



MODAMA

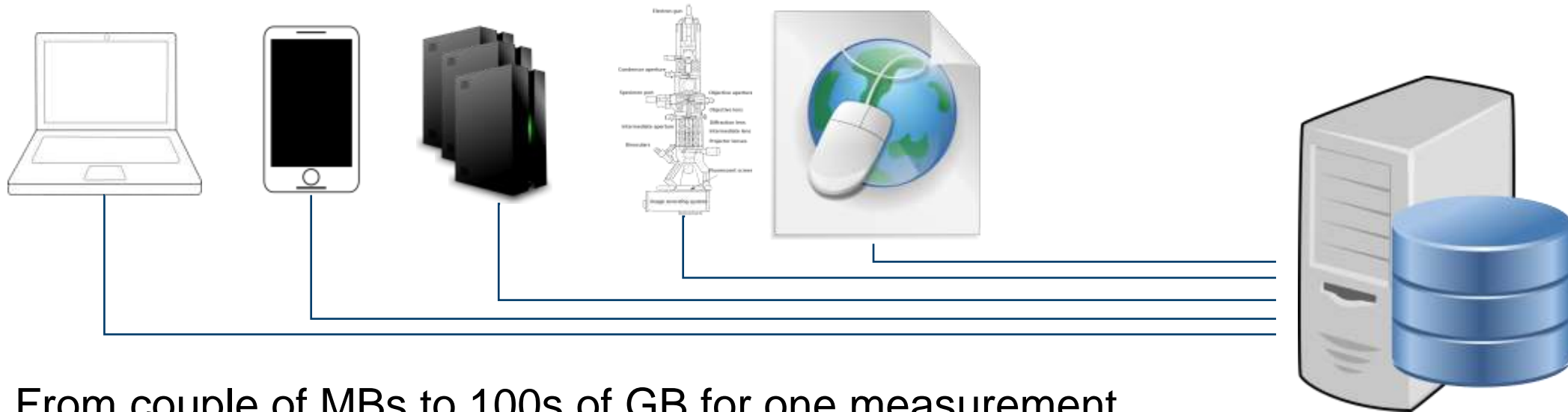
Modular Data Management System at ER-C

- 2025-08-26 | DIETER WEBER, ALEXANDER CLAUSEN, RITWIK SHANKER

DATA AT ER-C: COMPLEX USE CASE

Only partially solved in other set-ups

- People oversee their own data, different from other large facilities.
- External people participate, this is at the heart of a user center like ER-C!
- Combination of mobile devices, HPC, web applications



- From couple of MBs to 100s of GB for one measurement
- Very high throughput: GB/s processing speed even on ultrabook laptops!



<https://openclipart.org/>
https://commons.wikimedia.org/wiki/File:Scheme_TEM_en.svg

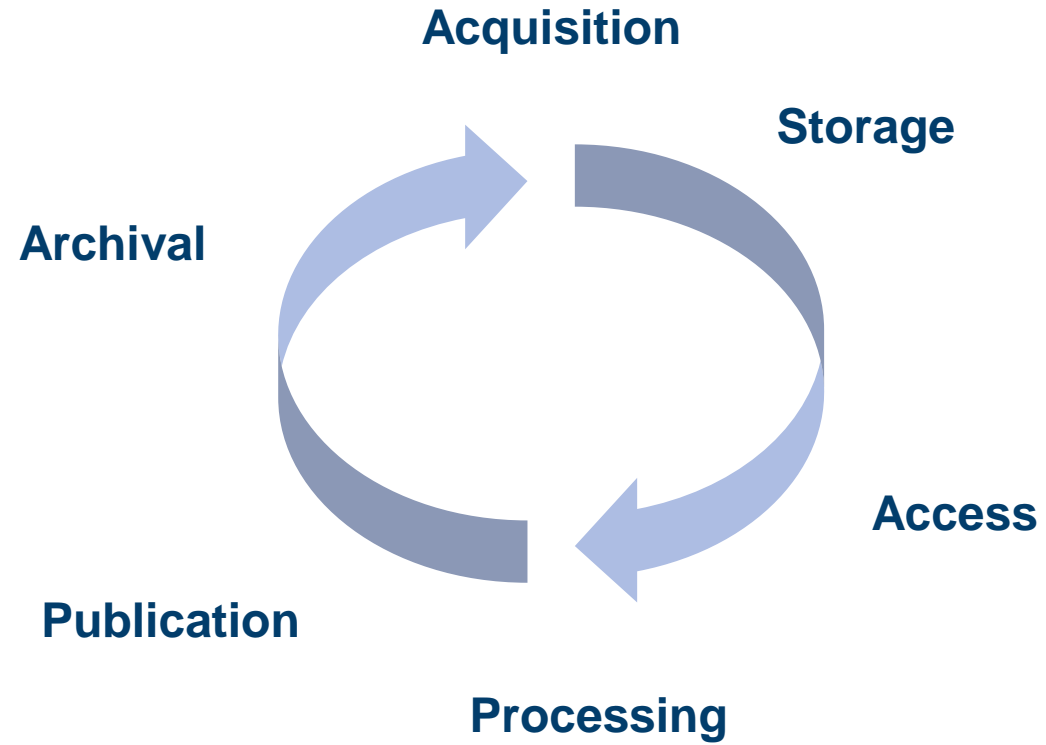
USER BASE AND APPLICATIONS

- Very diverse, incl short term visitors
 - Quick onboarding!
- From IT professionals to scientist without any IT background
- Full spectrum of EM methods and applications
- Method development: One-of-a-kind tools and projects
- Integrate with instruments and their peculiarities
 - TEMs have complex and individual IT systems, usually legacy systems
- Base line: external hard drives
- NAS and Nextcloud instances as well

Goals

- Seamless sharing and collaboration
- Support “big data” projects
- Data protection, Good Scientific Practice
- FAIR data (Findable, Accessible, Interoperable, Reusable)
- Support scientists, don’t get in their way
- Smooth transition

