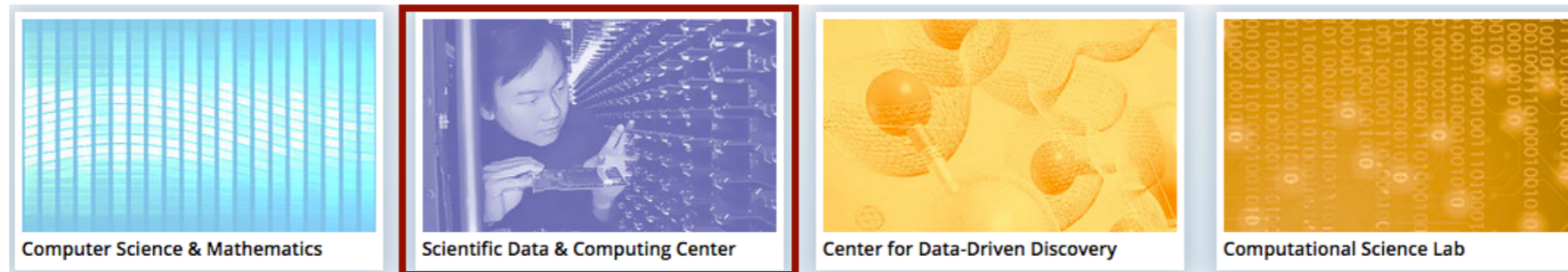


BNL BOX : Cloud Storage for Scientists

Ofer Rind
SDCC Tech Talk
June 23rd, 2020

- Intro to BNL Box
- Architecture
- Survey of Features — Current and Future
- Discussion

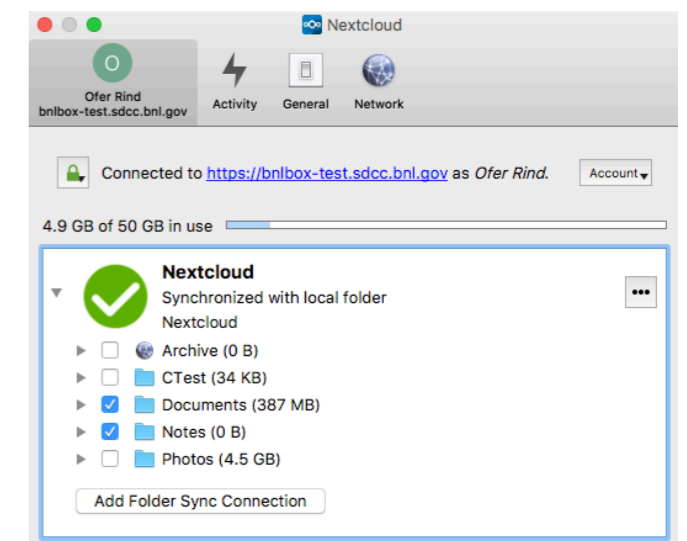
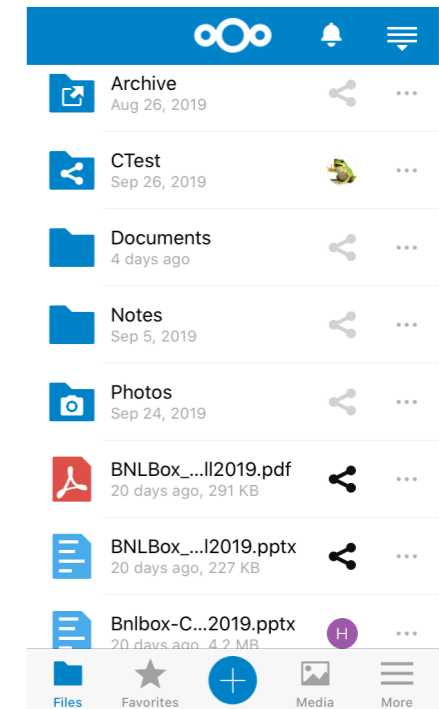
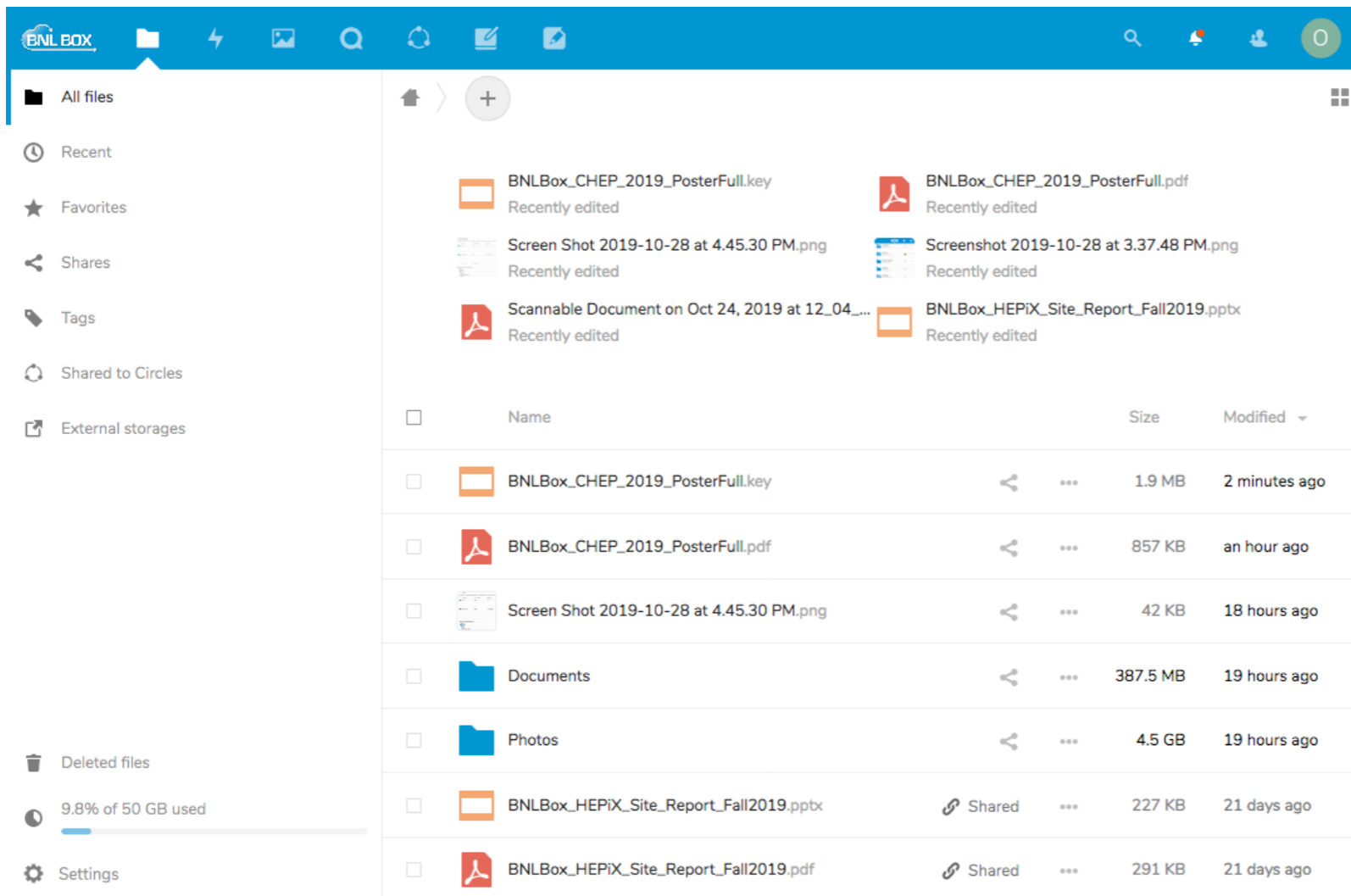
- The Scientific Data and Computing Center (SDCC) is a shared multi-program facility in Bldg. 510/515 serving ~2500 users from more than 20 projects with ~41 FTE
- A component of the Computational Science Initiative (CSI) that includes the RHIC & ATLAS Computing Facility (RACF) from the Physics Dept.



