

Keeping Research Data Safe with SWITCH

SWITCHcoldstor and beyond

SWITCH

Ann Harding / Sebastian Sigloch

CND Workshop 2019

About SWITCH



SWITCH ist **integraler Bestandteil der Schweizer Hochschulgemeinschaft.**

Basierend auf den **Kernkompetenzen**

- Network
- Security
- Identity Management

bietet SWITCH partnerschaftlich entwickelte ICT-Lösungen, die Anwender innerhalb und ausserhalb der akademischen Welt zu Spitzenleistungen im globalen Wettbewerb befähigen.

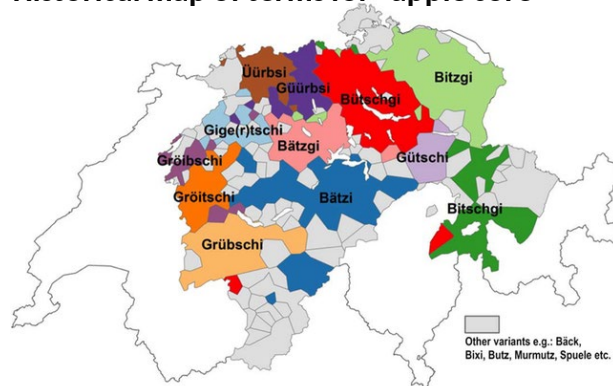
Data Growth in Research

DNA Sequencing in USB form factor



Everyone has data

Historical map of terms for "apple core"



<https://www.swissinfo.ch/eng/baeck--or--buetschgi--swiss-researcher-finds-new-ways-to-discover-dialects/45001776>