DATA LIFE CYCLE








DESIGN OF MODAMA

Combine NAS with Nextcloud, additional services

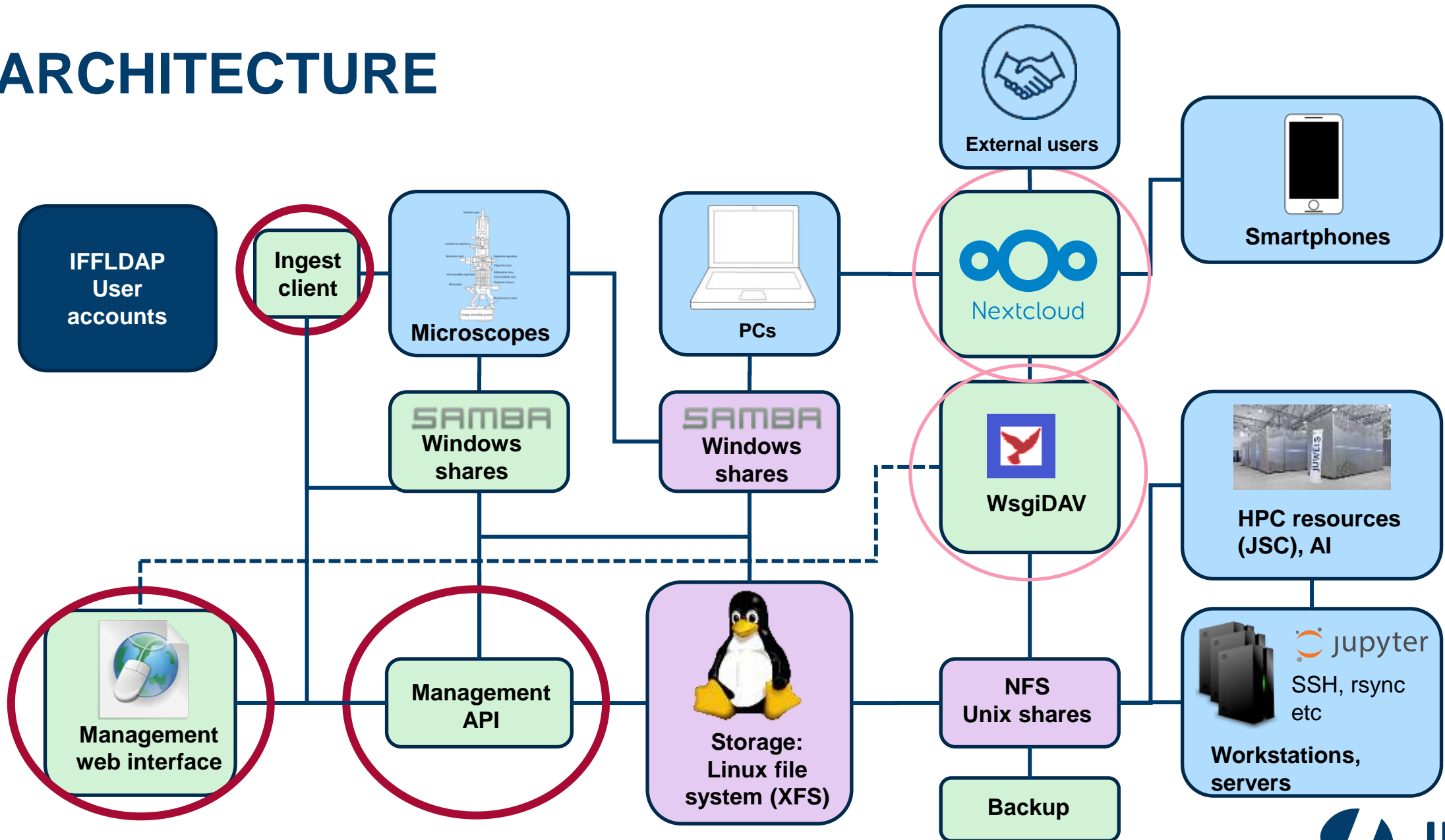
<https://er-c-data.fz-juelich.de/>

- Central file system
- Exported via CIFS (Samba), NFS and Nextcloud
- Available on shared workstations
- Self-service web interface to manage projects
- Web-based browse interface to link with other services
 - Electronic Lab Notebooks (ELNs), data viewers, Jupyter notebooks, ...

LANDING PAGE

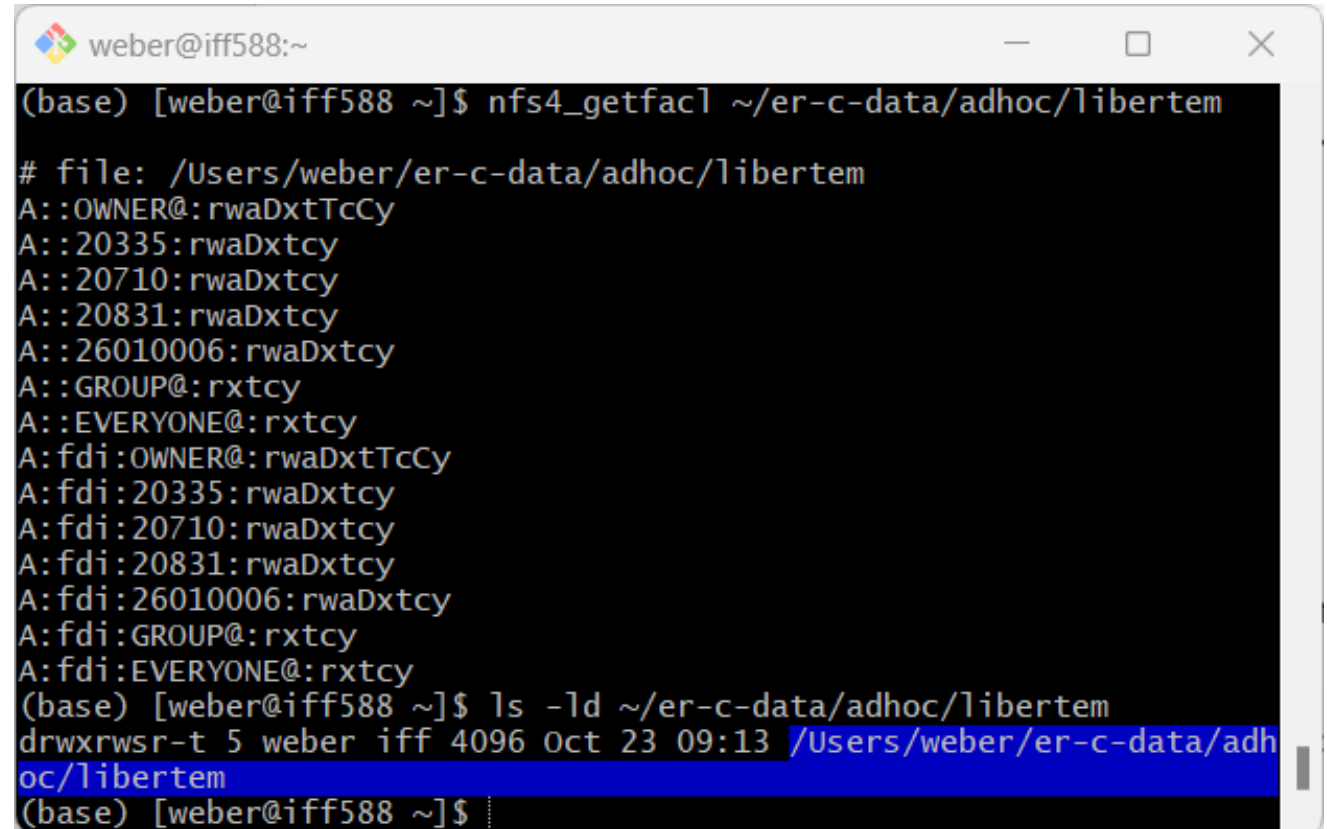
 Management Manage projects, sessions, ingest and access permissions (VPN/Intranet) MANAGEMENT INTERFACE	Nextcloud Synchronize, collaborate, ingest data via Nextcloud USE NEXTCLOUD	 Network Filesystem Directly access files from your favourite programs via network filesystems (VPN/Intranet) DETAILS
 Jupyter Notebooks Analyze and process data with Jupyter notebooks (VPN/Intranet) USE JUPYTER	File Transfer Use standard tools like scp or rsync to transfer data in or out TRANSFER FILES	 Ingest Data Use our Ingest GUI to store your session's data from the microscopes DETAILS
Synchronize locally Install and connect a native Nextcloud client INSTALL NEXTCLOUD	Large datasets Transfer large folders or files DETAILS	 SampleDB Learn how to access our electronic lab notebook an instance of SampleDB DETAILS

