# FROM HANDWRITING TO FOOTPRINTING

TEXT AND HERITAGE IN THE AGE OF CLIMATE CRISIS

ANNE BAILLOT



https://www.openbookpublishers.com

©2023 Anne Baillot

This work is licensed under a Creative Commons Attribution 4.0 International license (CC BY 4.0). This license allows you to share, copy, distribute and transmit the text; to adapt the text and to make commercial use of the text providing attribution is made to the authors (but not in any way that suggests that they endorse you or your use of the work). Attribution should include the following information:





Anne Baillot, From Handwriting to Footprinting: Text and Heritage in the Age of Climate Crisis. Cambridge, UK: Open Book Publishers, 2023, https://doi.org/10.11647/0BP.0355

In order to access detailed and updated information on the license, please visit: https://doi.org/10.11647/OBP.0355#copyright.

Further details about CC BY licenses are available at http://creativecommons.org/licenses/by/4.0/.

All external links were active at the time of publication unless otherwise stated and have been archived via the Internet Archive Wayback Machine at https://archive.org/web. Digital material and resources associated with this volume are available at https://doi.org/10.11647/0BP.0355#resources.

Every effort has been made to identify and contact copyright holders and any omission or error will be corrected if notification is made to the publisher.

ISBN Paperback: 978-1-80511-087-3 ISBN Hardback: 978-1-80511-088-0 ISBN Digital (PDF): 978-1-80511-089-7

DOI: 10.11647/OBP.0355

### Cover image:

Yarn texture by Tim Mossholder (2023), https://unsplash.com/photos/AQjSAPNLjGI

Cover design: Jeevanjot Kaur Nagpal

# 2. Publishing, editing, and their digital transformation

Archiving is only one process through which text is made available for reading. It is comparatively old and relies on technologies developed several centuries ago and applied to any form of portable writing. Since Gutenberg, printing devices have made wider dissemination in a print form possible. Text has become easier to carry around, and its duplication has contributed to a major increase in text mass.<sup>1</sup>

While archiving is about preserving, recording, and making available unique documents in one place, publishing is about multiplying versions of the same textual content, which is then distributed in an array of places, to a wide variety of readers. The symbolic as well as economic value of a single printed specimen is different from that of the unique archive: it is but one exemplar among many that are all similar to one another. The economic trade-off underlying print and the dissemination of its byproducts is more complex than that of archives. One of the reasons for this complexity is that more actors are involved in the dissemination process: a writer, of course, but also, at the very least, a printer, most likely also a publisher, a typesetter, copyeditors and critics. Publishing print products has been from the outset a capitalistic activity. Books, considered here as the epitome of circulating print, have an economic value as objects of trade, and they have a cultural value. The economic dimension and the cultural one are closely intertwined.

Compared to the confines of the archive, the transformative value of publishing lies in its distributing capacity. Historically, print made it possible to duplicate and disseminate a text widely, which archiving did not. Printing, and even more so publishing, involves a range of actors and techniques and is more organically integrated in cultural practices. In

<sup>1</sup> See Mac Luhan, the gutenberg galaxy [78].

<sup>2</sup> Book historians (see for instance Barbier, Trois cents ans de librairie [21] and Wittmann, Geschichte des Deutschen Buchhandels [104]) use designations that embrace this dimension ("librairie", "Buchhandel").

the context of western cultures, at least, one generally reads more books than one comes in contact with manuscripts.

In the following chapter, I explore how publication as a practice of text dissemination adds to the use of archiving to make text accessible in a modern socio-cultural context. While the previous chapter considered text in general, in this chapter I focus more on literary texts — not only for aesthetic reasons, but mostly because they concentrate questions of authorship and make it possible to illustrate a wide array of questions pertaining to access to text.

The first section is focused on the shifts brought about by the function and figure of the publisher in the European context. I will show how publishers became key actors during the 18th and 19th centuries, essential to the development of literature as we know it today. I delve into some detailed analyses, highlighting the relevance of the relationships between writers and publishers, and shedding light on the mechanisms at the core of modern literature as a process of transforming a text into a work of art or even an œuvre. In the second section, I will bridge the gap to digitisation processes and explore what digitisation means for the type of text representations involved in publishing and archiving as two cultural practices of text dissemination and preservation.<sup>3</sup> Looking into the digitisation of modern textual heritage, I argue that the changes created by the digital medium do not concern the quality, especially the media quality, of our approach to text, as much as they concern quantity.

### 2.1 From text to book to œuvre

Traditional archiving as I have presented it is set to preserve the primary materiality or media quality of the text in question. In the case of the old family papers found in a drawer or of the postcard collection that can be purchased at a flea market, it is the manuscript that is preserved in an archive — the postcard itself, or the correspondence where the pages are in a specific order — and not simply a transcription of their textual content on another device. As already discussed, this approach has the disadvantage of exposing the unique copy to decay, and with it the disappearance of material and textual content. Another way of preserving

<sup>3</sup> From a Human-Computer Interface perspective, see also Feinberg, *Beyond Digital and Physical Objects* [49].

text content is to duplicate it, which is usually done with the purpose of augmenting its impact, especially by reaching a wider audience. In this context, preservation is paired with dissemination. The publication process (making text available to an audience, making it public) involves additional layers of editing to the archiving process. The preparation of a text for publication follows different rules, sometimes complementary to those of archiving, and involves other actors, that I now present in some more detail. These observations emanate from analyses rooted in a specific context: that of modern textuality. I will set aside all digital notions in this first historical section, turning to digitisation processes only in a second step.

I will begin with turning back again to the old papers found in your family home. I will assume the sheet bundle looks like the manuscript of a novel, and consider that the great-great-great-grand parent who wrote it not only preserved the manuscript for you to eventually find it at the bottom of the drawer, but for a wider audience to read it. Let us assume that your hypothetical ancestor would have wanted to publish this novel.

Following the description I gave in section 1.2.1, I would say that here I am considering texts originally produced by a particular hand (maybe helped by other hands, such as a copyist's), and intended for a wider audience. Reaching this goal is facilitated by printing techniques that transform the medium of the text in such a way that it becomes easier to access for a larger readership. The transformative value thus achieved is reflected in the marketplace at large (what is the price of a book? what is its relevance in the concerned economic structure? what is the number of books necessary to achieve such a relevance?): a book has a place in a state's economy or even in a global one. On a cultural level, it affects reputation mechanisms that contribute to establishing a hierarchy in the types of texts that are circulating at a given period in a given area.

Such a hierarchical approach prevails in literary studies. A random text has no specific value on this scale until it has established its literary market value. A work (of art) in the form of a printed and distributed book deserves more reverence than a simple text, as it marks a greater

<sup>4</sup> What I am basing the following argumentation on is a very general presentation. While bringing complex mechanisms down to a generic description bears the risk of over-simplification, it facilitates an interdisciplinary perspective that encompasses a variety of aspects, as Darnton himself argues in his seminal *What is the History of Books?* [39].

degree of achievement and materialises a well-identified economic value and cultural capital. An œuvre consists of several works by an author that, considered together, have achieved a more superior status. In the context of an œuvre, even lesser texts (drafts manuscripts, correspondence — textual material with no primary market value) gain new significance.

Before the golden era of digital self-publishing, writers had seldom been in a socio-economic position from which they could themselves supply publications of their texts, by presenting them in such a way that a wider audience would have had access to them in exchange for payment. Writers in search of fame and money could hardly rely on a business model of that type. Publishing requires funds and skills in order to provide a context favourable to dissemination and to selling a literary work on a larger scale. This is where the publisher comes into play. Considering the complex relationships between writers and publishers also provides a shift of perspective in the history of literary text production.<sup>5</sup>

I will first analyse the mechanisms at play in the relationships between a writer and the actors that actively intervene in shaping a book for publication, focusing mainly on the publisher. Then I will then present two case studies based on early 19th-century German literature before I turn to questions concerning the digitisation of published textual heritage.

# 2.1.1 The deal between writer and publisher

Published texts whose superior literary quality is recognised are considered as works of art, and, taken as a whole, as lifetime œuvres. Not every text is part of an œuvre, but every literary œuvre is at its core composed of texts. The mechanisms underlying the transformation from text to book to œuvre, and the role of media transformation in book form, are not trivial. In the same manner that I worked out different stages of text constitution in section 1.2.1, I will now delineate the different stages in the evolution of a text transforming into an œuvre in an early modern context. The two temporalities of the evolving manuscript and the nascent book and their different stages, as we shall see, are partly intertwined.

Print applies to a range of formats: posters, journals, pamphlets, chapbooks, books, and more. Although books have not been the primary

<sup>5</sup> See also, in the perspective of social history, Schmidt, Die Selbstorganisation des Sozialsystems Literatur im 18. Jahrhundert [95].

print format to disseminate even literary content for a particularly long time,<sup>6</sup> I will talk of "books" here, not because I care only for books in the strict sense, but because the notion of a book encapsulates the cultural capital that is at stake in the media transformation from manuscript to disseminated print, especially in the case of literature.<sup>7</sup>

The process of publication can be schematically described as follows: a writer conceives and writes a text, then entrusts it to the publisher who brings this textual content to book form, duplicates it and hands it over to readers in exchange for money (via distributors such as booksellers), a part of which is then paid to the writer.<sup>8</sup>

In the first step of this process then, the writer conceives and writes down the text. As I have shown in section 1.2.1, realising this first step can take some time and effort. It is not always achieved at once, but more often than not requires drafting, re-writing, and a first copying phase that applies solely to the manuscript, not to the print version ("copy" has different meaning in the publication process, depending on whether it is applied to a manuscript or a print)9. You can still consider it as one step, one creative phase that includes all events happening between the moment when the writer has their idea for the text they want to write and them sending a text out to a publisher: the draft, the first clean version, copies, and edited copies are part of this process, which has a high potential for alterations of all sorts between its beginning (the idea) and its end (sending out a copy to the publisher). Some authors prefer to dispose of the drafts altogether, others write clean drafts, others dictate them. There is not one unique way to write; there are rather almost as many ways as there are writers.

<sup>6</sup> The printed press, with its feuilleton format for novels (in newspapers or magazines), was much more affordable than books and hence more popular. Only when the price of books could be significantly lowered, and literary publishers strove to reach a wider audience with high literature, did books become more popular. In Germany, the role of early 20th century *Kulturverleger* such as Samuel Fischer was key in that process. See Kuhbandner, *Unternehmer* [73].

<sup>7</sup> I am thus following the discipline concerned with the dissemination of print culture, which is called book history.

<sup>8</sup> The schema proposed by Darnton includes more actors [39]. I have chosen to simplify a fundamentally complex and changing set of relations, in order to provide general orientation. It does not account for historical and cultural variations and is strongly rooted in late 18th-century mechanisms when the publishing profession established itself.

<sup>9</sup> A copy of a manuscript is usually supplied by hand; a copy of a print is another print.

Whatever the preferred procedure, providing text is seldom a solitary activity, even before it is sent out to the publisher. It can happen that the writer does not disclose their writing activity to anyone else. <sup>10</sup> But I will assume that other people are involved (family or friends as test readers, for instance), as is usually the case. It can even happen that the publisher is already part of this process already. In the case of book projects by authors who are already in contact with the publisher with whom they wish to work on a specific book project, authors and publishers can discuss structure or content — even if it is only the number of pages envisioned — which is itself bound to have an impact on the final text. There might even be yet another person involved. Texts the publisher receives may have been copied by the author personally, or by a copyist. In the ongoing negotiation between writer and publisher regarding content, the intervention of the copyist provides room for assigning this third party responsibility in delicate cases.

A second step in the realisation of the print version of the text follows this first one of conception and redaction. On this second step, the writer has conceived and written the text, and now entrusts it to the publisher in order for them to bring this content into book form. This step too involves not just one, but a series of actions. The text received by the publisher is read and edited, at the very least in terms of typographical errors and obvious mistakes, possibly even with more substantial interventions. Additionally, the transformation process from a writer's copy into a typeset print template involves decisions regarding the font that will be chosen and its size, type of paper, and book format: all things necessary to generate a first printed version of the text. The publisher can be supported in this decision-making process by various professionals: printers, copyeditors, typesetters. Depending on the requirements in each case, copyeditors may check mainly for typos, or their intervention may extend to much more consequential parts of the text. Moral and/or political censorship remained the rule for a long period of time in the modern

The weight of social conventions is not to be underestimated there. Depending on period and social context, it could (and still can) be dangerous to be outed as a writer. Many women who published anonymously in contexts where it was frowned upon for people of their sex were likely to send their manuscripts directly to publishers, without showing them to anyone beforehand, even more so than men who were struggling to make a living as writers and had to produce texts under so much pressure that there was no time left for other input.

era and framed the copyediting process; I will come back to the question of censorship in some more detail below. Here, let me simply say that there is obviously already a likelihood that the publisher will need to get back to the writer once these interventions are realised. The typesetter's intervention is definitely one that leads to discussions with the writer too, if the writer is in a position to intervene at this point. Writers usually have precise expectations when it comes to the form of their work: is the font elegant enough, the paper white enough, the binding (where there is one) soft enough? Many a material question is to be discussed in the process of shaping a book.