- Tier-0 for RHIC, US Tier-1 for ATLAS, US Data Center for Belle-II, plus photon sciences, neutrino, astrophysics, LQCD, biology, EIC,
- High-Throughput and High-Performance computing including 90K+ CPU cores, ~90 PB disk storage, 180 PB tape storage, 3x100 Gbps network (ESNET)
- New data center under construction in Bldg. 725 (2021)
- Increasingly involved in adapting, developing and deploying collaborative tools



BNL Box is an Enterprise File Sync and Share Service (EFSS) integrated into the SDCC to provide flexible, easy-to-use, unified cloud storage for all BNL scientific users



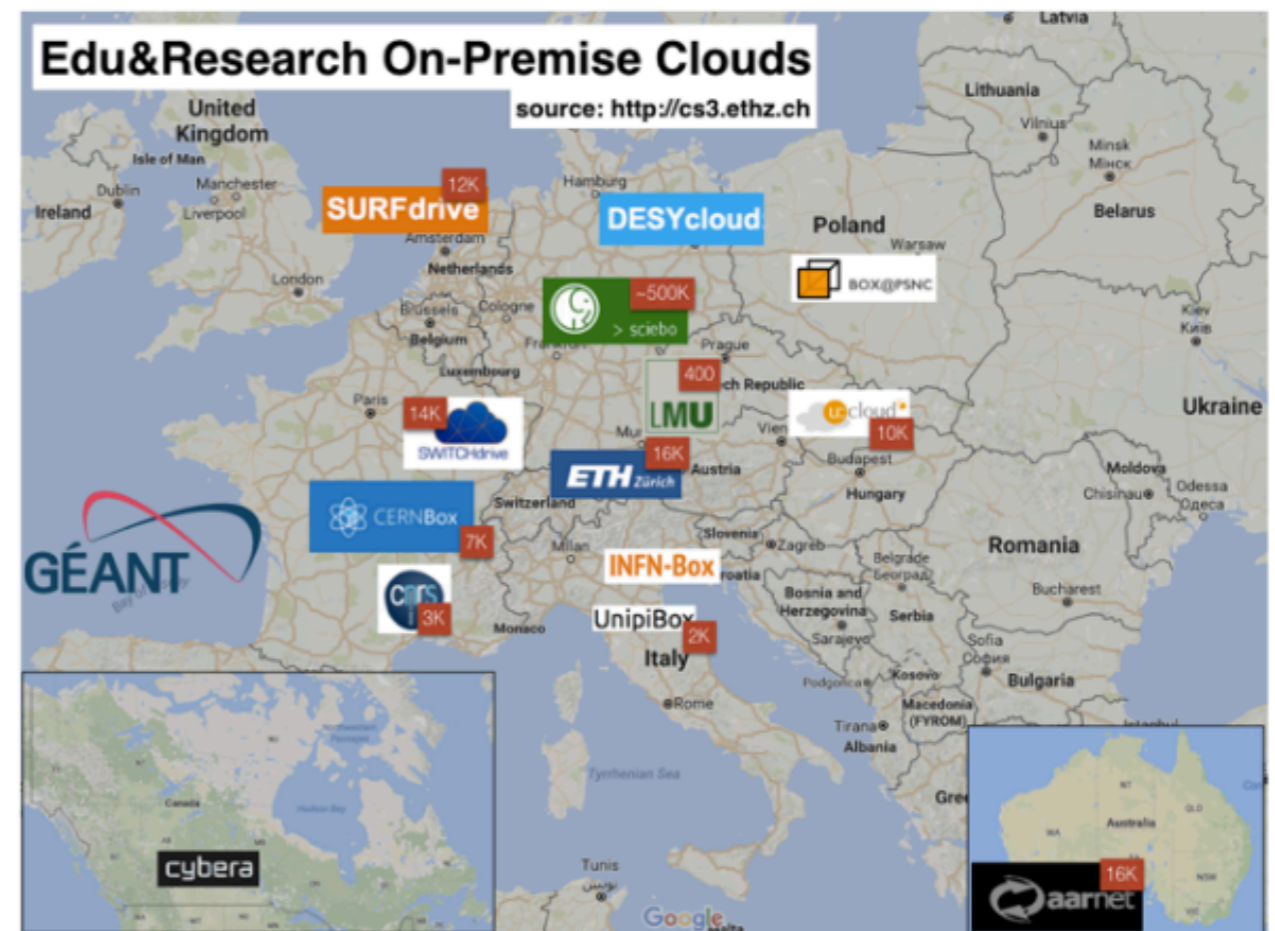
An integrated effort by many SDCC staff supporting back-end disk and tape storage, web and db services, AAI front-end, user interface, etc.

- Steadily growing cross-discipline community interest in EFSS driven by research needs and restrictions, for example:
 - Cloud Storage and Synchronization Services (CS3)
 - European Open Science Cloud (EOSC)

CS³ Site Reports ²⁰¹⁹

- **25** sites
- **5** EFSS technology stacks
- **400k** users
- **2.7 billion** files and directories
- **11.5 PB** storage
- Yearly storage increase: **5-250%**

Information provided on voluntary basis by CS3 site admins (January 2019)



