





STEPHAN HEUNIS, CHRISTIAN MOENCH, BENJAMIN POLDRACK, & MICHAEL HANKE,

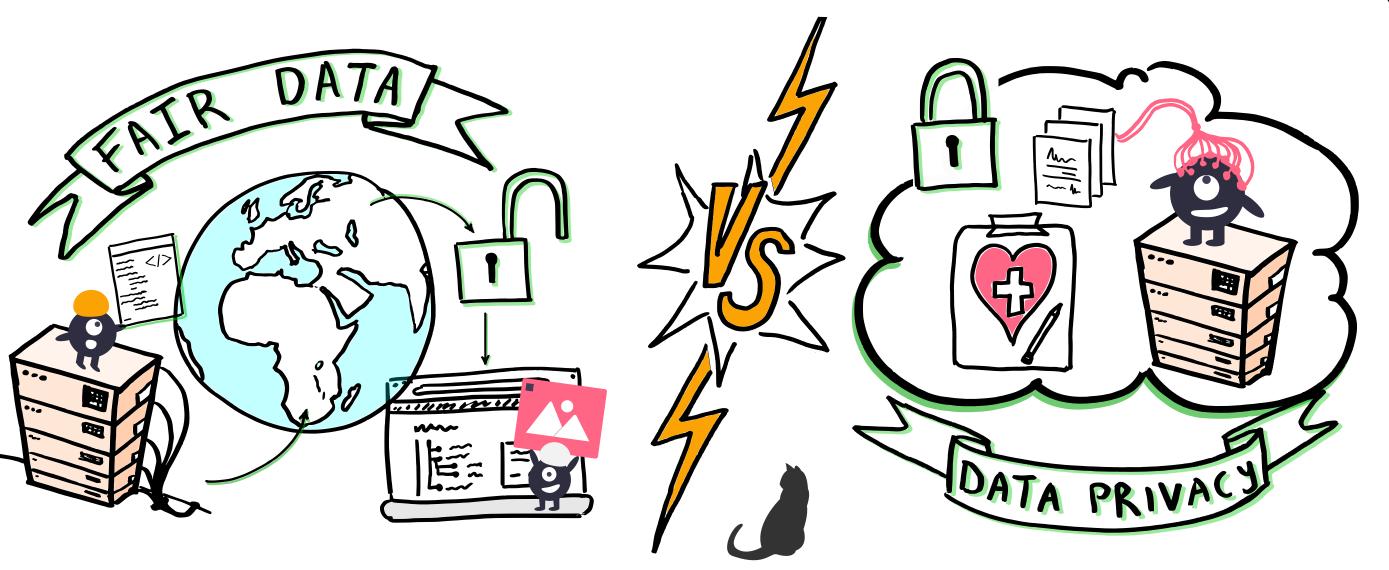
INSTITUTE OF SYSTEMS NEUROSCIENCE, MEDICAL FACULTY, HEINRICH HEINE UNIVERSITY DUESSELDORF, GERMANY

INSTITUTE OF NEUROSCIENCE AND MEDICINE, BRAIN & BEHAVIOUR INM-7, RESEARCH CENTER JUELICH, GERMANY

GENERATE A USER-FRIENDLY DATA BROWSER FROM STRUCTURED METADATA USING DATALAD-CATALOG

THE CHALLENGE-

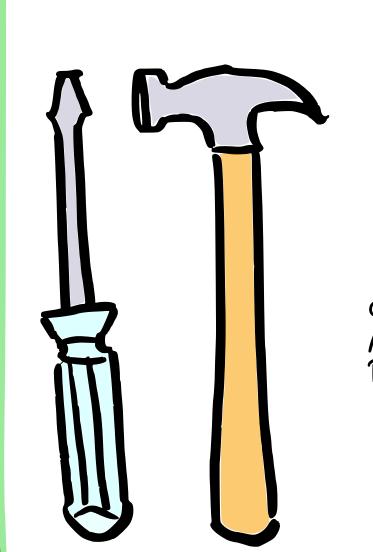
THE IMPORTANCE AND BENEFITS OF MAKING RESEARCH DATA FINDABLE ACCESSIBLE INTEROPERABLE AND REUSABLE (I.E. FAIR) ARE CLEAR'. BUT OF EQUAL IMPORTANCE IS OUR ETHICAL AND LEGAL OBLIGATIONS TO PROTECT THE PERSONAL DATA PRIVACY OF RESEARCH PARTICIPANTS.

