TECHNISCHE UNIVERSITÄT DRESDEN

MMT

Faculty of Computer Science, Institute of Software and Multimedia Technology, Senior Chair of Multimedia Technology

A Storage Ontology for Hierarchical Storage Management Systems

Sandro Schmidt, Torsten Wauer, Ronny Fritzsche, and Klaus Meißner

Carl Gustav Carus 1789 – 1869

DLR Deutsches Zentrum für Luft- und Raumfahrt e.V. Projektträger im DLR

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

DRESDEN concept Exzellenz aus Wissenschaft und Kultur

Malta, 26.09.2013

Outline

- Introduction and motivation
- Storage Ontology
- Evolution of the Storage Ontology in SENSE
- Conclusion

Introduction and Motivation

Why do we need an ontology for HSM Systems?

Hierarchical Storage Management (HSM) Systems

Typical storage media and devices

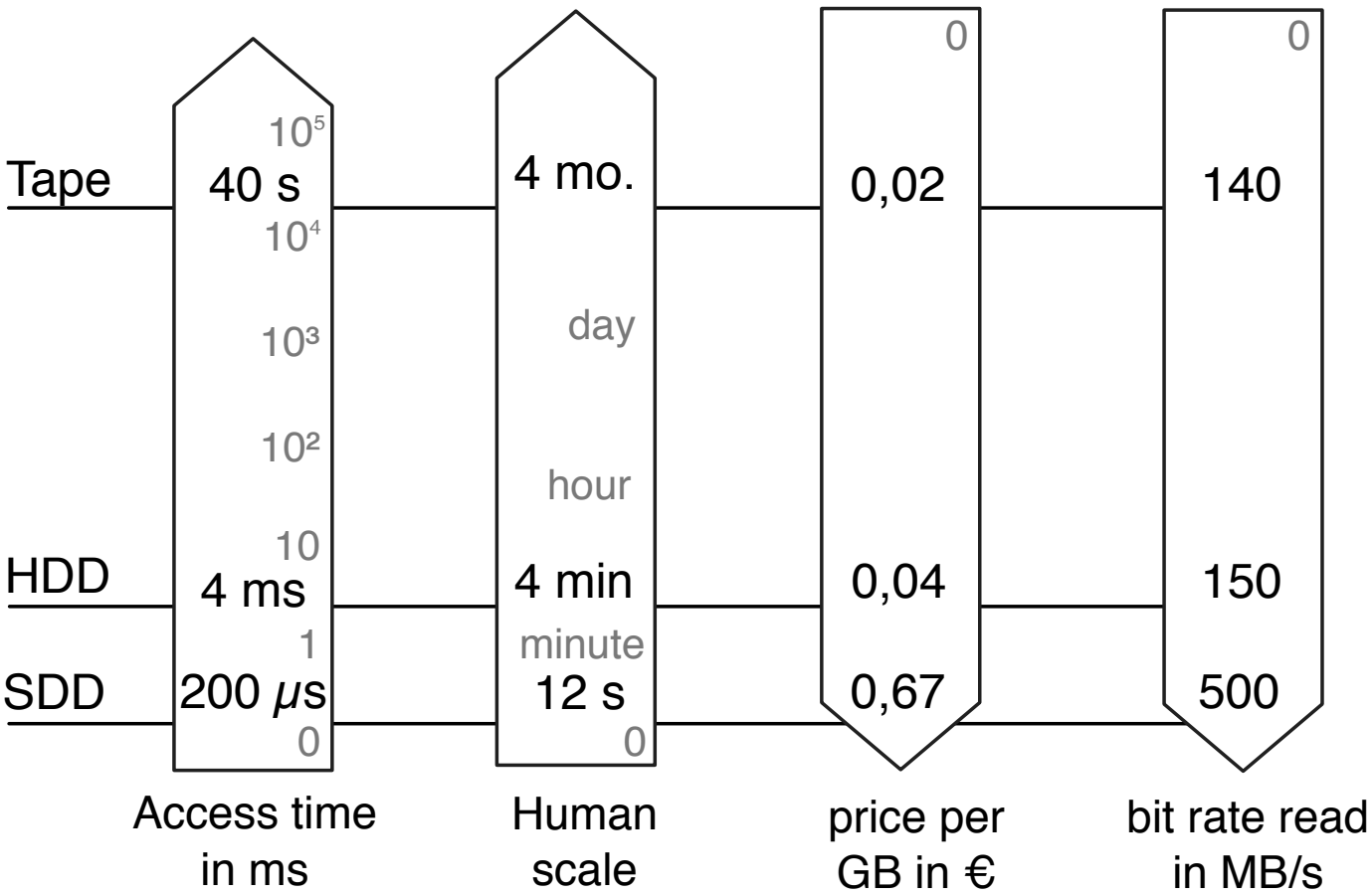
Capacity Cost/GB Access time

Performance tier

Capacity tier

Archive tier

HSM: tradeoff between costs and performance



Problems of classical HSM Systems (1)

