EXTENDING OURSELVES? ON THE CONCEPT AND FUTURE OF DIGITAL HUMANITIES

Digital Humanities is increasingly featured in philosophical discussions and has potential to significantly impact philosophy of science in practice. But what is Digital Humanities, what problems and questions can it help us address, and what are the prospects for the future?

Ariel Roffé and Sara Green here talk to Charles Pence and Henrik Kragh Sørensen, who are both frontrunners in this field.

15 - Charles Pence is a philosopher and historian of science and technology with a special interest in the life sciences. Charles works as Chargé de cours at the Institut supérieur de philosophie and the Faculté de philosophie, arts et lettres at the Université catholique de Louvain in Louvain-la-Neuve, Belgium.

16 - Henrik Kragh Sørensen is a historian of science, focusing in particlar on the history and philosophy of mathematics after 1800. Henrik is Professor at the Section for History and Philosophy of Science, Department of Science Education, University of Copenhagen, Denmark.

What is the Digital Humanities, in your view?

CHARLES: If you ask ten digital humanists this, you'll get twenty-five answers to the question. But I think the basic idea can be captured by the idea of using computers to answer the same kinds of questions that we would have traditionally been concerned with in the humanities. That might be as simple as managing a large electronic bibliography, or as complex as an online museum exhibit or a digital analysis of a corpus of texts.

HENRIK: Another divide in the DH community is where to focus on the deeply interdependent issues: The compilation of high-quality corpora through digitalization and annotation and the use of computational tools in answering questions about the topic domain. Analysis without high-quality data is empty, data without purpose and interesting questions is blind. But my particular competence is more in the computational aspects - extracting and combining existing metadata and using tools from machine learning to add nuance to known philosophical questions or raise new possibilities.

What can we use DH-tools for? What questions can they help us answer? Do you have any examples of works (either your own or of others) that illustrate this?

CHARLES: There are almost too many to mention! I'll stick with my own work, which has largely surrounded questions in text analysis. For some time now, I've been working to approach the

following question: how are changes in scientific concepts reflected in the scientific literature? To be sure, a massively difficult problem, not least for reasons that will be very familiar to readers of this newsletter: as we know after the development of the philosophy of science in practice, concepts are never clearly and exclusively reflected in journal articles, and so the interpretation required here is really significant! For now, we have a handful of analyses in process, around several different concepts – biodiversity, fitness, causal specificity – but getting from texts to answers to philosophically interesting questions is the work of years (after developing tools and corpora, which is *also* the work of years!).

HENRIK: The field of philosophy of mathematical practice deals with many of the same questions as SPSP, but focusing on mathematics, we often run into the additional challenge that parts of the key epistemic processes are private and mainly accessible *through* the mathematicians. Thus, we have been searching for ways to gain triangulation with empirically informed philosophizing. The field has been largely based on case studies and interviews, but about a decade ago, Inglis and Aberdein used Principal Component Analysis to analyze a questionnaire study in which they asked a medium size population of mathematicians about their perceived associations between adverbs describing mathematical results. Thus, their paper entitled 'beauty is not simplicity' showed that mathematicians were *not* associating the beauty of proofs mainly with simplicity as was the dominant philosophical analysis. Thereby, they added a quantitative calibration (they might even call it correction) to the philosophical analysis of aesthetics of mathematics which is (and was) an established topic in philosophy of mathematics.

HENRIK: In our group we have been working both in the direction inspired by 'beauty is not simplicity' of empirical triangulation about mathematical practice to try to avoid the small-N problem which is often raised as a criticism of other parts of PMP: 'exemplar philosophy' is a derogatory term we sometimes hear (see also Pitt's Dilemma about case studies). Thus, we have been active in broadening the case base, especially in the philosophical study of mathematical diagrams: As one of our first contributions, we trained a machine-learning agent to detect mathematical diagrams from (scanned) images of mathematical texts. That allowed us to gather far more diagrams whose epistemic roles we want to reason about and provided us with a historical overview that was difficult to get by other means. In fact, the strenuous efforts of my colleagues Mikkel and Josefine to scan 55.000+ pages of mathematical publications counting and classifying diagrams was the direct reason why I got involved: "I can do that faster with a computer!" And in the effort, we found that we could ask and answer so many additional interesting questions: Have the roles of diagrams changed over time? Do diagrams and their use vary between fields? What are good/typical/rare/… types of diagrams (in given contexts)? Trying to answer such questions will advance the study of mathematics *as it is practiced*, we believe, and are prerequisite to a practice-based epistemology of diagrams. And using computational techniques we have gone way beyond 55.000 pages and ventured into The Great Unread.

