Twin Talk: Bukvik+LitTerra+Colabo.Space - An Example of DH Collaboration Across Disciplines, Languages, and Style

Sasha Mile Rudan, Eugenia Kelbert, Lazar Kovacevic, Sinisa Rudan and Matthew Reynolds

sasharu@uio.no, ekelbert@hse.ru, lakinekaki@gmail.com, sinisa.rudan@gmail.com, matthew.reynolds@ell.ox.ac.uk

Abstract. This paper focuses on a long-term collaboration between the two poles of the ***DH-dipole***; the ***D-pole***: a CSCW (Computer-supported cooperative work) - Computer Science scholar, Sasha Rudan, and the ***H-pole***: a Comparative Literature scholar, Eugenia Kelbert. It involves work with a larger team as well, including this paper's co-authors, among others. Our research started through a mutual interest in the digital analysis of stylistic features of fictional texts, mostly novels. Eventually, it developed towards designing a new ecosystem for collaborative research in the textual and stylometric DH domains. From a practical research question in stylometry in translingualism, we evolved to developing new tools, a DH infrastructure, later a DH research collaboration ecosystem and meta-research questions addressing challenges of DH collaboration and its practical solutions. Here, we discuss the oppositions between the different disciplines involved, the challenges we faced on the road, and how we tried to avoid them by getting a level higher in our collaboration.

Keywords. DH collaboration, methodologies, workflows, stylometry, research challenges

1 Introduction

The primary participants in this collaboration already had experience working outside their field, and were prepared for the peculiarities of an interdisciplinary collaboration to some extent. Eugenia was working on a project comparing literary texts stylistically across languages and had reached the conclusion that a DH perspective would be complementary to the close analysis she otherwise based her argument on. She therefore took a course on computational linguistics (in Python) in the first year of her PhD program at Yale and later sought out another collaborator in Computer Science, William Teahan at Bangor University, with whom she worked on a conference paper in 2011, a few months before her and Sasha's collaboration started. She had also had some exposure to authorship attribution methods and probability theory, and was familiar not so much with contemporary stylometry as with the pioneering work by Shannon and the great Russian mathematician Kolmogorov. Sasha, in his turn, had always had an

interest in the Humanities, publishing poetry and performing slam poetry, as well as being active in literary campaigns in Serbia. At the same time, he was an active member of the DH community and worked, pre-DH, on various projects ranging from interactive text and media to visualizing novels and poems in an appealing way, juxtaposing writers, texts and facts. His dream was to get access to the archives of the Serbian Nobel Prize winner, Ivo Andrić (Иво Андрић) and understand him through the help of DH analysis. Similarly, he wanted to tame the wild metaphors of the Serbian neo-symbolist and surrealist, Branko Miljković (Бранко Миљковић).

1.1 Background Reflections

This previous history is key for the positive results of this collaboration, for two reasons. First of all, few collaborations start from scratch, and participants invariably bring in their agendas and experience. Establishing some vocabulary in common, and a mutual appreciation of the other field's methods, is perhaps the one prerequisite for any successful DH work in the long run. Such an appreciation can never be taken for granted in a DH collaboration, where both sides, however genuinely intrigued by the possibilities of working together, often have to overcome misunderstandings: the literary scholar may be skeptical about the extent to which scientific method can be usefully applied at the level of literary analysis, or feel threatened by such methods, and the computer scientist is liable to consider literary analysis to lack the formalism and the empirical grounding of a scientific approach.

Luckily, both collaborators had a degree of understanding of the other discipline's language and approaches, perhaps more than many starting off in DH. For example, Eugenia's knowledge of programming—albeit minimal—was invaluable. Unable to contribute to the code herself, she could understand it, when explained, and discuss it in some detail, which made a major difference to the project's progress. In this sense, we cannot stress enough the advantages of time invested in even the most basic acquaintance with the other collaborator's field of expertise, even if it appears meaningless (in Eugenia's case, for example, she may have not taken the course in Python thinking it would not be enough to code what she wanted on her own, and therefore not a good investment of her time; nevertheless, it was).

2 The birth of the project, or a DH research methodology

In terms of what each researcher brought to the project, it is important to note that such 'dipole' collaborations may be of three primary kinds. One [1] is where one of the parties has a project and enlists the other fully into it (for example, in a situation where the 'D' researcher hires literary experts to create training sets, or an 'H' researcher engages a programmer to create a tool for them). Its limitation lies in the fact that the enlisted party has no inherent motivation, may or may not contribute original thinking to the project, usually needs to be paid for their contribution and clearly there is no

interdisciplinary innovation involved. Another [2] is where one of the parties has a finalized corpus ('H') or tool ('D') the other decides to use, as for was the case in Sasha's collaboration with Biljana Dojcinovic (corpora of feminist literature) and Eugenia's collaboration with William Teahan (tools for textual compression), respectively. The limitation here is that the preexisting corpora/tool becomes a Procrustean bed that limits what the researcher can achieve significantly, and forces them to adapt the knowledge/method to what is available. This is, indeed, the issue with most stylometric projects relying on pre-existing tools, however flexible.