ARCHITECTURE



CORE: LINUX FILE SYSTEM

- Full range of FS features available
 - Direct I/O, random access, FS cache, memory mapping, ...
- Easier to export FS content as blobs than vice-versa
- File system permissions, ACLs, immutable flag, inotify to create desired behavior
- “Truth is in the file system”



```
weber@iff588:~  
(base) [weber@iff588 ~]$ nfs4_getfacl ~/er-c-data/adhoc/libertem  
# file: /Users/weber/er-c-data/adhoc/libertem  
A::OWNER@:rwaDxtTcCy  
A::20335:rwaDxtcy  
A::20710:rwaDxtcy  
A::20831:rwaDxtcy  
A::26010006:rwaDxtcy  
A::GROUP@:rxtcy  
A::EVERYONE@:rxtcy  
A:fdi:OWNER@:rwaDxtTcCy  
A:fdi:20335:rwaDxtcy  
A:fdi:20710:rwaDxtcy  
A:fdi:20831:rwaDxtcy  
A:fdi:26010006:rwaDxtcy  
A:fdi:GROUP@:rxtcy  
A:fdi:EVERYONE@:rxtcy  
(base) [weber@iff588 ~]$ ls -ld ~/er-c-data/adhoc/libertem  
drwxrwsr-t 5 weber iff 4096 Oct 23 09:13 /Users/weber/er-c-data/adh  
oc/libertem  
(base) [weber@iff588 ~]$
```

The management interface - create and update projects

oslo_ingrid	2022-06-16 18:23	mann	ljor mann	/adhoc/oslo_ingrid	
better_name	2022-03-24 12:12	lipinska	lipinska	/adhoc/better_name	
test	2022-03-23 16:30	4564654	jheinen markus kaiser 4564654	/adhoc/test	
ksca_2021	2022-03-23 15:22	barthel	barthel		

CREATE PROJECT...

Create Project

Projects that are created by users are directories in the /adhoc folder. The project owners and members can create files and folders everywhere within a project. Only file owners can delete files. Under the hood, project membership and write access for user-created projects is managed by the system.

