Faculty of Computer Science

universität wien

Department of Distributed and Multimedia Systems

Multimedia Information Systems Group

Lifting File Systems into the Linked Data Cloud with TripFS

Niko Popitsch, University of Vienna / Austria

niko.popitsch@univie.ac.at

Joint work with Bernhard Schandl, University of Vienna / Austria

bernhard.schandl@univie.ac.at

April 27, 2010

WWW 2010 Conference

Raleigh, North Carolina, USA

Introduction: Linked File Systems

- Major fraction of digital information stored in file systems

- File systems currently provide limited support for
  - **Data organization** (single hierarchy)
  - **Association of arbitrary meta data** with files (unstable identifiers)

- **Idea**: publish parts of a local file system as linked data
  - Files and directories become RDF resources
  - Data organization: single tree → semantic graph
  - Meta data: RDF data model

image: www.freeimages.co.uk

Representing File Systems as Linked Data

- Represent files and directories using **appropriate identifiers** (HTTP URIs) and RDF **vocabularies**
- Enrich RDF graph with **extracted meta data**
- **Link** to other (external) data
- Keep HTTP-URI / File-URI mapping **consistent**
- **Serve as linked data**

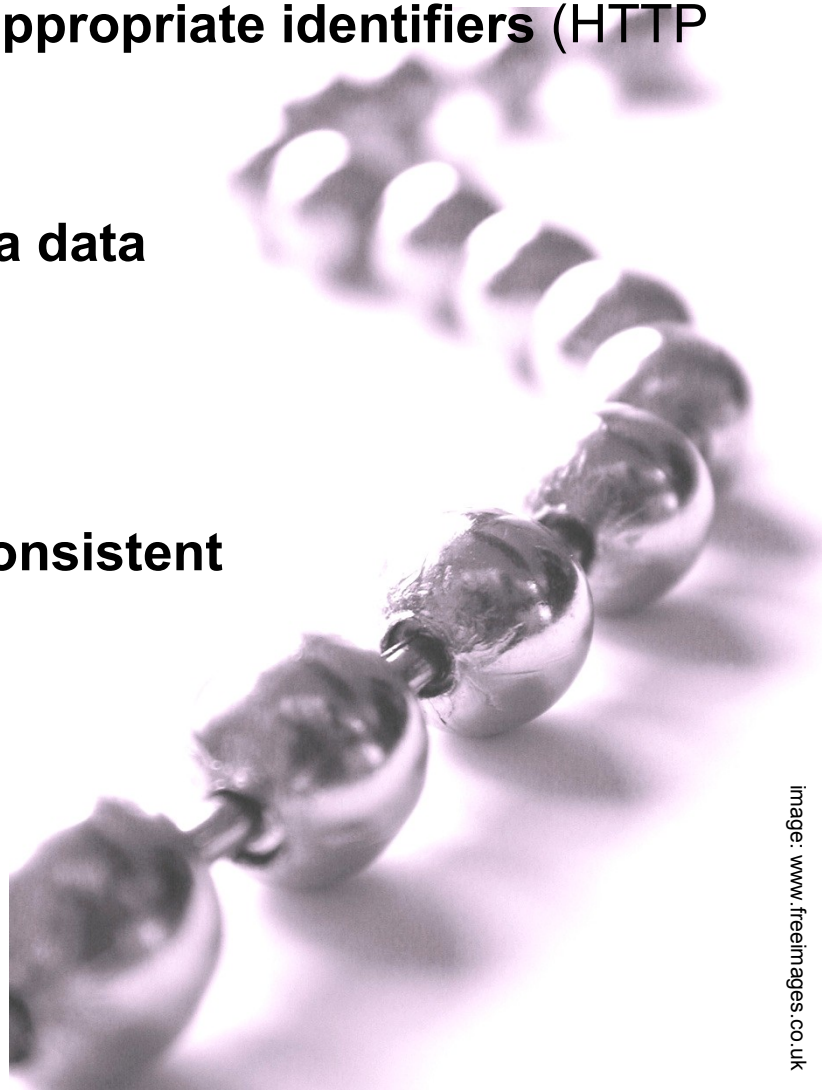


image: www.freeimages.co.uk

Identifying Files and Directories

