Towards Preservation of semantically enriched Architectural Knowledge

Stefan Dietze, Jakob Beetz, Ujwal Gañiraju, Georgios Katsimpras, Raoul Wessel, René Berndt
Goal

- Methods and tools for **sustainable** long-term preservation of architectural knowledge

Challenges

- **Diversity of data - interoperability**: low-level point clouds & legacy 3D models up to enriched Building Information Models (BIM), **higher-level semantics** and Web data / knowledge

- **Diverse stakeholders**: architects, building operators, urban planners, archivists, ...

- **Building, model and data evolution**: document temporal evolution to prevent information loss
Challenges

- **“Semantic” enrichment of architectural knowledge**: exploiting Web data and knowledge to enrich low-level architectural data.

- **Inconsistent vocabularies**: adopting state of the art (LD) vocabularies and schemas towards sustainability

- **Long-term readability / renderability of architectural models**: addressing digital decay (e.g., due to deprecated file formats) and model evolution

Goals and Challenges (2/2)
**UBO: Universität Bonn**
- Technical Coordinator
- WP4/WP5: change management, shape recognition

**Luleå University of Technology**
- WP8: dissemination/exploitation

**CITA, Center for Information Technology and Architecture Copenhagen**
- WP7: data, evaluation, test

**TUE, Department of the Built Environment, Eindhoven University of Technology**
- WP3: semantics & metadata

**Catenda, SME**
- User perspective, market requirements, evaluation

**Fraunhofer Austria**
- WP2: system specification & integration

**LUH: German National Library of Science and Technology (TIB) & L3S Research Center Hannover**
- Coordinator
- WP3 Semantic Enrichment
- WP6 leader, long-term preservation

**Consortium**

Stefan Dietze (L3S Research Center)
Why interlinking & semantic enrichment?

A very simplistic view on urban planning/architectural lifecycle today

1. research  2. design  3. monitoring (over time)

DURAARK approach - exploiting Web data to help architects and urban planners to answer questions like:

- What’s the **legal, social and environmental context** of a structure (sustainability policies etc)?
- How did **buildings and their contexts** (traffic, surroundings, usage and functionality, popularity, etc) evolve over time?
- How did an architectural change impact **surrounding traffic/environment**? (examples: bridges, airports)
- How did an architectural change impact **popularity and attractiveness** of a building?
- ....
Architectural Data Preservation

Building Information Models (BIM) = structured "Building Model Metadata"

DURAARK Scope

Archival

Semantic Enrichment

Linked Data sets
- Location
- Social Context
- History
- Usage
- Context
- Technical
- Legal

3D Models

Building Models

as-planned
- Architecture
- Structural
- HVAC

as-built
- Surveying

Point Clouds

Semantic Digital Archive

IFC/A

Geometric Enrichment

Point Cloud Data

IFC

Stefan Dietze (L3S Research Center)
- **Semantic enrichment** of low-level architectural models (gradual process)
- **Interlinking** of related models/data (across different abstraction levels, model types, datasets and repositories...)
- **Preservation & temporal analysis**: tracking the evolution of models, buildings and related data
Example: GDR’s People’s Palace - static vs evolving data/links

Social & Semantic Web for enrichment
Challenges

- **Selection** of suitable datasets from wealth of diverse datasets
- **Preservation**: dealing with evolution of distributed datasets (i.e. the semantics & context of the structure/models)
Data selection: too few information about too many datasets

- Lack of reliable dataset metadata but wide diversity (e.g., DBpedia vs traffic stats London vs ...)
  - Spatial and temporal coverage?
  - Dynamics? (evolution, frequency of changes...)
  - Resource types & topics? (policy documents vs traffic statistics)
  - Currentness, availability, provenance, ....

Enrichment & Preservation


329.527.661 triples

metadata

LOD cloud: 300++ datasets
DataHub: 6000++ datasets
Data preservation: handling evolution of distributed data

- Preservation needs to address evolution of distributed datasets / semantics of links
  - In RDF graphs (such as the LOD Cloud), „all“ nodes are connected:
    - Which datasets to preserve (only direct links or also more distant neighbours)? (semantic relatedness, see [ESWC2013])
    - Propagation of changes in LOD graph => measuring relevance of changes for specific entities
Data preservation: handling evolution of distributed data

- Preservation needs to address evolution of distributed datasets / semantics of links
  - In RDF graphs (such as the LOD Cloud), „all“ nodes are connected:
    - Which datasets to preserve (only direct links or also more distant neighbours)? (semantic relatedness, see [ESWC2013])
    - Propagation of changes in LOD graph => measuring relevance of changes for specific entities
  - Preservation strategies dependent on dataset dynamics
    - Simple linking (archiving) for static datasets (eg statistics over past periods in data.gov.uk)
    - Recurring link computation and graph archival for dynamic datasets (frequency?)

Traffic statistics (2013-…)
Energy efficiency policies

<dbp:Berlin>
<dbp:Berlin(east)>
<geoLatLong:52/13>
<dura:GDR Peoples Palace>
Approach: dataset profiling

- Enrichment & preservation = intertwined process!
- **Dataset selection & cataloging:** via DataHub.io (similar to LOD cloud)
  
- Dataset profiling: metadata about dataset dynamics, size, types, topics, evolution, temporal/spatial coverage etc
  
  => **Data observatory** (see also [ESWC2013], [ISWC2013])
- **Vocabulary curation** (expert-based)

Web Data Curation for Building-related Data

Stefan Dietze (L3S Research Center)
Dataset profiling: processing workflow

Goals:

- RDF catalog of datasets
- Tracking the evolution of datasets according to, eg, topics, dynamics, spatial coverage, accessibility
- Links and coreferences => unified view on data => Linked Building Data Graph
- Infrastructure & APIs for federated queries

Towards a Web Data "Observatory"
Web Data Observatory - ongoing work

Stefan Dietze (L3S Research Center)
Using dataset profiles for semi-automated data interlinking:

- Manual alignment of schemas & vocabularies into unified RDF graph
- Automated interlinking (and preservation) techniques
- Preservation metadata (PREMIS RDF?)
- Expert-based curation of building-related vocabularies
  - BuildingSmartDD (http://www.buildingsmart.org/standards/ifd)
  - OMNIClass, UNIClass
  - SFB-NL (http://nl-sfb.bk.tudelft.nl)
  - CROW Library for infrastructural objects (http://www.gww-ob.nl/)
  - …

Vocabulary Curation & Data Interlinking
Summary

- “Data Observatory” as generic platform and domain-specific instantiation (profiling building-related dataset aspects in DURAARK)
- Preservation/linking strategies for SDA based on dataset profiles (eg dynamics, relevance)

Outlook

- Dataset selection: populating DataHub-group
- Schema and vocabulary curation and alignment
- Dataset profiling: establishing LDO, considering range of metadata aspects
- Building SDA: data interlinking & dataset preservation

Conclusions
Thank you!

http://purl.org/dietze  |  @stefandietze

http://www.duraark.eu