

# Historical Source Criticism in the Digital Age

18.12.2017

University of Luxembourg

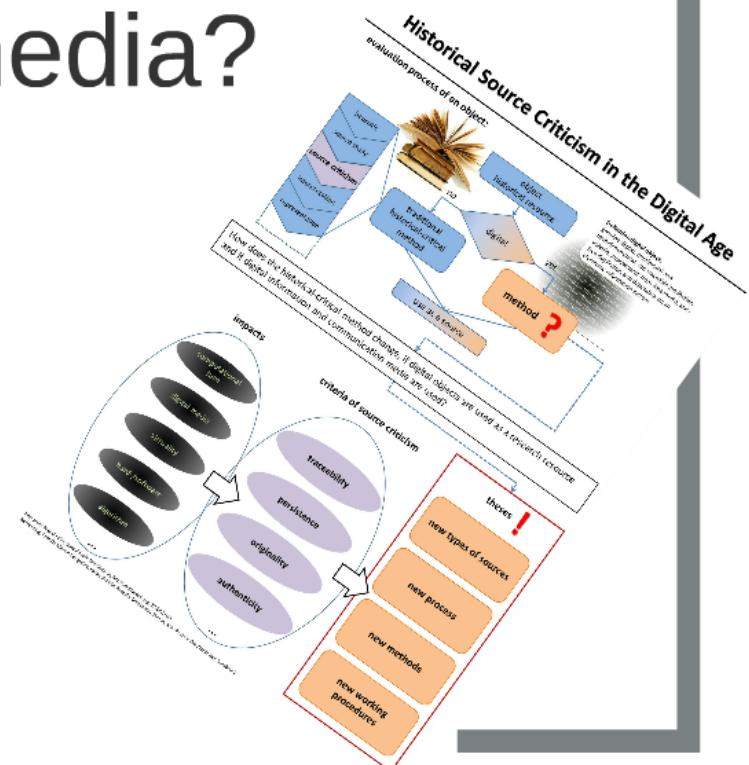


# Table of contents:

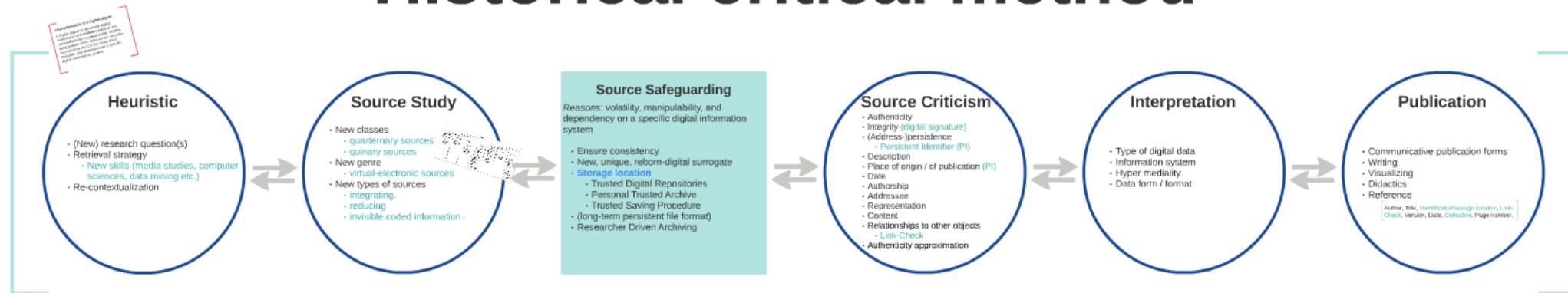
- Doctoral thesis (findings)
  - new kind of source
  - new process step
  - new skills
- Archives
  - new challenges
  - standards, collaboration
- Conclusion

# Research question:

How does historical source criticism change through the use of digital objects as a research resource as well as digital information and communication media?



