

Archiving in times of climate crisis. Selecting, digitizing, analyzing for tomorrow

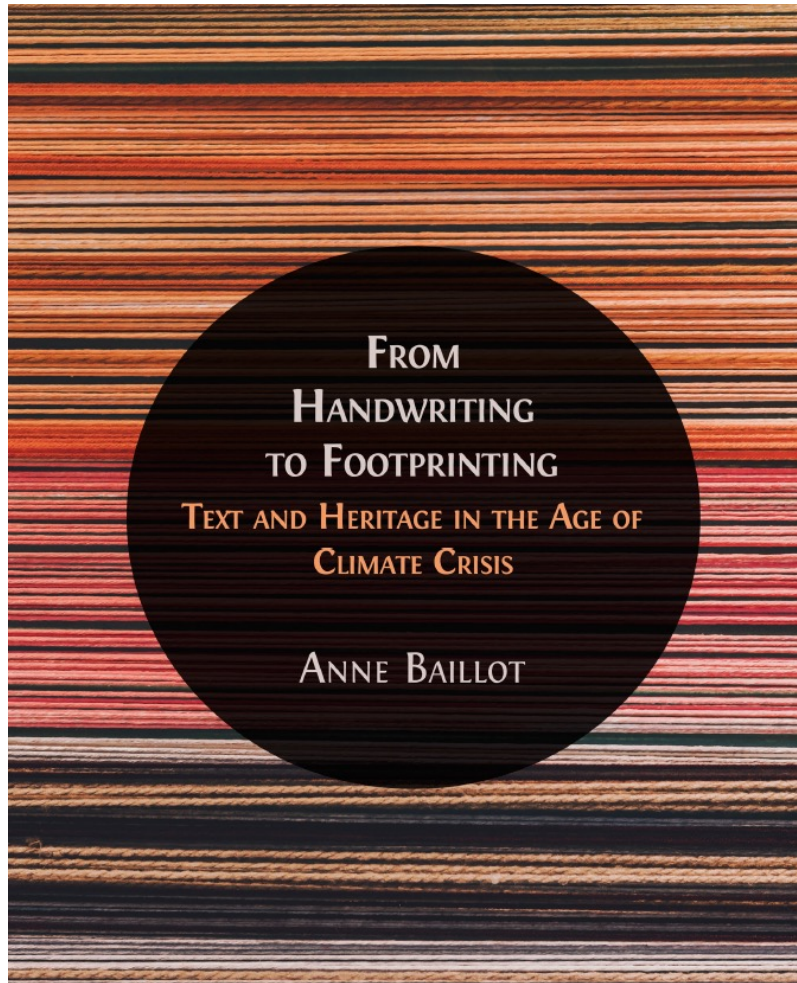
Anne Baillot

DARIAH-ERIC / ATRIUM project

September 2024

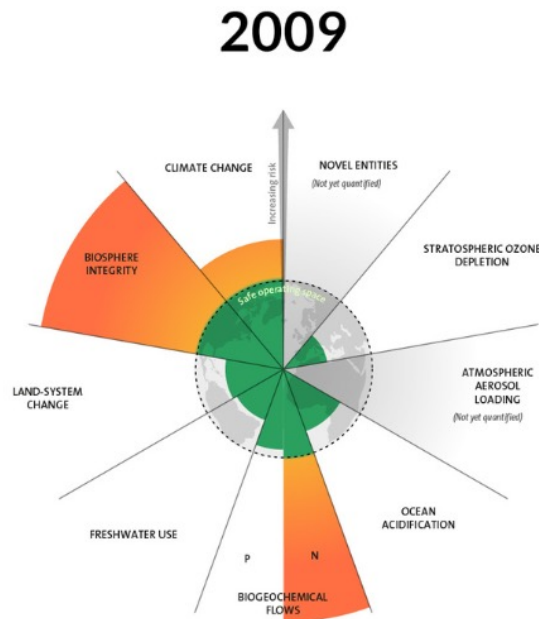


Hands-on as in implementing theory

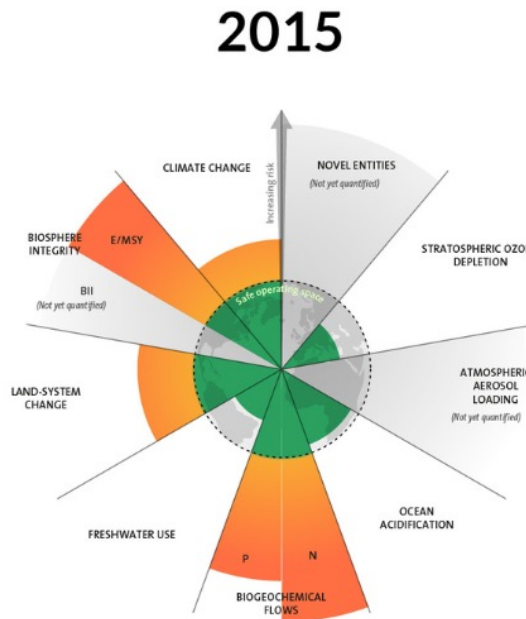


- A largely theoretical approach to a problem that is still mostly ignored
- Applied to my own practice, with a focus on the production of the book itself
- What about implementing the general concept on a bigger scale?

