



**HAL**  
open science

## How to learn and love digital text in four easy steps

Anne Baillot

► **To cite this version:**

Anne Baillot. How to learn and love digital text in four easy steps. DARIAH Friday Frontiers: Spring 2023 Series, Apr 2023, DARIAH-Campus, Germany. hal-04104225

**HAL Id: hal-04104225**

**<https://hal.science/hal-04104225v1>**

Submitted on 23 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# How to learn and love digital text in four easy steps

DARIAH Frontiers seminar, April 2023

//SLIDE 2// Things used to be better, they used to work better. Or at least you knew what you had to do and you knew what to expect. When you needed a text, you would go to the library, search the catalogue, get the book, check the table of contents or the relevant index, and find the passage you were looking for. Admittedly, it might not always have been easy to get hold of the right passage in the right book in the library at hand. You might have to check several books before finding the right one. And it might not have been available at your local library. But at least you could rely on the procedure, and on what you would have to expect when opening any book //SLIDE 3//: a title, an author (or several), a publisher, a date and a place for printing. There were certainly a few cases where there would be cheating of some sort. The printed name of the author would not be that of the person who actually wrote the book, the publishing house would be a cover for a forbidden one, the place of printing modified in order to circumvent censorship. But all in all, you knew what you were dealing with. A solid author, a solid publishing house, a clear historical context.

//SLIDE 4//And now? Now, nothing is what it seems. When you browse the internet, you can never really be sure why it is this specific text that your search engine is pointing to, sometimes there seems to be no way to know who the author of the text is. Maybe it's a robot, for all you know. Also, why is it that you keep finding the 12 pages before the passage you are looking for and the 240 pages after but not the 23 pages in the middle that you are actually interested in? It is as if the internet had been conceived for you not to find the text you are looking for – for you to find something else, to find other texts, to find videos, to find scans you won't be able to edit, to find memes, //SLIDE 5//oh look at that, what is this cat doing, isn't it crazy that people would record that kind of things, and what was it I was looking for again?

//SLIDE 6// One thing that makes it difficult to apprehend digital text is that it serves myriad of purposes. It is the metadata of the cat videos, it is the sentence you type in the plain text search field of your browser, it is the e-mail you receive from the supermarket on your birthday. It is also digitized heritage and state of the art science in preprint repositories. Text has become ubiquitous in the digital media, melting the Codex Manesse with advertisement. Text is not the book you read from front to end with a big cushion in your back any more. It is a jungle, it keeps growing and embarking



readers on a sprawling journey. In that, the digital media make it difficult to know what is a good digital text and, somehow, to keep faith in the fact that some good may come out of the internet jungle.

Looking back in time, this criticism turns out not to be all new. The introduction of radio in the early 20<sup>th</sup> century for instance was not welcomed with enthusiasm only. The lack of clear quality criteria for new media formats was more often than not translated into a lack of quality. For any change in media, it takes some time to develop quality criteria and to gain orientation. What seem to us to be an intuitive way to read and assess a book is the result of years of school training reinforced by our cultural environment, all developed alongside centuries of editorial practices. Digital literacy, on the other hand, is still in its infancy. While we cannot really imagine our lives without the internet, little is being done to make us actors of the digital world and conscious producers of digital text. We are comforted in our role as passive consumers by major tech companies, who exploit general ignorance and make money out of it, while it would not be that complicated, in fact, to gain agency – especially when it comes to something as simple (and as powerful) as text. The goal of this presentation is to facilitate this shift from consumer to producer of digital text.

Digital text comes in a variety of shapes, all interesting in their own right. My perspective is that of a literary scholar. What I will do today is to nail down quality criteria specifically for text that is relevant in a heritage perspective: text that pertains to culture at large, how to make it available online, how to preserve it for generations to come. I will argue that it comes down to four steps to gain agency over digital text in such a way that coming generations can actually build on it. Digital scholarly editions are major assets in that regard and they play an important role in my argument, as you will see.

