



Caterina Barillari, Juan Fuentes, Henry Lütcke & Carlo Minotti

Improving interoperability of Electronic Lab Notebooks and Data Repositories by enhancing API integration



Key concepts

Metadata

• *data about data*, i.e. the information about one or more aspects of the data.

Data repositories:

- Used for metadata organisation, documentation, sharing, security, storage.
- Examples in ETH domain: Envidat, Materials Cloud, ETH Research Collection, SciCat.

Logbooks and Electronic laboratory notebooks (ELNs):

- Logbooks: Used for jotting information about experiments, while they are running, and annotate unstructured information about them.
- ELNs: detailed experimental description.
- Examples in ETH domain: SciLog, openBIS, elog.



openBIS

Inventory management



Lab equipment Lab samples & materials O Q Global Search - 00 Collection: Mixers C Q Cichel Search + U Collection: Chemical admixtures W. 1144 + New Equipment Mixers More **W** Film ۰. ï + New Administrate More > E Lab Notebook > E Lab Notebook IC C 1-7 of 7) H Rows per page: 10 + COLUMNS FILTERS EXPORT · & Inventory 10 C 14 of 6 > >> Pows per page: 10 + COLUMNS FILTERS EXPOR · & invertory ~ Equipment Mixing Log Year of Code > Equipment Permid / Default Identifier Date of Code Name > Equipment Main AND + Filter Materials AND - Filter > Measurement Equipm CUPMENT_MODERS_6 EL10 20211115150104128- 10 L ~ Raw Materials LOG 135-21.005 Equipment Maters 2019 V. Preparation Equipment 2021-06-25 14:20:09 Sika I Additives ADM6 Skapard 705 L 20211115182205710 AMATERIALS/RAW, MATERIALS/RAM6 OTHER 900.0 I Mxer DEGUIPMENT_MIXERS_7 50 112 5L LOG 135-Equipment EQUIPMENT E Appregates 0029-04-20 21:00:00 +0000 > Lab Experts E Cement RASP > Materials C EQUIPMENT_MIXERS_1 evolution 20211115150104128-VICLUM 126 E Chemical admin LOG 135-29.006 Equipment Maters EQUIPMENT 2008 2017-09-27 07:00:00 +0000 > Methods E Suppler ADM8 Citre acid 20211115142205710 AMATERIALS/RAW, MATERIALS/RAM8 Relarder III Water EQUIPMENT_MOXEPS_2 VIOLUTION 20211115150104128-LOG 308-29.007 2016 Equipment Mixers EQUIPMENT 2019-12-15 15:48:33 5ika > 🖌 Utilities > Sold Sample a standard 20211115162205710 AMATERIALS/RAW_MATERIALS/ADMS Superplasticizer 2020-01-22 15:48:33 +0000 1080.0 O About > Methods +0000 LOG 135-21.003 Equipment Moters EQUIPMENT (Change of of - sxample) EQUIPMENT_MORERS_3 EININ R 20211115150104128- 75 L 1991 > Publications 8 18 2110.0 > F Utilities C EQUIPMENT_MODERS_4 ENCO RV 20211115150104128-LOG 135-21.004 1998 8 8 8 8 8 Equipment Maters O About ADM11 No.W.A. 130 20211115162205710- AMATER (0814-1), AMAURICE_BIOT/PROJECT [0814-2), More EQUIPMENT_MIXERS_5 00 W 150 104128- 75 L LOG 135- 2012 Box Name: Equipment Mixers EQUIPMENT Ber Bar or Propher 1 2 3 4 meas Lab procedures Samples' storage manager O Q Global Search · UI Collection: Shrinkage w. Film + New Shrinkage Protocol More ... -> E Lab Notebool Barcode reader IC (1 of 1)) Rows per page: 10 - COLUMNS FILTERS EXPORTS v & inventory > Fourpment time scale (of Permid / Default Shrinkage Barcode dimensionality SOP ID > Lab Experts Notes > Materials . Mathoda Operator AND + Filter **User rights** Measurement Protocol management III Creep MAURICE BIOT/PROJECT USHRINKAGE MEASUREMENT S/A Unterchand Periodol 2807.1 2021109133722556 Linear shrinkage fmf (range) fmssumg Smrkage METHODS ppsmdx 74 Linear shrinkage Imear shrinkage</t (Massurement 1), MAURICE_BIOT/PROJECT_1/SHRINKAGE_MEASUREMENT_gtm_admin 14:372 I Mass loss Shrinkag > Protocols > Publication > & Utilities **ETH** zürich O About Scientific IT Services 13.06.2023 60







ELN & data management

Data are always connected to experimental descriptions







- **Organize** the scientific data into datasets.
- Annotate the Datasets with administrative and flexible scientific metadata.
- Make the data **searchable/discoverable.**
- Provides the infrastructure for **publishing** the data, DOI generation.
- Can be used as frontend for **long term storage** (Archive) solutions of mass data (PB regime).
- Supports both open access and embargoed data.