# Historical critical method



Heuristic

Source Study

Source Safeguarding

Source Criticism

Interpretation

Publication

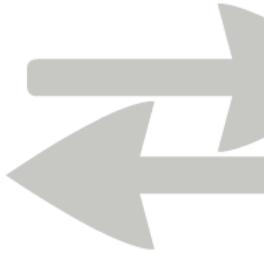
## Characteristics of a digital object

A digital object is *genuinely digital, multimodal and multidimensional, not comprehensibly manipulatable, volatile, independent of the data carrier, lossless reproducible and (at the same time) reusable and dependent on a specific, digital information system.*



# Heuristic

- (New) research question(s)
- Retrieval strategy
  - New skills (media studies, computer sciences, data mining etc.)
- Re-contextualization



# Source Study

- New classes
    - quarternary sources
    - quinary sources
  - New genre
    - virtual-electronic sources
  - New types of sources
    - integrating
    - reducing
    - invisible coded information

Elemente im organischen Raum		Material / physisch		Oberfläche		Werkstoff und / vorweg	
Umwelt	Umwelt	Biozid	Schiffchen	Akkord ausgewählt	Zustände	Elektronisch	
- adaptative - Rad- - Wachstums- - Wettbewerbs- - Fertigkeiten	- hochwertige - Monuments - Umweltarten	- Pflanzenscho- - Gesteins- - Körner- - Fasern- - Rinde- - Blätter- - Stämmen	- Mäuse- - Insekten- - Vögel- - Kinder- - Ameisen- - Chlamydia - Mikroben - Mücken und - Fliegen - Parasiten - Bakterien - Viren - Geschäftsfallen	- Erinnerung - Sitz- - Schwimmbein - Frucht- - Liege- - Schreie - Tint., Musik	- Silber- - Prozesse - Institutionen	- Darstellung - Graphiken - dynamisch - statisch - Wettbewerbs- - Rad- - Wachstums- - Fertigkeiten	- Verlust - Dynamik: - Bild- - Code- - Unterhal- - tungs- - Elementen

# Ein Bild sagt mehr...

⌚ 13. Juli 2015   📁 Off Topic, Problemfelder, Publikation   🔍 digitale Quelle, Hypertext, physisch, Wissenschaftliches Arbeiten

Kürzlich bin ich auf ein Bild gestossen, das sehr gut visualisiert, in welchem Dilemma die (Geschichts-)Wissenschaft an vielen Universitäten sich befindet:



Zwar wollen wir die digitalen (0 und 1) Möglichkeiten nutzen, sind aber oft an bestehende Restriktionen und Vorgaben gebunden, welche auf analogen Publikationsformen (Papierblatt) beruhen.

Diese Problematik wurde auf verschiedenen Veranstaltungen und in diversen Beiträgen diskutiert, weshalb auf weitere Ausführungen verzichtet wird. Aber das Bild musste hier publiziert werden, weil es das Thema pointiert darstellt.

Bild: PANTOGRAPH

# Source Study

- New classes
    - quarternary sources
    - quinary sources
  - New genre
    - virtual-electronic sources
  - New types of sources
    - integrating
    - reducing
    - invisible coded information

## Beispiel: mouseover Event-Handler

Bewegen Sie die Maus auf den Button und wieder weg, um den Event-Handler in Aktion zu sehen:

Klick mich!

## Beispiel: mouseover Event-Handler

Bewegen Sie die Maus auf den Button und wieder weg, um den Event-Handler in Aktion zu sehen:

Drüber! 

Ich bin dynamisch!

# Source Safeguarding

*Reasons:* volatility, manipulability, and dependency on a specific digital information system

- Ensure consistency
- New, unique, reborn-digital surrogate
- **Storage location**
  - Trusted Digital Repositories
  - Personal Trusted Archive
  - Trusted Saving Procedure
- (long-term persistent file format)
- Researcher Driven Archiving

# Source Criticism

- Authenticity
- Integrity (**digital signature**)
- (Address-)persistence
  - **Persistent Identifier (PI)**
- Description
- Place of origin / of publication (**PI**)
- Date
- Authorship
- Addressee
- Representation
- Content
- Relationships to other objects
  - **Link-Check**
- Authenticity approximation

# Never trust a digital object !

Authenticity approximation:  
probability of the authenticity of a digital  
object after its verification

- subjective value
- 7 rated and weighted issues

The screenshot shows a tweet from David M. Berry (@berrydm) dated 07 Oct 2015 at 03:10 AM. The tweet reads: "@pfhist digital source criticism is overrated http://bit.ly/1qdfLOj #postdigital". The tweet has 97 retweets and 50 favorites. Below the tweet is a table showing the 7 rated and weighted issues:

Issue	Comment	A	G	M
Integrity	<ul style="list-style-type: none"><li>Keine Integritätsmerkmale, Angaben nur von Twitter</li><li>Tweets nach Versand grundsätzlich nicht veränderbar</li><li>twitter.com: vertrauenswürdig (whois)</li><li>freier Zugriff</li><li>DRM nicht erkennbar</li></ul>	80	5	400
Persistence	<ul style="list-style-type: none"><li>Im Archiv von Twitter verfügbar</li><li>private Unternehmung</li><li>Tweet bei Suche im Archiv nicht wiedergefunden</li></ul>	60	1	60
Date	<ul style="list-style-type: none"><li>07.10.2015, 03.10 Uhr</li><li>Jeder Tweet wird vom Dienst automatisch mit dem Publikationsdatum versehen.</li></ul>	100	4	400
Authorship	<ul style="list-style-type: none"><li>David Berry ist bekannt als 'digital humanist' – plausibel</li><li>Herkunft von Berry plausibel (Bild stimmt mit Person überein – persönlich getroffen)</li><li>Berry kennt pfhist persönlich – plausibel</li></ul>	100	4	400
Addressee	<ul style="list-style-type: none"><li>Twitterer direkt (pfhist) – plausibel, da persönlich bekannt</li><li>Interessierte Personen (Followers von berrydm und am Thema interessierte postdigital) – Thema plausibel</li><li>Twitternutzer weltweit</li></ul>	100	1	100
Content	<ul style="list-style-type: none"><li>kritische Anmerkungen</li></ul>			

Issue	Comment	A	G	M
Integrity	<ul style="list-style-type: none"><li>Keine Integritätsmerkmale, Angaben nur von Twitter</li><li>Tweets nach Versand grundsätzlich nicht veränderbar</li><li>twitter.com: vertrauenswürdig (whois)</li><li>freier Zugriff</li><li>DRM nicht erkennbar</li></ul>	80	5	400
Persistence	<ul style="list-style-type: none"><li>Im Archiv von Twitter verfügbar</li><li>private Unternehmung</li><li>Tweet bei Suche im Archiv nicht wiedergefunden</li></ul>	60	1	60
Date	<ul style="list-style-type: none"><li>07.10.2015, 03.10 Uhr</li><li>Jeder Tweet wird vom Dienst automatisch mit dem Publikationsdatum versehen.</li></ul>	100	4	400
Authorship	<ul style="list-style-type: none"><li>David Berry ist bekannt als 'digital humanist' – plausibel</li><li>Herkunft von Berry plausibel (Bild stimmt mit Person überein – persönlich getroffen)</li><li>Berry kennt pfhist persönlich – plausibel</li></ul>	100	4	400
Addressee	<ul style="list-style-type: none"><li>Twitterer direkt (pfhist) – plausibel, da persönlich bekannt</li><li>Interessierte Personen (Followers von berrydm und am Thema interessierte postdigital) – Thema plausibel</li><li>Twitternutzer weltweit</li></ul>	100	1	100
Content	<ul style="list-style-type: none"><li>kritische Anmerkungen</li></ul>			

# and weighted issues



**David M. Berry**  
@berrydm

Follow

@pfhist digital source criticism is overrated  
<http://bit.ly/1qdfLOj> #postdigital

RETWEETS 97 FAVORITES 50

SEL ENA MIC HAE AALI YAH NICH OLA

Reply Retweet Favorite More

03:10 AM - 07 Oct 2015

Issue	Comment	A	G	M
Integrity	<ul style="list-style-type: none"><li>• Keine Integritätsmerkmale, Angaben nur von Twitter</li><li>• Tweets nach Versand grundsätzlich nicht veränderbar</li><li>• twitter.com: vertrauenswürdig (whois)</li><li>• freier Zugriff</li><li>• DRM nicht erkennbar</li></ul>	80	5	400