It is technically possible for the publisher to not consult at all with the writer, but this is rarely the case in the next step of the publishing process, which delivers the first version of the book, the so-called proofs. This first printed version of the text is provided for the author to check and correct. Proofs are printed in only one copy, with the sole goal of checking both printing techniques (layout, typesetting, etc.) and content (correct reproduction of the writer's handwritten or typewritten copy from step one). In general, at least one round of proofreading is planned for each publication, sometimes more. Carefully checking proofs is a delicate process requiring authors to immerse again in a text they might have entrusted to the publisher months before. If the first proofreading raises major issues, such as the insertion of several pages in the wrong place in the book, or anything that would massively disrupt the structure of the printed book, then a second, improved iteration of proofs has to be proofread as well.

A publisher's greatest wish, when it comes to proofs, is for the author to make as few modifications as possible. Even a change in a syllable or a word can lead to a shift in the page that will disrupt the layout, the placement of chapter beginnings on uneven pages, for example. When an author re-writes a text completely, based on the proofs, the publisher may have to start the entire typesetting and printing process all over again. This costs not only time, but also money: the publisher then needs to spend considerably more resources in the book-making process. Every modification is costly, and to go over several sets of proofs has a notable

<sup>11</sup> Some famous authors are known for completely redrafting proofs, to their publisher's despair. See for instance this digital version of the proofs of Baudelaire's *Flowers of Evil*: https://gallica.bnf.fr/ark:/12148/btv1b86108314/f23.item.

financial impact. In other words, this third step also provides room for negotiation of a balance between text accuracy and financial cost.

Even in this simplified description of the publishing process, you can see that the publisher's role is not limited to spotting typos. The publisher has a range of opportunities to contribute in-depth to the form and content of what they publish — room for intervention that can be interpreted as a form of authority over the text. Authors from the 19th century, such as Heine, have claimed that publishers' interventions equate to censorship. 12 In this case, the context was rather tense both in terms of the political situation and of the critical content of the writings. But more generally, if a publisher prevents a writer from using a specific, controversial wording, it can pass for censorship, whether it is officially branded as such or not. However, if an author does not even use a specific wording in the first place because they know that it would be frowned upon, and instead submit a manuscript to the publisher that avoids controversial wording altogether, should that be called self-censorship? Where does adjustment stop and censoring begin? It is in fact extremely tempting to brand as censorship any external element that leads to a modification of the text under some kind of constraint — but then, from this perspective, a large part of the production of a text would be considered to be modelled by censorship.

A text is always created in a specific socio-political context, with its moral, and often religious constraints. I prefer to use the term "censorship" in cases where an external organ is active in controlling the content of what is published. In 18th- and 19th-century German literature, this concerns mainly political censorship, but also religious censorship. Both church and state had appointed personnel to monitor the enforcement of the rules they had imposed on behaviour but also publications. Actual censors were commissioned, and publishers had to submit to them everything they intended to publish. This opened the door to a parallel, clandestine book market that would not be subject to official censorship rules. Such literature followed slightly different publications mechanisms. <sup>14</sup>

<sup>12</sup> On the complex relationships between Heine and his publisher Campe, and the way Campe navigated between official censorship and Heine's sense of what he could print, see Ziegler, *Julius Campe* [105].

On censorships mechanisms, see Kiesel and Münch, Gesellschaft und Literatur [72].

<sup>14</sup> For more information on this topic, see the numerous excellent publications by Robert Darnton, for instance his *Censors at Work* [40].

There is still one final step to mention in the process of negotiating text between writer and publisher, which relates to the final duplication of the text for readers in return for money, a proportion going to the author. While the actual quality of the book accounts generally for its success, it is also true that this success depends on the reputation of both publisher and author. If an author is already well-known, their production will attract buyers. Similarly, a famous publisher will have an established audience. An author's name or a publisher's name can serve as a selling point.

In the case of an established author, it is the publisher's mission to capitalise on their existing reputation. Print distribution will have to be well measured, contact with booksellers optimised, and advertisements targeted, in order to increase the benefit in relation to what the publisher invested into the production of the book.

If one looks at the way a book presents itself to its readers, author and publisher are traditionally both involved. When I open a book, what I see on the title page is the author's name, the book title, maybe a place and date, but also, even if it is in smaller print and somewhat lower on the page, the name of the publishing house. This is the face of the book, delivering its identity to its audience.

Now to the distribution process itself. Readers are essential to this final step in the book production. Readers (or rather, what author and publisher imagine them to be) are the horizon of the whole negotiation journey in each of the previous steps. The concept author and publisher have of the book's audience becomes vital in the distribution phase. It can well be that they have very different expectations, but in order for them to come to an agreement, a minimal overlap in their respective sense of what the reception of the book should look like will be essential.

In his History of Reading, Alberto Manguel writes:

Almost everywhere, the community of readers has an ambiguous reputation that comes from its acquired authority and perceived power.<sup>15</sup>

The focus is on the respect due to readers, a reverence that can be explained by the fact that the dynamics of any audience is unpredictable.

<sup>15</sup> Manguel, A History of Reading [80], p. 35.

The success of a book depends on audience reaction, and that reaction can only be channelled to a certain extent. Some authors consider that readers do not understand their work (mostly in cases where they have had limited success), and they insist on readers being wrong and the book being right. This measure of self-protection does not always make them want to publish less, however. But then again, the financial incentive of publishing books does not come from people reading books, but from people buying them.

This is, however, only the pecuniary point of view. In terms of reputation (and financial value feeds on reputation), the goal is not solely that readers buy a book, but that they encourage others to do so in order to multiply sales. A good reader in these terms is someone who makes others want to read too. Reviews, for instance in journals, can be a major asset in disseminating the reputation of a book and encouraging more people to buy it, but hearsay might be even more efficient. And literary critics oftentimes have a very different opinion from that of a wider audience.

Financial success and literary reputation do not always work in perfect combination. Commercial literature can attract a wider audience than high quality literature even if, in the end, it is the high literature that will pass into literary history and keep being read beyond its author's lifetime.

There is not one simple unequivocal way to define good literature. Is it that which receives good reviews, that which is awarded prizes, that which sells well and provides its author financial security — or that which will still sell 200 years later? These recognition processes are long and complex. Trying to disentangle them sheds light on how arbitrary judgements can become authority arguments if they align with socio-cultural values of a specific time and place.

In all stages of the preparation of a text for publication, intellectual production and financial issues are intertwined to the point that they cannot be considered independently from one another. Author and publisher play interconnected roles from the creation to the dissemination of a book. Depending on the period and the situation of the literary market, their relationships operate differently in the process of transforming a

In Germany, the so-called canon debate is a topic of scholarly, didactic and even public interest, regularly discussed in the press. For a systematic approach, see for instance Freise, *Literaturwissenschaftliche Theorien und Modelle der Kanonbildung* [54].

text into a book, and ultimately an œuvre. To some extent, working out a literary history based on the relationships between writers and publishers would also mean writing a new history of the book market. And it would shift focuses in traditional literary history, as I would like to show to close this section.

This narrative has its starting point in the late 18th century. Before that period, the book market was based on a barter trade that involved, apart from writers, printers and booksellers. The profession of publisher — the one person who incarnates the interface between printer and bookseller — emerged in the late 18th century together with the professionalisation of the writer job. Professionals who acted as publishers, that is, those who started investing capital in publications, were still often called printers or booksellers in the late 18th century. The emergence of the profession of publisher does not mean that all the other professions involved in the production of prints were at once strictly separated from each other. In fact, you could find publishers who were also writers, writers who had copyeditor jobs, copyeditors who translated and printed, publishers who were also booksellers, etc.

In the late 18th century, the expansion of the book market and its financial potential led to various abuses. Printing technologies had become cheaper, literacy was on the raise, and the book market promised to become an interesting source of income for professionals in the printing branch. All scholars were affected by this evolution. In his 1797 *Metaphysical Foundations*, Kant took a look at the book as a cultural artefact of primary relevance. In this text, he focuses on the book as a medium of transmission of text and positions first the writer (or author), then the publisher as actors in its transmission to an audience. His argumentation revolves around the materiality of the book itself insofar as it embodies the text. Three key actors are involved, and are situated in relation to this

<sup>17</sup> See Wittmann, Geschichte des Buchhandels for this and the following historical overview [104].

See Immanuel Kant, Metaphysische Anfangsgründe, 1797: "Was ist ein Buch? Ein Buch ist eine Schrift (ob mit der Feder oder durch Typen, auf wenig oder viel Blättern verzeichnet, ist hier gleichgültig), welche eine Rede vorstellt, die jemand durch sichtbare Sprachzeichen an das Publikum hält. – Der, welcher zu diesem in seinem eigenen Namen spricht, heißt der Schriftsteller (autor). Der, welcher durch eine Schrift im Namen eines anderen (des Autors) öffentlich redet, ist der Verleger. Dieser, wenn er es mit jenes seiner Erlaubnis tut, ist der rechtmäßige; tut er es aber ohne dieselbe, der unrechtmäßige Verleger, d.i. der Nachdrucker", quoted from Gelehrsamkeit ein Handwerk? [92], p. 230.

material artefact. The writer is the one-time producer of the languagebased signs that make up the text; the one who presents this text to a wider audience in the name of the writer is the publisher in the case where the duplication is legal, the reproducer in the case where it is illegal. Kant points here to a major debate of the late 18th century, which eventually led to the first (albeit comparatively late) copyright rulings in the German-speaking area. In the absence of an international legislation on reproduction rights for intellectual works, several printers set themselves up to reproduce books without having any kind of agreement with their authors. In this case, the author did not receive any financial compensation for the printer to reproduce their work. This dysfunctional modus operandi was partially corrected in some German states and regulated in others at the beginning of the 19th century, which contributed to establishing the reputation of specific states and cities as publishing hubs based on their legislations and how respectful they were of writers' financial and creative aspirations.

Questions of reputation have a different impact for authors and for publishers, which also depends on the type of book that is being published. My focus here is on what is called high literature, although I know very well that it made up for only a small part of the book market in the German-speaking area of the late 18th to early 19th century. Literature, just like any other published work, is dependent on the conditions of the book market in general. But, maybe more than any other domain in the book market, high literature is one that involves strong identification mechanisms with the published work, at any rate more than might be the case with more technical productions, such as household-related guidebooks that were a much more widely disseminated genre than high literature at the time. Strong identification with the text also led to more emotional responses to the money issues at stake in the process of producing a book. This was true for both writers and publishers, and added to the complexity of their relationship.

How authors and publishers interact can vary. The agreement between the two parties can be more or less formalised: in late 18th-century German-speaking areas, most authors, and most publishers, operated without any form of written contract. This left the question of the shared responsibilities open to some extent, and could lead to an unresolved situation between author and publisher. Who is in charge of deciding

on typesetting, on the number of copies, even on the wording, if this is not settled in a contract? The work, in the form of a book, is a common endeavour of both writer and publisher, requiring acknowledgement of what they share and how they share it.

Even in the cases where a formal contract exists, it does not necessarily remove all ambiguities, if only because things change with time. Contracts could very well deal with not simply the production of one book, but several or even all of them. If a publisher wanted to keep working with an author (and vice versa), specific agreements could bind them for more than one publication project. Hence the long-time associations of authors and publishers that become an intellectual, cultural, and economic tandem— and sometimes also the estrangement that becomes gossip for the cultural elite.

If the reputation or success of one of the agreement partners evolves differently from that of the other, an initially stable and clear relationship can be disrupted. But it can also be that the emotional investment of the partners varies over time. A particularly close relationship between a publisher and an author could be qualified as a "friendship" by either one or both of them at a certain point in their relationship. This choice of word to designate a work agreement involving money negotiation and transactions opens the door to a wide range of potential misunderstandings.<sup>19</sup>

Does that mean that the relationship between author and publisher consists of so much more than business that it cannot be embraced by a contract? Therein lies essentially the contradiction. Reciprocal trust is at the core of their relationship, and the business model itself requires both actors to identify themselves with each other: the author has to feel at home in the publishing house in order to be able to entrust their text to the publisher; the publisher has to see the text rooted in the publishing house's profile at the very least, if not in their own intellectual identity, for it to be published.<sup>20</sup> This leads to a sense of reciprocal identification effects, but also to possible conflicts of interest, disputes on competence areas, and more. In a way, authors have to separate from their text and its uniqueness in the moment when they agree to a contract, be it virtual or

<sup>19</sup> See Fischer, Merkwürdige Verbindung [52].

<sup>20</sup> In *Der deutsche Buchhandel* [67], Hiller says, in that sense, "Autor und Verleger leben in einer Art Symbiose" — author and publisher live in some sort of symbiosis (p. 77).

real, with a publisher. At this point, the publisher takes possession of the text to some extent. The artefact at the core of the transaction is charged with emotionality, not only symbolically, but almost essentially. As a result, the relationship between author and publisher is impregnated by authorship issues that no contract, however extensive it may be, can fully resolve.

