



# Publication of open research data with Sync & Share storage

**Andreas Ia Roi, Madeleine Fritschi,  
Gianluca Caratsch, Tilo Uwe Steiger**

ETH Library / IT Services  
March 11 2024, Meyrin

# Agenda

1. Why Libdrive?
2. Implementation as an extension
3. Libdrive in a nutshell
4. Libdrive example in Research Collection
5. Launch, storage space and usage of Libdrive
6. Lessons Learned and open points

# Why Libdrive?

- Repeatedly expressed need of ETH researchers for publication of large data sets
- ETH «only» offers [Long term storage \(LTS\)](#) without option for publication or active preservation measures
- Restrictions by repository software: DSpace is not designed for handling the publication of large data sets
  - Uploading large files leads to timeout problems
  - No folder upload
  - Current soft limits: 10 GB/file, 50GB/entry

# Implementation as an extension (loosely integrated)

- Integrated solution in DSpace not feasible
- Feasibility tests with internal ETH Library pilot instance
- Decision: loosely integrated external extension
  - Data hosted on a separate ownCloud platform
  - Advantage: later switch to a different solution

# Libdrive in a nutshell

- Project in cooperation with IT Services of ETH Zürich
- Instance of polybox hosting service
  - OwnCloud based filesharing platform
- Testing with pilot customers (2020-2021)
- Integration in upload workflow during submission process in DSpace
- Softlaunch of beta version during first half of 2022



# Libdrive example in the ETH Research Collection

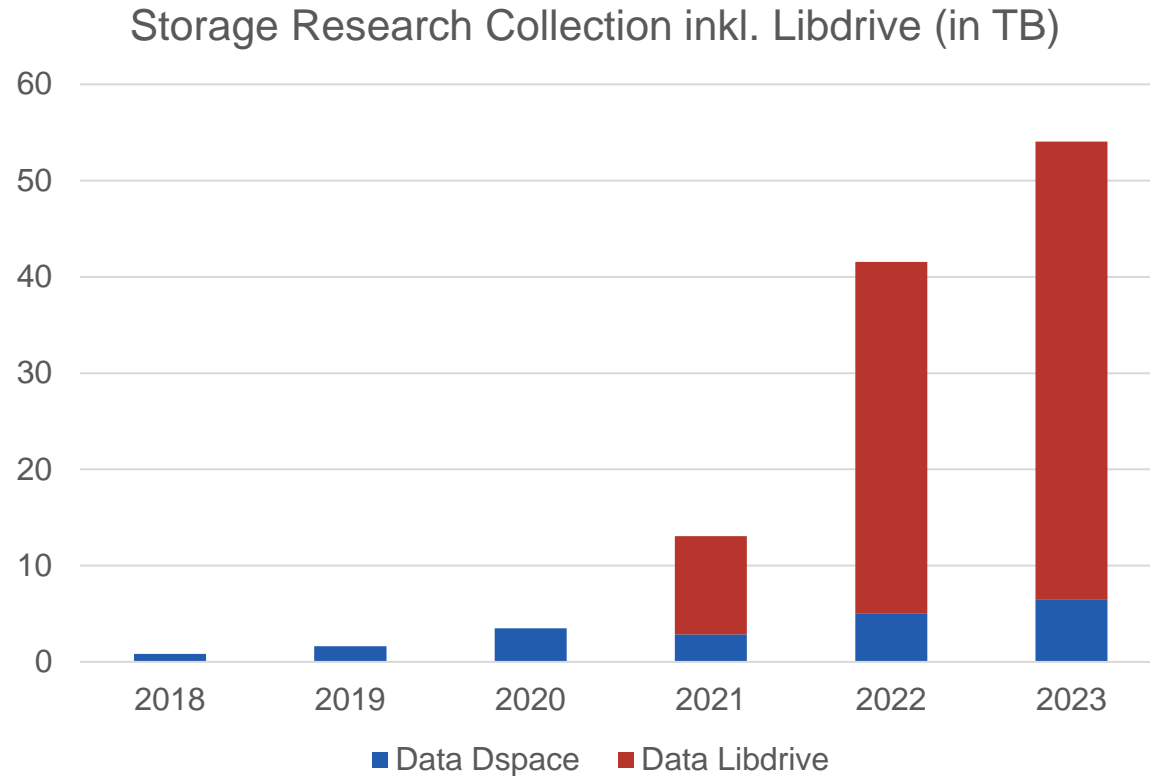
The screenshot shows the 'Research Collection' interface. The main title is 'Genome-wide RNAi screen for human 60S subunit biogenesis factors'. Below the title, there is a brief description: 'Genome-wide RNAi screen identifies novel players in human 60S subunit biogenesis including key enzymes of polyamine metabolism'. The page includes a 'Download' section with a link to 'README\_60Sscreen.pdf (PDF, 26.6Kb)'. A red box highlights the 'Data' section with the text 'Access the files'. Other sections include 'Rights / license' (Creative Commons Attribution 4.0 International), 'Abstract', 'Permanent link', 'Contributors', 'Publisher' (ETH Zurich), and 'Date collected' (2014-2019). The page also features a sidebar with navigation options like 'Browse', 'Publish', and 'Statistics'.

