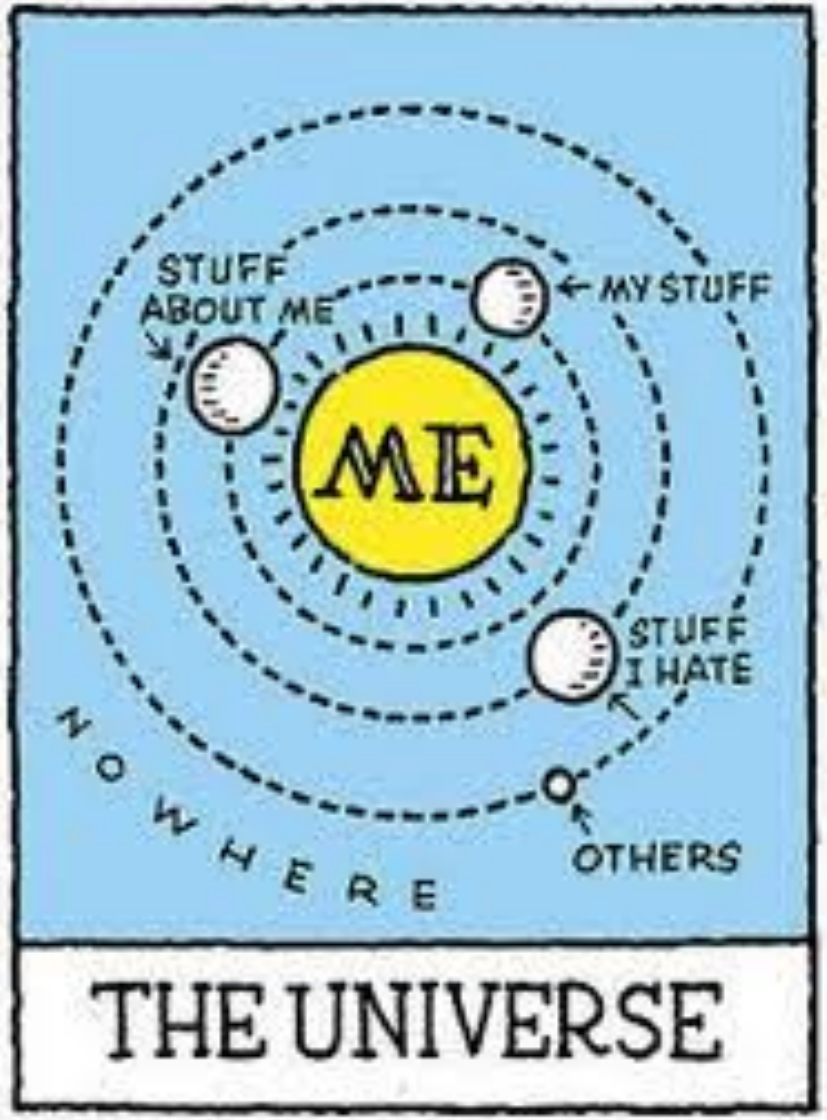
Blockchain to guarantee Authenticity and Integrity of Provenance data

Ronald Siebes
ronald.siebes@dans.knaw.nl

CLARIAH – Provenance workshop - Sept 3rd, 2018


TRUE
RDF

Is that you? - Is this my paper?



Provenance – Authenticity – Integrity

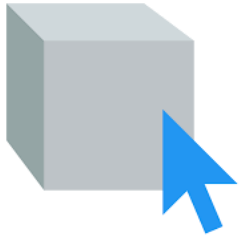
Provenance, broadly speaking, is documentation about the origin, characteristics, and history of an object (artifact); its chain of custody; and its relationship to other objects (artifacts).

Clifford Lynch[1]

[1] https://www.clir.org/pubs/reports/pub92/lynch/

data provenance data

- set of statements (triples) bound to an *artefact* and uttered by *agents*?



Artefact X

Agent Smith

Artefact X. prov-o:createdBy foaf:Agent-Neo
Artefact X. prov-o:creationDate ”Sept 8th, 2019”

Agent Neo

Artefact X. prov-o:alteredBy foaf:Agent-Smith

There is no truth, there are many

- No a-priori limitation of choice needed when we have a time-stamped monotonically growing Akashic record of utterances (statements – triples) by agents about artifacts

- Things to solve:
  - How to identify agents and artifacts
  - How to efficiently find statements from agents and about artifacts
  - How to persistently store the statements (e.g. danger of single point of failure)
  - How to guarantee nobody can secretly alter the records
  - …

Blockchain!

- The Etherium protocol allows to create a blockchain with custom code
- The idea is that an artifact is somewhere in the Ethereum Virtual Machine represented by its unique alterego (same as in the Matrix),with DNA as the metaphor for identification any relative or clone should by design of the matrix have a different DNA
- The miners in the matrix get Etherium coins for
  - checking the integrity of the artifacts living in the EVM
  - Monitoring the behavior of agents.
  - The way they do it is determined by the custom (but fixed) programs. The miners only have to spend energy to execute these programs.

Blockchain, ctnd…

- To summarize the idea, the EVM guarantees:
  - Digital artifacts in the ‘real world’ have a unique alterego on the EVM
  - The DNA is the binding element (e.g. MD5 hash, SHA-256, normalization first?)
  - agents can say things about digital artifacts
  - Agents are also uniquely represented (Wallet ID?)
  - Everything is recorded (monotonically increasing, distributed, P2P maintained, transaction log
  - nothing said can be removed
  - An agent cannot deny something what it said
  - An agent can prove that is did not say something

But…what about provenance?

- Since we now have a log of statements by agents on artifacts, we can use that to let agents make provenance statements about artifact
- Nobody can deny a made statement (authenticity)
- Nobody can delete a statement (integrity)