The Whole is Greater than the Sum of its Parts Analyzing Aristotle Commentaries in Collaboration Between Philology and Data Science

Michael Krewet[1][0000−0001−7807−6089], Danah Tonne[2][0000−0001−6296−7282], Germaine Götzelmann[2][0000−0003−3974−3728], Philipp Hegel[3][0000−0001−6867−1511], and Sibylle Söring[1][0000−0002−1698−3289]

[1] Freie Universität Berlin, Germany
[2] Karlsruhe Institute of Technology, Germany
[3] Technical University of Darmstadt, Germany

Abstract. This paper aims at presenting the surplus value of collaboration between philologists and data scientists in the research on medieval digitized manuscripts. Both the great potential and the challenges of such a collaboration will be addressed. The following case study originates from research which is conducted in the Collaborative Research Center "Episteme in Motion. Transfer from the Ancient World to the Early Modern Period" which is located at the Freie Universität Berlin and funded by the German Research Foundation (DFG). One of the goals of this collaboration is to advance research questions in which the data basis is complex or too complex for traditional research methods.
The case study presented in this paper will deal with the knowledge transfer and text transmission in manuscripts of Aristotle's ancient Greek treatises on logic, the so-called *Organon*, and will focus on the manuscripts of his work *de interpretatione* (*On Interpretation*) and on commentaries and explanations which are found as paratexts in the manuscripts.

Keywords: Information Infrastructure · Collaborative Research Centers · De interpretatione.

Our cooperative work is embedded into the Collaborative Research Center 980 "Episteme in Motion. Transfer from the Ancient World to the Early Modern Period". The majority of projects within this CRC can be considered to originate from so-called small disciplines, that are disciplines with a very limited number of professors and students at an institution. Hence, a forum for interdisciplinary discussion and definition of core terminology was implemented. The variety of scholarly approaches, perspectives, and interpretations is a distinctive characteristic of this interdisciplinary CRC but also causes the need for a specifically close collaboration to ensure its success. Despite this variety, looking closer at

CEUR-WS.org/Vol-2717/paper11.pdf

the collaboration between one of the humanities projects that deals with Aristotelian works on the one side and the information infrastructure in use in the CRC on the other side will illustrate the specific requirements and challenges that this kind of collaboration generates.

1   The Humanities Research Problem

Aristotle's writings have been transmitted in manuscript form (via codices) during the Middle Ages. Today, we still possess approximately 150 copies of his logical treatise *de interpretatione*, which date from the 9th to the 16th century. They are preserved in many different libraries worldwide. Together with the manuscripts of the *Categories*, another Aristotelian treatise which belongs to his logical writings, this is the highest number of conserved manuscripts of a pagan ancient Greek text. Although the number of 150 manuscripts does only represent the manuscripts which were not lost or destroyed in the course of the centuries, the comparatively enormous number of preserved manuscripts unveils an intensive occupation with the Aristotelian logic among scholars between the 9th and 16th century [13].

Another evidence for the scholarly occupation with the treatise is found inside the manuscripts themselves. The manuscripts do not only contain the text of the treatise but also various forms of explanations on the pages, a kind of the so-called paratexts ([9] for a general overview of the paratexts in the manuscripts of the Aristotelian treatises).[4] Even the pages of the earliest manuscripts had been structured from the beginning for the purpose of adding longer explanations by leaving e.g. a wide margin. Other manuscripts also show a big space between the single lines to give room for explanations of single words (the so-called interlinear glosses).[5]

On the basis of paleographical particularities it can also be observed that the margins or interlinear spaces were often filled from different scholars or scribes at different times, sometimes over the course of centuries. All these paratexts (glosses, logical diagrams, scholia and commentaries) show that scholars have worked with these manuscripts and that these manuscripts were the basis for teaching and learning Aristotelian logic within schools or erudite circles (Examples have been given by [1]).

Both the high number of preserved manuscripts of the Aristotelian treatise *de interpretatione* and its integration in the logical instruction have fascinated

---

[4] For an overview over the manuscripts and whether they contain paratexts cf. http://pinakes.irht.cnrs.fr/notices/oeuvre/2973/ (2020–08–13) and particularly https://cagb-db.bbaw.de/register/werke.xql?cRef=Int. (2020–08–13) although both lists do not contain all preserved manuscripts of Aristotle's treatise *de interpretatione*.

