Sharing, Versioning and Annotating SBML Models using the ***e!DAL*** Data Repository API

Daniel Arend (Research Group ‘Bioinformatics & Information Technology‘)
IPK Gatersleben - 19 September 2013

IPK - Leibniz Institut of Plant Genetics and Crop Plant Research

Gatersleben

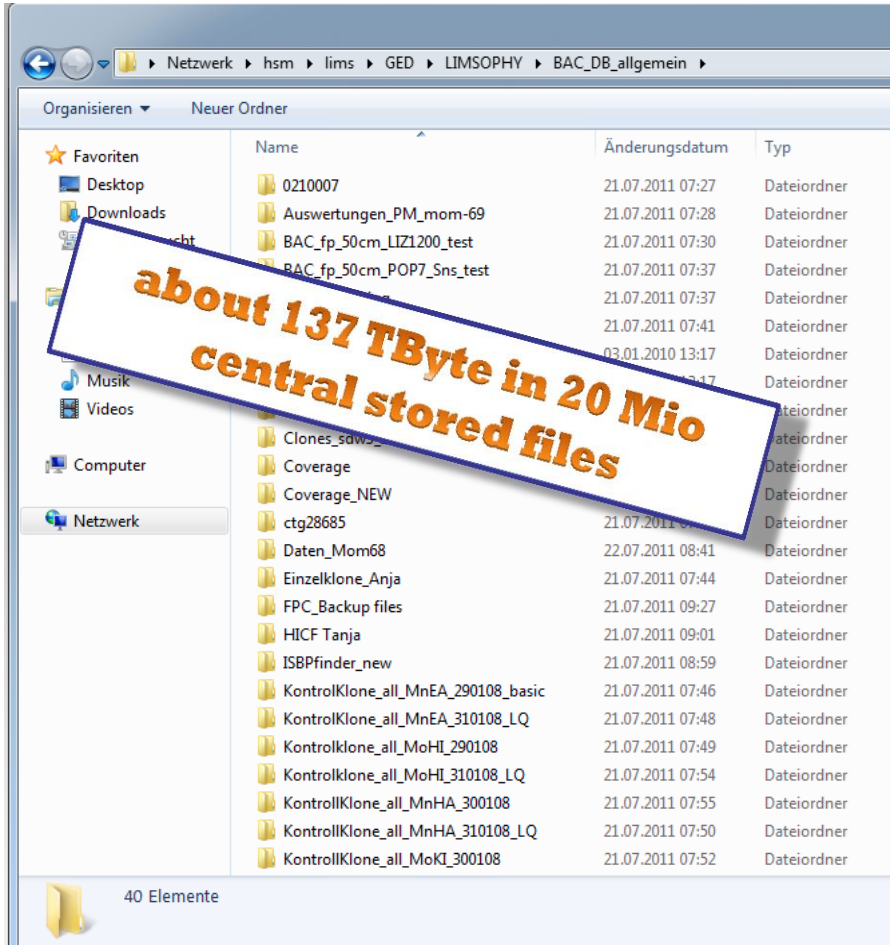
- IPK is 70 years old
- „**Magdeburger Börde**“: very fruitful soil
- source of the breeding industry in Germany
- total staff: ~550
- scientists: ~200
- 30 Research Groups

Research Topics:

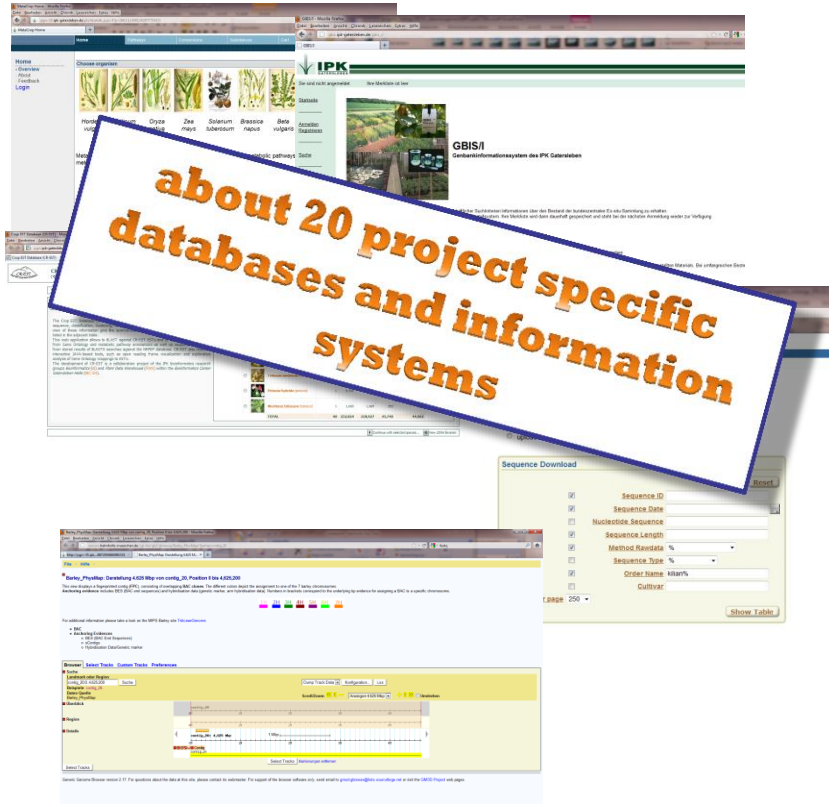
- **diversity of crop plants**
- **dynamics of plant genomes**
- **integrative biology of plant performance**

Status @ IPK: Data Access

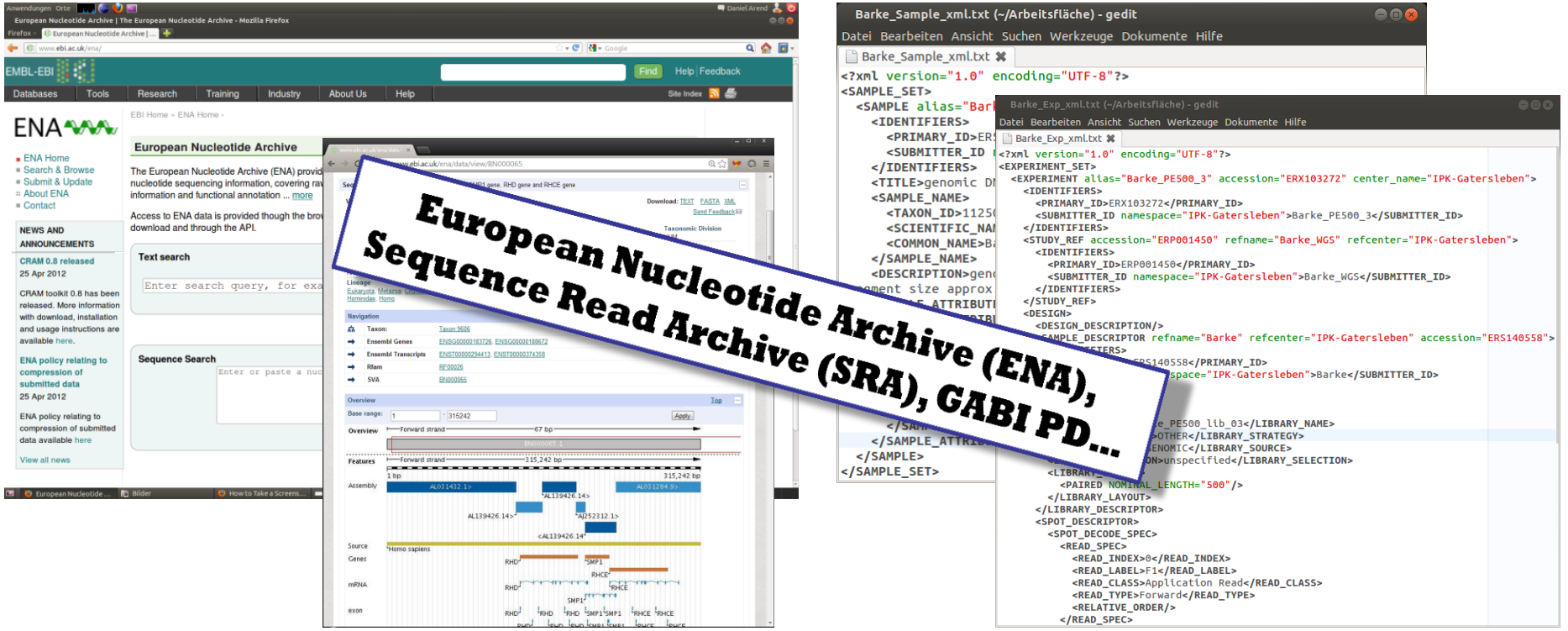
- File system exploration/versioning/indexing (desktop search)