- Linked Data: Identify resources with **HTTP-URIs**

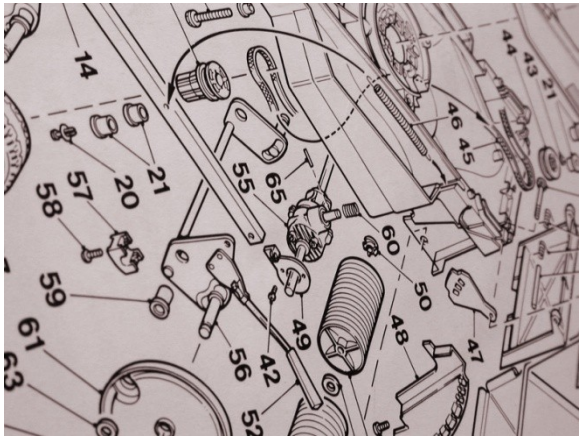


image: www.freeimages.co.uk

- File URIs are not suitable
  - Not stable
  - Not globally unique

- Our approach: **UUID-based URNs**
  - Random UUIDs can be used in global distributed context (uniqueness)
  - Universally Unique IDs are opaque (stable)
  - HTTP-URI Prefix + UUID = HTTP-URI

http://queens:9876/resource/urn:uuid:c1dd60bd-4050-4216-9455-a121efb0fe1b

Representing Files and Directories

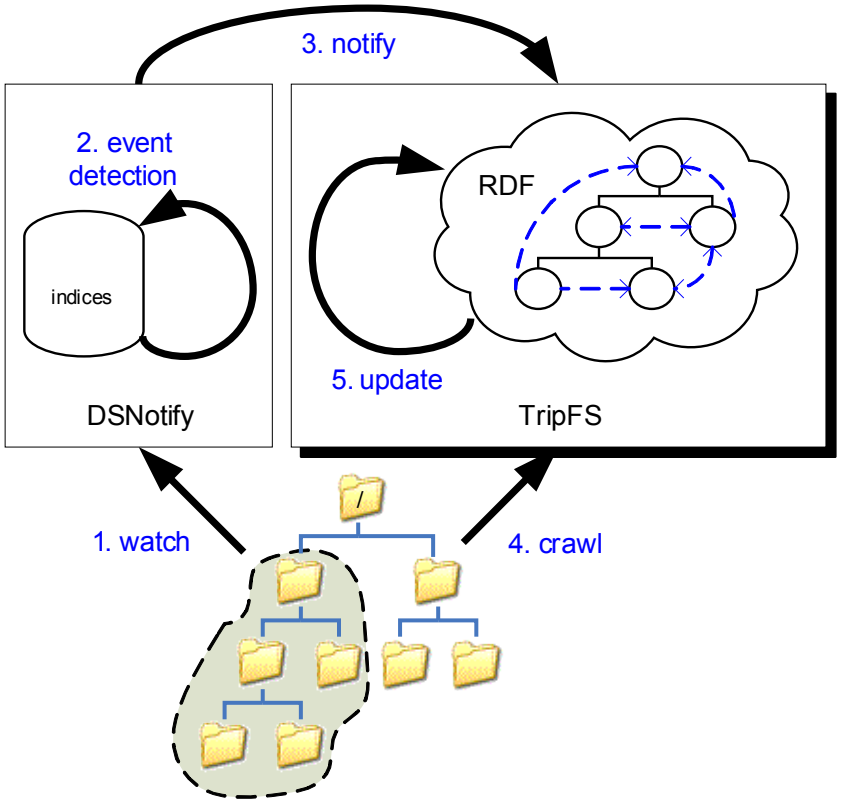
- **Low-level file system meta data** (parent/child relationships, path, size, creation date, …) are modeled using our vocabulary
  - http://purl.org/tripfs/2010/02#

- **Extractors** can be plugged into TripFS
  - Read files of certain format and extract RDF graph
  - Normally use/re-use existing semantic Web vocabularies
  - May extract whole entities related to a file
    - E.g., artist that created a certain piece of music stored in an MP3 file

- **Linkers** can be plugged into TripFS
  - May act on extracted meta data as well as on the file data itself
  - Return an RDF graph containing RDF links to external resources
  - May also interlink local files/directories

Stable TripFS identifiers

- **DSNotify** is a **change detection** add-on for data sources
  - Watch local FS
  - Report detected events
  - TripFS RDF model update