Even old data is subject to user demand, new analysis, legal/regulatory needs

Keeping Research Data Safe



Organise



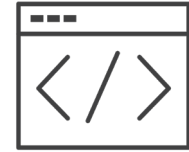
Identify



Automate



Lifecycle



Simple tools
Open formats

<https://www.explora.ethz.ch/en/s/surfing-the-sea-of-data/>

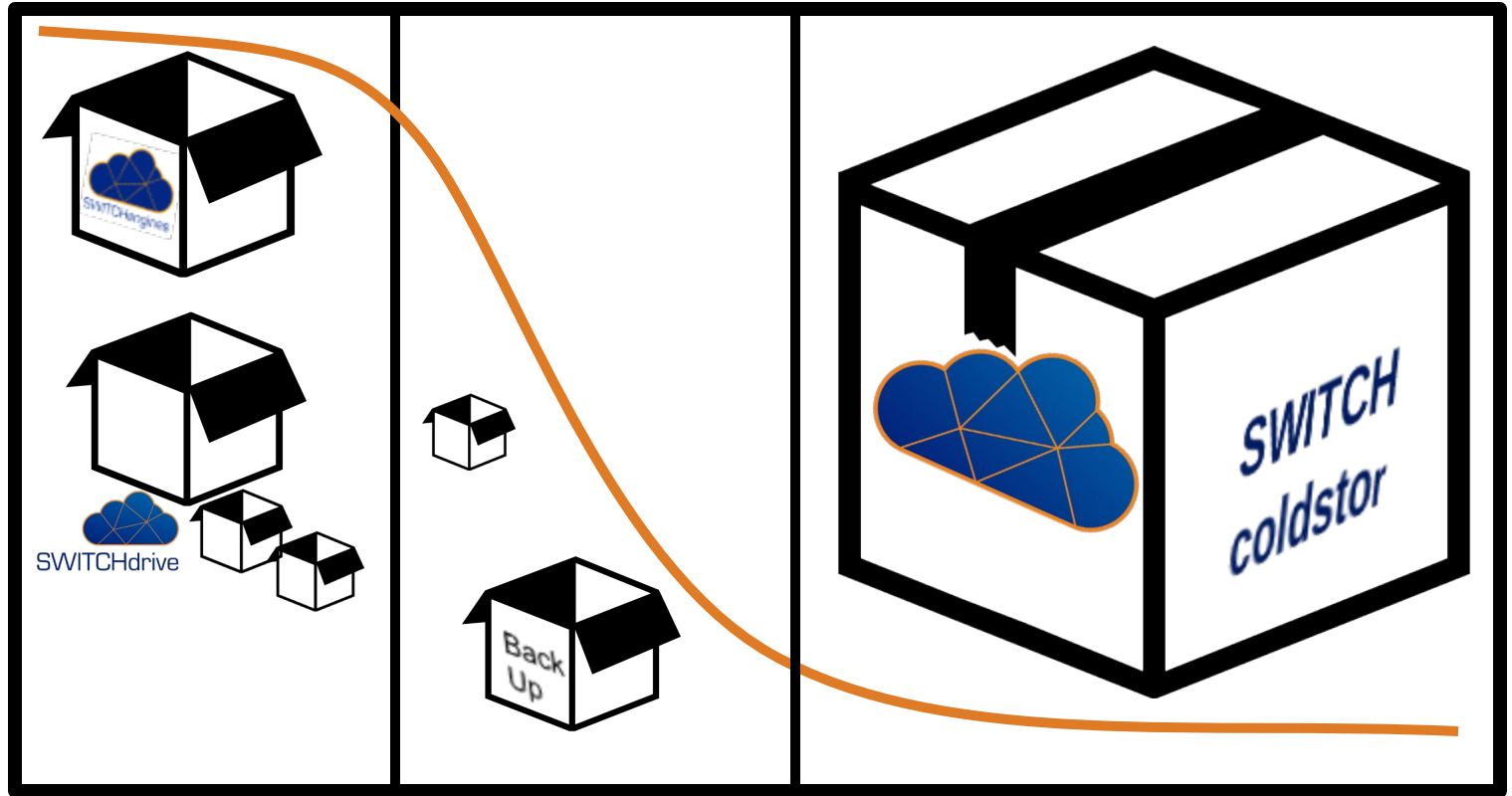
SWITCH



SWITCHcoldstor

Open Science &
Research Data
Connectome

How often do you need the data



Live

1yr

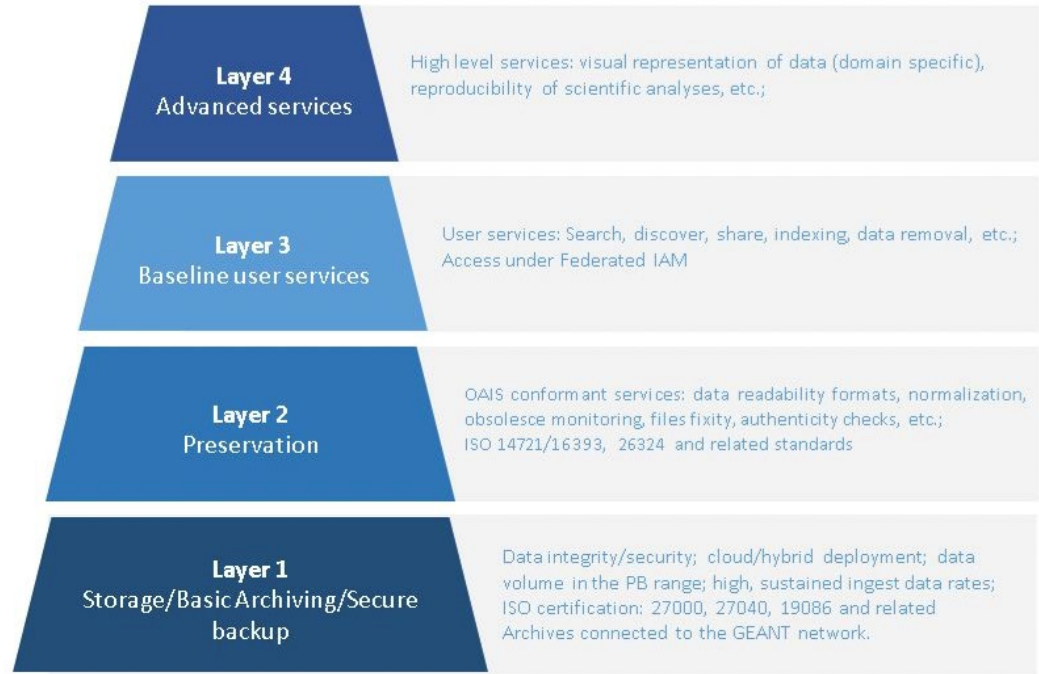
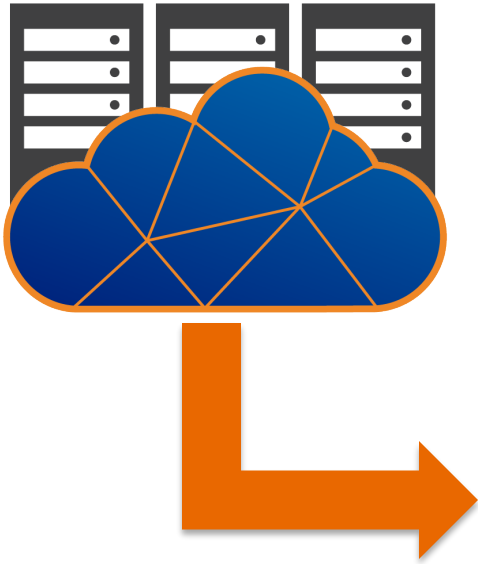
3yr

