

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

about 137 TByte in 20 Mio
central stored files

Existing Repositories – NGS Data

e.g. European Nucleotide Archive (ENA)

limitations :

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



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

Existing Repositories – System Biology Data

BioModels Database

Search for:

Models published in the literature

- Browse curated models
- Browse curated models using Path2Models
- Browse non-curated models

Path2Models

Submit a model

Links

- Main instance at EMBL, EBI, UK
- Mirror at Caltech, USA
- Project on SourceForge
- Web Services
- Download archived models

Kim et al. (2007). A hidden oncogenic positive feedback loop caused by crosstalk between Wnt and ERK pathways.

August 2013, model of the month by Vijayakrishnan Chelvan

Original model: [BIOCID000000149](#)

Crosstalk interactions that couple distinct signaling pathways, cellular mechanisms, remain poorly understood due to their coupling components result in significant functional shift of the molecular changes in individual proteins can cause further pathways they function in.

Several studies have identified the cause of abnormal activities of single proteins and the underlying mechanisms, remain poorly understood due to their coupling components result in significant functional shift of the molecular changes in individual proteins can cause further pathways they function in.

The Wnt and ERK pathways are found to play an important role in colorectal cancers. The crosstalk between these two pathways have been identified by several groups. Kim et al. (2007) [1] have observed a positive feedback loop embedded in the crosstalk between the Wnt and ERK pathways, by designing a pathway model constructed across integrating three experimental findings (refer to [1] for more details). The first signaling pathway is one of the well-known oncogenic pathways, which is of particular importance in colorectal cancer. The pathway is normally regulated by β -catenin protein levels. A protein complex comprising β -catenin, adenomatous polyposis coli (APC) and glycogen synthase kinase-3 (GSK-3 β) prevents translocation of β -catenin to the nucleus. When the APC/GSK-3 β complex is degraded, β -catenin is free to translocate into the nucleus, where it associates with T-cell factor (TCF) to form a transcription factor that can drive proliferation. Colorectal cancer cells commonly have a constitutively high β -catenin level caused by inactivating mutations in APC or less frequently by mutations of the GSK-3 β phosphorylation sites in β -catenin. This leads to aberrant accumulation of β -catenin and constitutive formation of the β -catenin/TCF transcription factor, which results in persistent proliferation.

The ERK pathway is also a well-recognized major cellular proliferation signaling pathway that mediates proliferation in response to extracellular stimuli. The pathway is found constitutively activated in approximately 30% of all human cancers including colorectal cancer. The activation of the pathway through extracellular growth factor receptor (EGFR) is most thoroughly studied. Ras proteins activated upon extracellular stimuli, initiate the activation (phosphorylation) of a series of kinases (Raf-1-MEK-ERK). Double phosphorylated ERK (ERK2p) translocate into the nucleus to phosphorylate other signaling molecules and transcription factors that participate in the regulation of proliferation. Mutation in Ras and Raf-1 are found in colorectal cancers.

Kim et al. (2007) [1] BIOCID000000149 have integrated the following experimental evidences into a pathway model. (1) Wnt signal activates Raf-1. (2) ERK2p phosphorylates GSK-3 β and primes it for inhibition. Design of a pathway model by integrating the above experimental findings have identified a hidden positive feedback mechanism that leads to stability in the system. The authors suggest that this can critically contribute to the pathogenesis of colorectal cancer. The diagram showing the Wnt and ERK signaling pathways is shown in Figure 1 and the summarized positive feedback loop embedded within the two pathways is shown in Figure 2.

The model describes the effects of activated ERK on the regulation of β -catenin/TCF levels as well as the effects of upregulated β -catenin/TCF on ERK activity. The model also simulates the effects of constitutive activation of the ERK or Wnt pathways. Levels of ERK2p and the β -catenin/TCF transcription factor were used as outputs of the respective pathways. Constitutive activation was imposed by (a) ensuring an enhanced expression of β -catenin, (b) using inhibitors of ERK phosphatases (shaded acid, CA), (c) inhibitors of GSK-3 β (SH-221513, SH-3) as a combination of SH-221513. These conditions functionally correspond to the constitutive activation of the Wnt pathway, ERK pathway or both. Figure 3 shows the simulation results of these mutated conditions, with respect to the normal condition. This illustrates that the positive feedback loop formed by the crosstalk induces a state transition from an inactive state to a constitutively activated state and maintains it. These predictions obtained by simulations, were also tested experimentally (Figure 4) and a

WikiPathways, Biomodels Database...

BIOMD0000000005 - Tyson1991_CellCycle_6var

Close

Edit comments | Replace | Update Simulation Result | Manage additional file(s)

Download SBML | Other formats (auto-generated) | Actions | Curation Comments

Model	Overview	Math	Physical entities	Parameters	Curation
Spatial dimensions: 3.0 Compartment size: 1.0					
EmptySet	Initial amount 0.0				
Compartment cell					
Annotations:					
SBO:	Set SBO Term	bgmodelis	CHEBI	CHEBI	Add
cdc2k	Initial amount 0.0				
Compartment cell					
Annotations:					
SBO:	Set SBO Term				

metaid="metaid_00000006"

SGFR" "compartment"

compartment="0"

http://w3.org/1999/02/22rdfsyntaxns#"

http://biomodels.net/biologyqualifiers/">

#metaid_00000006">

<rdf:RDF

xmlns:biomodels="http://biomodels.net/biologyqualifiers/">

<rdf:Description

<bqbiol:hasPart

<rdf:Bag

<rdf:li rdf:resource="http://identifiers.org/uniprot/P07522" />

<rdf:li rdf:resource="http://identifiers.org/uniprot/Q9QX70" />

</rdf:Bag

</rdf:Description

</bqbiol:hasPart

</rdf:RDF

</annotation

</species

[...]

e.g. BioModels Database

limitations :