Imagining a history of literature that focuses on the relationships between writers and publishers would make it possible to assess authorial processes in the production of text in the form of a book as a work of art, and in the process of creating an œuvre. It is also a way to consider the media transformation of the text as a process that fully determines how readers will access it. The publication process I have described above, although presented as a general framework, best applies to (early) modern literature. In order to make the transformative impact of the relationships between author and publisher at text level easier to grasp, I next elaborate on a few examples from the German context . From there, I will then move to the media transformation involved by digitisation processes and how they, too, generate shifts in access to text and its constitution.

# 2.1.2 Negotiating the œuvre

At first sight, relationships between authors and publishers appear to be a collaborative process that can be framed and regulated in a productive manner. But productivity comes at a price, and it is worth looking behind the curtain when it comes to transforming literary texts into valuable works and ultimately into a recognised œuvre. My two examples from the late 18th and early 19th century, Goethe and Tieck, shed light on different strategies.

Johann Wolfgang von Goethe and Ludwig Tieck followed different career paths, but they share a few biographical features. They are both considered major German writers, and they lived well into old age, Goethe died to the age of 83, Tieck 80, yet troubled for most of their life by a hypochondriac fear of dying. Their longevity did not bring reputation, money, and happiness alone: they both lost many dear friends in the course of their existence. In 1805, Goethe lost in Schiller his *alter ego*, and survived him for 28 years. In 1830, near the end of his life, the death of his

son August deprived him of a strong asset in the management of his late publications, too. Tieck lost his friend Wackenroder in his youth, then his young-adulthood friend Novalis, and later his mature-adulthood friend Solger, who all died at a relatively young age. These losses affected their view on life, and their vision of their own work.

Their strategies with publishers were, however, quite different, largely because of their respective work ethics and ability to capitalise on past literary successes. Tieck played different publishers off against one another, while one of Goethe's strategies led him to focus ultimately on a single publisher — the most powerful one of his time, Johann Friedrich Cotta. Book historian Reinhart Wittmann considers Goethe as exemplary for the way in which relationships between authors and publishers evolved at the beginning of the 19th century, and I will draw on his argumentation in what follows.<sup>21</sup>

Goethe's literary career starts in 1765, when a bookseller from his hometown Frankfurt takes him to the Leipzig book fair (already then an institution, and still one today) and introduces him to publisher Philipp Erasmus Reich. A first volume of poems by the young artist was published anonymously in 1769, but Reich refused to print Goethe's next manuscript. Goethe then produced his next works by self-publishing them, including his epic drama *Götz von Berlichingen* in 1773, which turned out to be a success and was immediately reproduced in unauthorised copies by numerous publishers. These reproductions were sold without bringing the author any money. Nonetheless, he had succeeded in gaining traction within the literary arena. For his next opus, an epistolary novel entitled *The Sorrows of Young Werther*, he found a publisher and made enough profit to wipe out the debt left by his self-publishing phase.

From this moment on, he resolved to systematically request very high fees from his publisher for every new work he wanted to publish, without even showing them the manuscript. In addition, he negotiated with publishers via a third party: it was someone else (a person he trusted) who actually contacted publishers and negotiated the conditions dictated by Goethe. At this point in his life, he did not technically need the income from his publishing activity as he had been made a minister in Weimar. His hardcore negotiation tactics were less a struggle for survival than a

<sup>21</sup> See Wittmann, Geschichte des deutschen Buchhandels [104], pp. 175-185.

symbolic gesture of revenge for the humiliation he had endured in his first years as a young writer.

Upon returning from a transforming trip to Italy in the late 1780s, Goethe tried to publish his first scientific work, the Metamorphosis of Plants, with publisher Georg Joachim Göschen, who had commissioned the first complete edition of his literary works. Göschen refused in somewhat crude terms.<sup>22</sup> It does not matter to me, Göschen said, whether Goethe wrote the book: a shopkeeper cannot be a philanthropist. Göschen's commercialism was hardly compatible with Goethe's intellectual aspirations, and even less with his self-esteem. A subsequent round of negotiations was undertaken with Johann Friedrich Unger, who had published a few of Goethe's books, but with so many printing mistakes that Goethe doubled his author fees. In order to compensate for his financial loss, Unger procured a parallel print (accordingly exempt of author fees), which he did not mention to Goethe. But Goethe noticed and refused to continue working with Unger. At the same time, Goethe had been discussing a much higher fee for his play Hermann and Dorothea, which was ultimately profitable to publisher Johann Friedrich Vieweg because he kept printing and selling it for years after the standard two-year contract had elapsed. This cat and mouse game is symptomatic of two circumstances: first, the lack of clear rules; second, the fact that both parties, author and publisher, were trying to negotiate financial advantage in relationships in which trust conflicted with the understanding each had of their own legitimate claims.

A glance at the discussions between Schiller and Goethe related to the publication of the journal *Die Horen* in the 1790s sheds light on cooperation with yet another publisher, Johann Friedrich Cotta. <sup>23</sup> The journal *Die Horen* was edited by Schiller and published by Cotta between 1794 and 1797, and became a major publication outlet of German Classicism. In a letter to Goethe from September 1794, Schiller underlines Cotta's "zeal and decisiveness", "tirelessly promoting the journal", and his "'punc-

<sup>22 &</sup>quot;Ob ein Goethe das Buch geschrieben hat, ob es die höchste Geisteskraft erfordert hat, darauf kann ich als Kaufmann keine Rücksicht nehmen. Ein Krämer kann kein Mäcen sein", quoted by Wittmann, Geschichte des deutschen Buchhandels [104], p. 179.

<sup>23</sup> Ironically, the correspondence between Schiller and Goethe, in which they extensively discussed negotiation tactics with Cotta, was published by Goethe with Cotta's publishing house a few decades later — magisterial proof of Cotta's ability to anchor his professional reputation on factors other than personal considerations.

tuality in delivering author and editor fees".<sup>24</sup> Schiller's praise of his publisher — aimed at convincing Goethe to contribute to the journal — was not only motivated by Cotta's qualities as a publisher, but also by the fact that Schiller needed the regular income provided by Cotta in exchange for the publication of the journal.

While it was Schiller's initial idea to publish contributions anonymously, Cotta insisted on having them signed by their authors, which would be a valuable selling point, considering these were Goethe, Schiller, Humboldt, Fichte — by then already known actors in the intellectual scene. Goethe refused categorically, pointing out that anonymity was the only way for him to remain free to write what he wanted.<sup>25</sup> Schiller and Cotta had to give up on their selling argument for the sake of Goethe's freedom.

Later discussions display a similar imbalance. Schiller writes to Goethe that Cotta wishes for more diverse, less abstract content. Goethe's answer disregards the publisher's concerns and focuses solely on the relationship between author and audience. Let us proceed on our path — we know what we can deliver, writes Goethe, adding: I have known the farce of German authors for twenty years now, it just has to be played. The show must go on, so to speak: accommodating this perspective as well as Cotta's was an acrobatic manoeuvre that required all of Schiller's diplomatic and editorial skills. The show must go on, so to speak that required all of Schiller's diplomatic and editorial skills.

Around 1800, Goethe engaged in a long-lasting personal work relationship with Cotta.<sup>28</sup> It was Schiller who had motivated Cotta to try and publish Goethe's works. Schiller states that Goethe is too unique to have a standard market value: and since he is priceless, any price will

<sup>24 &</sup>quot;Eifer und Entschlossenheit"; "unermüdete Tätigkeit in Verbreitung des Journals"; 'Pünktlichkeit im Bezahlen"; see Schiller-Goethe-Briefwechsel [96], p. 50.

<sup>25</sup> Goethe writes: "Cotta mag recht haben, daß er Namen verlangt; er kennt das Publikum, das mehr auf den Stempel als den Gehalt sieht. Ich will daher den übrigen Mitarbeitern die Entscheidung wegen ihrer Beiträge völlig überlassen haben, nur was die meinigen betrifft, muß ich bitten, daß sie sämtlich anonym erscheinen; dadurch wird mir ganz allein möglich, mit Freiheit und Laune, bei meinen übrigen Verhältnissen, an Ihrem Journale teilnehmen zu können"; Schiller-Goethe-Briefwechsel [96], p. 73.

<sup>26</sup> See Schiller-Goethe-Briefwechsel [96], pp. 103-104.

<sup>27 &</sup>quot;Lassen Sie uns nur unsern Gang unverrückt fortgehen; wir wissen, was wir geben können und wen wir vor uns haben. Ich kenne das Possenspiel des deutschen Autorwesens schon zwanzig Jahre in- und auswendig; es muß nur fortgespielt werden, weiter ist dabei nichts zu sagen."; Schiller-Goethe-Briefwechsel [96], p. 105.

<sup>28</sup> For more detail on the following analysis, see Fischer, *Cotta* [50], especially the chapter "Karlsbad und die Folgen".

be worth paying  $^{29}$  — the argument could not be less oblique compared to Göschen's. Cotta's reaction shows a surprising display of restraint ("ich war zu schüchtern, in dieser Hinsicht etwas zu erwähnen" — I was too shy to suggest anything), explicitly asking Schiller to intervene in the negotiation as the middleman. In this first negotiation step, Cotta also points to the fact that this connection should not be a one-off. The extent of Cotta's dedication to his star author was clear. Publishing Goethe's works turned out not to be profitable at all for Cotta. Goethe demanded extremely high fees and never really trusted Cotta. Evidently, Cotta deemed such a treatment tolerable in return for the prestige of calling himself Goethe's publisher.

After Schiller's death in 1805, negotiations between Cotta and Goethe became more complicated without Schiller's instrumental intervention, and ended up extremely troubled when it came to the production of the final edition of Goethe's complete works, the *Ausgabe letzter Hand*. The 1820s had seen the rise of new highly regarded publishers for works of literature, especially Georg Andreas Reimer and Friedrich Arnold Brockhaus, with whom Goethe had engaged in negotiations even though he was still under contract to Cotta. Goethe finally turned to Cotta in 1823 to discuss the ultimate edition of his œuvre, however. When Cotta answered with a request for additional information, this hurt Goethe's feelings.<sup>31</sup> What is more, the offer Cotta made for the 40-volume edition was lower than those the other publishers had already made. Goethe was irritated and did not react to his publisher's offer. In the absence of the now deceased Schiller as a middleman, Cotta commissioned art collector Sulpiz von Boisserée to enquire about Goethe's state of mind.

After Boisserée managed to sort things out and re-establish a constructive relationship between Goethe and Cotta, Goethe's son August (who was blinded by the — technically unrealistic — offers made by other

<sup>29 &</sup>quot;Ein Mann wie Goethe, der in Jahrhunderten kaum einmal lebt, ist eine zu kostbare Akquisition, als daß man ihn nicht, um welchen Preis es auch sey, erkaufen sollte", quoted from Wittmann, Geschichte des deutschen Buchhandels [104], p. 181.

<sup>30 &</sup>quot;Ich hege freilich immer den stolzen Wunsch, daß ein angefangenes Verhältnis der Art nie getrennt werden möchte, und ich werde daher immerhin das möglichste tun, es zu erhalten und diejenigen, die sich mit mir in solche Verbindung einlassen, es nie bereuen zu machen." — I have the proud wish that this bond shall never be broken, and I will do all that is in my power to maintain it: Goethe quoted from Fischer, Cotta [50], p. 113.

<sup>31</sup> Fischer, Cotta [50], p. 668.

publishers) came across an edition of his father's *Faust* by Cotta, which the publisher had not mentioned to the author and for which Goethe had accordingly received no fees.<sup>32</sup> Cotta's argument that the rights he had acquired for the complete works included the right to reproduce single works that were part of them further worsened Goethe's mood. Taking advantage of the weaker position into which his son had manoeuvred Cotta, Goethe managed to negotiate even more profitable conditions, and the contract for the *Ausgabe letzter Hand* between Goethe and Cotta was signed in 1826.

This brief glimpse into a rich and complex context demonstrates that even in the case of major actors in the literary field, whose survival, life or career did not depend on the outcome of the negotiation, work relationships hang by a thread, and trust remains the answer.

The case of Tieck presents a somewhat different situation, but leads to a similar conclusion. Tieck started his literary career by providing contributions for Enlightenment publisher Friedrich Nicolai's satirical collection, *Die Straußfedern* in 1794. Tieck was 21, and he continued working for Nicolai until 1798. The texts were published anonymously, without any title, as a series of numbered contributions. While the publisher was by then already famous, the writer was not only unknown to the public, but writing mostly on commission. He also translated French and English prose, which he did with the help of his sister Sophie and her husband, August Ferdinand Bernhardi. While Nicolai was well aware of young Tieck's literary talent, Tieck himself did not adhere to the satirical style of late Enlightenment à *la* Nicolai and this led to tensions. This period was one of the most productive in Tieck's life, however. He integrated these early texts into later editions of his works — subjecting them to a renewed editing each time, but keeping the basic stock.

