



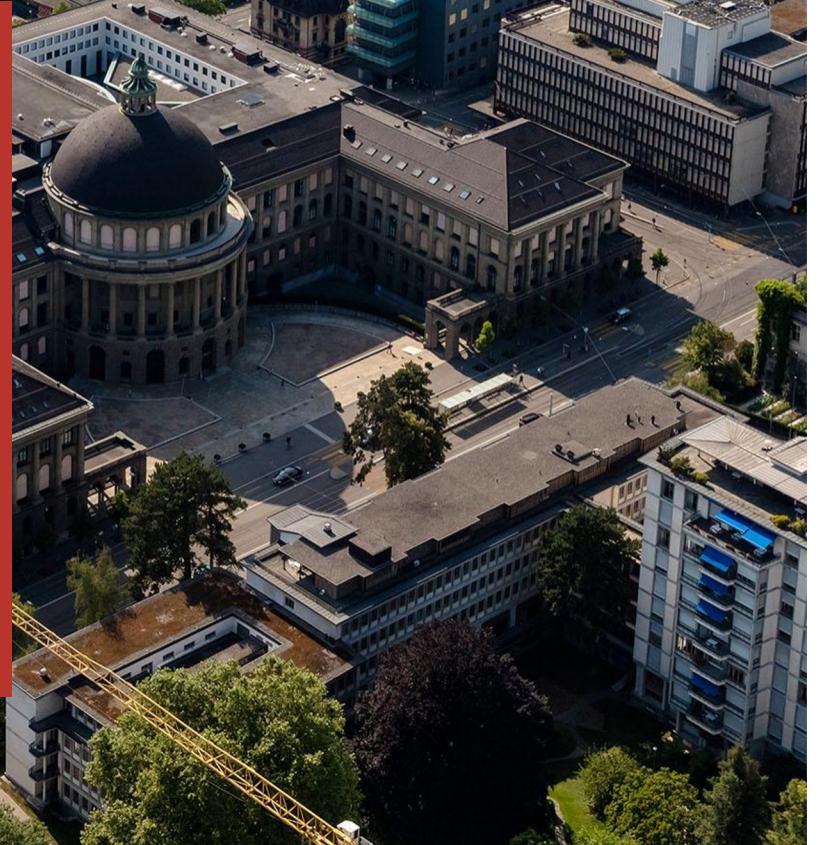
LabKey Server for Reproducible Biomedical Research

Natalia Chicherova

Biomedical Data Manager

2023-04-19, Zurich

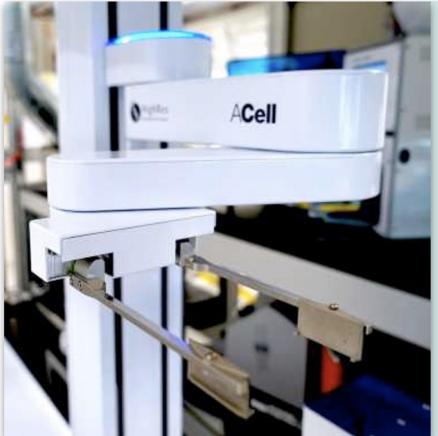
PUBLIC



About Us



We enable and accelerate the execution of R&D projects for the personalized health community!



Screening & Lab Automation

Organoids & spheroids
CRISPR
Phenotypic screening

Bioinformatics

Multi-omics analysis
Reproducible pipelines
Data management



Statistics

Experimental design
Multi-omics analysis
Clinical trials

Software Engineering

Web services
Clinical decision support
Method tool box



Data Management @ NEXUS



Tumor Profiler 1.0
Tumor Profiler Center



INTeRCePT
mTORUS



Swiss
Personalized
Health
Network



Strategic Focus Area
Personalized Health
and Related Technologies

National Data Streams

Swiss Personalized Oncology
IICU (Infections in Swiss ICUs)



Universitäts
Spital Zürich

Balgrist

Universitätsklinik

ETHzürich



Centre hospitalier
universitaire vaudois



Hôpitaux
Universitaires
Genève

Kantonsspital
Baselland



Universitätsspital
Basel



INSELSPITAL
UNIVERSITÄTSSPITAL BERN
HÔPITAL UNIVERSITAIRE DE BERNE

What is research data management (RDM)?