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

- 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

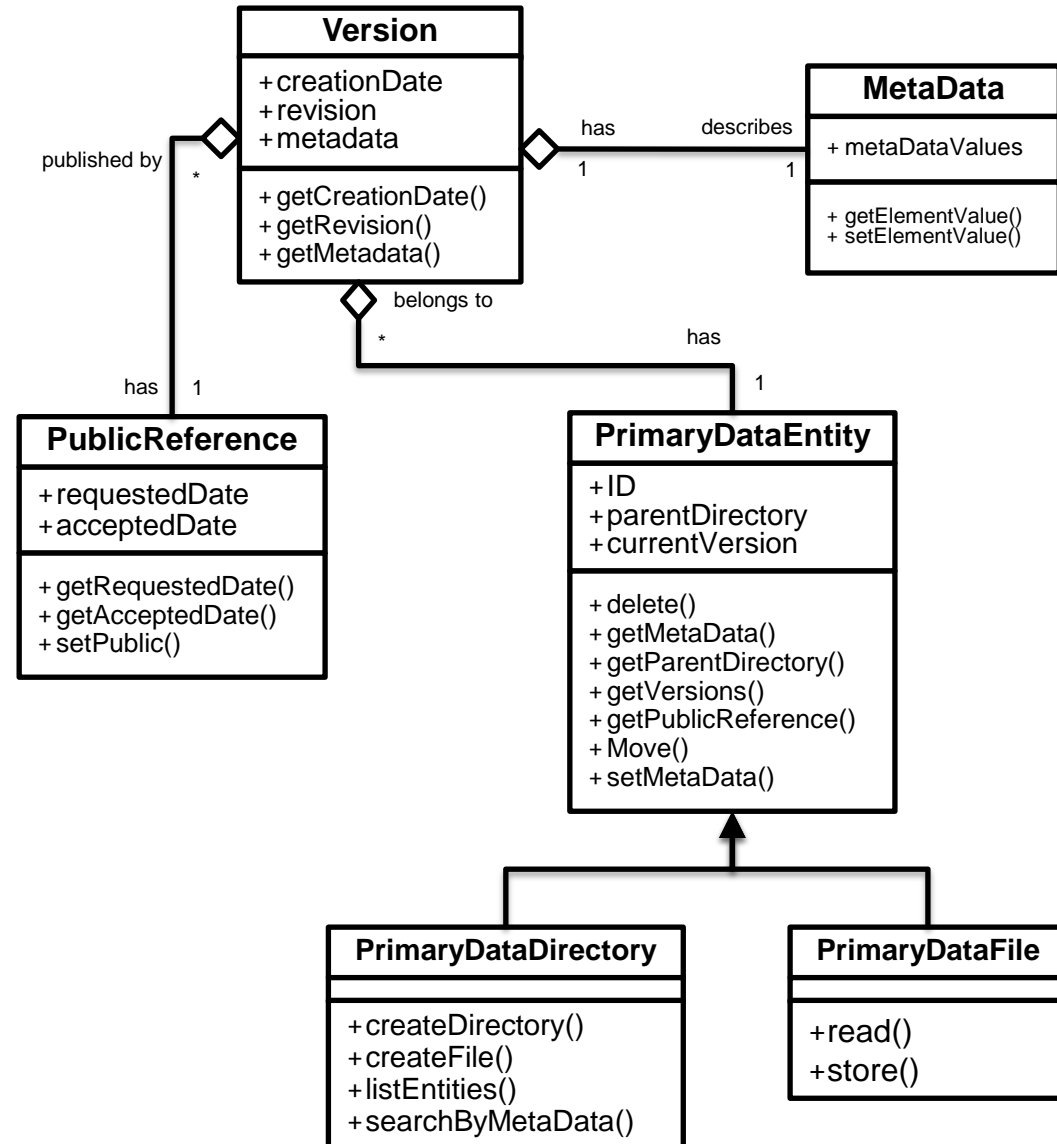
- *electronical Data Archive Library*
- enhanced file system like storage for any data, e.g. primary data
- features for long term preservation
- based on recommendation of *DataCite*, *DFG* and *nestor*
- Java 1.7 API



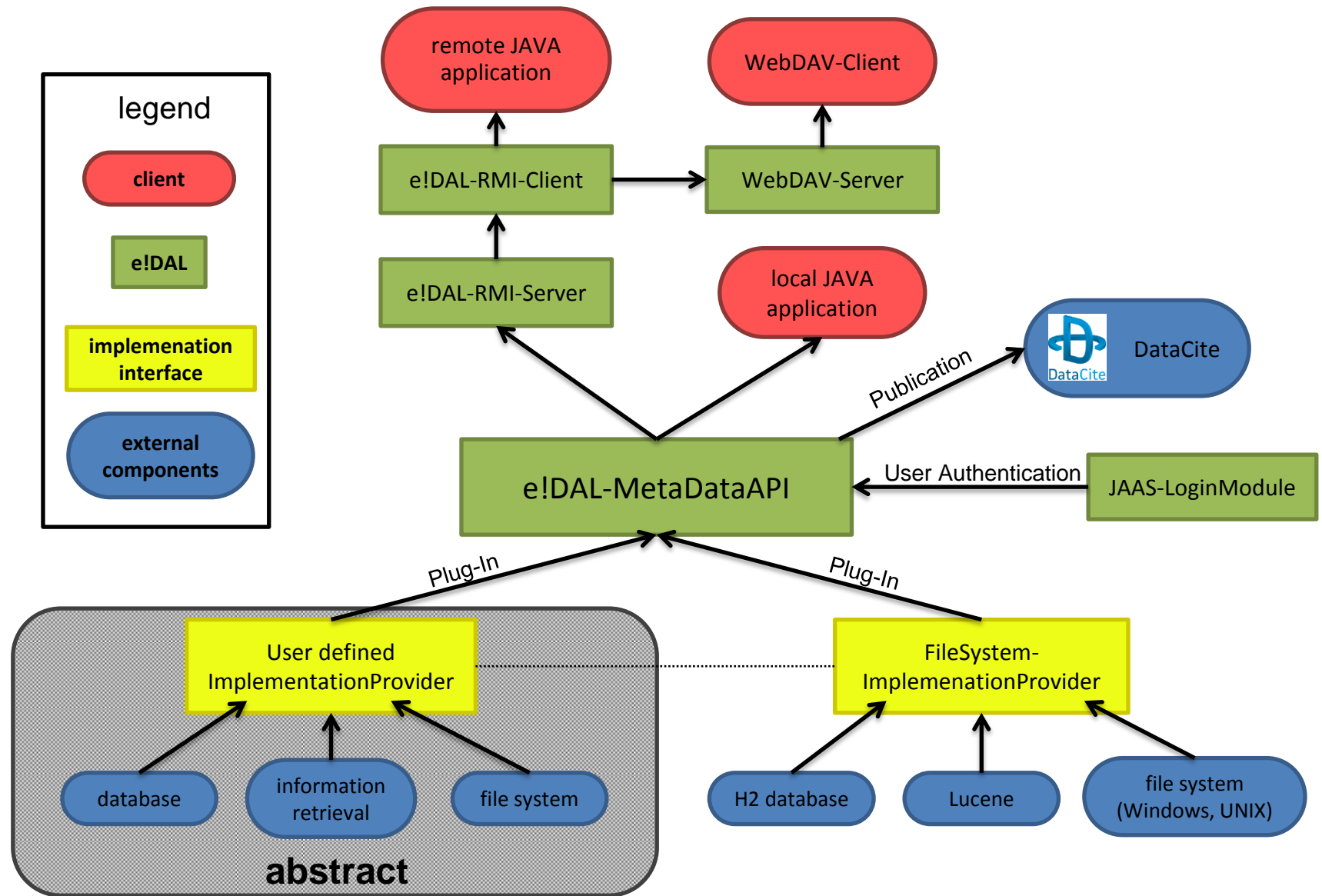
Arend D, Lange M, Colmsee C, Flemming S, Chen J & Scholz U: **The e!DAL JAVA-API: Store, Share and Cite Primary Data In Life Sciences.** In IEEE International Conference on Bioinformatics and Biomedicine 2012; 511-515;
DOI: <http://dx.doi.org/10.5447/IPK/2012/13>