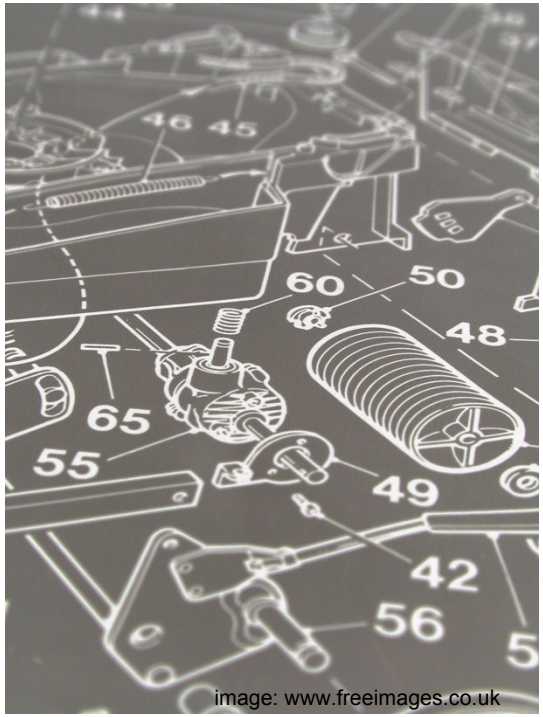


- Event detection based on **feature vector comparison** and **plausibility** checks

- Detects file **create**, **remove**, **move** (rename) and **update** events

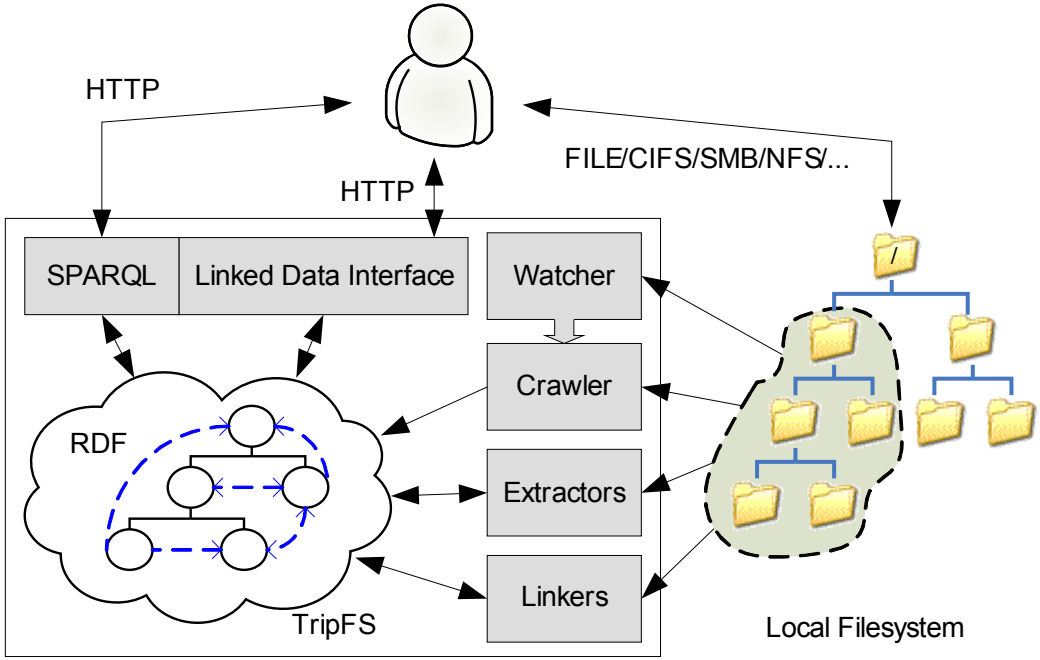
TripFS Change Detection 2

Feature	Datatype	Similarity	Weight
Last access	Date	Plausibility	
Last modication	Date	Plausibility	
IsDirectory	Bool	Plausibility	
Checksum	Integer	Plausibility	
Name	String	Levensthein	3.0
Extension	String	Major MIME type	1.0
Path	String	Levensthein	0.5
Size	Long	Equality	0.1
Permissions	Bitstring	Equality	0.1



- Extracted features, their data type and the strategy used by DSNotify to calculate a similarity between them.
- Some features are used only for plausibility checks