+10yr

Duration of storage

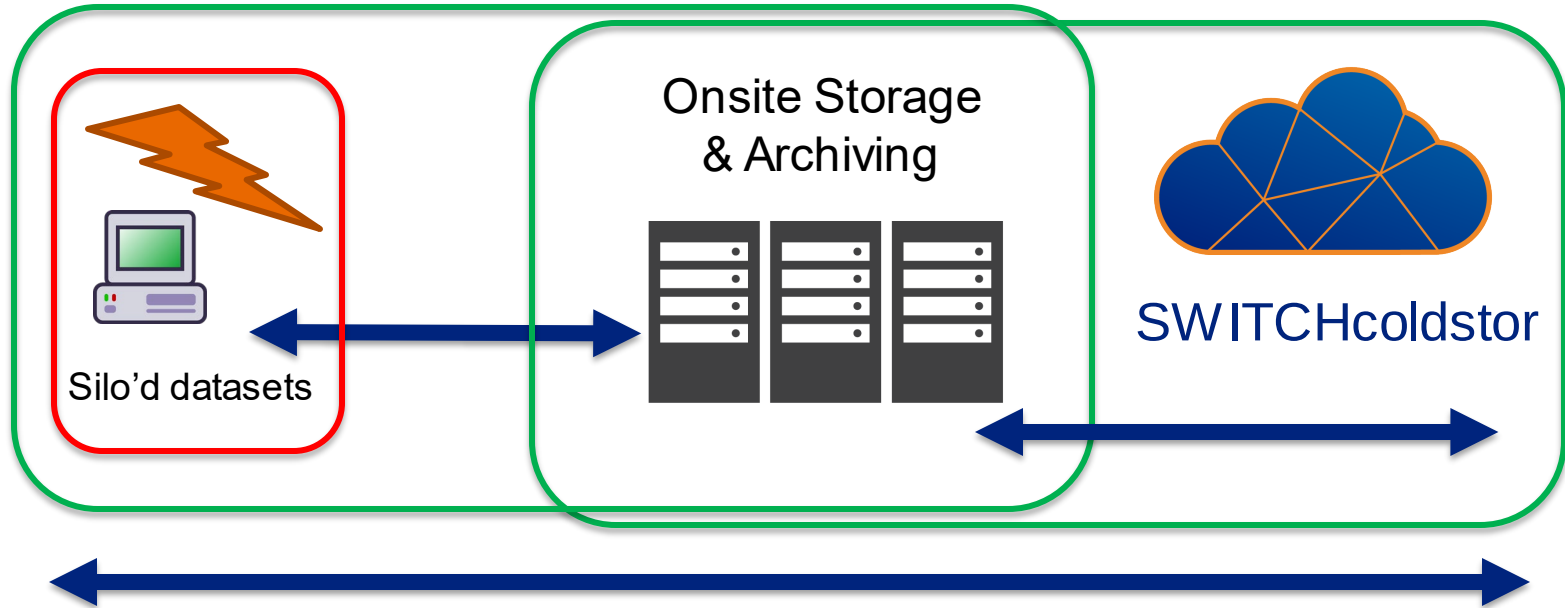
Where does long term storage fit?