<http://www.cs3community.org/>



København 2020



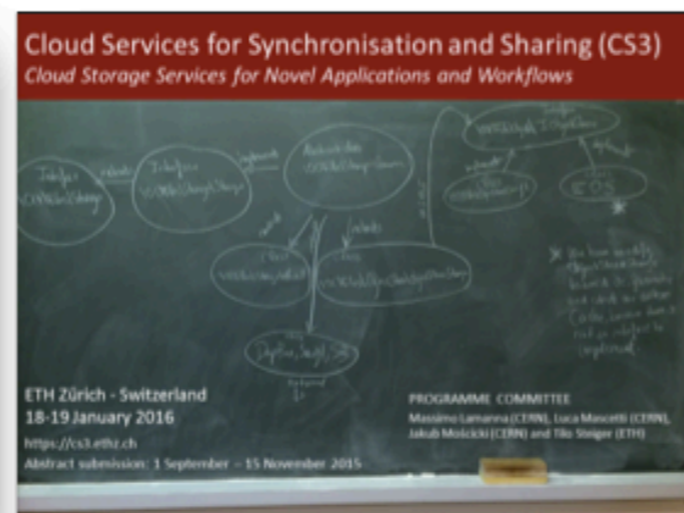
Roma 2019



Kraków 2018



Amsterdam 2017



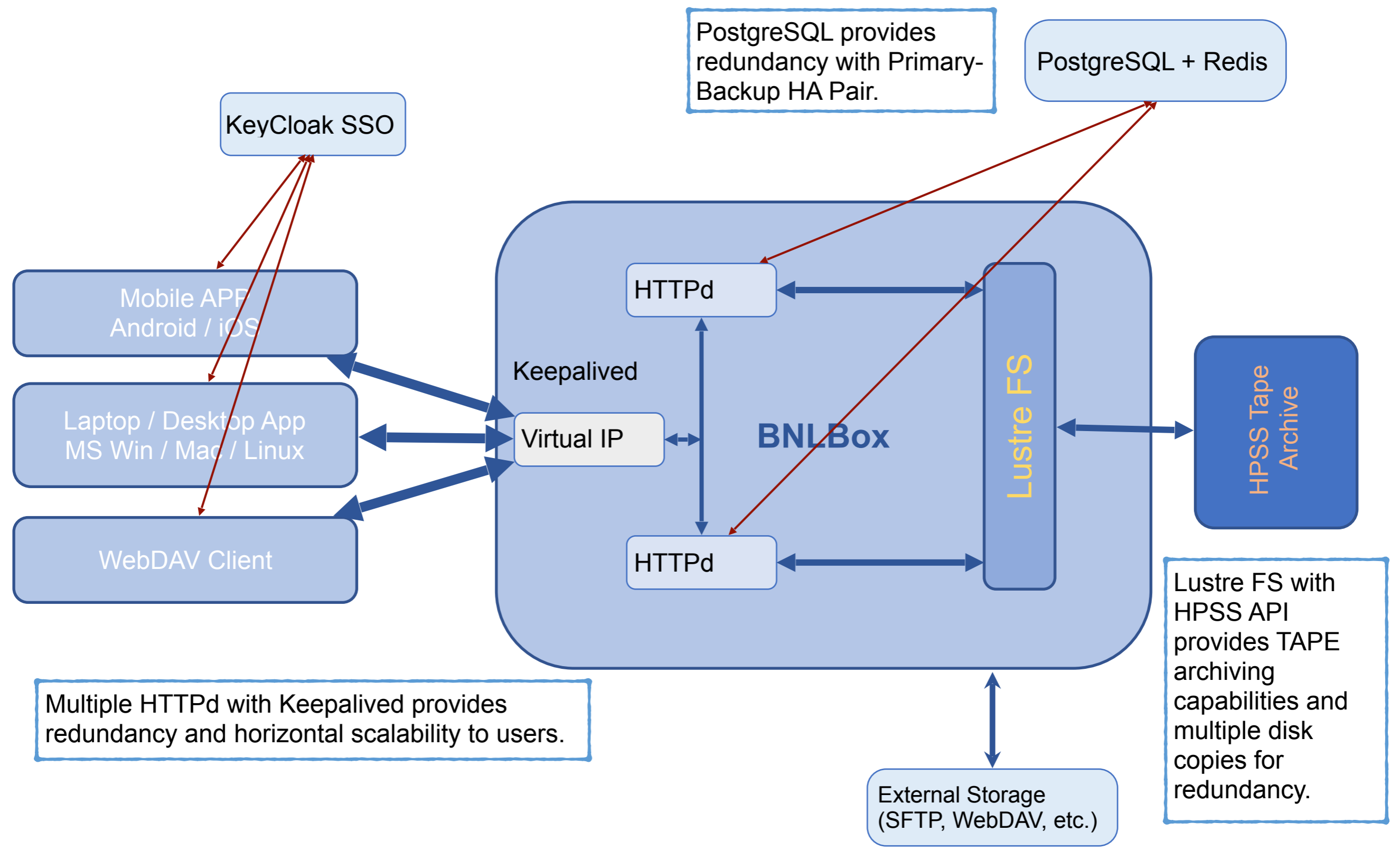
Zürich 2016



Genève 2014

Architecture

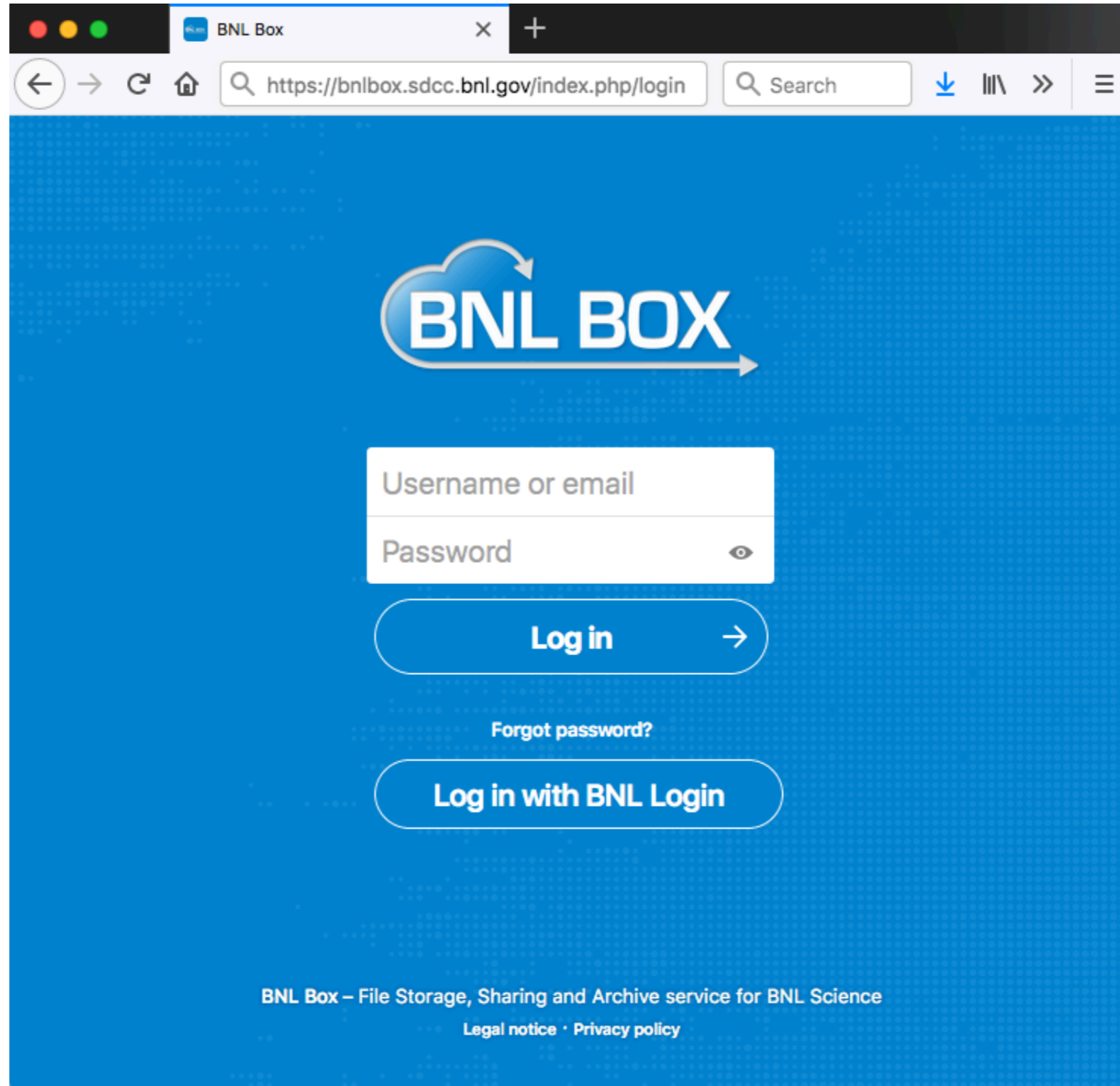
- BNLBox became a production service of the SDCC in Nov 2019, replacing the OwnCloud/Ceph prototype setup that had been running for the previous two years
 - Based on Nextcloud version 17. 100% free open source software.
 - PostgreSQL backend with a HA Primary/Backup pair for redundancy, and fast NVMe flash storage for performance.
 - Lustre file system, with multiple copies for redundancy, an HPSS API for archival storage and TSM backup. **Currently 1PB of usable storage**
 - Apache server front-end pair using KeepAlived for redundancy and load-balancing.
 - 40 Gbps WAN-to-storage network capacity
 - Authentication via KeyCloak OIDC integrated with SDCC LDAP and BNL AD
- **Account with 50 GB default quota automatically available to any BNL user**



Multiple HTTPd with Keepalived provides redundancy and horizontal scalability to users.

PostgreSQL provides redundancy with Primary-Backup HA Pair.

Lustre FS with HPSS API provides TAPE archiving capabilities and multiple disk copies for redundancy.



<https://bnlbox.sdcc.bnl.gov>

The screenshot displays the BNL BOX web interface in a browser window. The address bar shows the URL: `https://bnlbox.sdcc.bnl.gov/index.php/apps/files/?dir=/Phot`. The interface includes a navigation sidebar on the left with options like 'All files', 'Recent', 'Favorites', and 'Shares'. The main area shows a file list with columns for Name, Size, and Modified. The selected file, 'Pan_8478-81.jpg', is shown in a detailed view on the right, including its size (11.1 MB) and a sharing section with options like 'Share link' and 'Add to a project'.

| Name | Size | Modified |
|-------------------------|---------|---------------|
| Coast.jpg | 801 KB | 10 months ago |
| Hummingbird.jpg | 572 KB | 10 months ago |
| Nextcloud Commu... .jpg | 779 KB | 10 months ago |
| Nut.jpg | 933 KB | 10 months ago |
| Rainbow4862.jpg | 1.7 MB | 10 months ago |
| BNL BOX Logos_W... .svg | 220 KB | 3 years ago |
| BNL BOX Logos_W... .png | 116 KB | 3 years ago |
| Pan_8478... .jpg | 11.1 MB | 5 years ago |

Integrated apps for in-browser viewing of image files, video, and pdf

BNL BOX

File Storage, Sharing and Archive service for BNL Science

Nextcloud gives you access to your files wherever you are.
Our easy to use desktop and mobile clients are available for all major platforms at zero cost!

