries of humanities-based knowledge on cultural forms, languages and systems of meaning can meet the most current and forward-looking scientific methods, where new questions are asked and where novel conceptualis-ations can meet or emerge from rigorous empirical, analytical or numerical work. To set the stage in this article to inaugurate the new phase of the *Cultural Science Journal*, we would like to provide an example and a case study of how one subdomain of 'cultural science' (in the broadest possible sense), which we prefer to call 'cultural data analytics', should continue to evolve. To explain this, we first address the existing parallel concept and established domain of funding known as 'digital humanities'.

The rise and fall of digital humanities

'Digital humanities' has been one of the most successful buzzwords in humanities in the last 15 years. A very diverse and substantive body of research has been conducted under the banner, and it is safe to say that the study of digital humanities has shaped the current face of humanities more than any other new approach. However, in the context of this success story, it cannot be overlooked that 'digital humanities' is a problematic term in many respects. We propose that, while it was a very necessary flagship for the renewal of humanities research in the first two decades of this new century, by now the term 'digital humanities' has largely exhausted its potential and it is time to move on to more precise and concrete terms, while can also transcend the limitations of the present concept.

First, it might be prudent to ask: What is the problem with 'digital humanities'? To start with the most general point, digital humanities was haunted from the very start by difficulties in its definition. This may be best illustrated by the suggestion in the seminal reference work in the field, the *Handbook of the Digital Humanities*, published in 2004, which stated that 'digital humanities should be treated as a discipline in its own right' (Schreibman, Siemens & Unsworth 2004: xxiii). However, a dozen years later, when the same authors compiled a fully revised version of the compendium, titled *A New Handbook of Digital Humanities*, they self-critically acknowledged that 'it remains debatable whether

digital humanities should be regarded as a "discipline in its own right", rather than a set of related methods' (Schreibman, Siemens & Unsworth 2016: xvii). Despite many attempts, digital humanities scholars have not yet reached a consensus on how to define and delimit the new discipline (see most recently Luhmann & Burhardt 2021). A good illustrative example is the website https://whatis-digitalhumanities.com/, compiled by Jason Heppler, which brings together 817 definitions of digital humanities from 2009–2014.

To go into more detail, it is actually both sides of the word that are problematic: 'digital' and 'humanities'. Paradoxically, starting with the latter, 'humanities' is both too broad and too narrow a term to be used meaningfully. Too broad in the sense that there are few, if any, researchers today who identify themselves as 'humanities researchers' in general. It is also too broad an umbrella term to describe a specific field of research. Of course, it is possible to say that one's field of research falls within the humanities, but this needs to be followed by a specification – which field of the humanities specifically. There are very few professorships in the 'humanities' in the world, so it makes little sense to create them in the field of 'digital humanities', which is potentially no narrower a field. In addition, a closer look at digital humanities research quickly reveals that it does not cover all the humanities disciplines to the same extent. Most digital humanities studies are primarily concerned with textual data and philological research in the broadest sense of the word. However, clearly the poorest served are those areas of the humanities that work mainly with pictorial, material, spatial, sonic and multi-modal data. Digital art and music studies, or film and videogames studies, for example, are much less practiced than digital linguistics and literary studies. Thus, the term 'digital humanities' is somewhat misleading, as it promises more than it actually offers.

But 'digital humanities' is also too restrictive a term to describe digitally enriched research. It helps to keep alive the old, mainly 19th-century opposition between cultural and natural sciences, between the humanities and the sciences, which digital technology should help to overcome rather than perpetuate. Instead of limiting itself to the humanities, contemporary culture study should seek cross-disciplinary collaboration with computer science, network science, complex systems theory, machine learning and artificial intelligence research, and other 'non-humanities' disciplines (Barabási 2016; DeDomenico 2019; Russell & Norvig 2020). Thus, the aim should be to find appropriate names for the new field, which is still emerging, that do not follow the disciplinary boundaries of the century before last, but rather should encourage crossing the boundaries.

The prefix 'digital' is also problematic, as it inevitably raises an opposition to 'analogue' or 'traditional' humanities, yet this opposition is not so well-grounded. First, strictly speaking, every humanities researcher is a 'digital humanities researcher' in the sense that he or she uses digital tools in his or her research (be it word processors or online resources); the question therefore is only the extent to which digital tools are used. Second, it creates the illusion that digital humanities deal with a different subject matter from traditional humanities, even though digitising sources does not necessarily change their content or nature. Only a small section of digital humanities focuses exclusively on 'digital-born' information, while the majority of digital humanities work with digitised materials, the use of which generally requires the same professional skills as needed in 'analogue humanities'. Of course, under the aegis of digital humanities, a number of new methodologies have been introduced into humanities research; however, methodological innovation has been a feature of the humanities from the very beginning and does not necessarily imply renaming the field. Emphasising the prefix 'digital' can also lead to a situation where digital humanities researchers may be seen as support staff for classical humanities research, as a technical support service helping to calculate, visualise or model the work done by traditional humanities researchers. The current institutional organisation of work shows that many digital humanities specialists find jobs primarily in libraries and archives, rather than in universities, where they still have to search for jobs and fight for them by the skin of their teeth.

