

Introduction to Fedora 4 in collaboration with docuteam applications

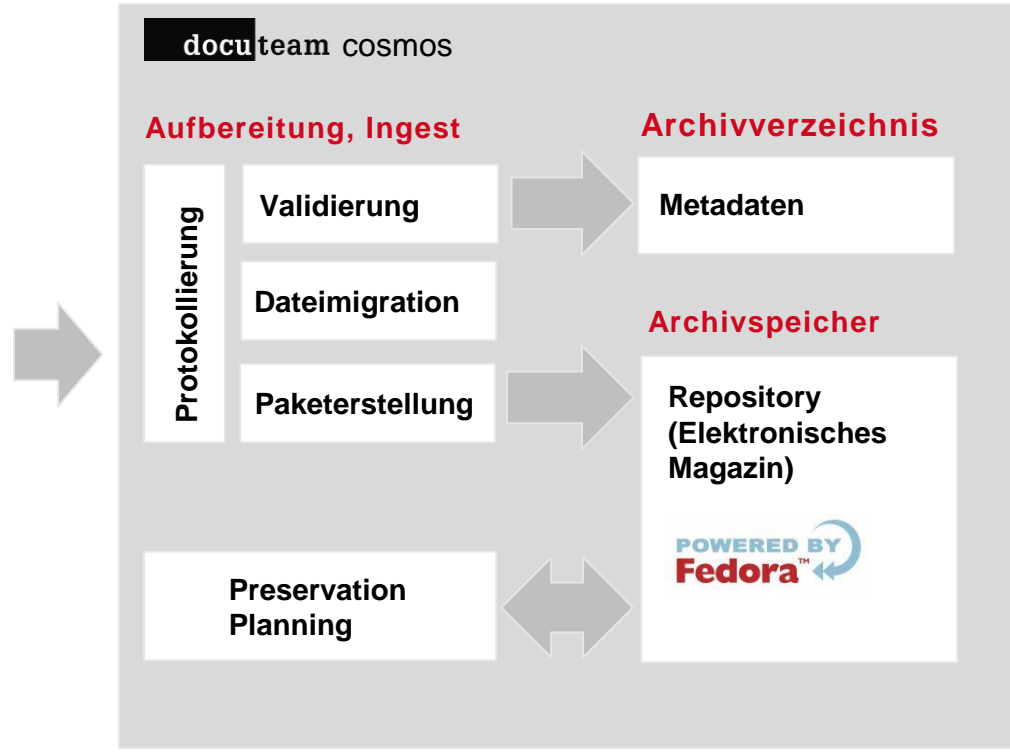
docuteam Community Day, 31. Oktober 2018
docuteam Community Day, le 31 octobre 2018
Andreas Nef, a.nef@docuteam.ch

docuteam cosmos: die Übersicht

docuteam cosmos: vue d'ensemble

Auskoppeln und abliefern

- docuteam packer
- Brain CONNECT
- CMI AXIOMA
- Acta Nova
- Open Text Content Server
- HERITRIX



docuteam webgate

Integration

- atom access to memory
- OAI-PMH

- cyberlink
- SWITCHengines

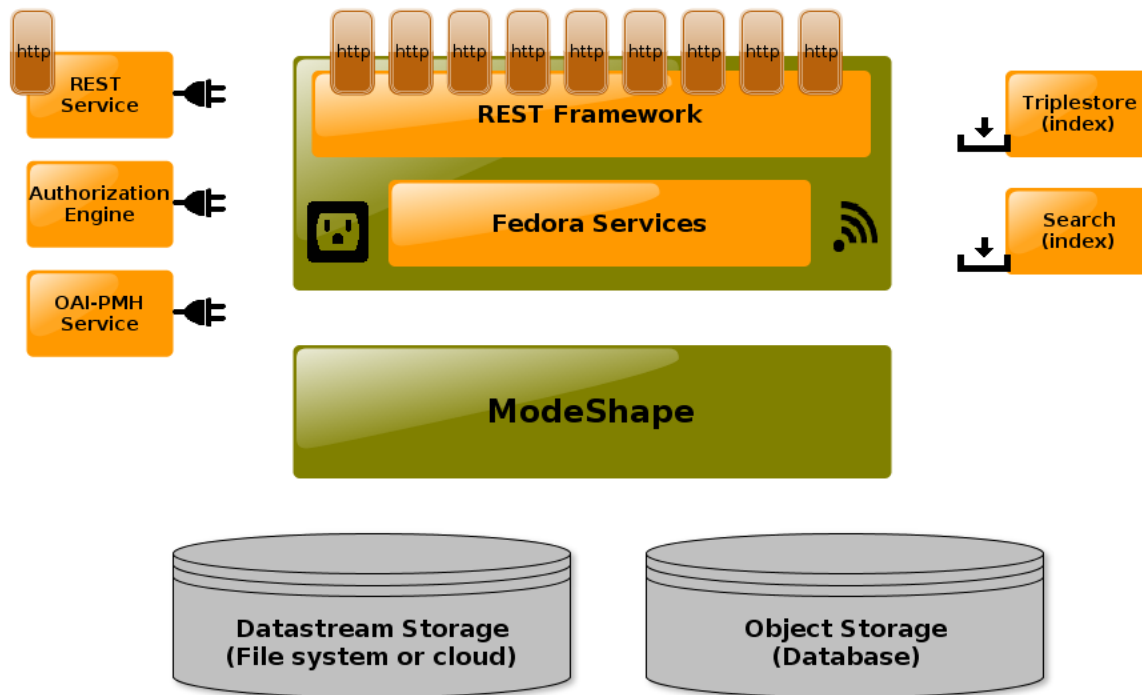
- **Fedora is the flexible, modular, open source repository platform with native linked data support.**
- What it is:
 - adding functionality and semantic to the storage layer
 - supporting concepts to model simple to complex objects in a flexible way
 - offering a standardized and secured access to the objects stored
 - solid base for complex digital resources platforms
- It's not:
 - an end-user product with a flashy user interface
 - an out-of-the-box digital reading room
 - an ingest workflow engine