Issue	Comment	A	G	M
Integrity	<ul style="list-style-type: none"> <li>Keine Integritätsmerkmale, Angaben nur von Twitter</li> <li>Tweets nach Versand grundsätzlich nicht veränderbar</li> <li>twitter.com: vertrauenswürdig (whois)</li> <li>freier Zugriff</li> <li>DRM nicht erkennbar</li> </ul>	80	5	400
Persistence	<ul style="list-style-type: none"> <li>Im Archiv von Twitter verfügbar</li> <li>private Unternehmung</li> <li>Tweet bei Suche im Archiv nicht wiedergefunden</li> </ul>	60	1	60
Date	<ul style="list-style-type: none"> <li>07.10.2015, 03.10 Uhr</li> <li>Jeder Tweet wird vom Dienst automatisch mit dem Publikationsdatum versehen.</li> </ul>	100	4	400
Authorship	<ul style="list-style-type: none"> <li>David Berry ist bekannt als 'digital humanist' – plausibel</li> <li>Herkunft von Berry plausibel (Bild stimmt mit Person überein – persönlich getroffen)</li> <li>Berry kennt pfhist persönlich – plausibel</li> </ul>	100	4	400
Addressee	<ul style="list-style-type: none"> <li>Twitterer direkt (pfhist) – plausibel, da persönlich bekannt</li> <li>interessierte Personen (Follower von berrydm und am Thema Interessierte postdigital) – Thema plausibel</li> <li>Twitternutzer weltweit</li> </ul>	100	1	100
Content	<ul style="list-style-type: none"> <li>kritische Aussage von Berry plausibel</li> <li>Unter der URL befindet sich eine nicht direkt mit dem Thema verwandte Webseite (falsch kopiert?):            <a href="https://www.zotero.org/support/quick_start_guide">https://www.zotero.org/support/quick_start_guide</a> </li> </ul>	70	3	210
Relationships	<ul style="list-style-type: none"> <li>Verwendung eines URL-Shorteners: selten von Berry genutzt</li> <li>Beziehungen zu anderen Nutzern (Retweets) plausibel, Berry hat über dreitausend Follower</li> <li>verlinkte URL verdächtig</li> </ul>	70	1	70
Sum			19	1640
Authenticity approximation				86.32%

(s) plausibel,

(s) plausibel,			
	19	1640	
	<b>86.32%</b>		



<http://tweetfake.com/>

# Interpretation

- Type of digital data
- Information system
- Hyper mediality
- Data form / format

# Publication

- Communicative publication forms
- Writing
- Visualizing
- Didactics
- Reference

[Author, Title, Identifier/Storage location, Link-Check, Version, Date, Collection, Page number.]