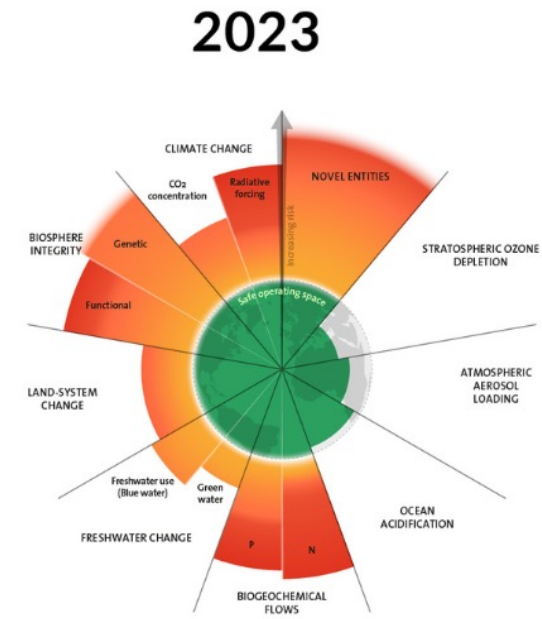
Our situation of polycrisis



7 boundaries assessed,
3 crossed



7 boundaries assessed,
4 crossed



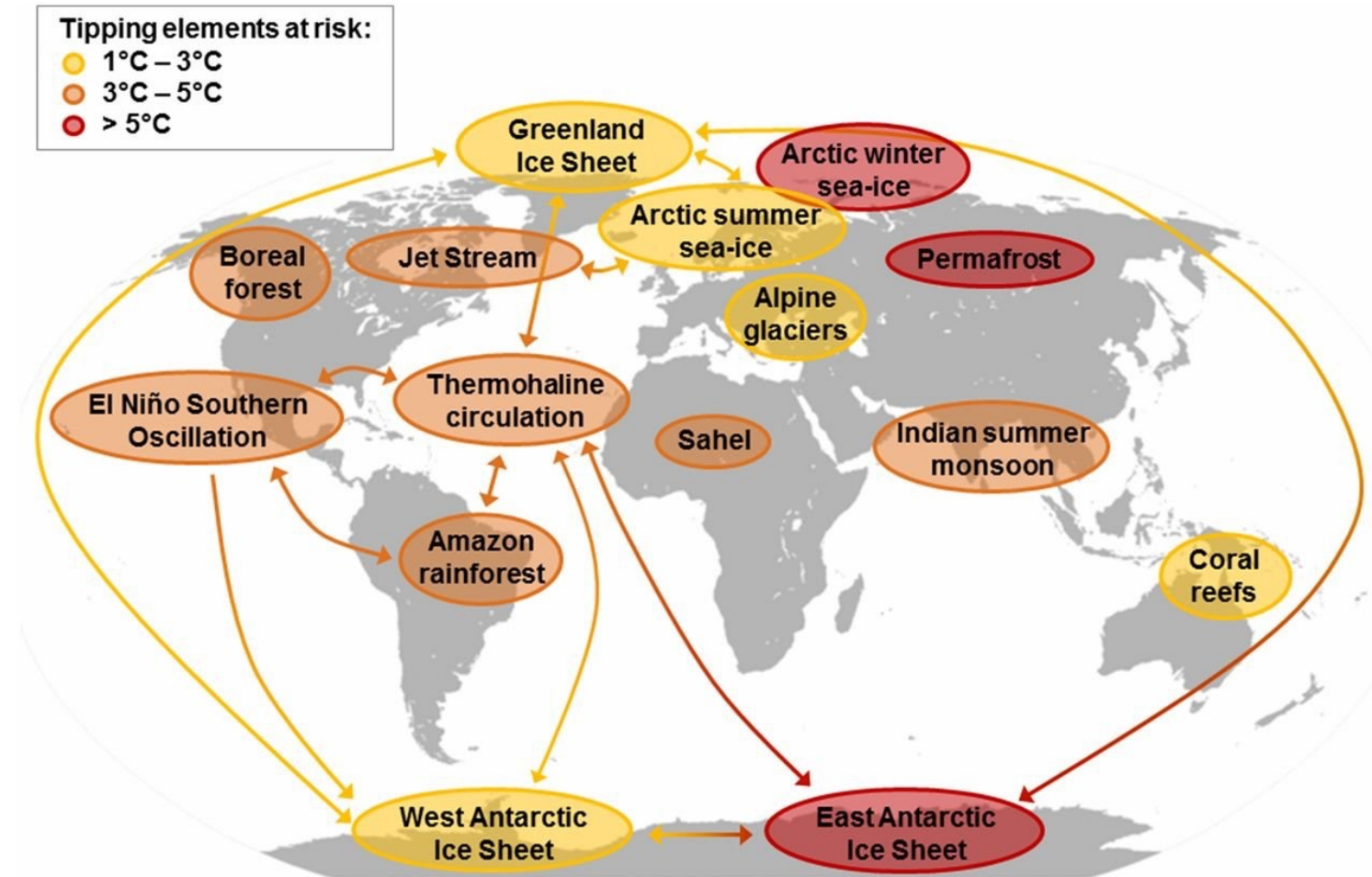
9 boundaries assessed,
6 crossed

The evolution of the planetary boundaries framework. Licenced under CC BY-NC-ND 