- Databases/Web applications



Existing Repositories – NGS Data



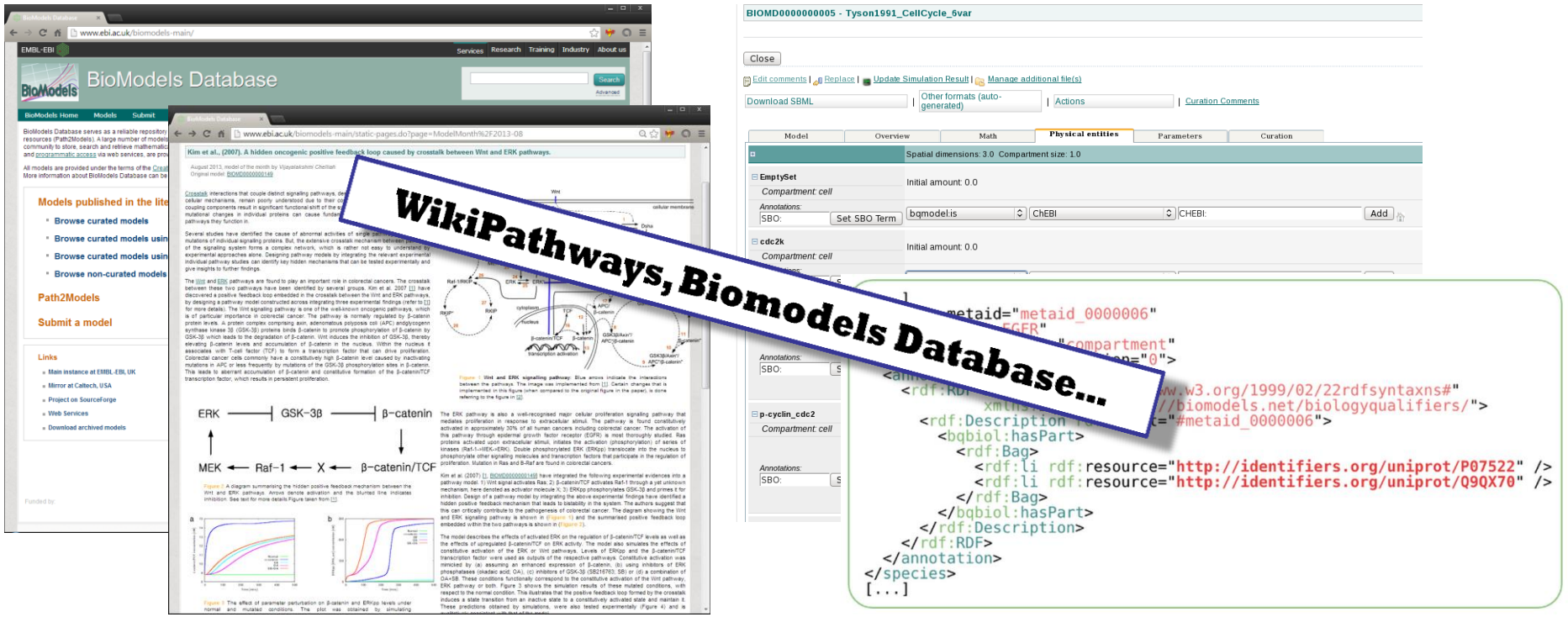
e.g. European Nucleotide Archive (ENA)

- next-generation sequence raw data
- proprietary metadata schema
- ENA accession numbers
- upload over FTP, Aspera protocol

limitations :

- for NGS raw data only
- 5 xml files -> difficult to maintain
- proprietary (none persistent) identifiers
- upload time-consuming

Existing Repositories – System Biology Data



e.g. BioModels Database

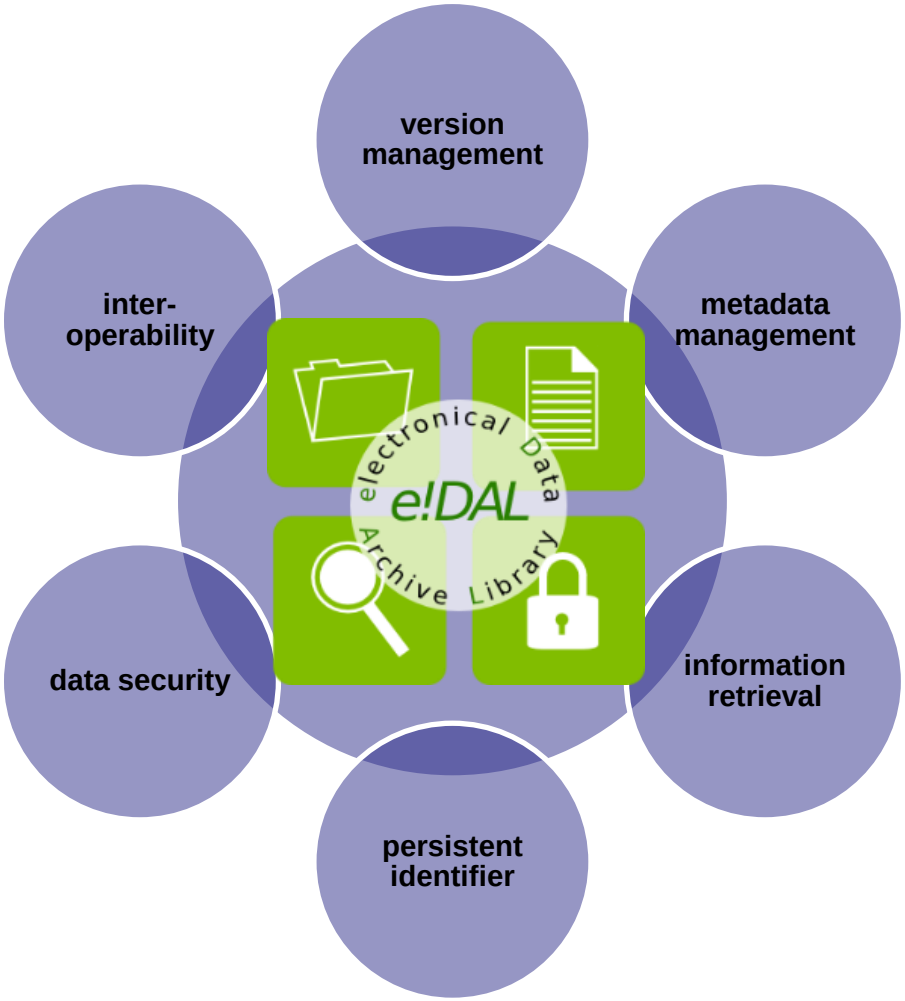
- for models of biological processes
- proprietary metadata schema
- BioModels ID
- upload using FTP protocol

limitations :

- dedicated to biological processes
- mixture of different schema (dublin core, vCard…)
- proprietary (none persistent) identifiers
- upload over web-interface -> no API available

e!DAL-MetaData API

- ***e****lectronical* ***D****ata* ***A****rchive* ***L****ibrary*
- enhanced file system like storage for any data, e.g. primary data
- features for long term preservation
- based on recommendation of *DataCite*, *DFG* and *nestor*
- Java 1.7 API