The screenshot shows the Libdrive interface. The top navigation bar includes the Libdrive logo, a search bar, and a dropdown menu for 'Hinzufügen libdrive.ethz.ch'. Below the navigation bar, there is a table listing folders and files. The table has columns for 'Name', 'Größe', and 'Geändert'. The folders listed are 'Ambion', 'Qiagen', and 'Validation'. The files listed include '\_README\_60Sscreen.pdf', 'Ambion\_assay well annotation.xlsx', 'Ambion\_control\_annotation.xlsx', 'doi.txt', 'Qiagen\_assaywells\_annotation.xlsx', 'Qiagen\_controls\_annotation.xlsx', and 'ValidationScreen\_assaywells\_controls\_... .xlsx'. The total size of the folders and files is 2.2 TB.

Name	Größe	Geändert
Ambion	1 TB	vor 7 Monaten
Qiagen	1.2 TB	vor 7 Monaten
Validation	43.3 GB	vor 7 Monaten
_README_60Sscreen.pdf	26 KB	vor einem Jahr
Ambion_assay well annotation.xlsx	2.6 MB	vor einem Jahr
Ambion_control_annotation.xlsx	13 KB	vor einem Jahr
doi.txt	< 1 KB	vor 10 Monaten
Qiagen_assaywells_annotation.xlsx	3.2 MB	vor einem Jahr
Qiagen_controls_annotation.xlsx	13 KB	vor einem Jahr
ValidationScreen_assaywells_controls_... .xlsx	28 KB	vor einem Jahr