Simple Tools, Open Formats



High Level Architecture of ARCHIVER R&D Challenges <https://www.archiver-project.eu>

The landscape



Overall Storage Strategy

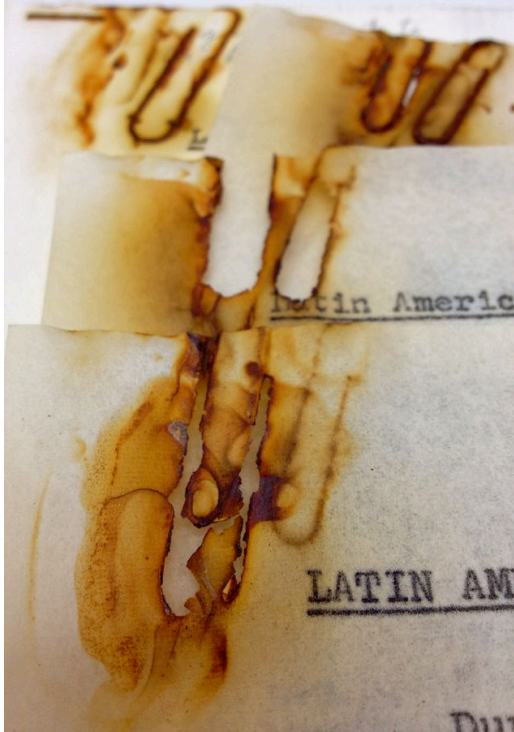
— Good solution
— High-Risk «option»

Organise, Identify, Automate

- S3 Object Storage
 - Any kind of data
 - Simple, flat structure
 - Add metadata to objects
 - Configurable access policies
 - Versioning
 - REST/HTTP
 - Critical mass - lots of tools available



Data Integrity



<https://blogs.library.duke.edu/preservation/2013/01/31/why-i-hate-mr-clippy/>

Photo data corruption; in this case, a result of a failed data recovery from a hard disk drive
https://en.wikipedia.org/wiki/Data_corruption#/media/File:Data_loss_of_image_file.JPG

Data Integrity



Data stored on DNA?










Disks are here today

For Recovery from 3 simultaneous Drive or Node Failures	
N+3	Efficiency
-	-
-	-
-	-
3+3	50%
4+3	57%
5+3	63%
6+3	67%
7+3	70%
8+3	73%
9+3	75%
10+3	77%

Erasure Coding: Data distributed across nodes, not just disks
8 data, 3 parity, recoverable from fragments



Checksum: Even if one byte of data has been altered or corrupted, checksums will not match

 	Data Integrity	Erasure coding
	High Volume	From 256 TB to 1 PB packages
	Location in Switzerland	Special region separate from SWITCHengines
	Application compatible	S3 API to integrate with your platforms
	1TB Object Store Free Trial <small>(available on request)</small>	Trial to test our solution on request
	65-85.- CHF / TB p.a.* <small>(VAT not included, tariff 2020)</small>	Unlimited writes, read for retrieval/restaging as active data

