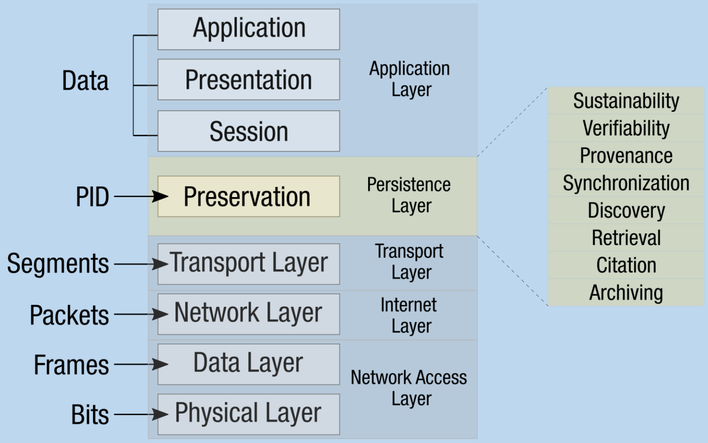
FactStack: Interoperable Data Management and Preservation for the Web and Industry 4.0

Lars Gleim (RWTH Aachen University)



Introduction

The management and preservation of the diverse and constantly evolving data on the Web remain an open challenge to date. Interorganizational and interoperable solutions are needed to power e.g. the digital manufacturing supply chain of the future.

With FactStack, we devised a uniform and interoperable persistence and preservation layer, inspired by principles from research data management, successful version control systems and implementable throughout the traditional data lifecycle. Employing global, persistent identifiers in combination with resource versioning and data provenance, FactStack enables the on-demand integration of persistence and provenance into existing resource-oriented architectures based on open Web technologies and the FactDAG data interoperability model. As such, FactStack facilitates agile, interorganizational and uniform data management and preservation, effectively supporting the creation of provenance-linked knowledge graphs of versioned and preservable resources as part of the classical data lifecycle.

Dated URIs[1]

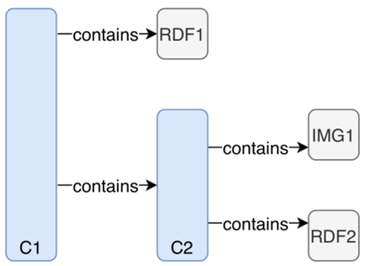
Persistent identification for arbitrary web resources

Identify the immutable state of the resource identified by `<embeddedURI>` at time `<timestamp>` *using dated URI*[1]:

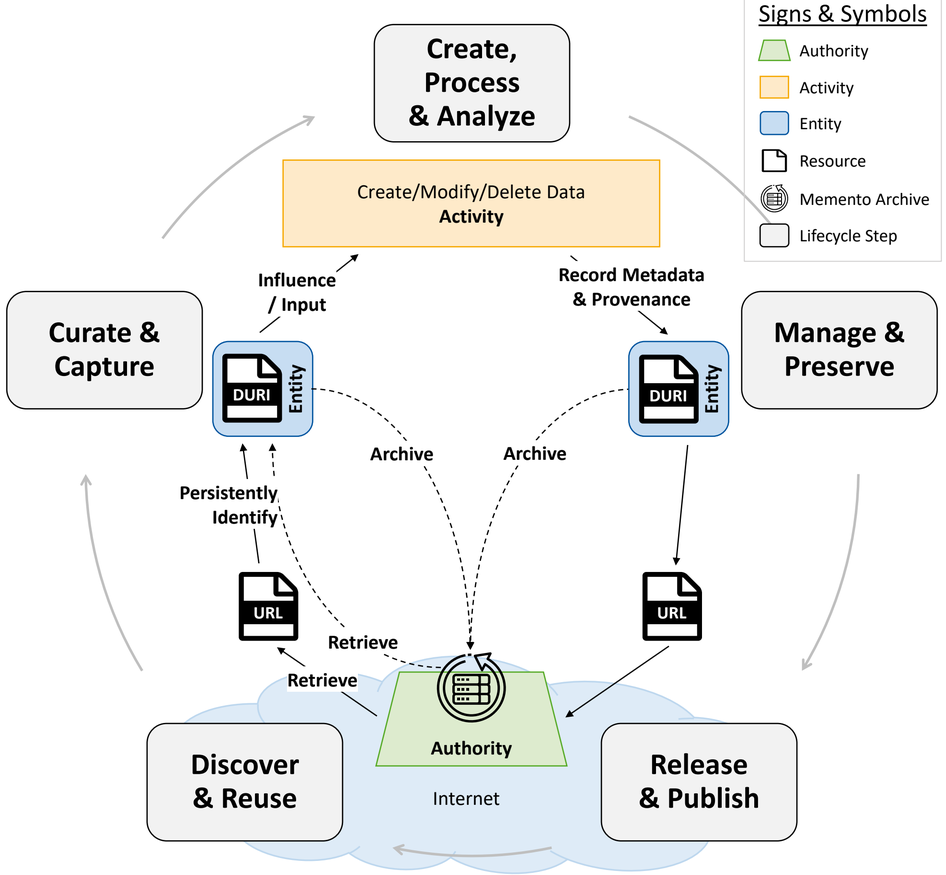
```
duri:<timestamp>:<embeddedURI>
```

`2020-11-03T13:59:59.99Z` `https://example.org/resource/path`

W3C Linked Data Platform[2]



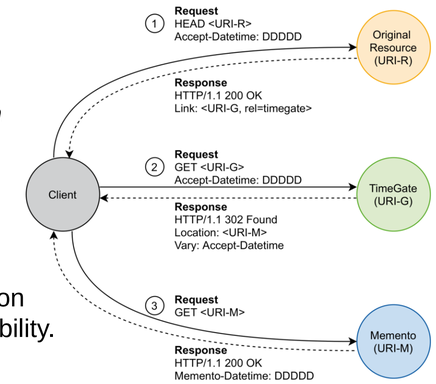
Standardized REST API for data discovery, R/W access, & hierarchical organization.



HTTP Memento Protocol[3]

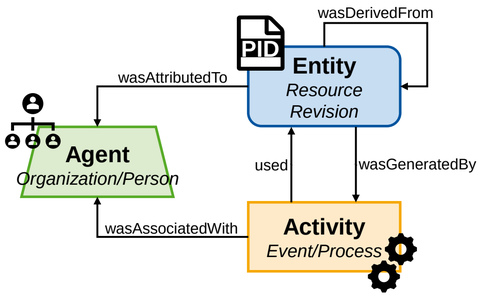
Time-based HTTP content-negotiation

Archive discovery, Memento retrieval, versioning, state synchronization & resource immutability.



W3C PROV Provenance Standard[4]

Metadata about
- data origins,
- influences,
- authors,
- its revisions,
- the processes leading to its creation.



Prototype: https://git.rwth-aachen.de/i5/factdag/factlibjs

[1]Masinter, L.M. (2012). *The 'tdb' and 'duri' URI schemes, based on dated URIs*, https://datatracker.ietf.org/doc/html/draft-masinter-dated-uri-10
[2]Speicher, S., Arwe, J., & Malhotra, A. (2015). *Linked Data Platform 1.0*, https://www.w3.org/TR/ldp/
[3]van de Sompel, H., Nelson, M. L., & Sanderson, R. (2013). *HTTP Framework for Time-Based Access to Resource States – Memento*, RFC7089
[4]Groth, P., & Moreau, L. (2013). *PROV-Overview. W3C Working Group Note*, https://www.w3.org/TR/prov-overview/

Exzellenzcluster „Internet of Production“ (IoP)
www.iop.rwth-aachen.de

Computer Science 5 - Information Systems and Databases
RWTH AACHEN UNIVERSITY