There is, of course, already a venerable tradition of critiquing 'digital humanities', both within the field and beyond. This is surely the reason why some authors have tried to play up the ambiguity of the title as one of its strengths. Matthew Kirschenbaum (2012), for example, refers to 'digital humanities' as a 'tactical term' that helps to achieve practical goals, such as obtaining funding or creating jobs. Digital humanities is often referred to as an 'umbrella term' (e.g. Jones 2014: 5) or as a 'big tent' under which many different types of activities can be accommodated (e.g. Svensson 2012; Weingart & Eichmann-Kalwara 2017).[1] James Smithies (2017: 2) refers to 'digital humanities' as a 'floating signifier'. Steven Jones (2014: 5) aptly sums up, 'It's still not clear whether digital humanities should be thought of as a field or an (inter)discipline, or just as a label of convenience for the moment for what all humanists will be doing relatively soon'. In another influential statement, Todd Pressner and Jeffrey Schnapp wrote in their 'Digital Humanities Manifesto 2.0' (2009): 'Digital Humanities is not a unified field but an array of convergent practices that explore a universe in which: a) print is no longer the exclusive or the normative medium in which knowledge is produced and/or disseminated; instead, print finds itself absorbed into new, multimedia configurations; and b) digital tools, techniques, and media have altered the production and dissemination of knowledge in the arts, human and social sciences'.

Increasingly, attempts are being made to refine the name of digital humanities with the help of a few complements. For example, some authors who see digital humanities as above all having a revolutionary potential that would play around with the current organisation of the humanities have called for the promotion of 'transformative digital humanities' (see Baldun & Deyrup 2020). Others have suggested integrating digital humanities with critical theory and have proposed the term 'critical digital humanities' (Berry & Fagerjord 2017: 136–150, see also Berry 2014), which would place a greater emphasis on critical reflection.

Yet all these attempts to play off the ambiguity of 'digital humanities' as a tactical strength or to limit it by adding new adjectives to the terminology do not seem very promising. We therefore propose that 'digital humanities', despite its popularity, should be seen as a transitional name, a temporary fighting slogan that has been useful in attracting attention and mobilising interest, but which, at this stage of its development, needs to be replaced by a more concrete name that is free of all the old contradictions.

We see a clear need to get rid of many of the divisions that are built into the digital humanities project. First, the distinction between the qualitative analysis of idiographic phenomena and the quantitative analysis of nomothetic phenomena must be overcome. Second, the dichotomy between the study of the individual and the general, the part and the whole, must be overcome. Third, the dichotomy between the digital and the analogue must be overcome, since there are many relevant computational and exploratory tasks that require non-digital solutions. Fourth, there is a need to move beyond the classical boundaries of the humanities and to develop dialogue with, for example, systems biology, evolutionary economics, machine learning, complex systems theory and many other areas of research. We propose that 'cultural data analytics' could be one of the research areas to be developed beyond the digital humanities and could help to overcome the formulated distinctions.

Making sense of cultural complexity

On the way towards our new approach, which we call cultural data analytics, we must first highlight the centrality of the concept of complexity and how it could be seen to link the different approaches to the study of culture; that is, the so-called nomothetic and idiographic perspectives to cultural

1   Big tent DH' was coined as the official theme of the international DH 2011 conference at Stanford University. https://dh2011.stanford.edu/.

phenomena and processes. Striving towards a deeper understanding of cultural phenomena and their development over time, the humanities have been handicapped and often actively held back by the assumption that their focus is and should be on *specifics*, often artificially contrasted with the supposedly exclusive focus of the natural sciences on more *general phenomena*. To the lay reader or specialist this may seem exaggerated, almost comical. Yet, as any multidisciplinary scholar could testify, there is a brutal reality underlying this statement. This reality, known as the so-called "two cultures" (Snow 1959), is not rooted in the factual correctness of the statement – the premise is wrong indeed – but in the function of the statement as a weapon in peer-review, in hiring committees, and in funding decisions.

For a broader audience it makes sense to recapitulate the origins of the argument about two supposedly mutually exclusive scientific worlds. The ideological separation of the humanities, including implicitly digital humanities, from the natural sciences is rooted in 19th century philosophy, specifically that of Wilhelm Dilthey, who contrasted the German terms *Geisteswissenschaften* versus *Naturwissenschaften* (Dilthey 1883); literally "sciences of the mind/spirit" versus "sciences of nature". It is notable from the outset that this separation of a seemingly quasi-divine mind/spiritual world from the physical world not only predates massive advances in all sciences, but even in the 19th century was contrary to perspectives such as that forwarded by Goethe, as cited in the very first issue of *Nature* in 1869, that "even the most unnatural is still nature" (Huxley 1869). While *Nature* ironically went on to mostly ignore cultural scholarship, Wilhelm Windelband, another German scholar, furthered the separation of humanities and natural sciences, coining two words to characterise the supposedly inseparable difference (Windelband 1894). The natural sciences, he said, are primarily "nomothetic" (literally, "law-giving") as they seek to generalise patterns into laws. An obvious example would be that physical objects "simply" follow the laws of motion identified by Newton. The sciences of the mind/spirit, Windelband said, are on the other hand "idiographic", focusing on the specific. According to this view, historical events always differ, much like the never fully repeatable situation in which we are writing or in which you are reading this article. Adopting such a perspective, which fits well with more recent perspectives in post-structural thought, it must seem that quantifying cultural phenomena aims to enter forbidden territory, as the only goal could be to establish nomothetic laws of culture. In the context of broader conceptualisations of digital culture, one example has been Friedrich Kittler's critique of Niklas Luhmann's systems-theoretical sociology. Kittler developed a new approach (also known as "media archaeology") to the study of unique contingencies that change historically according to the material and technical resources at one's disposal at any time. This led him to a radical historicism that sought to dissolve the universality of concepts such as "media" or other cultural institutions of meaning-making and communication. Based on this, he also opposed broader law-like theories of social evolution, including those of Luhmann (see Kittler 1994, Winthrop-Young 2000: 411).