Get the apps to sync your files

Desktop app
Windows, OS X, Linux

ANDROID APP ON
Google Play

Download on the
App Store

Connect your desktop apps to BNL Box

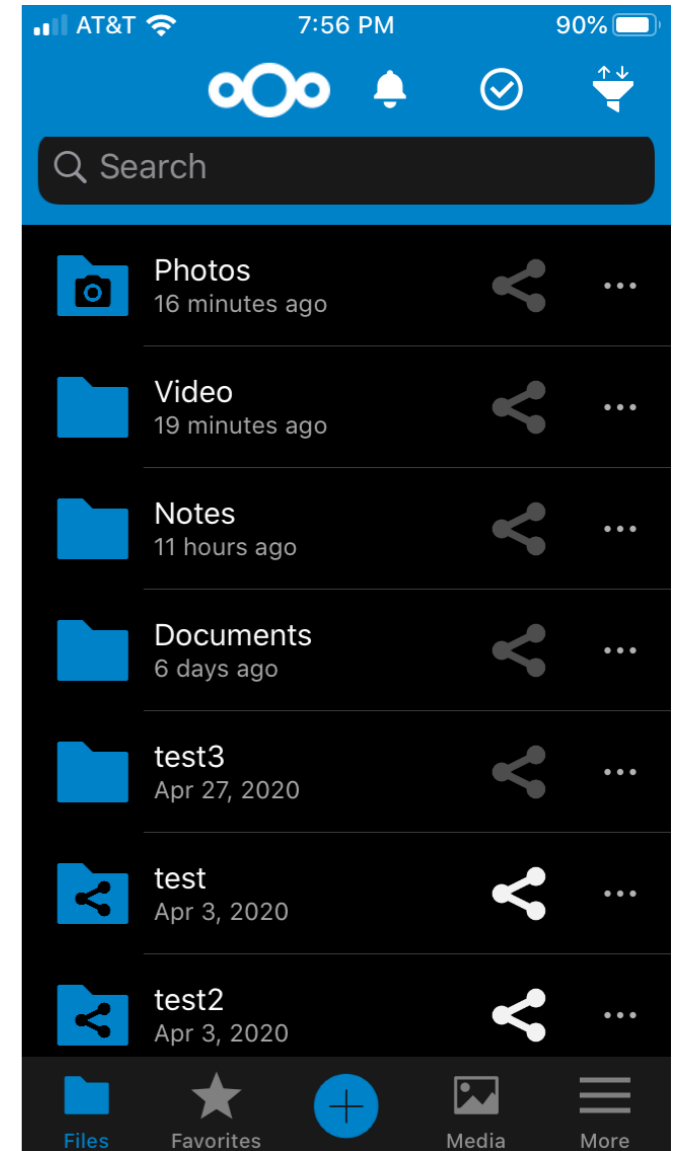
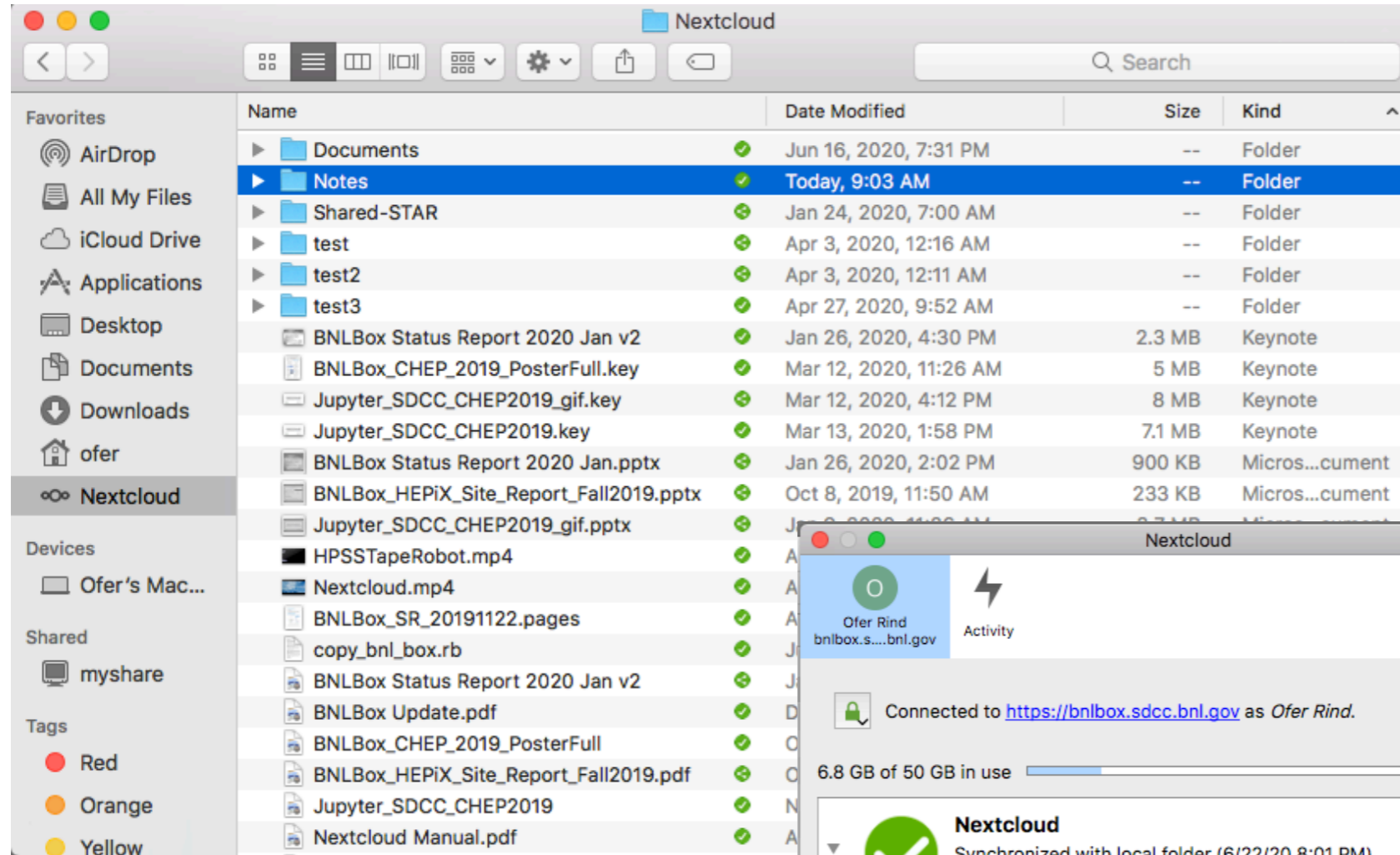
Connect your calendar