- Few information for migration:
  - limited file attributes
  - defined rules on folders / file types / by date
  - users demand (purge)

- No information about the location of a file:
  - No information how much time it takes to load a file

Problems dealing with HSM Systems (2)

- File system access for metadata of files:
  - Possibly to slow
  - Not sufficient for reasoning

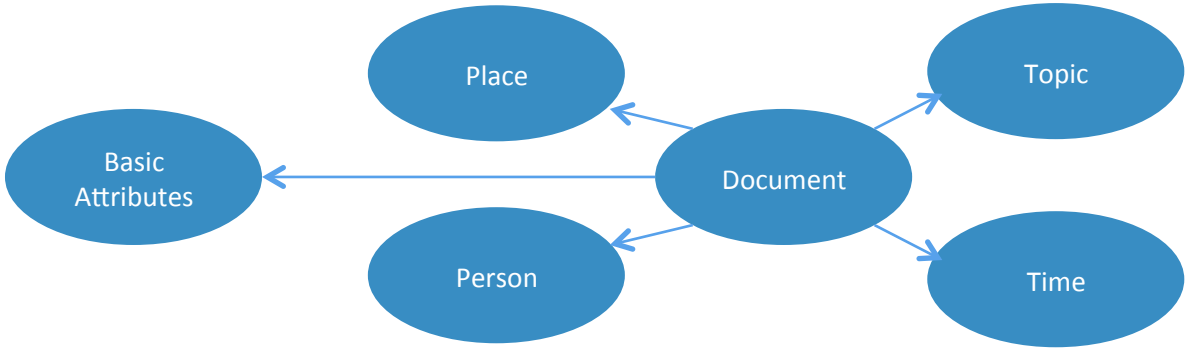
- Migration is not based on:
  - Content of files
  - Relations between Files

Storage Ontology (SO)

The Concept.

Managing stored information using an ontology (1)

- In 2011 [1] first idea of using an ontology to optimize file storage in HSM systems
  - HSM Extension which used a ontology to store basic semantic information