3.0 (Credit: Azote for Stockholm Resilience Centre, Stockholm University. Based on Richardson et al. 2023, Steffen et al. 2015, and Rockström et al. 2009) Click on the image to download.

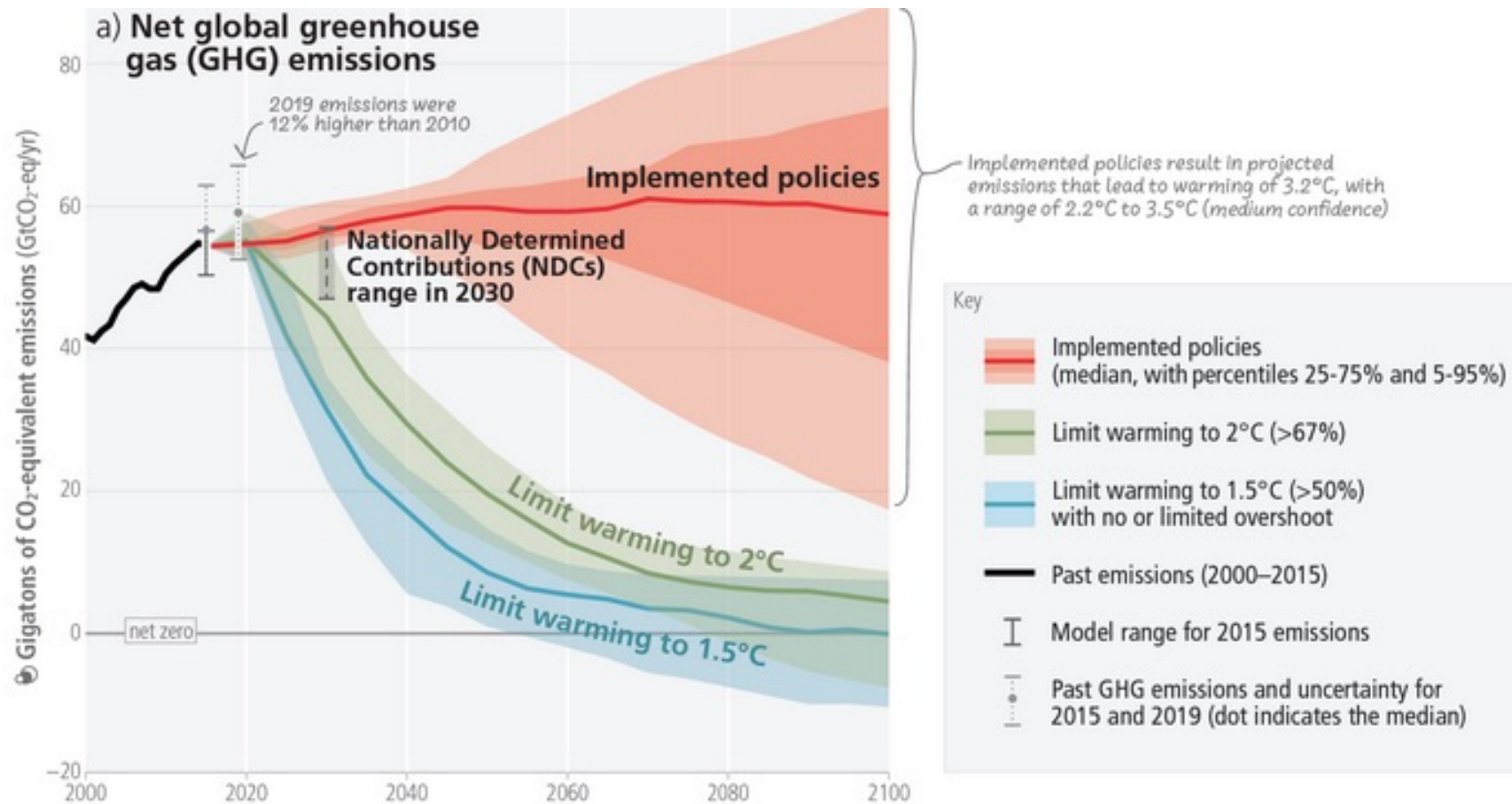
<https://www.stockholmresilience.org/research/planetary-boundaries.html>