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)

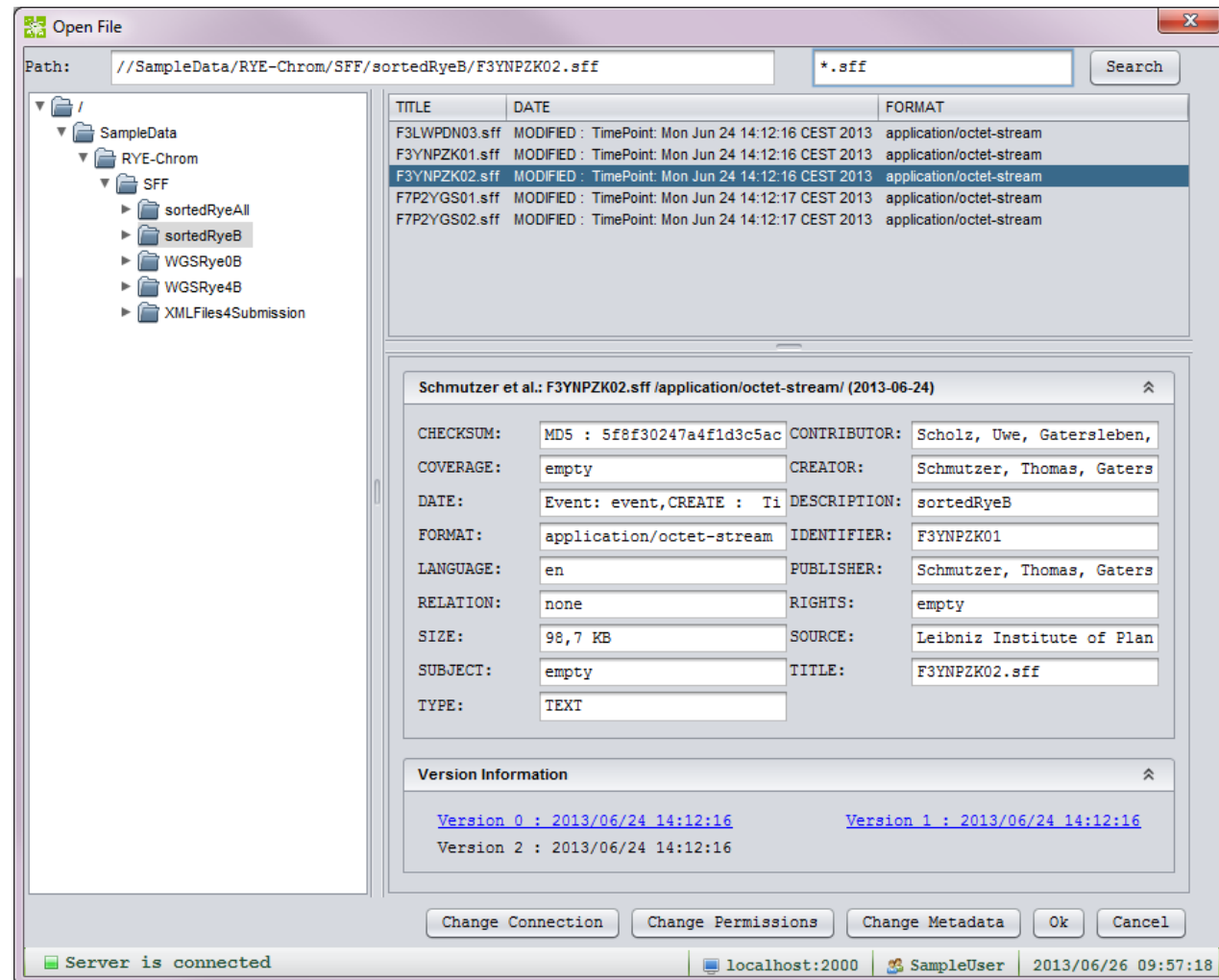


e!DAL – Architecture & Interoperability



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

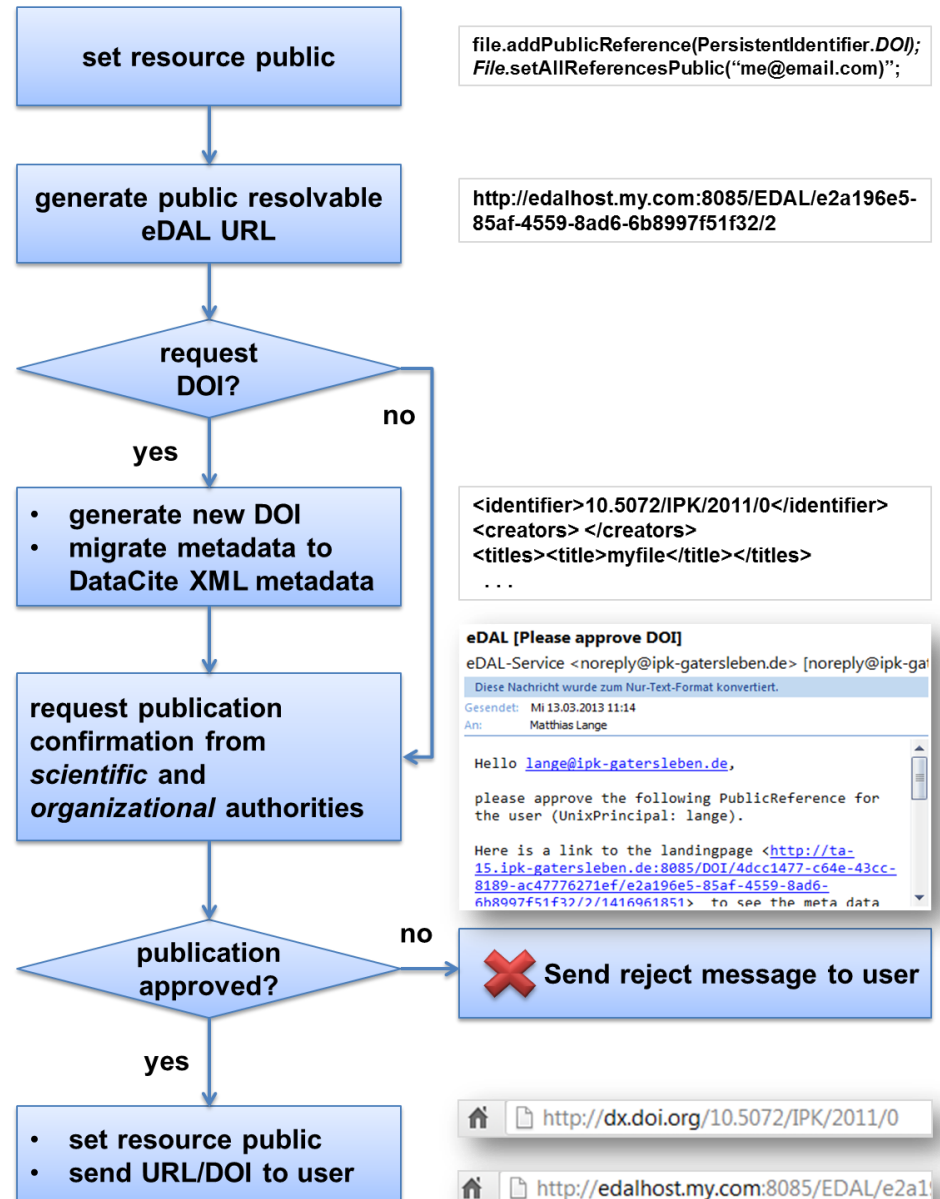


graphical user interface – e!DAL-FileChooser

¹ <http://dublincore.org/>

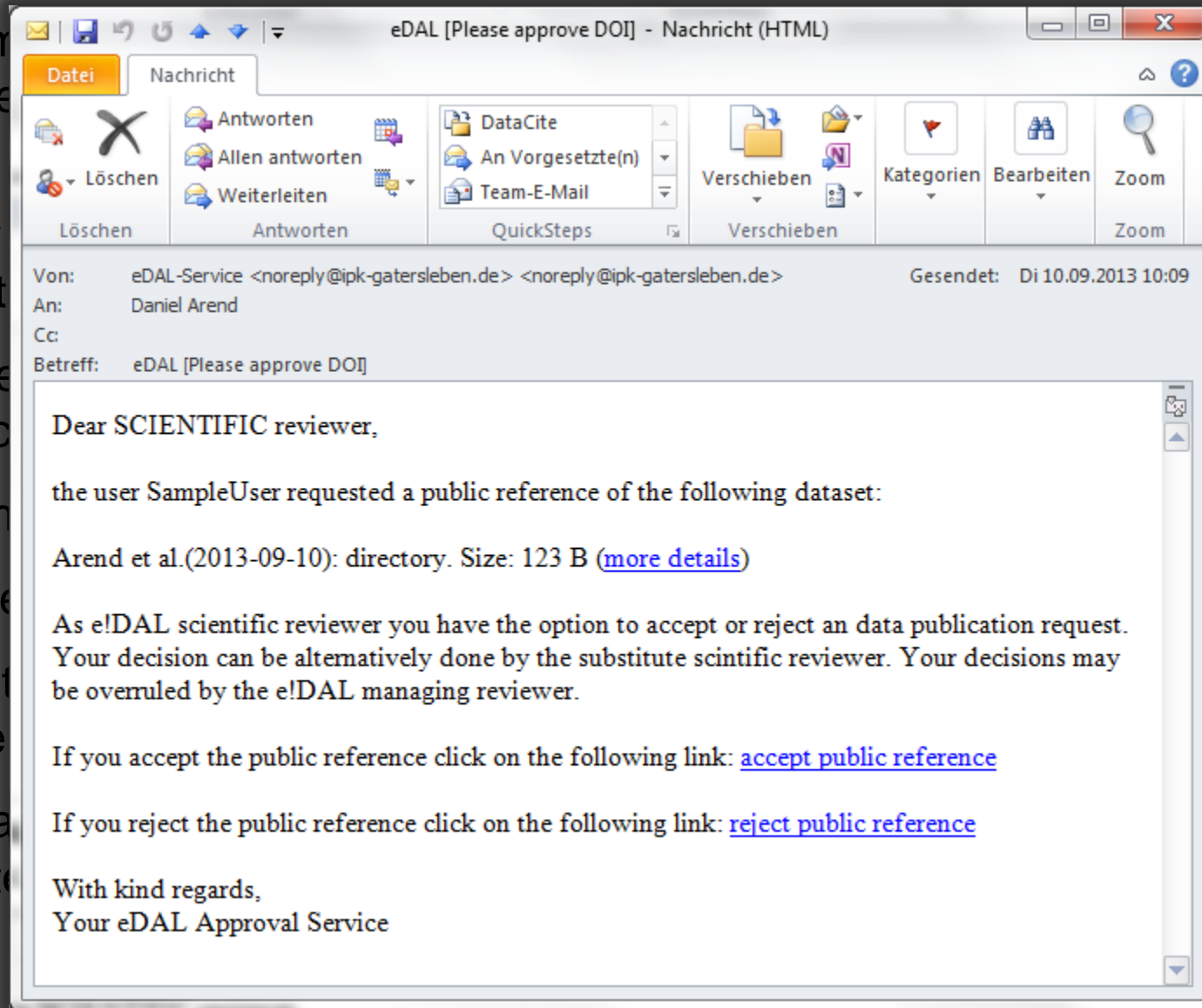
e!DAL – Persistent Identifiers

- 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

- supplement published (tables, ...)
- primary publication
- published (DB access)
- problem
- interface
- registration identifier
- approval request



istentIdentifier.DOI);
me@email.com)";

5/EDAL/e2a196e5-
32/2

1/0</identifier>

</titles>

ben.de> [noreply@ipk-ga-
vertiert.

