



Tools and Workflows for Data & Metadata Management of Complex Experiments

Building a Foundation for Reproducible & Collaborative Analysis
in the Neurosciences

Julia Sprenger

Schlüsseltechnologien / Key Technologies

Band / Volume 222

ISBN 978-3-95806-478-2

Forschungszentrum Jülich GmbH
Institute of Neuroscience and Medicine
Computational and Systems Neuroscience (INM-6 / IAS-6)

Institut für Neurowissenschaften und Medizin
Decoding the human brain at systemic levels (INM-10)

Tools and Workflows for Data & Metadata Management of Complex Experiments

Building a Foundation for Reproducible & Collaborative
Analysis in the Neurosciences

Julia Sprenger

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies

Band / Volume 222

ISSN 1866-1807

ISBN 978-3-95806-478-2

Contents

1	Introduction	1
1.1	Data and metadata models	3
1.1.1	Hierarchical metadata in the <i>odML</i> model	4
1.1.2	Generic data organization via the <i>Nix</i> model	8
1.2	Thesis overview	11
2	Sharing data	13
2.1	Relevance to the field	16
2.2	The experiment	18
2.3	The metadata structure	24
2.4	Data and metadata processing pipeline	24
2.5	Data loading and enrichment with metadata	28
2.6	Shortcomings of the <i>odML</i> generation pipeline	32
2.7	Requirements for maintainable and reproducible metadata management	37
2.7.1	Evaluation of presented metadata pipeline	39
3	Metadata management	43
3.1	Introduction to <i>odMLtables</i>	43
3.2	Software description	47
3.2.1	Tabular representation of the <i>odML</i> format	49
3.2.2	Software functionalities	49
3.2.3	Software architecture	52
3.3	Embedding <i>odMLtables</i> in data acquisition and analysis	53
3.4	Discussion	57
3.4.1	Performance estimation	59
3.4.2	<i>odMLtables</i> as conversion tool	60
3.4.3	Relation to electronic laboratory notebooks	61
3.4.4	Outlook	62
4	Data representation	67
4.1	The <i>Neo</i> Python package	69
4.1.1	Feature updates and current development	70
4.1.2	<i>Neo</i> object structure	76
4.2	<i>Neo</i> usage examples	78

4.2.1	Loading & visualization	79
4.2.2	Annotation of data with metadata from <i>odML</i>	79
4.2.3	Saving data & format conversion	82
4.3	Comparison of <i>Neo</i> and <i>NWB:N</i>	85
4.4	Summary	87
5	Workflow management	89
5.1	Workflow management tools - <i>Snakemake</i>	92
5.2	Practical application	97
5.2.1	The Vision-for-Action project	97
5.3	Metadata workflow in the Vision-for-Action project	103
5.3.1	Discussion	110
5.3.2	Workflow evaluation	113
5.4	Summary & guidelines	115
6	Discussion	117
6.1	Comparison of Reach-to-Grasp and Vision-for-Action workflows for data and metadata handling	121
6.1.1	Experimental design	121
6.1.2	Concept for metadata aggregation	122
6.1.3	Changes due to software updates	122
6.1.4	Usability	123
6.1.5	Pipeline and workflow approach	124
6.2	Outlook	124
6.2.1	The future of <i>odMLtables</i>	124
6.2.2	The future of <i>Neo</i>	125
6.2.3	Automated workflow management	125
6.2.4	Data analysis	126
6.2.5	Published datasets	126
6.2.6	Lessons to learn	128
6.2.7	Concept extension	128
6.2.8	Dissemination of a data and metadata workflow system	129
6.2.9	Looking further ahead	131
A	Supplementary description of the Reach-to-Grasp experiment	133
A	Experimental apparatus	133
A.1	Behavioral control system	135
A.2	Neural recording platform	138
A.3	Origin of the channel IDs	140
B	Data preprocessing	140
B.1	Translation of digital events to trial events	142
B.2	Preprocessing of behavioral analog signals	142
B.3	Offline spike sorting	143

B.4	Code availability	144
C	Data records	145
D	Technical validation	150
D.1	Correction of data alignment	150
D.2	Quality assessment	150
D.3	LFP data quality	151
D.4	Spike data quality	152
E	Usage notes	153

Schlüsseltechnologien / Key Technologies

Band / Volume 222

ISBN 978-3-95806-478-2