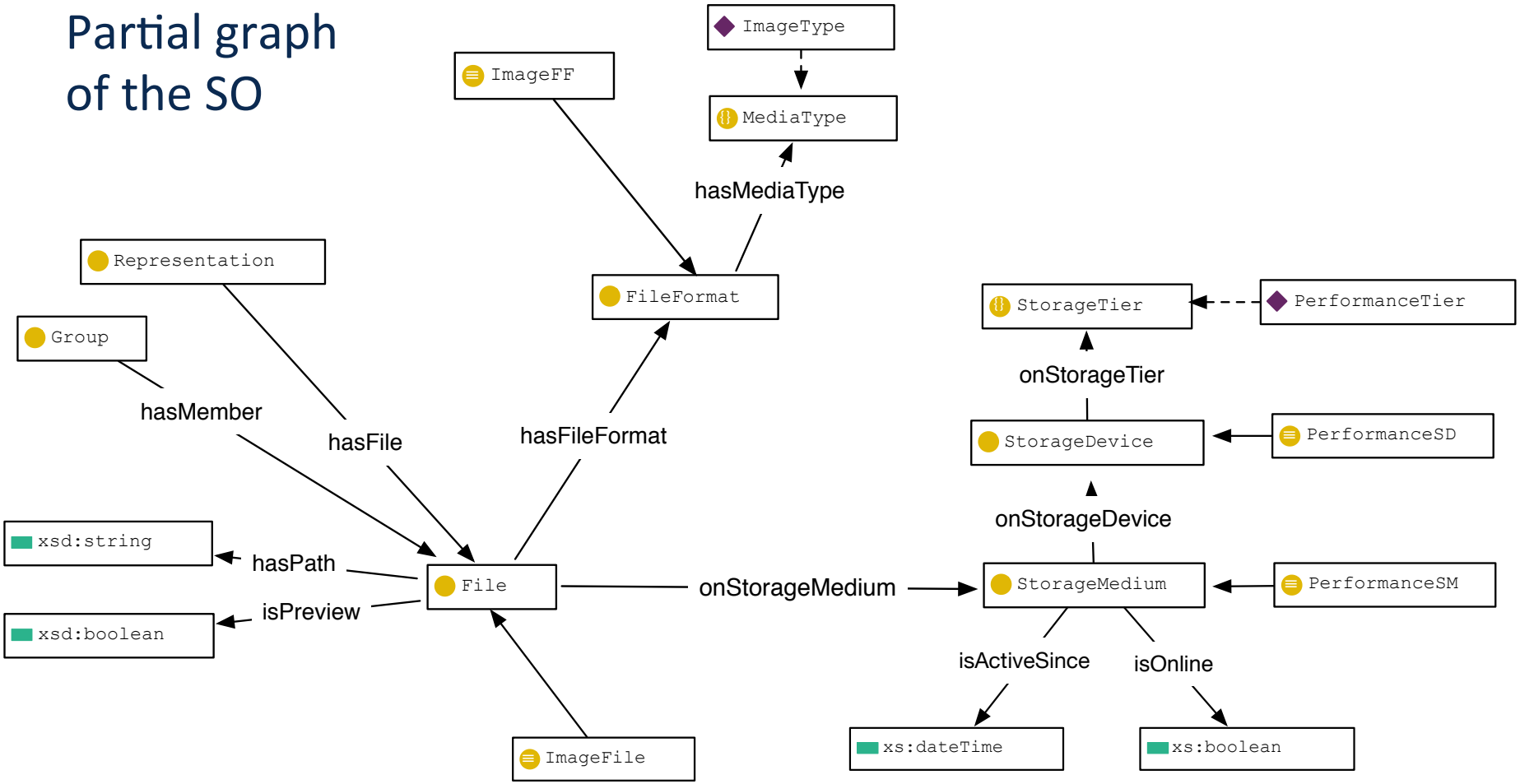


Managing stored information using an ontology (2)

- Information about files itself
- We also need information about:
  - Location in the HSM system
  - Storage conditions

- Consequences: an ontology describing …
  - **Files** and relations between them
  - **HSM** structure / architecture
  - **Relation** between Files and HSM

Partial graph of the SO



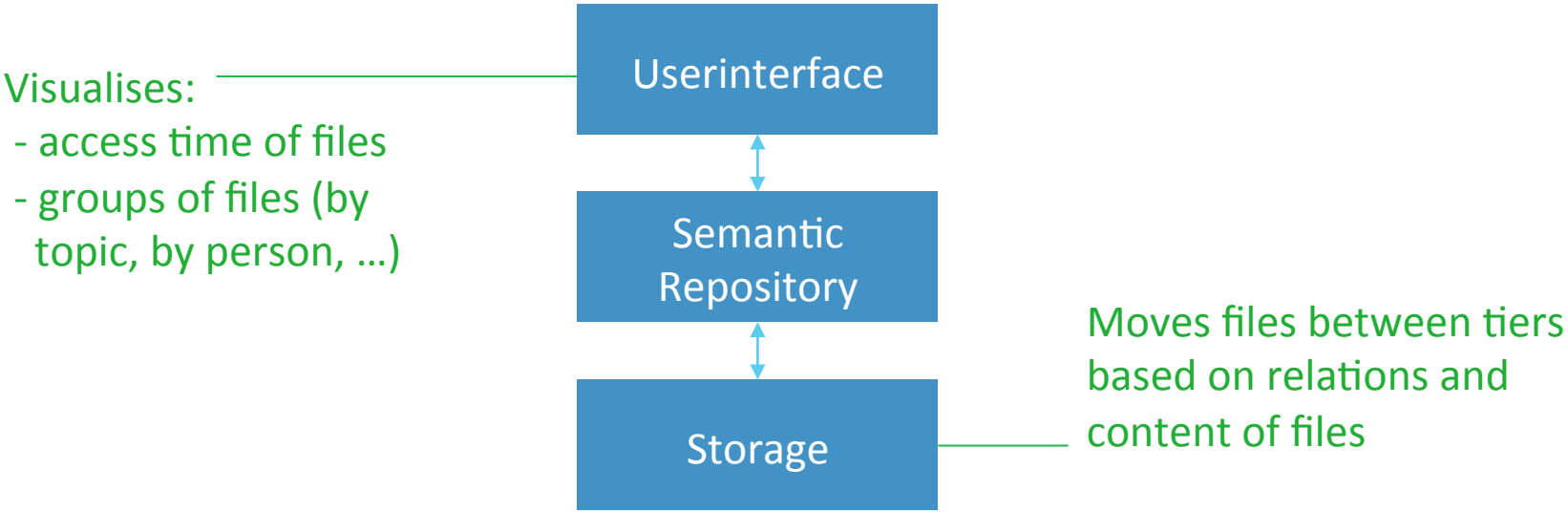
Evolution of the SO in SENSE

The Framework around the Storage Ontology.