blicReference for

ge <<http://ta-1dccc1477-c64e-43cc-4559-8ad6->
see the meta data

ssage to user

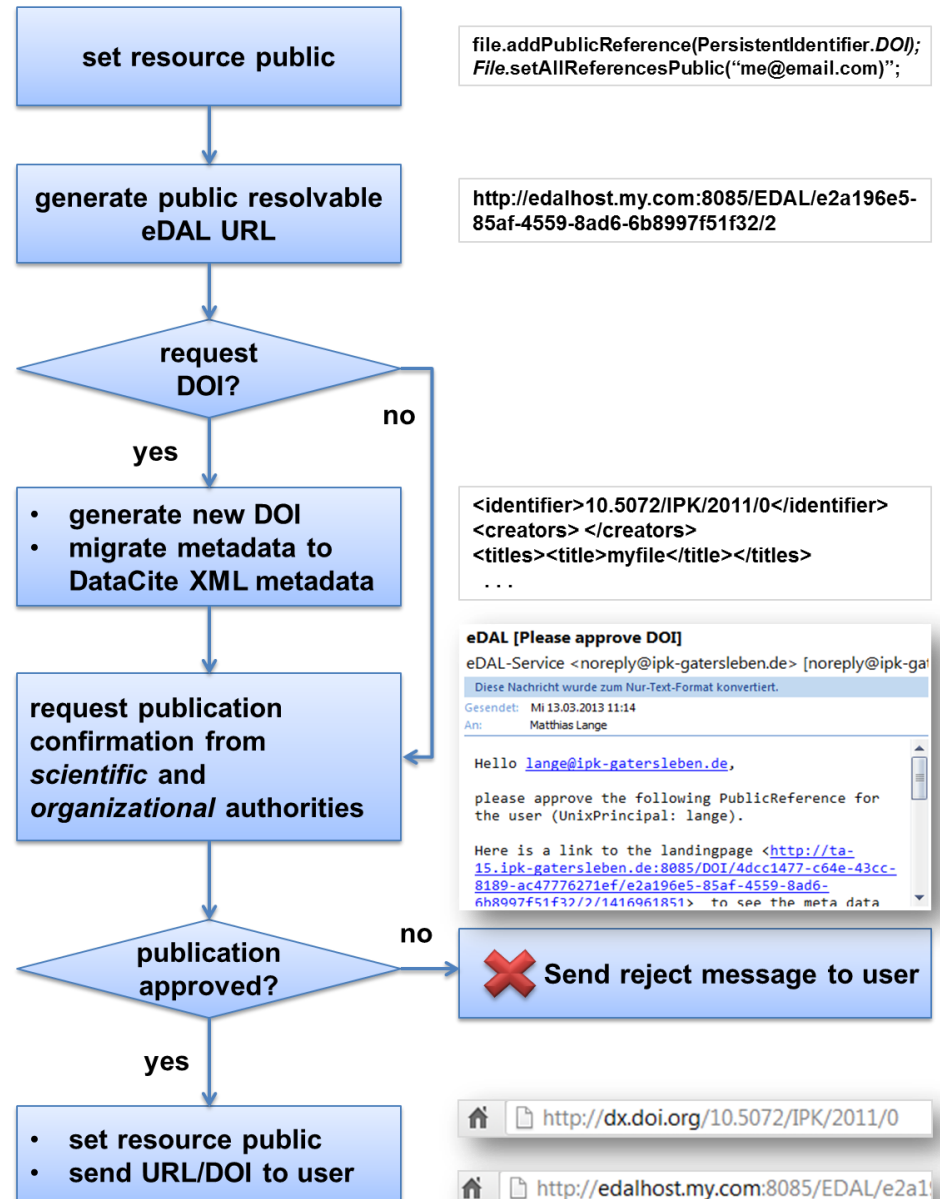
5072/IPK/2011/0

• send URL/DOI to user

<http://edalhost.my.com:8085/EDAL/e2a196e5-32/2>

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
- (tak
- pri
- pub
- pub
- (DE
- pro
- inte
- reg
- ide
- app
- req