Finally, perhaps the most promising but also the most complex scenario is what the present collaboration ended up to be, namely two or more researchers who each has a stake in the mutual project and is therefore internally motivated.

2.1 Dimensions of Freedom (or Interests)

Initially, our work started with a range of **different dimensions,** or rather interests that were at the same time challenging, and opened new opportunities and improved both our individual and collaborative research processes. Below, we present some of these dimensions and the researchers' "place" along them.

1) Tools: the D-pole: to understand how the DH stylistic distant-reading process may be improved to provide better and more targeted/useful results and new insights, and the H-pole: to use available DH tools to get insights into the style of bilingual writers compared to native-speaker writers.

2) Languages of interest: Eugenia Kelbert's main languages of interest were English, Russian, French, and German, and Sasha Rudan's languages of interest were English, Serbian (and other former Yugoslavian languages), and Russian where the former Yugoslavian languages were under-supported languages (in the NLP+stylometry scope). Both of them had a general interest in languages well-supported in the NLP+stylometry domain.

3) Collaboration scale: the D-pole was customized to a higher-scale real-time collaboration with various stakeholders with a high interest in inter-disciplinary collaboration. On the other hand, the H-pole tends to support lower-scale collaborations, and less real-time collaborative work, and is usually less used to inter-disciplinary collaboration.

4) Close reading: in our collaboration, the H-scholar's expertise lies in the close reading of bilingual writers (among others), while the D-scholar's competence comes from his undergraduate education, as well as from being a writer of poetry and short stories.

5) Distant reading: in our collaboration, the primary D-scholar's expertise is in NLP and data analysis, and system modelling especially for collaboration, and the H-scholar's competence lies in introductory programming courses and stronger mathematical background.

6) Research workflow tools: the D-scholar's research interest lies in optimizing teams' face-to-virtual workflows and enhancing knowledge federation. On the other hand, the H-scholar had a basic knowledge of the Python ecosystem and higher than average computer literacy, but no exposure to elaborative digital research workflows.

7) Research methodologies: The D-scholar's are statistical quantitative and qualitative methodologies, comparative evaluation and participatory design and, partially, action research. The H-scholar's preferred research methodologies fall within the fields of qualitative analysis and archival research.

8) Infrastructure evolution: finally, the D-scholar aimed to design, research and optimize the workflow, while the interest of the H-scholar was in the availability, consistency, and reliability of the research workflow.

Challenges and their resolutions

True ***inter-disciplinary collaboration*** of two researchers with equal stakes in the project comes with several benefits, but also significant challenges. Coming from different disciplines, two (or more) researchers bring rich innovation dimension to mutual work and much stronger overall expertise and likeliness of correct and successful project finalization. On the other hand, given distinct, and usually not strongly overlapping, research agendas, they are liable to have wildly divergent investment in the project, leading to unexpected developments and inevitable compromise.

The challenges of collaboration in our case lay mainly in two categories; [1] the "***collaboration***" category relating to different practices and previous experience in collaboration and the "***research-interests***" category relating to different research interests in the project and overall collaboration – for example; Sasha's strong research interest was in the continuous evolvement of the DH tools and methodologies through participatory design and action research. While this is an interest Eugenia eventually came to share, her primary interest is in using tools and conducting stylometric research. This means that she was especially invested in workflow stability, which opposed Sasha's research interest. This bipolarity of the our skills and research interests, which were largely complementary, introduced inevitable tension during the project's critical milestones. However, we had respect for the methodologies each of us brought to the project, were keen to expand the range of languages covered, and wanted to improve the tool to be both powerful and, crucially for both, flexible to evolve over the long term as competing technologies evolved. In other words, despite tension coming from non-aligned interests and collaboration practices, we had mutual goals in terms of the resulting set of tools and methodologies, which largely helped with the ongoing success of the project and the collaboration itself.