# **Archives: Problems and approaches**

Receiving digital objects

- Data volume
- Proprietary software

Preservation

- Data storage media
- Data formats

Access

# Data Volume

IDC: 163 zebibytes in 2025

163'000'000'000'000'000'000 bytes

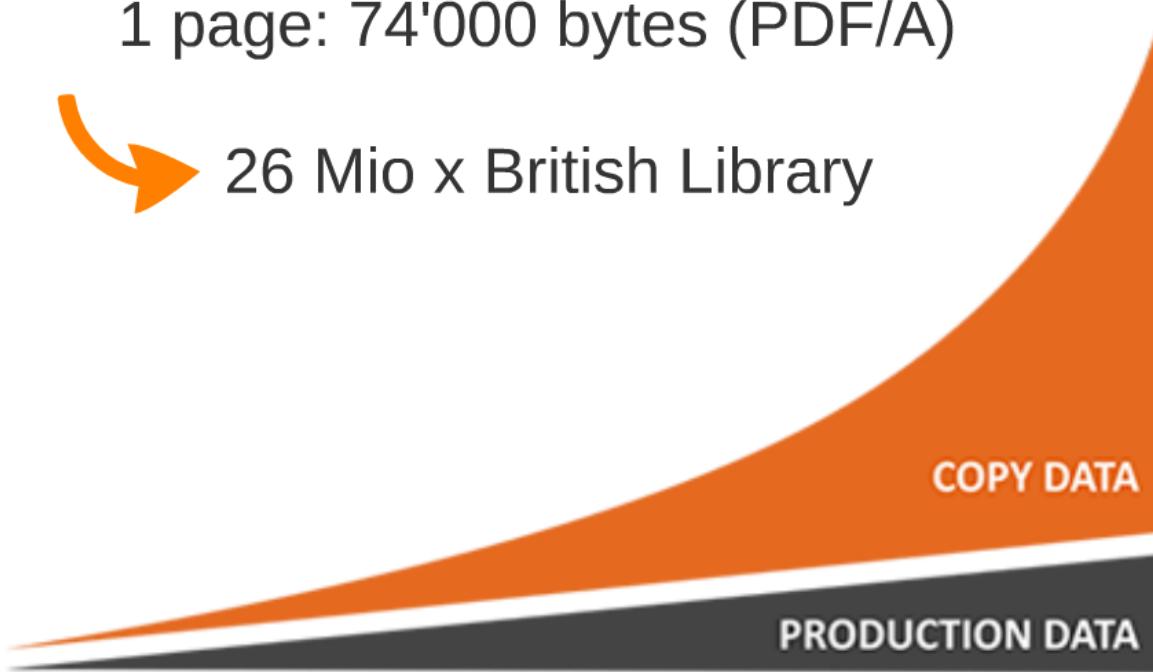
British Library: 170 Mio books

1 book: 500 pages

1 page: 74'000 bytes (PDF/A)



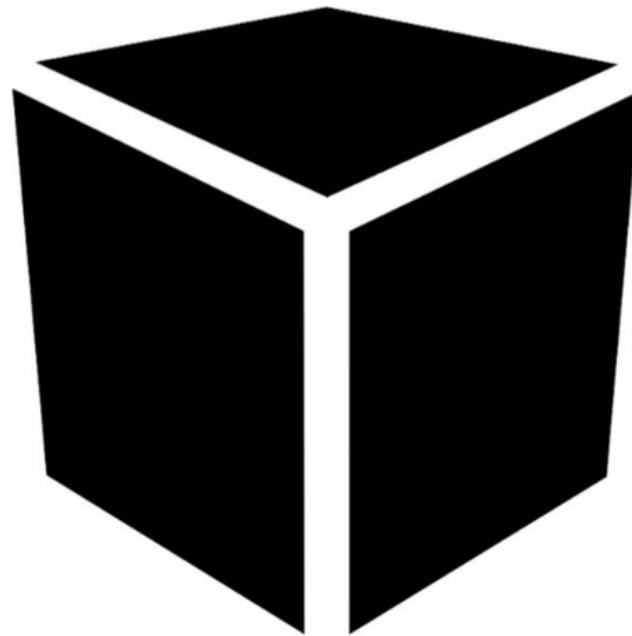
26 Mio x British Library



Receiving digital objects

# Proprietary Software

- Export?
- Data format?



