

Evaluation of the semantic research data management system CaosDB in glaciology

Alexander Schlemmer^{1,*}, Henrik tom Wörden^{1,*}, Johannes Freitag²,
Timm Fitschen^{1,*}, Johanna Kerch², York Schlomann², Daniel Hornung^{1,*},
Ilka Weikusat², Ulrich Parlitz¹, Frank Wilhelms² and Stefan Luther¹

MPI for Dynamics and Self-Organization, Göttingen (1)

AWI for Polar and Marine Research, Bremerhaven (2)

IndiScale GmbH i.G. (*)

2019-06-05



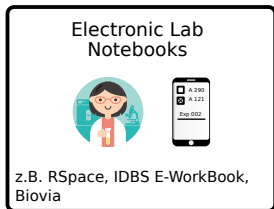
CaosDB

Important Requirements for Research Data Management

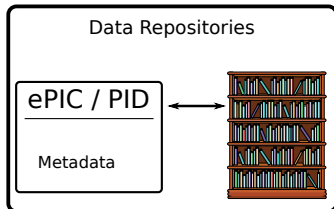
- Simple but expressive search functionality
- Ability to store every data format, at any file size
- Possibility to store, link and retrieve raw data, processed data, analysis results and documentation
- Support all kinds of data analysis software, from simple scripts to high-level software
- Minimally invasive workflow
- Scientific environments change often: Need for flexible data model

Research Data Management during Data Analysis

Data Acquisition



Data Publication



Data Analysis



Scientist picture: 201705 Scientist bench F.svg from commons.wikimedia.org/wiki/Category:Life_science_images_from_DBCLS, CC-BY 4.0

Bookshelf: <https://openclipart.org/detail/289378/bookshelf-with-blue-books>

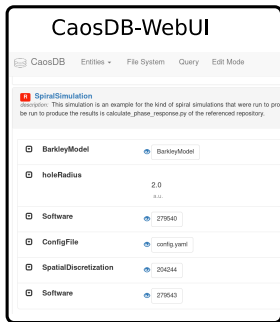
CaosDB Overview

Data Acquisition:

Use your desired workflow!

Data Files

FileSystem



CaosDB-WebUI

CaosDB Entities File System Query Edit Mode

SpiralSimulation
description: This simulation is an example for the kind of spiral simulations that were run to produce the results in calculate_phase_response.py of the referenced repository.

BarkleyModel	BarkleyModel
holeRadius	2.0
Software	279640
ConfigFile	config.yaml
SpatialDiscretization	204244
Software	279643



CaosDB-Python-Interface

```
salexan@salexan-x1 ~ % ipython
In [1]: import caosdb as db
In [2]: results = db.execute_query("FIND Simulation")
```

RESTful
XML-Protocol

CaosDB-Crawler

Automatic
file indexing



OpenSource! <https://gitlab.gwdg.de/bmp-caosdb>

State and Future of the Project

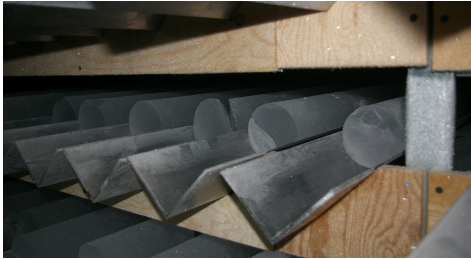
- CaosDB is a scientific project at the Research Group Biomedical Physics (MPI Göttingen)
- It is developed since ≈ 8 years and running stable since ≈ 2016
- CaosDB has been released as OpenSource software in 2018
- Service Company: IndiScale GmbH i.G.
- CaosDB is currently tested in other workgroups in and outside of Göttingen
→ Very positive feedback

CaosDB in Glaciology at the AWI in Bremerhaven

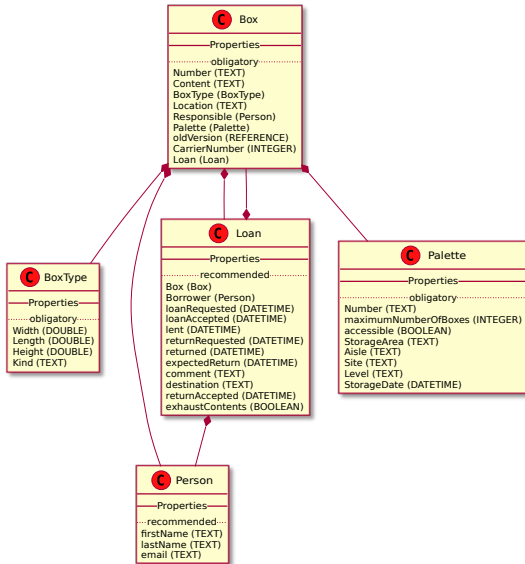
Aims of the Collaborative Project

- System for efficient management of ice core boxes
 - Future extension to sample management
- Testing of the flexible data model of CaosDB in the context of polar research
- Evaluation of CaosDB in a different scientific environment
- Feedback from new users