DFG Deutsche Forschungsgemeinschaft

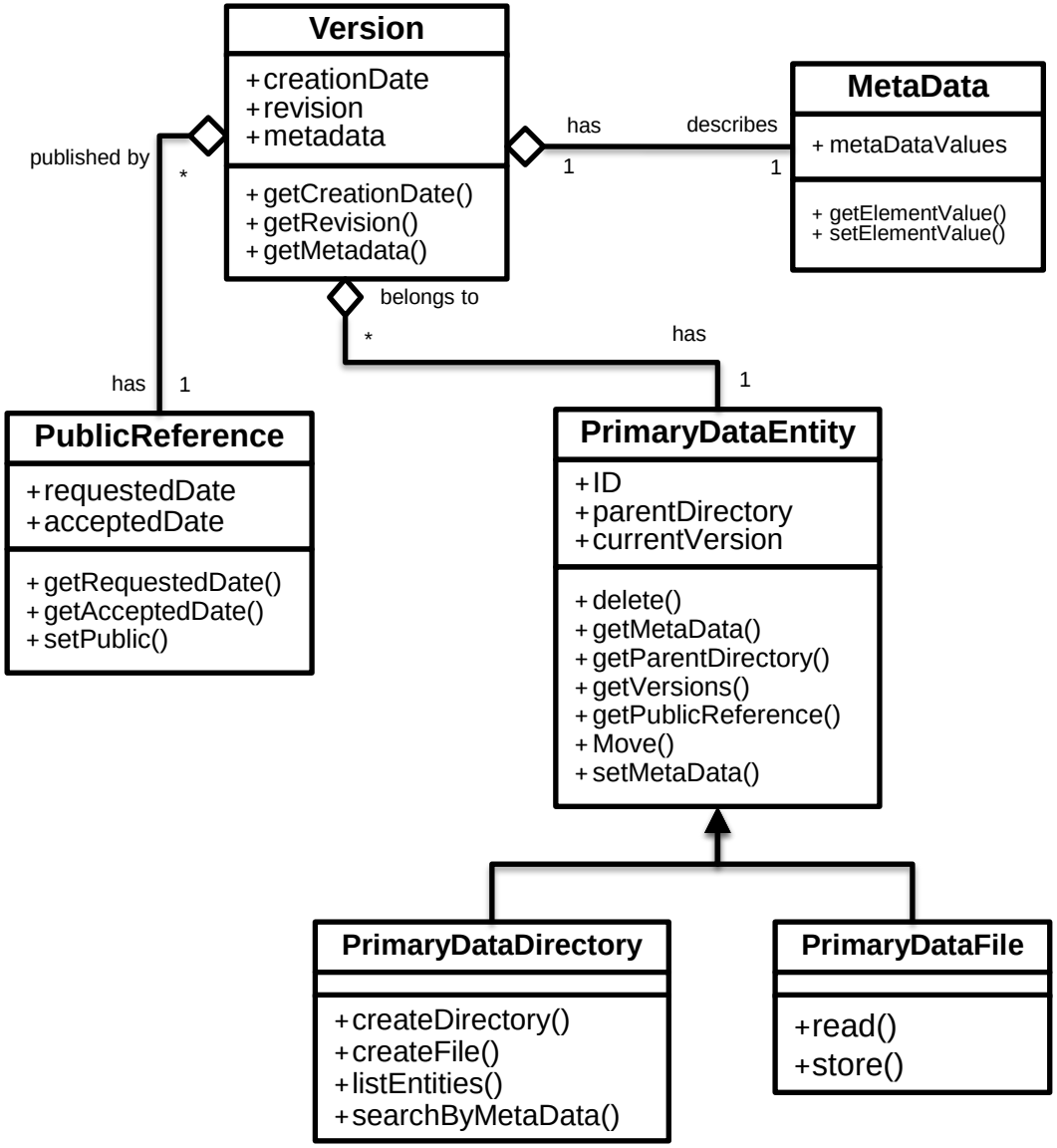
nestor

DataCite

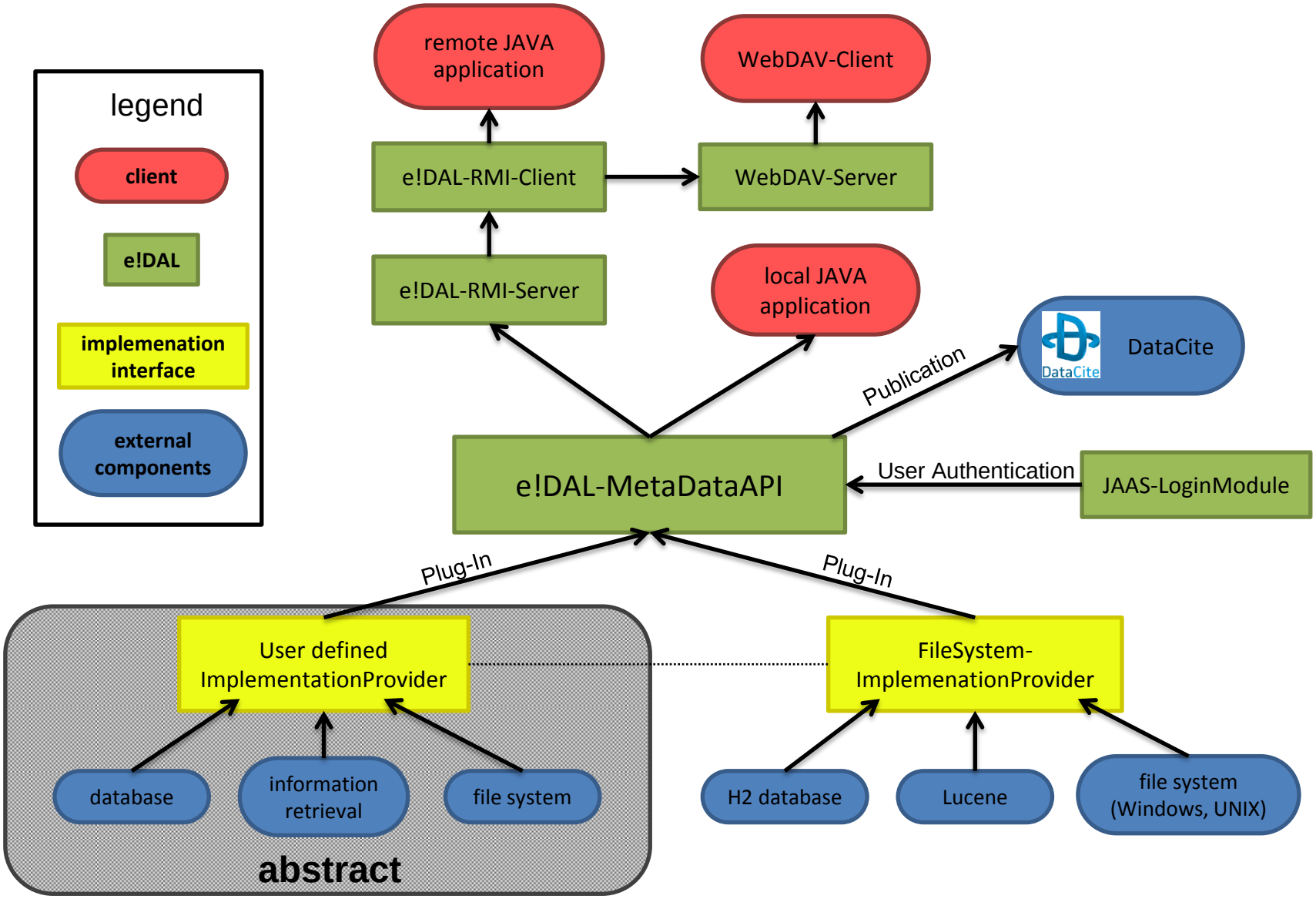
Arend D, Lange M, Colmsee C, Flemming S, Chen J & Scholz U: **The e!DAL JAVA-API: Store, Share and Cite Primary Data In Life Sciences.** In IEEE International Conference on Bioinformatics and Biomedicine 2012; 511-515;
DOI: http://dx.doi.org/10.5447/IPK/2012/13

e!DAL – Core Features

- file system like organization
- versions and metadata
- client server and embedded
- mountable as remote file system
- support homogeneous authentication and authorization (e.g. Windows/Unix/Kerberos)
- assign persistent identifiers (e.g. DOI)