Connect your contacts

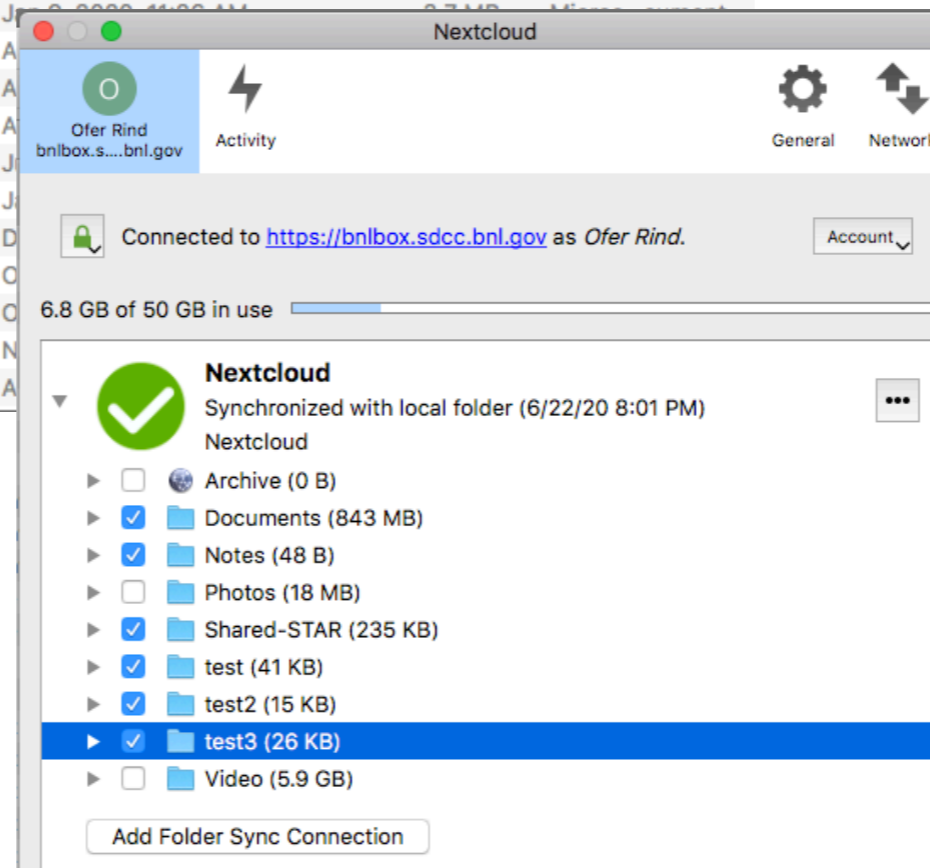
Access files via WebDAV

Download macOS/iOS configuration profile

Available at first login or from “About” link in upper-right menu



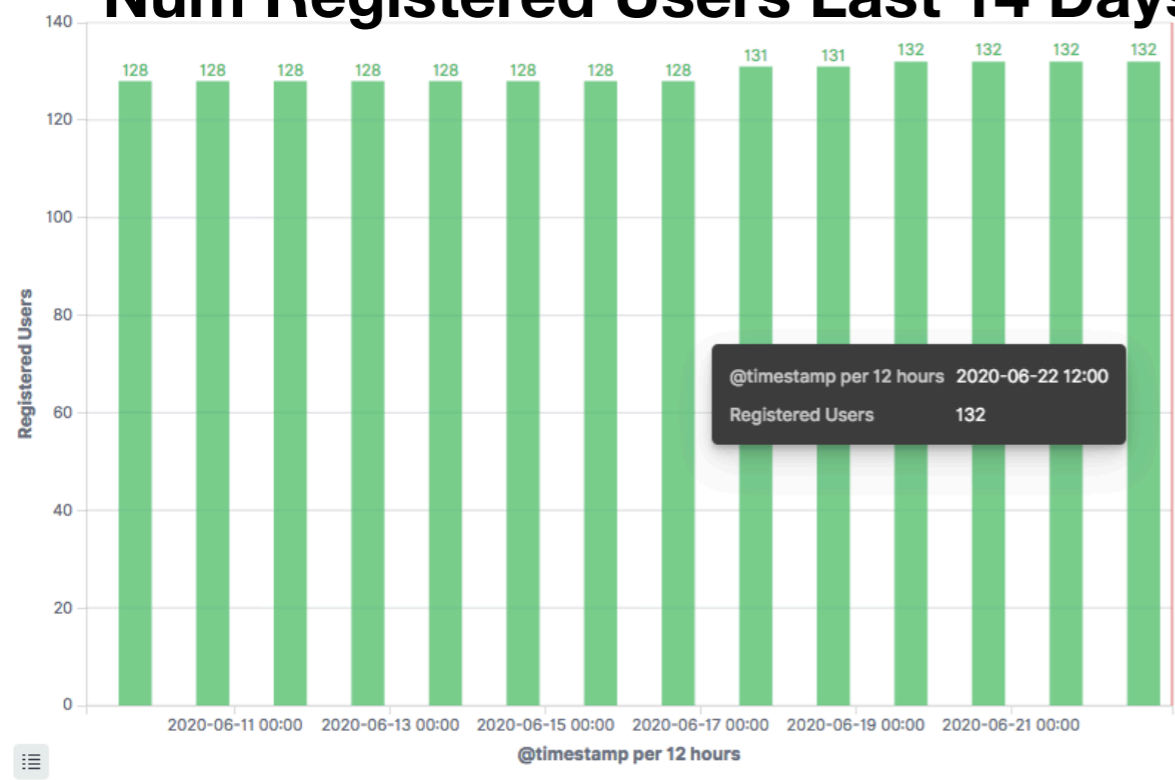
Cloud storage mounts with drag-and-drop functionality



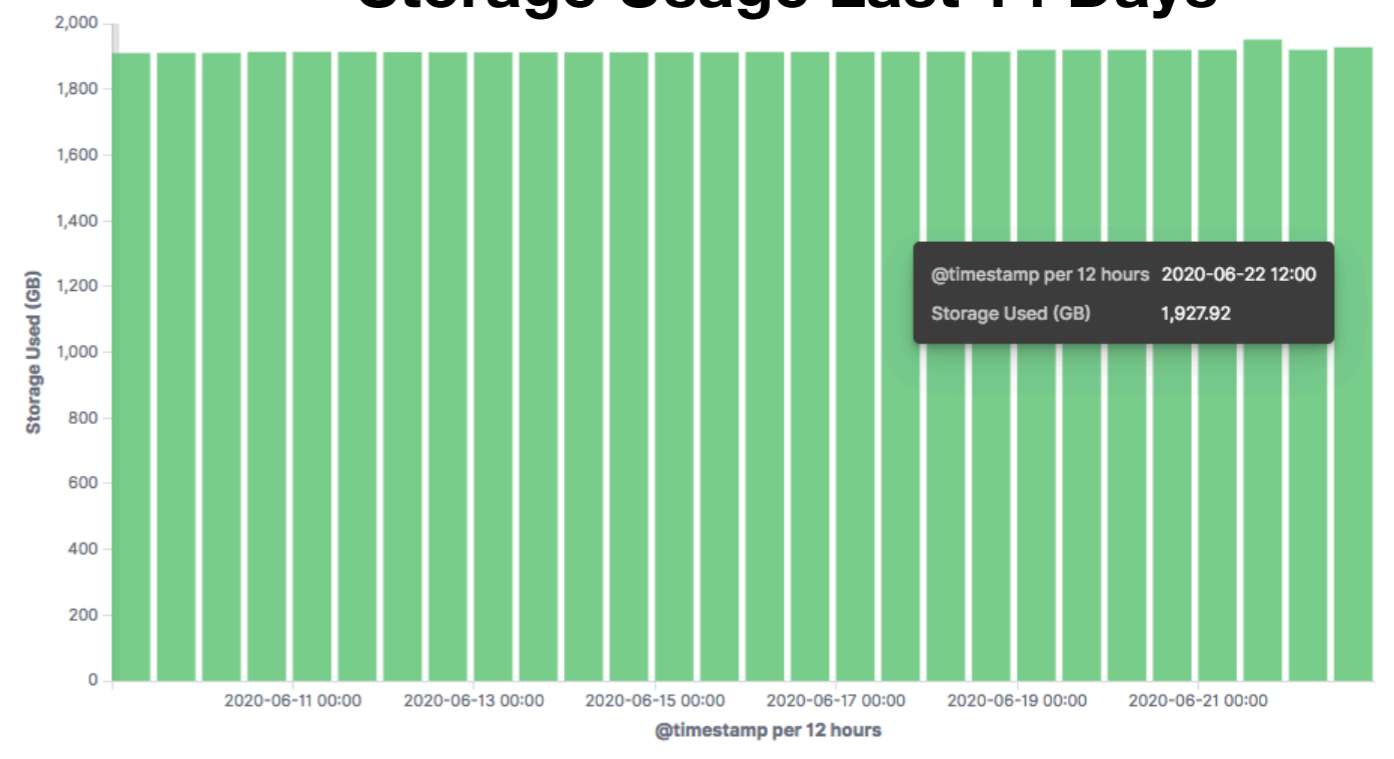
<https://nextcloud.com/install/#install-clients>

- Currently ~140 users, including guest accounts
- ~2 TB of user data comprising 3.2M files (90% of space used by top 5 users)
- Standard Ganglia and Nagios system monitoring
- Instrumenting/testing ELK dashboards using metricbeats, filebeats, custom DB queries to monitor overall usage as well as by user (current SULI student project)