Turning his back on Nicolai proved profitable for Tieck's literary ambitions. They ceased to work together after a quarrel regarding the publication rights of one of Tieck's works — a quarrel that became public, ensuring that the young author made a name for himself on the literary scene. Now he was young, famous, and a champion of true poetic inspiration. This prestige called for much higher fees than those he had

<sup>32</sup> Fischer, Cotta [50], p. 671.

<sup>33</sup> For the biographical part of this analysis, I draw on Roger Paulin's standard work *Ludwig Tieck. Eine literarische Biographie* [86] as well as on the chapter on "Tieck und seine Verleger" by Philipp Böttcher in the *Tieck-Handbuch* [27], pp. 148-164.

received from Nicolai. Tieck moved to Jena where he became part of the famous Romantic circle.

It was Friedrich Schlegel, his Jena roommate, who introduced him to publisher Georg Andreas Reimer in 1802 in the context of an edition of Novalis' works. Reimer had been active in Berlin as a bookseller since 1800, and selected his publications in relation to his political orientation, which, in the context of the Napoleonic wars, was decidedly German and patriotic. Reimer published both Novalis' works (which sold very well) and Tieck's (which sold modestly). At this point, Tieck had started repeating the same procedure he had experimented with when working with Nicolai: he asked for credit, again and again, for books that either did not sell well or which he never even delivered.

In the 1810s,<sup>34</sup> Tieck published his famous *Phantasus*, a collection of youth writings, comprehensively edited and framed with a narrative in the style of Boccaccio's *Decamerone*. He was financially supported by a patron — a former schoolmate who ensured his and his family's subsistence. Tieck kept asking for high fees from Reimer, but Reimer had seen through him and found a strategy by which he would not suffer excessive losses. Reimer always paid Tieck a little bit more of the advance fee he had asked for, in order to trick him into feeling guilty and actually writing and sending the promised texts. But on the other hand, he only paid him the fees due on sold volumes with lengthy delay, even years of deferral. All in all, Reimer did not engage in a highly risky business with Tieck.

Tieck started to notice the late fee delivery, and for several years, their correspondence stopped. Reimer kept sending money, but the reciprocal trust was damaged and, in the 1820s, two other publishers set out to win over Tieck and make him their house author: Josef Max and Friedrich Arnold Brockhaus. Tieck had by then settled in Dresden with his enlarged family, where he enjoyed a new creative phase. He tried to take advantage of the concurrence situation among the three publishers. In a first step, Reimer remained in charge: he took over the contract for Tieck's complete works (signed in 1827) but, in order to prevent further losses if Tieck did not deliver, he included the right to publish, five years after their first

<sup>34</sup> On this period in Tieck's production, see Baillot, "Das Bild Shakespeares" [5] as well as Jochen Strobel in the chapter on "Dresden, Berlin und Potsdam" in the *Tieck-Handbuch* [97], pp. 108-111.

publication, all works by Tieck that had appeared in other publishing houses. Further, Tieck was to pay 100 thaler per month per manuscript for each one he failed to deliver on time. For the Shakespeare translation, Reimer opted for the opposite solution, and offered an additional 1500 thaler per volume delivered on time.

In 1838, Tieck wanted to break away from Reimer. This was extremely complicated as the contract was binding, tying him firmly to the publisher. He turned to the Brockhaus publishing house: from 1826, Tieck published novellas in their pocket book series, *Urania*. He received rather high fees for this work, but Brockhaus did not complain about the cost. He printed in small font, forcing Tieck to deliver more text content per page than was usual and in addition, the format was pre-defined and the author was compelled to stick to it. Tieck complained bitterly, saying it was "barbaric" not to be able to decide for himself where textual content would be positioned on line and page.<sup>35</sup>

The real crux in the relationship between Tieck and Brockhaus turned out to be Tieck's book collection.<sup>36</sup> Although he constantly lacked money, Tieck had managed to gather an impressive library of books over the years. But in 1839, he needed money so badly that he sold it to Brockhaus. It was really more a lease than a sale, as Brockhaus had agreed that Tieck could keep his books and receive a yearly amount of money in exchange for Brockhaus' right to dispose of the collection after Tieck's death.

As advantageous as the agreement was — especially as Brockhaus did not have any major financial plans for the book collection, which he only intended to eventually make available for research and teaching — Tieck still sold his book collection a second time, in 1849, to auctioneer Adolf Asher. Scandal could only be diverted with the help of a friend who bought Tieck's books back from Brockhaus in order for Asher to be able to proceed with the auction. Brockhaus' gesture of offering such a generous lease, an exceptionally friendly offer, was not honoured by Tieck. Here, again, trust appears to be a most fragile device.

Goethe and Tieck are only two examples, and from a time when there was little legal regulation available to frame the business relationship

<sup>35</sup> See Böttcher, Tieck und seine Verleger [27], p. 160.

This discussion relies strongly on the chapter "Tiecks Bibliothek" by Achim Hölter in the *Tieck-Handbuch* [68]. See also the digital edition of the catalogue of Tieck's book collection at https://tieck-bibliothek.univie.ac.at/.

between authors and publishers. The loose legal framework of the early 19th century sharpened identification with the published work on the author's as well was on the publisher's side. These examples also show how unproductive it is to assume that there is only one authoritative version of any one work. Both Goethe and Tieck produced several editions of their works. The publishers influenced the output, but it was also authors themselves who kept re-drafting their works — based, inevitably, on the critical assessment of previously printed versions.

The relationships between writers and publishers shed a singular light on the history of literature. They do not tell a story of definitive versions and of authors' final words, but one of common writing processes. Manuscripts of literary works can be better understood when considered in the context of the correspondence between writers and their first reader circle, but also between writers and their publishers, as well as such commissioners as Boisserée in the case of Goethe's *Ausgabe letzter Hand*. All these contributing hands shaped the text into the work, then into the œuvre that we know today — because at the present point in the history of western literature, it is mostly famous works and famous authors that we read about.

But one could also look at this from the point of view of the history of print and book. What novel insights could be gained there, drawn from what material? One would not only need the correspondence between writers and their friends and colleagues to understand the shape in which prints have reached us. One would also need to systematically study book fair catalogues, censored drafts, account books of publishing houses, printers' correspondences, and accounts of overarching institutions such as the *Börsenverein des Deutschen Buchhandels* in the German context.<sup>37</sup> One would need publishers' correspondence to be fully integrated into the analysis of the creative process, maybe even that of their spouses, who often contributed a great deal to the family business. This is studied as a branch of history, but not in an encompassing manner as a branch of literary history.

The Börsenverein is an association that was founded in 1825 to secure the interests of publishers in the book market, and remains an active actor to this day. A wealth of scholarly works have been dedicated to its history; a short overview is provided on their website at https://www.boersenverein.de/boersenverein/ueber-uns/geschichte/.

The history of literature does not provide the tools to address this material, partly for pragmatical reasons. Archives of publishing houses are prone to destruction. On the one hand, it is a business in which the destruction of material is a question of survival: given a finite room to stock material, there is only so much that can be kept and, in general, it makes more sense to stock material that can be sold (such as recently produced journals and books) than what cannot (as is the case with archival material). On the other hand, especially for larger, established publishing houses willing to keep traces of their history, this destruction process is not always encouraged or facilitated. But publishing houses have been exposed to the most precarious conditions during the 20th century: they, and with them their archives, were specially targeted by shelling during both world wars.<sup>38</sup> Their strategic value was recognised, which endangered them in critical political contexts.

Literary critics at the time of publication too could be integrated much more explicitly into the transmission of literary texts than it is the case today. Considering that writers often redraft previous versions over time, the chances are that vocal critics, or particularly convincing ones, may have an impact on the editorial evolution of a text. Admittedly, such interventions are not easy to grasp. They require an excellent knowledge of the period, of its means of communication, and of influences: this is what frames literary quarrels and is where sensitive spots can be found. Critics often carry some responsibility in the posterity of a literary work, and hence contribute to the construction of literary history at large. The relationships between publishers and critics are another field in which one would be likely to find traces of the evolution of texts on their way to becoming œuvres, such as letters they exchange, and mandates they undertake. As I mentioned in section 1.2.2, the reception of a text is also part of a text genesis, just as text genesis is a reaction to the pre-existing reception of other texts or works. This intertextuality is at play in a variety of textual traces, and inherently questions the notion of a unique version of a text, as well as that of the unique authority over it. Placing the relationships between writers and publishers under closer inspection is one way to shed light on the many hands at work in a text, even when the writer is a famous, established author, and even without taking into

<sup>38</sup> This is the case for the Insel Verlag; see Kuhbandner, *Unternehmer* [73] and Ziegler, *Buchfrauen* [106].

account all the ancillary work done at home by those who surround him (more seldom: her), and whose traces are even more volatile and difficult to grasp.

This information, gathered by scholars from various disciplines and with various interests in textuality, is brought to the attention of general readers through two major outlets. The first one is school material that provides an initial approach to literary works and to literary history, with varying levels of precision depending on the teaching level. Goethe's poems can be studied early in a school curriculum. In fact, both Goethe and Tieck wrote short poems that provide a simple access to poetry and that are used in classrooms, making them basics in German literature. Another essential way of channelling the reception of literary works are the editions of their works. Once authors are dead, these editions can be undertaken by family members, initiated by publishers, and/or be commissioned as scholarly endeavours. Editorial tradition is as old as print techniques, even older than them. Preserving and transmitting texts from Antiquity has been ongoing across centuries. It took on a new dimension in the early 19th century, when choices of modern authors to edit (who? how?) were not to be reduced to philosophical issues (should we follow Aristotle or Plato?), but steered by political orientations. Editorial traditions of the 19th century evolved differently even within Europe. Some emphasise text genesis, others context, yet others engage with the definition of a unique version of reference. Our knowledge of literature and literary history owes its key orientations to European editorial practices of the 19th and 20th centuries. This is where digitisation found us: struggling to cope with editorial traditions and to make sense of endeavours that had become scholarly or economic behemoths — 20 years or more for a team to edit the œuvre of one author in books so expensive that almost nobody can afford them.

In the next section, I will argue that the shifts brought about by digital media make it possible to reassess these orientations, and leads to a new convergence of archiving and publication practices.

# 2.2 Text quality from scan to digital edition

Much has happened between 1830 and the early 21st century. Political and social evolutions, shifts in economic leverage instruments and actors and,

with the development of new forms of media, also novel ways to engage with readership.<sup>39</sup> I will not undertake a history of editorial practices since 1830, but rather take a big leap to examine what happens with the relationships between authors and publishers, and the hierarchy between text, work and œuvre, once text is made available online. In terms of textual materiality, I will remain focused on those literary works that are considered relevant heritage, as they tend to be included in digitisation programs and illustrate my argument incisively.

In the following section, I consider the way in which the shift to a digital environment modifies representations of text and discuss to what extent this leads to renewed forms of its reception. Does the fact that anyone can post anything online lead to a complete disintegration of the literary canon? Does the multiplicity of the forms of presentation available mean that we have stopped reading linearly and do not rely on the page structure anymore? I argue that the media change to digital formats has not modified our reading habits as drastically as we may think, and certainly not as much as might have been possible considering the extent of the structural change that has taken place. In the final section of this chapter, I discuss models of online access to text and explain why Open Access is susceptible to major leverage in the renewal of access to text, and more specifically to cultural heritage material. I begin by unrolling once again the process that has been presented so far: that of the constitution, editing, and publishing of text, but this time, in a digital setting.

### 2.2.1 Emancipation from the page?

For someone coming from traditional text or book studies, or even from the heritage domain, assessing online textuality is nothing short of intuitive. I first consider the different types of text that can be found online and how they can be transformed into genuinely textual formats. My goal is to delineate as clearly as possible why formats matter and what impact they have on access to text in general. Understanding the different representations of text available online is the initial step to acknowledging (and eventually applying) quality criteria to digital text.

<sup>39</sup> See Mac Luhan's seminal *Understanding Media* [79], 1964.

While archives, libraries, academies, and publishers, to name only some of the key actors involved in text dissemination processes, were attributed roles with respective boundaries that crystallised over the past decades and centuries, the borders between these players and their activities are somewhat blurred in the digital context. For a layperson browsing the web, it is not easy to distinguish the difference between a webpage that is set up by a random individual, a site that contains valuable archived material, and a literary work of art. In order to understand the mechanisms that regroup such varied textual material under the overarching roof of digitality, I will consider it not only from the point of view of the user (how do I, as a reader, assess the type of material I am consulting?), but also from the point of view of a producer (how do I, as someone who wants to make textual material available online, transform text into a digital archive, or into a digital edition?). <sup>40</sup>

In the physical world — or rather in the analog world, since the digital is itself material — text presents itself on a page. It might be a manuscript page, a postcard, a typewritten document: as long as the page contains letters or even just signs, it can be considered a text. This is a piece of paper, a page on which legible characters are drawn or written, and form to some extent a coherent unit of sense.