There have also been counter-movements and intentions, most notably by Raymond Williams when he articulated the necessity for "cultural studies" as a new approach to the study of culture. Williams had taken his lead from the Birmingham Centre for Contemporary Cultural Studies (CCCS), where the early focus had been on the import of Continental "cultural sciences" to English language academia (see Shuttleworth 1966). Following this, Williams gave a speech in 1974 where he similarly proposed a new approach to what he first called "cultural science" and later corrected to "cultural studies" ('which is English for "cultural science"' – Williams 1974). In that speech he too built on the German intellectual tradition of Dilthey, Weber, and Marx. But he wanted to move further; to identify the "central problem" of cultural science for the latter part of the twentieth century as "the relations between different practices" (see Hartley 2009). While cultural studies mostly remained focused on the micro-contexts of culture, it is perhaps now with the emergence of cultural data analytics that such systematic study of relationships in culture can be imagined and carried out. However, it must be emphasised that humanities research has never been free of "law-like" generalisation. Indeed,

humanities textbooks and education are full of intuitive generalisations, types, and categories, which at least implicitly presuppose some kind of convergent average over a quantity of examples. The gothic arch on the 20-euro banknote is such a generalisation. Similarly, we can imagine an average gothic cathedral, even though there is no such average. The first cathedral that comes to your mind is probably an outstanding and not an average example. The prime example is typically surrounded by a larger set of increasingly less typical examples. Indeed, the feedback between specific observations and such generalisation is an integral part of the humanities, known as the so-called hermeneutic circle, which is another concept of German philosophy connected with Dilthey (Bod 2014).

The issue is obviously not either specification or generalisation. Instead, it is clear that inquiry of both specifics and generalisation are essential activities if we are to make sense of cultural phenomena. The true issue is that cultural phenomena are complex, not necessarily following simple quantitative laws with simple meaningful averages. This is crucial, and must be explained a little further. "Complex" in this circumstance is not equivalent to "complicated" or "depressing", which are the two most common notions of the word, familiar to a broad audience (Zinoviev 2016). Indeed, "complex" is used here in the sense of complexity science (Mitchell 2009, De Domenico 2019), in some sense contrasting with the common notion of complication. An individual historical event, much like a single painting or a visualisation in a biology paper, can be complicated, necessitating keen observation and differentiated qualitative inquiry to further our understanding. On the other hand, a larger set of seemingly simple individual cultural interactions may result in a complex situation, subject to some regularity while concurrently subject to specific variation. Following complication to the bottom, we can get lost in a myriad of specific yet different details, while complexity may emerge from a few simple generating mechanisms. The murmurations of a flock of starlings, all similar and following a number of relatively simple rules, are an example of such "organised complexity" (Weaver 1948). Urban structure and dynamics emerging from the local activities of urban dwellers are another (Jacobs 1961, Barthelemy 2019). Similar phenomena, where organised complexity emerges from the local activities of small entities, permeate our world. Nuclear physics gives rise to molecular chemistry, which gives rise to cellular biology and organisms, next giving rise to cognition and social interaction, eventually giving rise to tangible and intangible products of culture (Anderson 1972, Riedl 2000).

At each level, where complexity emerges from local activity, "more is different", as Anderson pointed out. This results in a perhaps surprising opportunity for qualitative inquiry. The result of quantification in the humanities will likely not be simple quantitative laws, but novel forms of quality. New and interesting clusters, patterns of confusion, and a zoology of global statistics with intriguing exceptions are emerging from the concert of local specifics. At the meso level and the macroscopic level we find identifiable qualitative phenomena that are neither perfectly regular nor random, much like the urban street patterns or the recognisable continental topographies that become visible as an aeroplane climbs to cruising altitude on a long-distance flight (Lee et al. 2017, Mandelbrot 1982). From large amounts of simple local interactions, such as connecting the birth and death locations of one hundred thousand individuals, the meta-narrative of cultural history and long sought-after patterns of *moyenne durée* emerge at the meso level, as implicit in the dataset, while the demographic "laws of migration" become evident on the macro level (Schich et al. 2014, Ravenstein 1885, Hexter 1972). Systematic bias, beyond the natural cognitive limit of individual humanistic expertise, becomes evident through the analysis of multiple large datasets in a comparative approach, while all of the above must remain opaque to a more exclusive focus on local specifics.

At the same time – and this is crucial – a shift in focus takes place, from "substance" to "function", from "objects" to "interaction", from "structure" to "dynamics", and from "synthesis" to "analysis" (Schich 2019, Cassirer 1910). While one may argue that the humanities were always interested in contexts and the unfolding of events, it is undeniable that several core disciplines, from art history and literature to media studies, still primarily define themselves and therefore remain centred around substance,

such as artworks, texts, film, or other forms of media. Meanwhile, unprecedented progress has been made towards a deeper understanding of interaction, with the multidisciplinary science of "complex networks" increasingly permeating all areas where complexity is shown to emerge from local activity, from physics to culture (Barabasi 2016, Schich et al. 2016, Ahnert et al. 2021).

As we quantify the emerging complexity arising from cultural interaction, traditional hermeneutic research in the humanities is not at risk of being replaced with shortsighted mechanistic or positivistic models of culture. Instead, quantification of emergent complexity has the potential to provide a corrective within the collective hermeneutic circle, as the consensus of humanistic logic and intuition is confronted with quantitative evidence that cannot be imagined without a rigorous quantitative cartography of cumulative evidence.

The Singular and General Constitute a Semiosphere

In a search for a new data-analytical basis for a cultural studies research programme, one answer could be Juri Lotman's cultural semiotics, and in particular his semiosphere theory (1990), a comprehensive framework for analysing complex and dynamically evolving cultural processes. We propose this systems-theoretical and holistic approach to cultural dynamics as a good basis for the study of contemporary digitally mediated or enabled cultural processes (see Hartley, Ibrus and Ojamaa 2021).