Num Registered Users Last 14 Days



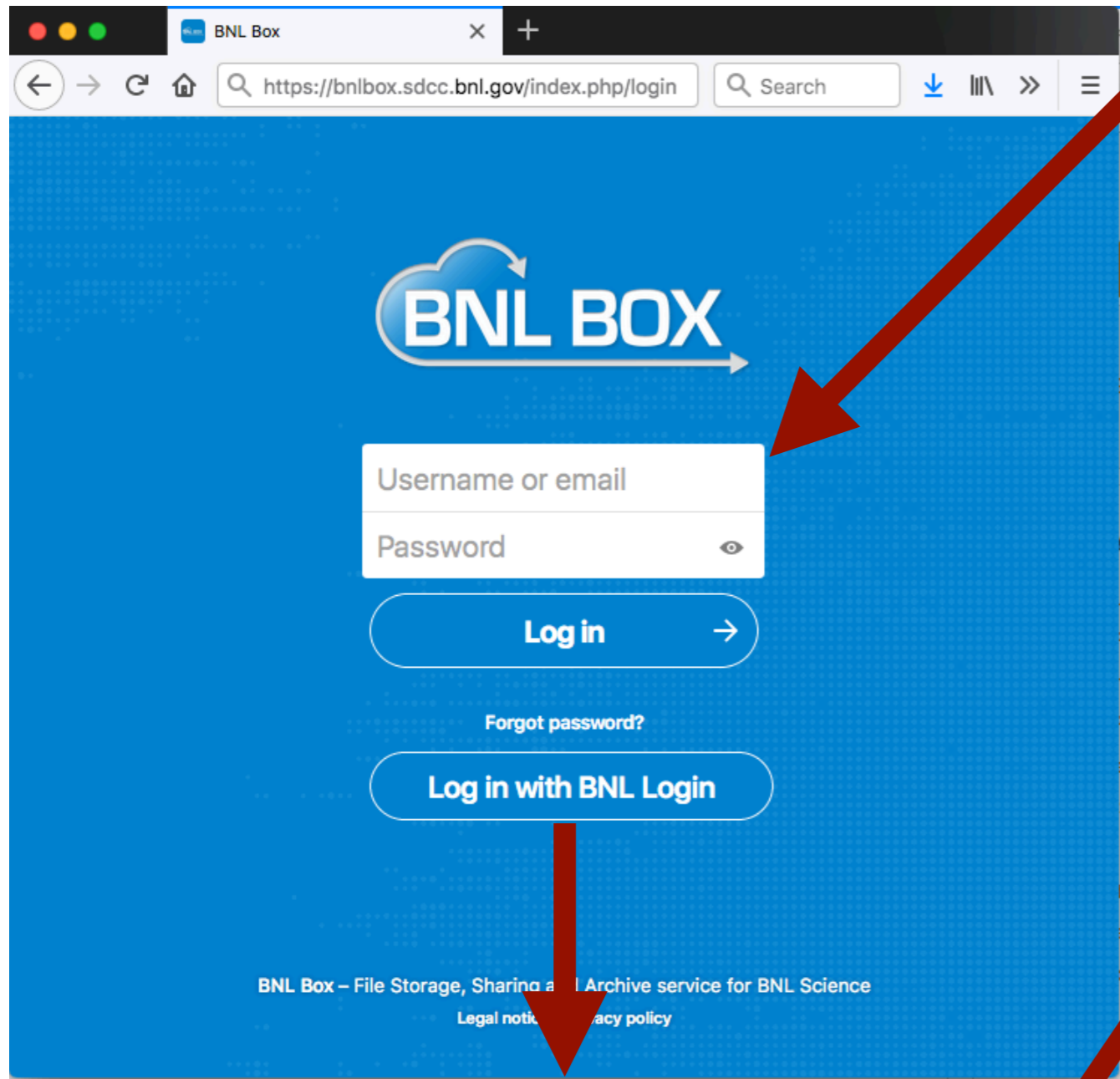
Storage Usage Last 14 Days



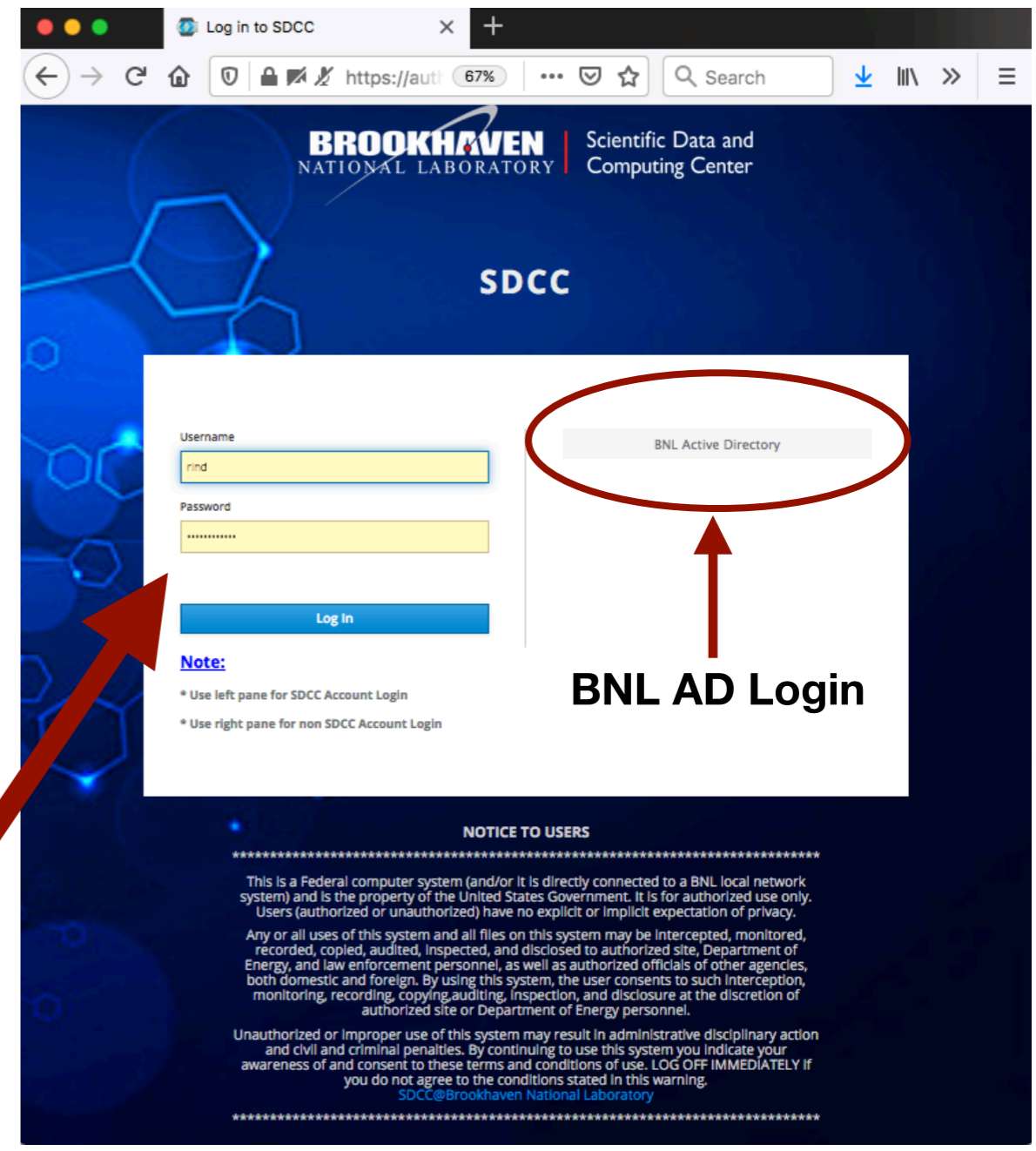
GeoIP Last 14 Days

Features

- *Locally-hosted*, secure, cloud storage, providing easy access to user data via browser, desktop & mobile clients, including automated synchronization
- Flexible file-sharing methodologies for sharing data with external collaborators
- Multiple access points from local user accounts, including from batch system, interactive hosts, etc.
- Tape archiving capability (via Lustre & HPSS) for large data sets
- Open to all users with SDCC or BNL AD accounts (via Keycloak based SSO) with straightforward capabilities for Federated ID



Used for guest account login



Login via Keycloak OIDC

Automated account creation upon first access
 Unique UUID avoids username clashes
 Duplicate prevention keyed to email address

Option of enabling MFA and/or federated account access in future

The screenshot shows a web browser window displaying the BNL BOX file management interface. The address bar shows the URL: <https://bnlbox.sdcc.bnl.gov/index.php/apps/files/?dir=/Photos&fileid=815>. The interface includes a sidebar with navigation options like 'All files', 'Recent', 'Favorites', and 'Shares'. The main area displays a list of files in the 'Photos' directory. The file 'Rainbow4862.jpg' is highlighted, and its sharing dialog is open on the right. The dialog shows the file name, size (1.7 MB), and modification date (10 months ago). It includes options for 'Share link' and 'Add to a project'. A search bar for 'Name or email address...' is also present. The dialog is titled 'Nextcloud Metadata'.

| Name | Size | Modified |
|----------------------------------|---------|---------------|
| IMG_4750.JPG | | |
| Coast.jpg | 801 KB | 10 months ago |
| Hummingbird.jpg | 572 KB | 10 months ago |
| Nextcloud Community.jpg | 779 KB | 10 months ago |
| Nut.jpg | 933 KB | 10 months ago |
| Rainbow4862.jpg | 1.7 MB | 10 months ago |
| BNL BOX Logos_White_gradient.svg | 220 KB | 3 years ago |
| BNL BOX Logos_White_gradient.png | 116 KB | 3 years ago |
| Pan_8478-81.jpg | 11.1 MB | 5 years ago |