Metadata - poster.pdf x

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

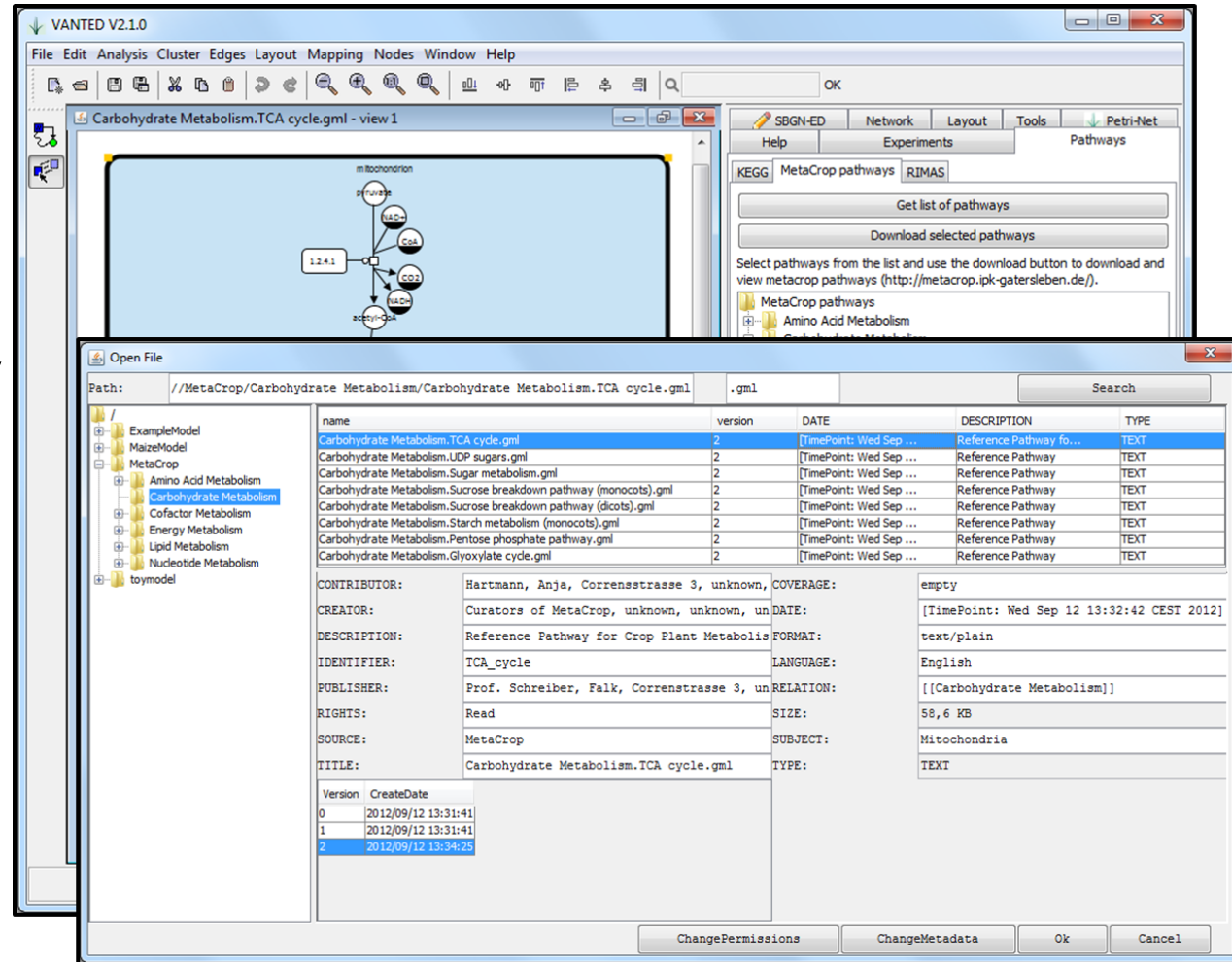
Meta Data

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	

2013-SEPTE

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

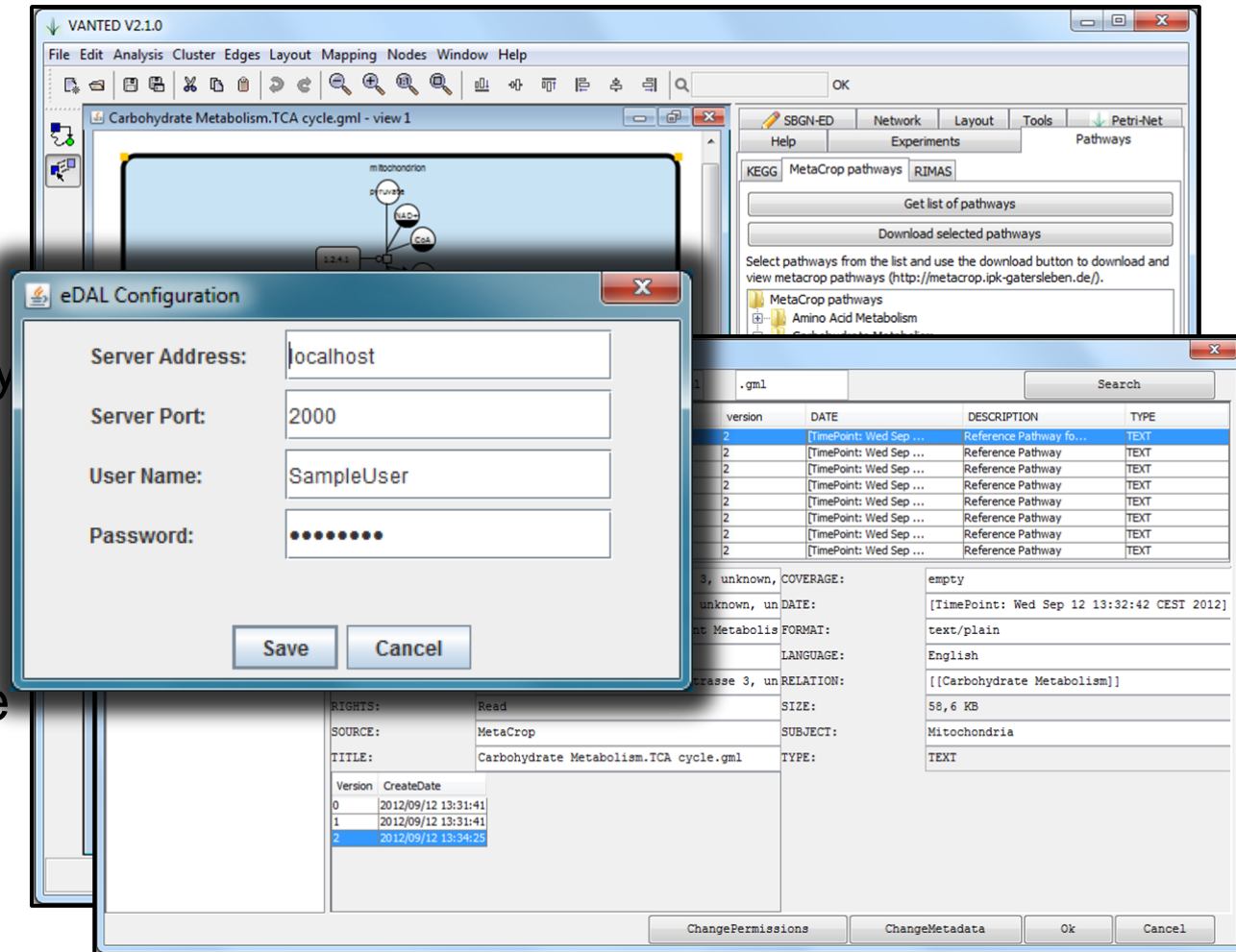


Embed into tools: e.g. VANTED pathway editor¹

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

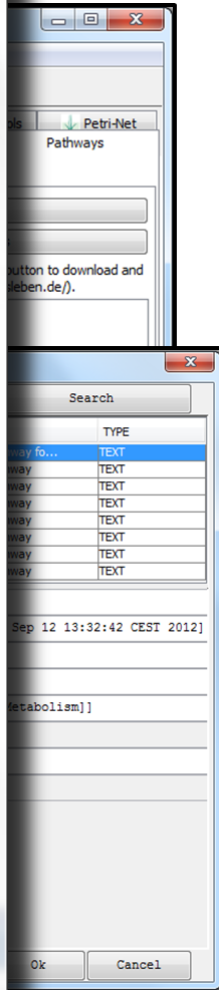
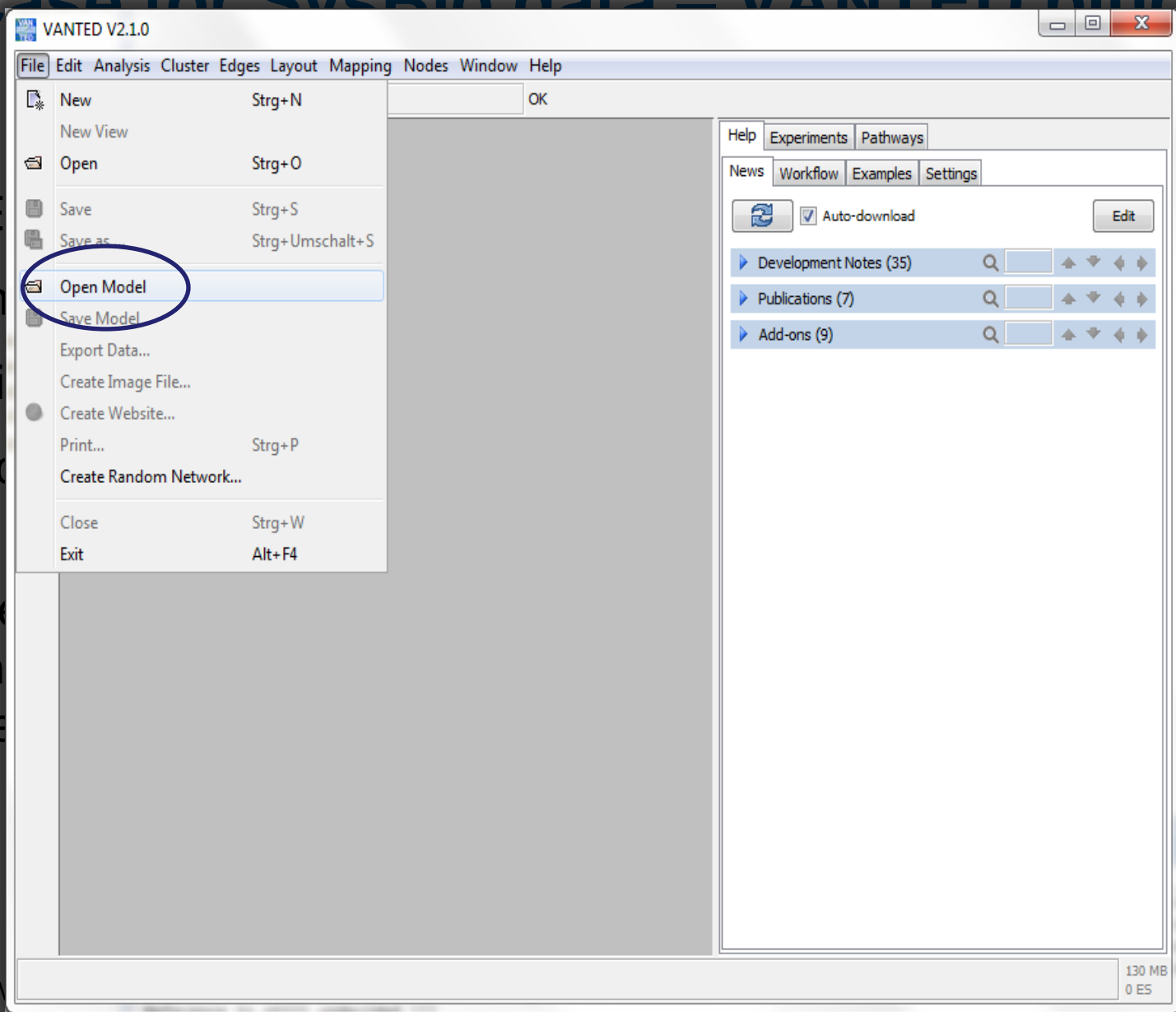


Embed into tools: e.g. VANTED pathway editor¹

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

Use case for SysBio data – VANTED plugin

- embed VANTED
- use remote
- central
- secured access
- manage version with me



r1

¹ (Rohn et al. BM

Use case for SysBio data – VANTED plugin

VANTED V2.1.0

Open File

Path: //MetaCrop/Carbohydrate Metabolism/Carbohydrate Metabolism.TCA cycle.gml .gml Search

ExampleModel
MaizeModel
MetaCrop
 Amino Acid Metabolism
 Carbohydrate Metabolism
 Cofactor Metabolism
 Energy Metabolism
 Lipid Metabolism
 Nucleotide Metabolism
 toymodel

name	version	DATE	DESCRIPTION	TYPE
Carbohydrate Metabolism.TCA cycle.gml	2	[TimePoint: Wed Sep ...]	Reference Pathway fo...	TEXT
Carbohydrate Metabolism.UDP sugars.gml	2	[TimePoint: Wed Sep ...]	Reference Pathway	TEXT