The four steps I will present concern 1) how to represent text in a digital environment, 2) how to connect information, 3) how to open it and 4) how (and why) to share it. With these four steps, I intend to show you that there is such a thing as beautiful digital text, and that it can contribute to a more harmonious relationship between our cultural advances and our natural resources.

Now let's embark on the first of the four steps: how to represent text.



## //SLIDE 7//Represent

The book format that is currently dominating the market, and with it, the imaginaries, is that of the modern era, that of the **Gutenberg Galaxy**, as Marshall McLuhan put it.

// SLIDE 8// It has a unique block of text in the middle, which is to be read in one direction only. This block of text is surrounded with blank margins. The whole device is conceived so as to make it possible to flip through pages. Each page hence has a recto and a verso. Page numbers and overhead chapter indications allow for orientation in the text.

Yet this is only one conception of the page. Other artifacts addressing the same type of cultural practices could include, for instance, several blocks of text, or no margin, or information on what comes before or what comes next for each page or each column. Anthony Grafton has shown how history, usage and aesthetics of the page are deeply intertwined, and how complex the concept of page.

Still, when we talk of a page, we usually have this type of page in mind: the page what we learn to read in school.

//SLIDE 9// “N’espérez pas vous débarrasser des livres” – don’t even fathom to get rid of books : the title of this volume by Umberto Eco and Jean-Claude Carrière gives a sense of the widespread confidence in the cultural supremacy of the book format among textual and literary scholars. When affirming “Book was there”, Andrew Piper canvasses the **media-based specificity** of print devices more in-depth, and articulates a comparison with electronic texts. But the use of **equivalent media categories** for the printed and the digital text leads to what can be seen as misunderstandings. //SLIDE 10// Piper writes for instance: “Digital texts lack feeling”, or “Digital texts are somewhere, but *where* they are has become increasingly complicated, abstract, even forbidden”(p. 15). These quotes suggest that the specifically digital materiality is bound to elude any analysis, as if digital text was by essence virtual.

//SLIDE 11// And yet, there is a **materiality in digital text**. While it is true that the page paradigm is remaining dominant, as pointed out by all the scholars I have mentioned so far as well as by many others, I would like to argue that it is a **richer one**, when it comes to text that is available in a digital format, than the comparison with the printed page might suggest.

//SLIDE 12//Just like we usually talk about a modern page like the one I have just shown when we talk about “a page”, we usually mean linear reading when talking about “reading”. Now, linearity is an **impediment**, when it comes to the digital representation of text. Piper writes (p. 17) “What lies *after* the digital page? An abyss.” I think it is



worth actually looking into what comes after, beyond and related to the page, into the abyss.

Let's dive into this question, which is that of **text representation**, and look at it in particular in the context of digital editions of heritage material.

// SLIDE 13// What you would call "a text" in an analog context can be represented in a variety of ways in a digital setting: it can be an **image** (the scan of a book page for instance), //SLIDE 14// it can be **text** (like on zeno.org), and the image as well as the text can be enriched. //SLIDE 15// In this case, **additional information is embedded** in the visual or textual basis version, in a format that is readable by the computer. In general, enrichments (additional information) are themselves in the form of text. They are data or metadata, in a specific computer language, connected to the initial dataset that itself represents the concerned text.

These different forms of textual information (raw, enriched, connected or not) have primarily to do with **the structure** integral to digital text, not with the way in which said text is displayed online. Also, what we see when we find text online will not necessarily give at first sight a clear sense of the nature and quality of the source.

//SLIDE 16//But let me linger a minute on this notion that what we call "a text" is in fact, in a digital context, one specific iteration of one specific type of representation of the text that is concerned. What we would, in an analog context, call "the same text", is likely to come in different representations in an electronic context.