Open the sharing dialog

File shared with me

Nextcloud Metadata

Features: File Sharing With Other Users



Rainbow4862.jpg
★ 1.7 MB, 10 months ago Tags
Chat Activity Comments **Sharing** Versions

- Name or email address... →
- O OferGuest
 - W William Strecker-Kellogg
 - H Hironori Ito
 - H Hiro
 - H hollowec@bnl.gov
hollowec@bnl.gov

Dialog autocompletes known users and guests

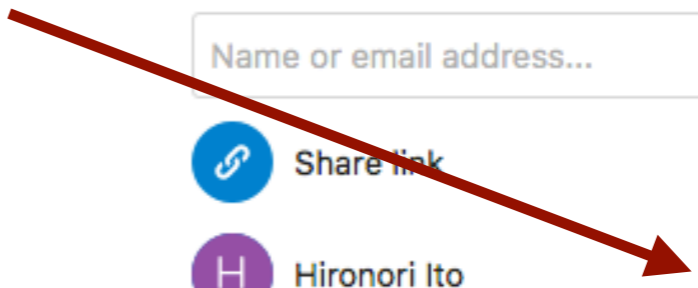


Rainbow4862.jpg
★ 1.7 MB, 10 months ago Tags
Chat Activity Comments **Sharing** Versions

- Name or email address... →
- Share link
 - H Hironori Ito Can edit
 - Add to a project
Connect items to easier to find
 - Can reshare
 - Set expiration date
 - Note to recipient
 - Unshare

Default Read-only

Sharing Options



Features: File Sharing By Link



Rainbow4862.jpg

★ 1.7 MB, 10 months ago Tags

Chat Activity Comments **Sharing** Versions

Name or email address...

Share link

Add to Connections easier

Password protection for links is mandatory

..... ✓

Password is among the 1,000,000 most common ones. Please make it unique.

Generates unique https link
Password enforced

Link enforced read-only



Rainbow4862.jpg

★ 1.7 MB, 10 months ago Tags

Chat Activity Comments **Sharing** Versions

Name or email address...

Share link

Add to Connections easier

Allow editing

Hide download

Password protection enforced

.....

Password protect by Talk

Set expiration date

07-07-2020

Note to recipient

Delete share link

+ Add another link

Default expiration date



Rainbow4862.jpg
★ 1.7 MB, 10 months ago Tags
Chat Activity Comments **Sharing** Versions

- tony@bnl.gov →
- T** tony@bnl.gov tony@bnl.gov ✉
- 🔍 Search globally
- C** Create guest account for tony@bnl... 👤

Option of creating a limited access guest account for more flexibility

Generates unique https link sent via email



Rainbow4862.jpg
★ 1.7 MB, 10 months ago Tags
Chat Activity Comments **Sharing** Versions

- Name or email address... →
- S** Share link 📄 ...
 - H** Hironori Ito Can edit ...
 - T** tony@bnl.gov (email) Can edit ...
 - A** Add to a p... Connect ite... easier to fir

Default Read-only

Options to set password protection and expiration date

- Password protect
- Password protect by Talk
- Set expiration date
- N** Note to recipient
- U** Unshare

Ofer Rind via BNL Box


Ofer Rind shared »Rainbow4862.jpg« with you

To: ofer@bnl.gov,

Reply-To: Ofer Rind

Inbox - rcf.rhic.bnl.gov 11:56 AM




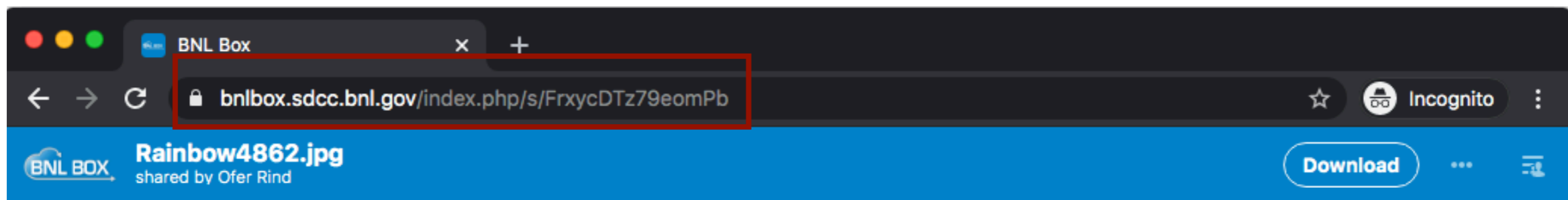


Ofer Rind shared »Rainbow4862.jpg« with you

Ofer Rind shared »Rainbow4862.jpg« with you. Click the button below to open it.

Open »Rainbow4862.jpg«

BNL Box – File Storage, Sharing and Archive service for BNL Science



Discuss this file

Join conversation

BNL Box – File Storage, Sharing and Archive service for BNL Science
[Legal notice](#) · [Privacy policy](#)
[Get your own free account](#)

Search circles

No filter

STAR
Closed (Member, Member)
Owner: Jerome Lauret

A boring public circle
Public (Unknown, Not a member)
Owner: Terence Healy

TEST
Closed (Unknown, Not a member)
Owner: Hironori Ito

TEST
Closed circle

Join this circle

Settings

| Username | Level | Status | Joined |
|--------------|-------|--------|---------------------|
| Hironori Ito | Owner | Member | 2019-12-11 20:53:55 |

Personal circles

Public circles

Closed circles

Secret circles

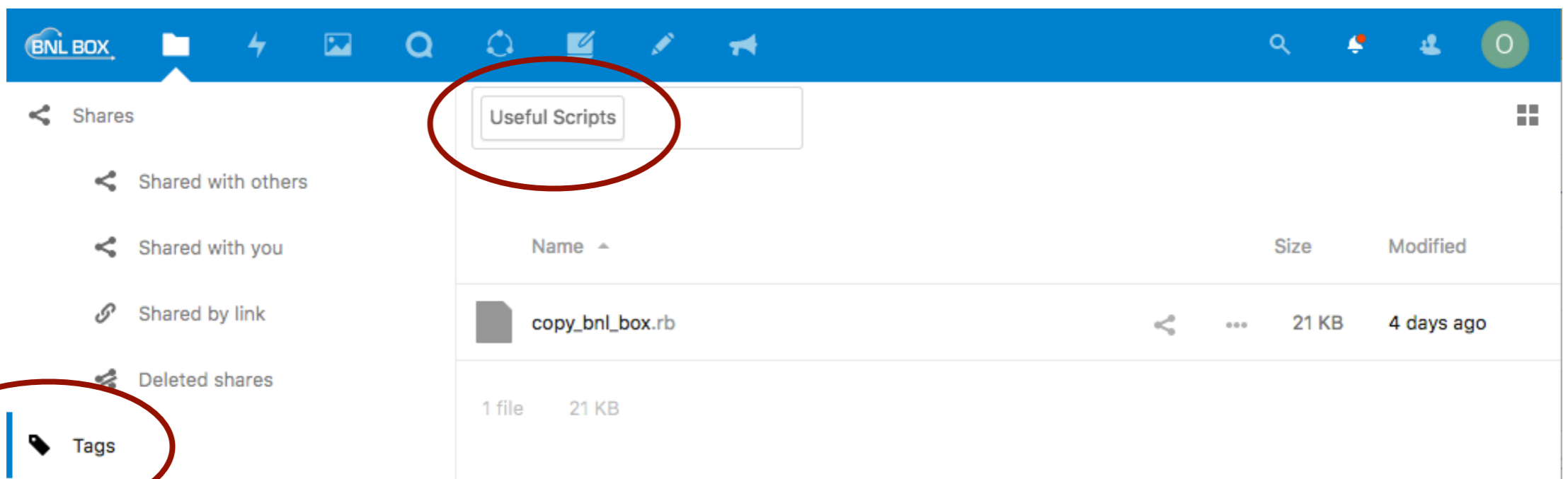
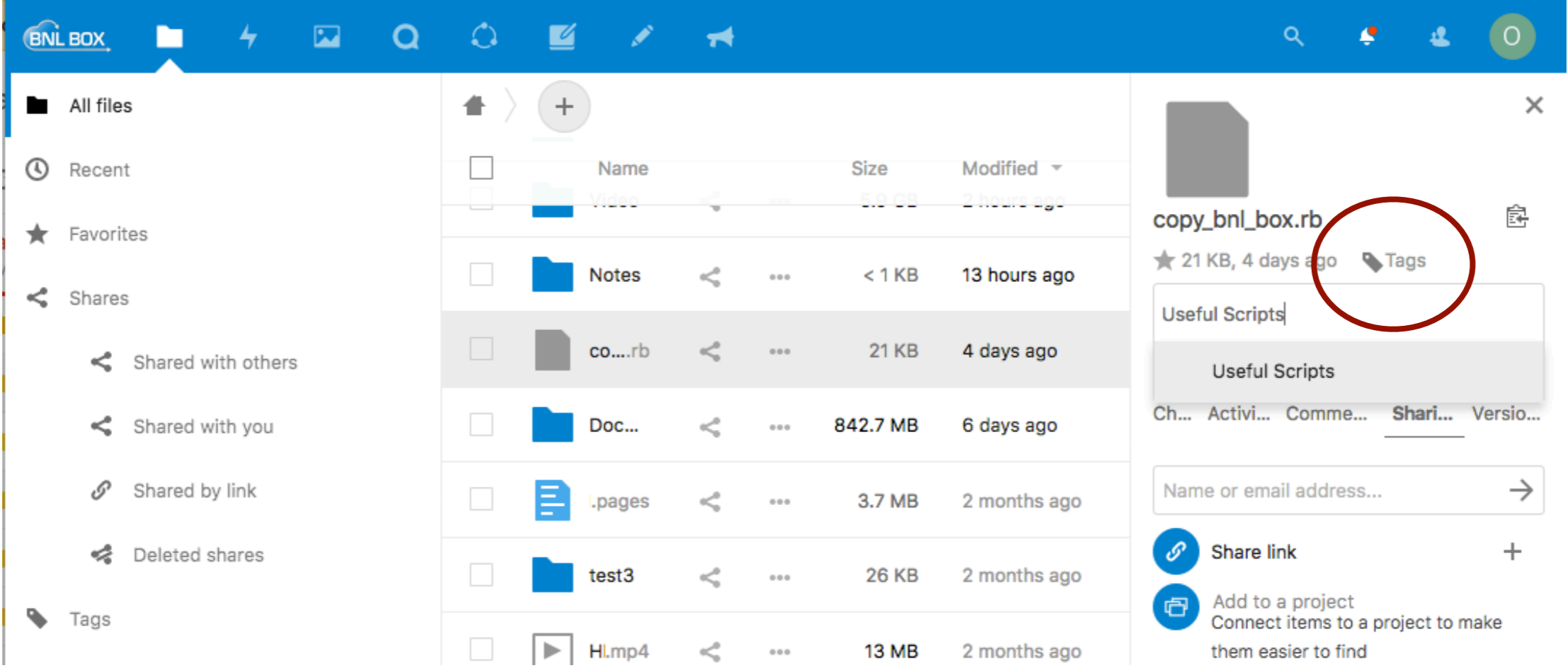
All circles