- ❑ **RDM** is the process of storing, organizing and maintaining the data created and collected in a study.
- ❑ **RDM** is concerned with handling of research data throughout their entire life cycle and this way tightly connected with reproducibility of the study.
- ❑ The **data management process** includes a combination of different functions that collectively aim to make sure the data in the study is accurate, available and accessible.



<https://library.ethz.ch/en/researching-and-publishing/data-management-and-policies/research-data-management.html>
<https://www.techtarget.com/searchdatamanagement/definition/data-management>

What does RDM mainly consist of?

- Systematically collect, name and store research data (raw & derived) in one place
- Collect and organise metadata and ensure it is complete
- Ensure data security
- Collect standard operating procedures (SOPs)



Our research data management approach is based on 4 main blocks:

1 – Sample, experiment, and analysis tracking

→ LabKey web server for comprehensive metadata collection

2 – Data repository

→ a secure server (BioMedIT) for data upload and sharing

3 – User access management

4 – Documentation

→ up-to-date information on project data management available online

→ SOPs in project git repository

What is LabKey?



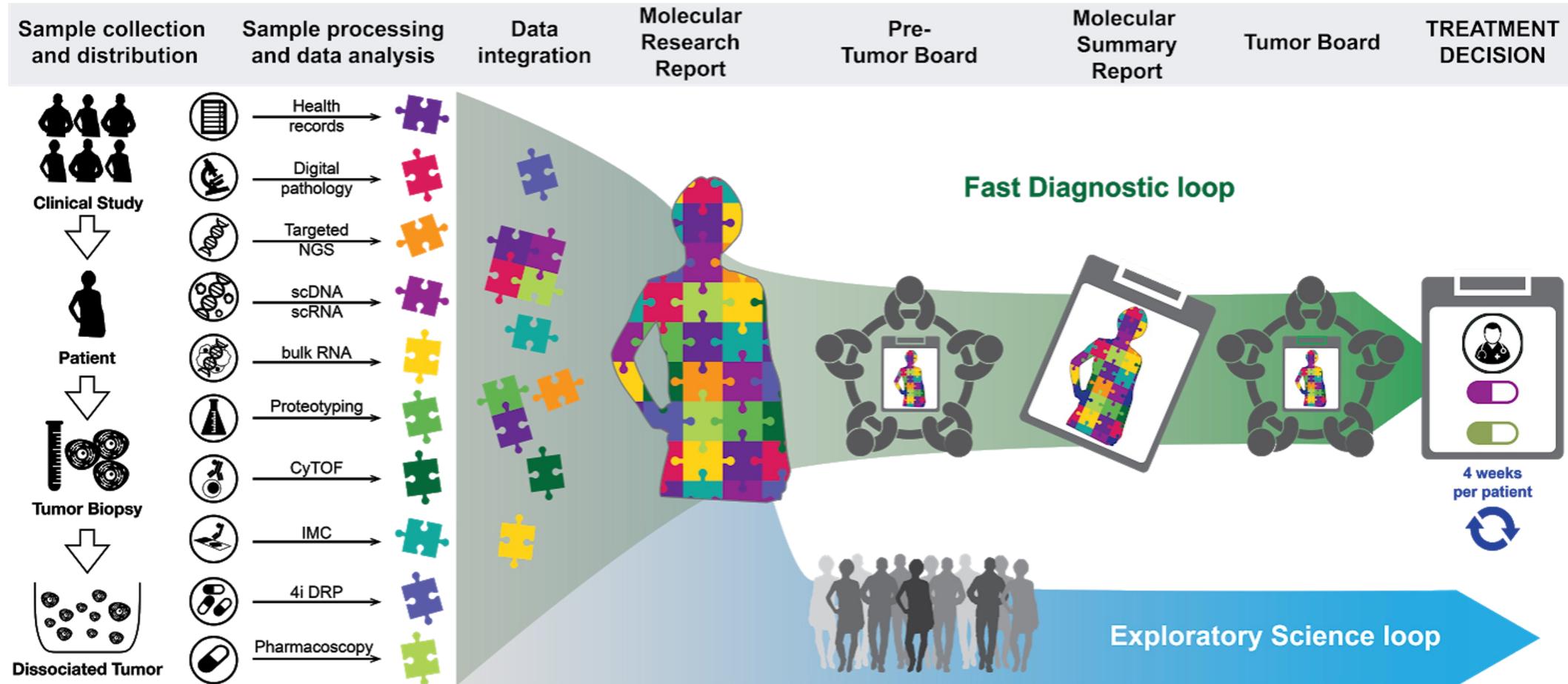
- LabKey Server is a **Scientific Data Management System** that helps to manage and analyze scientific data.
- It was developed at the **Fred Hutchinson Cancer Research Center (Seattle)** to manage proteomics data.
- LabKey Server provides a **metadata repository** that allows **web-based querying, reporting, and collaborating**
- LabKey offers a **free community edition** (presented here) and more powerful paid editions