In a digital setting, the same textual document can be accessed either in an image format (the scan of the manuscript), as a transcription of its content (the raw text), or in an annotated format with underlying information that is readable by the machine without being necessarily being visible to the human eye. Annotations can be added to images as well as raw text. An example of such an annotation would be, for instance, if abbreviations are expanded and additional information is supplied. As a reader, you would then know who the person mentioned in the manuscript is when the manuscript simply says "Mr R.", or in what century a text was written when the manuscript date reads "7 June 36". The additional information, whether names, dates or something else, is part of the metadata of the document concerned.

An image, a raw text, and a somewhat enriched (annotated) image or text are the most basic representations of text in a digital context. But

<sup>40</sup> On the functional overlaps between reader, user and producer in the context of the web, see Rob Wilkie, *The Digital Condition* [102] and Elizabeth Bird, *Are We All Produsers Now?* [24].

one can also engage in the production of other, more complex representations of text.<sup>41</sup> In the case of literary works, statistical analysis of word occurrences can be calculated based on raw or annotated text and a visual representation of the result can be generated based on this calculation, for instance, in the form of graphs. The output, graph or otherwise, is then a new way of representing the text, one that includes the result of statistical exploration or analysis. Word clouds, displaying the words that are the most important statistically in a text in a font size relative to how often they occur in that text, are such a (rather simple) representation.

Following this method of calculating connections between elements of a text, you can represent networks of figures in a theatre play so that the relative importance of the relationships between them is represented visually. Such graphs are made available online, sometimes very prominently on a webpage, as a way for the reader to gain an overall visual impression, leading them, in a way, on a pre-determined path. <sup>42</sup> These representations of text require an additional level of abstraction as they are the result of visual choices, which are themselves derived from mathematical calculations based on the source text. But they can be considered as representations of text as well.

Yet, the difficulty lies not in considering that specific forms of visualisation are representations of text, but in knowing how to deal with them in terms of interpretation. How do you read a graph? And what does it tell you about the plot, the style, and the characters in a text? In fact, digital representations of text require different skills from their reader than analog representations do. This statement will not come as a surprise in the case of visualisations so complex that they require skills in statistical analysis, but it is to some extent true for all types of digital representations of text.

When you have a traditional, analog page in front of you, be it handwritten or printed, you have been taught in which direction to usually read it. You have been trained in school since first grade to do precisely that. For most readers directing their attention to languages they know, it is as though the instructions have been inscribed on the page since

<sup>41</sup> See Baillot and Lassner, Von Graphen [18].

<sup>42</sup> See the homepage of the digital edition of August Wilhelm Schlegel's Correspondence (https://august-wilhelm-schlegel.de/briefedigital/), or looking for instance into Goethes Faust on the dedicated page of the DraCor platform: https://dracor.org/ger/goethe-faust-in-urspruenglicher-gestalt.

they have been familiarised with deciphering them since their childhood. They know the code. The content will make sense if they read the series of characters in a specific order that was taught to them in school and has been reinforced in their everyday environment ever since.

Things are different when it comes to digital text. There are several forms of digital text representations for which literacy taught in school will not suffice for literate readers to generate meaning. I will examine this more closely, starting with what is maybe the most important step in the production of digital text: the shift from an image format to a text format. This is a process in which machines have been intensely trained in the past decade. Computers have become better and better at recognising letters from images of printed pages (optical character recognition, OCR), or even from manuscripts (handwritten text recognition, HTR). They are trained to identify first what makes up a line, then to duplicate the same line structure in a text format for the output. Within the line blocks, they then isolate groups of characters, which generally form words and, within these units, their next job is to identify, on the one hand, the shape of the characters and match it with existing characters, and on the other hand the combinations of characters that are possible. For instance, if the computer's character recognition system identifies that a unit that looks like a word and is composed of something that looks like four letters, with the first one being an "r" and the last one being a "d", the computer will look up a dictionary where frequent words are listed (ideally, a list tailored to the context of the text concerned, which suggests high probabilities of usage for the vocabulary specific to this text) and indicate a textual transcription that would be, for instance, the word "read", with a high probability of occurrence if the concerned text is a primary school manual in this case.

The process of generating a complete transcription of a text based on an image requires the image to be of extremely high quality and the computer to have been programmed and trained for the specific typography or handwriting that is being used. It also needs tailored dictionaries that facilitate statistical assumptions for word probabilities. Even with this technical apparatus, the output will not look like what is generally considered a text. There will be mistakes, words that are not recognised, diacritics or punctuation marks that will be identified as letters and vice versa. The textual output will need to be improved, in

the end, manually by a human, in order to comply with the standards of what is considered a legible text. Generating legible text from a digital source requires special skills, just as deciphering old manuscripts does.

The process of automated text recognition<sup>43</sup> is in many ways exemplary of the necessity of setting up modes of collaboration between humans and computers in order to gain legibility. It is also symptomatic of some misunderstandings that may arise along the way, which is why I will now look more precisely into the process of acquiring text from an image.

The scan of a text page is a visual representation of this text. Unless it has been either programmed to recognise the characters or informed regarding the textual meaning of the image it is displaying, a machine is unable to derive text from the image of a text. This also has consequences at metadata level: if the image of a text is not connected to textual information (be it the wording in textual form or information about this text in the form of metadata), it cannot be found on the internet unless one knows its precise URL. An image without comments is invisible in the digital context. While it might be too tedious to fully transcribe the content of a scanned page, it requires only a little effort to describe it in a few words: "this is an image of a page in this or that script, written by this person in that year, dealing with topic so and so, and which can be found in this place". With these few accompanying words, the image is provided with metadata, and it can be found online.

Now let us assume that the scan of a text can be found online thanks to its metadata. Many readers will say the image suffices as long as they can read it. They will consider that it is legible if the image resolution is of a good enough quality. Poor reprint editions are doing nothing else than taking for granted the fact that a good enough image of a page is a text, a published text even. Readers, deeming themselves happy with the online scan, can then either read the scan cursively online, download it onto their device (in order to read it on an e-reader, for instance), or even print it. In that way, they will have been able to gain access to a text that may not have been within their physical reach, if, for instance, it is preserved only in an archive or a distant library. In terms of access to text, the reader has gained a lot.

<sup>43</sup> Automatic text recognition, ATR, includes both OCR and HTR.

But the procedure is still problematic in two ways. First, not any random metadata will make it possible to search the web with success for the text the reader is looking for. How is it that by searching for the title of the book, sometimes even an erroneous form of it, you still manage to find it? Because the underlying information is structured — with both positive and negative consequences. A whole economy is based on promoting specific internet resources as a result of user requests. The algorithms used by major companies active in this area have been programmed to recognise, analyse and filter metadata in such a way that users keep seeing their products, with web browsers generating information that will in turn foster the economic value of the information offered. While this primarily concerns other types of (cultural) consumables, even domains like text dissemination have fallen prey to it.

Access to quality text is a particular challenge because quality is not the primary criterion for the companies that decide what it is that your browser-based search will lead you to. Once you have found a text, you cannot easily know whether it is reliable or not. And is this text really what the metadata tells you that it is? Sometimes the result of an internet search appears to have very little to do with what you are actually looking for. The unsorted wealth of textual material that can be found online requires proper training for the reader or user to recognise online text quality, especially considering the current lack of standardised ways to display who the author is, what the title is, in what context the content has been created, and whether it is acceptable to reuse it for whatever purpose. This information can be retrieved in an online library or archive catalogue, but not on every webpage. There are initiatives to make these catalogues more visible, and to direct readers towards them when scrolling the web for information. But making them even more visible would require a massive investment, similar to that which, for instance, the institutionalisation of archives must have necessitated in their time — which was not solely an issue of time, money and space, but also of political will.

I would like to go into more detail about what makes for quality in a digital text. Findability is one of the key criteria. Legibility by a human reader is another, as is its legibility by a computer. If a human reader is able to read the poor result of an OCR output, for instance, they may be able to improve it, and in that manner make it accessible to even more

readers. This means that improvability of the text that is made available to readers is another quality criterion.

This is one major difference between born-digital and digitised texts. From the traditional perspective on a work of literature, one would identify a specific iteration as the authoritative one: the one and only version of reference. Compared to this version, there can be only lesser versions or variants. A born-digital text, on the other hand, can very well have several versions. These can be identified with a version marker (in the sense of the master versions I mentioned in section 1.1.2) and a time stamp. The time stamp is not temporally unidirectional: it does not assume that a later version is necessarily better than an earlier one. But it makes it possible to identify versions unambiguously and distinguish versions generated at different moments in time from one another. Applied to the inherent history of the text that I presented in section 1.2.1, it means that it is possible to preserve different layers of the constitution of one text, with metadata providing documentation on the status of these different layers. Digital editions of textual heritage can take advantage of this plasticity to elaborate a more dynamic concept and form of text.

There is yet another quality that is required from a digital text: it should be connectible. The power of hyperlinks lies in their ability to foster discoveries not simply from one website to another, but also from one specific element of a webpage to another digital resource. A website containing many stable hyperlinks is a door to a wealth of other resources. The best-known way to do so in the most common browser language (HTML) is to have a hyperlink underlying a word or a group of words: if you click on the highlighted word or group of words, you are redirected to the resource corresponding to the underlying hyperlink. But this is not by any means the only way to connect two digital resources with one another. There are other types of connections from one text to another text, some requiring the use of other computer languages.

There is more to this question of the digital connection of text-based resources (text-based here meaning that there is always text involved, either at data or at metadata level, or both), and it is worth a closer inspection for several reasons. One of these is that it shows the type of effort that can be undertaken to balance out mechanisms based on the economic value of information exchange and dominated by big tech companies.

Some digital information is not connected via simple hyperlinks, but through an elaborate system of unique identifiers that are being used by a variety of actors of the web in such a manner that they serve as standards, either because they are official standards<sup>44</sup> or because they are widely used and have become de facto standards. Cities for instance, can be identified via geo-coordinates. If your great-great-great-grand parent mentions Paris in the old papers you found in a drawer of the family home, how can this be linked to additional digital information? Geo-coordinates can serve as a pivotal information for the machine to connect information from one specific webpage with, for instance, a platform like Wikidata, which centralises basic information. 45 For this information to be recognised by the machine (and interpreted as: "this word is Paris, meaning the city of Paris, France"), it matters to integrate geo-coordinates as underlying information or metadata to the related passage in the text in this case, the word "Paris" in the original text, and to do it in a standardised manner that will be legible to the computer and make it possible to generate a connection to a central platform like Wikidata.

If the old papers in the drawer are not the manuscript of a novel, but something more likely, such as letters or ledgers, the chances are that not only places will be mentioned, but also people: friends, family, and public figures at the time of writing. In the case of persons that are mentioned in a text, the connection to information is even more refined. Librarians all over the world have joined forces in creating a colossal online catalogue in which writers, and to some extent also other personalities in the book market such as publishers and critics, are each assigned a unique identification number. This resource is growing by the day and is less and less limited to book market-related figures, encompassing more and more historical figures in general. The goal of this huge index is for all digital resources to be able to use a unique identification number when referring to a specific historical person, facilitating the connection of available information about this person. Digital resources that use this person identification system, such as libraries or digital editions, can register on

<sup>44</sup> Like ISO Standards, see https://www.iso.org/standards.html.

<sup>45</sup> See https://www.wikidata.org/wiki/Wikidata:Main\_Page.

<sup>46</sup> For a more technical description of the following, as well as potential applications, see Baillot and Busch, *Vernetzung* [13].

another Wikipedia-based service called Beacon, which connects all of the registered users through the uniquer identifier, every time they use it — and so on and so forth for every single occurrence of every single person mentioned and tagged with a common unique identifier.<sup>47</sup> In return, each resource can automatically point to the other resources that use the same system. With the Wikipedia Beacon, biographical data can be aggregated while keeping a precise track of their origin.

Catalogues containing biographical information about the people involved in the production of books (writers, publishers, printers, translators, illustrators and critics) were initially curated at a national level. This valuable information stock is now merged into an international digital catalogue. 48 Today, we can resort to a world-wide standardised resource identifying writers — meaning here people that are relevant to cultural heritage institutions and involved in writing processes in one way or another — and more and more other (known and lesser known) historical figures. Of course, things are more complicated in practice than they may seem on the paper. There now are so many data providers that the catalogue contains many doublets. Quality control is more complicated on a world-wide scale than when dealing with a small institution's index of persons. Additionally, more and more non-librarians and non-archivists wish to contribute information, to generate unique identifiers themselves when they gather information about interesting figures of the past, especially scholars working on earlier periods for which it is necessary to process archives in order to gain novel biographical information. Their input is not always as standardised as that of professional librarians and can generate some confusion.