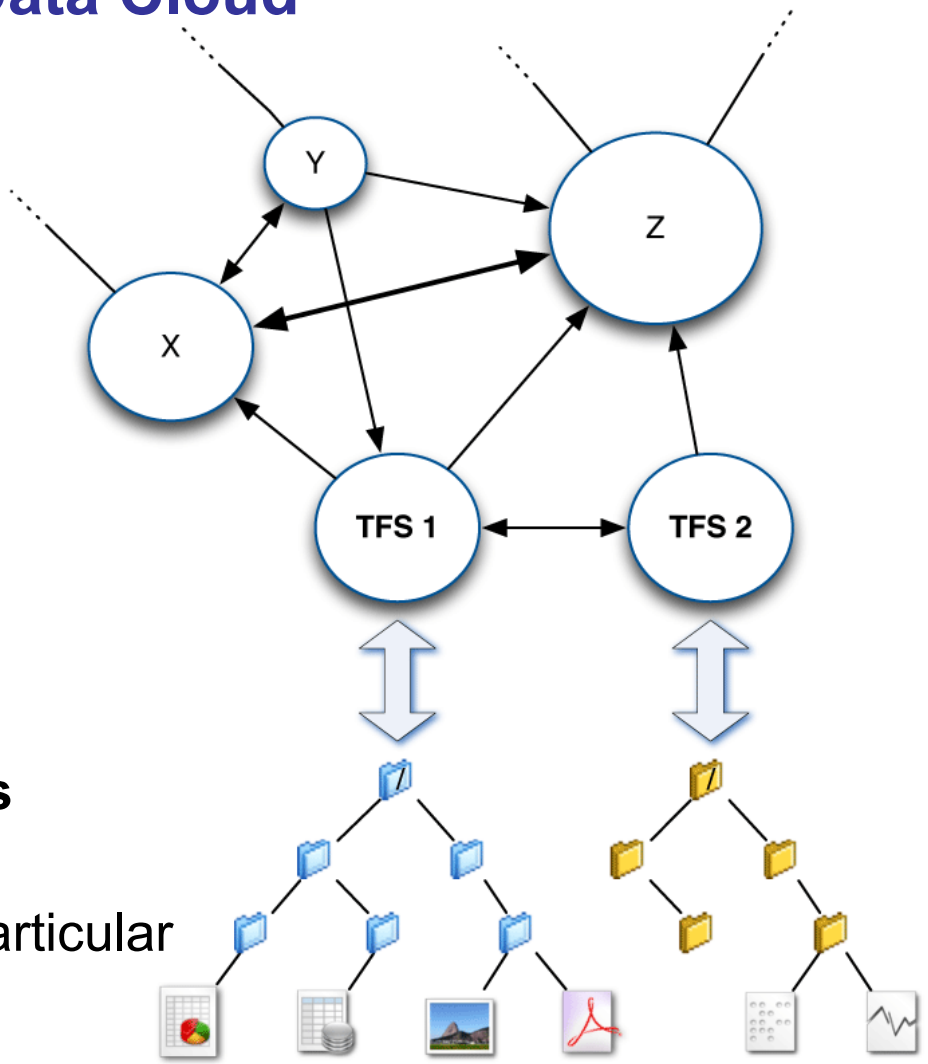
TripFS Architecture and Implementation



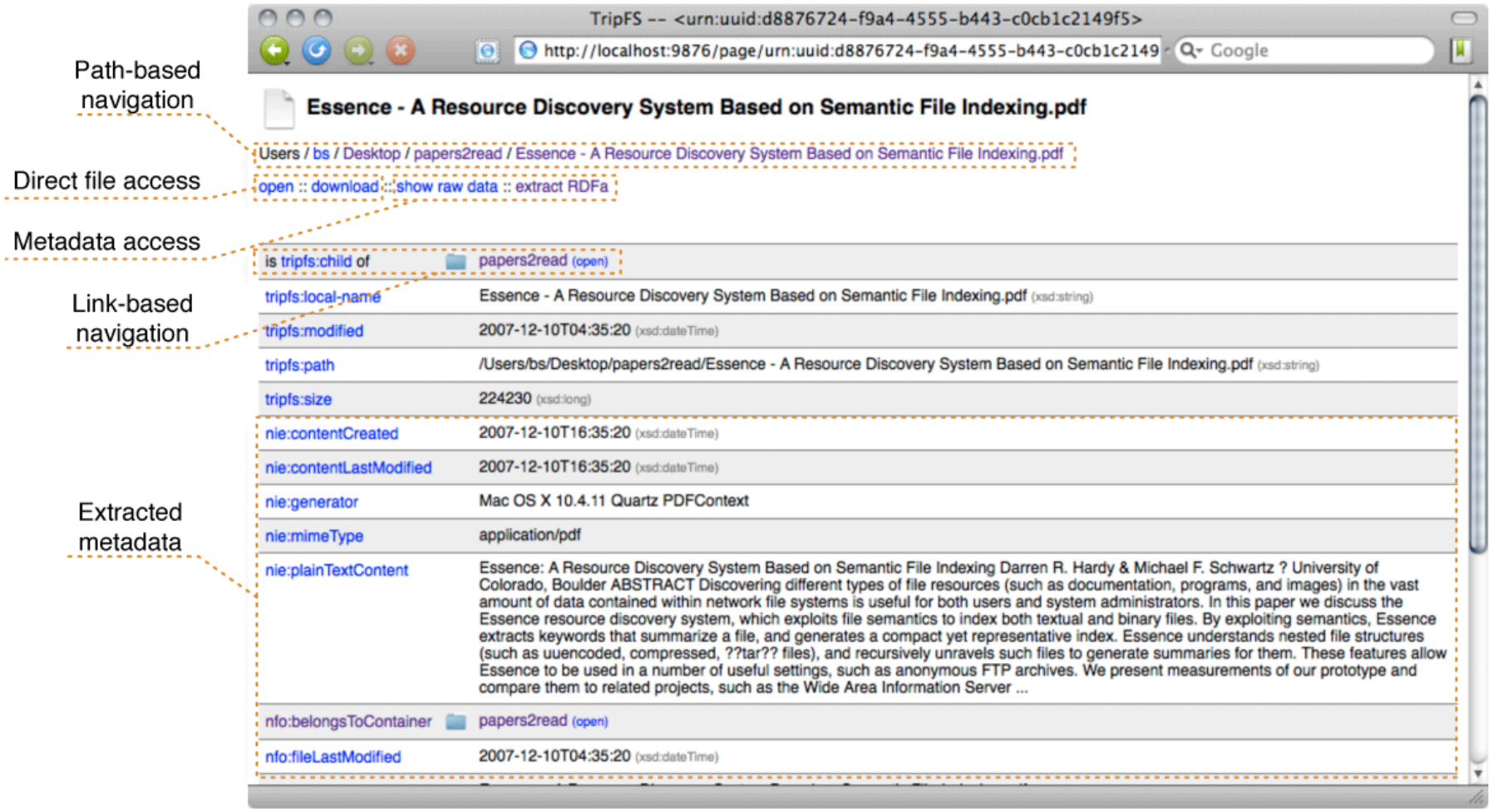
- **Plug-in** concept (Several extractors, linkers, watchers are already implemented)
- **SPARQL** endpoint
- **Linked data interface**
- Technologies: Java, Jena, Jetty, Aperture, DSNotify

TripFS and the Linked Data Cloud

- Each TripFS instance is a “bubble”
- Possibly **transient** but **stable** data
- Links between
  - **TripFS** instances
  - other **LD sources**
  - other **remote resources** (e.g., Web pages, etc.)
  - **Local resources** in a particular TripFS instance



TripFS Interface



RDF Example

```
<urn:uuid:887d728e-bc12-4f28-a497-7d66439086e9>
   a tripfs:File ;
   rdfs:label "eswc2009-schandl.pdf" ;
   tripfs:local-name "eswc2009-schandl.pdf"^^xsd:string ;
   tripfs:path "/Users/bs/.../eswc/eswc2009-schandl.pdf"^^xsd:string ;