In terms of the ***DH tools and infrastructure***, Sasha's interest and that of another collaborator he brought into the project, Lazar Kovacevic, was in language-agnostic (when possible, or multilingual when there were language specific requirements) solutions, scalable to work with a high and reproducible volume of research. For example, our LitTerra infrastructure deals with the whole Gutenberg corpus counting over 45'000 texts with various intertextual and intratextual analyses. On the other hand, what Eugenia wanted was a set of tools supporting her project, since existing tools either did not satisfy her needs, or were too hard to unite into a consistent workflow and/or unequal to tackling large corpora in several languages systematically. Sasha's answer to that challenge was not to deal with and maintain every single tool in a conceptually consistent research workflow, but rather to propose a "*one ring to rule them all*." In this way, he could avoid unhealthy maintenance of separate tools, but also provide a reproducible environment for parallel experiments against multiple corpora. Sasha's interest as a researcher was in workflows and systems facilitating collaboration and knowledge federation, so that the tool itself had for him an added value as a case study of such a system. The result was that, on the level of the research workflow, the project took on a life of its own as a workflow-based system rather than a simple toolkit, but with the capabilities required by the initial project. In other words, a great deal of flexibility, as well as patience, was required of both parties to accommodate each other's research needs. During this process, each collaborator became a contributor to the theoretical and methodological aspects of the other's research pursuit.

An interesting disbalance and semantical inequality of the D and H disciplines lay in the fact that our D-related work resulted in a rather generic tool that could be used by H-scholars without relying on a D-researcher. However, the H-scholars' results were not "reusable" for D-researchers. For example, for stylistic analysis of Former Yugoslavian authors, there was not much help (apart from certain methodological aspects) from material associated with the writers Eugenia was interested in. On the other hand, collaboration on designing and conducting stylometric research at generic and meta levels helped Sasha and the other D-contributors (Sinisha Rudan and Lazar Kovacevic) to transfer practical and tacit knowledge and conduct research on Former Yugoslavian authors (Rudan et al, 2019-Torun) as well as ongoing research with Matthew Reynolds on his Prismatic Jane Eyre project.

Differences in working styles when it came to ***collaboration*** proved to be another major, and unexpected, challenge. In our case, this seemingly innocuous difference, which one would have expected to be a lot less of an apple of discord than, say, methodological differences, became one of the hardest issues to overcome in our work together. We were both open-minded and willing to learn and to accommodate the other discipline's methods and approaches. We were, however, a lot less willing to change our day-to-day workflows. A humanities scholar tends to do most of their work at their own pace, and have entrenched ways of working, and can be resistant to the practices of structured collaboration, brainstorming, regular meetings, documentation, etc. These are only partially personal differences of style; they are largely down to divergent cultures of research within the different disciplines. Even in writing this paper, after seven

years of working together, we experienced tension over Sasha expecting Eugenia to write her part of the paper in bullet points, and Eugenia insisting to formulate her thoughts writing in full sentences from scratch. A compromise we arrived at was that she wrote her parts first but then highlighted the internal structure for Sasha to use the highlights as 'bullet points' of sorts to integrate into the overall argument. Even minor factors such as using different textual editors, or Sasha's insistence on Markdown format and GitHub for documentation as a form that allowed for easier integration with the coding environment, added to the cognitive load of adapting not to one tool (the one we were developing together), but to several different interfaces and ways of working.

On the larger scale of ***project development***, it was a challenge for Eugenia to write user documentation for the program we created as the form was alien to her, and once she learned how to do one task or another, she did not feel the need for a separate record. This, in turn, made Sasha's work harder, since omissions in documentation meant he had to repeatedly not only re-teach his collaborator after a break in the project, but also often re-teach himself, as he would also forget the parameters in running a given version of the tool. For Eugenia, on the other hand, it was a source of frustration that the procedure of running the program and the interface—not intuitive for an H-scholar—had to be **relearned for each version** as the system improved or needed to be restructured.

Finally, and perhaps crucially, both collaborators had very different tacit assumptions about the **development process** (the infrastructure-evolution dimension). For Eugenia, the very concept of the coding workflow took time to absorb. On the other hand, she had to deal with discomfort when she recognized with time that the tool, once functional, was never set in stone but kept developing, both as it grew and improved, and also as the external libraries and tools it relied on also changed, triggering the need for several instances of top-to-bottom refactoring. For a D-scholar, this was the normal—indeed expected—price of a system's evolution and progress. From an H-scholar's perspective, however, it came as a surprise that our work depended on external—and evolving—systems and that a function that already worked seamlessly could easily require an upgrade five months later.

As these brief profiles demonstrate, much in what we had to bring to the project shares core attributes with those of an average literary scholar who is not a novice in digital humanities (i.e. who has a traditionally humanities research agenda and experience working with stylometrical tools, perhaps some instruction in the area) and those of an average computer scientist interested in the humanities (personal interest and background but little formal training). Perhaps more unusual, in our case, was the focus on stylometrical tasks across languages and, for the D-scholar, the research interest in system architecture, which he brought to the project. On the whole, our experience, and that of finding a mutual research language and procedure, illuminates both the core challenges and the potential of close inter-disciplinary collaboration as a solution to existing challenges in Digital Humanities as a field.