Ice Cores





UML diagram of the data model


















Querying the Data Model

- FIND Record Box with Number = "0258"
- SELECT Number, Content from Box with Content like "ice"
- FIND Record Box with palette with number = "12"
- FIND Fabric
 - which references Section
 - which references Bag
 - which references PPStrip
 - which references Core with name = "EGRIP"

 **Box**  Backref

[Update XML](#) [Borrow Box](#)










 Number	1227	
 Content	Eiskern-Archiv B33-013.EK01 B33-014.EK01 B33-015.EK01 B33-016.EK01 B33-017.EK01 B33-018.EK01 B33-019.EK01 B33-020.EK01 Comment: Archiv	
 Location	Fischereihafen-103-18-04	
 Responsible	 <input type="text" value="Thomas Gieseler"/>	
 BoxType	 <input type="text" value="EKK01"/>	
 Palette	 <input type="text" value="288"/>	

Comments [add new comment](#)

Implementation of the Loan Workflow

R Box Print

[Update XML](#) [Borrow Box](#)

☐ Number	1227	
☐ Content	Eiskern-Archiv B33-013.EK01 B33-014.EK01 B33-015.EK01 B33-016.EK01 B33-017.EK01 B33-018.EK01 B33-019.EK01 B33-020.EK01 Comment: Archiv	
☐ Location	Fischereihafen-103-18-04	
☐ Responsible		
☐ BoxType	 EKK01	
☐ Palette	 288	

Implementation of the Loan Workflow

R **Box** ← Backref

[Update XML](#) [Borrow Box](#)

Borrow Box

Please enter your full name and your email address. If you are doing this for the first time, a new user record will be created.

First Name:	Last Name:	Email:
<input type="text" value="Alexander"/>	<input type="text" value="Schlemmer"/>	<input type="text" value="alexander.schlemmer@ds."/>
Expected Return Date:	Comment:	Destination:
<input type="text" value="11 / 29 / 2018"/>	<input type="text" value="Will be used for ..."/>	<input type="text" value="Department for ..."/>

Are you going to completely exhaust the contents of this box?

<input type="checkbox"/> Number	1227	
<input type="checkbox"/> Content	Eiskern-Archiv B33-013.EK01 B33-014.EK01 B33-015.EK01 B33-016.EK01 B33-017.EK01 B33-018.EK01 B33-019.EK01 B33-020.EK01	

Implementation of the Loan Workflow

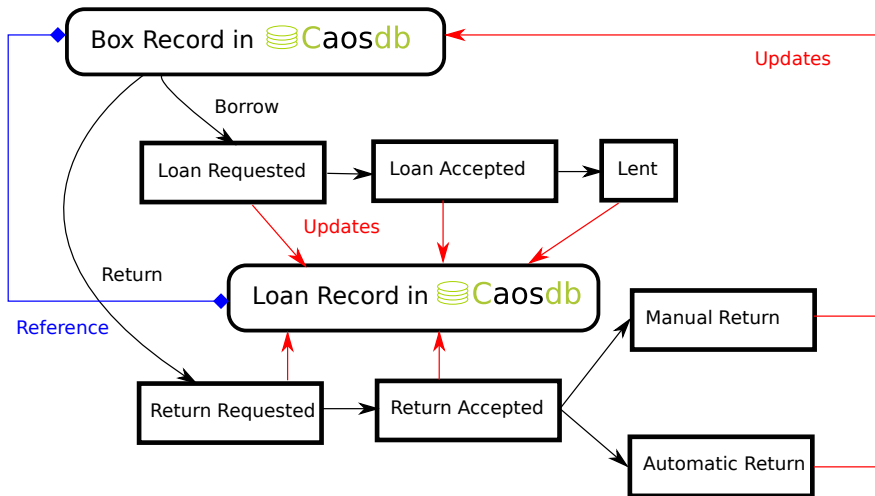
R Box ← Backref

[Update XML](#) [Accept Borrow Request](#)

<input type="checkbox"/> Number	1227	
<input type="checkbox"/> Content	Eiskern-Archiv B33-013.EK01 B33-014.EK01 B33-015.EK01 B33-016.EK01 B33-017.EK01 B33-018.EK01 B33-019.EK01 B33-020.EK01 Comment: Archiv	
<input type="checkbox"/> Location	Fischereihafen-103-18-04	
<input type="checkbox"/> Responsible	<input type="text" value=""/>	
<input type="checkbox"/> BoxType	<input type="text" value="EKK01"/>	
<input type="checkbox"/> Palette	<input type="text" value="288"/>	
<input type="checkbox"/> Loan	<input type="text" value="Borrowed by Alexander Schlemmer (loan requested)"/>	

Comments [add new comment](#)

Full Ice Core Box Loan Workflow



Automatic Return

CaosDB Query Upload Storage File New Box New Palette New User Template Help salexan

Ausgewählte Datei: Beispiel_einlagerung_test.xlsx Submit

Das Hochladen und Verarbeiten der Excel Datei kann einige Minuten dauern.