   tripfs:size "425561"^^xsd:long ;
   tripfs:modified "2009-03-11T02:38:45"^^xsd:dateTime ;
   tripfs:parent <urn:uuid:35069c61-451e-4688-98f5-080924b261f4> .

<urn:uuid:a998272d-45f0-4814-8f15-be5db5fe811a>
   nie:mimeType "audio/mpeg" ;
   nid3:title "Bohemian Rhapsody" ;
   nid3:leadArtist [ nco:fullname "Queen" ] ;
   nid3:length 355106 .

<urn:uuid:887d728e-bc12-4f28-a497-7d66439086e9>
   owl:sameAs <http://dblp.l3s.de/d2r/resource/publications/conf/esws/SchandlH09> .

<urn:uuid:a998272d-45f0-4814-8f15-be5db5fe811a>
   rdfs:seeAlso <http://musicbrainz.org/track/c7faf83f-9cb3-4de4-a39f-1c1f98b8d81a> ,
                <http://musicbrainz.org/track/95ebc842-9926-4658-8012-12c358247946> ;
   owl:sameAs <http://musicbrainz.org/track/bbd5a2e7-9814-4988-8f5a-dc38c208eeea> ,
              <http://musicbrainz.org/track/064c440c-4eba-47a6-83c4-c91a979eeb4b> .
```