name

fancy-new-method

members

✕ weber

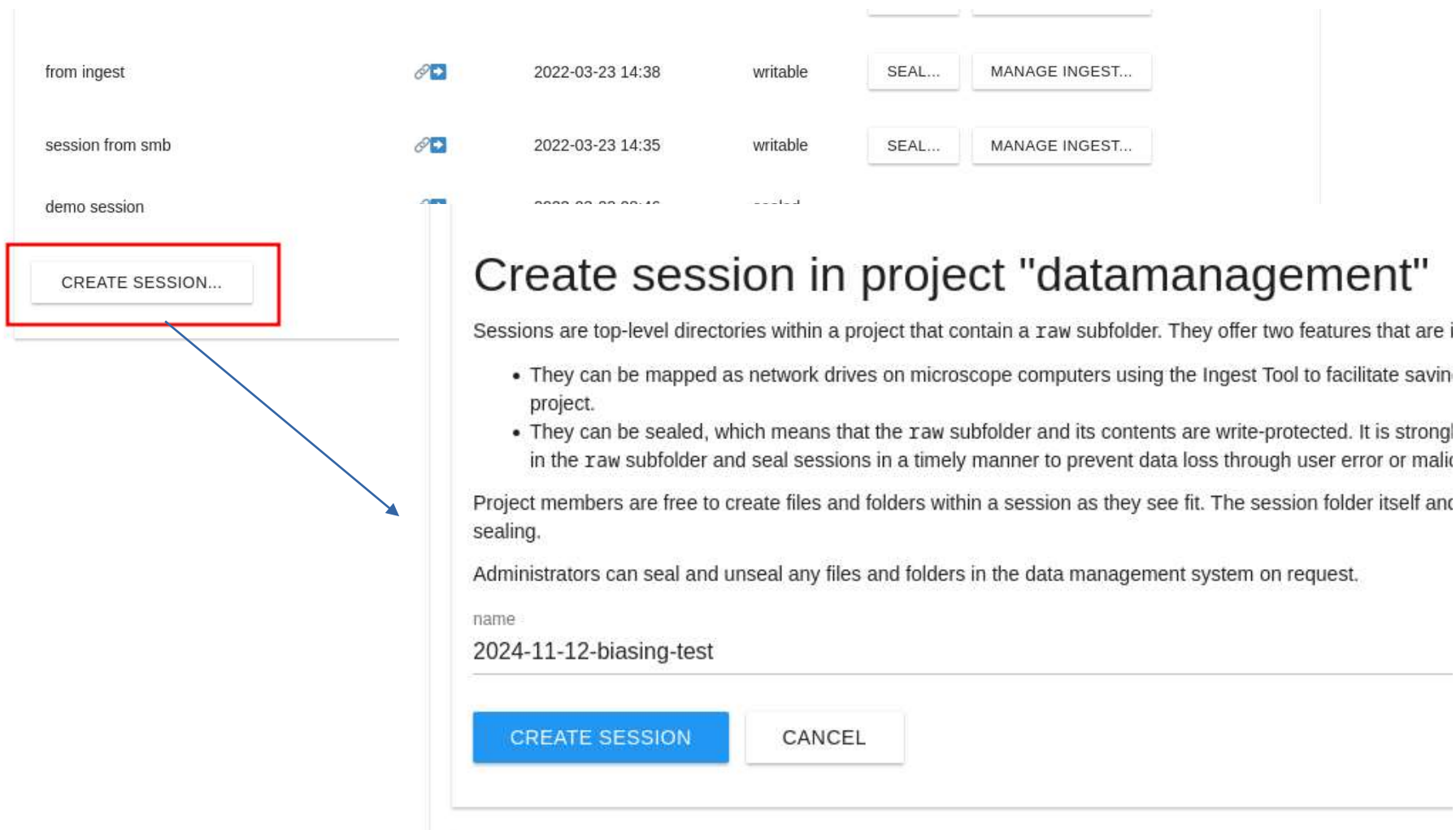
jvas

jvas

CREATE PROJECT

CANCEL

The management interface – create session



The screenshot displays a web-based management interface. On the left, a sidebar contains a list of sessions: 'from ingest', 'session from smb', and 'demo session'. Below this list is a button labeled 'CREATE SESSION...'. A red rectangular box highlights this button, and a blue arrow points from it towards the right-hand modal window.

The modal window, titled 'Create session in project "datamanagement"', provides detailed information about sessions. It states: 'Sessions are top-level directories within a project that contain a raw subfolder. They offer two features that are i'.

- They can be mapped as network drives on microscope computers using the Ingest Tool to facilitate saving project.
- They can be sealed, which means that the raw subfolder and its contents are write-protected. It is strongly recommended to seal sessions in a timely manner to prevent data loss through user error or malicious activity.

Project members are free to create files and folders within a session as they see fit. The session folder itself and its contents are sealed.

Administrators can seal and unseal any files and folders in the data management system on request.

At the bottom of the modal, there is a text input field labeled 'name' containing the text '2024-11-12-biasing-test'. Below the input field are two buttons: 'CREATE SESSION' (in blue) and 'CANCEL' (in white).

Session Name	Created At	Status	Actions
from ingest	2022-03-23 14:38	writable	SEAL... MANAGE INGEST...
session from smb	2022-03-23 14:35	writable	SEAL... MANAGE INGEST...
demo session	2022-03-23 14:35	sealed	

The management interface – overall folder structure

- /storage/er-c-data/adhoc/<project>/<session>/
- Session contains a “raw” subdirectory
 - This should contain the raw experimental/simulation data
 - It can be sealed to write-protect it
 - Good for integrity: can’t (accidentally) change data!

The management interface – seal session

Sessions

Name	Browse	Modified	Status	
2024-11-12-biasing-test		2024-11-12 17:33	writable	<div><div>SEAL...</div><div>MANAGE INGEST...</div></div>

[ER-C-Data](#) / [Projects](#) / [datamanagement](#) / Seal session

Confirm seal session "2024-11-12-biasing-test" in project "datamanagement"

Sealing a session sets the raw subfolder of a session and all its contents immutable. Administrators can seal and unseal any files or folders on request.

SEAL SESSION '2024-11-12-BIASING-TEST'

CANCEL

The management interface – browse and preview

ER-C-Data / Projects / skyrmion-creation-time-resolved

Project "skyrmion-creation-time-resolved"

Owner

clausen

Created

2024-11-06 09:39

Modified

2024-11-06 09:39

Size

42.9 GiB

Members



zingsem, clausen, valuedgatancustomer, jvas

MANAGE PROJECT MEMBERS...

Browse

/adhoc/skyrmion-creation-time-resolved

Sessions

Name	Browse	Modified	Status
2024-11-06-second-test		2024-11-06 17:55	sealed
2024-11-04-first-test		2024-11-04 16:11	sealed

CREATE SESSION...

The management interface – browse and preview

ER-C-Data / Management / Browse

skyrmion-creation-time-resolved

Project

skyrmion-creation-time-resolved

Created

2024-11-06T09:39:07Z

Modified

Wed, 06 Nov 2024 09:39:07 GMT

Entries

2

Access this file or folder

Nextcloud

Access folder with Nextcloud

BROWSE IN NEXTCLOUD

JupyterHub

Access folder in JupyterHub

OPEN JUPYTERHUB

LiberTEM

Open file or folder with LiberTEM

OPEN IN LIBERTEM

SampleDB

Search in SampleDB

SEARCH IN SAMPLEDB

Rsync push to er-c-data with optimized settings:

rsync -av --info=progress2 -e "ssh" -T -c aes128-ctr -o Compression=no -x" "<src_path>" clausen@moellenstedt.iff.kfa-juelich.de:/storage/er-c-data/adhoc/skyrmion-creation-time-resolved

SFTP with optimized settings:

sftp -c aes128-ctr -o Compression=no clausen@moellenstedt.iff.kfa-juelich.de:/storage/er-c-data/adhoc/skyrmion-creation-time-resolved

SMB (Windows):

\\er-c-data-smb.fz-juelich.de/er-c-data/adhoc/skyrmion-creation-time-resolved

SMB (Mac, Linux) — connect:

smb://er-c-data-smb.fz-juelich.de/er-c-data/adhoc/skyrmion-creation-time-resolved





























Browse

/ er-c-data / adhoc / skyrmion-creation-time-resolved

Name	Nextcloud	JupyterHub	LiberTEM	View	SampleDB
------	-----------	------------	----------	------	----------