This means that what you see online when you look for let's say, a text you want to discuss with your students, is like **the surface of a mirror**, on which you only see a **specific reflection** (determined by the frame of the mirror, maybe also by its enhancing qualities) while the actual, complete information that is available is stocked behind the mirror. What you can find behind the mirror is **the source – the actual textual basis –**, and what you see online is a specific way of leading the reader to a selected set of the information contained in the source; it is a rendering of the source that could, in other circumstances – for instance if the editor were addressing an audience of 3-yr old instead of adults – look different.

The source is written in a language that is understandable by the machine – in a computer language – so that computers are able to process the information it contains and eventually transform it into what human readers or users will consult online along the lines of what the editors wish to display.

Especially from the point of view of the source, there are major differences between a random electronic text and a full-fledged edition, even more so a digital scholarly



edition, even if what you see displayed on your monitor is, in both cases, mainly computer text that is legible by a human.

While a reader might go online to read the full text of a blog post, the consultation of a digital scholarly edition usually does not amount to an exhaustive, linear reading. //SLIDE 17// A digital scholarly edition provides access to content about the historical, linguistic, geographic, philosophical, etc. context of the text you are consulting, all of which break the mold of linear text reading. Digital Scholarly Editions concern themselves with **enriched texts**, for which the book format always bore major **material limitations**, since it is not really possible to represent all the (historical, linguistic, geographic, philosophical, or other) information that belong together on one printed page, or on even two if you consider the “diptych” (Piper) //SLIDE 18// that frames the analog mode of reading, with one uneven and one even page laid out together. Although the monitor primarily emulates a page, it does not simply **reproduce its content**.

When you consult or realize a digital scholarly edition of heritage material, you should hence take into account: 1) quality of the source (code), 2) solidity of the transformation scenario (database structure, that will make it possible to transform the source code into content legible by a human) and 3) usability of the presentation interface. These are, if you want, three levels of textuality. With them come also (at least) three levels of reading, even though most readers are only aware of the most superficial one, the one that appeals most immediately to the eye.

### //SLIDE 19//Connect

The internet has developed as a vision-based space. The vast majority of textual information is presented in a visual format (much more than audio). As a consequence, dealing with ways to present text online amounts in general to issues of **visualization**. One particular type of visualization has gained traction over the past decade and has become particularly dear to Humanities scholars: //SLIDE 20// graphs, and more specifically, networks. Networks connect information but they can only do so based on source data that itself contains information, that is: rich data.

Visually connecting information is also a way of inducing certain types of reading of the source material. //SLIDE 21// Even a simple word cloud like the one on the homepage of the August Wilhelm Schlegel Edition of Letters engages the reader into a particular reading based on a visualization that puts the relative importance of correspondence partners front and center. In this case, it is simple corpus statistics that provides the entry point to the text.



Digital scholarly editions rely on the way in which the information provided in **the core text** – that is, the source – can be connected to information that is otherwise available online (such as bibliographical, biographical or geographical data). This is where the combination of source data and thought-through interface conception beats the print page in book format. //SLIDE 22//In a print context, voluminous indices have to be scrolled and mentally put in relation with the concerned page through an abstraction effort of the mind. In an online interface, a click or a mouse-over suffices to provide additional information: it is presented next to the relevant text bit. //SLIDE 23// The digital interconnectivity of information, when provided in reliable formats, expands the page as though through myriads of windows. In that sense, **the network** is certainly an accurate figuration of the manner in which connected information can be solicited in a digital context.

Connecting the dots with available information makes it possible to increase knowledge in a common effort. By connecting the information each and every one disposes of with, for instance, wikidata, we generate a web of information that makes it possible to expand the space that can potentially be explored. But such a mechanism relies on the fact that the effort is, in fact, common. Striving towards the same goal looks brilliant on the paper, but it comes with a series of difficulties, some technical, some more philosophical. In order to connect the dots, the available information has to be retrievable in a format that facilitates aggregation – a point that is key to understanding the interdisciplinary reuse of data in the context of digital scholarly editions.