Carbohydrate Metabolism.Sugar metabolism.gml	2	[TimePoint: Wed Sep ...]	Reference Pathway	TEXT
Carbohydrate Metabolism.Sucrose breakdown pathway (monocots).gml	2	[TimePoint: Wed Sep ...]	Reference Pathway	TEXT
Carbohydrate Metabolism.Sucrose breakdown pathway (dicots).gml	2	[TimePoint: Wed Sep ...]	Reference Pathway	TEXT
Carbohydrate Metabolism.Starch metabolism (monocots).gml	2	[TimePoint: Wed Sep ...]	Reference Pathway	TEXT
Carbohydrate Metabolism.Pentose phosphate pathway.gml	2	[TimePoint: Wed Sep ...]	Reference Pathway	TEXT
Carbohydrate Metabolism.Glyoxylate cycle.gml	2	[TimePoint: Wed Sep ...]	Reference Pathway	TEXT

CONTRIBUTOR: Hartmann, Anja, Corrensstrasse 3, unknown, COVERAGE: empty

CREATOR: Curators of MetaCrop, unknown, unknown, un DATE: [TimePoint: Wed Sep 12 13:32:42 CEST 2012]

DESCRIPTION: Reference Pathway for Crop Plant Metabolis FORMAT: text/plain

IDENTIFIER: TCA_cycle LANGUAGE: English

PUBLISHER: Prof. Schreiber, Falk, Correnstrasse 3, un RELATION: [[Carbohydrate Metabolism]]

RIGHTS: Read SIZE: 58,6 KB

SOURCE: MetaCrop SUBJECT: Mitochondria

TITLE: Carbohydrate Metabolism.TCA cycle.gml TYPE: TEXT

Version	CreateDate
0	2012/09/12 13:31:41
1	2012/09/12 13:31:41
2	2012/09/12 13:34:25