[5] To mention only two manuscripts to give some impressions, it can be referred to the codices Vatican, BAV, Urb. Gr. 35 (9th century), f. 55v.: https://digi.vatlib.it/view/MSS_Urb.gr.35 (2020-08–13), Paris, BNF, Par. Gr. 1845 (end of 13th / beginning of the 14th century), f. 33v.: https://gallica.bnf.fr/ark:/12148/btv1b107218100/f70.image (2020–08–13).

scholars over centuries from a philological, philosophical and historical perspective. Particularly two research fields and challenges can be stressed: (1) The philological research concerning the question what the original Aristotelian text looked like. (2) The challenge to follow the paths in which this Aristotelian treatise has spread over the centuries. Determining the way the manuscripts travelled and detecting epistemic centers where many manuscripts were copied is equivalent with getting to know historical times and places with a particular interest in the Aristotelian logic.

The research on the first philological question has encountered significant problems and challenges which have not yet been resolved. It is a long tradition in the field of philology, particularly in the classical philology, to make endeavours to edit a text of an antique author which comes as close as possible to the original. Related manuscripts usually contain the same mistakes or the same textual variants. While copying the text scribes committed mistakes, e.g. omitting unintentionally words. Text editors have to find and eliminate such mistakes and opt for the textual variant which they consider to be the authentic one.[6]

The problem which the philologists have encountered in the case of Aristotle's work *de interpretatione* is a so-called highly contaminated transmission of the text. Contaminated means that there were scribes or scholars who did not only copy the text from one older copy, but from more than one older copy, whereby these older copies showed different variants in single passages. In such cases copying the text accompanied a philological and philosophical interpretation and examination of the text as it can be found within an erudite school or teaching context, for which the texts were also copied and annotated with paratexts. Due to the fact that such a contamination is widely spread in the manuscripts of the Aristotelian logic [8, 14] one of the greatest experts in the field of Aristotelian manuscripts studies has stated that the retracing of the relations of the manuscripts of the Aristotelian logical writings – including *de interpretatione* – is such a complex task that it cannot be solved within the life of a scholar [11, pp. 57–69].

Since scholars who have dealt with this question have come to a negative result, the detailed research data which derived e.g. from applying the philological methods, as shortly outlined above, and which had led them to their conclusion has never been published. It is simply gone with the end of the research project or the end of the scholar's life. This is the second problem which everyone faces who deals with this research question: She or he has to start again from scratch.

But also scholars conducting research on the paths in which the knowledge of Aristotelian logic spread e.g. over Europe during approximately 800 and 1600 and on epistemic centers in which the Aristotelian logic was taught and studied and in which the Aristotelian logical treatises were copied face major challenges,

---

[6] For the stemmatological method, which is also named after its founder Karl Lachmann (1793-1851) 'Lachmannsche Methode' c. f. in detail already [10]. Cf. further [6, 5, 12] Applying this method also relations between the manuscripts can be detected if they share e.g. the same mistakes. For a categorization of different mistakes which often occur while copying a text cf. [15].

because for the majority of the manuscripts we do not possess information about the exact year and the place where the manuscript was copied. Due to paleographical and codicological particularities a rough approximation to the date or region when or where the manuscript was copied becomes possible in many cases (Cf. [1] for a good approach). But for tracing the exact paths it is indispensable to find additionally, if possible, relations of manuscripts of which we do not know the exact provenience with manuscripts of which we know this exact provenience due to corresponding notes which scribes or scholars left in the manuscripts.

One method which can be applied in this case is the same which has been outlined above. If, for example, two manuscripts have to be dated in different centuries and if it is, moreover, clear that they originate from different regions (e.g. one from the Greek East, the other one from southern Italy) due to their paleographical or codicological particularities, these particularities would suggest that there is no relation between the two manuscripts. But it can become evident e.g. that the southern Italian manuscript, which has to be dated later, shares the same significant mistakes and textual variants with the older manuscript from the Greek East and contains additionally some own mistakes. Thus, it can be concluded that this later southern Italian manuscript is a copy of the earlier Greek manuscript and that the Greek manuscript was brought to southern Italy by scholars. One path of the dissemination of the Greek manuscripts can be detected in this way. If the exact place, where the southern Italian manuscript was copied, can be determined, the historical research can now examine – if further evidence can be found – whether there was an epistemic center at this place which might have been the reason why manuscripts were brought to it for being copied.