The good fit between cultural semiotics and cultural data analysis is not accidental. When developing his approach, Lotman drew heavily on cybernetic theories (notably the work of Norbert Wiener and Ross Ashby) and various other approaches (including Ilya Prigogine's (2018) work on dissipative structures and self-organisation), which are now considered precursors to both systems and post-humanist theories. We also highlight here the influence of Jakob von Uexküll's work (1909) on further developments in cultural semiotics after Lotman (Kull and M. Lotman 1995, Kull 1998). In the context of the systemic study of digital culture, it is important to mention that Uexküll's work also strongly influenced Ludwig von Bertalanffy in the development of his "general systems theory" (1951)[2].

Both the aforementioned and the contemporary developers of cultural semiotics have been influenced above all by the concept of the *Umwelt* (perceived environment). This suggests that the selves of different species of wildlife, as of people, can differ greatly, depending on the specific affordances of their senses. A primary shared system of meaning emerges in the translations between *umwelts* or self-worlds. The more such translations or communicating entities that there are, the more there are separate systems of meaning, all of which together form what is called a semiosphere – a shared, co-produced system of meanings (plural). It should be stressed, however, that for Uexküll and later for Lotman, the formation of any larger system, including culture, began with the auto-communication of its subsystems (Kull and Lotman 2012, Torop 2008). For Lotman, auto-communication is the ability of a system to talk about itself, to distinguish and define itself in its environment, and thereby to create itself communicatively. It is also the ability to create self-referential texts, to translate what appears to be meaningful from the environment into the system, to integrate it continuously with its existing codes and thus to create a new system of its own. As different systems – for example, different artistic genres, political ideologies, or communities of start-ups – form an environment for each other, the development of culture thus involves processes whereby these systems interpret each other and thus recreate their own systems in a changing environment, adapting them to the surrounding environment. Such systems can range from the very small (e.g., a music genre created by a few artists and lasting for a short period of time) to the medium (e.g., a national culture) to

---

2   Bertalanffy in turn is understood to have been influenced by Aleksndr Bogdanov – see the special section of this issue titled "Eisenstein, Bogdanov, and the organization of culture".

the very large – the whole of human culture, for example, or the audiovisual culture co-created on YouTube by billions of people.

However, the central idea of cultural semiotics is that all the levels are mutually conditioned. For example, a single new musical genre could not come into being without the rest of the cultural space having existed before it. It is meaningful, specific, and "new" only in the context of the rest of culture. It is causally dependent on all that has gone before. At the same time, however, when a new cultural form or grouping and the discourse that carries it are born, they in turn disrupt the entire existing cultural system. All the other cultural sub-systems must also, directly or indirectly, take account of it and adapt to it, and so the entire global semiosphere is ultimately transformed. Another guiding principle of cultural semiotics relates to the above: systems at different levels, following Lotman, are "isomorphic", in the sense they are structured in a similar way, and by analysing one you can draw conclusions about the other. It is precisely on the basis of this principle that it is possible, for example, to draw important conclusions about the structural foundations of past cultures from individual material fragments. It is here where humanists have dealt with emergent complexity for a long time, and where complexity scientists may find novel challenges. For instance, when building on concepts such as renormalisation derived from physics (cf. the attempts of Fáth Savary 2005 and Thompson et al. 2018).

Hence, semiosphere theory is one possible compromise or unifier between the nomothetic and idiographic approaches described above. Its premise is that one conditions the other, that by studying one it is possible to draw conclusions about the other. More importantly, by studying the auto-communicative dynamics in different cultural subsystems, both very small and very large, and the translations between them, one can systematically explore the causal links between the developments of different cultural subsystems over time. We explained above why complexity is an important concept for understanding and contextualising cultural phenomena at different levels, that complexity can be considered both at the micro level of culture, when studying "specific" phenomena, and at the meso or macro level, where interactions between micro-level processes result in the emergence of new quality. From a cultural-semiotic perspective, cultural complexity is the result of and is catalysed by the auto-communicative or modelling action of systems at different levels. Small systems are simultaneously embedded in a number of larger cultural systems, which are nomothetic to the smaller ones, but which together bring complexity to their environment or context. At the same time, the auto-communication and modelling activities of small systems produce new complexities that can lead to changes in large systems. Such emergence of complexity at different levels is already being explored, and could be further explored by new methods of cultural data analysis. Until now, such a research programme, based on the ideas of cultural semiotics, was only hypothetical or possible on a small scale. But now, with the support of large datasets and an evolving arsenal of cultural data analysis methods, it is becoming a reality.

We now highlight some key concepts in cultural semiotics that could and should be used in different ways in the analysis of cultural data. The first of them does not derive directly from cultural semiotics, but from "cultural science" – the very approach to which this journal is dedicated. This line of inquiry was originally established by the Australian scholar, John Hartley, and his colleagues. It draws on cultural semiotics and cultural studies, complexity science, and evolutionary economics (see Hartley and Potts 2014, Hartley 2020). The term is a "deme", which refers to an auto-communicating group of people united not only by a shared self-referential discourse and the texts that carry it, but also by common media channels, forms and means of communication. Examples of this today are shared Telegram or Facebook groups, or perhaps tightly integrated Twitter networks. An older example could be the European clergy before the advent of the printing press, as they created and read shared literature and were united by a common medium, common genres, and common ideas and discourses that gradually evolved over time. Twitter-based memes are nowadays easy to study

using network analysis methods, but using the metadata corpus of the Estonian Film Database, for example, the historical networks and memes of Estonian filmmakers can be and have been analysed just as successfully (Ibrus and Ojamaa 2020).