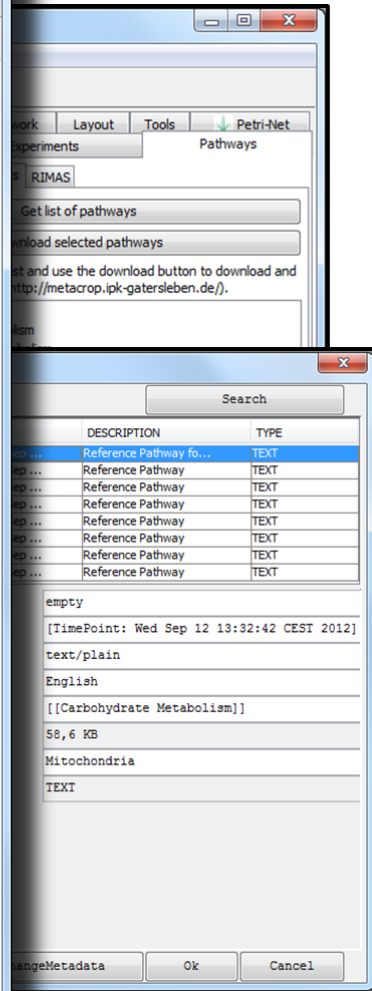
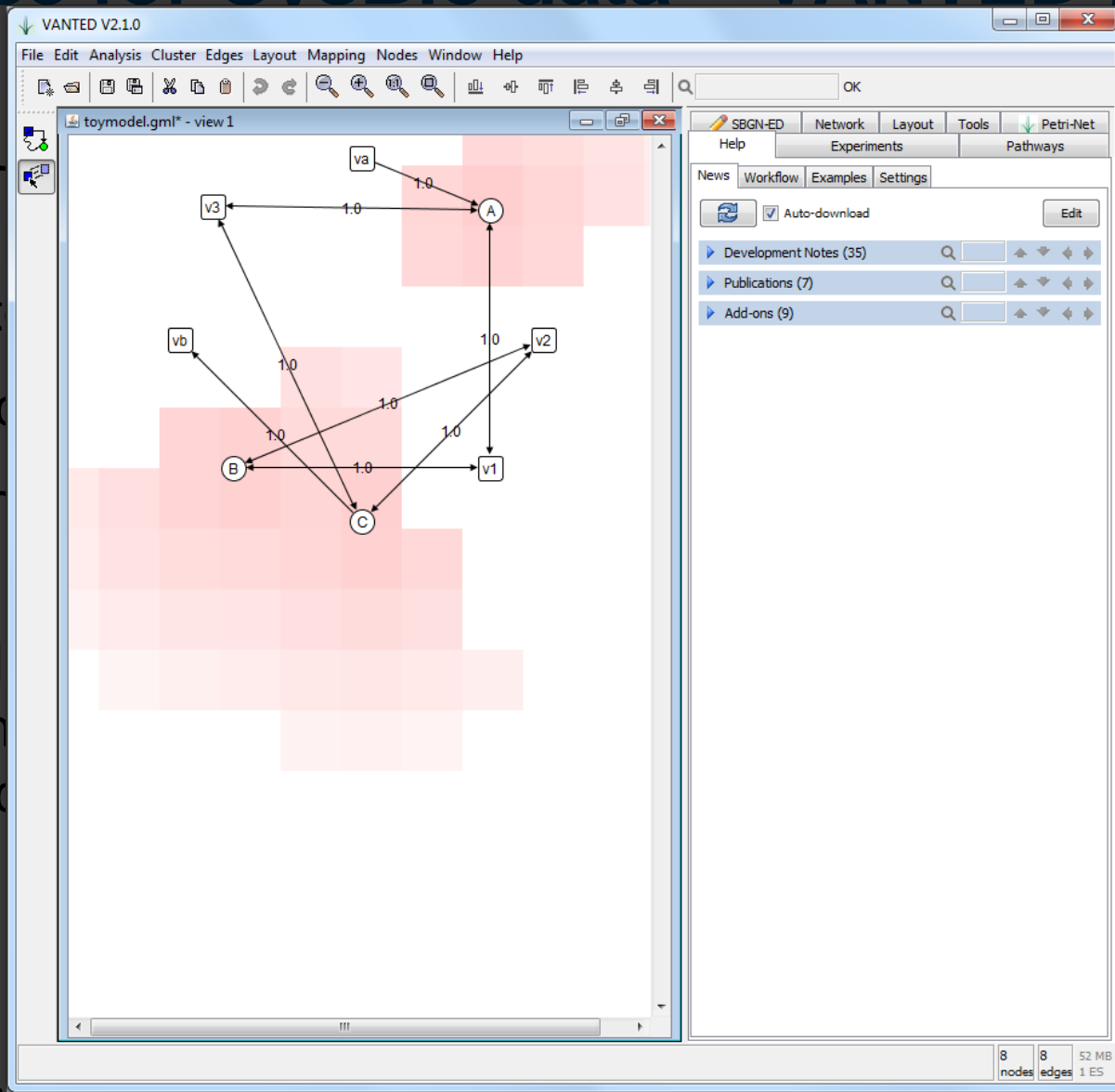
ChangePermissions ChangeMetadata Ok Cancel

Embedded into tools: e.g. VANTED pathway editor

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

Use case for SysBio data – VANTED plugin

- embed GUI into VANTED
- use remote SysBio data
- centralized data storage
- secured metadata access
- manage data version and metadata with meta-data



ay editor¹

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!

The screenshot shows a Firefox browser window displaying the eDAL-MetaData-API website. The address bar shows the URL `projects.ipk-gatersleben.de/eDAL-Project/`. The page title is "eDAL - MetaData - API". The header includes the IPK Gatersleben logo and the text "Last Published: 2013-09-09 | Version: 2.0". The left sidebar contains a navigation menu with sections: INTRODUCTION (Welcome, Download), SCIENTIFIC PUBLICATIONS (Papers, Talks, Poster), DOCUMENTATION (Samples, Core API, Configuration, Authentication, MetaData, Data Publication, Security, JavaDoc, Client-server API, Server, Client), WebDAV interface (GUI: FileChooser), MAVEN DOCUMENTATION (Project Information, Project Team, About, Project Summary), and a Google Custom Search bar. The main content area is titled "eDAL-MetaData-API - store cite and share primary data". It describes eDAL as an electronic Data Archive Library and provides a list of features: file system and database independent pure JAVA persistence store of data entities, enrich meta data to every entity (support the Dublin Core standard) and manage different versions, save and read data as JAVA streams, search across files and meta data using Apache Lucene, and integrate user authentication and authorization into existing infrastructure (Kerberos, LDAP, Unix native, Windows native) by supporting JAAS technology. It also mentions two modes of use: Embedded and Client-Server. A "Demo-Application" section describes a demo/install wizard. At the bottom, there is a Java Webstart logo and a small inset image of the eDAL installer window.

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

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