							? Help	i	About 📮	 	■ Details	Datafiles	Re	lated Datasets	Attachme	ents	🗘 Lifecycle					
tasets /										Datasets Files	pyter Hub		\mathcal{I}									
Search 🔳 Clear	My Data All Public Data All Archivable Retrievable Work	In Progress System	Error Use	Error				1	3 "	Instruments	General Info	mation										
ID			_		lterns per pag	n: 1000 💌 1	1 – 1000 of 8138	1<	< 15- 1	lobs	lame	2022_12_05_5l	iL5_2									
ext Search	c 🖗 🕴	٢			•	*	<u></u>		o =	Policies	escription	-										
	2022_12_05_SL5_2	/wakonig_k	14 MB	2022-12-07 Wed 17:33	derived	Proposal ID	a-35487	0	e 🕹 F	Proposals	'ID	20.500.11935/bi	bab8f984-cc6d	-450a-9f11-bc2a1	775bedc						۵	
roup	documentation	a10/e17642	988 KB	2019-03-06 Wed 12:03	raw	0.11935/2018	p17642	0	re 🔳 F	Published Data	vpe	derived										
p17642 8130	photographs	a10/e17642	69 MB	2019-03-06 Wed 09:54	raw	0.11935/2018	p17642	0	re X S	Samples	reation Time	2022-12-07 17:	:33									
p17502 7	S08000-08999/S08163	999/508163	725 GB	2019-03-05 Tue 21:41	raw	0.11935/2018	p17642	۵	re 🌣 S	Settings	eywords											
a-35487 1	508000-08999/508162	999/508162	4 GB	2019-03-05 Tue 21:41	raw	0.11935/2018	p17642	۵	re → L	Logout										1	Edit	
tart Date - End Date 🗈	508000-08999/508160	999/508160	2 GB	2019-03-05 Tue 21:33	raw	0.11935/2018	p17642	0	retrievable											_		
+ Add Condition	508000-08999/508159	999/508159	92 MB	2019-03-05 Tue 21:33	raw	0.11935/2018	p17642	6	retrievable		Creator Infor	Fuen Augustin										
	S08000-08999/S08158	999/508158	49 GB	2019-03-05 Tue 21:26	raw	0.11935/2018	p17642	٥	retrievable		wner Group	วงศา ผนชูเวรเท ล-35687										
	S08000-08999/S08157	999/\$08157	4 GB	2019-03-05 Tue 21:26	raw	0.11935/2018	p17642	0	retrievable													
	S08000-08999/S08156	999/508156	57 GB	2019-03-05 Tue 21:22	raw	0.11935/2018	p17642	0	retrievable		File Informat	on	~									
	508000-08999/S08155	999/508155	74 GB	2019-03-05 Tue 21:02	raw	0.11935/2018	p17642	۵	retrievable		ource Folder	/das/home/wako	onig_k									
	508000-08999/508154	999/508154	4 GB	2019-03-05 Tue 21:01	raw	0.11935/2018	p17642	0	retrievable		ize	14 MB										
Creators*					_			these da observe during u	ta sets we ob that the satu nsteady state	bserve that the uration of the ga ie.	pore scale dynamics evolve as plateaus out before the di	as the macroscopic flow trans erential pressure across the	nsitions from unst e core. This sugge	eady state to steady sta sts that gas phase is mo	ite. We ire mobile	S) 0.5	across sam	al	MMMMMM	M - 0.5	e (10'	
Catherine Spurin C Tom Bultreys Maja R	ücker 🕲 Gaetano Garfi 🕲 Christian M. Schlepütz 🕲 Vladimir Novak (Steffen Berg S	Martin J. Bl	unt 🕲			Drop	Pub	lication det	tails						- 4.0 uratio	alt			- 0.4	sssure	
None'								DOI Resource	DOI https://doi.org/10.16907/46a4d882-4dec-4097-8289-816331a4aa36 Resource Type derived							- c.o sat	0,00	×××*****	· · ·	0.3	al pre	
PSI Resource Type *					_			Related Publica	tions Dis	. Spurin, T. Bultr istinct Pore-Scal ttps://doi.org/10.	eys, M. Rücker, G. Garfi, C. M e Dynamics During Transient 1029/2020WR028287	Schlepütz, V. Novak, S. Berg, and Equilibrium Subsurface I	g, M. J. Blunt, and S Multiphase Flow.	5. Krevor. Real-Time Ima Water Resour. Res. 56, 4	ging Reveals 33 (2020).	ලී 0.2	percolation		Кеу	- 0.2	erenti	
derived Description *					Ť			Dat	acote							0.1	across FoV		 Pressure Saturation 	n 0.1	Diff	
Insp postenet data conceton contrast the obtained by X-ray tromographic microscopy of a carbonate rock sample 5 mm in diameter and 20 mm in length. Both brie and nitrogen gas are injected into the sample at total Blow the marked 0.1 m/lm/im (bit is marked 0.1 km/lm) (bit is marked 0.1 km/lm) (bit Surces. The First three datasets contain the scanned volume reconstruction during unsteady-state dynamics, while last two datasets contain the same scanned volume during steady state dynamics.								Dut	This pu diamet	ublished data co eter and 20 mm	ollection contains five datase in length. Both brine and nite	is obtained by X-ray tomogra	nple 5 mm in n (the brine	0	1 2	3 4	5 6	7 8				
					111	0.7 - u	insteady-state	Descrip	tion made in The first the same	up 85% of this t rst three datase ime scanned vol	otal flow rate). Data were co ts contain the scanned volur lume during steady state dyn	lected and processed at the e reconstruction during unst amics.	e TOMCAT beamlin steady-state dynai	e X02DA of the Swiss Li mics, while last two dat	ght Source. asets contain	0	i ime	2 manual di la constanti di constanti di constanti di constanti di constanti di la constanti d	on (nours)			
