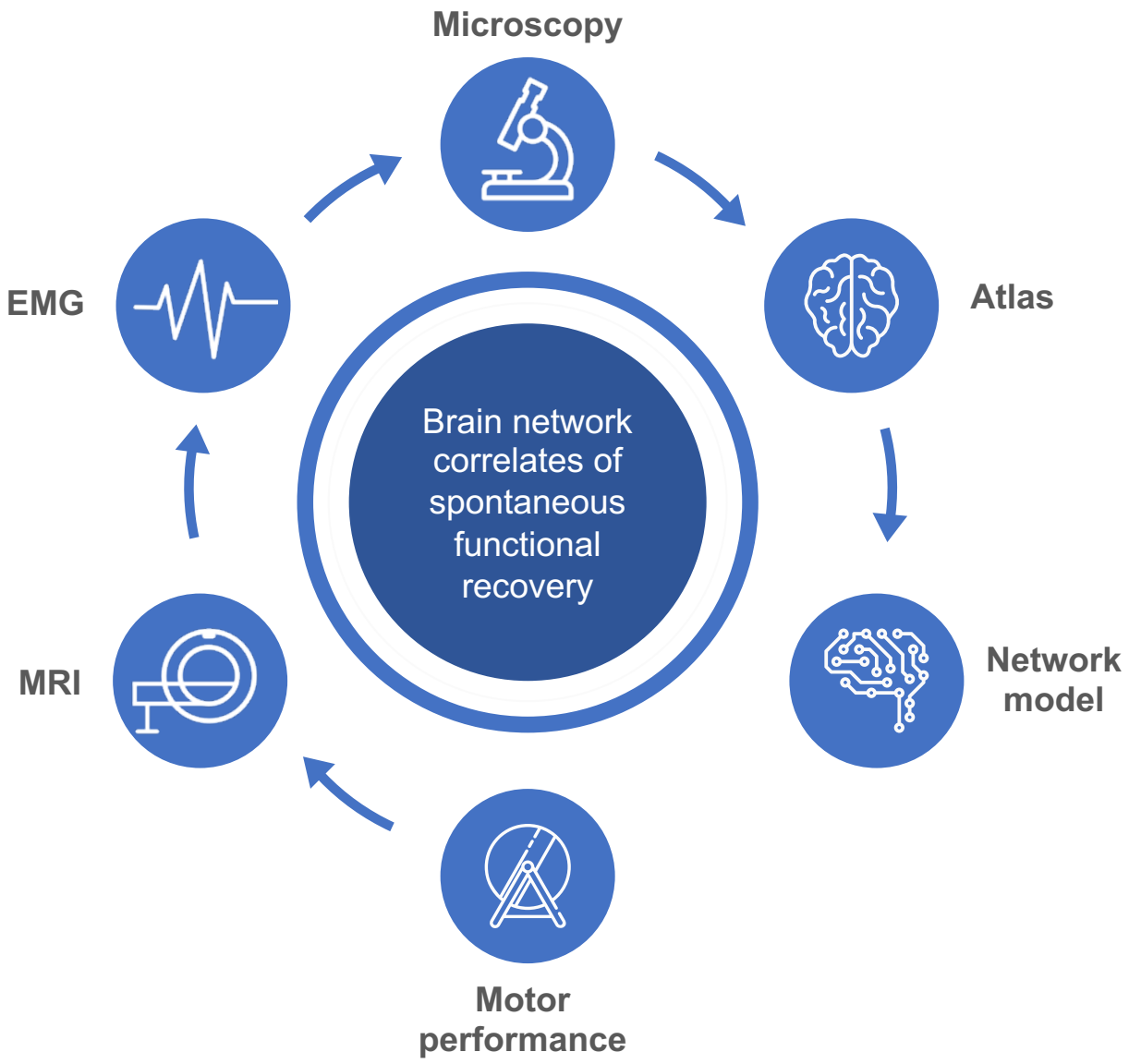
How to set up a FAIR workflow for multimodal rodent imaging data in a no-code environment of a translational stroke lab

Markus Aswendt & Aref Kalantari

University Hospital Cologne, Dept. of Neurology

September 20, 2023

The solution



The burden of working multimodal…

Techniques: 8

Locations: 3

Team: 10

File types: 7

Data size: 2+ TB/study

The problem

I have Converted this value from 82 to 97.2 (Equation is 82/100*120)