<https://doi.org/10.3929/ethz-b-000528381>

# Launch, storage space and usage of Libdrive



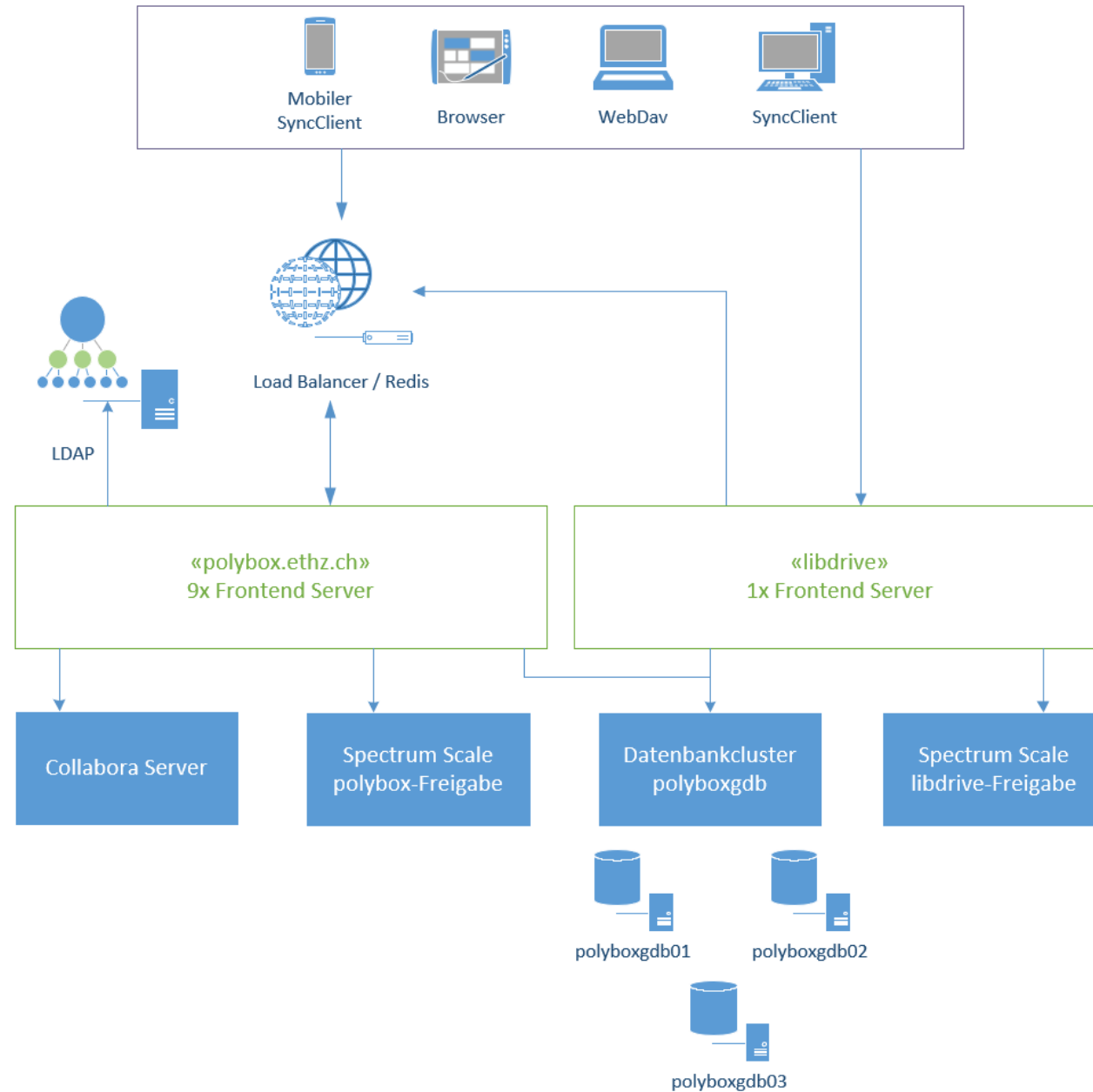
- Launch
  - Softlaunch March – August 2022
  - Communication campaign in September with Launch on 20/09/2022
- Storage space:
  - Status 08/03/2024:
    - 41.2 TB on Libdrive
    - 45 published datasets
    - 2.7 – 16 TB per dataset (Ø 1.2 TB)
- 146 TB of downloaded data in 2023

# Challenges with Libdrive

- No long-term preservation
    - Integration in ETH Data Archive (Rosetta) not possible
  - Risk of exponential data growth
    - ETH Library commits to data hosting
    - Quality of data
  - No steering via cost allocation > decision by ETH Library management
- Solution: publication on Libdrive only for 10 years. Long-term preservation delegated back to customer (using other product e.g. LTS storage)



# Libdrive infrastructure



# Lessons learned and open points

- Lessons learned
  - Customer needs of sharing unpublished data with reviewers can be fulfilled by libdrive
  - Easy to use solution
- Open points
  - Software bug causing a lot of backend traffic
  - Process a bit cumbersome: increased mail correspondence
  - Solution for long-term preservation
  - Future with free of charge storage space
  - Seamless integration, automated workflow

Andreas la Roi  
E-Publishing  
[andreas.laroi@library.ethz.ch](mailto:andreas.laroi@library.ethz.ch)

ETH Zürich  
ETH Library  
HG H31.5  
Rämistrasse 101  
8092 Zürich

[www.library.ethz.ch](http://www.library.ethz.ch)

Gianluca Caratsch  
Storage  
[gianluca.caratsch@id.ethz.ch](mailto:gianluca.caratsch@id.ethz.ch)

ETH Zürich  
ID Infrastructure  
OCT G 19  
Binzmühlestrasse 130  
8092 Zürich

[www.id.ethz.ch](http://www.id.ethz.ch)