By looking at the Facebook groups of Estonian subcultural music communities and their auto-communicative practices (Järvekülg and Ibrus 2021), it is possible to examine how self-identity is created in these subcultures, how they differentiate themselves from their environment, and how this is done over long periods of time. By looking for the links between groups and by examining the similarities and differences in the auto-communication that takes place within these groups, it would be possible to further explore how different subcultures are linked, how information moves between them, and how they interpret each other. Drawing on Lotman's cultural semiotics and on Hartley and Potts' cultural science, the study of deme differentiation, even conflict, is equally important. For in differentiation and conflict, important alternatives of different eras are expressed. They also produce explosive developments that can result in new dialogues or novel cultural systems operating on new foundations. Hartley, Ibrus and Ojamaa (2021b), for example, show the conflict between global climate movements and online communities of conspiracy theories, which, paradoxically, have also become dialogical – these systems are co-evolutionary. Perhaps, by exploring the dialogic practices, conflict, translation, and auto-communication at different scales through data analysis methods, a broader picture of cultural dynamics will eventually begin to emerge.

We foresee that such a type of research programme could be developed into applied, forward-looking research on a range of contemporary challenges (the spread of misinformation, rapidly escalating online conflicts, etc.). This means that our proposed research programme could also have a cultural and media policy dimension. Let us think back to the development of meteorology. It took nearly two centuries to reach its current level of accuracy. Initially seen as a hopeless folly, it is now one of the most advanced fields, demonstrating the success of the rapid processing of large volumes of data, while providing a public good accessible to all. In the mid-19th century the first experimenters were mocked. Today it is information essential for planning the global economy and the daily lives of societies. Indeed, it is an area of growing importance as the climate crisis unfolds. Could cultural data analysis also evolve into a foresight science (i.e., a science that is calculative or anticipatory rather than predictive with certainty)?

Research in public universities aimed at unlocking this potential is all the more important because the private sector has been doing it in one way or another – as became clear to many people with the Cambridge Analytica scandal. The latter, however, was of relatively limited impact and scope. Most of the data about our social lives, our connections and interests, our cultural meaning-making, is collected and stored by large global platforms as well as state actors. Couldry and Mejias (2020) call this practice "data colonialism". While the first wave of colonialism involved Western countries annexing territories to obtain raw materials, the new wave of colonialism, they argue, is whereby large platforms and their associated "data industries" colonise people's lives, both private and social, to extract value (Sadowski 2019). With the spread of 5G technology and the Internet of Things, and with the building of an all-encompassing "metaverse", this commodification of lives through data mining could become widespread, especially when datafication and platformisation models characteristic to Web 2.0 are not hindered.

The natural aim of cultural data analysis based on cultural semiotics should be to undermine the processes described above. It should be borne in mind that all modelling, including scientific modelling, adds patterns to the wider cultural space. As Peeter Torop (2015) has shown, it is the awareness of the results of one's actions that characterises cultural semiotics. This does not mean that one should refrain from taking action, but rather contribute with useful models. If the thesis of cultural semiotics has been that the development and good functioning of a culture depends on its intrinsic diversity,

then analytical activities must contribute with models that enhance this diversity, opening up new realities as well as pointing to the plurality of cultural subsystems.

In addition, semiospherical modelling can be used to show how knowledge and values in a culture are created and spread through the actions of many people. Such modelling is in itself illuminating about the mechanisms of culture, potentially reducing superstitions, mistrust, and inter-group fears. But at the same time it helps to predict periods of conflict and turbulence. If we accept that human culture, or the semiosphere, will influence and shape other systems (biosphere, geosphere, atmosphere) in the Anthropocene era, then the planetary functioning of the semiosphere must be understood as well as possible in order to assess its risks to the living environment. This implies that the foresight of global cultural processes should aim first and foremost at creating "public value" (Benington and Moore 2011, Mazzucato 2018) – producing and sharing knowledge that is accessible to all actors. This means, above all, forecasting for the sake of good governance.

Emphasising all the above potentials, it must be acknowledged that such modelling of global cultural dynamics also has its important drawbacks. Let us recall Borges' (1975) short novel, *On the Exactitude of Science*, in which the cartographers of an unknown empire produce a detailed 1:1 map of their entire country. Its inhabitants found it useless and the map was destroyed. With 5G mobile communications, the Internet of Things, and augmented reality applications fast becoming a reality, the idea of creating a "mirror world" or "digital twin" of all physical space is being toyed with again. However, in this context, we should remember that when (meta)data are used to model social or cultural or physical reality, the models they create do mimic, but they also interpret and simplify. As such they are inaccurate, and inevitably "lie". We refer here to Umberto Eco's (1977: 7) famous assertion that semiotics studies everything that can be used to lie. Indeed, as we now know from studies of data bias (e.g., Milan and Treré 2019, Perez 2019) and data injustice (e.g., Heeks and Shekhar 2019, Dencik et al. 2017), data are often misleading, leading to injustices of various kinds and thus also to new conflicts. A particular stance of data studies based on cultural semiotics in dissecting such data relationships could be that if "lying is inevitable" (i.e., if any representations, including data-based representations, are always selective, representing only a certain part of reality to the exclusion of the rest), then the focus should be on how these choices affect cultural dynamics at different levels of culture – at the micro, meso, and macro contexts. As an example of a micro-context, this would entail the question of how metadata schemas in specific databases can affect the insights users gain from historical events (Coleman 2020, Ibrus and Ojamaa 2018). At the meso-level, data discrepancies can affect, for example, comparisons of national cultures (Fisher and Poorting 2018, Erlin et al. 2021). At the macro-level they can affect global summaries, as eloquently described by Lev Manovich (2020), or as exemplified in "a network framework of cultural history" by Schich et al. (2014).