			hBeam SPEED				hBEAM DISTANCE				vBEAM score				NS(General + Focal)					MRI new and final 2D analysis	new and final analysis ctx/str	new and final analysis hipp/thal	size forebrain level (GFAP)	%infarct size STR level (GFAP)	size Hip/THL level (GFAP)	GFAP integrated density forebrain	GFAP integrated density striatum	integrated density hippocampus	TISSUE processed for...		Speed slope	Distance slope
			Baseline	D4	D8	D14	Baseline	D4	D8	D14	Baseline	D4	D8	D14	Baseline	D1	D4	D8	D14		STR	Hipp/THL	forebrain	striatum	hippocampus	forebrain	striatum	hippocampus				
			-1	4	8	14	-1	4	8	14	-1	4	8	14	-1	1	4	8	14													
SP2.3.1	Sham (n = 4)		8.15789	9.45455	7	5.83333	120	120	120	120	4	4	4	4	0	5	0	0	1										Western	no	-0.3489	0
SP2.8.2	Sham		7.33333	5.07576	7.85455	11.3542	120	117.5	120	117.667	4	4	4	4	0	0	0	0	1										Western	no	0.62432	-0.0154
SP3.1.1	Sham		11.7556	25.5	11.6667	14.5714	97.2	115	120	120	3	2	3	4	0	0	1	0	0										Immuno	Serum	-0.9684	0.46053
SP4.1.4	Sham		9.57372	11.6667	9.37121	2.98095	119.5	120	119.5	56	4	3	4	4	0	0	0	2	0										Dissection	test	-0.8841	-6.7303
average			9.20513	12.9242	8.97311	8.68497	114.175	118.125	119.875	103.417	3.75	3.25	3.75	4	0	1.25	0.25	0.5	0.5												-0.3943	-1.5713
stdev			1.9357	8.81984	2.04598	5.24222	11.3191	2.39357	0.25	31.6302	4	4	4	4	4	4	4	4	4												4	4
											0.5	0.95743	0.5	0	0	2.5	0.5	1	0.57735											0.73237	3.44641
											0.48999	0.93826	0.48999	#NUM!	#NUM!	2.44995	0.48999	0.97998	0.56579											0.71771	3.37742
SP2.4.2	MCAO	BAD	12.6364	0	0.16667	0	120	0	0.33333	0	4	1	4	4	0	17	22	13	3	43.4									Western	no	-0.0022	-0.0044
SP2.5.2	MCAO	BAD	11.4545	1.47143	1.36667	1.01398	120	13	13	17.3333	4	1	4	4	0	2	14	5	2		33.93%	39.49%	9.31253	45.0159	54.0648	0.07942	0.76311	0.40151	Immuno	no	-0.0468	0.45614
SP2.6.2	MCAO	BAD (Char	6.86275	0	3.01838	3.72029	120	0	39.5	51.6667	3	0	4	4	0	29	23	3	5		22.25%	50.75%	33.1109	44.2647	58.017	0.46863	0.78059	1.20223	Immuno	no	0.35189	4.91886
SP2.6.3	MCAO	BAD	10.6667	2.5	1.91667	4.67571	120	7	24	82.3333	4	0	3	4	0	18	21	16	5	47.9									Western	no	0.2367	7.70614
SP3.1.5	MCAO	BAD	12.6984	0	1.1875	3.17484	120	0	9.5	55.5	4	1	3	4	0	17	15	6	6		30.51%	44.77%	18.1594	41.0013	43.7311	1.02174	0.48325	0.63803	Immuno	Serum	0.31857	3.71711
SP3.2.1	MCAO	BAD	11.8056	0	0	0.33333	120	0	0	0.33333	3	1	1	1	0	12	17	6	6		31.64%	43.62%	50.8684	49.4683	61.9886	1.76003	0.945	0.74657	Immuno	Serum	0.03509	0.03509
SP3.3.5	MCAO	BAD	22.5	0	0	0	120	0	0	0	4	1	3	1	0	21	16	5	5		34.82%	59.18%		52.3965	74.4714		0.69072	0.98827	Immuno	no	0	0
SP3.4.5	MCAO	BAD	14.2857	3.77143	6.87879	4.91667	120	33	68	69.5	4	3	3	3	0	7	12	3	3		39.27%	28.62%	44.2355	41.2399	47.9497	0.60707	0.68286	0.82727	Immuno	Serum	0.07967	3.38158
SP4.1.1	MCAO	BAD	11.0294	0	0	0	120	0	0	0	4	3	3	2	0	19	6	4	10	31.9									Immuno	no	0	0
SP4.1.3	MCAO	BAD	8.61244	2.00694	4.61111	4.33333	120	36.5	38	26	4	1	2	3	0	8	6	1	5	42.7									RNA	Serum	0.21062	-1.125
SP4.2.4	MCAO	BAD	6.32143	0.16667	0.77949	0.97571	114	0.5	26.5	51.5	4	3	2	3	0	17	21	9	8	50.1									Immuno	test	0.0771	5.02632
SP4.3.2	MCAO	BAD	5.92941	3	0.61875	1	120	12	5.5	4	4	2	3	3	0	22	19	9	4	57.7											-0.1792	-0.7566
SP4.3.3	MCAO	BAD	14.026	0	0.5	3.73571	120	0	1.5	43.5	4	3	3	4	0	11	16	13	9	46.4									RNA	Serum	0.38665	4.55921
SP4.3.4	MCAO	BAD	12	3.83333	3.4375	2.83333	120	18.5	27.5	14	4	4	3	3	0	9	8	2	7	36.7									RNA	Serum	-0.1001	-0.5921
SP4.3.5	MCAO	BAD	27	0	2.25	3.41626	120	0	4.5	64.5	4	4	3	4	0	20	3	8	6	43.7									Immuno	test	0.33	6.73026