But even as imperfect as its implementation at a global scale may be, the benefit of such a standardised approach cannot be emphasised enough. As a reader, one does not need to engage in intensive research on each person that is mentioned in a given historical context, one simply has to use the identification number and rely on linked resources. The remarkable level of connectivity this entails makes it possible for digital texts to lead from one information source to the next: they draw a path

<sup>47</sup> See https://de.wikipedia.org/wiki/Wikipedia:BEACON. See also Baillot, Krux [9] on methodological challenges to the work with biographical data for network analysis in textual studies, and for a less historical approach including a use case for the Wikipedia Beacon, Baillot, Weisheit der Massen [7].

<sup>48</sup> See https://viaf.org/.

through the web for the reader to follow, in the footsteps of a writer or another historical figure they want to discover, even if this person has been dead for centuries. With the original resources linked as entries for each identification number, it is always possible to fact-check any given piece of information by clicking back to its source.

The large-scale realisation of this ideal web of knowledge will only be possible, though, if textual resources are also made available online in such a way that they are accessible for the reader. <sup>49</sup> This is best illustrated by the example of the book scan I mentioned earlier, assuming that it is only available in an image format, accompanied by minimal metadata that connects biographical information on the author to the scan of the document. In such a configuration, readers can bridge the gap between biographical information and text, but they cannot do much with the text itself, apart from reading it from the beginning to end. If there is a raw machine-based transcription underlying the image, readers can search for specific passages based on keywords or phrases, and jump to the parts of the scan that are likely to be of interest to them, without having to read the book full to find the relevant passages. OCRd books can also feature hyperlinks that connect the table of contents with the corresponding chapters, allowing the reader to move swiftly to the passages they are most interested in.

But, based on an image of a text, the possibility of connecting it with other digital resources remains limited. An image contains no information that can be interpreted by the machine in order to generate a connection with another text. To reach a better permeability between textual resources, it is important that underlying machine-readable information is added to them. This information has to be structured in such a way that the machine is able to interpret which part of the text it refers to (the whole text, a paragraph, or a single word), what type of information it provides (for instance external input, semantic information, linguistic structure, details on text genesis), and whether that information should be matched with external resources, as was the case for the geo-coordinates or the referenced persons I mentioned above. This should be done in a computer language compatible with those used in

In the following, I present the technical aspects of this accessibility issue, while the economical and philosophical dimensions are addressed in section 2.2.2.

other digital resources with which one wants to generate a connection or a link.

The ability for programming languages to communicate and interact with one another can be roughly subsumed under the umbrella concept of interoperability. There are, of course, different ways for digital resources to interoperate, at different levels, and not all interoperability leverages work the same way. 50 What matters here is the general idea that it is possible to conceive digital texts in such a way that the information contained in one can be integrated with another, that they can enrich each other reciprocally. This supposes, of course, that authors authorise the reuse of the text that they have produced. In the current European legal framework, copyright rules favour reuse of older data (by authors that have been dead for at least more or less 70 years, depending on national jurisprudences); living authors are required to state reuse conditions explicitly. But for my purposes I will assume that reuse is possible, and that there is no reason why someone would not want others to use the information either written a long time ago or provided online with a reuse authorisation, in the form of a license.

With these constraints in mind, what can you do with the old papers you found in a drawer in the family home? The local archive does not want to have them because they are not old enough, or relevant enough, or maybe you do not even want your local archive to have them. You think that these old papers are incredibly interesting; for whatever reason you might have, it matters to you that they are preserved and made available to a wider audience. What do you do in order to make this happen on your own? It is entirely possible to do it with everyday equipment and simple technical skills.

Step one, you scan the old papers and add metadata to the scans, even if only a file name and a date. Next, you need a repository where you store your images: it can be a hard drive or a cloud server, for instance. Then you need a way to give access to the information housed on your repository. At this point, you cannot do without some use of computer languages. The easiest way to provide access is to use a simple content management system (like Wordpress), which avoids confronting readers unfamiliar with computer language with the code underlying the editing

<sup>50</sup> See for instance the subtleties introduced by Syd Bauman as early as 2011 in his seminal *Balisage* paper [23].

process of a webpage. You will have to check regularly on your content management system, though, to make sure that the links remain active through all the various system and browser updates that will come up over time. But this solution has advantages that make the curation effort worth while. In your content management system, you can add hyperlinks to external resources and, in that manner, generate a connection between your resource and other resources. This greatly increases the findability of your digital resource. Now your minimal webpage presents pictures of the documents that you took with your smartphone, a simple description of the documents and links to complementary resources, and it is ready for everyone to access.

What will you have provided then? In a way, a digital archive, albeit not a very sustainable one. You will have made your documents available and, by adding metadata and hyperlinks, you have made it findable. If you add a license authorising reuse, you will have made it reusable, too. What you have not provided if you remain at this level, however, is a digital edition of your resource.

A digital edition would mean that you had supplied a text (and not simply images and metadata). Such a text can be of variable quality: it can be a raw transcription; it can even be an OCR output that is almost illegible to the human eye. But this is a first step, and can be considered a digital edition if it is conceived in such a way that it can be improved and referenced, and that you have explained how it was conceived. Of course, digital scholarly editions are validated through further quality insurance mechanisms.<sup>51</sup> But, by and large, having established a text (again, not simply scanned an image) makes you an editor. This means, conversely, that it is necessary to provide information about the person (and/or algorithm) that generated the text: an edition is always procured by someone or something that has to be named and characterised since it situates the editorial endeavour. One piece of information that is essential to any digital edition is who is its editor: if that information is not available to you when you consult a digital resource, you cannot take for granted that you are dealing with an actual digital edition.

<sup>51</sup> For Digital Scholarly Editions for instance, a formal catalogue of quality criteria has been developed with the goal of providing the backbone for evaluation processes. An English version is available at https://www.i-d-e.de/publikationen/weitereschriften/criteria-version-1-1/.

What happens at cultural heritage institutions when they are in a similar situation, having acquired, for instance, a new manuscript that they would want to make available to a larger audience? They have one key advantage: infrastructure. They have professional scanning material and skilled personnel, maybe also expensive automated text recognition software, server capacities, and a web interface. They can also rely on a set of standardised metadata that are integrated in digital catalogues, making the resource easy to find online. They can provide information on the content of their manuscript and connect it with the VIAF catalogue (that records writers of all times and other historical figures). They also have the opportunity to add an entry in this database and create an identification number if the author is not already registered.

But again, what is then provided is a digital archive, not a digital edition. A digital edition requires one to procure a text, in this case, at the very least, a transcription of the manuscript, and ideally a transcription that gives further information about the text. This information can relate to the way the text is distributed on the manuscript page, whether part of it has been erased and rewritten (if so, how), and whether it contains allusions to another piece of text, be it explicitly or implicitly (and then, if possible, point to it). In fact, in a digital scholarly edition, it is possible to gather as much information as you want about what I called the geography and the history of the text in section 1.2.1.

This is because, in a digital edition, you are not bound by the constraints of the page format. It is possible to integrate a wealth and variety of information in the digital source document of the edition — more than can be technically displayed on one page or on one monitor screen in a way that could be interpreted at a glance by the human eye. How this information is displayed is another question, and to some extent a secondary one. The core of a digital scholarly edition is not what you can see on your monitor; it is the information contained in the annotated text data (the encoded source file).<sup>52</sup> It is all the more difficult to appreciate what a solid digital edition really is when what you get to see online is not primarily the source code, but one way of presenting it.

<sup>52</sup> The Text Encoding Initiative has been developing guidelines for the sustainable and interoperable annotation of text since the 1980s. It fulfils all the requirements I mention above and facilitates those I mention below. For an introduction to the TEI, see Burnard, What is the Text Encoding Initiative [31]. The TEI Guidelines can be consulted and browsed at https://tei-c.org/guidelines/.

This way of presenting the information will inevitably be partial, constrained by the limitations of the monitor interface. But whatever the design choices digital editors make, they have to explain what is presented and how, and that, again, takes up space. As a result, one crucial accomplishment in the design of digital editions consists in finding a balance between, on the one hand, the intuitiveness of an online representation, which is key for the reader to gain orientation in the presented text and, on the other, all the visual options one might have available to represent the complexity of the textual phenomena one wants to account for. This might be interlinear, non-textual information, for instance, or changes in colour and disposition.<sup>53</sup>

To some extent, and as opposed to other devices such as a smartphone or a tablet that can be moved and turned, personal computers (PCs) still make editors largely dependent on the page format when editing primary material that is displayed on a page in an original manuscript or print (the analog source). If you want to edit text that is written perpendicularly to the rest of the page in order to display it on a PC monitor, there is no choice but to turn it around in your online presentation so that it is aligned and in the reading direction relative to the rest of the page. Otherwise, readers would have to turn their necks 90 degrees to read it in a position corresponding to the one it has on the manuscript (which is not a problem with a book, as you can turn a book through all 360 available degrees).

In that sense, when it comes to digital editions of textual heritage, we are only partially emancipated from the page format. In fact, many digital resources reproduce even the gesture of thumbing through pages when presenting scanned books. But this is only a cosmetic issue. A much less cosmetic question concerns the way to cite a digital text. When it comes to citation, the page is most certainly convenient. Or, more accurately, citation practices have historically developed alongside page and book formats, and not staying with such units in the digital world makes text citation much more complicated.

Depending on the device you use and on your personalised settings, an electronic publication will be displayed on a varying number of pages, which means that the page disposition can change according to your

<sup>53</sup> See Bleier et al., Digital Scholarly Editions as Interfaces [25].

personal settings. Page 3, for example, could refer to a variety of contents and not, as is the case with books, to a stable block of text that is printed (or, for that matter, handwritten) on the same page as the number three. In order to refer to the same block of text independently of the device on which it is displayed, it becomes necessary, in a digital context, to rely more strongly on overarching structures such as chapters and on underlying units such as paragraphs. It has therefore become good scholarly practice to quote paragraphs from publications. At first sight, applying this system to, for instance, literary texts, seems a bit dry. But the only alternative at present is to rely on the URL, independently from the length of the text one can find under this URL. The notion of page remains subject to many variations when applied to the web.

The fact is that the emancipation from the page is only partly achieved to this day, and that no alternative practice of constructive scrolling, no serendipitous move from one online text to the next, has yet established itself as a recognised cultural practice.<sup>54</sup> How to read online text is still something of a mystery, also because readers often do not know how to read, given all the features, columns, and banners, and how to know what it is they have read (fake news or valid statement?).

This consideration brings me back to the question of the qualities a digital text should display. We now know this much: it should be findable, interoperable, enrichable, citable, and legible for man and machine. At this point, it is legitimate to wonder bluntly to what extent these quality criteria are compatible with those developed for print editions or traditional archiving practices. A digital text does not need to be completed, it simply needs to state where it stands in the realisation process. A digital text does not need to be authored by one person, but the roles and contributions of all the contributors should be stated. A digital text needs to be curated in terms of the software necessary to access it, otherwise it might cease to be consultable. A digital text cannot subsist in the long run without a stable infrastructure, in which it is different from books but similar to archival material. One of the consequences of these shifts pertains to the constitution and transmission of texts that are recognised as central

<sup>54</sup> In the article entitled "Lecture" (reading) of the philosophical dictionary L'interprétation, Denis Thouard assumes an ongoing renewal of reading practices as a fundamental cultural device: "Dans son universalité, la lecture est un mode d'orientation fondamental de l'être humain" [99], p. 259. See also François Moureau, La plume et le plomb [83]; and Brigitte Ouvry-Vial's stance on reading seen as commons [84].

by the concerned cultures. In particular, it fosters a reassessment of the literary canon and how it is positioned as a cultural practice.

The orientation provided by the literary canon especially over the past decades and centuries relies on archiving and publishing practices. The examples related to Goethe that I have presented in sections 1.1.1 and 2.1.2 are explicit in that regard: there is a political and cultural coherence in the choice of literary works that are considered essential in a given cultural context, and this coherence can be better monitored through control, at state level, of the infrastructures that are responsible for the way these texts are provided to the public. If the Goethe-Archiv is funded by the state, it will be appraised in educational schemes: there is an expected return on investment. Looking at the current state of digitisation, most of the countries do not invest vastly in large national schemes including targeted investment and infrastructure, while, on the other hand, initiatives like Google Books provide a mass of digitally available and, to some extent, searchable text, which it is impossible to compete with at state level. This raises the question of who should serve as gatekeeper of textual culture at large, and to what extent it is problematic — or not — if it is not a public actor like a state, a region or a city that decides and implements what is worth digitising and what is not, and in what quality.

It has been a major effort of literary studies scholarship to take advantage of digitisation in the past decade to propose a revised view of the literary canon, at a European level at least. Much energy was invested in a form of self-criticism, as well as an analysis of the biases in the traditional literary canon, looking for ways in which these biases could be counterbalanced.<sup>55</sup> By doing so, one inevitably creates a new canon. At the time I am writing this book, digitisation of textual heritage is at its most extensive generated by big tech companies. States, and to some extent the EU, try to sponsor specific branches of cultural heritage by funding the digitisation of specific collections. Scholars strive to analyse and channel efforts towards a better-balanced representation of linguistic and/or stylistic minorities, as well as politically and culturally less dominant figures.