The Tallinn Approach of Cultural Data Analytics

Taking into account all aspects as outlined above, it is possible, and in line with the original ideas for cultural science, to establish a systematic approach of cultural data analytics that transcends the problematics of digital humanities, desegregates the so-called two worlds, and harnesses the joint opportunity of cultural semiotics and complexity science. Indeed, based on the identified compatibilities, we can enact a holistic spectrum of approaches, allowing individual researchers to maintain their specialities, while being embedded in a systematic framework that can aim for a multifaceted yet eventually integrated understanding of the semiosphere. The necessary expertise (i.e., the epistemic disciplinary communities to be integrated) include cultural semiotics, art history, cultural history, media studies, policy research, creative industry research, educational technology, evolutionary economics, network science, computational social science, computational linguistics, machine

learning, user experience design, cultural physics, and generative art. A multidisciplinary team, involving many of those disciplinary knowledge fields, has been assembled at Tallinn University's Cultural Data Analytics Open Lab. In this lab, researchers make use of a large heterogeneity of source materials, including audiovisual, image, text, and numerical data, unstructured and structured, in the form of databases or knowledge graphs, including open data sources and data from public and private institutional stakeholders. The challenge and the promise lies in the establishment of a joint systematic approach. The methodological risk can be mitigated as researchers can bring expertise in their areas, while having the opportunity to go beyond their respective states-of-the-art through multidisciplinary collaboration.

The potential of such an approach is similar to the emergence of systems biology, where an equivalent integration of qualitative inquiry and quantification led to unprecedented progress in understanding and a continuous stream of benefit to society. Similarly, this novel approach to cultural analysis, properly nurtured, could result in a sustained groundswell, beyond the fashion fad of novel terminology. To summarise, we propose that a more integrated approach to cultural analysis, that transcends the artificial segregation of the so-called two worlds, is a way forward. Theoretically, this approach is rooted in the so-far unexploited resonance of cultural semiotics (Tamm 2019, Tamm and Torop 2022) and cultural complexity science (Mitchell 2009, Schich et al. 2014, De Domenico 2019), and can build on the "cultural science" approach (Hartley and Potts 2014, Hartley 2020, Hartley, Ibrus, and Ojamaa 2021a). Empirically, the approach could be supported by a growing corpus of research which harnesses the common roots of research in cultural history, networks, higher-order topology, and computation (Schich 2019).

Acknowledgements

References

Ahnert, RS.; Ahnert, SE.; Coleman, CN. and Weingart, SB. 2021. *The Network Turn: Changing Perspectives in the Humanities*. Cambridge: Cambridge University Press.

Anderson, PW. 1972. More is different. *Science*, 177(4047): 393–396.

Baldun, MM. and Deyrup, MM. (eds) 2020. *Transformative Digital Humanities: Challenges and Opportunities*. London and New York: Routledge.

Barabási, A-L. 2016. *Network Science*. Cambridge: Cambridge University Press.

Barthelemy, M. 2019. The statistical physics of cities. *Nature Reviews Physics*, 1(6): 406–415.

Benington, J. and Moore, MH. 2011 *Public Value: Theory and Practice*. London: Palgrave Macmillan.

Berry, DM. and Fagerjord, A. 2017. *Digital Humanities: Knowledge and Critique in a Digital Age*. Cambridge: Polity.

Berry, DM. 2014. *Critical Theory and the Digital*. London: Bloomsbury Academic.

Bertalanffy, L. v. 1951. General system theory: A new approach to unity of science. *Human Biology*, 23(4): 302–312.

Bod, R. 2014. *A New History of the Humanities*. Oxford: Oxford University Press.

Borges, JL. 1975. On exactitude in science. In: JL. Borges, *A Universal History of Infamy*. London: Penguin.

Cassirer, E. [1910] 2000. Substanzbegriff und Funktionsbegriff. In: E. Cassirer, *Gesammelte Werke*. Hrsg. B. Recki. Hamburg: Felix Meiner Verlag.

Couldry, N. and **Mejias, UA.** 2019. The *Costs of Connection: How Data is Colonizing Human Life and Appropriating for Capitalism*. Stanford: Stanford University Press.

De Domenico, M; Brockmann, D.; Camargo, C.; Gershenson, C.; Gershenson; Goldsmith, D.; Jeschonnek, S.; Kay, L.; Nichele, S.; Nicolás, JR., Schmickl, T.; Stella, M.; Brandoff, J.; Martínez Salinas, AJ. and **Sayama, H.** 2019. *Complexity Explained*. DOI: doi.org/10.17605/OSF.IO/TQGNW

Dencik, L; Hintz, A.; Redden, J. and **Treré, E.** 2019. Exploring Data Justice: Conceptions, Applications and Directions. *Information, Communication & Society*, 22(7): 873–881.

Dilthey, W. [1883] 1959. *Gesammelte Schriften*. Vol. 1. Stuttgart: Teubner.

Eco, U. 1977. *A Theory of Semiotics*. Bloomington: Indiana University Press.

Erlin, M.; Piper, A.; Knox, D.; Pentecost, A.; Drouillard, M.; Powell, B and **Townson, C.** 2021. Cultural Capitals: Modeling 'Minor' European Literature. *Journal of Cultural Analytics*. DOI: doi.org/10.7910/DVN/SDMGRC.

Fáth, D. and **Miklos, S.** 2005. A renormalization group theory of cultural evolution. *Physica A: Statistical Mechanics and its Applications,* 348: 611–629.

Fischer, R. and **Poortinga, YH.** 2018. Addressing Methodological Challenges in Culture-Comparative Research. *Journal of Cross-Cultural Psychology,* 49(5): 691–712.

Gold, MK. (ed.) 2012. *Debates in the Digital Humanities*. London and Minneapolis: University of Minnesota Press.

Gold, MK. and **Klein, LF.** (eds) 2016. *Debates in the Digital Humanities*. London and Minneapolis: University of Minnesota Press.

Gold, MK. and **Klein, LF.** (eds) 2019. *Debates in the Digital Humanities*. London and Minneapolis: University of Minnesota Press.

Hartley, J. 2009. From Cultural Studies to Cultural Science. *Cultural Science*, 2(1): 1–16.

Hartley, J. 2020. *How We Use Stories and Why That Matters: Cultural Science in Action*. New York: Bloomsbury.