In this paper, we discuss our findings and the co-evolution leading us toward these findings. As we were introducing additional collaborators to our research team, adding

additional projects and participating in external grants, we understood the importance of a proper collaboration strategy and even more, of developing a collaboration ecosystem.

3 Research Questions Trajectory

3.1 How a new tool has born - Bukvik (research)

While brainstorming potential DH tools for our first mutual project, the D-scholar had the initiative to establish an internal DH infrastructure. His primary reasons were to ensure uniform analysis of texts and user experience, which would be agnostic of the tools used, provide continuous research workflow and work as a reproducible research environment for comparative stylometric analysis. This is how the Bukvik infrastructure was born and presented at the SCLA Conference in Zagreb, Croatia, 2012. From that moment, we embraced Bukvik as our internal infrastructure that helped us to incorporate some aspect of our collaboration in practice, and evaluate future needs. It became the playground for our future tools, a prototype of our understanding of what a DH-framework should be.

3.2 Initial Research Questions

Two main research questions we started our collaboration journey with were in the domain of: 1) translingual stylometry and 2) flexible corpora analysis infrastructures.

The emerging field of stylometry is still far from being able to fully grow out of methods it inherited from authorship attribution and distant reading, which shaped it with their own aims and priorities. This takes both time and a different generation of computational tools that would focus on stylistic features for their own sake rather than for the sake of clustering and identification. Our goal with our central project, ***Bukvik***, has been to fill this gap, first, in terms of relying on *a custom-made tool* with an initial focus on cross-lingual textual comparison. Secondly, it extends the principle of *multidimensional analysis*, identified by Jockers, to a potential stylistic profile: the sum total of quantifiable stylistic features for each text or body of texts that, together, constitute a multidimensional model of the given *writer's style* with reference to a balanced corpus of fiction in the given language. It supports, further, a novel method of textual analysis based on the visualization of individual words in a literary text as a network. This work relies on interdisciplinary collaboration to enable the development of original tools. The tool's modular structure ensures its relevance beyond the features that we are capable of tracking today and extends the relevance of the stylistic profile model beyond the specifics of the current project (cf. Jockers, 2013; Hoover, 2014).

The translingual stylometry aspect of the collaboration seeks practical solutions to quantifying those of the possible stylistic markers that current language processing tools are already capable of tracking and contextualizing this work within the

theoretical framework of comparative literature. Is style separate from the linguistic norms of a given language? Is content? Eugenia's dissertation on bilingual writers (Yale University, 2015) strongly suggests that it is not, or at any rate not fully. The featured bilingual authors' corpora were used for the initial digital comparison in the collaboration. We aim to extend it to other corpora, notably translated texts with their originals and corpora in the NLP-underdeveloped languages (like Former Yugoslavian languages as part of South Slavic languages, although the scene dramatically changed in this aspect in the past few years with dedicated research like The CLARIN Knowledge Centre for South Slavic languages (CLASSLA) and more universal tools like Adobe's Cube NLP, Universal Dependencies framework and treebanks).

3.3 Secondary Research Questions: Methodology

Our study draws on an original methodology that aims to make a real contribution to computational stylistics or stylometry. This approach complements Moretti's more popular method of *distant reading*. The system is conceived as an aid for automatic zooming: unlike the "distant reading" approach where statistics replaces reading and helps process large corpora, we see Bukvik as a non-automatic *augmenting* framework that will ultimately aid and direct close reading. "Close reading at a distance" is one way to describe the idea behind the methodology. This goes together with the research approach we refer to as *Qualitatively Augmented Quantitative Analysis*. The goal is for the two approaches to interact and inform each other: qualitative data will shape and instruct the quantitative component in analysis leading to more relevant results. Having this flexibility, Bukvik allows scholars a variety of tasks, such as the analysis and differential parallel comparison of translations of the same book, of an original with a translation, of corpora of two writers' work, as well as comparing texts within a language or across languages, and comparing variations from respective corpora in each language.

4 The birth of further tools (and eventually, an infrastructure)

Out of this multi-dimensionality and polarity, the understanding was emerging that we had to essentially *design and structure our own collaboration* in order to fulfill the requirements and expectations of each pole of the DH-dipole. We realized, further, that our collaboration exemplified many of the general core challenges of DH collaboration more generally, and the need to provide a more articulated and rigid framework for DH practice.