“Crossing boundaries increases the risk of generating large-scale abrupt or irreversible environmental changes. Drastic changes will not necessarily happen overnight, but together the boundaries mark a critical threshold for increasing risks to people and the ecosystems we are part of.”

Global Map of potential tipping cascades



GHG emissions reduction scenarios



The environmental footprint of digital technologies

- Global scale: digital technologies responsible for 2-4% of GHG emissions (similar to civil aviation). Tendency: growing.
- Strong rebound effect: development of more efficient technologies leads to a huge increase in their use, hence no impact reduction
- Fosters global injustice: benefits mostly the Northern hemisphere, South carries most of the burden (pollution, indecent work and life conditions, political oppression for production's sake)
- Materiality is key (production of devices, data centers, communication networks, end of life of devices) but largely invisible for consumers of digital technologies

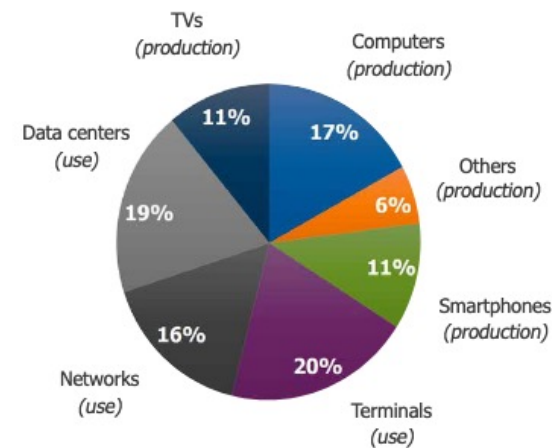
Where does the environmental footprint of IT come from – example of a smartphone

FACTS & FIGURES

Producing a smartphone weighing 140 grams (approximately 0.3086 lb) demands about 0.700 GJ of primary energy whereas, according to ADEME, about 85 GJ to produce a gasoline powered car weighing 1,400 kg (approximately 3,086 lb) (ADEME, 2013). Therefore, it is necessary to consume 80 times more energy to produce "a gram of smartphone" than to produce "a gram of car". **It is noteworthy that miniaturization also increases energy consumption during recycling, since the energy needed to separate the metals increases as a function of the complexity of the assembly.**