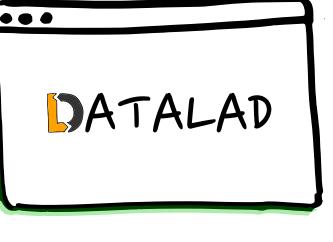


SO WE ARE STRUCK WITH THIS APPARENT CONTRADICTION: HOW CAN WE SHARE OUR DATA OPENLY YET KEEP IT SECURE AND PROTECTED? SHOULD WE ERR ON THE SIDE OF FAIR DATA OR OF DATA PRIVACY? OR HOW CAN WE BALANCE THE ASSUMED TRADEOFF BETWEEN THESE ASPETCS?

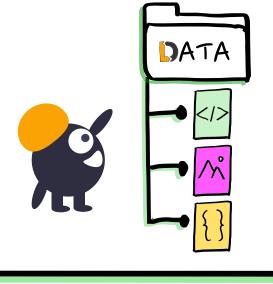
THE TOOLSET-

THESE IDEALS ARE ACHIEVABLE IN PRACTICE WITH A FREE AND OPEN SOURCE TOOLSET. DATALAD CAN BE USED FOR DECENTRALISED MANAGEMENT OF DATA AS LIGHTWEIGHT, PORTABLE AND EXTENSIBLE REPRESENTATIONS. DATALAD-METALAD CAN EXTRACT STRUCTURED HIGH- AND LOW-LEVEL METADATA AND ASSOCIATE IT WITH DATASETS OR WITH INDIVIDUAL FILES. AND AT THE END OF THE PIPELINE, DATALAD-CATALOG CAN TURN THE STRUCTURED METADATA INTO A USER-FRIENDLY DATA BROWSER!



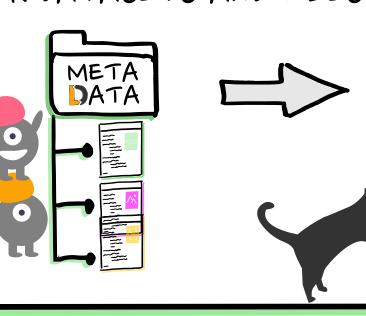


CREATES A LIGHTWEIGHT AND ABSTRACT DATASET REPRESENTATION





EXTRACTS AND AGGREGATES METADATA FOR DATASETS AND FILES





Data

GENERATES

A DATA

FROM

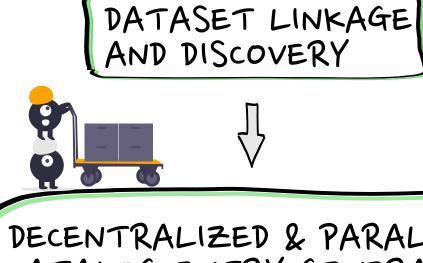
BROWSER

METADATA

5 THE DECENTRALIZED & COLLABORATIVE WORKFLOW—

WHILE EXISTING CATALOGING TOOLS OFTEN RELY ON CENTRALISED INFRASTRUCTURE WITH ACCESS TO THE DATA THE PORTABILITY OF THE DATALAD-BASED TOOLSET ALLOWS DECENTRALIZED AND COLLABORATIVE CATALOG GENERATION AND MAINTENANCE AS THE OPTIMAL STANDARD. INSPIRED BY THE "FAIRLY BIG" PROCESSING WORKFLOW? THE FAIRLY BIG CAT WORKFLOW DEMONSTRATES HOW THE STEPS OF DATASET LINKAGE AND DISCOVERY, PARALLEL METADATA EXTRACTION AND CATALOG ENTRY GENERATION, AND ULTIMATELY CATALOG MAINTENANCE CAN BE AUTOMATED WITH THE USE OF TOOLS LIKE DATALAD AND COMMON WORKFLOW MANAGERS.