- Good experiences with Fedora 3
 - Stable and reliable product used for more than ten years
 - Active/international community
 - Focus on core functionality
 - Platform independent, flexible infrastructure requirements
 - Good support of community standards (XML based)

But...

- New standards evolve in metadata management and object modeling
 - Semantic web
 - Ontologies (e.g. PREMIS, MODS)
- Technology stack is getting deprecated or lacks functionality/standardization/performance of more up-to-date solutions

- continued community support
- even more focus core functionality
 - modularizing/externalizing
 - using reliable technologies for core concepts:
 - Memento (Versioning)
 - ACL (Permissions)
- native support for Linked Data
 - Fedora 4 is a LDP server, i.e. implementing the W3C rules on handling web resources (<https://www.w3.org/TR/ldp/>)
 - Metadata not only as literals,



- Fedora 5 is already in the pipeline
- primarily an alignment of the API and a reference implementation
 - Resource Management (Linked Data Platform)
 - Resource Versioning (Memento)
 - Resource Authorization (Web Access Controls)
 - Notifications (Activity Streams)
 - Binary Resource Fixity (HTTP headers)

New data model required: From Matterhorn METS to Matterhorn RDF

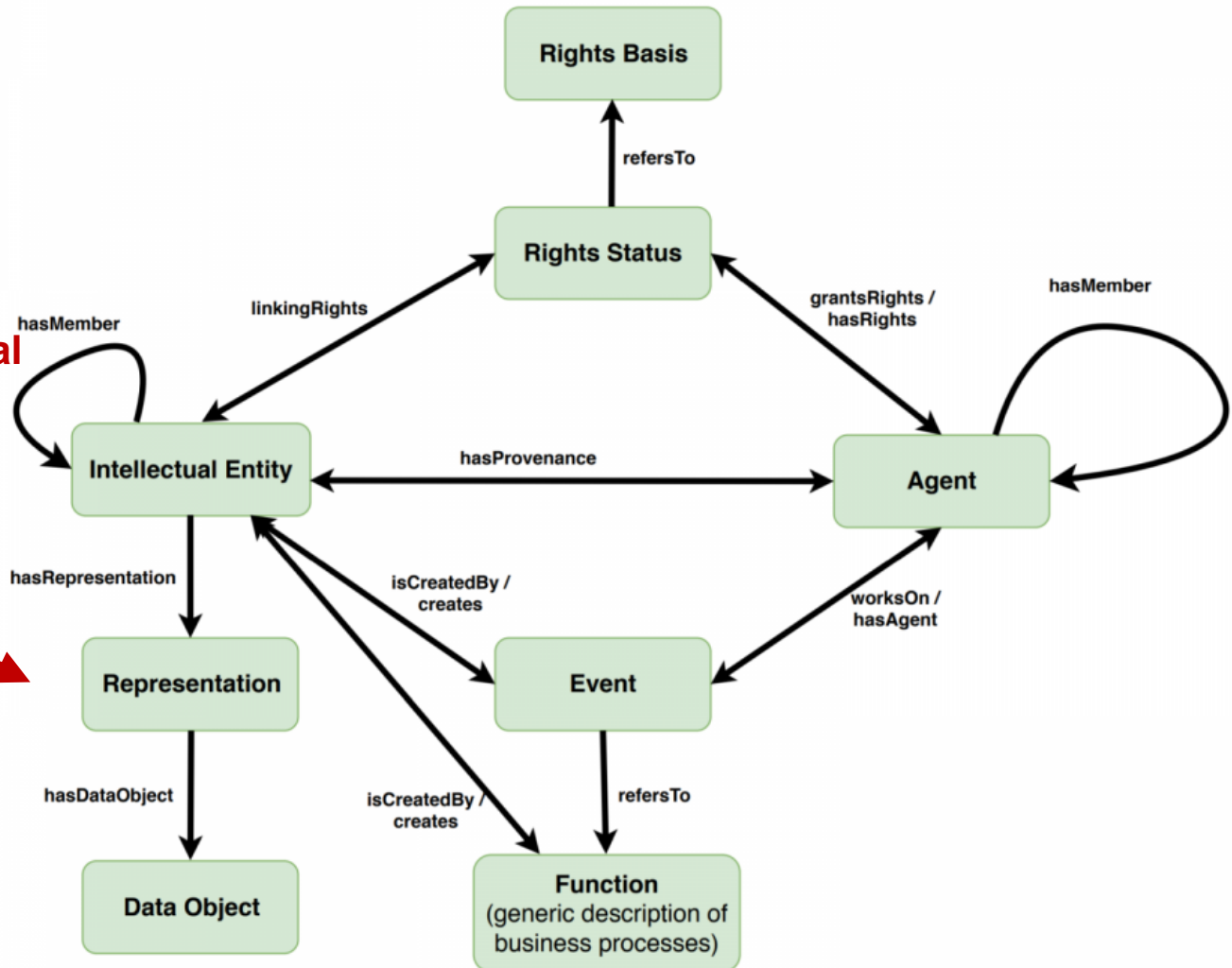
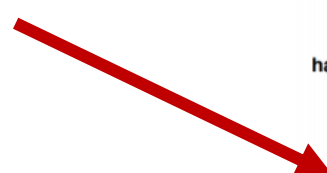
- From XML to RDF
 - Fedora 3: Matterhorn METS with METS, EAD, PREMIS
 - Fedora 4: A new RDF-based data model is required
- Rework the Matterhorn METS Profile in view of the possibilities offered by Linked Data and Open Linked Data, for which digital preservation is a core motivation
- A larger context (libraries, Wikipedia etc.) for archives
- Make use of existing resources to improve the precision of archival description

- ICA «Expert Group for Archival Description» (EGAD) works on «Records in Context» since 2012 to replace ISAD / ISAAR / ISDF
 - **EGAD’s nongeneric approach:** Developing an RDF standard specific to archives, with gateways to library and museum standards
 - **The Matterhorn RDF Data Model’s generic approach:** Based upon RDF existing and consensual international standards, allowing to model Records in Context.
 - In contrast to EGAD, the Matterhorn RDF Data Model is based on existing ontologies. It follows the best practices propagated by the W3C: «**It is best practice to use or extend an existing vocabulary before creating a new vocabulary.**»