Software

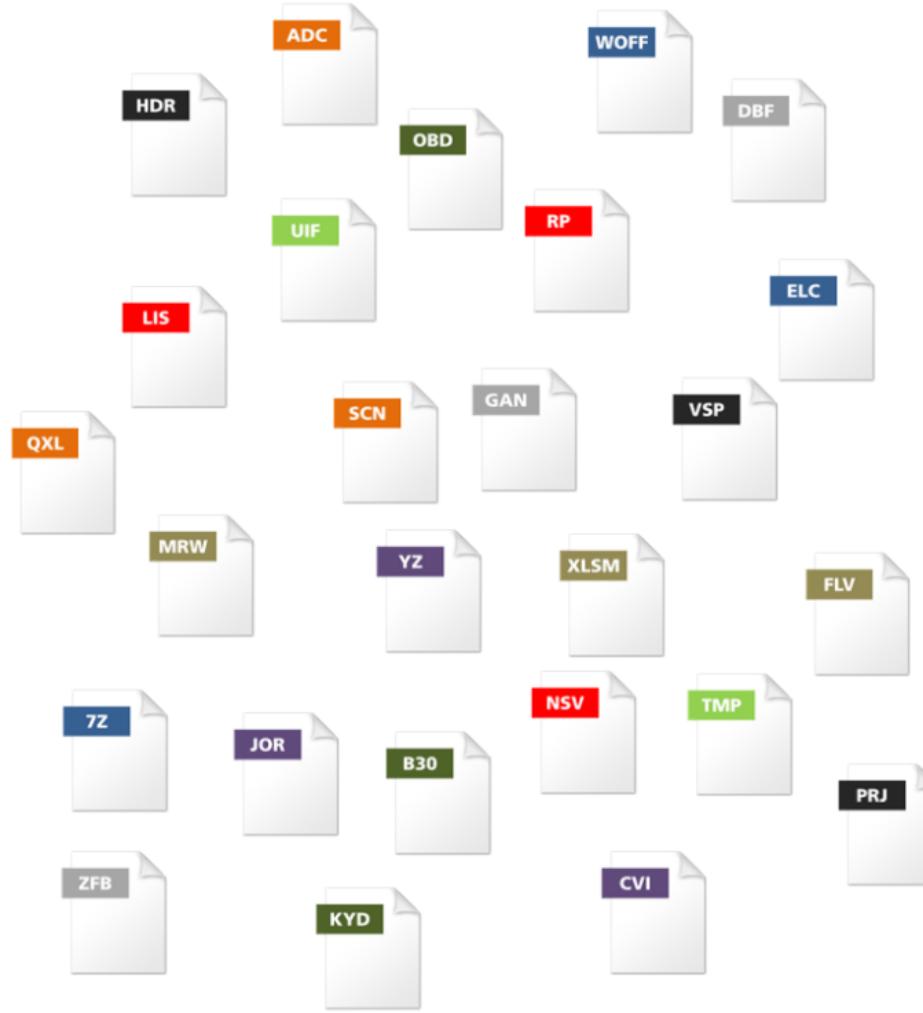
# Data storage media

- Durability
- Readability



# Data format

- Obsolescence



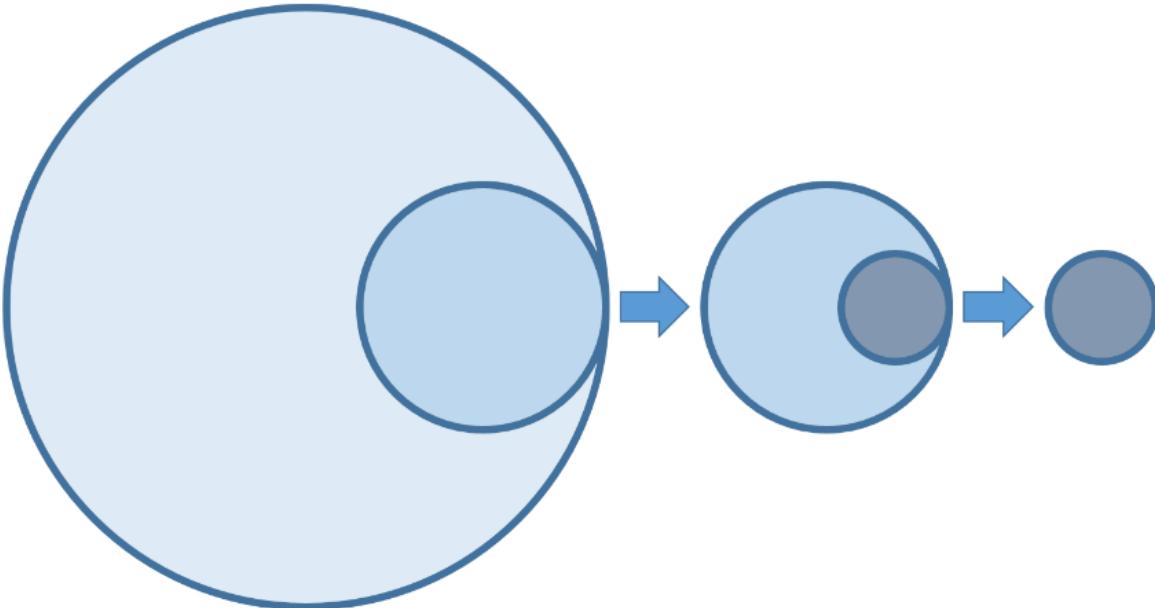
# Access

- Who (permission) ?
- When (blocking period) ?
- How (system) ?
- Format ?
- Copyright