Hartley, J.; Ibrus, I. and **Ojamaa, M.** 2021a. *On the Digital Semiosphere: Culture, Media and Science for the Anthropocene*. London: Bloomsbury.

Hartley, J.; Ibrus, I. and **Ojamaa, M.** 2021b. Emergent self-mediating classes in the digital Semiosphere: Coronavirus conspiracies and climate justice movement. *Nordic Journal of Cultural Studies*, 3: 118–140.

Hartley, J. and **Potts, J.** 2014. *Cultural Science: A Natural History of Stories, Demes, Knowledge and Innovation*. London: Bloomsbury.

Hexter, JH. 1972. Fernand Braudel and the *Monde Braudellien… Journal of Modern History*, 44(4): 480–539.

Heeks, R.; Shekhar, S. 2019. Datafication, development and marginalised urban communities: an applied data justice framework. *Information, Communication & Society*, 22(7): 992–1011.

Huxley, T. 1869. Nature: Aphorisms by Goethe. *Nature*, 1: 9–11. DOI: doi.org/10.1038/001009a0.

Ibrus, I. and **Ojamaa, M.** 2018. Newsreels versus Newspapers versus Metadata: A Comparative Study of Metadata Modelling the 1930s in Estonia. *VIEW: Journal of European Television History and Culture,* 7(14): 123–137.

Ibrus, I. and **Ojamaa, M.** 2020. The Creativity of Digital (Audiovisual) Archives: A Dialogue Between Media Archaeology and Cultural Semiotics. *Theory, Culture & Society*, 37(3): 49–70.

Jacobs, J. 1961. *The Death and Life of Great American Cities*. New York: Vintage Books.

Järvekülg, M. and **Ibrus, I.** 2021. Auto-communicative reconstruction of meaningfulness in musical randomness: reclaiming musical order on Facebook. *Media, Culture & Society*, Online First, DOI: doi.org/10.1177/01634437211045513.

Jones, SE. 2014. *The Emergence of the Digital Humanities*. London and New York: Routledge.

Kirschenbaum, MG. 2012. Digital Humanities As/Is a Tactical Turn. In: MK. Gold (ed.), *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, pp. 415–428.

Kittler, FA. 1994. Wenn die Freiheit wirklich existiert, dann soll sie doch ausbrechen. In: Kittler, FA., *Am Ende vorbei*. Hrsg. R. Maresch. Wien: Gespräche, pp. 95–129.

Kull, K. 1998. On semiosis, Umwelt, and semiosphere. *Semiotica*, 120 (3/4): 299–310.

Kull, K. and **Lotman, M.** 1995. Semiotica Tartuensis: Jakob von Uexküll ja Juri Lotman. *Akadeemia*, 7(12): 2467–2483.

Kull, K. and **Lotman, M.** 2012. Semiotica Tartuensis: Jakob von Uexküll and Juri Lotman. *Chinese Semiotic Studies*, 6(1): 312–323.

Lee, M.; Barbosa, H.; Youn, H.; Holme, P. and **Ghoshal, G.** 2017. Morphology of travel routes and the organization of cities. *Nature Communications*, 8(1): 1–10.

Lotman, Y. 1990. *Universe of the mind: a semiotic theory of culture*. Bloomington and Indianapolis: Indiana University Press.

Luhmann, J. and **Burghardt, M.** 2021. Digital humanities – A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape. *Journal of the Association for Information Science and Technology*, online first, DOI: doi.org/10.1002/asi.24533.

Mandelbrot, BB. 1982. *The Fractal Geometry of Nature*. Vol. 1. New York: WH Freeman.

Manovich, L. 2020. *Cultural Analytics*. Cambridge, Massachusetts: MIT Press.

Mazzucato, M. 2018. *The Value of Everything: Making and Taking in the Global Economy*. Milton Keynes: Allen Lane.

Milan, S.; Treré, E. 2019. Big Data from the South(s): Beyond Data Universalism. *Television & New Media*, 20(4): 319–335.

Mitchell, M. 2009. *Complexity: A Guided Tour*. Oxford: Oxford University Press.

Perez, CC. 2019. *Invisible Women: Exposing Data Bias in a World Designed for Men*. London: Chatto & Windus.

Pressner, T. and **Schnapp, J.** 2009. *The Digital Humanities Manifesto 2.0.*, http://manifesto.humanities.ucla.edu/2009/05/29/the-digital-humanities-manifesto-20/.

Prigogine, I. and **Stengers, I.** [1984] 2018. *Order Out of Chaos: Man's New Dialogue with Nature*. London: Verso Books.

Ravenstein, EG. 1885. The laws of migration. *Journal of the Statistical Society of London*, 48(2): 167–235.

Russell, S. and **Norvig, P.** 2020. *Artificial Intelligence: A Modern Approach*. 4th ed. Boston: Pearson.

Riedl, R. 2000. *Structures of Complexity: A Morphology of Cognition and Explanation*. Berlin: Springer.

Sadowski, J. 2019. When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6(1): 1–12, DOI: doi.org/10.1177/2053951718820549.

Schich, M. 2016. Figuring Out Art History. *International Journal for Digital Art History*, 2, DOI: doi.org/10.11588/dah.2016.2.24761.

Schich, M. 2019. *Cultural analysis situs*. ART-Dok eprint. Open access: doi.org/10.11588/art-dok.00006347.

Schich, M.; Song, C.; Ahn, Y-Y.; Mirsky, A.; Martino, M.; Barabási, A-L. and **Helbing, D.** 2014. A Network Framework of Cultural History. *Science*, 345(6196): 558–562.

Schich, M., Meirelles, I.; Malina, R. (eds) 2016. *Arts, Humanities, and Complex Networks. 5th ed.* [Kindle Edition]. Cambridge: MIT Press.