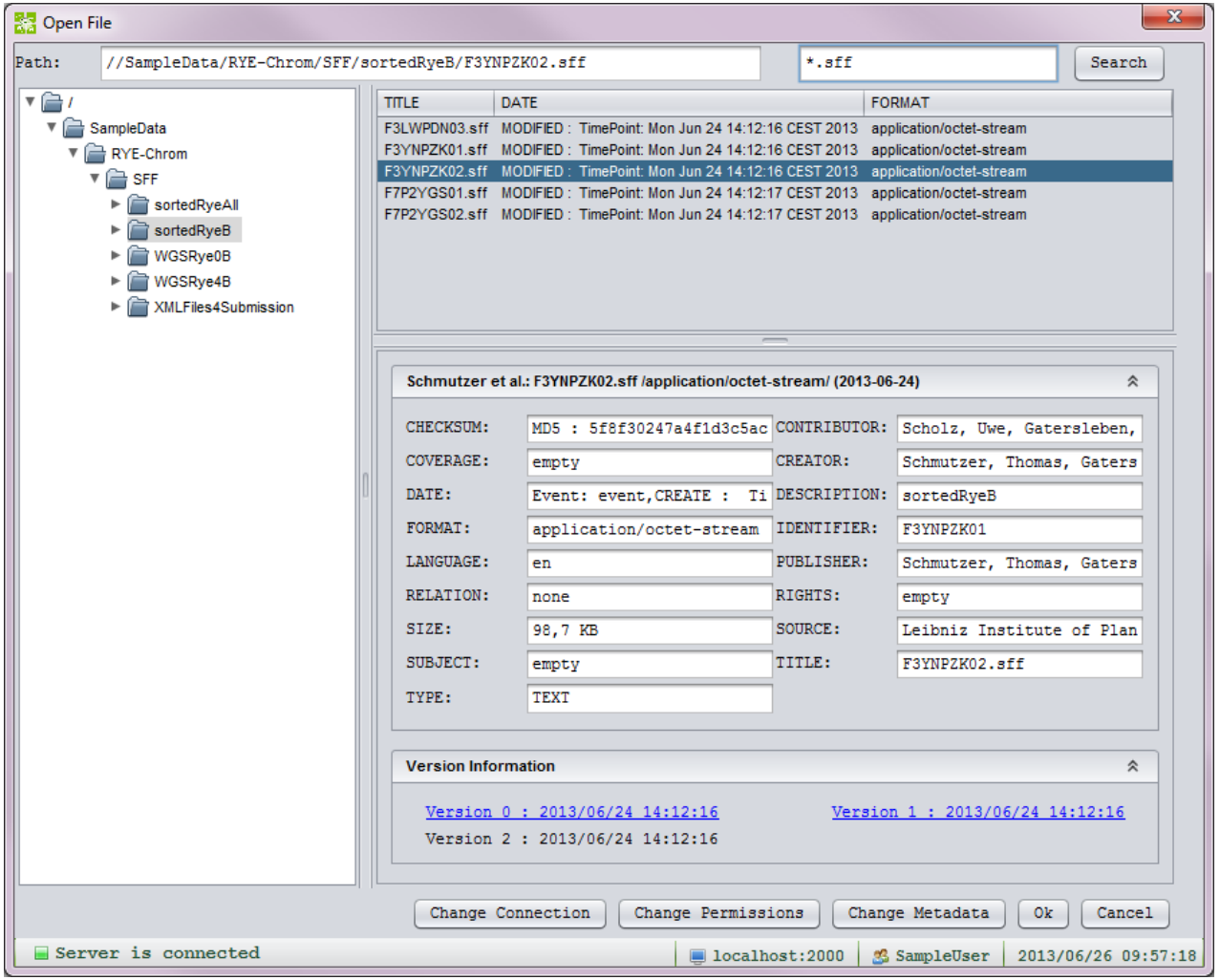


e!DAL – Architecture & Interoperability



e!DAL – Metadata & Information Retrieval

- annotation with technical meta data
- *DublinCore*[1]
  - 15 elements
  - FORMAT, AUTHOR, DESCRIPTION…
- general full text search
- element specific search filtering
- increase reusability & retrievability