Future Work and Discussion

- Linked file systems could become **bubbles in the linked data cloud**
- They could
  - improve data organization on the desktop
  - help in various application scenarios like
    - **Enterprise data integration**
    - **Ad-hoc sharing** of resources and context
    - **Annotation** of local data with semantic Web tools

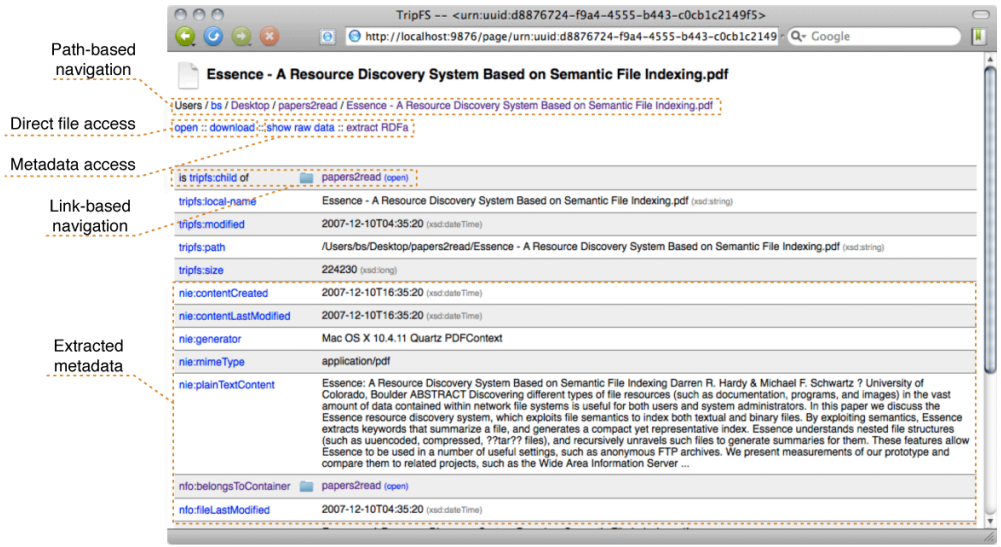
TripFS is a first linked file system prototype

- Future work:
  - Evaluate TripFS regarding **scalability** and **performance**
  - **Accuracy** of the **change detection** solution (DSNotify)
  - Introduce fine grained control for
    - **What** is exposed via TripFS
    - **How** it is exposed and
    - **Who** may access it
  - Integration with desktop tools (e.g., file explorers)

Thank you !

Demo and Discussion

http://demo.mminf.univie.ac.at:9876/



bernhard.schandl@univie.ac.at
niko.popitsch@univie.ac.at

Related work

- **Semantic file system prototypes**
  - AttrFS: attribute-based access to files
    - prototypical implementation based on user-level NFS server
    - Query files by building conjunctive/disjunctive logical expressions
    - Also: computed attributes (e.g., “age in days”)
  - LiFS: attributed links between files
    - FUSE-based prototype
    - Accessible via enhanced POSIX interface
  - Many more! SFS, Presto, LISFS, SemDAV, …
- **Tools for extracting / converting RDF descriptions**
  - Aperture, PiggyBank, Virtuoso Sponger, …
- **Tools for exposing data representations as linked data**
  - D2R, Triplify, OAI2LOD, XLWrap, …
- **iNotify** could be used on Linux as change detection component in DSNotify