DER - Digital environmental repository : Production Phase					
Impacts	Hadwares				
	Laptop	Smart-phone	Server (Data center)	Connected TV	
Primary Energy (MJ)	6640	717	/	/	
GHG (kgCO _{2e})	514	61	588	441	
Metals	Gallium [Ga] (mg)	8	0.5	/	200
	Indium [In] (mg)	20	7	/	12,000
	Tantalum [Ta] (mg)	500	50	/	/
	Copper [Cu] (mg)	170,000	20,000	/	885,000
	Cobalt [Co] (mg)	12,000	6,000	/	/
	Palladium [Pd] (mg)	1	5	/	/
Ore Extracted Volume (L)	7	2	/	200	

[Source: "[Lean ICT Materials] DER", tab "DER Prod Phase". Produced by The Shift Project]



Distribution of energy consumption per source for the production and use of digital devices in 2017.

[Source: The Shift Project 2018, as of Andrae & Edler 2015]

The environmental footprint of research activities: challenges, epistemological consequences

- Digital technologies: part of the research process
- Academic dynamics of innovation and competition: encourages polluting practices like excessive plane travel, redeployment of technologies similar to existing ones, splitting work in project units for fundability, concurrence instead of collaboration, not pulling resources together, falling prey to big tech.
- Core practices of Academic activity are embedded in a system that is incoherent with the scientific reality of the polycrisis => cognitive dissonance, lack of leverage
- Instead, we should consider the consequences for our understanding of our research practices and of our discipline

Three elements in historian work

- relying on archives & our own archival practice
- digitizing archival material
- analyzing it with digital tools

What archiving is about

- Historians rely on archives
- Archives as institutions have a dedicated mission according to which they select material, preserve and record it
- Archives as written pieces of paper require specific heat, humidity, light conditions to be preserved from material decay.
- Loss has always been part of the archiving process: some losses intentional and documented, some losses circumstantial
- Political instrumentalization of loss inherent to institutionalized archival practice: mechanisms of invisibilization

Authority

- Archiving one iteration of a text over another: a choice, expression of an authority
- Scholarship is in part such an authoritative gesture – taking responsibility for knowledge creation, transmission but also valorization of a narrative over another (e.g., narrative of powerful white males)
- Archivists and scholars have a different approach to authority over information -> mostly not a problem in an analogue context, more complicated when it comes to curating data.

What is different with digital archiving?

- Memory institutions: champions of key concepts of digital archiving:
 - have always worked with metadata (not always standardized though)
 - record versions
 - document selection criteria
- Scholars on the other hand are not always familiar with the notions of metadata, versioning and documentation
- First intellectual hurdle: consider your research corpus as data 😊

Digital archiving is not only done by archives

- Archiving: central to the work with research data
- Scholars have to concern themselves with the archiving of their project data
- What is long term archiving anyway?
 - different answers possible (10, 20 years – very short compared to analogue archiving)
 - different accessibility possible (cold/warm, licensing)
 - online findability? Role of catalogues & thesauri
- Who should be archiving what and to what end?

In practice

- Digital Scholarly Edition *Letters and Texts. Intellectual Berlin around 1800*
- www.berliner-intellektuelle.eu
- For 10 years, carried by the idea (illusion?) that it contributes to advancing science (following Open Science and FAIR principles)

LETTERS AND TEXTS: INTELLECTUAL BERLIN AROUND 1800

This edition presents texts that were until this day either completely unpublished or published in an abridged form. Conceived and developed in cooperation with the [archives](#) that preserve them, this edition provides a digital access to the original manuscripts. The editorial and scholarly work on the manuscripts is realized by the junior research group "[Berlin intellectuals 1800-1830](#)".

The junior research group led by [Anne Baillot](#) works on the participation of scholars, writers both male and female, and publishers in public life in Berlin around 1800: What forms does this kind of participation take on? What significance does it have on a political, on a literary level? The question of the strategies developed to communicate in that context is of major importance. It is a goal of this edition to explore them.