But before I move on to this, I would like to discuss another drawback of the many possibilities there exist to fill the digital page. Over the past decades, efforts have been made to get a better grasp on the reading behavior of users of digital scholarly editions, and for a good reason. In fact, digital editions too encounter a **limitation** due to the **excess of information** that can be presented at once. The human mind has its limitations in terms of how much information it can process at a time – very different limitations than those of the computer. While the machine can process massive quantities of source code to generate, for instance, complex graphs, the human mind is not spontaneously able to process the same quantity of data, sometimes even to interpret the result of these computer compilations. This means that issues of visual design have to be taken into account when presenting information online in order for readers to be at all able to receive the information you want to transmit.

More broadly, the fact that digital editions, and particularly scholarly ones, are accessible for anyone to read come with an **educational challenge**. Not everyone used to consult big folios and their appendices. Only learned scholars did: people who had been trained in the specific way of reading that is required to gain information from



such complex contraptions. With online content being virtually accessible to anyone, the issue of digital literacy at large has become a major social challenge.

Indeed, this does not only concern digital scholarly editions, but digital resources in general. The general inability to consider the different levels in text representation, to differentiate between source code and visualization, between raw data and enrichment, to distinguish metadata from data, is a disappointing failure of North-Western countries to update educational and cultural practices at the same pace as they develop technologies. Universities have a key role to play in instilling these basics. This does not only concern students of Digital Humanities, but should to my eyes be integrated, for instance in the form of introductory data-based approaches, in methodology classes of all Humanities disciplines, particularly the philological ones.

### //SLIDE 24//Open

The ability to read interfaces (or rather the lack thereof) is per se an issue. But there is more. When publishing online content that has heritage value, it is not only necessary to take into account graphical design challenges, but also requirements that are those of scholarly and cultural endeavors. Online content does not only have to look good and be user-friendly, it also has to display coherence with the artifact it presents, to inform about its context. This is what digital scholarly editions do. They take a text that has a recognized (or recognizable) heritage value and transform it into data that is then presented online by connecting it to other data that is available online. //SLIDE 23// Digital scholarly editions present editors' **readings** both in the literal sense (the editors present the text they have read and established) and with regard to the stance they take on their findings. This dimension of a digital edition is not located at the sole surface of the mirror: it also concerns all that is hidden behind it, the data at its source. A digital edition can only present that which is integral to its source code. Much is at stake, both in terms of representation and of graphical implementation, in the way the data is handled.

This brings me to the third element that is necessary to apprehend digital text. I have discussed how to represent it, then what the connection of information embedded in digital text brings, and I would now like to show that nothing of all this can be done if the data as well as the ways to access it are not open.

In order to be presented online, the source information has to be **processable by a computer**: ideally in such a manner that it can be processed by any computer, at any point in time, as this will guarantee that the data will remain accessible. Poor choices in that regard are largely responsible for the webpage cemeteries that we all have come to encounter: empty URLs, contents that disappear from one day to the next... in many cases, the content is actually still there, it is simply not accessible any more.





The technical choices that are made when deciding on formats and languages are not simply technical. They engage the whole concept of the way in which content is to be presented online. And even if I am not able to program my edition from start to finish single-handedly, I need, as an editor, to be aware of the implications of the technical decisions that are made. Editorial work for digital text is team work where team members educate one another and build on each other's knowledge.

//SLIDE 26// Over the past decades, textual scholars have worked towards establishing best practices to provide guidelines for the online presentation of textual sources. These concern both the surface of the mirror – the data as it is presented or visualized online – and, more importantly, the source code, the data behind the mirror. These guidelines turned out to be in line with what became the FAIR principles of Open Science, promoting data that is findable, accessible, interoperable and reusable. I would like to discuss in particular two elements that are key to thus contributing to the necessary openness: what standards are for and what Open Access at large really facilitates.