The management interface – browse and preview

/ er-c-data / [adhoc](#) / [skyrmion-creation-time-resolved](#) / [2024-11-06-second-test](#) / [raw](#) / [measurement-06](#)

Name	Nextcloud	JupyterHub	LiberTEM	View	SampleDB
...					
 frame-00000.dm4					
 frame-00001.dm4					
 frame-00002.dm4					
 frame-00003.dm4					
 frame-00004.dm4					
 frame-00005.dm4					

The management interface – browse and preview

[/ er-c-data /](#) [adhoc /](#) [skyrmion-creation-time-resolved /](#) [2024-11-06-second-test /](#) [raw /](#) [measurement-06](#) [🔗](#)

Name	Nextcloud	JupyterHub	LiberTEM	View	SampleDB
📁 ..					
📄 frame-00000.dm4	🔗		🔗➡	🔗➡	🔗➡
📄 frame-00001.dm4	🔗		🔗➡	🔗➡	🔗➡
📄 frame-00002.dm4	🔗		🔗➡	🔗➡	🔗➡
📄 frame-00003.dm4	🔗		🔗➡	🔗➡	🔗➡
📄 frame-00004.dm4	🔗		🔗➡	🔗➡	🔗➡
📄 frame-00005.dm4	🔗		🔗➡	🔗➡	🔗➡

The management interface – browse and preview

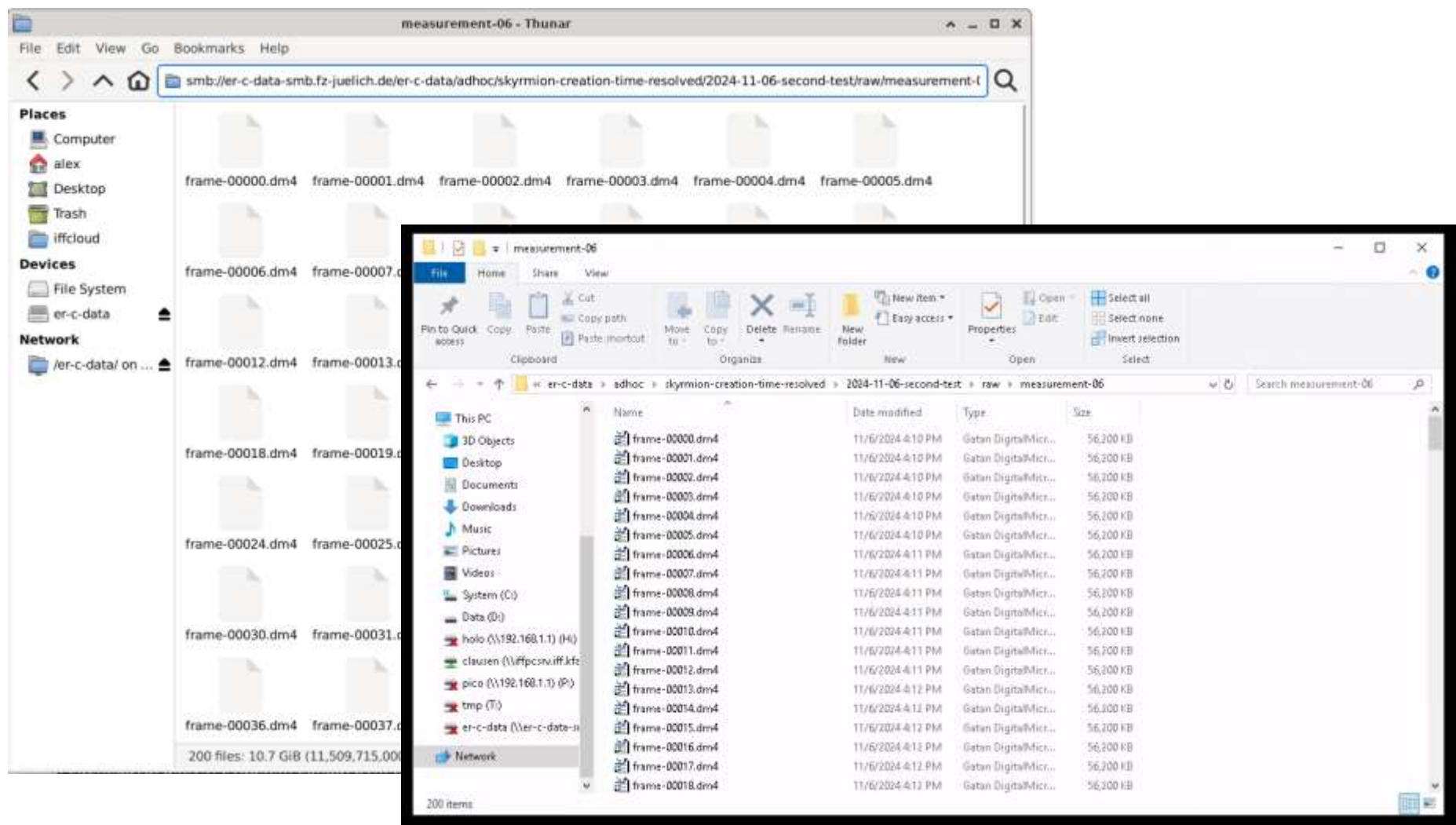
Image Viewer

File: /storage/er-c-data/adhoc/skymion-creation-time-resolved/2024-11-06-second-test/raw/measurement-06/frame-00005.dm4



VPN/internal only!