SP4.4.3	MCAO	BAD	17.5	0	0	0	120	0	0	0	4	1	1	2	0	17	17	6	5	45.2									Immuno	no	0	0
SP4.4.4	MCAO	BAD	16.0714	0	1.1	1.95714	120	0	5.5	17.5	4	1	0	2	0	18	17	11	12	59.3											0.19154	4.76974
SP4.5.3	MCAO	BAD	16.6667	0	1.5	1.73481	120	0	3.5	45	4	1	4	4	0	33	17	14	10	26.3											0.16288	4.69079
SP4.6.1	MCAO	BAD	7.72727	0	0	0	120	0	0	0	4	1	4	3	0	16	18	6	8	46.8											0	0
SP4.6.3	MCAO	BAD	11.4545	3.56034	3.26812	3.80263	120	66.5	56.5	61	4	3	4	4	0	14	14	7	10	35.9											0.02935	-0.4474
SP4.6.4	MCAO	BAD	8.15789	2.78357	2.77091	2.23333	120	114	90	29	4	4	4	3	0	10	7	1	6	52.4											-0.0578	-8.6316
SP4.6.5	MCAO	BAD	18.5714	0.75	1	3.9127	120	1.5	6	42	4	1	3	4	0	8	4	4	10	60.7											0.32963	4.20395
SP4.7.2	MCAO	BAD	6.55087	1.83333	1.46525	1.98413	120	5.5	34.5	16	2	3	0	0	0	19	14	8	7	22.7									Immuno	FACS	0.02072	0.72368
SP4.7.5	MCAO	BAD	12.1212	0	0	1.41667	120	0	0	34	4	4	3	2	0	14	8	6	7	58.8									Immuno	FACS	0.14912	3.57895
average			12.6104	1.06988	1.57649	2.13211	119.75	12.8333	18.9097	30.1944	3.83333	1.95833	2.79167	2.95833	0	15.75	13.9583	6.91667	6.625	44.9222	32.0702	44.4048	31.1374	45.5644	56.7038	0.78738	0.72426	0.80065			0.10515	1.74753
stdev			5.19766	1.43353	1.72081	1.66706	1.22474	26.8845	24.6312	25.9099	0.48154	1.33447	1.21509	1.16018	0	6.92349	5.90857	4.10638	2.58444	11.053	5.69228	10.323	17.3891	4.54823	10.9291	0.63959	0.15122	0.2778			0.15816	3.48286
SP2.6.1	MCAO	GOOD (n =	8.30769	3.97727	5.75668	10.2778	120	62.5	109	120	4	4	4	4	0	7	5	4	2	48.4	37.84%	36.41%	10.7917	26.9965	46.8384	0.07296	0.27583	1.1943	Immuno	no	0.6398	5.44079
SP2.7.1	MCAO	GOOD	9.45455	0	3.06196	5.47619	120	0	64.5	120	4	1	1	4	0	10	22	14	3	45									RNA	no	0.53615	11.7829
SP3.1.2	MCAO	GOOD	14.5833	2.33333	2.875	19.5	120	10	23	120	4	3	3	3	0	15	7	2	2		32.71%	22.18%	24.4995	42.7965	51.1013	0.40145	0.6663	0.79923	Immuno	Serum	1.79989	11.4079
SP3.2.4	MCAO	GOOD	20	4.33333	2.875	17.1429	120	13	53.5	120	4	3	2	4	0	20	1	4	4		36.78%	section de	26.2819	48.2118	36.238	0.37652	0.98521	0.47314	Immuno	Serum	1.36756	10.7303
SP3.2.5	MCAO	GOOD (Ch	25	3	10.0699	9.74026	120	3.5	120	120	4	1	4	4	0	20	11	7	7		17.27%	39.72%	12.2919	53.4542	49.2317	0.14769	1.38419	0.49188	Immuno	Serum	0.61648	10.7303
SP3.3.2	MCAO	GOOD (Ch	13.3929	1.16667	1.94444	6.33946	120	2	25.5	108.5	4	1	4	1	0	19	20	7	7		23.58%	38.35%	50.1995	50.8686	46.4976	1.21709	1.03817	0.63522	Immuno	Serum	0.53427	10.9013
SP3.3.3	MCAO	BAD	22.5	12	12.6167	12.1154	120	120	119.5	120	1	1	4	3	0	15	8	5	5		25.89%	39.25%	34.786	49.9432	43.9034						0.00403	0.00658
SP3.4.3	MCAO	GOOD	12.8788	0	3.20536	8.5119	120	0	35.5	102.5	3	1	4	3	0	22	13	5	5		30.67%	34.82%		56.2393	57.9976	N/A	1.04458	0.79612	Immuno	Serum	0.85381	10.3224
SP4.4.2	MCAO	GOOD	12.6667	5.14286	8.52941	10.6667	116	92	120	120	4	4	4	4	0	11	7	3	0	39.6									RNA	Serum	0.53689	2.57895
SP4.6.2	MCAO	GOOD	13.3333	3.875	11.2	8.48214	120	17	112	109.5	4	4	4	4	0	5	8	5	5	33.2									RNA	Serum	0.38858	8.48684
average			15.2117	3.58285	6.21345	10.8253	119.6	32	78.25	116.05	3.6	2.3	3.4	3.4	0	14.4	10.2	5.6	4	41.55	29.2499	35.1233	26.4751	46.93	47.4011	0.44314	0.89905	0.73165			0.73752	8.54971
stdev			5.49565	3.45605	4.02502	4.44605	1.26491	43.5947	41.8152	6.60576	10	10	10	10	10	10	10	10	10	6.64204	7.43404	6.60055	14.7173	9.74588	6.67471	0.45538	0.38102	0.26698			0.54033	4.26386
											0.96609	1.41814	1.07497	0.96609	0	5.92921	6.54557	3.33999	2.26078													
											94.4444	0.87895	0.66626	0.59878	#NUM!	3.6749	4.05691	2.07011	1.40122													