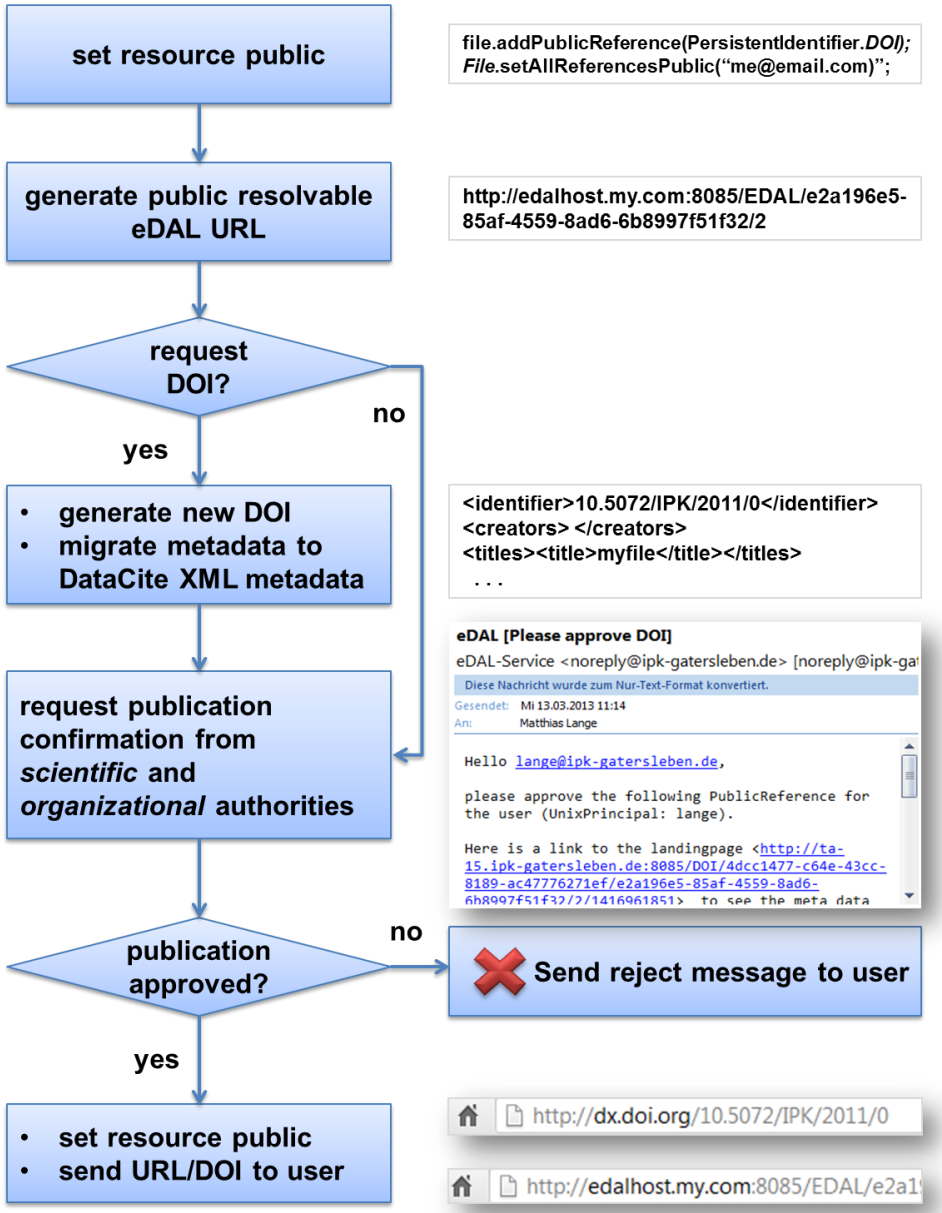


graphical user interface – e!DAL-FileChooser

[1] **http://dublincore.org/**

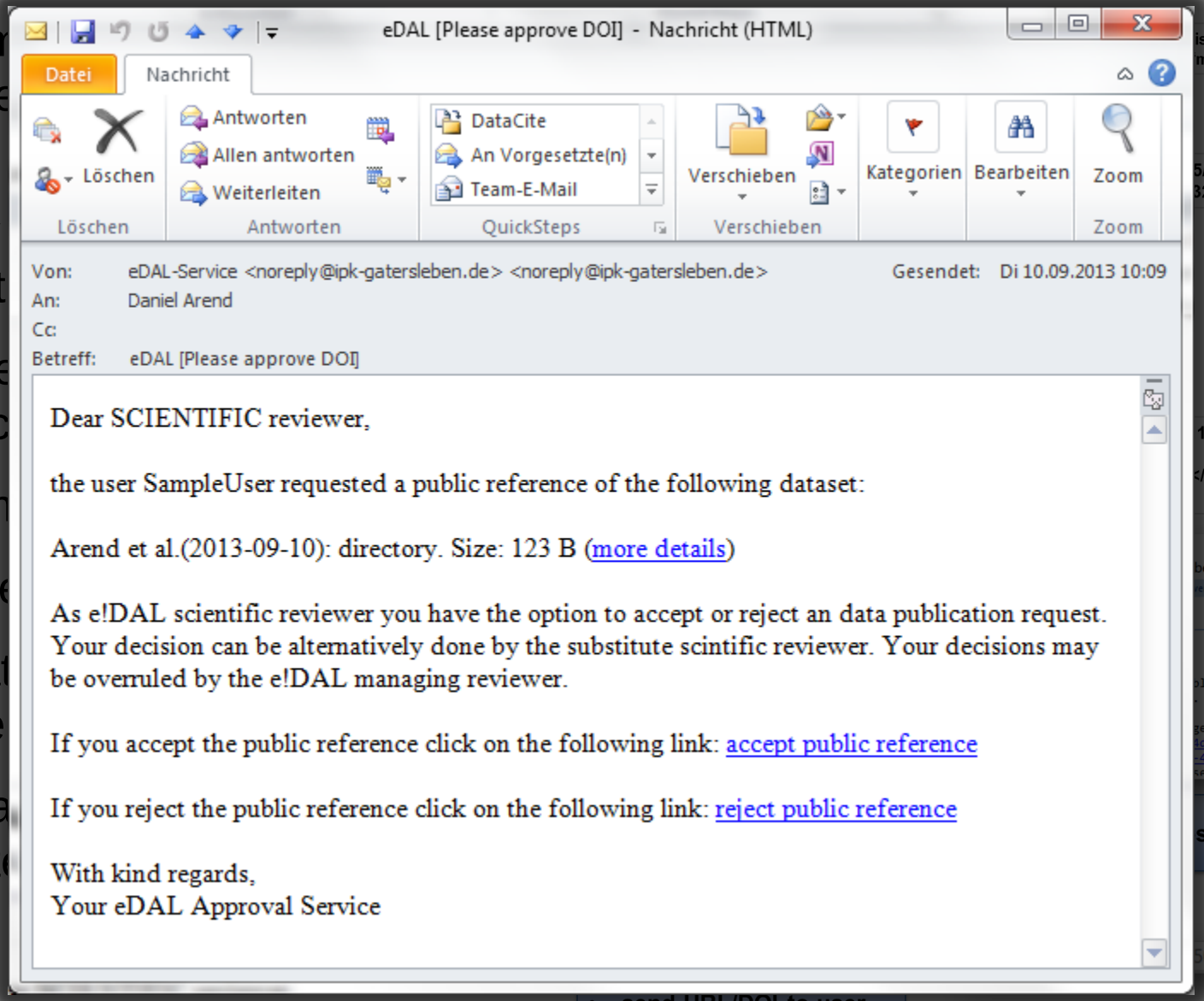
e!DAL – Persistent Identifiers

- supplementary data frequently published in condensed form (tables, charts)
- primary data rarely part of a publication
- published as proprietary identifier (DB accessions, URLs)
- problem in long-term stable citation
- interface: ***PublicReference***
- registration of global persistent identifiers, e.g. DOI
- approval process to validate requested identifiers



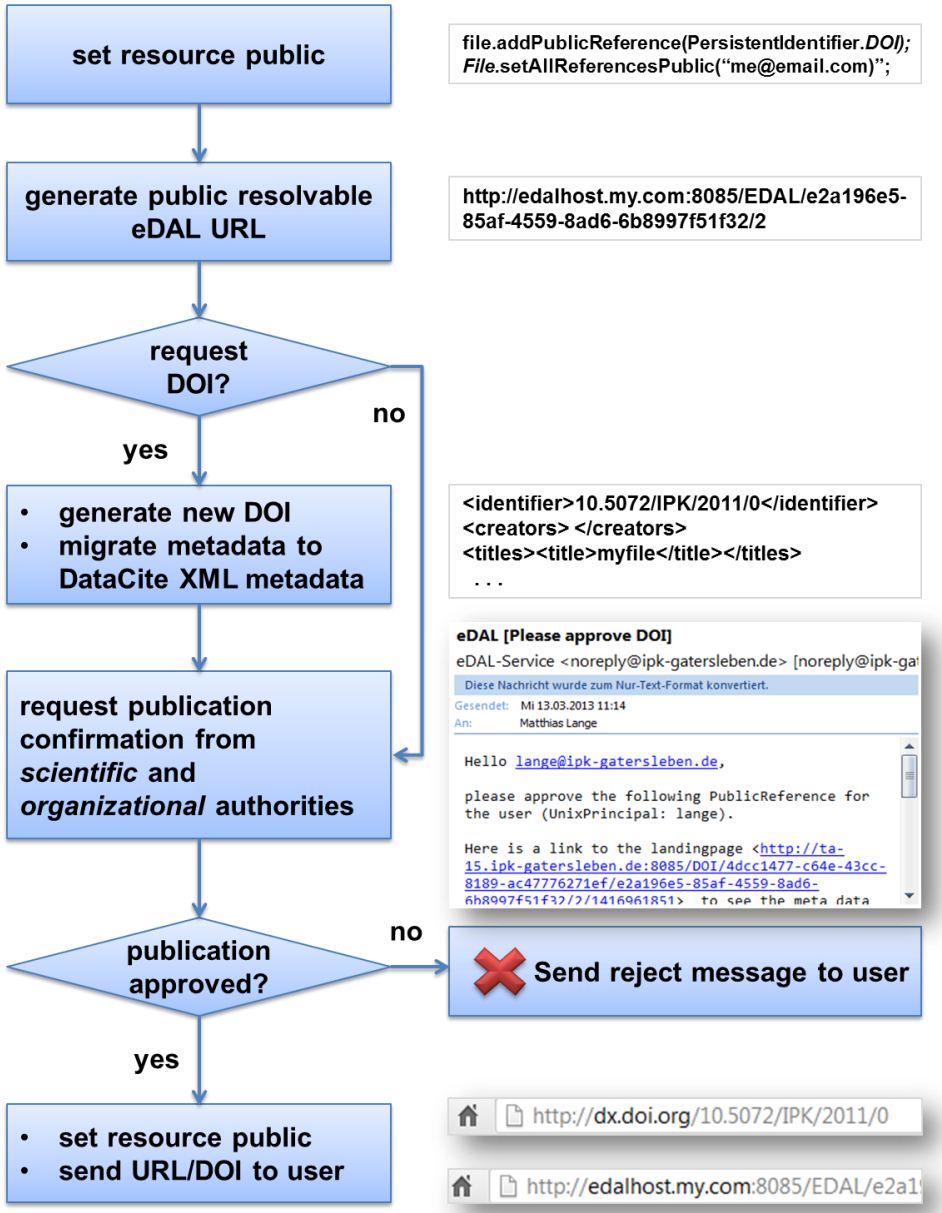
e!DAL – Persistent Identifiers

- supplem
  publishe
  (tables,
- primary
  publicat
- publishe
  (DB acc
- problem
- interface
- registrat
  identifie
- approva
  request



e!DAL – Persistent Identifiers

- supplementary data frequently published in condensed form (tables, charts)
- primary data rarely part of a publication
- published as proprietary identifier (DB accessions, URLs)
- problem in long-term stable citation
- interface: ***PublicReference***
- registration of global persistent identifiers, e.g. DOI
- approval process to validate requested identifiers



e!DAL – Persistent Identifiers

- sup
  pub
  (tab
- prir
  pub
- pub
  (DE
- pro
- inte
- reg
  ide
- app
  req

Metadata - poster.pdf

194.94.136.249:8085/DOI/abf8ce47-7e7b-48a5-9e39-ef396a65026d/97ff653e-8856-47a6-b00e-f6ece3991b33/1/1847940088

e!DAL - store, cite and share primary data

Data File Description (always quote citation when using data !)

Citation: Arend et al.(2013-09-10): poster.pdf. Size: 6,1 MB;

Abstract: This is just a longer example description to test the API

Download Data

Download: poster.pdf (application/pdf)