What motivated you to get started in this field?

CHARLES: I was looking (with my graduate advisor, Grant Ramsey) to better understand what exactly it is that scientists say about "fitness." We have a collection of a dozen or so papers that philosophers of biology always read, but it's hard to know on that basis whether we have really captured the "spirit" of the biological literature as a whole. This is exactly the sort of thing that DH tools should be able to help us with, and now a decade after we got started, I'm finally able to clearly frame that kind of question in our technical system.

HENRIK: As mentioned above, the particular challenge of automating the counting of diagrams got me interested in the first place. Since then I have been able to re define my research focus, combining four expertises that sum up my focus: history and philosophy of mathematical and computational sciences. And getting to write code again, doing a bit of statistics, studying actual mathematics, and raising and addressing historical developments has made me feel like I have really found a home and a niche for myself.

CHARLES: I really agree with this. I've found it enormously mentally satisfying to be able to put down the world of prose and deep philosophical analysis and use a different part of my brain every once in a while – I think this kind of work could be really attractive if you're looking for a little more variety in your everyday philosophical work!

HENRIK: Another great thing I have found in my new niche is that it is very easy to form collaborations - with colleagues and students - because interesting and unanswered research questions can be 'parcelled out' quite easily. But that raises the greatest challenge: Time (see below).

How do you view the epistemology of Digital Humanities, compared to traditional philosophical analysis?

CHARLES: I could say *a lot* here, and have a few recent papers doing just that. But what I'll say briefly is that we often forget how complex and difficult the epistemology of old-fashioned close reading actually is, just because we've all been doing it for so long! So we have lots of catch-up work to play in thinking about how digital work can inform an empirically informed approach to the philosophy of science.

HENRIK: Again, I completely agree with Charles - and it is a question that occupies me (and most DH'ers, I guess): How are we adding? Whose norms are we trying to meet? Are we to be seen merely as a tool? Personally, I am most interested in doing DH *for* a practice-informed philosophy (of mathematics). That's why I called my agenda DH4PMP: Digital Humanities for Philosophy of Mathematical Practice. A successful project often arises when we can make three ends meet: An interested philosophical problem or field to add to (which provides relevance and 'hook'), a high-quality corpus to study (which provides validity), and a relevant computational tool to apply (which is typically the key to saying something *interesting* that goes beyond data itself, trying to avoid too much idiosyncrasy).

CHARLES: I really like Henrik's idea of framing our work in terms of audience. We don't want to create a sibling discipline, off doing its own work without any connection to the problems and debates that brought us to philosophy in the first place.

What kinds of reactions do you get to your work in this area? Have you encountered any resistance to the use of quantitative methods in HPS?

CHARLES: Of course. DH's loudest proponents sometimes talk as though we can "commit to the flames" anything that's not founded on large sample sizes. But that's silly. I think the way we build bridges is in saying that what we get from these tools are ways to supplement our traditional close readings with quantitative studies, giving us access to kinds of complementary knowledge that just weren't available before.

HENRIK: I agree that DH is not opposed to small-N studies - on the contrary. But adding the 'ordinary' (in history of literature often referred to as The Great Unread following Franco Moretti) to the corpora we can study, we actually get a fairer picture of the *practice*. Thus I find DH more relevant to the practice-based approaches which are also often more methodologically varied already. I would

not consider myself a philosopher, but certainly a scholar in the study of mathematical practices and cultures and for that DH is my new telescope.

HENRIK: Within the philosophy of mathematical practice, I have thus far mainly (only) encountered great enthusiasm about the DH approach - and a wonderful curiosity about what it (and our group) can add to existing projects.

What do you see as the greatest challenges - in your own work and/or in the field of Digital Humanities in general?