I input %infarct based on CD68, here.
%infarct = (Infarct area with CD68 positive cell accumulation)/(Area of Ipsilesional hemisphere) *100

I added GFAP infarct size data here, but couln't find the intensity data. Seems like not big deal, though.
Masa 2016/8/26
I left out GFAP intensity for now as we are not using it for a figure...
Markus 2016/09/13

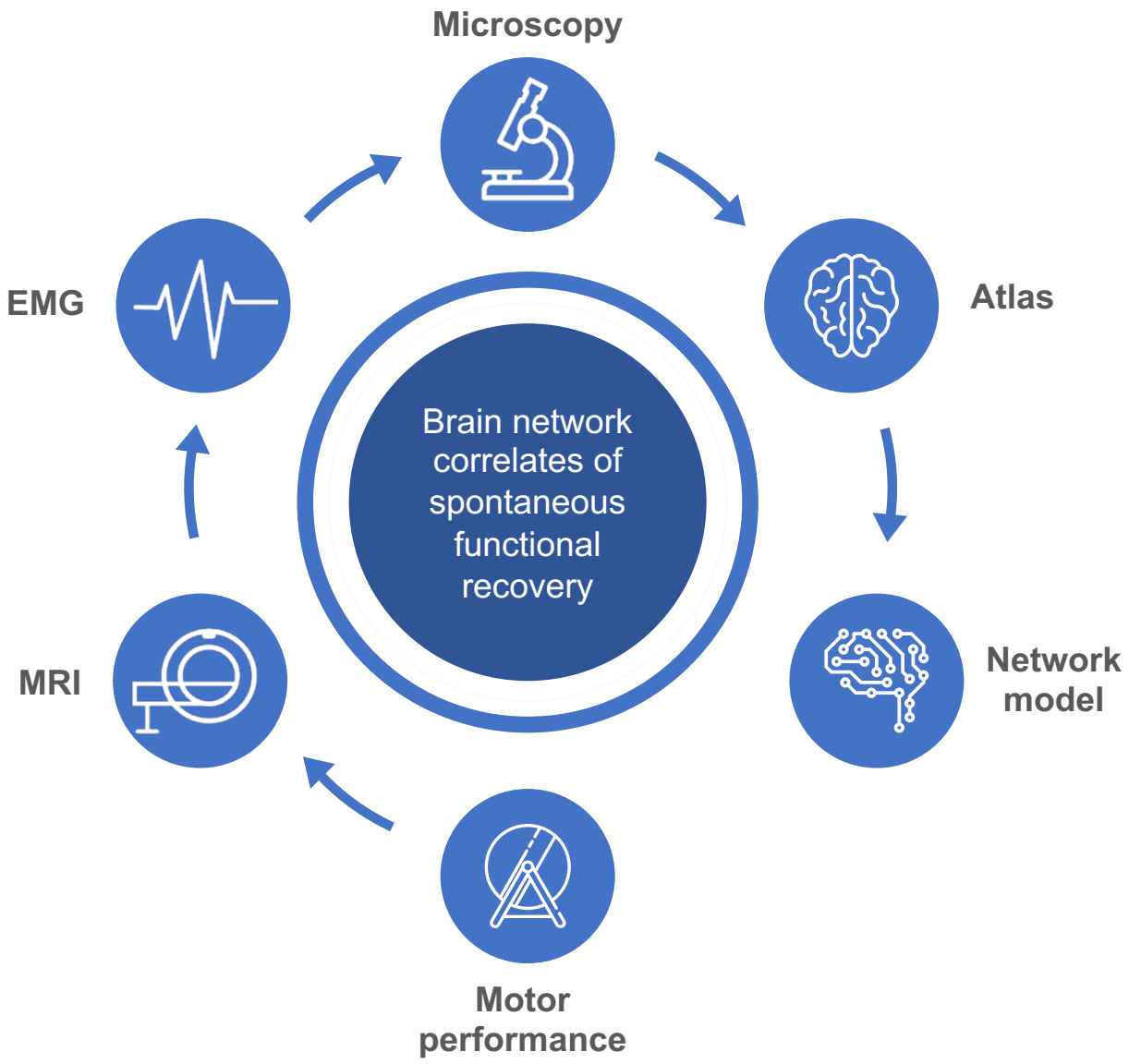
https://www.ophtek.com/wp-content/uploads/2015/03/hard-drive-pile.jpg https://www.faz.net/aktuell/wirtschaft/rezept-gegen-unordnung-simplify-your-office-11527408.html
https://stock.adobe.com/de/images/scientist-or-pharmacist-with-helpless-confused-gesture-and-question-mark-in-thought-bubble-hand-drawn-cartoon-sketch-vector-illustration-whiteboard-marker-style-coloring/336222671

The problem

https://www.hungryshredder.co.uk/how-to-avoid-drowning-in-paper/

The solution

The solution



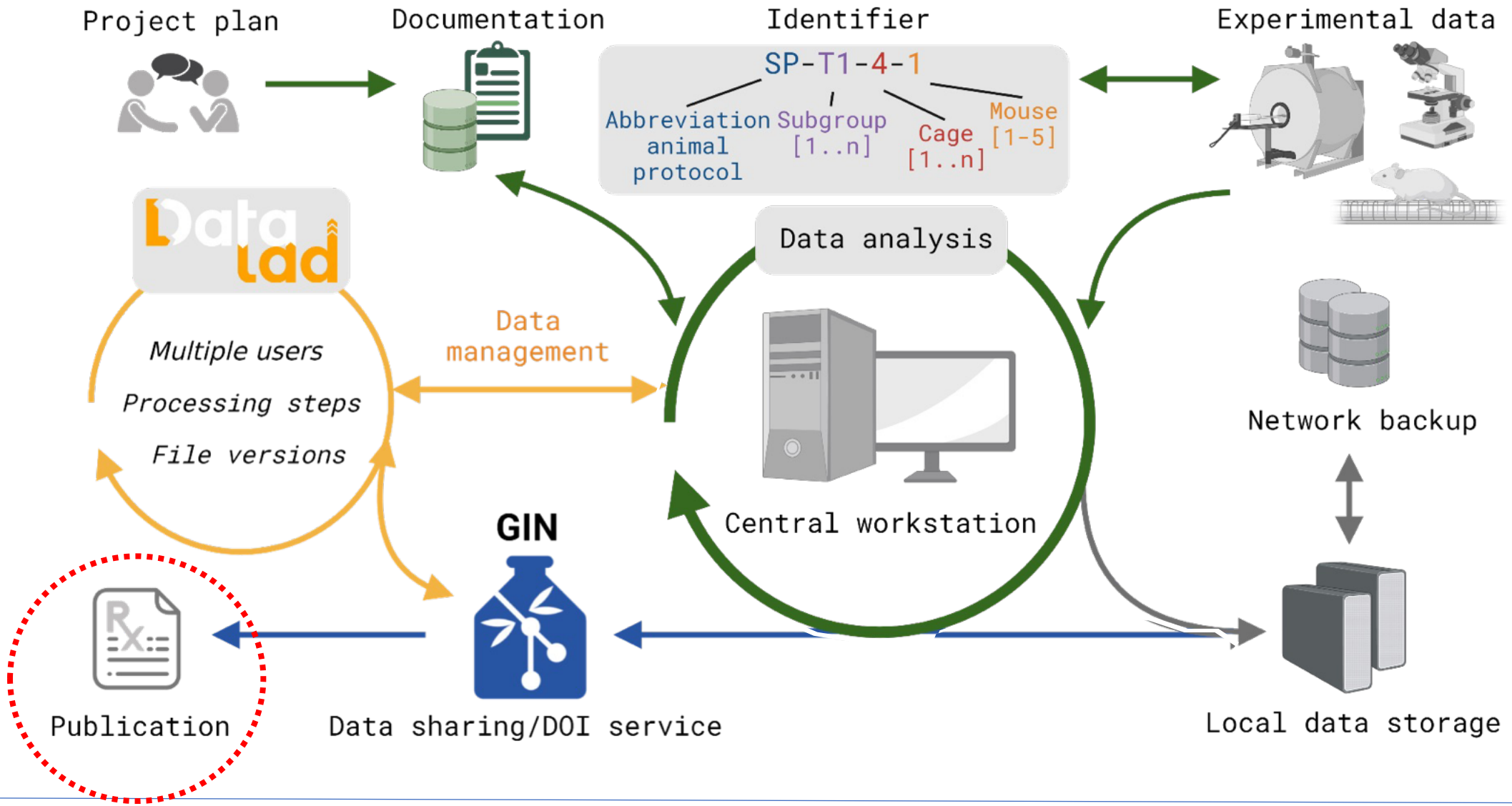
Concept for multimodal small animal data

ARRIVE FAIR IMPROVE

Adopting existing standards

1) RDM strategy => Aref
2) Electronic lab book
3) Standard operating procedures
4) Automated processing

How does the workflow look like?