Data access - Windows Network Drive (Win/Mac/Linux)



Data access – Nextcloud (sync and **share**; external read/**write** access)

<https://er-c-data.fz-juelich.de/nextcloud/>

The screenshot shows the Nextcloud web interface in a browser window. The address bar displays the URL <https://er-c-data.fz-juelich.de/nextcloud/apps/files/?dir=/ER-C-Data/adhoc/raw/measurement-06>. The interface includes a left sidebar with navigation options: All files, Recent, Favorites, Shares, Tags, External storage, Deleted files, and a storage usage indicator (0 B of 500 MB used). The main content area shows a breadcrumb path: ER-C-Data > adhoc > raw > measurement-06. Below the path is a table of files:

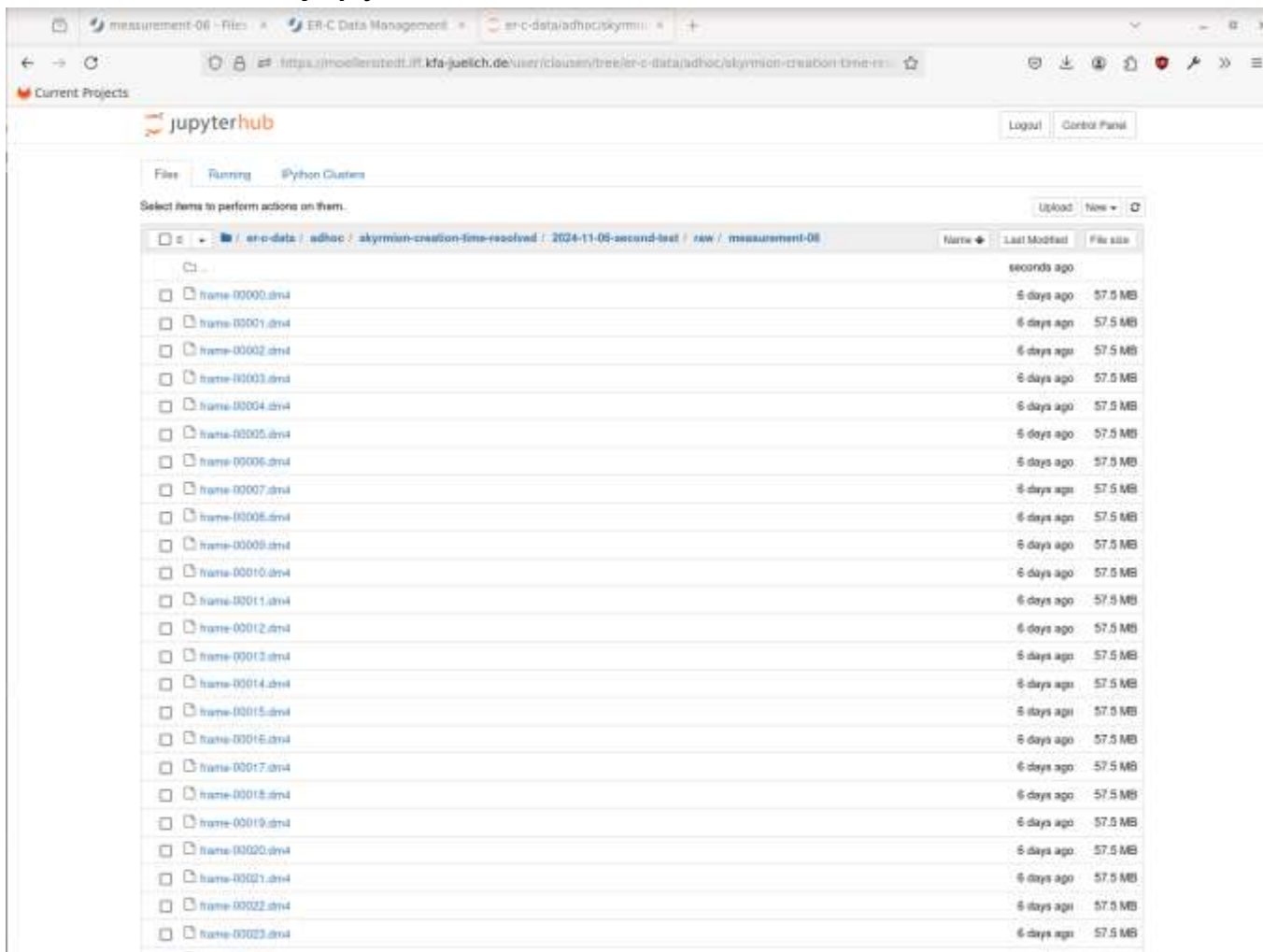
Name	Size	Modified
frame-00000.dm4	54.9 MB	6 days ago
frame-00001.dm4	54.9 MB	6 days ago
frame-00002.dm4	54.9 MB	6 days ago
frame-00003.dm4	54.9 MB	6 days ago
frame-00004.dm4	54.9 MB	6 days ago
frame-00005.dm4	54.9 MB	6 days ago
frame-00006.dm4	54.9 MB	6 days ago
frame-00007.dm4	54.9 MB	6 days ago