CHARLES: This stuff just takes *time.* I have difficulty recommending to junior scholars to get started in it: I've been very lucky to have worked in contexts where people were happy to let me cultivate these projects with a very, very long view to corresponding results. It's taken us literally around ten years of coding to wind up with a really nice, clean system for analyzing texts and a corpus that's worth analyzing. I think we need to think very hard about how to build smaller projects and make this kind of work much more accessible to people new to the field!

HENRIK: I would also say that *time* is the critical bottleneck - but time here means many different things: The time to train the required computational competences, the time to developing operationalizations of philosophical questions in computational terms, the time to develop, revise and implement the pipelines and tools that we rely on, the time to manage collaborations and projects, the raw time required to download and process huge data sets, etc.

How do you view the future of Digital Humanities?

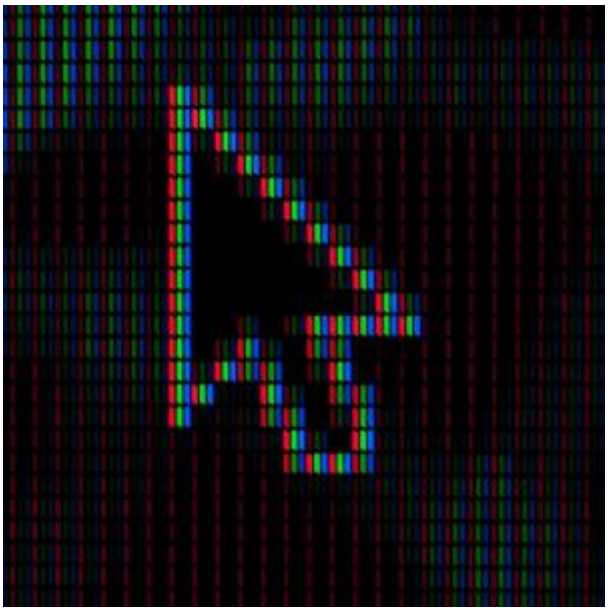
CHARLES: I think things are really bright. Whenever I see the kinds of work being done by colleagues here, I'm continually floored by what amazing stuff people are up to. I ran an online meeting in 2021 and a quick glance over the topics is, I think, extremely exciting for where this area might lead!

HENRIK: My personal ambition would be to make DH another possibility in our toolbox for doing philosophy of mathematical or scientific practices and integrated history and philosophy of science. That requires that we provide recognizable contributions using these tools and that we can collaborate both within a DH specialty and between DH and the larger domain. But DH is (and should, I think) be one among a number of ways of providing triangulation for the humanities. I have also found that studying highly normalized epistemic domains (such as mathematics) through computational tools make computer scientists interested in what we want to achieve.

How could one start using DH-tools? Is there any specific/technical knowledge required? (for instance, programming languages, statistics and/or bibliometric analysis terminology, etc.)

CHARLES: So for all that I mentioned this as a huge challenge, I think that more and more there *are* really nice ways to get started using existing tools and existing platforms. I'll plug our own system, Sciveyor, which is publicly accessible for anyone who wants to analyze our corpus of 1.9M scientific journal articles. But there's also a few other user-friendly systems that you can use to look at your own texts: Voyant Tools is a great example. There's also ways to get trained: I can't recommend strongly enough the Digital Humanities Summer Institute (DHSI), with training in everything from Python to project management.

HENRIK: I would only add that it might be fun to reach out to one of the groups already working with computational tools to explore possible ways of applying these tools to your research questions.



What kinds of collaboration would you be open to? Where should people contact you?

CHARLES: We are really at the moment where we have this system, we've done a lot of really great infrastructure work, but we haven't extracted enough results from it yet! So I'm always open to further collaborations, proposals of research topics that people think would be usefully answered by the corpus that we've put together. Send me a message at <charles@charlespence.net>!

HENRIK: I am also very open to collaborations and like Charles, we have different things to offer to researchers in philosophy of mathematical and scientific practices: We have pipelines for a number of relevant corpora such as the arXiv or bibliographic data sets, we have experience in various methods of (semantic) information extraction, we have a methodology (and ideology!) of operationalizing research questions for DH. To learn more, you can visit DH4PMP; and please send a message to henrik.kragh@ind.ku.dk if you are interested or have questions!