Researching these paths becomes more complicated when e.g. the southern Italian manuscripts does not only show the significant mistakes or textual variants of the older Greek manuscript but also significant mistakes or textual variants from another older e.g. Greek manuscript, whereby the two older Greek manuscripts do not mutually share their mistakes. In the case of a text transmission which is accompanied by a very erudite occupation with this text as it is the case with Aristotle's treatise *de interpretatione*, such examples suggest that an erudite scholar has copied the text from two older manuscripts. This scholar did not only want to copy a text but compared two templates with different variants and decided for the textual variants which he considered to be the authentic Aristotelian one. Moreover, the scholar can also correct the variant which he considers to be wrong in the respective template. Such a case is a culturally contextualized example of the already mentioned contamination of the text. If another scribe or scholar copies from the older copy with the correction of the scholar into a new copy, this correction cannot be noticed as a correction anymore by a reader who has only this new copy in front of him.

For the scholar today, who tries to research the path of copies for gaining knowledge about the way in which this text spread, it becomes very difficult to explain this contamination unless he succeeds in finding the original copy with the correction. To detect the origin the scholar has to find the crossing point at

which a previous scholar had two manuscripts – as in the given example – with two different textual versions. Moreover, it is possible that an erudite scholar tried to correct obvious mistakes, which he detected in the template for his copy, but that his correction originates only from his own thought. This so-called conjecture can, at the end, be different from the authentic variant and, thus, also cause contamination. And even more complicated cases can be found when an erudite scholar copied his text form more than one already contaminated older copies of the text. And, lastly, a scholar – particularly if she or he was erudite – was also able to insert paratexts such as glosses, scholia, extracts of commentaries or logical diagrams or to copy them (or some of them) from one of this witnesses into the other if she or he found them in only one of his templates. The latter fact becomes particular significant for detecting later contacts and relations between manuscripts. This potential has not been exploited yet.

To conclude: The complexity of the outlined tasks and challenges to research a textual transmission which took place in connection with a highly erudite scholarship concerning the text is enormous. The difficulty particularly consists in finding exactly these crossing points of manuscripts with different textual variants and mistakes. These difficulties and the complexity of research has raised big problems for scholars to proceed in this field of research so that it was concluded that the life of a single scholar is too short to solve it.

2 The Solution

Although the described situation poses a big challenge, it has never been doubted that significant progresses can be made by widening the research data and by extending the method of research. Regarding the first point, it is a promising approach to take into account also similarities of the paratexts in different manuscripts. If e.g. two or more manuscripts contain the same paratexts or the same false logical diagram, which have been added in the manuscript in later years or centuries compared to the copy of the text itself, this can be a strong argument that these manuscripts came together at the same place at a later date when these paratexts were added. It might seem to be paradoxical and counter-productive to further increase the complexity of the already complex research field. And it makes, indeed, only sense within an extended methodological approach which builds upon the great potential of a digital research infrastructure and digital research tools for the specific research problems outlined above.

Thus, a two-way approach is promising. Firstly, researching the relations between the manuscripts by means of the described philological method, and secondly, researching the relation between the manuscripts by means of the comparison of the paratexts in manuscripts. A digital research infrastructure can be the key for handling the described problems of contamination and complexity. The first way resembling the traditional method will now be complemented by the possibilities offered by a digital research infrastructure as it will be explained in the following. The second way, on which the research data will be enriched via including the paratexts, represents a useful supplement. There is a signif-

icant amount of cases in which relations between manuscripts as far as their later mutual contact is concerned cannot be proved via the mere comparison of significant mistakes and textual variants. On the contrary, the outcome of such textual comparisons in the narrow sense can be that these manuscripts are not related to each other because they do not share significant mistakes or textual variants. Thus, including the second way is significant to find later contacts between manuscripts and crossing points when e.g. an erudite scholar caused a contamination as explained above.

If such points of contamination of two or more manuscripts can be determined and if we also possess reliable metadata for all or at least some of these manuscripts, it becomes possible to reveal epistemic centers in which manuscripts of different origins with different textual variants circulated particularly due to a broad scholarly philological and philosophical interest in the Aristotelian logic in general and in the treatise *de interpretatione* in particular. The interest in copying the text and editing a manuscript with a reliable text which comes as close to the original as possible can be a second reason for the circulation of different copies. And the will to create the best text for a didactical or philosophical reasons can be another reason for different textual variants.