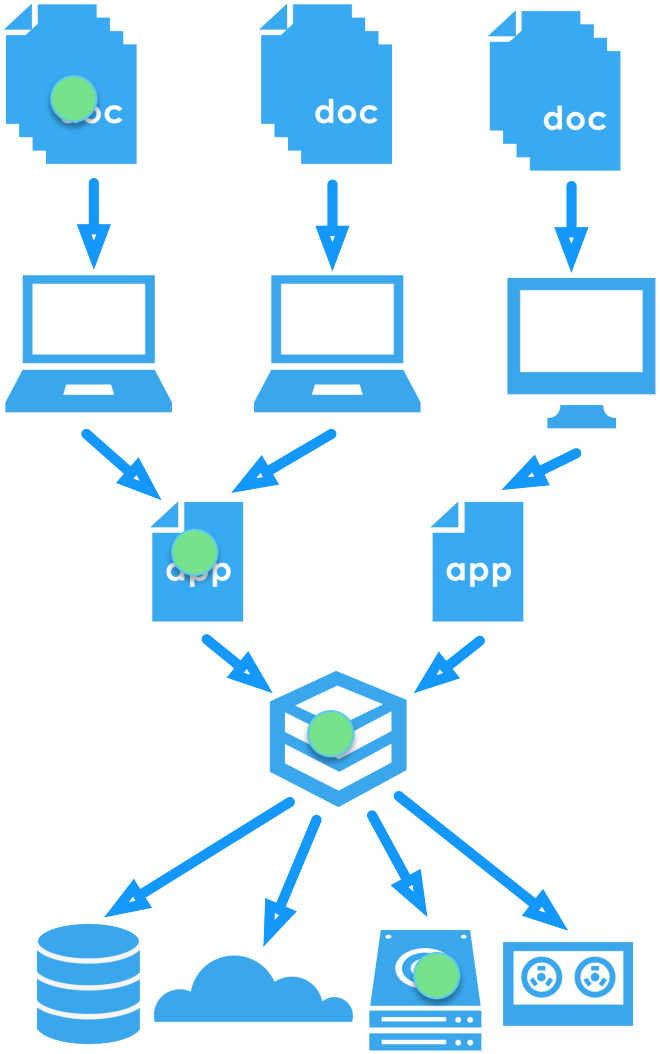
The SO within SENSE

SO is the basis for work and research within the SENSE project [2]

- Aim: offer a continuous semantic description for all application layers

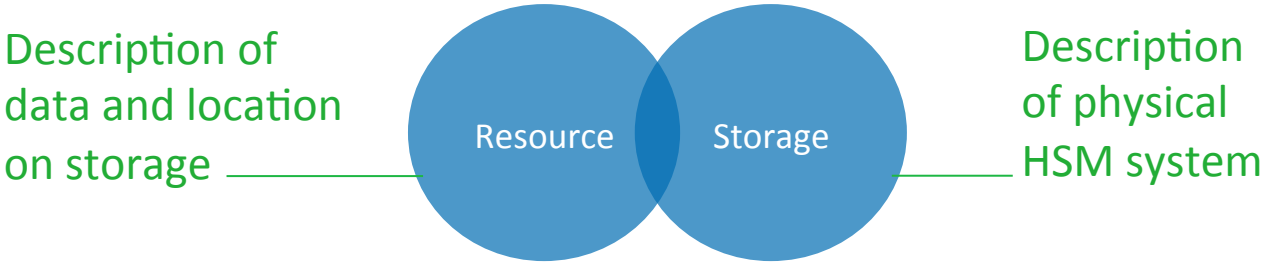


SENSE: The idea



Splitting the SO

- Splitting of the SO in cooperation with PoINT Software and Systems GmbH (Gundolf Gremler) into two seperate ontologies