Number	1227	
Content	Eiskern-Archiv B33-013.EK01 B33-014.EK01 B33-015.EK01 B33-016.EK01 B33-017.EK01 B33-018.EK01 B33-019.EK01 B33-020.EK01 Comment: Archiv	
Location	Fischereihafen-103-18-04	
Responsible	<input type="text"/>	
BoxType	EKK01	
Palette	288	
Loan	Borrowed by Alexander Schlemmer (return accepted)	

Comments [add new comment](#)

CaosDB Query Upload Storage File New Box New Palette New User Template Help salexan

Ausgabe des Scripts: Code: 0

Standard Messages:

16 Boxes successfully updated.



FIND Record Box with Number = "0258"



Shortcuts

Show the box with number

Generate table with boxes that have a content with

Find all boxes with name that contains

Generate table with all boxes borrowed by (last name)

Generate Table with state of all lent boxes

Show loan state for box with number

Show location of box in Fischereihafen Storage

Find all fabrics from core (name)

Show open loan requests

Fulltext search

Find all boxes stored on palette with number



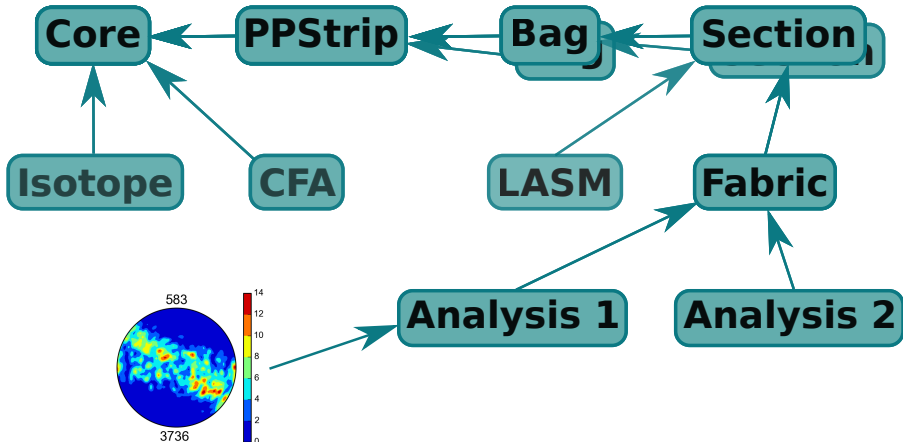


Table of selected fields

[Download TSV](#)

Borrower	expectedReturn	lent	returned
Alexander Schlemmer	2018-11-28	2018-11-19T22:00:03+0100	
Alexander Schlemmer	2018-11-21	2018-11-19T21:52:06+0100	
Alexander Schlemmer	2018-11-27	2018-11-19T21:14:58+0100	
Alexander Schlemmer	2018-11-28		
Alexander Schlemmer	2018-11-15	2018-11-19T22:03:30+0100	
Henrik tom Wörden	2018-11-21	2018-11-20T09:27:04+0100	
Alexander Schlemmer	2018-11-22	2018-11-20T10:27:00+0100	
Alexander Schlemmer	2018-11-22	2018-11-20T10:42:07+0100	2018-11-20T11:43:51.824019+0100
Alexander Schlemmer	2018-11-29		2018-11-20T12:24:59.828999+0100


From Box Management to Sample Management



Data Analysis in CaosDB

CaosDB Query Upload Storage File New Box New Palette New User Template Help salexan

cAxesAnalysis [Backtest](#)
[Update XML](#) [Analyse Data](#)

- Fabric** 5870
- grainsFile** grains.bt
- boundariesFile** boundaries.bt
- BoundaryImage**
 - Image** [Backtest](#)
 - path: /var/ta/EGRIP583_5_20/cAxes/boundary_rgb.png
 - size: 489865
 - checksum: 583A0485B8CFEBA951C8C619A0FB324
 - 

CaosDB Query Upload Storage File New Box New Palette New User Template Help salexan

cAxesAnalysis [Backtest](#)
[Update XML](#) [Analyse Data](#)

Submit

Data analysis can take some minutes.

- Fabric** 5870
- grainsFile** grains.bt
- boundariesFile** boundaries.bt
- BoundaryImage**
 - Image** [Backtest](#)
 - path: /var/ta/EGRIP583_5_20/cAxes/boundary_rgb.png
 - size: 489865
 - checksum: 583A0485B8CFEBA951C8C619A0FB324

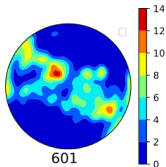
Data Analysis in CaosDB

Ausgabe des Scripts:

Errors:

No handles with labels found to put in legend.

Standard Messages:



5873

stereoAnalysis ← Backref Update XML

- stereoFile** stereo.txt
- AnalysisResult**
 - Image AnalysisResult** ← Backref
 - path: raw/ita/583-stereo_py_con0.png
 - size: 40217
 - checksum: 01C8F63CDE2328EAAE649BBBB9931E0
 - Thumbnail heatmap visualization with a color scale from 0 to 14. The number 583 is above the heatmap and 3736 is below it.
 - Comments add new comment



- OpenSource Project: <https://gitlab.gwdg.de/bmp-caosdb>
- Paper on ArXiv: <https://arxiv.org/abs/1801.07653>, under review in Data¹

Thank You!

¹Addendum: The paper is now published, <https://doi.org/10.3390/data4020083>