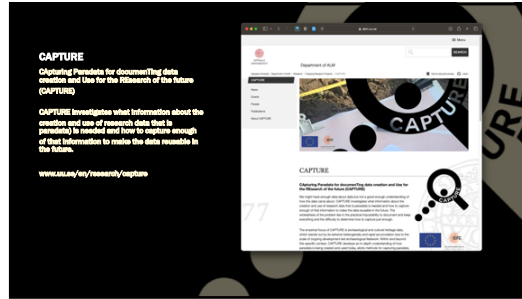
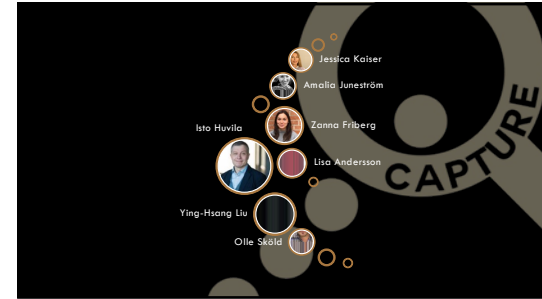




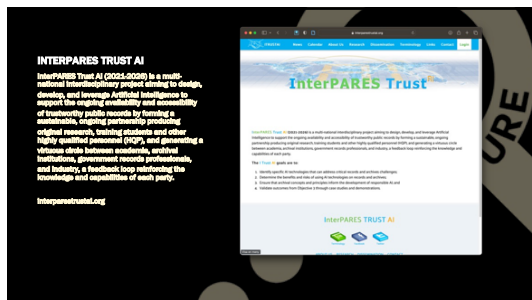
1



2



3



4



5



6

Data sharing is FAIR
Wilkinson et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, 160168.

<p>Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.</p> <p>FINDABLE</p>	<p>Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.</p> <p>ACCESSIBLE</p>
<p>Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.</p> <p>INTEROPERABLE</p>	<p>Data and collections have a clear usage license and provide accurate information on provenance.</p> <p>REUSABLE</p>

https://www.nature.com/articles/sdata0160168/1/1/cover-image

7

Data is MEAN!
**miscellaneous,
 exceptional, arbitrary and
 nonconformist**

Huvila, Isto (2017). Being FAIR when archaeological information is MEAN. Presentation at Centre for Digital Heritage 2017 conference, Leiden.

8

**Data is kept for
 (different) reasons
 and purposes.**

9

Keeping data for

- **Record-keeping (evidence)**
 - Proof that something had happened
 - Validation of earlier results, decisions
 - Contesting earlier decisions
 - ...
- **Knowledge production (research)**
- **Applied purposes**
 - Product development
 - Policy-making
 - ...
- ...

10

Problems in using data

- Data is difficult to find
- Many people like to work with their own data
- Data is not well-documented
- Context and provenance of data are not well-documented
- Available documentation is not especially relevant for its users

11

Problems in managing data

- (Relevant) data is difficult to identify
- Data is spread throughout different systems and difficult to capture
- Data is not well-documented
- Context and provenance of data are not well-documented
- ...

12

From archives perspective:
Data are not records

13

**If AI is an answer,
what is the question?**

14

AI as an opportunity
AI for data management

15

AI as a complication
AI generated data and records

16

**Records on
automatic
decision-making.**

17

What is AI capable of doing?
processing large quantities of information
calculating and predicting
learning and adapting responses
recognising and classifying objects

18

making data searchable and retrievable
classification
appraisal
handling sensitive information
metadata extraction and generation
automatic content extraction and indexing
collections development

19

Making searchable

Search Finnish Court Records
 Search and browse Finnish court records with the help of algorithmic text recognition

20

Classification

AEOLIAN Network
 Home About Blog Outcomes Events Case Studies Team Join

Case Study 1: The National Archives (UK)

Events

- Online Workshop 6: New Horizons in AI and Machine Learning for Libraries and Archives. This workshop will be held online and hosted by the University...
- Online Workshop 5: Making More Sense With Machine AI/ML Methods for Interrogating and Interrogating...

21

Appraisal

THE NATIONAL ARCHIVES

Using AI for digital selection in government

Also in Research and collaboration

Substituting the cluttered digital memory

22

"... aid the task of record selection in semi-structured and unstructured collections"

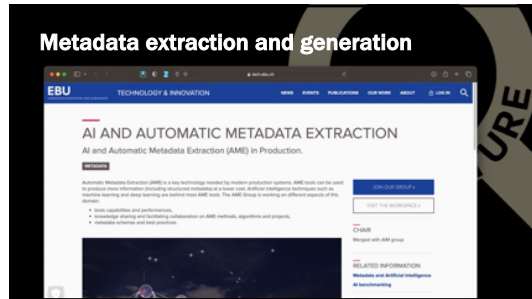
National Archives UK

23

Handling sensitive information

- Personal information content assessment (Suderman in ITRUST AI)
- Case Study on Extraction and Identification of Records containing Personal Data and Sensitive Personal Data for Long Term Preservation (Barnard in ITRUST AI)
- ...