Data access – data processing workstations

```
Terminal - clausen@iff588:/storage/er-c-data/adhoc/skyrmion-creation-time-resolved/2024-11-06-second-test/raw/measurement-06
(base) clausen@iff588:~$ cd /storage/er-c-data/adhoc/skyrmion-creation-time-resolved/2024-11-06-second-test/raw/measurement-06
(base) clausen@iff588:/storage/er-c-data/adhoc/skyrmion-creation-time-resolved/2024-11-06-second-test/raw/measurement-06$ ls
frame-00000.dm4 frame-00025.dm4 frame-00050.dm4 frame-00075.dm4 frame-00100.dm4 frame-00125.dm4 frame-00150.dm4 frame-00175.dm4
frame-00001.dm4 frame-00026.dm4 frame-00051.dm4 frame-00076.dm4 frame-00101.dm4 frame-00126.dm4 frame-00151.dm4 frame-00176.dm4
frame-00002.dm4 frame-00027.dm4 frame-00052.dm4 frame-00077.dm4 frame-00102.dm4 frame-00127.dm4 frame-00152.dm4 frame-00177.dm4
frame-00003.dm4 frame-00028.dm4 frame-00053.dm4 frame-00078.dm4 frame-00103.dm4 frame-00128.dm4 frame-00153.dm4 frame-00178.dm4
frame-00004.dm4 frame-00029.dm4 frame-00054.dm4 frame-00079.dm4 frame-00104.dm4 frame-00129.dm4 frame-00154.dm4 frame-00179.dm4
frame-00005.dm4 frame-00030.dm4 frame-00055.dm4 frame-00080.dm4 frame-00105.dm4 frame-00130.dm4 frame-00155.dm4 frame-00180.dm4
frame-00006.dm4 frame-00031.dm4 frame-00056.dm4 frame-00081.dm4 frame-00106.dm4 frame-00131.dm4 frame-00156.dm4 frame-00181.dm4
frame-00007.dm4 frame-00032.dm4 frame-00057.dm4 frame-00082.dm4 frame-00107.dm4 frame-00132.dm4 frame-00157.dm4 frame-00182.dm4
frame-00008.dm4 frame-00033.dm4 frame-00058.dm4 frame-00083.dm4 frame-00108.dm4 frame-00133.dm4 frame-00158.dm4 frame-00183.dm4
frame-00009.dm4 frame-00034.dm4 frame-00059.dm4 frame-00084.dm4 frame-00109.dm4 frame-00134.dm4 frame-00159.dm4 frame-00184.dm4
frame-00010.dm4 frame-00035.dm4 frame-00060.dm4 frame-00085.dm4 frame-00110.dm4 frame-00135.dm4 frame-00160.dm4 frame-00185.dm4
frame-00011.dm4 frame-00036.dm4 frame-00061.dm4 frame-00086.dm4 frame-00111.dm4 frame-00136.dm4 frame-00161.dm4 frame-00186.dm4
frame-00012.dm4 frame-00037.dm4 frame-00062.dm4 frame-00087.dm4 frame-00112.dm4 frame-00137.dm4 frame-00162.dm4 frame-00187.dm4
frame-00013.dm4 frame-00038.dm4 frame-00063.dm4 frame-00088.dm4 frame-00113.dm4 frame-00138.dm4 frame-00163.dm4 frame-00188.dm4
frame-00014.dm4 frame-00039.dm4 frame-00064.dm4 frame-00089.dm4 frame-00114.dm4 frame-00139.dm4 frame-00164.dm4 frame-00189.dm4
frame-00015.dm4 frame-00040.dm4 frame-00065.dm4 frame-00090.dm4 frame-00115.dm4 frame-00140.dm4 frame-00165.dm4 frame-00190.dm4
frame-00016.dm4 frame-00041.dm4 frame-00066.dm4 frame-00091.dm4 frame-00116.dm4 frame-00141.dm4 frame-00166.dm4 frame-00191.dm4
frame-00017.dm4 frame-00042.dm4 frame-00067.dm4 frame-00092.dm4 frame-00117.dm4 frame-00142.dm4 frame-00167.dm4 frame-00192.dm4
frame-00018.dm4 frame-00043.dm4 frame-00068.dm4 frame-00093.dm4 frame-00118.dm4 frame-00143.dm4 frame-00168.dm4 frame-00193.dm4
frame-00019.dm4 frame-00044.dm4 frame-00069.dm4 frame-00094.dm4 frame-00119.dm4 frame-00144.dm4 frame-00169.dm4 frame-00194.dm4
frame-00020.dm4 frame-00045.dm4 frame-00070.dm4 frame-00095.dm4 frame-00120.dm4 frame-00145.dm4 frame-00170.dm4 frame-00195.dm4
frame-00021.dm4 frame-00046.dm4 frame-00071.dm4 frame-00096.dm4 frame-00121.dm4 frame-00146.dm4 frame-00171.dm4 frame-00196.dm4
frame-00022.dm4 frame-00047.dm4 frame-00072.dm4 frame-00097.dm4 frame-00122.dm4 frame-00147.dm4 frame-00172.dm4 frame-00197.dm4
frame-00023.dm4 frame-00048.dm4 frame-00073.dm4 frame-00098.dm4 frame-00123.dm4 frame-00148.dm4 frame-00173.dm4 frame-00198.dm4
frame-00024.dm4 frame-00049.dm4 frame-00074.dm4 frame-00099.dm4 frame-00124.dm4 frame-00149.dm4 frame-00174.dm4 frame-00199.dm4
(base) clausen@iff588:/storage/er-c-data/adhoc/skyrmion-creation-time-resolved/2024-11-06-second-test/raw/measurement-06$
```

Standard location: /storage/er-c-data/

Data access – jupyterhub



The screenshot shows the JupyterHub web interface. At the top, there's a browser address bar with the URL <https://moellenstedt.iff.kfa-juelich.de/user/clauser/tree/er-c-data/adhoc/skymion-creation-time-resolved-2024-11-05-second-test/raw/measurement-06>. Below the browser, the JupyterHub logo is visible, along with 'Logout' and 'Control Panel' buttons. The main interface has tabs for 'Files', 'Running', and 'Python Clusters'. The 'Files' tab is active, showing a file browser view. The path in the breadcrumb is `/ er-c-data / adhoc / skymion-creation-time-resolved : 2024-11-05-second-test / raw / measurement-06`. The file list shows a directory containing many files, all named `frame-00000.dcm4` through `frame-00023.dcm4`. Each file is 57.5 MB and was last modified '6 days ago'. The interface also includes a 'Select items to perform actions on them.' prompt and an 'Upload' button.

<https://moellenstedt.iff.kfa-juelich.de/> (VPN/internal only)

DISCUSSION

- Simple and flexible – make a lot of data a bit more FAIR
 - Familiar and mature components, standard interfaces
 - POSIX file system, NFS, CIFS, WebDAV, Nextcloud
 - Easy onboarding – most users familiar with network drives or Nextcloud
 - Modular design
 - Robust – core keeps running even if components have issues
 - Extensible without central coordination
 - Flexible solutions via diverse interfaces
- Don't create a complete system, but a starting point and conditions for it to grow
- Positive user feedback, steady use

CURRENT DEVELOPMENTS

- Member of EU-wide EOSC Data Commons project

- <https://www.eosc-data-commons.eu/>



- Ingest and integrations
- Reorganize storage
 - Likely Ceph cluster, remote block device
- Improve interface and services
 - Web interface overhaul
 - Private projects
- Integrate ER-C-3 (CryoEM)
- Continue growth

ACKNOWLEDGMENTS

- IT group of PGI/JCNS
 - Markus Consoir, Florian Rhiem, Josef Heinen
- People at ER-C
 - Alexander Müller, Juri Barthel, Daniel Mann, Thomas Heidler
- External partners
 - Steffen Brinckmann and team (IEK-2)
 - Matthew Bryan (CEA-LETI, AIDAS)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823717 (ESTEEM3)