4.1 How a further tool was integrated - LitTerra

Soon we realized that for successful DH research we needed a "space" to map our research findings and provide them to other scholars. A key element of this component would be the visualization of findings that would facilitate both cross-references to the

texts analyzed and data analysis. That is how we integrated another infrastructure in our research workflow; LitTerra - an infrastructure for augmentation of texts with various digital content, founded at a similar time by the D-scholars in the project, Sasha Rudan and Lazar Kovacevic (Rudan et al, 2013; Rudan et al, 2019).

The most important consequence of such an integration lay in the understanding that there was a much wider audience for Bukvik than we were aware at the beginning. In the language of business models, we discovered additional user personae. We have also isolated research analysis (*Bukvik*) from research presentation (*LitTerra*) making it possible to extend to other texts and corpora. This allowed finalized research to co-exist with related texts and other relevant research and be made available for the end user to explore holistically. Eventually, it helped us in terms of the availability and shareability of Bukvik results.

Ongoing work with Matthew Reynolds on the Prismatic Jane Eyre project (prismaticjaneeyre.org) enforces the standardization and scalability of the Bukvik+LitTerra systems. It additionally pushes the multi-lingual and collaboration aspects as the Prismatic Jane Eyre project involves dozens of translations of "Jane Eyre" and a large community of researchers.

4.2 How new tools were integrated - Collaboration (dialogue and knowledge federation)

After a few years of practice supported with the Bukvik and partially LitTerra infrastructures, we felt the need to formalize our work and methodologies, and that is how we approached the discipline of CSCW (Computer-supported cooperative work) for answers. The first concept from CSCW we introduced in our practice were ***boundary objects*** (BO), as "spaces" of common understanding, that had a reasonably clear meaning for most of the stakeholders (Star & Griesemer, 1989; Star, 2010). In an inter-disciplinary collaboration such as this one, building a dialogue space is indispensable; without such a space, however limited, no collaboration could continue. Hence, we felt, the importance of what we have referred to above as the meta-discussion of a collaborative process one is part of, and consequently, of a theoretical basis for this discussion. To technically integrate boundary objects into our research workflow, we came to the ***Colabo.Space*** ecosystem as a part of Sasha's PhD dissertation and Sinisha Rudan's research and development, supported with Dino Karabeg's *Knowledge Federation* initiative. Colabo.Space provided the ***knowledge federation*** component of the DH-ecosystem which could natively support the concept of boundary objects together with fuzzy-knowledge and multi-truth. This helped our collaboration in the incremental development of the initial (*fuzzy*) knowledge starting from the commonly-understood concepts (expressed with the *boundary objects*).

Additionally, integrating the Colabo.Space ecosystem was intrinsically feasible as its main principle is ***puzzlebility*** (i.e. modularity, *fig. CF-example*). Thus, we could federate Bukvik and LitTerra with an instance of the Colabo.Space ecosystem adjusted to our requirements.

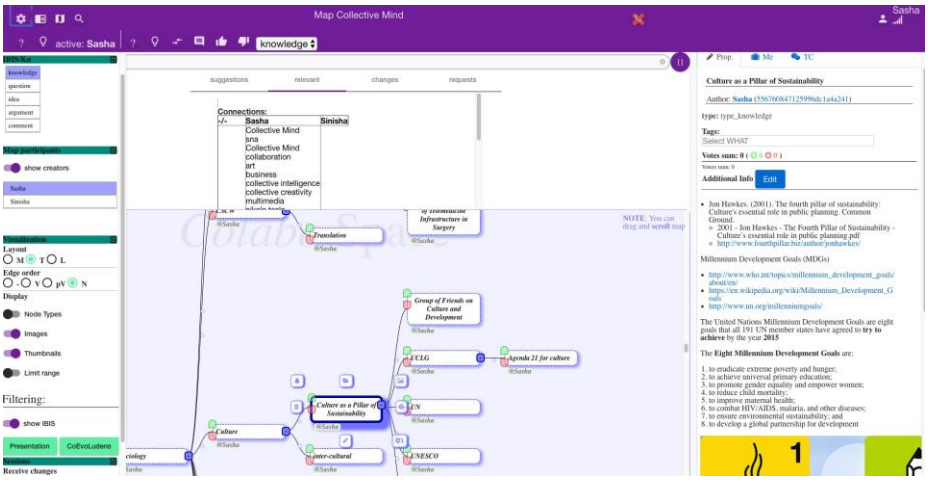

Figure CF-example: An example of the Colabo.Space ecosystem in use

4.3 Search for sustainable research evolution (ColaboDialogue)

However, we still lacked a healthy mechanism for dialogical collaboration which would organically evolve into a next round of research questions, actions and solutions.