//SLIDE 27// More than the principles themselves, it is the philosophy that the FAIR principles serve that can give a notion of the operational framework in which the online accessibility of digital text is inscribed. When it complies by **standards**, a digital resource is easier to connect to other digital resources. Conversely, choosing not to use standards means isolating digital text in a silo. Only by using standards that are processable by a computer does the provider of digital text actually open it to the point that it can be part of the web in a general sense. The benefit of using standards operates on two levels. First, users who work with standards are concerned with a common interest, they are intrinsically motivated to improve state of the art data for everyone. It creates a sense of community in the curation of data and its infrastructure. Second, standards actually make it possible to provide access to a lot of information, as they facilitate its aggregation and dissemination. Standards are good for the people and good for the data.

Admittedly, not everything is perfect in the world of standards. For starters, not everyone acknowledges the same standards and it tends to defeat the purpose of standardization to consider there are several standards in use for a given procedure. Also, standards are not simply out there, like an Eden we would be born in. They are the result of negotiations, of developments, they take time to elaborate. This is on the one hand really nice: we are going where we need to, we discuss, we weigh up, we compromise, and then we concur.

//SLIDE 28// When it comes to standards for digital text, the one key player is the Text Encoding Initiative. The TEI community has an amazing ability to critically frame its own



topicality and to keep building on existing standards and interoperable languages. But it does not solve all problems magically. This is, on the other hand, the **challenging part** of standards in the making: even in the TEI, there are more often than not several ways of accounting for one particular textual phenomenon. And while the TEI guidelines and documentation are open source, by far not all TEI-based resources are available in Open Access.

Still, standards make it possible for digital text to be **sustainable**. They make data curation easier, in particular because they rely on communities of interest, and with that, they contribute to minimize data loss. Standards facilitate **access** to the source code whatever the system used to retrieve it. When it complies by standards, data can be improved, enriched, reused – this is true for Heritage textual data just as well as for any data. To my eyes, standards are integral to the concept of data openness especially when it comes to data reuse.

//SLIDE 29//Facilitating access is key in terms of reusability, and the wide acceptance of Open Access has been a real game changer in that regard. Open Access is not simply a principle any more. It is policy, it is legislation, and it is a business model – more accurately, it is a range of business models. Some forms of Open Access are still widely under the grip of large tech companies and their partners in the publishing world, making life quite hard for scholars, heritage professionals and small publishing houses who want to disseminate their content online not only openly, but also following quality criteria of their communities of practices which, as I have been arguing today, most users of the concerned resources are not aware of.

Strangely enough, though, this is not what people – even Humanities scholars – talk about when they talk about Open Access. I am certain that you all have had these conversations in which you have to explain (again) that no, giving open access to a text will not make plagiarism easier: a license and a time stamp suffice to prevent it and it actually makes it easier to prove authorship than if the document had remained in your drawer. Or you have to convince your students that finding a text that is available online does not deprive it from its author either.

These discussions shed light on how little is known about the work that has been completed to advance quality open access over the past decade. They also show the gap between online practices that are, in theory, established, and their application.

//SLIDE 30//Yes, it is possible to cite a digital resource and all of its contributors. Yes, it is possible to cite a specific version of a digital resource, and it even is necessary because digital resources can by definition be enriched. Yes, this means that mistakes can be corrected (yay!), but also, that there is nothing like a ultimate authoritative



version of a digital textual resource. No, there are no page numbers, but what about numbering paragraphs?

**Pragmatic solutions** do exist but they require to be applied more widely, which in turn can only work if we liberate ourselves from a textual habitus that is still defined by the linearity of the book.

The benefits of standards and Open Access are widely acknowledged in Digital Humanities circles, but they still are anything but intuitive to other text-based Humanities disciplines. They require, in fact, to see **text as data** – still an epistemic challenge for many Humanities scholars – and to consider that data is not destined to meet their fate once and for all in the hands of a given scholar. Data is to go on living: this is the core principle of Open Access – and Open Source. Once you have realized that, there really is no turning back to the “old” way of providing access to heritage material.