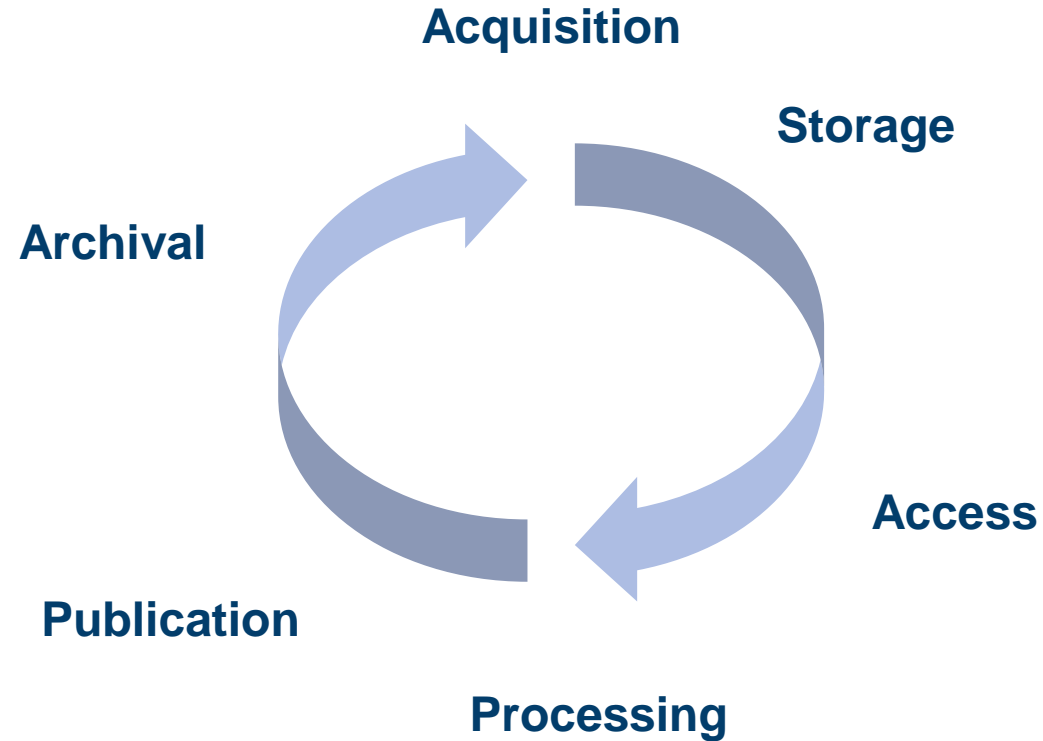
Cooperation within framework of AIDAS



This project has received funding from the European Union's Horizon Europe Research infrastructures programme under grant agreement No 101188179 (EOSC Data Commons)

REQUIREMENTS DATA MANAGEMENT

Acquisition – Storage – Access – Processing – Publication – Archival



REQUIREMENTS DATA MANAGEMENT

Acquisition – Storage – Access – Processing – Publication – Archival

Absorb the data coming from the detector and store it somewhere

- Usually provided by detector vendors in some form
- Some form of data access with sufficient speed

Currently: Stored on instrument systems

WIP: Store directly in data management

REQUIREMENTS DATA MANAGEMENT

Acquisition – Storage – Access – Processing – Publication – Archival

Inspect data quickly during session

- Decide next steps of experiment
- Possibly live analysis

Currently: Install inspection software locally (→ LiberTEM project)

WIP: Use web services

REQUIREMENTS DATA MANAGEMENT

Acquisition – Storage – Access – Processing – Publication – Archival

Keep notes, meta-information linked to raw data files

- Create a stable reference: File paths or other identifiers
- Relevant data scattered on multiple systems!
 - Microscope, spectrometer, camera, in-situ control, ...
- Easily create new projects – avoid later data movement or catch-all megaprojects
- Move data into management system ASAP
- Make available on microscope systems (WIP)
- Link to raw data via landing page URL e.g. in ELNs, presentations, GitHub issues, ...

REQUIREMENTS DATA MANAGEMENT

Acquisition – **Storage** – Access – Processing – Publication – Archival

Transfer to permanent storage location

- During or at end of session
 - Correct ownership, provenance tracking
 - Microscopes usually run under system accounts!
 - Keep connection to notes, other meta-information
-
- Ingest: WIP (later discussion if interested)
 - Also web-based ingest with Nextcloud
 - Fast network access: WIP
 - Allows storing all kinds of files, incl. notes etc. – not limited to microscopy datasets

REQUIREMENTS DATA MANAGEMENT

Acquisition – **Storage** – Access – Processing – Publication – Archival

Prevent data loss

- Main risk are you and I – ask me how I know 🤖
 - Write protection of source data
 - Keep associated mutable user data close
 - Analysis scripts, figures, paper drafts, ...
 - Replication, backup, snapshots
-
- Incremental daily backup of the central file system
 - Allow making data immutable in the file system – not overridable on clients
 - Restrict write access to own files

REQUIREMENTS DATA MANAGEMENT

Acquisition – Storage – **Access** – Processing – Publication – Archival

Collaboration, sharing, exchange

- Selective access control read/write
 - Internal access for team members
 - Includes collaboration on projects
 - Sharing with external partners
 - Microscope systems
 - Shared workstations
 - PCs
-
- Standard protocols
 - Wide range of access pathways
 - External sharing via Nextcloud
- Mobile devices
 - Services
 - External storage media
 - Which software, OS, protocols?

REQUIREMENTS DATA MANAGEMENT

Acquisition – Storage – Access – **Processing** – Publication – Archival

Shared compute systems

- High performance computers with fast data access
 - Computation close to the data
 - Remote desktop, Jupyter, LiberTEM web GUI
 - Gateway for transfers
-
- NFS (Linux), CIFS (Windows, MacOS)
 - Fast network (WIP, currently small “island” but extension imminent)

REQUIREMENTS DATA MANAGEMENT

Acquisition – Storage – Access – **Processing** – Publication – Archival

PCs, laptops

- Testing, development, travel, data exchange, contingency
- Helpful if user's devices allow basic data analysis
- Work pretty well for smaller data

- Nextcloud local sync
- CIFS, SSHFS

REQUIREMENTS DATA MANAGEMENT

Acquisition – Storage – Access – Processing – Publication – Archival

Keep data that is not in active use anymore

- Store safely for at least the mandated period of time
 - Possibly, perpetual storage
 - National archives and such?
 - Keep all relevant meta-information connected
 - ELN entries, notes, software, provenance, specimens, ...
 - Information to allow others to understand what is stored there when the original authors are gone
 - Make sure there's a record of what is in the archive
-
- Currently deferred, keep everything in hot storage for now