TEXT CORPORA

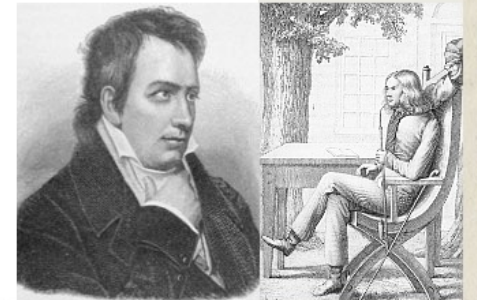
The texts edited here shed light on intellectual life in Berlin in the late 18th and early 19th century. They give insight on the genesis of German Romantic Literature, but also into idea and culture transfers such as they can be retraced in this politically and literary troubled period in the Prussian capital city.

This edition is conceived to put at the core of the project the relationships between the many intellectual circles that were at work in Berlin around 1800: university, academies, societies, Salons, publishers, journals. The users are welcome to use all the [search options](#) developed to that aim.

Authors
Genres
Topics
Repositories
Dates

Search
Editorial guidelines
Partners
Imprint
Nachlassprojekt

de | en | fr



Why the environmental footprint of my digital edition has been a burden

- The dream of an universal Open Access comes with the assumption that every interested reader has a good internet connection, a device equipped with the necessary hardware and software, data literacy, language competence
- => a Eurocentric view, completely stuck in a backward Enlightenment approach
- Basically I have been doing the opposite of what I wanted to do, eating away planetary resources in order to provide close to useless information accessible only to the elite

How to make a digital scholarly edition sustainable

- 3 main elements to leverage: the scans, the transcription and the database structure underlying the edition
- Transcription output in TEI-XML: lightweight in term of environmental footprint
- Move to TEIPublisher App: reduces impact of the overall structure – pages only generated when needed
- But the scans!
- And does it need to be available 24/7?

Challenges in the digitization of archival material

- Tendency to digitize in highest possible quality, to save everything in multiple duplication, based on the idea that we should archive as much as possible in as good a quality as possible because we have the capacity
- Collaboration with Memory Institutions still challenging when it comes to digitizing historical material: information channels, metadata exchange (standardization, authority issues), selection and quality criteria
- Not as simple to optimize/improve at it seems

And then came AI

- Popularization of AI and LLM: new orders of magnitude in the work on data.
 - Data centers used for generative AI: resource-intensive in their production+use (important impact on water+biodiversity)
 - Useful in automatic text recognition of digitized manuscripts: saves time, work
 - But ChatGPT for instance also presented by Universities as “another tool we have at your disposal” for students and staff
- > another interesting case of cognitive dissonance. We know AI is terrible for the planetary resources. But we (we in the Northern hemisphere) are encouraged to use it (productivity increase).

Sharing services – but which ones

- Use an existing model or train one's own, what has the biggest footprint?
➔ NOT the right question to ask
- Use of AI has a massive environmental footprint, also on people involved in the production of the material (data centers, cables, etc.).
➔ Question to ask: Do you need to use it at all?
- We need shared solutions but ethical ones: role of public infrastructures is central in achieving sustainability
- And we need to move away from a culture of productivity and competition

For a culture of reuse (Open Science and FAIR principles are meaningful after all, thank goodness)

- Infrastructure sharing divides the environmental footprint by the number of users, also fosters reuse (of course it doesn't look as good since all users have to work within the same framework. But at least we're not eating the planet away)
- Data sharing => standards and resource-saving formats are key
- Models can also be shared and reused, as is the case in the context of automatic text recognition
- Next step: making actionable workflows reusable

For the Historians of tomorrow: It is the data that matters

- Thinking of the post-docs and PhD candidates that will want to do research when I am long gone: what can I archive for them that would make sense?
 - It is all about the data:
 - quality data
 - actionable, reusable, contextualized (explain selection criteria!)
 - as resource-saving and as informative as possible
 - and a framework for collaborative approaches to data curation
- > our approach to archiving low-resource reusable data, models and workflows is essential. Why don't we focus on it more?**
- The DHd working group Greening DH has issued good practices guidelines for data management with concrete suggestions: https://dhd-greening.github.io/rdm/empfehlungen_dmp

Towards digital sobriety

- We will not reach sustainability without sobriety (at best), in the sense of sufficiency
- GHG and water budgets will eventually apply to our institutes
- We can either wait for it to happen or prepare for it:
 - assess costs and benefits of our use of digital technologies (using for instance efficiency measures)
 - think of archiving as a gift to the coming generations, & how to wrap up the parcel as nicely as possible (there is beauty in this!)
 - slow down our crazy pace



Thank you for your attention!