For those who have embraced this core idea of openness – namely, that it contributes to a **theoretically endless data lifecycle** –, a new challenge is arising with the climate crisis. Indeed, the advent of Open Access has led to – or has accompanied – an inflationary approach to data. In order to be able to reproduce, in order to give the greatest possible access, to preserve as much as possible, it has become good practice to digitize in the best possible quality and make all possible data available in that maximal quality. This is still considered as the condition for guaranteeing reproducibility and sustainability. Inevitably, the data that is made available in this inflationary manner has, when you sum it all up in the end, an environmental footprint that might not be as large as the one of astrophysics data, but one that is, all in all, not really conceived so as to be minimized.

So, is there a contradiction between Open Access on the one hand and a sober approach to digital information on the other hand? And what role do infrastructures and digital text have to play there?

[//SLIDE 31//Share](#)

The fourth and last action point will provide an answer to both questions. Sharing, and the infrastructure that we need for sharing, are the final building blocks needed to envision digital text as a perspective for the future.

While it is true that producing data and making it available online in a range of formats in order to facilitate access (at least to some people) and reuse has a negative impact in terms of producing content that will generate greenhouse gas emissions, there is a



simple way to reduce the environmental footprint generated by this wealth of resources: it consists in actually sharing them, //SLIDE 32// that is, in them being used by as many people as possible. This means going further than simply making your data available for reuse with a nice CC-BY license, further even than curating reusable data. It consists in 1) having your data actually **be** reused and 2) conceiving new data and/or corpora based on existing data, that is, **reusing what already exists**. We lack academic incentives for both approaches, which is why they are surprisingly rarely implemented although when you think about it, it is clear that this is the only rational way to go in a situation of energy and resources shortage. It is also an interpretation of (green and diamond) Open Access that is actually aligned with the values of openness and sustainability. In terms of where we stand with regard to the planetary resources, there should be much greater rewards for the people who build on what exists than for those who use a great deal of energy in order to start from scratch. The more we emphasize innovation and novelty in funding schemes today, the less future generations will be given the chance to have similar room for innovation.

This way of sharing is key to sustainability. What I am talking here is a **different kind** of sustainability than the one we are used to talk about in the Northern hemisphere. It is not about preserving everything on two or three different servers, as tiff and as pdf and also as thumbnail because hey why not, about having a digital resource available 24/7, about providing executables everywhere. I am talking about the kind of data sustainability that we will need if we want to still be able to work with our research objects in three, five, ten years from now, and to do so in a digital and sober way. I am talking about the kind of sustainability our students will need from us if they ever want to pursue, in one way or another, what we are doing today. I am talking about a sustainability practice that invites us to look up how our colleagues from the Southern hemisphere are used to work.

//SLIE 33//I mentioned it at the beginning: the digital is not virtual. It is material, even if textual scholars are not always able to grasp, describe and measure this materiality since it does not amount to running meters of books or of boxes in a storage room. We are not used to consider the lifecycle assessment of the equipment we are using when producing digital resources. Precisely evaluating the greenhouse gas emissions generated by digital activity, even more so their impact on biodiversity or on water resources, is to this day a challenge. It is a technical challenge – methods to measure them vary, information is difficult to retrieve, we are not familiar with orders of magnitude. But what makes it a real challenge is not the technicalities, it is its **systemic character**. By giving access, by advancing knowledge, by preserving heritage without



questioning the way in which we do so, we keep a colonial and patriarchal habitus alive that will be depriving the coming generations of all that and more. The environmental footprint of the production, curation and use of a digital edition can vary. But Open Access or not, many of them are luxury products, accessible only with an excellent internet connection, energy-consuming software, recent hardware.