To enable incremental evolution when it comes to the capacity for dialogical intervention through knowledge changes and actions, we needed to introduce a reflective and proactive mechanisms of dialogue and knowledge evolution—to balance the unbalanced. Unfortunately, the majority of technologies and tools used to support dialogue (including IBIS systems and Wikimedia) lack the possibility of automatic and continuous evaluation and evolution of dialogical outcomes—interpreting dialogical results and intervening either in the knowledge space or in the real-world. In other words, the sustainability of the *dialogue-knowledge-action* loop was broken.

Therefore, we embraced ***ColaboDialogue***—a concept that unites all the three spaces (dimensions), i.e. dialogical, knowledge and action spaces, into a single continuum where interactions across domains are natural, fluent and frictionless. In essence, the main or rather the most solid and long-term dimension is the ***knowledge*** dimension, which evolves continually —it represents the collective memory of our collaborative research effort. The aim of each DH community is to evolve its collective memory. That evolution can run solely across the knowledge dimension, but it can be supported by expansions into other dimensions. These expansions (based on their nature, evolution and life-time) we call **bubbles**.

A ***dialogical bubble*** bubbles out as a need to discuss an issue in the knowledge space, for example, the “insight D” in the knowledge space initiated the “bubble 1” in the dialogical space (*fig. colabo-dialogue*). At the same time, the dialogical bubble is reflective (for example “supports” reflection) on the knowledge space (as can be seen on the same *fig. colabo-dialogue*). One important feature of the multidimensionality of the

ColaboDialogue is that *the dialogical bubble lives in a separate dimension* and does not *pollute* the knowledge space. At the same time, it is strongly coupled with the knowledge space and can support, change and reflect the knowledge artifacts (Insight D, Claim A, …). After a period of time, the dialogue in the bubble *matures* and it can usually be considered as "*resolved*." Consequently, following the real-world and social model of artifact lifetime, it "*fades out*". It is important to notice that it *remains available* to enable arguing a particular knowledge evolution (decision) and avoid "*knowledge-wars*" (well known in the Wikipedia discourse).

On the other hand, a dialogue provokes (real-world) *actions* and creation of an ***action bubble*** (i.e. Question 1 → Action 1). The whole process naturally continues through interactions across domains—an action outcome can affect the knowledge space (Action 3 → Fact 3) or the (original) dialogical bubble (Action 2.2 → Idea 2). In this way, actions introduce changes into the system and provide new information that calls to be processed and understood. The ultimate goal of the process is to go the whole way back and evolve the original knowledge space.

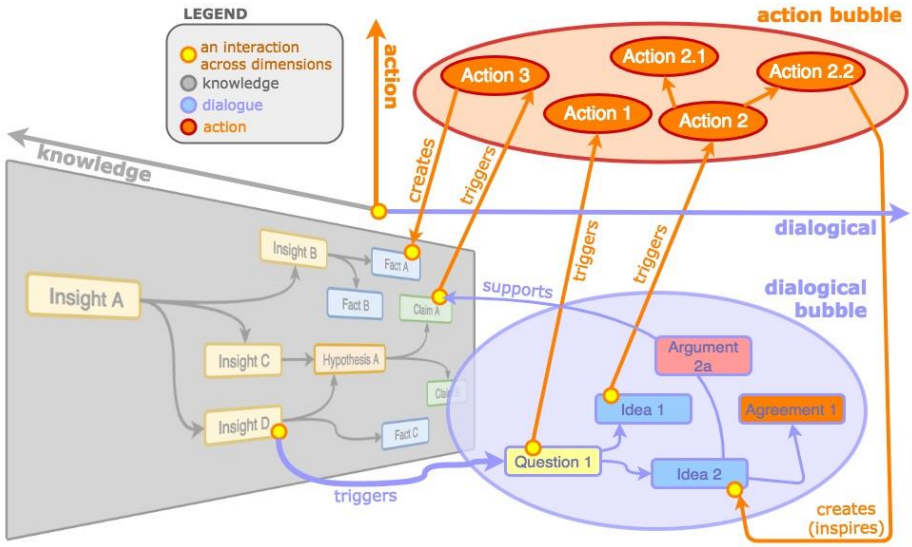

Figure colabo-dialogue: A detailed view of the three dimensions of the ColaboDialogue - an example of the dialogical and action bubble 1

As a result, dialogue does not "hang in the air," but reflects back and transforms knowledge and potentially neutralizes the tension or open question (in the knowledge space) that initiated the dialogue at the first place. We can say that *dialogue provides healing support for knowledge management*.