A digital research infrastructure can help to handle methodologically all mentioned problems, if it is built in a bottom up process facing concisely the research problems. As a first step a data repository for the digitized manuscripts of the Aristotelian treatise was set up with restricted access to researchers associated with the Aristotle archive at Freie Universität Berlin. This archive comprises the collection of microfilms of all known (approximately 1000 preserved) manuscripts. Digital reproductions of said microfilms in black and white, in few cases accompanied by digitized manuscripts in color, were put in the data repository. To these digital reproductions metadata files (TEI-XML)[7] were added. These files contain scholarly descriptions which particularly base on the single descriptions of every manuscript as it is found in the respective library catalogues or scholarly articles on the respective manuscript. The files mainly focus on the provenience, the date of the manuscript and the material and size of the codex. These data are complemented with research results of the current project itself concerning the relation between single manuscripts.

As a second step, several digital tools to face the research problems and challenges were newly developed or adapted. Firstly, a semiautomatic layout analysis focuses on measuring the size of a manuscript page and of the text area and margin area [2]. We know that the pages of the codices were structured in advance before the text was copied, because a lining for the text (and in many cases also for the paratexts) can be seen on the single pages.[8] Therefore, it has to be concluded that this layout was organized before the text was copied. Furthermore, it is obvious that manuscripts within a school often got the same

---

[7] http://www.tei-c.org/Guidelines/P5/ (2020–09–03).
[8] Cf. e.g. the codex Oxford, Magdalen College, Fonds principal Gr. 15, f. 1r.: https://digital.bodleian.ox.ac.uk/inquire/p/b4890d13-f697-494f-8bea-4ee3cee103d8 (2020–08–13).

layout. Thus, if an automatic layout analysis can detect the same layout in different manuscripts and if we possess e.g. the metadata of only one manuscript out of a group of manuscripts with the same or a very similar layout, we can conclude that the other manuscripts originate from the same place. These results help us to detect the provenience of a manuscript which is the prerequisite for following the path which this manuscript has traveled.

Secondly, an annotation tool for enriching the data on the digitized manuscripts in the data repository has been developed [3]. This tool enables scholars to describe, translate and – if necessary or helpful – also describe paratexts (glosses, scholia, diagrams and commentaries) as well as a semantic tagging of the paratexts. Furthermore, the transcription and semantic tagging of mistakes and significant textual variants in the Aristotelian text itself will soon be possible with this tool for scholarly use.

Thirdly, search and analyzation tools have been set up. By means of these tools it becomes possible to search and find the same glosses, scholia, diagrams or commentaries in different manuscripts – also at the backdrop of the metadata of the manuscripts, particularly relations which cannot be found by means of the traditional analog method (one example will follow).

The first evident advantage of this collaboration between philologists and data scientists compared to every single previous research in this field is that any research data will not, as it often is the case, disappear at the end of the project, but it will be stored and made accessible in a repository. Due to the application of standardized models, the reusability of the data is guaranteed. The second big advantage is that scholars from different places can research the outlined question together thanks to the digital infrastructure, making data, tools and services accessible independently of their location and operation system. Thus, competences and research power of humanities scholars interested in this question, can be pooled on an international level. A third big advantage is that all manuscripts can be aggregated via this digital infrastructure. This enables to combine results of search and analyzation tools directly with the respective digital reproductions of the manuscript pages on which the searched or analyzed items are found, have been transliterated, tagged etc. Handwritings can be compared in this way directly with each other. Scholars can see e.g. whether the search results originate from manuscripts of the same region, of a different region, of the same or different time, whether there are more similarities between the results etc.

Just to give two concrete examples: (1) Both the paratexts to an important chapter of the Aristotelian treatise and the textual variants and mistakes in the text of this chapter have been transliterated and tagged by scholars. If a scholar encounters during her or his research on the manuscripts and the paratexts among the transliterated diagrams a false logical diagram in the *de interpretatione* manuscript Vienna, ÖNB, Suppl. Gr. 67, f. 117v., it is possible to search via these tools whether a second manuscript contains the false version of the diagram. The result of the search shows us that this false diagram is also contained in two other manuscripts (Paris, BNF, Par. Gr. 1971, f. 34v. and Vatican,

BAV, Urb. Gr. 56, f. 88r.). Additionally, it can be analyzed via this tool, in the next step, if these manuscripts contain also further similar paratexts (in this case glosses, because glosses are richly found in all three manuscripts). The result of this automatic analysis shows that these three manuscripts do also contain a comparatively high degree of matches among the glosses. In a further step the scholar can test whether there are other manuscripts besides these three which share significant variants of glosses with these three manuscripts. The result of this last analysis is negative: there are no matches.

