

Negotiating provenance

@MaxKemman

www.maxkemman.nl

University of Luxembourg



Distinguishing provenance

Interview with history professor

“In the metadata of objects subtleties disappear if you never follow the catalogue card. So we solved that with provenance information, and **that is why it is so important to distinguish between the concept of provenance in computer science and in historical science**”

Distinguishing provenance

- How are 'historical provenance' and 'CS provenance' different?
- Which type(s) of provenance are implemented in digital research infrastructures?
- How does this affect historical practices?

Provenance perspectives

W3C PROV Model Primer

- **Agent-centred** (who generated or manipulated data)
- **Object-centred** (from what documents is data extracted)
- **Process-centred** (what steps were taken to generate data)

Data provenance

- The provenance of a piece of data is the process that led to that piece of data
- **‘where-provenance’** – where does a given piece of data come from
- **‘why-provenance’** – why is it in the database
- **‘how-provenance’** – the structure of proof explaining how an output is derived

Process-centred

Buneman et al. 2001, Moreau 2010

Archival provenance

Dutch Manual 1898

- the ‘**respect des fonds**’ (this principle dictates that archival materials, when transferred to archival custody, remain as distinct collections catalogued and filed according to their creator or office of origin)
- the ‘**respect de l’ordre**’ (this principle demands that records in these distinct collections are maintained in their original order).

Agent-centred & Object-centred
Arguably Anti-Process

Fickers 2012

Interview 1

- “The provenance of the data shows which **processing layers** there have been, and to what extent the data you see before you, what manipulations led to this level of aggregation of the data. So the search trail goes back so to say [...]
- While in historical science, that is in the real world. The **sources**, the question of provenance, takes place in the real world, with **many different actors, with movements in space and time**”

Data provenance → process-centred

Historical provenance → object-centred &
agent-centred wrt the source

Interview 2

“For me it is a reference to the original source. And for a computer scientist it can be a reference to the tool that has been used”

Data provenance → process-centred

History provenance → object-centred

| | Agent | Object | Process |
|-------------------------------------|--|---|--|
| Historical provenance | Dutch Manual Interview 2  | Dutch Manual Interview 1 & 2 | |
| Data provenance | | | Buneman et al Moreau Interview 1 & 2 |
| Today's Infrastructure d provenance | Piek Vossen Maarten v Gompel | Docker? Source files? Ronald Siebes |  Daan Broeder Slava Tykhonov Marjo van Koppen Jan Odijk Menzo Windhouwer Liliana Melgar Piek Vossen Richard Zijdeman Katrien Depuydt Nelleke Oostdijk |

Provenance in historical practice

“The most essential use of provenance in historical method is to locate materials that have been preserved in their original groupings. [...] Access is always available through provenance” (Grigg 1991)

“[T]here is the issue of provenance; can the document be traced back to the office or person who is supposed to have produced it, or could it have been planted?” (Tosh 2015)

Dependence on static provenance

Negotiations

- Archival provenance is anti-process
- Traditional historical method gave little attention to provenance except as
 - Heuristic
 - Authority
 - Assumes static provenance
- Data provenance in contrast is process-centred
 - Covers authority (Nelleke Oostdijk)
- Infrastructured provenance
 - Covers authority as per data provenance
 - Does it cover heuristic?
 - Is it static?

Negotiations

“I never thought about provenance before”

- The relation of the digital copy to the paper source is still debated (Peter Doorn)
 - Problem-oriented
 - Source-oriented
- Historical method will need to include provenance?
 - Habits (Marjo van Koppen)
 - Why are changes made? (Jan Odijk)
 - Reproducibility is not a tradition
- Process-centred digital source criticism?
 - Turtles all the way down...
 - Which parts are meaningful?

*Thank you for
listening*

www.maxkemman.nl
[@MaxKemman](https://twitter.com/MaxKemman)
max.kemman@uni.lu