With ColaboDialogue, we could safely perform "*Close reading at a distance*" and integrate the *Qualitative Augmented Quantitative Analysis* research approach into our collaborative workflow.

4.4 Seeking a mutual language - ColaboFlow

Much of the existing research in Digital Humanities relies on either scholars of literature adapting their approach to existing scientific methods and tools, or computer science scholars working on literary texts. In both cases, competence is necessarily one-sided, and we have not yet come to a defined language that would allow the two competences to be orchestrated, together, to address the same problem. Our collaboration is, among other things, an experiment in establishing such a language. This brings the last key player in our research workflow: ColaboFlow, founded by the D-pole (Sasha Rudan and Sinisha Rudan). ***ColaboFlow*** is a visual language for brainstorming, designing, visualizing, and, most importantly, executing research workflows, and finally exploring and visualizing their results. It is based on an extended subset of the BPMN language. As a visual language, it became our language of collaboration, the *Lingua Franca* of DH research collaboration. *Fig. ColaboFlow* shows an example of the ColaboFlow used in the Prismatic Jane Eyre Project.

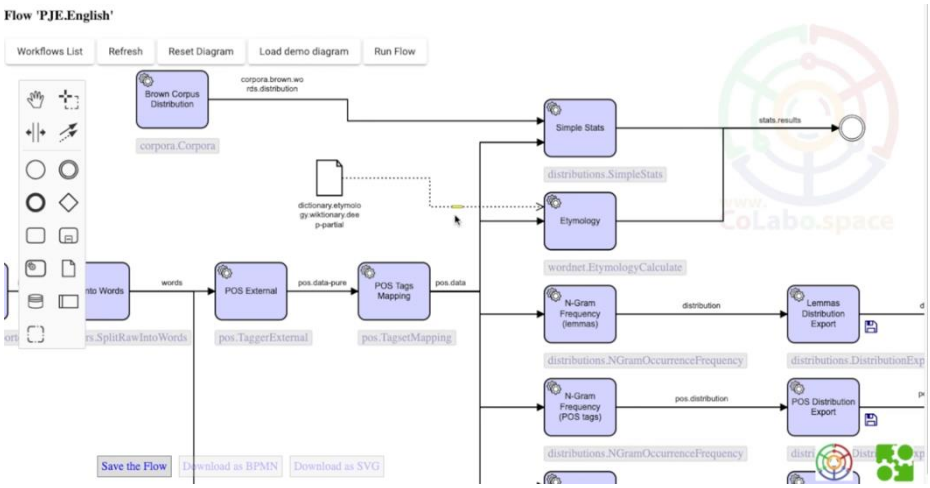

Figure ColaboFlow: An example of ColaboFlow used in the Prismatic Jane Eyre Project

The DH holistic research workflow and ecosystem presented here helped us to practice it in the open/real-world at workshops, in various projects and campaigns.

DH being a relatively young discipline (although hand-counted authorship analyses and Markov chains were demonstrated for the first time in literary analysis before digital computers were discovered), many DH scholars are H-scholar new-comers from an H-discipline (literature, history, music, art, etc). With the (*fig. DH-research-workflow*), we present a standard workflow of a DH-scholar. As one can see from this diagram, such a research flow is not that different from a similar science research flow, and reasonably different from regular humanities research (e.g. a close-reading research flow).

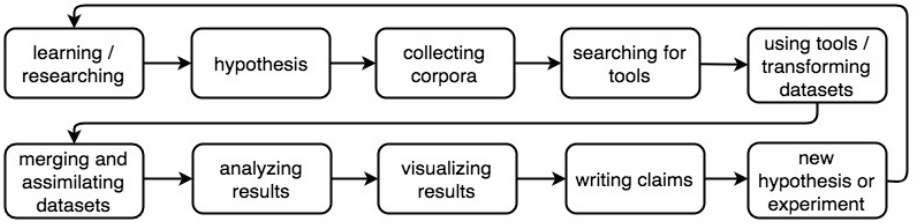

Figure DH-research-workflow DH research workflow

On the one hand, this means that a DH-scholar is often faced with unforeseen challenges. On the other hand, for the H-scholar new to DH, this field represents a new world expanding their disciplinary horizons toward new visually exciting and interactive forms of research. That being said, DH researchers may find DH research rewarding without it always being innovative or rigid in digital terms. In the case of a transdisciplinary team collaborating on a DH project, this difference will bring conflict in the way the D- and H- parts of the community work, or even in their research interests. At the (*fig. DH-research-challenges*), we present a set of common challenges H-scholar may face when they enter the DH world.