Path: //directory/poster.pdf

Parent Directory: directory

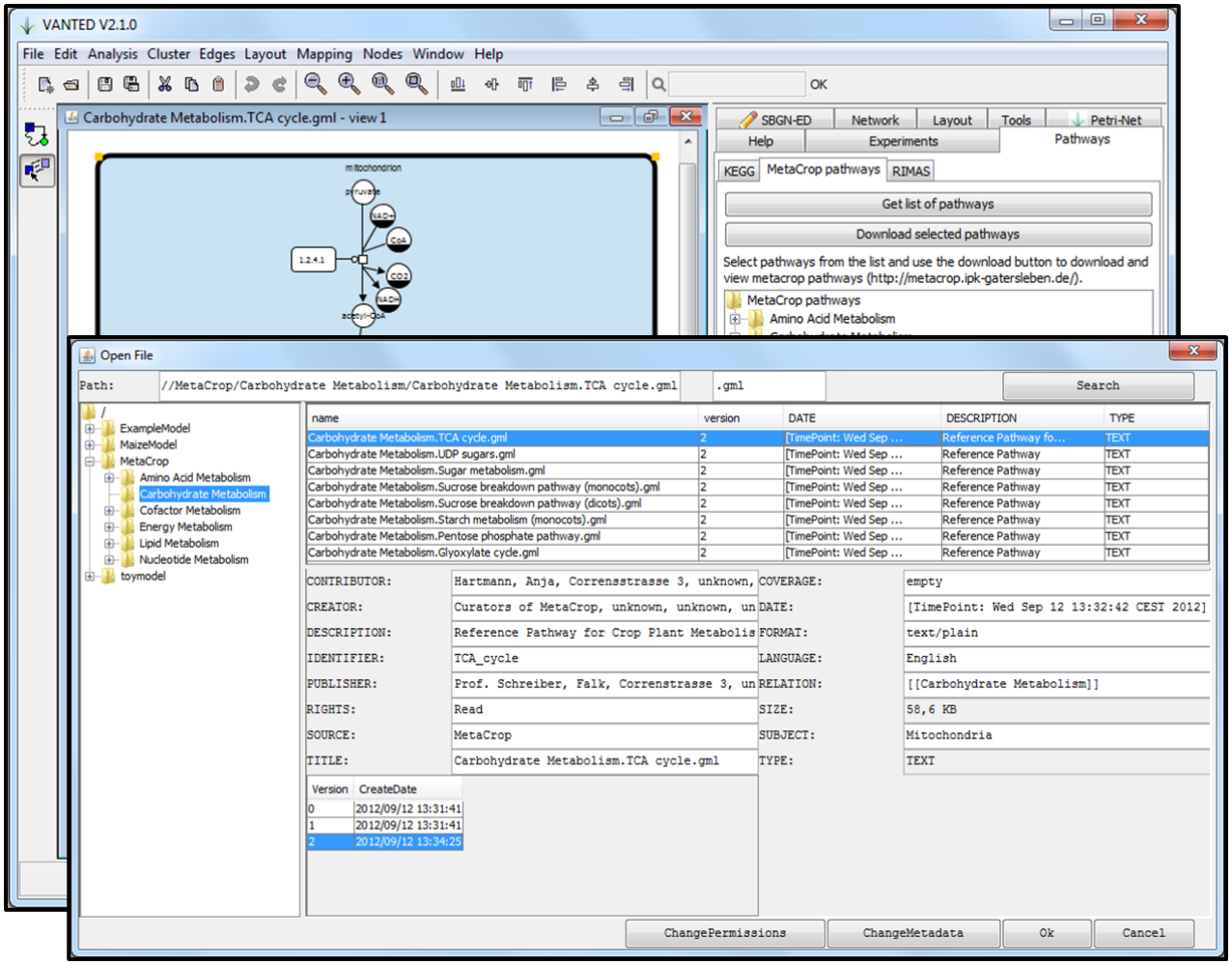
Revision: 1 **CreationDate:** Tue Sep 10 10:08:29 CEST 2013 **RevisionDate:** Tue Sep 10 10:08:30 CEST 2013

Meta Data

CHECKSUM:	MD5 : 334aa36695bca683f53ee87c5ddd658	CONTRIBUTOR:	unknown
COVERAGE:	empty	CREATOR:	Arend, Daniel, Corrensstr. 3, Gatersleben, 06466, Germany
DATE:	Event: event CREATE : TimePoint: Tue Sep 10 10:08:29 CEST 2013 MODIFIED : TimePoint: Tue Sep 10 10:08:29 CEST 2013	DESCRIPTION:	This is just a longer example description to test the API
FORMAT:	application/pdf	IDENTIFIER:	Unknown_ID
LANGUAGE:	de_DE	PUBLISHER:	IPK, Gatersleben, 01234, germany
RELATION:	none	RIGHTS:	opensource
SIZE:	6,1 MB	SOURCE:	empty
SUBJECT:	empty	TITLE:	poster.pdf
TYPE:	TEXT		

Use case for SysBio data – VANTED plugin

- embed GUI into VANTED
- use remote API
- centralized repository
- secured multiuser access
- manage different version and annotate with metadata

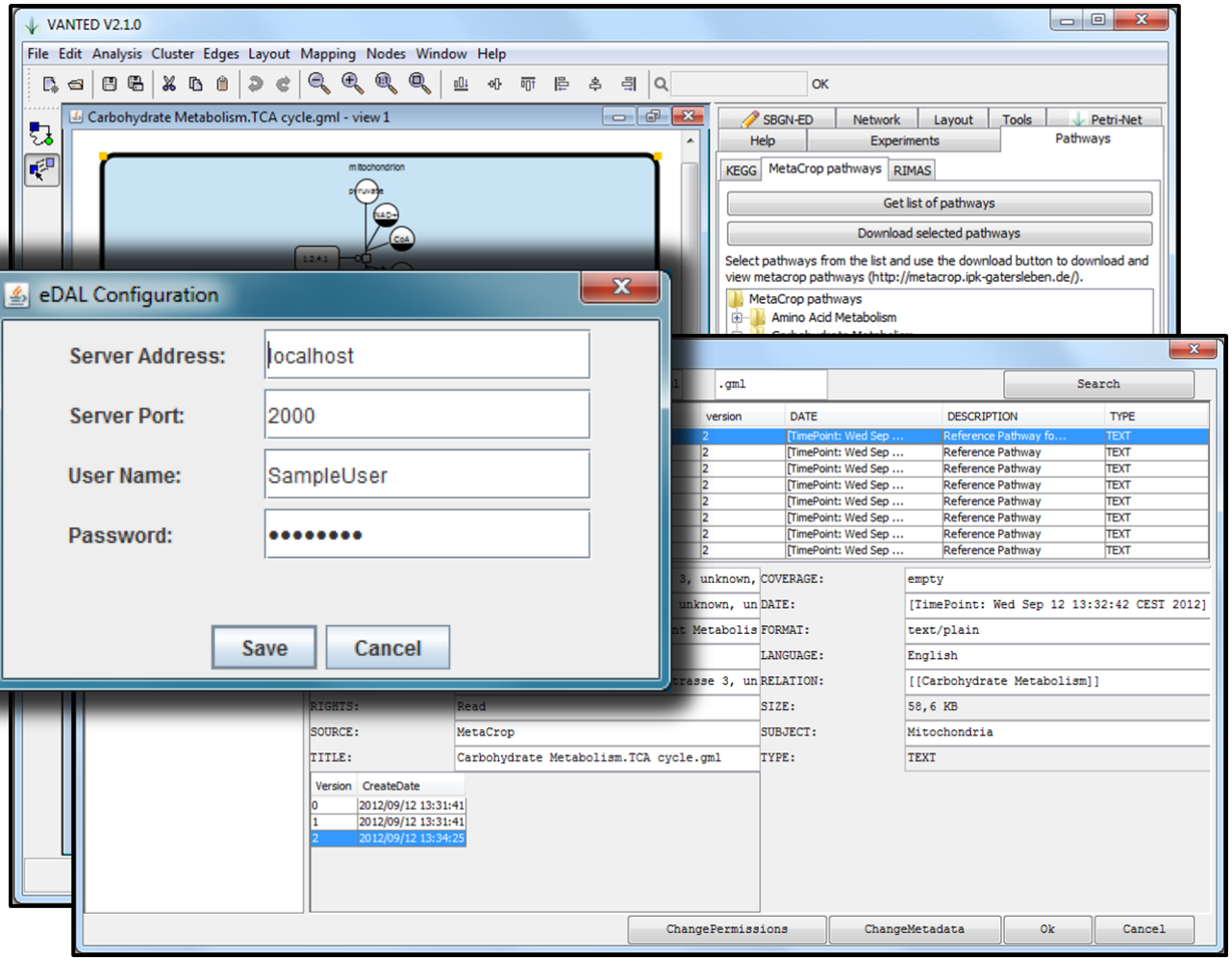


Embed into tools: e.g. VANTED pathway editor[1]

[1] (Rohn et al. BMC Systems Biology, 2012); DOI: 10.1186/1752-0509-6-139

Use case for SysBio data – VANTED plugin

- embed GUI into VANTED
- use remote API
- centralized repository
- secured multiuser access
- manage different version and annotate with metadata