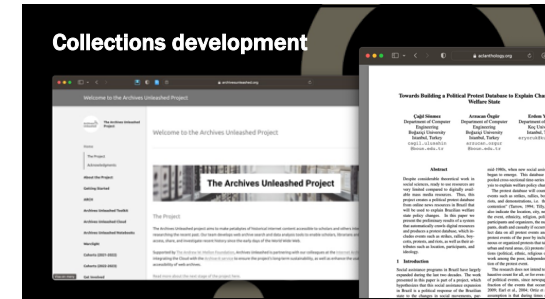
24



25



26



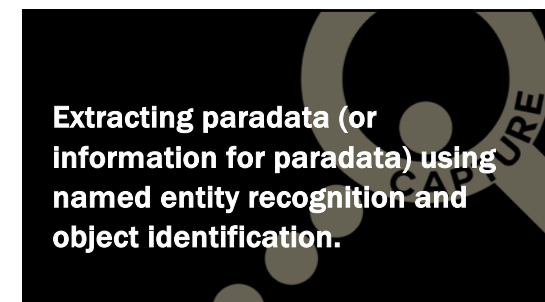
27



28



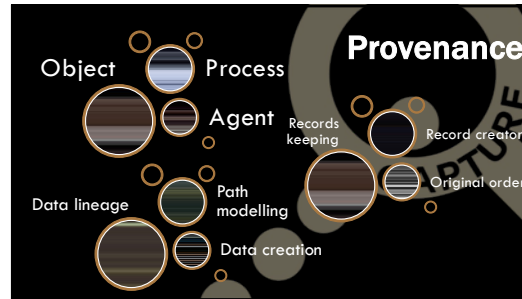
29



30



31



32



33

A slide with a black background and a 'CAPTURE' watermark. The text reads: "Duranti: 'Provenance metadata is a part of record. Paradata is not.'"

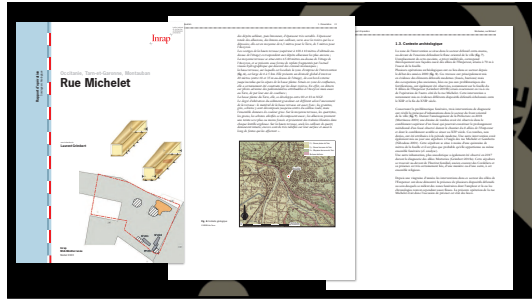
34

A slide with a black background and a 'CAPTURE' watermark. The text reads: "Wait!?" followed by "If paradata is not a part of the record...?" in orange.

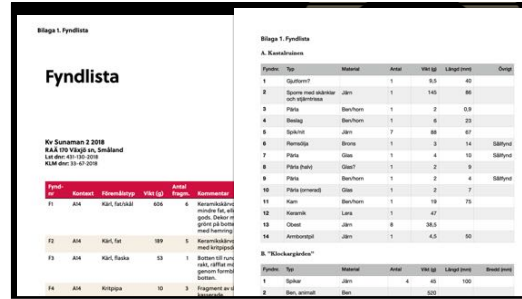
35

A slide with a black background and a 'CAPTURE' watermark. The text reads: "Trials on archaeological investigation reports."

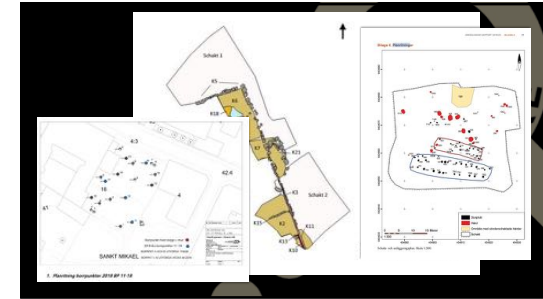
36



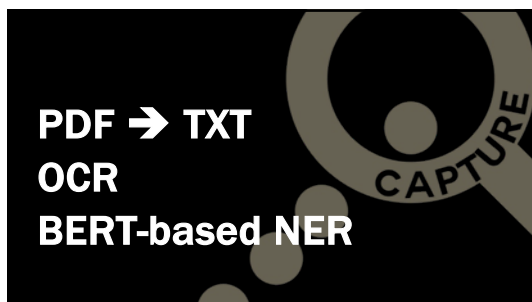
37



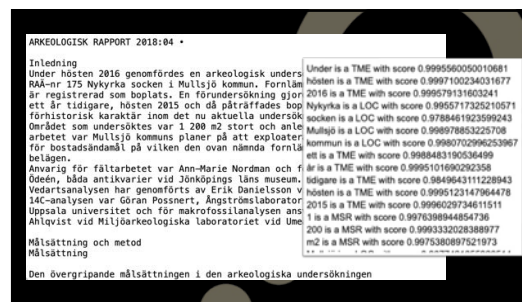
38



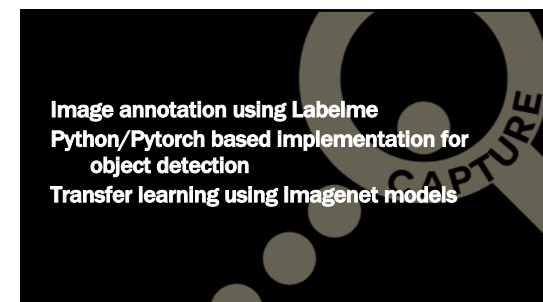
39



40



41



42



43

Possible to identify information relevant as (a basis for) paradata.

44

Much of the relevant information is contextual and inferential.

45

Better understanding of paradata phenomenon and its instances needed. (but still promising)

46

Conclusions?

47

AI as a problem and a solution
Ethics
Risks
Bias

48

Current guidance and legislation are limited.

49

**Critical thinking.
Distinguishing AI and non-AI interventions.**

50

Opportunity and complication?
Efficiency
Difficulty to anticipate technology development

51

Use of AI In Data Management

Isto Huvila
Department of ALM | Uppsala University
@ihuvila | www.istohuvila.se | isto.huvila@abm.uu.se



UPPSALA UNIVERSITY   This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement No 810210.

52