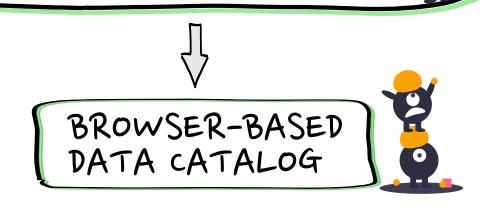


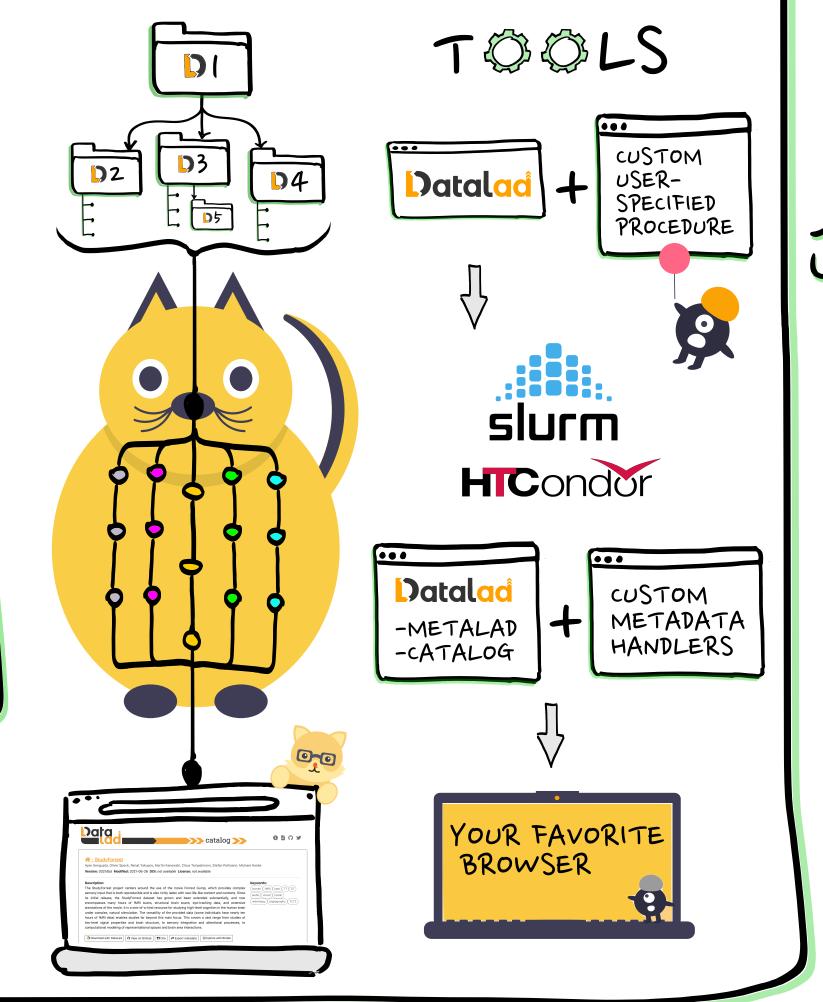
DECENTRALIZED & PARALLEL CATALOG ENTRY GENERATION IN SHORTLIVED WORKSPACES:

- BRANCH PER DATASET
 - (1) CLONE DATASET
 - 3 TRANSLATE METADATA

2 EXTRACT METADATA

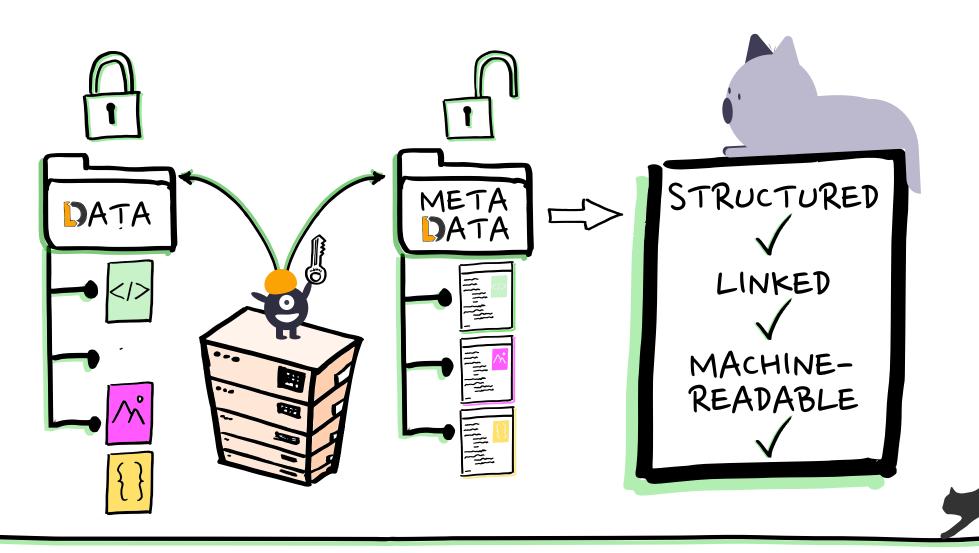
- 4 ADD TO CATALOG
- OCTOPUS MERGE





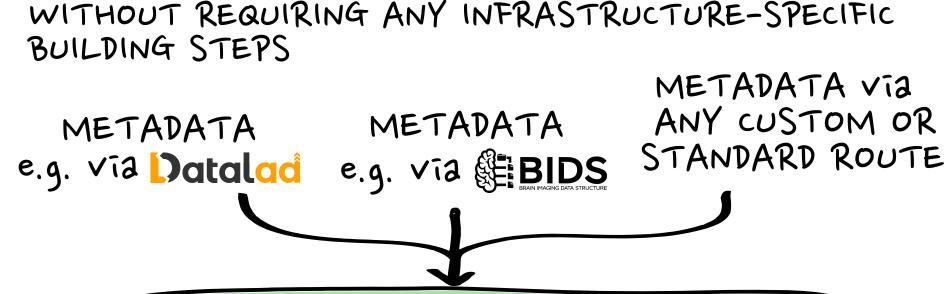
2 THE OPPORTUNITY -

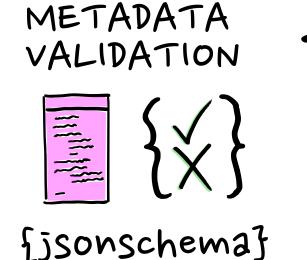
STRUCTURED, LINKED, AND MACHINE-READABLE METADATA PRESENTS A POWERFUL OPPORTUNITY TO ADDRESS THIS CONTRADICTION. METADATA PROVIDES BOTH HIGH-LEVEL INFORMATION ABOUT OUR RESEARCH DATA (SUCH AS STUDY DESIGN PARTICIPANT INFORMATION AND DATA ACQUISITION PARAMÉTERS) AND LOWER-LEVEL DÉSCRIPTIVE ASPECTS OF EACH FILE IN THE DATASET (SUCH AS FILENAMES, RELATIVE PATHS, SIZES, AND FORMATS). WITH THIS METADATA, WE CAN CREATE AN ABSTRACT REPRESENTATION OF A FULL DATASET THAT IS SEPARATE FROM THE ACTUAL DATA CONTENT. THIS MEANS THAT THE CONTENT CAN BE STORED SECURELY TO MAINTAIN PRIVACY WHILE THE METADATA CAN BE SHARED OPENLY TO COMPLY WITH FAIR PRINCIPLES.



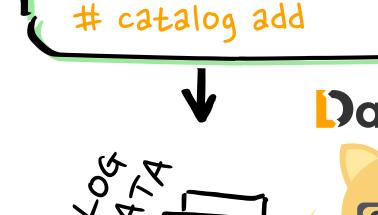
THE CATALOG-

METADATA EXTRACTED FROM VARIOUS SOURCES CAN BE AGGREGATED AND ADDED TO A CATALOG. INCOMING METADATA IS FIRST VALIDATED AGAINST A CATALOG-SPECIFIC SCHEMA BEFORE THE CATALOG ITSELF IS GENERATED OR INDIVIDUAL ENTRIES ARE ADDED. THE OUTPUT IS A SET OF STRUCTURED METADATA FILES AND A VUEJS-BASED BROWSER INTERFACE THAT HANDLES AUTOMATIC RENDERING. THIS CAN BE HOSTED AND SERVED AS STANDALONE CONTENT WITHOUT REQUIRING ANY INFRASTRUCTURE-SPECIFIC





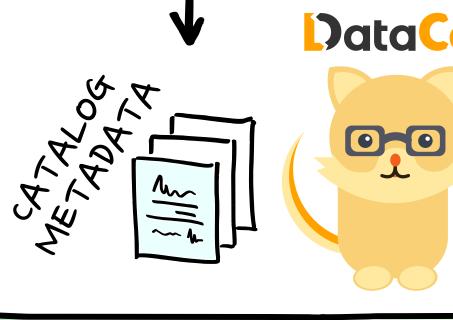






CATALOG

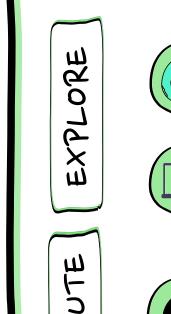
GENERATION





6 TRY IT OUT!

DATALAD-CATALOG IS A FREE AND OPEN-SOURCE SOFTWARE TOOL WITH A COMMAND LINE INTERFACE AND PYTHON API. THAT ALLOWS THE CREATION OF USER-FRIENDLY DATA BROWSERS FROM STRUCTURED METADATA. INSTALL. EXPLORE AND CONTRIBUTE VIA THESE LINKS:



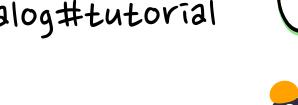
SERVE

PUSH

DEMO: datalad.github.io/datalad-catalog







TUTORIAL

INSTALL





FAIRLY-BIG-cAT: github.com/jsheunis/fairly-big-catalog-workflow









FUNDING







WILKINSON, M. ET, C. ET AL. (2016). DOI: 10.1038/SDATA.2016.18

REFERENCES

² HALCHENKO, Y.O. ET AL. (2021). DOI: 10.21105/JOSS.03262 3 WAGNER A.S. ET AL. (2022). DOI: 10.1038/S41597-022-01163-2





s.heunis@fz-juelich.de

CONTACT



TALK



DEMO





CODE