Memory in the related publication to these data sets, we explore the flow dynamics for two-phase flow in a porcos medium (a bioclastic carbonate rock). We use state-of-the-art synchrotrom X-ray tomography to capture the fluid dynamics within the porce space, with a scan time of 1 second and a temporal resolution (car respection rated of 2. Thereod sample was included as a state of who here (0) was recondent with the porce as ware included as a state of the second and a temporal resolution (car respection rated of 2. Thereod sample was included as a state of the second as the second as a temporal resolution (car respective) and the fluid memory and the second as a temporal resolution (car respective) and the fluid memory and the second as the second as a temporal resolution (car respective) and the second as a temporal resolution (car respective) and the second as a temporal resolution (car respective) as a state of the second as a temporal resolution (car respective) as a state of the second as a temporal resolution (car respective) as a state of the second as a temporal resolution (car respective) as a state of the second as a temporal resolution (car respective) as a state of the second as a temporal resolution (car respective) as a state of the second as a temporal resolution (car respective) as a state of the second as a temporal resolution (car respective) as a state of the second as a temporal resolution (car respective) as a state of the second as a temporal resolution (car respective) as a state of temp						0.6 (S ⁰ 0.5	vercolation	20.50	0.11935/64	4af1e80-c539	4a90-a051-b7db5e6e71	d						Sat				
stanzation oscillates about a constant mean value, this is referred to a stady state. There are 5 data sets 3 of which capture the utspace lyster, and the capture the utspace lyster, and the capture the utspace lyster, and the capture the utspace lyster between						0.4			0.11935/e1	151f4d6-198a	7e7-ac63-0b258ef36ed3						C. no.	All ale			Flow	
dawniad in a https://doi2.psi.ch/dataset5/das/work/p177/p17614/Data10/dosk1/						0.3 0 3 att	*****	20.50	0.11935/b9	9782901-be3t	-40fe-91d0-3e0a784337	4				1		in the		Í	1 mm	
related publications (C. Spurin, T. Bultreys, M. Rücker, G. Garfi, C. M. Schlepütz	z, V. Novak, S. Berg, M. J. Blunt, and S. Krevor, Real-Time Imaging Reveals Distinct Pore-	Scale Dynamics During T	ransient and Eq	uilbrium Subsurface		0.1 percol across	ation s FoV	20.50	0.11935/58	899a0eb-7e3b	-451f-b01e-17ddfc0d093	3				4	(9	6			
Multiphate Flow, Water Resour, Res. 56, 433 (2020). ht	tps://doi.org/10.1029/2020WR028287					0 1	2 3	Acti	ions							3	C I	R.			Flow	
Update Published Data Cancel						0 6 7 7	lime after i	To access	s the data as:	ssociated with t	his DOI click below and follow	the instructions					and the second	A REAL		I	J.1 mm	
								0.0	0	/												
						11-4-		This wor	k is licensed	under a Creativ	e Commons Attribution-Shar	Alike 4.0 International Licen	nse									
															/							







The interoperability/Missing API problem

- ETH Domain scientists use multiple tools during and after the experiment lifecycle:
 - Logbooks/ELNs to annotate and jot information during experiment run
 - Data catalogues to label raw data during and after the experiment
- No connectivity between these tools, requiring the scientists to jump from one to the other and needing to remember what was and goes where



Use cases

A scientist wants to:

- 1. export information from ELNs/logbooks to data catalogues once experiment is finished (e.g. openBIS to SciCat, SciLog to SciCat)
- 2. move/link raw data from openBIS to the SciCat data catalogue and vice versa
- 3. view, find or create metadata from data catalogue in logbook, e.g. see metadata describing experiment from SciCat in openBIS, SciLog
- 4. have better version control in SciCat, for derived data
- 5. automatically push Jupyter files changes to a remote git registry



Key challenges

- 1. Implement a publishing integration from openBIS to SciCat, relying on the existing data standard
- 2. implement a data transfer service, relying on a common data standard
- 3. implement widgets in openBIS/SciLog that allow finding, viewing and creating data from/into SciCat
- 4. improve the functionality that tracks metadata and derived data changes and stores it in SciCat, SciLog, openBIS
- 5. support git functionalities from openBIS, SciCat, SciLog