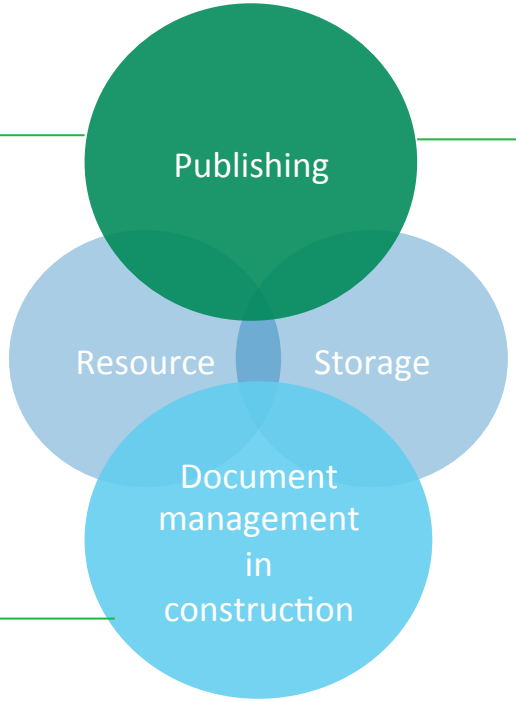


Use cases and tests

Integration of application ontologies (publishing and construction)

Fink&Partner Medienservice Gmbh

First test set with
- 400.000 files
- 6.9 Million triples



planConnect

Example 1

- Updating the information about the used tier

```
PREFIX res:<https://svn.mmt.inf.tu-dresden.de/SENSE/ontologies/resource#>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>

INSERT { ?res res:isStoredOn sto:ArchiveTier }
WHERE {
   ?res res:ResourcePathName PATH^^xsd:string .
   OPTIONAL { ?res res:isStoredOn ?tier }
}
```

Example 2

- Example: SPARQL-Query to load images from HSM

```
SELECT ?mpimgu ?hpimgu ?mtier ?htier ?previewuri
 FROM <media>
 WHERE {
    ?highres a schema:ImageObject.
    ?highres fup:hasPreviewImage ?preview.
    ?preview res:ResourceURI ?previewuri.
    ?highres res:ResourceURI ?highresu.
    OPTIONAL
    { ?highres fup:hasPreviewMidresImage ?mpimg.
    ?mpimg res:ResourceURI ?mpimgu.
    ?mpimg res:isStoredOn ?mtier. }
    OPTIONAL
    { ?highres fup:hasPreviewHighresImage ?hpimg.
    ?hpimg res:ResourceURI ?hpimgu.
    ?hpimg res:isStoredOn ?htier. }
    Filter(?preview = res:cb09804b-e667-44a4-9c42-2ba76fdf330c ||
       ?preview = res:371b70fa-7268-4d6d-a776-cef2c68f3d8f}
```

Load high resolution image and preview

If available, load other resoltions of the image

Example 3

- Find all related documents to documents, that are not stored in the performance tier

```
PREFIX res:<https://svn.mmt.inf.tu-dresden.de/SENSE/ontologies/resource#>
PREFIX sto:https://svn.mmt.inf.tu-dresden.de/SENSE/ontologies/storage#

SELECT ?path
WHERE {
    ?a res:ResourcePathName PATH^^xsd:string ."
    ?c res:ContainsResources ?a .
    ?c res:ContainsResources ?d .
    ?d res:ResourcePathName ?path .

    MINUS {
        ?d res:ResourcePathName PATH^^xsd:string .
    }
    MINUS {
        ?d res:isStoredOn sto:PerformanceTier .
    }
}
```

Conclusion

What have we achieved?

Conclusion

Results:

- Genric set to describe documents with an ontology
- A semantic model to describe the architecture of storage sub systems
- Benefits of reasoners to gain indirect knowledge

Known problems:

- SPARQL-Performance
- Quality of results depends on information extraction

Thank you for your attention.

Questions?

Acknowledgements

Vizualisation and exploration

OvGU Magdeburg – Data and Knowledge Engineering Group

fink&Parnter Media Services GmbH
(Usecase publishing)

planConnect
(Usecase construction)

Storage optimisation

PoINT Software & Systems

Seniorprofessur Multimediatechnik

References

[1] Axel Schröder, Ronny Fritzsche, Sandro Schmidt, et al. „A Semantic Extension of a Hierarchical Storage Management System for Small and Mediumsized Enterprises".

[2] Intelligent Storage and Exploration of large Document Sets. (http://www.sense-projekt.de)