Receiving digital objects

## Data Volume

- Preselection

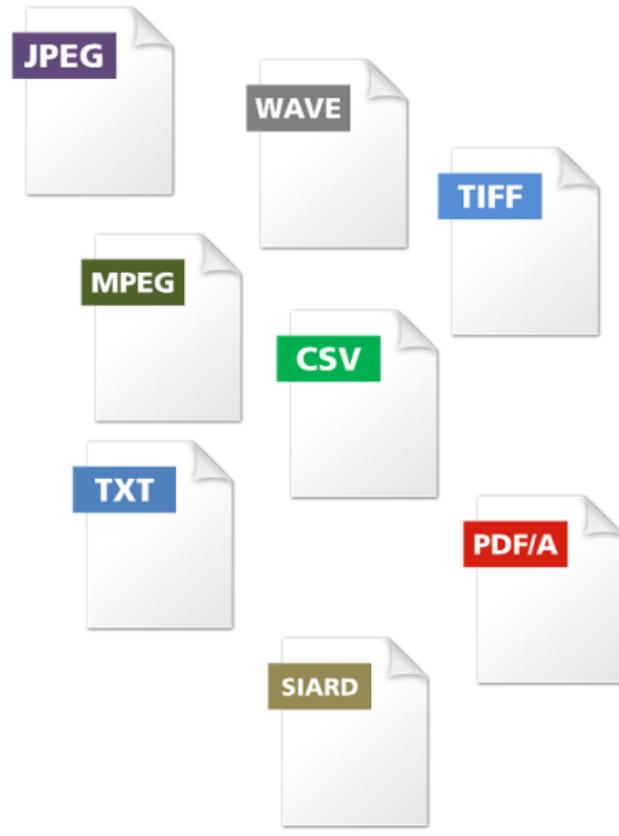


- Researcher Driven Archiving

Receiving digital objects

# Proprietary Software

- Open Source
- Export in accepted data formats



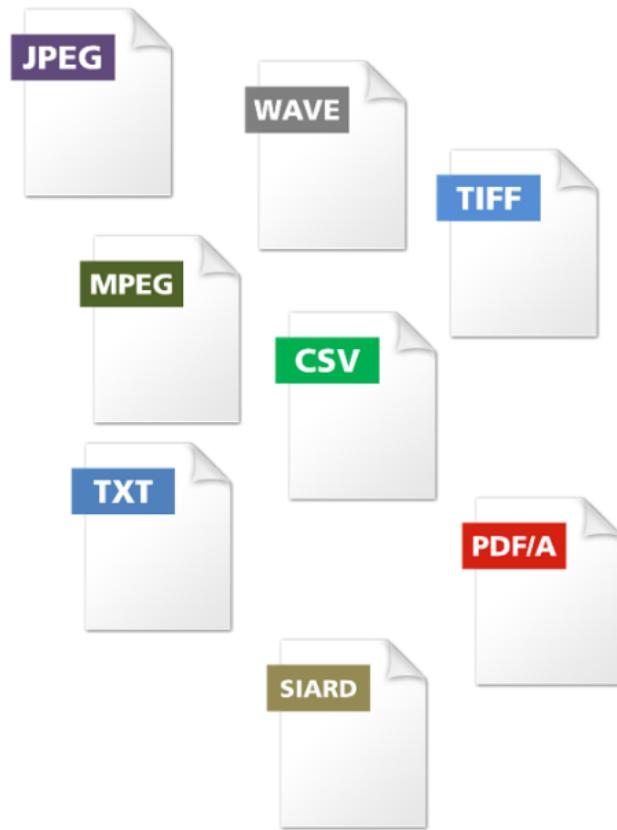
# Data storage media

- Migration every 3-5 years !



# Data formats

- simple, long-term
- Migration when necessary



# Access

Tendency to grant access only inhouse, especially because of data and privacy protection.

With a simple online search, a disproportionate amount of data can be found, correlated and (mis-) interpreted.

# *A lot of metadata*

- information about:
  - the 'original' information system
  - how was the object generated
  - how was the object used
- every migration
- every access?
- Linking

# Standards

Only collaborative solutions and the use of standards can ensure the continued use of digital research resources across all branches of science.

# Conclusion

New process step

New skills

New methods and procedures

New infrastructure

Standards, standards, standards

# THANKS !

01010001 01110101 01100101 01110011 01110100 01101001 01101111 01101110 01110011 00100000 00111111

(Questions ?)

Pascal Föhr

[hsc.hypotheses.org](https://hsc.hypotheses.org)  
PFHist  
[pfhist@foehr.ch](mailto:pfhist@foehr.ch)