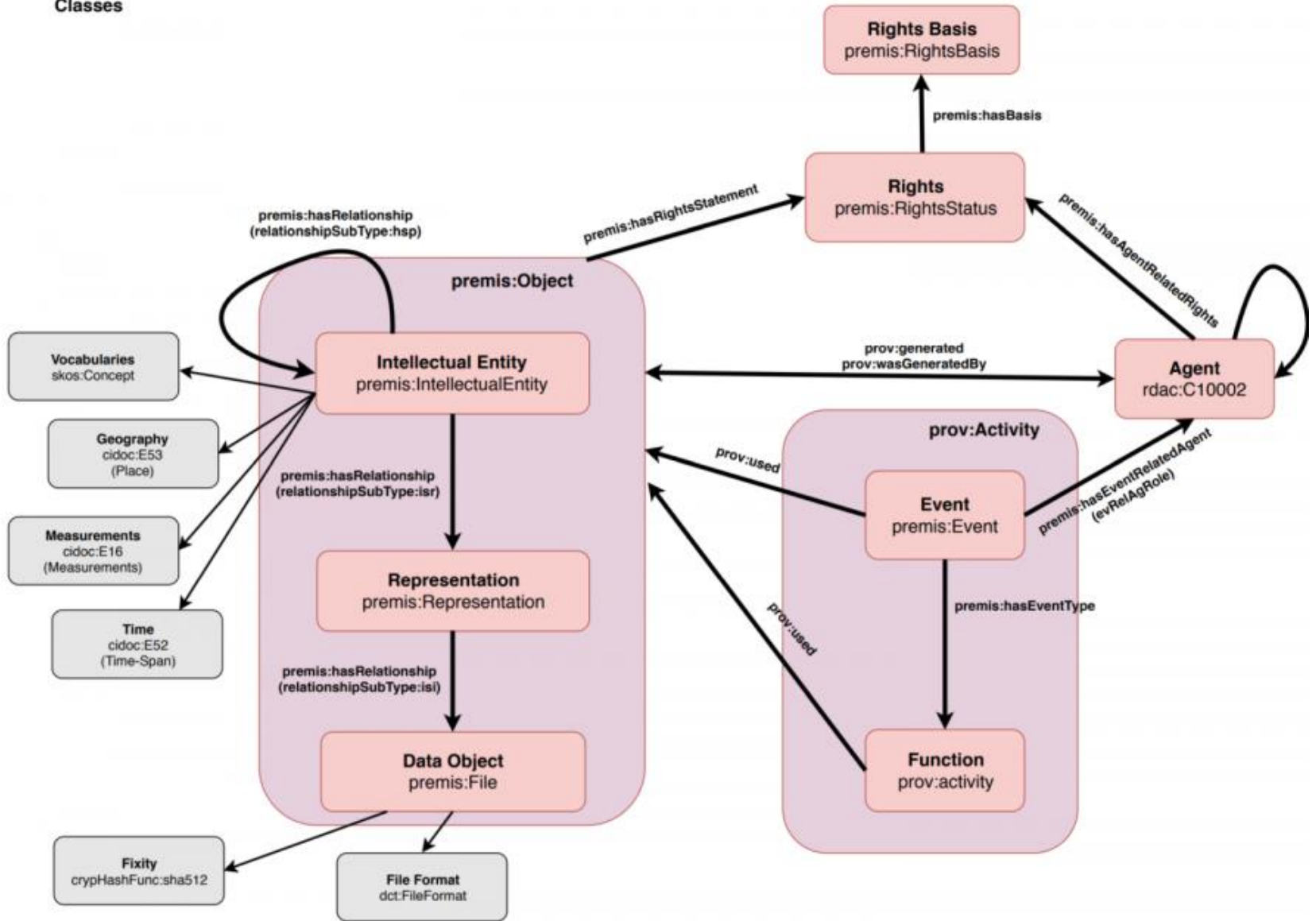
- The Matterhorn RDF data model is again a joint development by the [State Archives Canton of Wallis](#) and docuteam. It is based both on the ICA standards for descriptive metadata (now ISAD/ISAAR/ISDF, in the future Records in Context (RiC)) and the OAIS Information Model.
- The class model of Matterhorn RDF is mainly based on the PREMIS 3 ontology and the library standard RDA (but not the FRBR data model). The PREMI - classes are enriched with attributes from other ontologies to be able to model descriptive metadata.
- Most important ontologies:
 - PREMIS 3
 - RDA (Ressource Description and Access)
 - DC, DC Terms, CIDOC, SKOS
- Matterhorn RDF will also be used for the AtoM 3 – Proof of Concept by Artefactual
- All informations are online:
<https://wiki.docuteam.ch/doku.php?id=docuteam:matterhornrdf>

Conceptual Model

Most important conceptual change compared to Matterhorn METS: Representations!



Classes



- F3:
 - quality assurance: assigning PID for submission root node
 - [ingest: ...]
 - storage:
 - prepare METS (primarily dealing with XML namespaces)
 - create intermediary Fedora files (FOXML), remember PIDs
 - upload intermediary Fedora files

- F4:
 - One take: Go through nodes and upload one after the other, remembering identifiers

- SIP preparation as usual, for example using packer or automated through feeder
- Ingest using feeder
 - Storage as part of the ingest process
 - Persistent identifiers are stored in the mets.xml as usual

Yes, you can:

- https://wiki.docuteam.ch/doku.php?id=docuteam:packer_500
- Upload your example to http://bit.ly/1_inbox (pwd: docudemo)
- Wait for ingest using feeder
- Check output in http://bit.ly/4_output (pwd: docudemo)

- (Semi-)automated process
- Needs to convert XML to triples
 - Metadata as Literals
- In F3, we're dealing with objects and their (optional) datastreams (the binaries).
- In F4 (or more general: LDP), we're talking about resources in general, and more specifically about
 - Containers/RDF source (the F3 objects)
 - Non-RDF source (the F3 datastreams)

- In Fedora 3, the PID had a semantic note attached to it (organizational context), and at the same time some of the storage adapters also used PIDs
- The persistent identifier is used internally as the public reference in the sense of DOI or ARK
- Fedora 4 offers more flexibility, both for creating the internal identifiers as well as adding individual „minters“ or external services.