Thus, by applying the respective components of the digital infrastructure on the paratexts, the scholar concludes that there is a close relation between these manuscripts. Considering that one of these three manuscripts does not share the significant mistakes or textual variants of the other two – a result which the automatic analysis can also show us perspectively – makes the importance of this results, that we owe to the digital research infrastructure, even more evident. Since the result that the codices Vienna, ÖNB, Suppl. Gr. 67 and Paris, BNF, Par. Gr. 1971 do not share the same significant mistakes shows that the codex Vienna, ÖNB, Suppl. Gr. 67 cannot be the copy or the template of the Parisian codex, and since the paratexts in the Viennese codex have been added later in comparison to the text, due to paleographical reasons, there must have been a later contact between these codices. If a scholar goes into further detail, this context does also help to explain contamination in the codex Vienna, ÖNB, Suppl. Gr. 67 which contains textual corrections. The corrected versions are the same which are found in the codices Paris, BNF, Par. Gr. 1971 and Vatican, BAV, Urb. Gr. 56. Thus, one crossing point and point of contamination of manuscripts can be detected in this way. And to cut short: taking also into account the metadata of these manuscripts, we can follow the paths the Viennese codex has most probably travelled (from Greece to Italy) [4]. And similarly, the research can proceed in many other cases by uncovering the relations between the manuscripts.

(2) The advantages of such digital methods resulting from the intensive collaboration of humanities scholars and data scientists cannot be stressed enough, because the scholar's attention can be drawn again and again – also randomly – to relations which could hardly have been detected without this digital infrastructure and research tools. During her or his research the scholar can e.g. notice a seldom textual variant which was given as citation of the text by a commentary found at the margin of a manuscript.[9] The scholar can again search via a fuzzy search whether he finds this textual variant in another paratext or manuscript. The astonishing result shows that among other relevant results even logical diagrams can be found with a lettering which derives from this textual variant. Since the digital infrastructure also enables the scholar to synoptically compare results with the images of the diagrams on the respective manuscript pages, she

---

[9] Cf. for such a case the codex Vatican, BAV, Urb. Gr. 35, f. 55r, where a fragment of Olympiodorus' otherwise lost commentary on *de interpretatione* is transmitted in form of a scholion at the margin of the manuscript page. This fragment shows a textcritically important long variant of the definition of the noun.

or he can more or less randomly notice that this diagram does not only occur in more than one manuscript, but that two of these diagrams were drawn by the same hand, because the handwriting is identical (in the codices Milan, BA, Ambr. Q 87sup., f. 54r and Genova F VI 9, f. 68v). Thus, a contact between these two manuscripts has been proved and the contact between these two manuscripts can be further analyzed, although one might not be the copy the other.

3   The Collaboration Experience

The given example of a humanities problem and its solution with digital tools and methods may seem to be rather peculiar. Due to its potential for methodological advances it has become a leading example of our collaboration approach between humanities scholars and data scientists. However, in the scope of a Collaborative Research Center, which provides the specific academic context for the given collaboration experience, this is only part of a bigger picture.

In this CRC, although mostly traditional research is carried out and digital methods are not in the focus of many participants, more and more data are produced. For the preservation of digital results and insights, an 'information infrastructure project' (INF project) was established after four years of research activity, and needed to be integrated into an established environment. It is expected of a successful information infrastructure approach in this context to be applicable for research data of all involved projects. Basically there are two options for this: (1) providing a generic data storage solution for all projects or (2) exploring the potential of joint research. In our case a generic data repository was set up to manage the research data from the CRC in a structured way and to serve as an access point for further tool integration. But the main focus is implementing a concept of joint research and development where both sides aim to profit in regards to pushing boundaries of the respective (disciplinary) research – be it established philological methods or the application of big data technologies on heterogeneous humanities data.