A major impulse was given by Moretti with *Distant Reading* [82], as well as Jockers, *Macroanalysis* [71]. For more recent progress, see the work on the "European Literary Text Collection" ELTeC, https://www.distant-reading.net/eltec/.

It is difficult to predict which of these influences will prevail in the decades to come. The tension is also one between the quantity and quality of digitised information. The current efforts speak for a better recognition (and establishment) of texts that are preserved in a more sustainable manner: texts that are more findable, more enrichable, and more shareable.

In that regard, the way in which human readers adapt to reading more than just linear sequences of pages may play quite a decisive role. As important as the cultural technique of linear reading of a book page may be,<sup>56</sup> media studies have shown that other forms of reading have gained traction with the popularisation of web-based content dissemination and, with them, other perspectives on textuality. In the next section, I would like to show how Open Access has the potential to be instrumental in shaping new quality criteria for textuality in the context of wide dissemination and enriched forms of reading.

## 2.2.2 Access made Open

In the previous sections, I have elaborated on the quality of digital text at what is a rather theoretical (albeit occasionally technical) level, without taking into account the economic trade-offs that digital access to text relies on. I will now turn to discussing access to text taking account of that point of view. This includes an analysis of the way in which digital archives, digital editions, and digital publications in general are interconnected, which provides an opportunity to present all that an open access to textual heritage would ideally entail. In fact, I intend to show all the good that digital media have done to the access of text. I begin by paying closer attention to the philosophy of knowledge sharing, but I will ultimately circle back to technical feasibility.

The modalities of consultation of web content that I have previously mentioned were subsumed under an assumption that I have left unexplained until now. All digital connection processes, whether connecting via hyperlinks, enabling web browsing with the support of geocoordinates, or providing a wealth of biographical data, only work if the resources the links point to are accessible, that is if they do not require

<sup>56</sup> On the negative effects of digitisation on bibliophily, see Piper, *Book was There* [89]; and Carrière and Eco, *N'espérez pas vous débarrasser des livres* [35].

the user to pay in order to access them. Many resources, however, are not available for free, but hidden behind a paywall.<sup>57</sup>

To some extent, it makes sense to have to pay to access content. You pay to purchase a book or a newspaper, and you sometimes pay to visit a museum: the personnel and infrastructure that provide content and service are working to make it possible, and work has to be paid for. But you do not always pay to use a library (or only very little), and you seldom pay to consult documents in archives. This means that mechanisms giving free, or almost free access exist, even to material that is produced in a private context and involved funding in the production of information. And there are certainly mechanisms to grant free access to material that is produced in a public context. Archives curate documents produced by administrations whose work is paid for with taxpayers' money. In return, taxpayers are provided with the opportunity to consult them.

Over the last centuries, such trade-offs (for information that is generated by public servants) or forms of leverage (for information produced by free market actors) were developed to foster access to information. They related to material of a textual nature in the case of shared resources that are considered of general political and/or cultural interest. Political interests then translate into economical means to facilitate access: for example, states subsidise cultural heritage institutions so that these can share their material. Not all heritage institutions share material under the same conditions: the balance between public service and economic sustainability is not easy to maintain when one has to take care of a building, employ personnel, pursue an acquisition policy, curate collections, and enrich catalogues. Investment choices depend on political agendas, but generally, it is considered one of the tasks of the public administration to monitor and facilitate access to relevant textual heritage material, whether it is in the form of archives, libraries, or museums.

What is true in the analog world is also true in the digital context. But since the digital has not been available for as long a time as public archives, libraries, and museums, it still lacks, at least to some extent, safeguards that would make the question of access one that is easy to answer. Access to text involves a series of actors and mechanisms whose

To my great dismay, this is the case with many of the resources I am pointing to in the bibliography of this book. I would have liked to be able to provide only fully accessible references, but it would have required the sacrifice of many important ones.

interests and whose business models are likely to follow different rationales. As a consequence, there are a variety of ways to compensate for the costs corresponding to the effort of making text available online. These generally depend on the type of text and on the type of institution that provides it.

If public archives wish to make their material accessible, their chief goal has to be to ensure sustainable access to metadata, in such a way that they can be integrated into a meta-catalogue, for instance, a national finding aid of all literary archives such as Germany has,<sup>58</sup> and in the best case to images of their textual material also. If libraries want to do the same, the requirements are similar: an online catalogue and an online collection of book scans make it possible for readers to find and read the books they are looking for.<sup>59</sup>

None of this relates to an edited text in the sense of a digital edition. In order to make a digital edition available, one needs to give access to full text in a text format and not simply as an image, and to acknowledge editorial choices explicitly: why this version rather than another, why information is displayed the way it is and, ideally, with what other online resources the digital text is connected. The digital infrastructure as well as the human skills necessary to produce an edition are different from those needed for a digital archive or library. But they are not necessarily disconnected from one another. Indeed, they are complementary and should be connected. When you find a digitised book, you might want to be able to quote passages without having to transcribe the whole book for yourself, and you might want to know why this version has been digitised rather than another, and also, perhaps, understand why this library has this book, in the case of a rather rare volume.

This is all the more true for archives. Finding aids there usually contain such information as type of document, number of pages, date, place, writer(s) and acquisition history. Ideally, this information is detailed for every single item, but even such a minimal set of intelligence requires quite some effort to assemble. Generally, archives receive archival material in the form of a folder or a box, in which papers are in a specific order

<sup>58</sup> See https://kalliope-verbund.info/en/index.html, a unique centralised resource in the German context where federated structures usually lead to a wider distribution of resources.

<sup>59</sup> See, for instance, the French Gallica website: https://gallica.bnf.fr/accueil/en/content/accueil-en?mode=desktop.

that is not necessarily the most logical one (letters, for example, might be thrown together without regard of their chronological order). Following the provenance principle mentioned in 1.1.1, archivists need to record the way in which the folder was handed to them, and number its elements, then consider the form in which the folder has been acquired and align it with conservatory or archival logic, and then record this information. Even if only to proceed to this work step, they need to carefully consider each textual unit and its relationship to the whole folder. Recording information for each single document adds one more work step. And making them available not simply through the in-house catalogue, but for a wider audience via catalogues that are freely accessible online, for instance, requires yet further effort, and a complex one too, since it is necessary to coordinate with other archives in order to align finding aid systems. <sup>60</sup> Needless to say, not all archives can afford all of these work steps.

When you look for a precise piece of archive, you rely heavily on the quality of the recording achieved by archivists. If you do not know which institution is likely to host that piece of archive, it will be crucial that their in-house finding aid is integrated in meta-catalogues that regroup finding aids from archives with similar interests, otherwise you will first have to list the relevant archives, then call or e-mail them, or even visit them, in order to see if the manuscript you are looking for is actually there. When I first started working on 19th-century manuscripts, state of the art technology was no more than a central Berlin address to which you could write a request letter and who would photocopy and send you home the index cards of the manuscripts you were looking for. Their major performance rested on having copies of index cards from all over Germany. We lived in a world of index card drawers. To make sense of them, one needed to know what item one should look up in the first place. There was no such thing as a plain text search: the user had to be familiar with entry points and bibliographical conventions.

I received a list of the manuscripts relevant for my research around the year 2000. I then visited archives, touring north and east Germany, on my manuscript quest. For some documents, it was not really worth the effort of traveling to consult them, when only a couple of pages proved to

<sup>60</sup> Obviously, coordinating at an international level is even more challenging.

be relevant to my research. For those, I acquired black and white copies that were sent to my home by the archives for a moderate fee. In general, this sufficed for the quality of transcription I was aiming at. I mostly simply wanted to transcribe, not procure a full-fleshed edition. But from the moment on that you start making decisions in the way you render the characters you can see on the manuscript page, you transform into an editor. Should I expand abbreviations? Indicate who I think it is when names pop up? Explain what date is meant in the mysterious doodle on the corner of the page?

It did not occur to me at the time that it would be possible for my transcription to be easily placed in relation to the index cards in the grandiose drawer room in Berlin. What I was chiefly interested in was to find a way to present the comparison between the text I had found in the manuscripts and the edition that was published in the early 19th century, in which passages were missing, names were omitted, and whole parts of the document folder were ignored. I was all the more eager to make my editorial work known to a wider audience because I could see evidence that the person who had published the altered edition had had the exact same manuscripts in their hands as I did myself. They left editorial marks, strike-throughs, comments in the margin, most of them corresponding to the editorial choices made in the print edition. <sup>61</sup> The only things I was really missing in order to fully understand how this edition had been conceived were the proofs of the edition and the correspondence between editor and publisher. I have not found them to this day, but I have come to realise, also, that the work I do has to be conducted in the context of a structural lack of information. I will never be able to fully understand it all. But I am able to understand enough, from what I can marshal, to improve knowledge on a variety of topics, such as intellectual networks, editorial processes, and correspondence rhetorics, for instance.<sup>62</sup>

But then what was it worth to advance this knowledge for the handful of scholars interested in Romantic studies who would have read the scholarly articles I wrote? Why would it not be possible for people interested

<sup>61</sup> Anna Busch shows in *Visualisierung* [33] how this contrasting information can be extracted from the different iterations of the text and digitally displayed. See also a comparison of different visualisation tools for textual variants by Torsten Roeder in the *RIDE* journal [93].

<sup>62</sup> See for instance Berliner Intellektuelle als Programm [12], Netzwerke des Wissens [4], Das Netzwerk als Kunstwerk [8], Wissen, Lieben [6].

in the transcription to get to see the manuscript and understand all the different layers of corrections, and to compare editions? And why should archives not be able to provide the transcription I made that could surely be of assistance to other scholars, perhaps in such a way that it becomes interesting not only for Romantic studies scholars? While these were speculative remarks I made to myself — fantasies, really, twenty years ago — making that big dream possible has been greatly facilitated by the popularisation of digitisation processes developed from the 2010s on. If a scan is online, and a finding aid is online, and a library catalogue is online, and a manuscript transcription is online, all relating to the exact same manuscript, then why not connect them all together? What was materially impossible with index cards in drawers, with manuscripts hidden in boxes in dark, cold rooms that could only be consulted individually, and with editions that were expensive books that did not even bring much reward in terms of academic reputation, suddenly became an evidence in the digital context. Once you have admitted that all of these — the scan, the catalogue entry, the transcription — are iterations of the representation of a text, and that what you want is to make visible all the knowledge that they entail when brought together, in order for a wider audience to be able to see and read that text anew, then it becomes very simple to define and reach a goal — the idealistic goal of providing access. And providing access means contributing to a massive background infrastructure that supports the circulation of complex information in such a way that the reader can find, display, and use it as easily as possible. 63

In a sense, the digital media provide the platform to fulfil the promises of Enlightenment, bringing knowledge to a vast array of readers. The crux is that it requires more than simple alphabetical literacy to be able to process online content, and it does not suffice to make quality information accessible online for it to automatically increase knowledge, rationality, and the state of the world. It requires informed digitally literate users. Nonetheless, technical solutions exist that make it possible to connect information and provide access to different representations of text.

This does not mean that we now have all knowledge of things just because we can gather it. It has more to do with arranging the pieces

<sup>63</sup> On why an archiving and publishing ecosystem encompassing research, cultural heritage institutions, and research infrastructures is necessary, see Anderson, *What are Research Infrastructures* [2] and Borgman, *Scholarship in the Digital Age* [28].

of a puzzle than actually completing it. What is more, the different pieces of the puzzle do not necessarily fit all that well. Some are broken; some do not adjust precisely. The formats and languages used to present information vary from one area to another. Archives have different digital standards than do editors.

One major challenge consists in making archival metadata and scholarly metadata interact with one another. While both are eager to know who wrote a letter and when, archivists and scholars have different ways of making this information readable by the computer and of displaying it online for a human reader. They have different quality control mechanisms (what, for example, if someone made a mistake typing a date: how do you verify and correct that?). They also have different ways of adding new information as it becomes available. Delayed synchronisation results in a great loss: it would be desirable for metadata to be actualised at the same time in the different resources presenting one specific textual resource. If a scholar realises that the text was not written in 1837, but in 1836, because the writer made a mistake, for example, they will know that because of information contained either in the text or outside of it, but it is information that is not necessarily processed and validated by archives. In that sense, while it may be very simple to add a hyperlink or a pointer from one digital resource to another as a way to connect different iterations of the same text resource, the fact that web content is bound to evolve and change in terms of its content makes the connection technically challenging. And it obviously does not come down to finding a technical solution alone: human decisions are involved — in this case, philological ones — in order for the connection process to be fully acknowledged and recognised by all actors. Generally, this is about changing our understanding of truth (additional information can change our knowledge of facts) and the mechanisms through which we validate the way in which we advance knowledge (not all insights scholars may have about texts may be considered as valid).