What are the main tools needed?

Datalad **is a data management tool for scientific research**

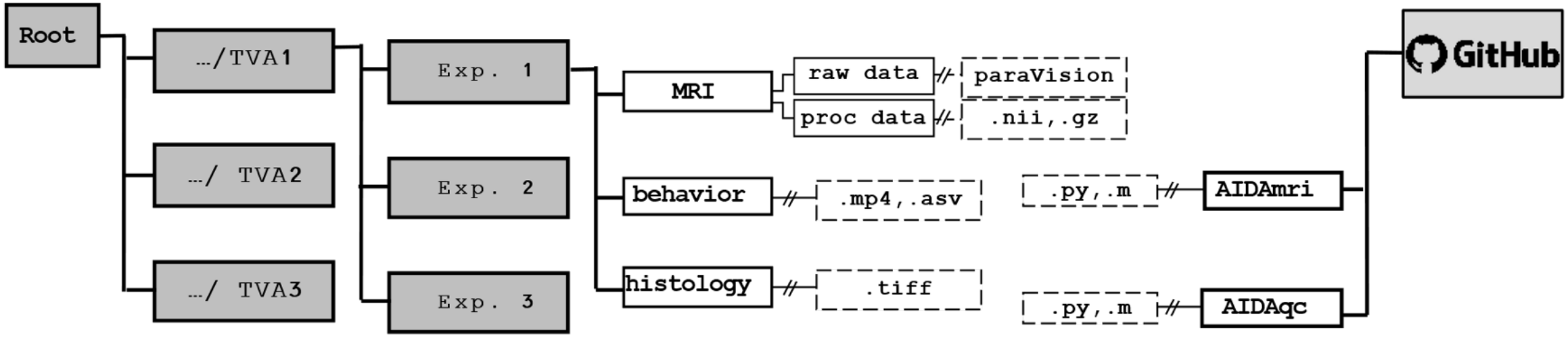
GitHub **is a web-based platform for version control of scientific coding**

Gin.g-node **is a web-based platform for version control of scientific data**

© 2016-2023 G-Node
Hosted by: LMU | Funded by: Federal Ministry of Education and Research | Registered with: re3data.org | Recommended by: SCIENTIFIC DATA PLOS eLife

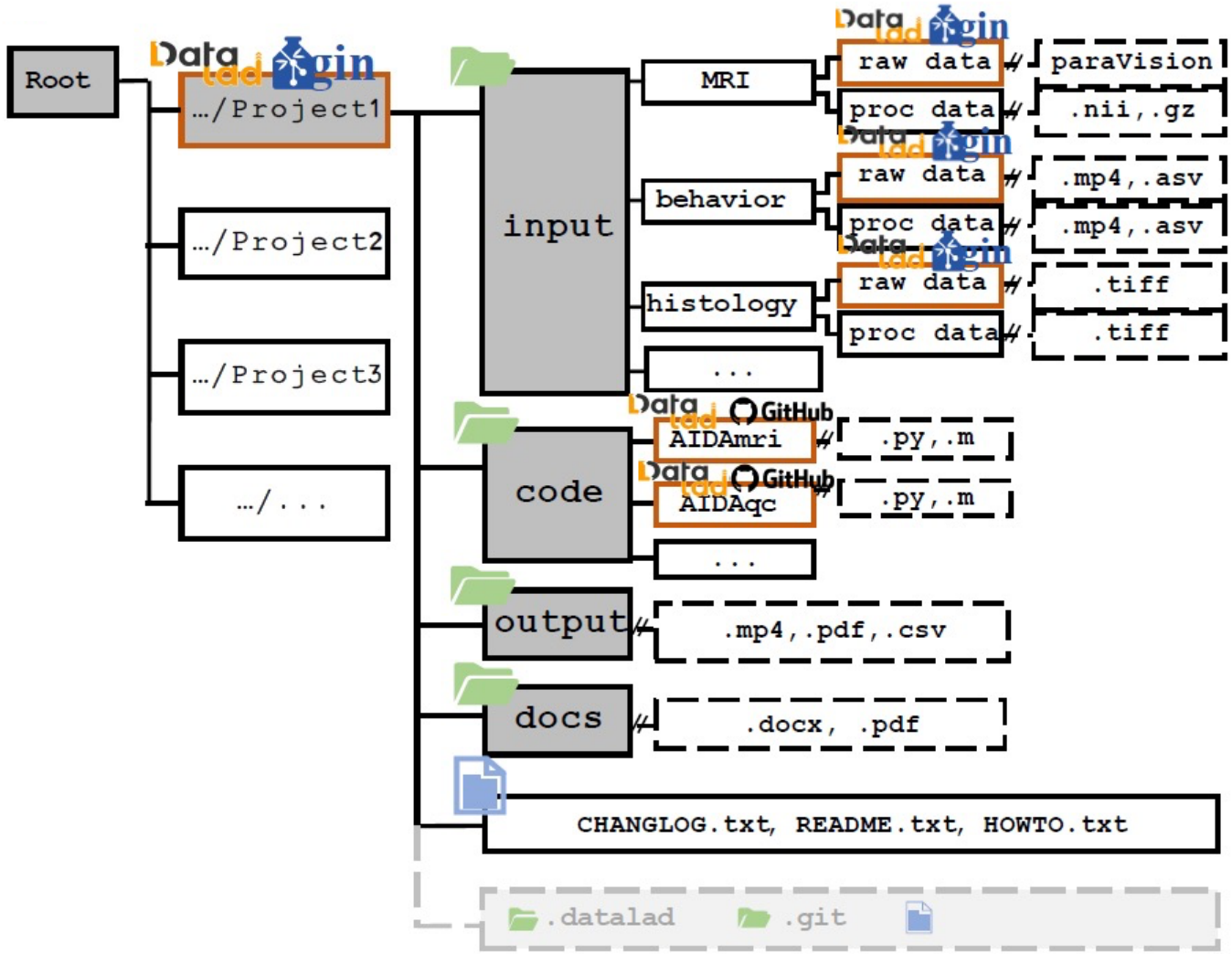
What has happened to our data structure?