Due to their heterogeneous disciplinary affiliations and research topics, it is quite challenging to focus on their specific subjects, research questions and methods, and to provide generic solutions beyond individual project scopes at the same time. The focal point of interest common for all is that of concepts of knowledge transfer. Thus, the data science collaboration also aimed for a focus on those concepts, to bridge the gap between peculiar research interest and information infrastructure but to also provide an additional point of view on those concepts. Multiple types of knowledge transfer were identified early on, enabling application and development of technical components like established metadata standards (TEI, CIDOC-CRM, etc.) or visualization techniques (spatio-temporal visualization, visualization of multidimensional relations and similarities) with high potential to become a methodological complement for as many of the projects as possible.

The entry barrier for such a close, reciprocal collaboration is certainly higher than for simple data storage. From the perspective of the humanities scholars one

of the most crucial starting points is to explain their specific research problem clearly to the data scientists and to outline what would have to be done to make progress. The data scientists need to break down logical concepts which work great in theory and apply them to the 'real world' and thus – as regards the humanities – to often messy data. Afterwards they can carefully discuss where research and infrastructure needs could meet to benefit both sides.

In the case of the CRC, four of the existing projects volunteered as 'pilot projects' for this new form of interdisciplinary research, one of them being the project on *de interpretatione*. All of the projects were already proficient in explaining their research to scholars from other fields due to the CRC's structure, but this cannot be taken for granted in general and needs to be actively encouraged. Still, in spite of this advantage, the results range from conceptual examples of possibilities to comprehensive digital enrichment of data resulting in new research insights. The reasons for this outcome are manifold and complex, just to name a few:

– In the beginning some scholars were unsure if they could 'fulfill' the requirements of an infrastructure, e.g. being able to provide a sufficient amount of data. In this case it worked quite well to start with a very small data set, show prototypes with a few functionalities early and then extend both sides iteratively. In the example of the outlined case study we have begun with a data set of only 50 out of 150 manuscripts. This procedure gave us the opportunity to train the involved humanities scholars in the specific metadata-management for manuscripts. After having adapted the annotation tool for the scholars' specific needs this limited data set opened the opportunity of integrating the tools into their daily research early on. By using the infrastructure, the scholars did not only become familiar with the tools, but also acknowledged the potential for progress in their research questions. Thus, they also started uttering wishes for further functions of the tools. Since all projects in the CRC are concerned with the research on knowledge transfer, the annotation tool in particular became of interest also for another pilot project. Thus, as a next step the generic parts of the tool could be adapted again to the special issues of this second project, also beginning with a small data set.

– On the other hand data might not be collected to the full extent initially aimed for, due to for example limited time or the need of comprehensive research work. While this is not an unusual development in qualitative research, there is a special risk to invalidate any additional quantitative research, if it cannot be performed on a complete or coherent dataset. Joint research – and here especially the data collection part – is resource intensive and tends to be underestimated. For small projects it is nearly impossible to do it on top of their regular workload. It seems particularly important that the humanities scholar plans time for the collecting and structuring of his research data. The collaboration can only work if there is a careful matching of resources and manpower on both sides.

– Even if data already exist they might not fit the requirements of the digital methods and tools in use, e.g. compliance to standards for an (semi-)automatic analysis or a lack of metadata. This is typically the case if the data were collected in advance without any intention to include a digital infrastructure into the research process afterwards. An adaption of the tools or an adaption of the data or even of both is needed if a digital analysis should become possible.

All in all one has to conclude that the effort for joint research is high. As research on all sides is an on-going process, the required communication and the need for flexibility should not be underestimated. Additionally in the case study above the humanities scholar contributed all his competence regarding paleography and codicology by collecting and adding information to the existing descriptions of the codices as well as transliterating, transcribing and tagging the Greek paratexts. As the infrastructure was still in development during that time it required a huge amount of trust on his side that it will be beneficial for his research in the end.

Our experience shows that exemplary case studies as shown are crucial for collaboration in a diverse project environment. Although a small project will probably hesitate to commit if a comprehensive example is shown, in general it is fruitful to demonstrate the full range of collaboration and research possibilities and afterwards define the specifics. While discussing the case studies within the CRC the projects developed ideas how the presented method and tools could add something to their research. Thus, data scientists can develop the most sophisticated infrastructure – it won't be used unless there are fascinating use cases and enthusiastic scholars that spark interest in colleagues.