All the connecting, validating, and presenting of work steps involve time and effort, which someone, in some way, has to pay for. Archival finding aids are paid for by archives, that is, in general, taxpayers' money. This goes for library catalogues too. Scanning their respective holdings already opens a Pandora's box, as it has only recently become part of their missions and generates important costs. Some institutions partnered with

Google in the context of the Google Books project to foster the digitisation of their stocks. As for editions, they too can also be funded by taxpayers' money, but generally not on the basis of a long-lasting mission: scholarly funding is still aligned on print productions that seal the achievement of a work process — a tangible end product. In other words, digital editions conceived in the context of so-called "projects" have no way of funding the long-term digital infrastructure that is necessary for them to remain accessible in a few years' time. On the one hand, we have institutions whose core is funded primarily for missions others than digital access; on the other hand, we have knowledgeable editors who do not have the required infrastructure to make the result of their efforts accessible in the medium and long term. And in the middle, we have a reader who does not necessarily have the skills required to get orientation and find their way around the overall mass of text that is the internet. So let us ask again: how can a reader recognise a good text from a poor one, and how can we foster access to the former and/or discourage readers from turning to the latter?

A good digital text is one that provides enough information to be citable: it provides a stable URL or a unique DOI, 64 an author, a date, and a title stating the nature of the resource as a bare minimum. But if the rest of the information is only available upon payment, the value of referencing these core elements is limited. The type of payment required to access a textual resource varies considerably. Some archives, for instance, established the rule that the first user to request the scan of a manuscript would pay for the digitisation process, which is not always very expensive unless you want to scan large amounts, and the archives would then add the realised scan to their digital collection for anyone to access for free. This has two positive consequences. First, it makes the reuse of existing scans all the more appealing, as users do not have to pay for scans that are already available: it fosters the reuse of digitally available data. Second, making these texts available in high quality and with quality insurance as part of a collection hosted by a reliable institution makes them more visible. The users who paid for digitisation will have

DOIs are based on a registered ISO standard; see https://www.doi.org/. Current dissemination practices lead to the attribution by publication aggregators of several DOIs to the same text entity, which is extremely detrimental for archiving, cataloguing, and dissemination.

contributed to the integration of this textual content into robust digitally accessible resources. In that sense, they took an active part in shifting the textual canon towards what is accessible freely and of good quality.

In other cases, you have to pay either for a single visit to an online resource, or for long-term access. Some offer the opportunity to grant access at an institutional level, in general for a rather high price and as a package in combination with other resources. Universities can thus acquire sufficient entry points to some resources that all their personnel and students can access them, as in the case of scientific journals. In that scenario, individual readers or institutions pay for the service that is provided to them, and thus for the maintenance of the publishing infrastructure. This is a situation very similar to that of the Ancien Régime where only a few fortunate rich or well-connected people could have access to cultural resources. The centuries that separate us from that period are precisely those in which public institutions like archives and libraries have been established in order to disseminate access to knowledge and culture more widely. Just as it has become possible to organise sustainable solutions for public libraries that provide access to valuable books, and for archives that provide access to manuscripts, it should be possible to have a sustainable digital model for the public accessibility of textual material even if its production comes at a cost. These models are the ones fostered by Open Access.<sup>65</sup>

Open Access strives to make information available and accessible online for any user. Some forms of Open Access are not as open as they hope to appear in the sense that they simply shift the weight of the cost from the user's shoulders to the producer's.<sup>66</sup> It is the person who wants to display content online that has to pay for the access to be made open and free for other users. Drawing a comparison with the classical library, this would mean that it would not be the library that pays the publisher for the purchase of a book, but the writer who has to pay the publisher for them to deliver the book to the library.

The shifts involved in the relationships between actors in the field of text production affects the mechanisms that frame the collaboration

<sup>65</sup> On Open Access in general, its different models and their implementation, see the standard work by Peter Suber, *Open Access* [98]; and more recently Avery et al., Special Issue on OA of the *Journal of Scholarly Publishing* [3].

<sup>66</sup> On the different models especially in the Humanities, see Eve, *Open Access and the Humanities* [47].

between author and publisher. The added value of a publisher's work — producing a valuable print artefact — is shaken by the potential dissemination power that digital media gives back to writers. Authors can master editing tools and produce legible texts on their own with their personal computers. In terms of dissemination, new media make it possible to catch the attention of readers without relying on a publisher's contacts and advertising. Moreover, readers' expectations in terms of how professional textual output is, as regards typesetting, fonts, and layout, can be lower than the standards expected by book lovers. All in all, you can independently write your text, edit it on a blog, advertise it on twitter, and get a fantastic readership. Monetising this, of course, calls for additional skills. But, by and large, there are ways for an individual to achieve even this to a certain extent, without the help of a publisher.

Even in a context where anyone can publish anything, and perhaps even more so in that context, there remain differences in the types of texts that are published independently, certainly when we consider literary texts. Cultural heritage narratives tell another story than that of a blogger writing fan fiction. And this brings me back to the question of the canon.

Different mechanisms contribute to the inclusion of a text in a literary canon. In referring to texts becoming part of a canon, I mean texts that are to be made available to a large audience based on their relevance for culture and society at large. Obviously, textual quality and overall message play a role, but publishing and archiving strategies do too. Digitisation has become another factor in this process, entailing shifts on account of the opportunity it offers to make extremely large amounts of text accessible to extremely large amounts of readers. Not all of those who can technically access digital text collections will have the skills required to read the digitised text, however. It will be of little to no use for a reader to access a scan of a text written in a language they do not know, or in a script they cannot read. This is an extremely limiting factor if you consider that, even for German texts of the 19th century, for instance, there are fewer and fewer readers proficient and trained in reading the Old German script that was used at the time. How much more difficult it is to make sense of texts from other cultural areas and earlier periods then!

One of the interests of making a text version (and not simply an image version) of such texts available is that they can be annotated in any

language with semantic explanations or even with a translation. *Linguae francae* such as English or, in some parts of the world, French or Spanish, can bridge the gap to help lower-resourced cultures gain traction and visibility. Admittedly, relying on colonial infrastructures is not likely to shift the canon very much.

But for such a wide recognition of cultural material, it requires that it be accessible and free around the world, be it only because money has a different value in different places around the globe. The few dollars that an affluent scholar can easily afford in the USA are an excruciatingly high price to pay for someone living in the Global South. So, what would it take for the dream of an open access to cultural heritage to come true?

An internet of cultural heritage should, ideally, be a meta-archive and a meta-library in one, and facilitate the way in which users navigate from one text to another. It should, especially at metadata level, add explanations in languages that are used for global communication: even if these languages are markers of a colonial past, they are often our only tool to ease communication and transmission across cultures. It should also generalise the use of standards in order for these resources to be easily combinable. Ideally, when you find old papers in your family home, you would then take a picture of them with your phone, add the information you have (date, number of pages, and what you can gather from the writer and the purpose in writing), and place it online in such a way that an archive would be able to provide a link to it if it complemented its own holdings. Of course, the photo you took with your cell phone will not be as good as the high-resolution scans archives can display and with which your material can be connected. But, in terms of identifying where sources of information are, and whether it might actually be worth starting a digitisation campaign for your manuscripts, it still helps. One of the big shifts that has taken place over the past years, in terms of digital resources that cultural heritage institutions host and curate, is that they now sometimes engage in preserving and presenting scans of material that does not physically belong to them, but that is a valuable complement to their own stock. Digital archives and libraries are institutions, but they are also infrastructures, and they can connect material that goes way beyond images of what they possess under their own roofs.

Again, this circles back to the question of the strategic choices that are made. Which text should be valued and integrated into digital libraries? Who pays for this? How do we make it sustainable in the long term, and what about choices one comes to regret later? Maintaining a digital collection comes at a high cost and can be subject to re-evaluation. It may well be that, at some point, a library decides that a specific digital corpus is not worth keeping on their server, but should make space for a more relevant one. There is always room for shifts in what is made accessible and what is not, and hence room for the disappearance of heritage material, even though it would be technically possible to preserve it. Digital media do not alleviate the dialectics of preservation and destruction that is inherent in the curation of heritage material, even for a basically two-dimensional item like text that can readily be handled in formats that are easy to store, easy to connect, and easy to share.

If one compares, on the one hand, a digital text published by a publisher, made available only if the user is willing to pay money (that is, hidden behind a paywall), and, on the other hand, a digital archive or library collection providing free access to scans and metadata, the second is the most sustainable one, because it relies on shared, public infrastructure that itself follows standards and is conceived to address a long-term preservation mission. In terms of the depth of the textual information, it is very likely that the published text behind the paywall provides more insights than the simple catalogue with the scans. But there is no way to easily connect the publication behind the paywall to other resources, whereas it is possible to provide free access to digital editions that are built on top of, or at least linked to, a digital library or digital archive. The interaction between these different points of access to heritage textual material is facilitated by the use of common infrastructures and digital service providers. This means that all the technical decisions that are taken at this level are much more than technical decisions: in the middle and long term, they impact what will become the digital canon of reference, and hence what we teach, and what we know.

A digital text of high quality is one that is freely available to all, in a format that makes adding further information possible, in the form of annotations, of hyperlinks, of transcriptions, of visualisations, of sonifications, and more, and which can be submitted to quality control, such as scholarly quality assessment. In order for this reciprocal enrichment

of information to work, another condition has to be fulfilled: reusability must be authorised. You will remember how annoyed Goethe was when publishers reproduced his works without him being informed about it.<sup>67</sup> Copyright has made some progress since then, but without evolving very much in the digital context. Authors who are still alive might want to provide reuse rights, but sometimes they are not able to do so because they have, per contract, handed over the rights to their work to a publisher. Works of authors that are dead enter public domain some 70 years after their death, and can then generally be reused without any restriction.

But this concerns the text, and not the single copy or the physical media form that supports it. For the single copy of a book or a manuscript, the person who possesses it also has something to say. It can very well be that your cousin, who owns the piece of furniture in which you found the old family papers, does not agree that you can take a picture of them and put them online, for whatever reasons. The artefact itself belongs to someone. Similarly, it is very likely that the publications behind a paywall are not conceived to be reused, but protected by restrictive rights limited to the publishing house.

While some textual material is undoubtedly sensitive, for personal, historical, or any other reason, and it is also true that not everything should be made easily available, it often happens that access to text is restricted for reasons that have nothing to do with the security or integrity of individuals. In western civilisations, the fear of being deprived of one's creation, and robbed of its value, often trumps the joy of sharing it, probably as a side effect of decades of fierce capitalism. That is not only sad; it is problematic.

If we could assume that the basic conditions for the reuse of a text or of the image of a single copy are that the author has to be cited when referring to it, things would be simple. But in fact, most of the time, when you click through a website, it contains no information whatsoever about the conditions of reuse of its content. Can you make a screenshot? How long a passage can you quote from it? Who is the author? Can you harvest and process the data?<sup>68</sup> If you cannot do any of this, even a text that is presented in a text format is not of much more use than a scan of a text in an image format.

<sup>67</sup> See section 2.1.2.

<sup>68</sup> See Baillot et al., Publishing an OCR [17].

This was the last element to add to our assessment of what quality digital text is. A quality digital text is found online in a processable format; it is easy to find through meta-catalogues, and easy to access for free; it relies on stable infrastructures, uses standardised computer languages, and can be enriched with additional information; and it informs the reader about its reuse conditions. Digital texts follow quality criteria that go beyond and add up to the editorial norms that were developed over the past centuries. Quality digital texts are sophisticated, and they reflect the construction mechanisms of society at large. I cannot simply write that this is what the future will judge us on because there is so much more to our culture and society. Yet undoubtedly one of the things on which future generations will be entitled to judge us on is our ability to preserve and transmit quality digital text.

In an analog context, access to text is provided to a much larger array of people through publication than it is through archiving. Textual decisions are taken at each step of the writing, preserving, and distributing of a text, involving a range of techniques and of actors. In the digital context, the limits that separate the different missions in this editing and disseminating process fray. A new division of roles, of functions, and of the money invested, but also of the goals pursued at a political level, could give hope for a future in which high quality text, defined along the lines of sophisticated technical and editorial criteria, could be available to anyone.<sup>69</sup>

The historical perspective in this chapter has shown how a wealth of actors have always been involved in shaping textual products. These collaborative processes have long been made invisible by authority concepts that have dominated the European book market. In the 20th century, media shifts have paved the way for an acknowledgment of iterative and distributed approaches in text production, constitution, preservation, and dissemination. While educational schemes have not kept up with technical progress, we can now marshal all necessary leverages to embrace digital philology at large and champion the cultural, political, and social benefits of access to text.<sup>70</sup> Recent perspectives on the literary canon

<sup>69</sup> See also Adema, Living Books [1].

<sup>70</sup> See also the image of the leaf falling on the river in the introduction to the excellent *Digital Technology and the Practice of Humanities Research* by Jennifer Edmond [46].

demonstrate an authentic effort of critical self-reflection which bodes well for future generations.

In the next chapter, I will explore the limits of this — admittedly somewhat optimistic — point of view in the context of the climate crisis.