Old Structure



What has happened to our data structure?

New Structure



YODA Structure

YODA Principles

Real Example 1: GIN

Dashboard | Issues | Pull Requests | Explore | Help | News

Aswendt_Lab

New Repository

Example_Dataset_Kalantari_Paper ★ 0 ⑂ 1
This is an example Dataset created to see how the dataset is structured based on YODA principles with some small history changes. Use DataLad to clone this dataset.
Updated 4 days ago

2023_Ielacqua_MRIphantomSOP ★ 0 ⑂ 0
Updated 4 days ago

2023_Kuhnhaeuser_PSS ★ 0 ⑂ 0
Updated 5 days ago

2023_Kuhnhaeuser_PSS_behavior ★ 0 ⑂ 0
Behavior videos of the main Project 2023_Kuhnhäuser_PSS
Updated 5 days ago

CRC_WP1 ★ 0 ⑂ 0
The whole project with integrated subdataset
Updated 2 weeks ago

People 5 >

Invite Someone

Teams 1 >

Owners
4 members · 17 repositories

Create New Team

Aswendt_Lab / 2023_Kuhnhaeuser_PSS

How to publish | Add DataCite file

Unwatch 4 | Star 0 | Fork 0

Files | Issues 0 | Pull Requests 0 | Wiki | Settings

No Description

16 Commits | 2 Branches | 0 Releases

Branch: master | 2023_Kuhnhae... | New file | Upload file | GIN | HTTPS | SSH | git@gin.g-node.org:/Aswe

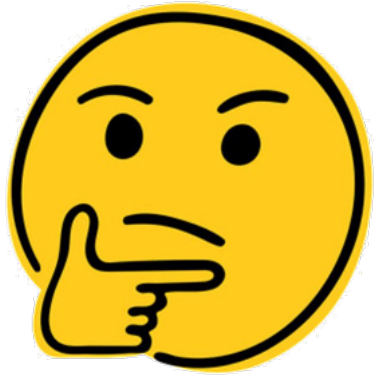
franzizio	dc6b87c453	new videos added	5 days ago
.datalad	715e6e9519	[DATALAD] new dataset	2 months ago
code	880af38b25	update /code	2 weeks ago
docs	c01cb869e1	update /docs	2 weeks ago
inputs	dc6b87c453	new videos added	5 days ago
outputs	9afb408936	initialize rest all	2 months ago
.gitattributes	715e6e9519	[DATALAD] new dataset	2 months ago
.gitignore	62f8e93f2f	initial save git ignore	2 months ago
.gitmodules	cee3702f70	saved git modules	2 months ago

Commit History Search commits | Find

Author	SHA1	Message	Date
franzizio	dc6b87c453	new videos added	2 months ago
franzizio	7dcc8273fc	new videos added	2 months ago
Markus Aswendt	880af38b25	update /code	3 months ago
Markus Aswendt	c01cb869e1	update /docs	3 months ago
Markus Aswendt	e7110d9a1b	add code/video_crop	3 months ago
arefks	cee3702f70	saved git modules	5 months ago

CRC Dataset 4 TB (MRI & Behavior):

60 Mice → Pipeline AIDAmri (Skull-Stripping, Bias field correction, Atlas Registration, …) → Processed Data



Raw Data —Code 1→ **Processed Data** —Code 2→ **Table A, Figure A**

Hmmm, I want to process the Data with some other initial parameters

Raw Data —Code 1'→ **Processed Data** —Code 2→ **Table A', Figure A'**

Concept for multimodal small animal data

Neuroimaging & Neuroengineering Lab

1) RDM strategy => Aref

2) **Electronic lab book**

3) Standard operating procedures

4) Automated processing

Cloud-based relational database for managing large amounts of multimodal animal data

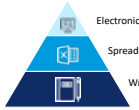
Objective

Data management becomes prone to user-errors when working with extensive multimodal and longitudinal datasets. Although outlined in the GLP (Good Laboratory Practice) of the WHO[1], most labs store their data not in a standardized way and lack way behind clinical standards such as GCP (Good Clinical Practice) compliant data management[2]. Researchers tend to underestimate the importance of a centralized and smart data handling, which is interfering with their efforts in basic research and translational approaches[3]. Here, we present a database specifically designed to meet the requirements for large multimodal imaging studies with the aim to provide an advantage compared to conventional paper lab notebooks, Excel sheets as well as data management tools (e.g REDCap[4]) and electronic laboratory notebooks (eLNs)[5].