A screenshot of the LabKey web application interface. The top navigation bar includes tabs for Overview, Participants, and Samples, along with a search bar and quick links for Melanoma Samples and All Participants. The main content area shows a 'Study Overview' section with counts for Number of patients (2) and Number of melanoma samples (1). Below this is a 'Sample Types' section containing a table with four rows of data. The table columns are Name, Description, Created, Created By, Modified, Modified By, Folder, and Sample Count. The data is as follows:

Name	Description	Created	Created By	Modified	Modified By	Folder	Sample Count
BioBank		2023-04-05 15:28	chicherova	2023-04-14 12:50	chicherova	LabKey Example	1
IMC_Samples		2023-04-14 11:06	chicherova	2023-04-14 11:06	chicherova	LabKey Example	0
Melanoma_Samples		2023-04-05 15:23	chicherova	2023-04-14 12:57	chicherova	LabKey Example	1
Ovarian_Samples		2023-04-14 11:17	chicherova	2023-04-14 11:25	chicherova	LabKey Example	1

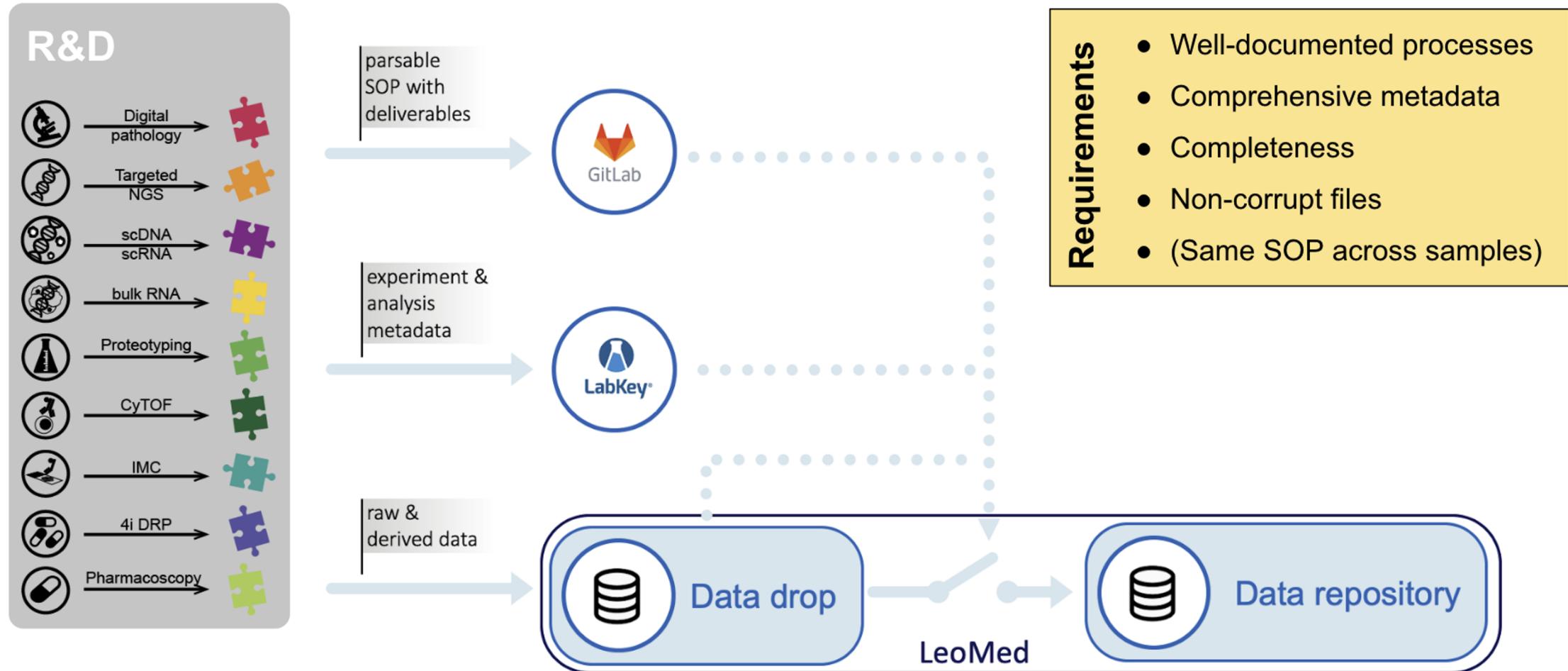
<https://www.labkey.com/products-services/labkey-server/>

LabKey is used in the Tumor Profiler Study



Irmisch et al., Cancer Cell, 2021

Ensure Creation of a High-Quality Research Dataset



Why LabKey?

1. Good visualisation of parent-child lineage for sample tracking
2. Customisable (python API, JS API)
3. User access audit log
4. Importing csv, tab-separated
5. Export Excel files
6. User-friendly interface