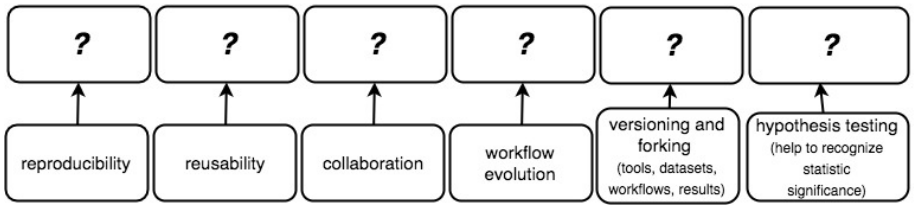

Figure DH-research-challenges: DH research challenges

It was to provide a safer environment for conducting DH research and in order to enable the dialogue across different disciplines (sub-communities) of the DH community (sometimes represented in the single DH team conducting particular research), we have designed and implemented Bukvik and evolved it into a DH-framework as presented in this paper.

5 Conclusion

From practical research questions in the domain of 1) translingual stylometry and 2) flexible corpora analysis infrastructures, we came to developing new tools, a DH infrastructure, and eventually a DH research collaboration ecosystem and meta-research questions addressing challenges of DH collaboration and its practical solutions and prototypes.[i]

There are various deadlocks a DH team can face in its lifetime, and not all such teams survive long-term due to incompatibility, losing energy or a lack of resources (financial or otherwise). A team may also not necessarily be interested in developing as a DH-dipole unit, i.e. in a search for answering new research questions using DH tools, and focus on developing and maintaining the tools they initially introduced.

All these scenarios were possible in our case, but we continued toward a successful collaboration with external partners and external grants supporting our work. Our solutions lead to new research questions to answer and a better understanding of DH challenges and possible solutions.

As already hinted in this paper, we are heading toward a DH-framework as a set of tools and methodologies that would ultimately help other DH researchers in their work, but this remains a topic for another paper.

6 Bibliography

Pfeiffer, S. I. (1981). “The Problems Facing Multidisciplinary Teams: As Perceived by Team Members.” *Psychology in the Schools*, 18 (3), 330-333.

Matthew Jockers, *Macroanalysis*, University of Illinois Press, 2013

David Hoover, ed. *Digital Literary Studies*, Routledge, 2014

Sasha Mile Rudan, Lazar Kovacevic, Eugenia Kelbert and Sinisa Rudan, (2019) “LitTerra, by augmenting literature with meaningful connections, turns readers into explorers and researchers”, ELO2019, Cork

Sasha Mile Rudan, Eugenia Kelbert, Lazar Kovacevic, Sinisa Rudan, Tamara Butigan, Miroljub Stojanovic, (2013) "Project LitTerra: New Travel Through Augmented Digital Book (the next step after digitization)," NCD'13 Conference, Belgrade, Serbia, November 2013.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology,translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. Social studies of science, 19(3), 387-420.

Leigh Star, S. (2010). This is not a boundary object: Reflections on the origin of a concept. *Science, Technology, & Human Values*, *35*(5), 601-617.
Sinisha Rudan, Sasha Rudan, Lazar Kovacevic, Eugenia Kelbert (2019), “Poetry on the Road: An Intercultural And Multidisciplinary IT-Augmented Dialogue on the Topic of Refugees and Migrants”, Comparing e/migrations: Tradition – (Post)memory – Translingualism, 2019, Toruń

Kelbert, Eugenia. "Acquiring a Second Language Literature: Patterns in Translingual Writing from Modernism to the Moderns" (PhD thesis, Yale University, 2015).

[i] The results of this collaboration and the DH infrastructures involved have been presented at several international venues, such as the selected examples below:

"Visualizing Mademoiselle O's Emigration Trajectory: a Stylometric Approach to Nabokov," Stockholm University, 2 March 2018, Stockholm, Sweden

Two workshops at the Digital Humanities in the Nordic Countries conference, 15-17 March 2016, Oslo, Norway: "Bukvik, a DH Scholar's Environment for Stylistic Analysis" and "Tools and methodologies of Collaborative and Scientifically Structured DH Research."

"Bukvik and Cross-Lingual Stylistic Comparison," Max Planck Institute for Empirical Aesthetics, 14 September 2016, Frankfurt-am-Main, Germany
"Use of Digital Humanities Techniques in the Context of (Self–)translation and Bilingual Writers," Encompassing Comparative Literature: Theory, Interpretation, Perspective, 24-26 October 2014, Belgrade, Serbia

"Visualizing Dynamics of Narrative in Fiction," 15-18 June 2015, SCALE, Malta

"Bukvik: A Literary Scholars' Environment for Running Visualized, Social-Augmented, Collaborative Research," MLA, 8-11 January 2015, Vancouver, Canada