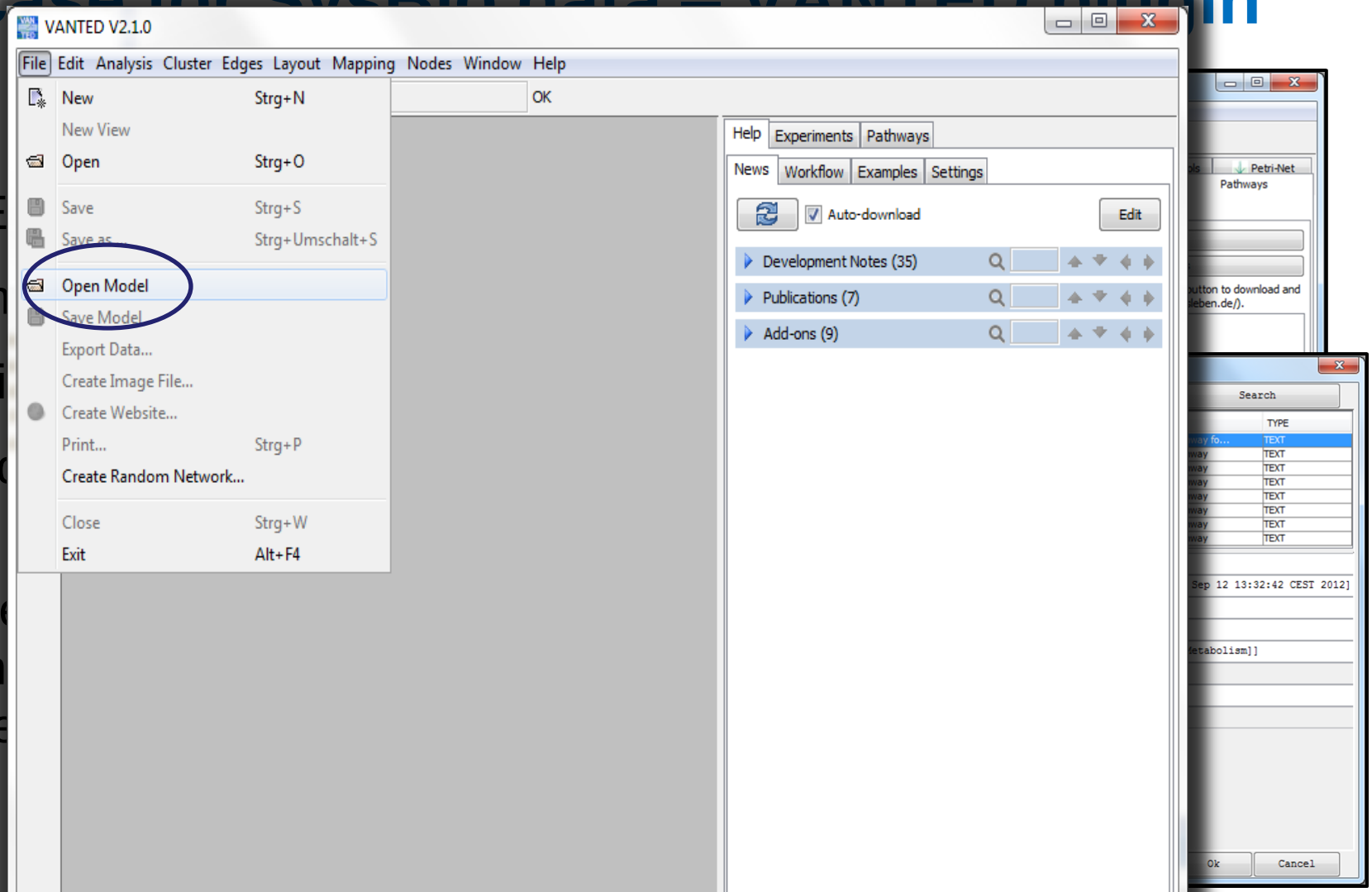


Embed into tools: e.g. VANTED pathway editor[1]

[1] (Rohn et al. BMC Systems Biology, 2012); DOI: 10.1186/1752-0509-6-139

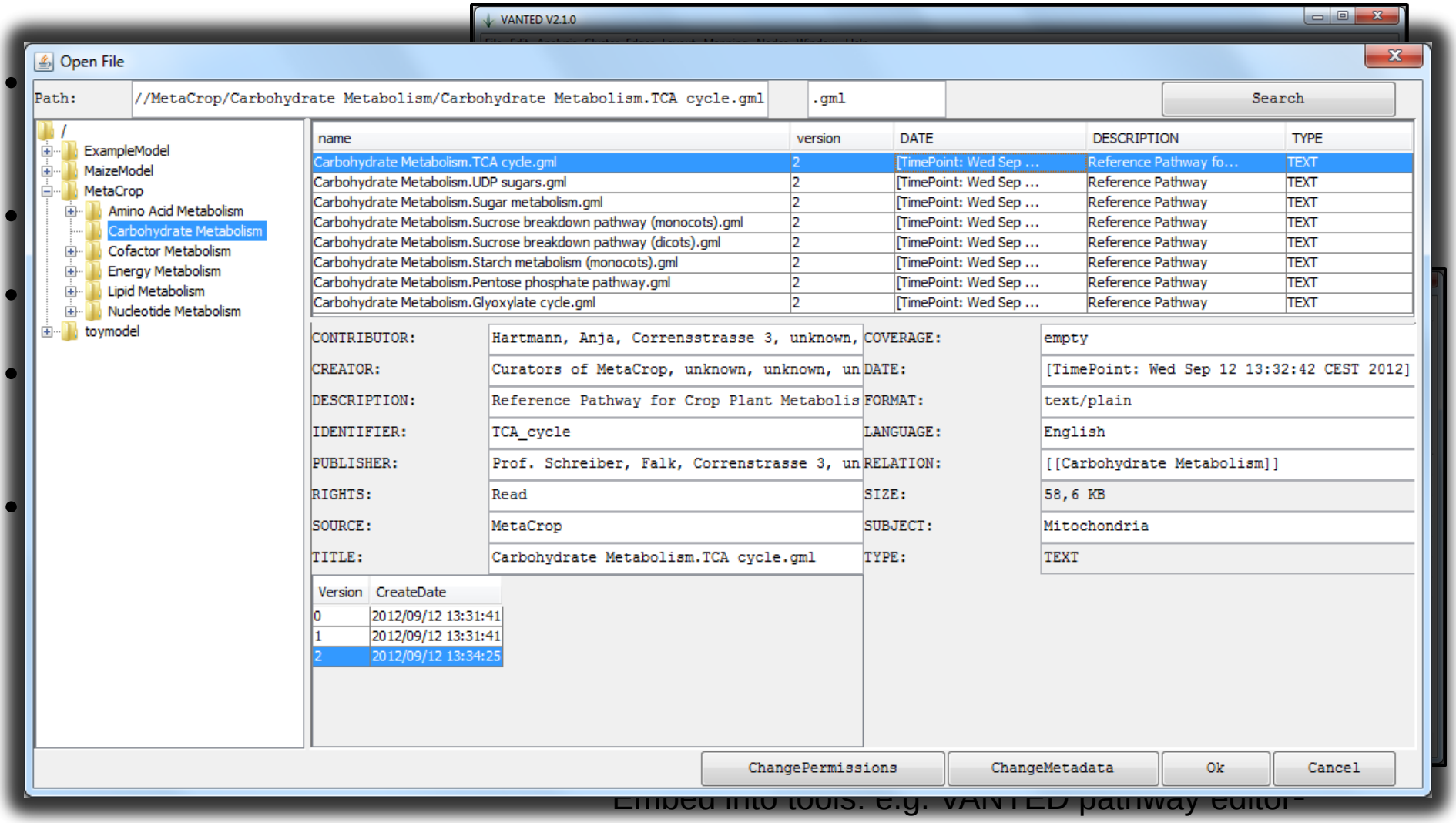
Use case for SysBio data – VANTED plugin

- embed
  VANTE
- use ren
- centrali
- secured
  access
- manage
  version
  with me