- 5 year package, available from 2020

Keeping Research Data Safe



Organise



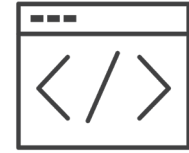
Identify



Automate



Lifecycle



Simple tools
Open formats

<https://www.explora.ethz.ch/en/s/surfing-the-sea-of-data/>

SWITCH



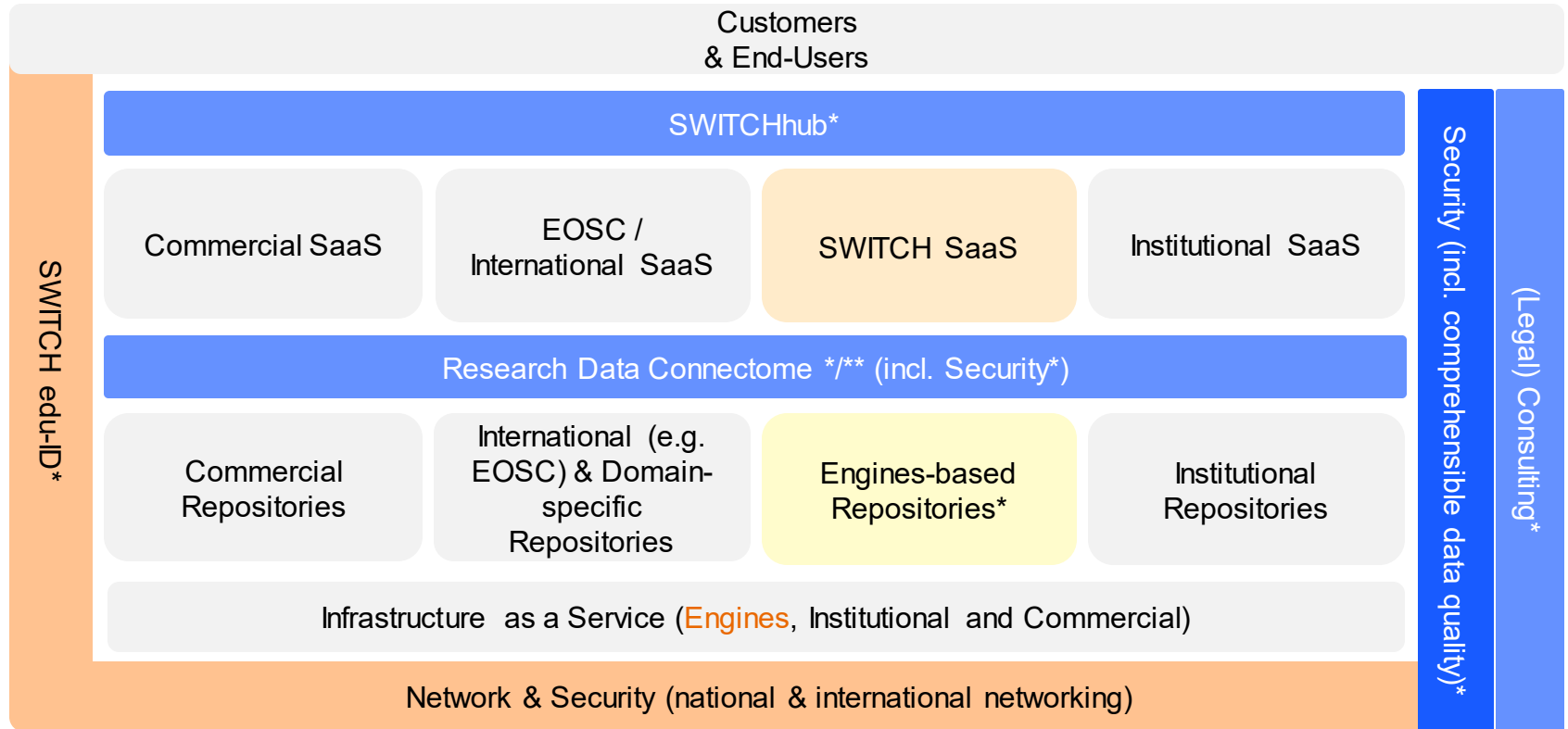
SWITCHcoldstor

Open Science &
Research Data
Connectome

Recap Methodology

- Terminology & Desk Research.
- 21 Expert Interviews (12 Institutions)
- Further Desk Research & NREN Discussions.
- Ideation and Clarification of Positioning.
- Prioritisation of ideas.
- Launch of next steps.

Positioning Open Science (Draft)



■ Established core competencies
■ no core competencies
■ Including ecosystem

■ New engagements, Activities exist
■ New engagements, no activities yet.

* Open Science Positioning
 ** SWITCH as Partner

Vision for a Research Data Connectome

“Our valuable (national) scientific information is stored decentralized and not accessible or (re-)usable for relevant actors. The Connectome organizes this research information across disciplines to make it widely accessible, interoperable and useful.” – Connectome Pilot Description

Video: <https://swit.ch/connectome-video>

Current Activities

Innovation Lab

Comprehensive
Data Quality (**active**)

satw it's all about
technology

Sprint: 09/19 – 01/20

Innovation Lab

Technologies for a
Research Data
Connectome (**active**)

**UNI
FR**
UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

xi
eXascale Infolab

Sprint: 07/19 – 11/19

Connectome
Pilot (**in planning**)*

 SAGW
ASSH

FORS 
explore.understand.share.

 Data & Service
Center for
the Humanities
DasCH

satw it's all about
technology
SDSC

EPFL

M

**UNI
FR** **xi**
eXascale Infolab

 CERN

zenodo

USI/SUPSI
IDSTIA

Start: 01/2020

* Innovation Project

Contact

Innovation@SWITCH:

Dr Sebastian Sigloch

sebastian.Sigloch@switch.ch

Research Data Connectome:

connectome@switch.ch