Note: LabKey is not suitable to store large amounts of data (FASTQs, DICOMs, ...)

Let's check our own local LabKey Server

- ❑ git clone <https://github.com/ETH-NEXUS/labkey-dockerized-app>
- ❑ start the containers with docker
- ❑ go to localhost:8080

Main features we will show in LabKey

- Sample sets
 - Lookup fields
- Datasets (participants)
 - Participant view
 - Cohorts assignment
- DataClasses
 - Lineage with Sample Sets
- Add query view
- Add external modules
- Add pages – successful registration
- User groups

Take-home messages

1. It is important to push researchers towards documenting their data because **the path to research findings is just as important as the findings themselves.**
2. For large volume of research data and collaborations among multiple groups it is essential to have a **data management strategy** for study reproducibility.
3. LabKey is a powerful and flexible tool for the management of research metadata.

https://dimewiki.worldbank.org/Research_Ethics

Thank you!

Tumor Profiler Data/IT Team

Natalia Chicherova

Christian Kunze

Kjong Lehmann

Gunnar Rätsch

Vipin Sreedharan

Daniel Stekhoven

Nora Toussaint

Shuqing Yu

Marc Zimmermann

Data Management @ NEXUS

Natalia Chicherova

Vipin Sreedharan

Nora Toussaint

Shuqing Yu



Natalia Chicherova
Biomedical Data Manager
chicherova@nexus.ethz.ch

ETH Zürich
NEXUS Personalized Health Technologies
Wagistrasse 18
8952 Zürich (Schlieren)

www.nexus.ethz.ch