Amount of data scales with number of subjects, subgroups and measurements:

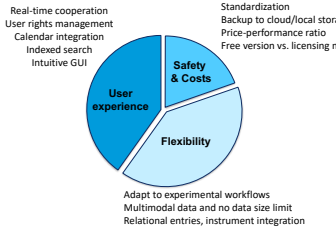
*100 mice in 2 subgroups x 6 tests (e.g. in vivo imaging and behavior) x 5 time points = **3000 data points***

Common ways to organize experimental data

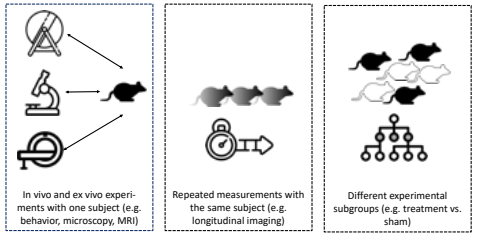


Requirements

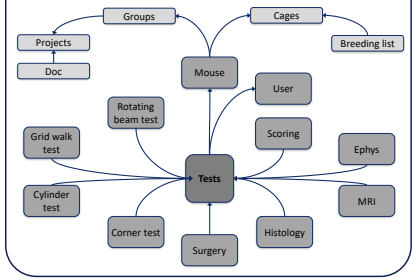
General requirements for storing data in a central database



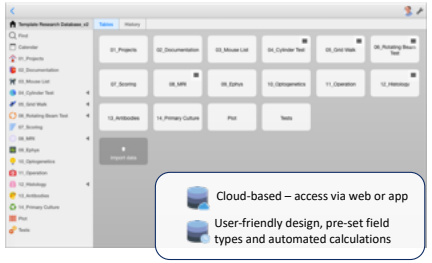
Specific requirements for in vivo, longitudinal data



Relational database model

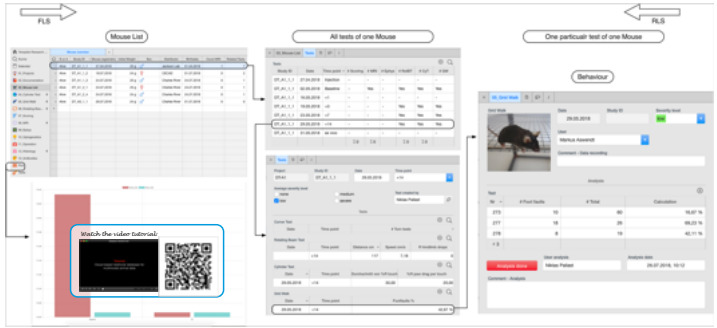


Web interface



- Cloud-based – access via web or app
- User-friendly design, pre-set field types and automated calculations

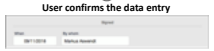
Data entry



1. User enters data –> Necessary fields highlighted in red

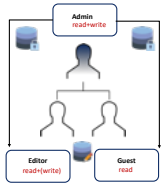
2. When last field is finished, the new button Done? appears

3. User confirms the data entry



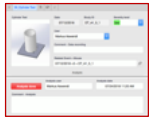
4. Only the admin keeps control to re-set confirmation

User management

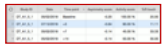


1. User enters data –> Analysis done

2. User has no access to the data

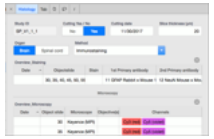


3. Data remains visible in table view

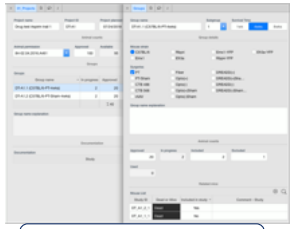


Filtering and sorting

Histology



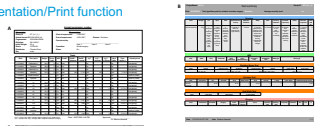
Animal protocol-relevant information



- Automated calculation of animal numbers per project/group/sub-group in relation to animal protocol permission

Plot function

Documentation/Print function



Conclusion

- Keep track of experimental data & groups, number of animals used per project
- Quick and platform-independent multi-user access, blinded experimenter
- Only minimal programming skills necessary to modify/extend the database
- Secure sharing of experimental raw & processed data with colleagues and guests
- Comprehensive documentation and animal protocol reporting

References

1. Organization, W. Handbook: good laboratory practice (GLP): quality practices for regulated non-clinical research and development. (2010).
2. Step, I. Integrated addendum to ICH E6 (R1): Guideline for Good Clinical Practice E6 (R2). *Current Step (2015)*.
3. Lapchak, PA & Zhang, JH. Data standardization and quality management. *Translational Stroke Research* (2017).
4. Harris, P. A. *et al.* Research electronic data capture (REDCap) a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of biomedical informatics* **42**, 377–381 (2009).
5. Dirnagl, U. & Przesdzing, I. A pocket guide to electronic laboratory notebooks in the academic life sciences. *F1000research* **5**, 2 (2016).
6. Francis Crick (1958/59). "The Replication of DNA. Wellcome Library for the History and Understanding of Medicine. Francis Harry Compton Crick Papers, licensed under a Creative Commons Attribution 4.0 International license.
7. http://www.faz.net/aktuell/wirtschaft/rezept-gegen-unordnung-simplify-your-office-11527408/wo-ist-noch-mal-die-rechnung-11528426.html

Icons from https://www.flaticon.com/.

Acknowledgements

This work was supported by the Friebe Foundation. The database runs on Ninox software (www.ninoxdb.de). The authors received no financial support by the Ninox GmbH.

Pallast et al. Database 2018

Concept for multimodal small animal data

Neuroimaging & Neuroengineering Lab

1) RDM strategy => Aref

2) Electronic lab book

3) SOPs

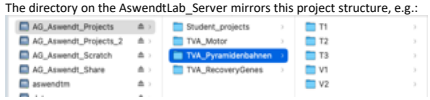
4) Automated processing

How to use save data	Version:	1
	Effective per:	11/14//2022
	Pages:	2
	Signature author:	Markus Aswendt