References

- Sasha Ames, Nikhil Bobb, Kevin M. Greenan, Owen S. Hofmann, Mark W. Storer, Carlos Maltzahn, Ethan L. Miller, and Scott A. Brandt. LiFS: An Attribute-Rich File System for Storage Class Memories. In Proceedings of the 23rd IEEE / 14th NASA Goddard Conference on Mass Storage Systems and Technologies, 2006
- William Y. Arms. Uniform Resource Names: Handles, PURLs, and Digital Object Identiers. Commun. ACM, 44(5):68, 2001
- Sören Auer, Sebastian Dietzold, Jens Lehmann, Sebastian Hellmann, and David Aumueller. Triplify: Light-weight Linked Data Publication from Relational Databases. In WWW '09: Proceedings of the 18th international conference on World wide web 2009
- 621{630, New York, NY, USA, 2009. ACM.
- Arati Baliga, Joe Kilian, and Liviu Iftode. A Web-based Covert File System. In Proceedings of the 11th Workshop on Hot Topics in Operating Systems, 2007
- Tim Berners-Lee. Linked Data. World Wide Web Consortium, 2006. Available at http://www.w3.org/DesignIssues/LinkedData.html
- Chris Bizer, Richard Cyganiak, and Tom Heath. How to Publish Linked Data on the Web, 2007. Available at http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/
- Sanjay Ghemawat, Howard Gobio, and Shun-Tak Leung. The Google File System. In 19th ACM Symposium on Operating Systems Principles, 2003.
- Bernhard Haslhofer, Wolfgang Jochum, Ross King, Christian Sadilek, and Karin Schellner. The LEMO Annotation Framework: Weaving Multimedia Annotations with the Web. International Journal on Digital Libraries, 10(1), 2009.
- Niko Popitsch and Bernhard Haslhofer. DSNotify: Handling Broken Links in the Web of Data. In 19th International WWW Conference (WWW2010), Raleigh, NC, USA, 2 2010. ACM.
- Leo Sauermann and Sven Schwarz. Gnowsis Adapter Framework: Treating Structured Data Sources as Virtual RDF Graphs. In Proceedings of the 4th International Semantic Web Conference (ISWC 2005)
- Bernhard Schandl. Representing Linked Data as Virtual File Systems. In Proceedings of the 2nd International Workshop on Linked Data on the Web (LDOW), Madrid, Spain, 2009
- Julius Volz, Christian Bizer, Martin Gaedke, and Georgi Kobilarov. Discovering and Maintaining Links on the Web of Data. In Proceedings of the 8th International Semantic Web Conference (ISWC 2009), 2009

BACKUP

Scenarios for Linked File Systems

- Enterprise Data Integration
  - Uniform data access to heterogeneous enterprise data
- Ad-hoc data sharing
  - publish files + semantic meta data
  - Exchange not only the data but also its (semantic) context
- Weave local data with semantic Web
  - Use semantic Web technologies on local data
  - Example: annotate and interlink local files with semantic annotation tools

Random UUIDs

- Java UUIDs have 122 random bits

$$p(n) \approx 1 - e^{-\frac{n^2}{2x}}$$

- Probability of accidental clash after generating *n* UUIDs
- Sources: http://www.h2database.com/html/advanced.html#uuid, Wikipedia

n	**probability**
68,719,476,736 = 2^{36}	0.0000000000000004 (4×10^{-16})
2,199,023,255,552 = 2^{41}	0.000000000004 (4×10^{-13})
70,368,744,177,664 = 2^{46}	0.0000000004 (4×10^{-10})

Demo: Connecting to shared folder

TripFS Demo running at: http://xx.xx.xx.xx:9876/

DSNotify Demo running at: http://xx.xx.xx.xx:8100/

Shared folder at smb://xx.xx.xx.xx/tfs

xx.xx.xx.xx =

Demo: Connecting to TripFS and WebDAV server

TripFS Demo running at: **http://xx.xx.xx.xx:9876/**
DSNotify Demo running at: **http://xx.xx.xx.xx:8100/**
Webdav Demo running at: **http://xx.xx.xx.xx:8080/dav/**

- **Mac OSX**
  - Finder → Go → Connect to Server
  - Enter address → enter username/pwd
- **Linux** (tested on ubuntu 9.10)
  - Places → connect to server → select WebDav → enter address and username/pwd
  - Or mount it using davfs or fusedav…
- **Windows XP**
  - NOTE: Windows support is bad! Editing files might not work, but copying and directory creation should..
  - Possibility 1
    - Open **iexplore 7**
    - File → Open → enter address and click “Open as Web Folder” checkbox
    - Enter username/pwd
  - Possibility 2
    - Use explorer
    - Tools → Map Network Drive → “**Sign up for online storage or connect to a network server”** → next → Choose another network location
    - enter address → next → enter username/pwd