[1] (Rohn et al. BM

Use case for SysBio data – VANTED plugin

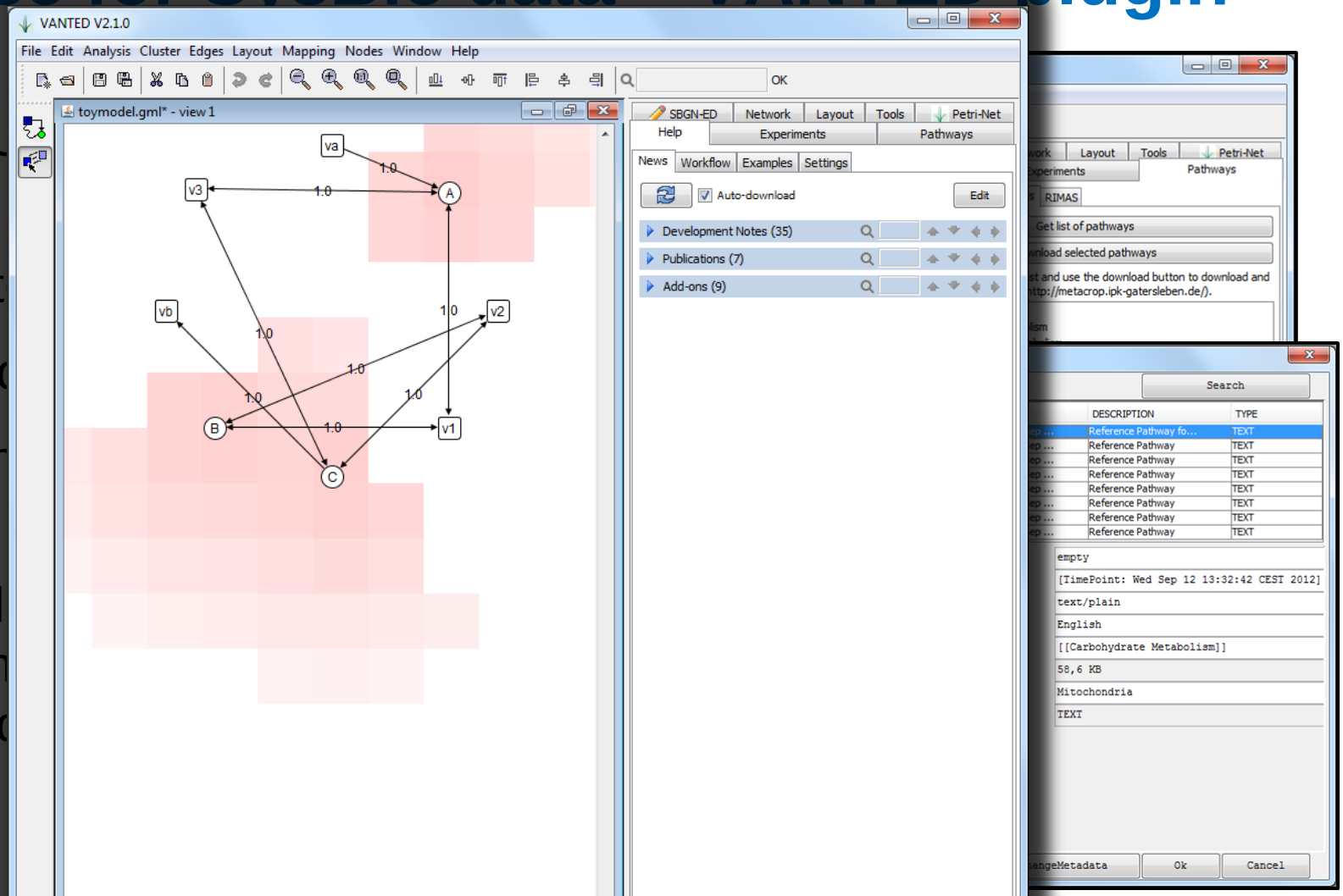


- Embed into tools: e.g. VANTED pathway editor

[1] (Rohn et al. BMC Systems Biology, 2012); DOI: 10.1186/1752-0509-6-139

Use case for SysBio data – VANTED plugin

- embed GU
  VANTED
- use remot
- centralized
- secured m
  access
- manage d
  version an
  with metad



ay editor[1]

[1] (Rohn et al. BMC Systems Biology, 2012), DOI: 10.1186/1752-0509-6-139

Summary

Advantages of e!DAL:

- easy to use → file system organization + version management
- universal → create public repository + provide persistent identifiers for your data / run as local repository e.g. within a workgroup to share data
- long-term stable → standardized metadata + functions for information retrieval
- available as API → integrate into existing infrastructures/tools (JAVA-based)
- **final release of *VANTED-e!DAL* plugin in the next few months**

Acknowledgements

Bioinformatics and Information Technology

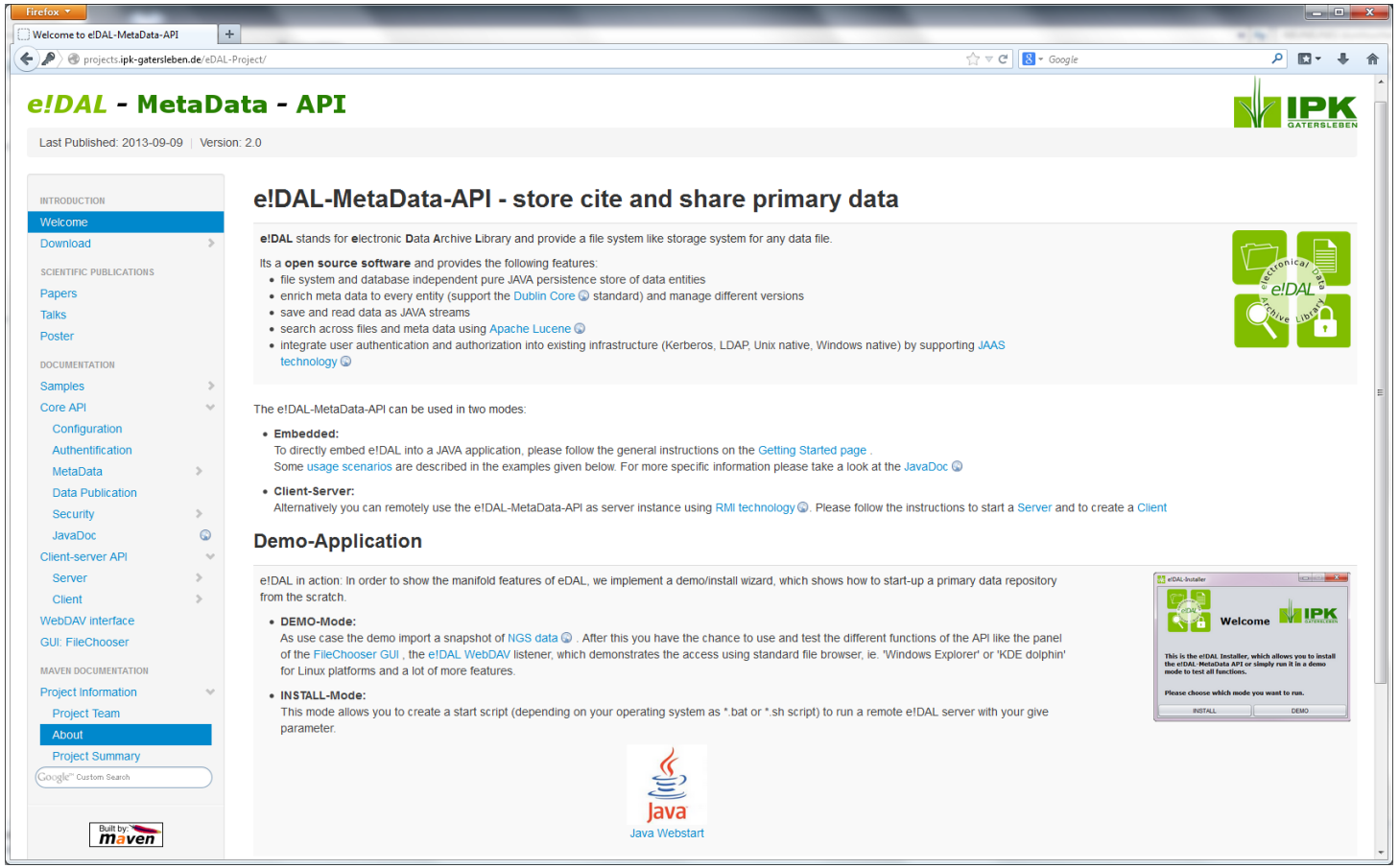
- Matthias Lange
- Jinbo Chen
- Christian Colmsee
- Steffen Flemming
- Denny Hecht
- Uwe Scholz
- Heiko Miehe
- Thomas Münch

Plant Bioinformatics

- Anja Hartmann
- Tobias Czauderna
- Falk Schreiber

This work is supported by the German federal Ministry of Education and Research (FKZ: 031A053A/B/C) and the European Commission within its 7th Framework Program, under the thematic area “Infrastructures", contract number 283496.

Thank you for your attention!



http://edal.ipk-gatersleben.de/

https://code.google.com/p/edal-metadata-api/