4   Conclusions and Recommendations to Educators

This paper explores the possibilities and pitfalls of interdisciplinary collaboration in the context of a Collaborative Research Center. The presented case study on Aristotle's *de interpretatione* illustrates a concept of joint research and development between disciplines, in particular philology and data science.

In the given example our experience shows that it is advisable for the philologists to get an insight into the management of their research data, either as a part of the regular curriculum or by specific courses at the beginning of the collaboration. It is very helpful if they not only use tools, but understand some of their basics and are able to handle e.g. the XML modeling of their metadata to some extent. Such an education allows the assessment of possibilities as well as limitations while applying digital methods, thus raising awareness to foster reuse and reproducibility. Reciprocal understanding of both sides can for example help to find creative solutions as far as categorizations and vocabularies are concerned in order to avoid too much free texts if a systematic and federated search is aspired. This understanding is indispensable in order that both sides – humanists and data scientists – profit from the research and work of the

other. If this comes off well, the success also increases the motivation for further collaboration on both sides.

The CRC 980 has thrived on interdisciplinary research between various so considered 'small disciplines'. Thus, the later extension to include computer scientists into the CRC was particularly easy in this well-established interdisciplinary context. In our experience especially those small humanities disciplines can benefit tremendously from such a collaboration and larger digital infrastructure while offering unique and challenging use cases for digital approaches. All in all we argue that it is absolutely crucial to strengthen openness to other disciplines' questions, tools and methods and to advance interdisciplinary profiles – creating a research environment where the whole is truly greater than the sum of its parts.

References

1. Arnesano, D.: Aristotele in Terra d'Otranto. I Manoscritti fra XIII e XIV secolo. Segno e Testo 4 **2**(5), pp. 149-190+Tav. 1–16 (2006)
2. Busch, H., Chandna, S.: eCodicology. The Computer and the Mediaeval Library. In: Busch, H., Fischer, F., Sahle, P. (eds) Codicology and Palaeography in the Digital Age 4, pp. 3–24. Books on Demand, Norderstedt (2017)
3. Götzelmann, G., Tonne, D.: Aristoteles annotieren – Vom Handschriftendigitalisat zur qualitativ-quantitativen Analyse. In: Hastik, C., Hegel, Ph. (eds.) Bilddaten in den Digitalen Geisteswissenschaften, pp. 53–66. Harrassowitz, Wiesbaden (2020)
4. Krewet, M., Hegel, Ph.: Diagramme in Bewegung. Scholien und Glossen zu Aristoteles' de interpretatione. In: Hastik, C., Hegel, Ph. (eds.) Bilddaten in den Digitalen Geisteswissenschaften, pp. 199–216. Harrassowitz, Wiesbaden (2020)
5. Kristeller, P. O.: The Lachmann Method. Merits and Limitations. TEXT - Transactions of the Society for Textual Scholarship 1 **2**(5), pp. 11–20 (1981)
6. Maas, P.: Textkritik. Teubner, Leipzig (1927)
7. Minio-Paluello, L.: Aristotelis categoriae et liber de interpretatione. Oxford University Press, Oxford (1949)
8. Montanari, E.: La sezione linguistica del Peri Hermeneias di Aristotele. 2 Vol. Università degli Studi di Firenze, Firenze (1984)
9. Moraux, P.: Aristoteles Graecus. Die griechischen Manuskripte des Aristoteles. Erster Band: Alexandrien – London. De Gruyter, Berlin / New York (1976)
10. Pasquali, G.: Storia della tradizione e critica del testo. Le Monnier, Firenze (1934)
11. Reinsch, D.: Fragmente einer Organon-Handschrift des zehnten Jahrhunderts aus dem Katharinenkloster auf dem Berg Sinai. Philologus 145 **2**(5), pp. 57-69 (2001)
12. Sahle, P.: Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 1: Das typographische Erbe. Book on Demand, Norderstedt (2013)
13. Trizio, M.: Reading and Commeting on Aristotle. In: Kaldellis, A., Siniossoglou, N. (eds.) The Cambridge Intellectual History of Byzantium, pp. 397–412. Cambridge University Press, Cambridge (2017)
14. Weidemann, H.: Aristoteles. De interpretatione. De Gruyter, Berlin / Boston (2014)
15. West, M. L.: Textual Criticism and Editorial Technique: Applicable to Greek and Latin Texts. Teubner, Stuttgart (1973)