Shuttleworth, A. 1966. *'A Humane Culture' and 'Max Weber and the "Cultural Sciences', Two Working Papers in Cultural Studies.* Birmingham: Centre for Contemporary Cultural Studies, University of Birmingham.

Snow, CP. [1959/1964] 2012. *The Two Cultures.* Cambridge: Cambridge University Press.

Schreibman, S; Siemens, R.; Unsworth, J. (eds) 2004. *A Companion to Digital Humanities.* Malden, MA; Oxford: Blackwell.

Schreibman, S.; Siemens, R. and Unsworth, J. (eds) 2016. *A New Companion to Digital Humanities.* Malden, MA; Oxford: Wiley.

Smithies, J. 2017. *The Digital Humanities and the Digital Modern*. London: Palgrave Macmillan.

Svensson, P. 2012. Beyond the Big Tent. *Debates in the Digital Humanities*, edited by Matthew, K. Gold. Minneapolis: University of Minnesota Press.

Zinoviev, D. 2016. The Pain of Complexity. *Leonardo*, 49(5): 450–450.

Tamm, M. and Kull, K. 1996. Toward a reterritorialisation of cultural theory: Estonian theory from Baer via Uexküll to Lotman. *History of the Human Sciences*, 29(1): 75–98.

Tamm, M. 2019. Introduction: Juri Lotman's Semiotic Theory of History and Cultural Memory. In: M. Tamm (ed.), *Juri Lotman – Culture, Memory and History: Essays in Cultural Semiotics*. Basingstoke: Palgrave Macmillan, pp. 1–26.

Tamm, M. and Torop, P. (eds). 2022. *The Companion to Juri Lotman: A Semiotic Theory of Culture*. London: Bloomsbury Academic.

Terras, M.; Nyhan, J.; Vanhoutte, E. (eds) 2013. *Defining Digital Humanities: A Reader.* Farnham: Ashgate.

Thompson, W.H.W.; Wojtowicz, Z.; DeDeo, S. 2018. *Lévy Flights of the Collective Imagination,* arXiv:1812.04013.

Torop, P. 2008. Translation as communication and auto-communication. *Sign Systems Studies*, 36(2), 375–397.

Torop, P. 2015. Kultuuriteooria kui kultuuri eneseteadvus [Cultural Theory as Cultural Self-Consciousness]. *Akadeemia*, 27(4), 626–649.

Uexküll, J. von. 1909. *Umwelt und Innenwelt der Tiere.* Berlin: J. Springer.

Weaver, W. 1948. Science and Complexity. *American Scientist,* 36: 536–567.

Weingart, SB. and Eichmann-Kalwara, N. 2017. What's Under the Big Tent?: A Study of ADHO Conference Abstracts. *Digital Studies/Le Champ Numérique*, 7(1), DOI: doi.org/10.16995/dscn.284.

Wiener, N. [1948/1965] 2019. *Cybernetics or Control and Communication in the Animal and the Machine.* Cambridge, Massachusetts: MIT Press.

Windelband, W. 1894. *Geschichte und Naturwissenschaft: Rede zum Antritt des Rectorats der Kaiser-Wilhelms-Universität Strassburg.* Strassburg: Heitz & Mündel.

Williams, R. 1974. Communications as cultural science. *Journal of Communication*, 24(3): 26–38.

Winthrop-Young, G. 2000. Silicon Sociology, or, Two Kings on Hegel's Throne? Kittler, Luhmann, and Posthuman Merger of German Media Theory. *The Yale Journal of Criticism*, 13(2): 391–420.

Authors information

Indrek Ibrus is Professor of Media Innovation at Tallinn University's (TLU) Baltic Film, Media and Arts School (BFM), Estonia. He curates also BFM's doctoral programme. His research interests include media innovation, the evolution of the ubiquitous internet, emergence of contemporary metadata formats, broader evolution of modern creative industries, and the implications of cultural heritage digitisation. He has published on mobile media, media innovation/evolution, metadata evolution, transmedia and cross-media production. He is a co-editor (together with Carlos A. Scolari) of Crossmedia Innovations

(Peter Lang, 2012), editor of Emergence of Cross-Innovation Systems (Emerald, 2019) and co-author (with John Hartley and Maarja Ojamaa) of On the Digital Semiosphere (Bloomsbury, 2020).

Maximilian Schich is a Professor for Cultural Data Analytics and the CUDAN ERA Chair holder at Tallinn University. A multidisciplinary researcher, he aims to understand the nature of cultural interaction via a systematic combination of critical and creative aesthetics, qualitative inquiry, quantitative measurement, and computation. Ongoing research builds on a background in art history, network science, computational social science, and an applied experience in cultural "database pathologist". In 2014, "A Network Framework of Cultural History" in Science Magazine made international impact. In recent years, he has focused on the upcoming "Cultural Interaction" book, which will outline a systematic science of art and culture based on two decades of work. In June 2020, he moved to Estonia to build, manage, and sustain the CUDAN research group and leading the ERA Chair project.

Marek Tamm is Professor of Cultural History at the School of Humanities in Tallinn University. His primary research fields are cultural history of medieval Europe, theory and history of historiography, and cultural memory studies. He has recently published The Companion to Juri Lotman: A Semiotic Theory of Culture (co-edited with Peeter Torop, Bloomsbury 2022); A Cultural History of Memory in the Early Modern Age (co-edited with Alessandro Arcangeli, Bloomsbury 2020), Juri Lotman – Culture, Memory and History: Essays in Cultural Semiotics (Palgrave Macmillan 2019), Rethinking Historical Time: New Approaches to Presentism (co-edited with Laurent Olivier, Bloomsbury 2019), and Debating New Approaches to History (co-edited with Peter Burke, Bloomsbury 2018).

OPEN ACCESS