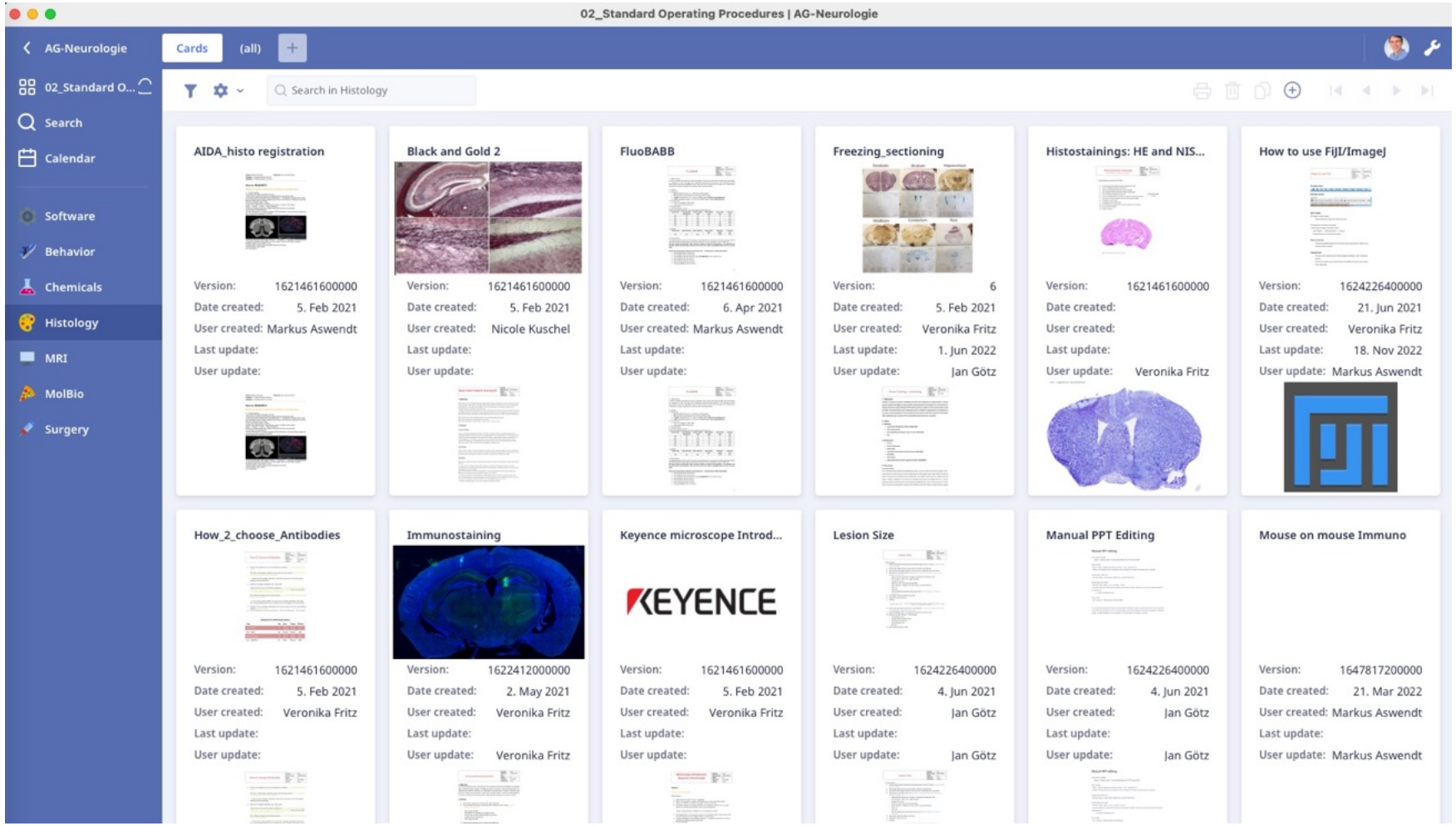
Concept raw data

- All raw data is stored in folders, which are organized according to the animal protocol[1]
  - To get an overview of all subprojects, see 01_Projects (Ninox database):

- The directory on the AswendtLab_Server mirrors this project structure, e.g.:

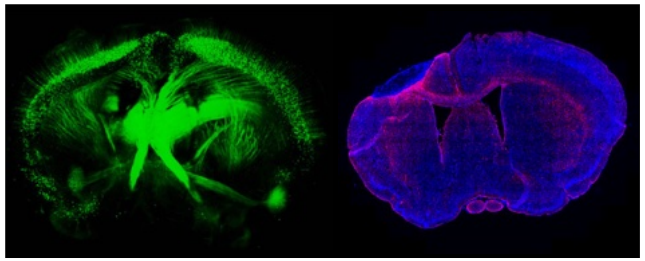


- Each subproject contains on the next lower level the modalities (Behavior, MRI, etc.). The modality folders contain the timepoints (e.g. baseline, P3, etc.). The timepoint folder contain submodalites or cages/groups of mice.

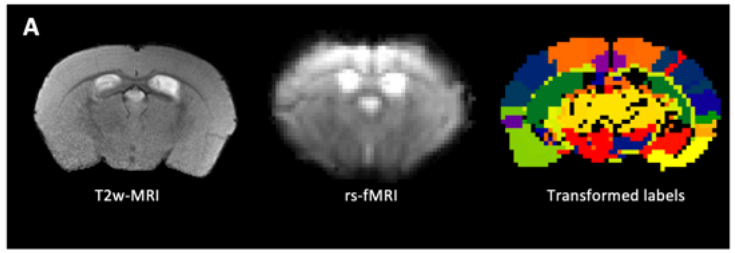


Concept for multimodal small animal data

1) RDM strategy => Aref
2) Electronic lab book
3) Standard operating procedures
4) **Automated processing**



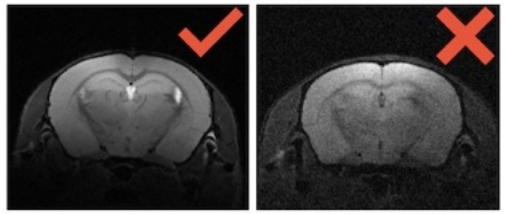
HISTO



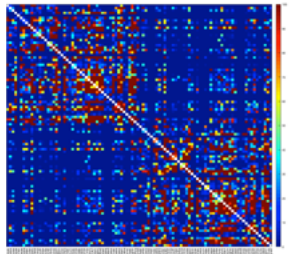
MRI

AIDA

QC



CONNECT



GitHub

https://github.com/aswendt-lab

Kalantari et al. in preparation, Scharwächter et al. Neuroimage 2022
Pallast et al. Neuroimage 2020, Pallast et al. Front Neuroinform 2019