In countries of the Northern hemisphere, the ecological cost of digitization concerns society as a whole. It concerns the production of digital text in such a way that what used to be considered to be good practice should be reassessed. As long as there is no systemic approach to this systemic issue, the only solution we have is to have this ecological weight supported by as many shoulders as possible.

This is why sharing is the word. Sharing means reducing the individual impact. Now, how do we share? By putting in common. We need **infrastructures** that make sharing possible (just like libraries are also a way to alleviate footprint). That means large-scale endeavors that make it possible to share necessary equipment like servers, software – infrastructure that can grow to forums in which to discuss developments that will serve the community. //SLIDE 34//Now you see why I am giving this presentation in the context of the DARIAH seminars. I have a couple of to-dos for DARIAH on my agenda for today.

When it comes to digital editions of textual heritage, this means developing and using infrastructures and database instances as common goods, using frugal computer languages like XML and getting used to ponder upon every single step along the way. Is the benefit from what I am going to create worth its ecological footprint? Do I really need to make everything accessible in high quality all the time? Which datacenters are located in geographically advantageous zones? Maybe a minimal, text-based version of my resources can be made available by default, and heavier resources like images only when renewable are available?

Now I am well aware that it might take a while before “less is better” becomes an actionable funding scheme.

Still, we should not wait for funders, especially since we can discuss this with infrastructures and because it is our competences as Arts and Humanities scholars and as Heritage professionals that are necessary to define the quality criteria of digital **sustainable** text - that which tomorrow’s digital collections will be made of. Because the collections that will be preservable, accessible and reusable at a low energetic cost are **tomorrow’s canon**. These are the collections that will be used in schools, that will be referred to in Wikipedia. These are the collections that will shape the future. And these collections will certainly not be made of high resolution scans, of a tiff version



and also another pdf version of the same high-quality image. These collections that will be the core of tomorrow's Arts and Humanities will very certainly rely on text, for several reasons.

- First, because the technologies we need to make digital text actionable at a low environmental cost are already there. They are open source, they need little development, they mostly require community-based curation and they are at the core of the work of major research and culture infrastructures.
- //SLIDE 35// Second, because one essential pillar of these collections and their low environmental footprint will be their reuse, which in turn is based on discoverability. In other words, structured metadata should be the cornerstone of sustainability. And metadata are, again, well, text.

In other words: When it comes to data-based approaches to Arts and Humanities, there are ways to combine digital media and sobriety. They call for an encompassing vision of digital text as leverage, they call for a common traction by Arts and Humanities scholars and Heritage professionals. And they by no means require to give up on Open Science, simply to pragmatically reframe its core principles.

### //SLIDE 36//Quality criteria for digital text

As a conclusion, I would like to sum up all the qualities we should look for in a digital text, both as users of such texts and as their producers.

A quality digital text is a text that is found online in a processable format. It is easy to find through catalogues, and easy to access for free. It relies on stable infrastructures, uses standardized computer languages, can be enriched with additional information, informs the reader about its reuse conditions and facilitates them. Digital text follows quality criteria that go beyond and add up to the editorial norms that were developed over the past centuries. As a social construction, it reflects on the mechanisms of society at large and provides a self-assessment of its position in its respective context of production and/or (re)use. A digital text that lives up to these requirements is definitely a text we can take pride in. Incidentally, it shows that there are things that we can do better with a computer, for today and tomorrow.

We have the technological, intellectual and infrastructural means to produce and disseminate digital text along the lines of these quality criteria, which also bring us closer to a sustainable approach to culture and science.

I am convinced that one thing future generations will be entitled to judge us on is our ability to preserve and transmit quality digital text. The four action lines I proposed today are my suggestion for a general framework towards a frugal digital philology,



one that will make it possible for many generations to come to relate to and enjoy cultural heritage, hopefully in a more global and more equitable manner than it is the case today.

//SLIDE 37//