<https://bnlbox.sdcc.bnl.gov/index.php/apps/circles/>

Flexible user-defined groups of users — usable by any other app for sharing purposes

<https://apps.nextcloud.com/apps/circles>

Features: Organizing Files by Tagging



Link different types of information together (files, chats, etc.) across a group of users. Files added to a Project can be shared or not.

The screenshot displays the Nextcloud web interface. On the left, a sidebar shows navigation options: All files, Recent, Favorites, Shares, and Tags. The main area shows a file list with columns for Name, Size, and Modified. The file 'Berlin.jpg' is selected and highlighted. On the right, a detailed view of 'Berlin.jpg' is shown, including its size (4.5 MB) and modification date (4 months ago). Below the file details, there are options for Chat, Activity, Comments, Sharing, and Versions. A search bar is present for sharing. A red box highlights the 'Add to a project' section, which shows a list of projects and files associated with 'Project Starfish'. The files listed include 'and a final test file.md', 'colors.jpeg', 'Project Starfish', 'Feature description.md', 'Berlin.jpg', and 'Project Starfish chat'.

| Name | Size | Modified |
|--------------------------|---------|--------------|
| airplanes | 35.5 MB | 4 months ago |
| Pexels | 52.8 MB | 4 months ago |
| Berlin.jpg | 4.5 MB | 4 months ago |
| Coast.jpg | 801 KB | 5 months ago |
| git-remote.jpg | 21 KB | 4 months ago |
| Hummingbird.jpg | 572 KB | 5 months ago |
| Nut.jpg | 933 KB | 5 months ago |
| pexels-photo-166680.jpeg | 76 KB | 4 months ago |
| small car.jpeg | 551 KB | 4 months ago |
| 2 folders and 7 files | | 95.7 MB |

Image: F. Karlitschek

Integrated Chat and Calling feature for discussion of documents, etc. (mobile apps available)

The screenshot displays the Nextcloud Talk interface within a web browser. The browser's address bar shows the URL `https://bnlbox.sdcc.bnl.gov/index.php/call/xr6m26ct`. The interface features a blue navigation bar with icons for Files, Activity, Gallery, Talk (highlighted with a red box), Circles, Quick, Notes, and Anno. Below the navigation bar, a sidebar on the left lists various chat conversations, including 'New conversation ...', 'OferGuest', 'Software & Computing Round T', 'Hironori Ito', 'TEST Group Chat', 'Talk updates' (with a green checkmark), 'Jerome Lauret', 'John S. De Stefano Jr.', and 'Kevin Casella'. The main chat area shows a conversation with 'John S. De Stefano Jr.' and 'Ofer Rind'. The chat history includes messages from Ofer Rind at 2:34 PM and 3:01 PM, and John S. De Stefano Jr. at 3:02 PM. A timestamp '10 months ago, September 6, 2019' is visible. The chat area also shows a message from John S. De Stefano Jr. at 10:39 AM with a link to `www.racf.bnl.gov/docs/services/cloud-storage/bnlbox-mobile`. At the bottom of the chat area, there is a text input field labeled 'New message ...' with a send button (arrow) and a plus sign. On the right side, a panel titled 'Conversation with John S. D...' is open, featuring a prominent blue 'Start call' button, icons for 'Participants' and 'Projects', and a list of participants: 'Ofer Rind (moderator)' and 'John S. De Stefano Jr. (moderator)'.

The screenshot displays the Nextcloud Talk interface. At the top, a browser window shows the URL <https://bnlbox.sdcc.bnl.gov/index.php/call/xr6m26ct>. The main interface is divided into three sections:

- Left Panel:** A list of conversations including "John S. De Stefano Jr.", "OferGuest", "Software & Computing Round T", "Hironori Ito", "TEST Group Chat", "Talk updates" (with a checkmark), "Jerome Lauret", and "Kevin Casella".
- Center Panel:** A video call in progress with the title "Features: Nextcloud Talk". The call area shows a "Your screen" label and a video feed of a man wearing a green cap. Below the video are icons for microphone, video, and screen sharing.
- Right Panel:** A chat window titled "Conversation with John S. D...". It includes a "Leave call" button, tabs for "Chat", "Participants", and "Projects", and a message history. The chat history shows a message from John S. De Stefano Jr. at 3:02 PM saying "For sure", a timestamp "10 months ago, September 6, 2019", another message from John S. De Stefano Jr. at 10:39 AM with a link www.racf.bnl.gov/docs/services/cloud-storage/bnlbox-mobile, a timestamp "Today, June 22, 2020", a system message "You started a call" at 10:03 PM, and a message from Ofer Rind. At the bottom is a "New message ..." input field.

<https://nextcloud.com/talk/>

The screenshot displays the Nextcloud Notes application interface. At the top, a blue navigation bar contains icons for Files, Activity, Gallery, Talk, Circles, Quick, Notes (highlighted with a red box), and Anno. Below this, a sidebar on the left shows a '+ New note' button, a 'Categories' section, and a list of notes including 'Note on Notes' and 'For_Terry'. The main content area shows a note titled '# Notes' with the subtitle '## A safe home for all your data'. The note text describes the app's purpose and features, such as supporting Markdown syntax and being accessible via mobile apps for Android and iOS. A 'Category' dropdown menu is visible, with 'DEMO' selected (circled in red). Below the note, a file list shows 'Note on Notes.txt' with a size of '< 1 KB' and a modification time of '2 minutes ago'. A red circle highlights the breadcrumb navigation 'Notes > DEMO' and a '+' button. On the right, a 'Note on Notes' panel shows the note's details, including the category 'DEMO' (circled in red), and options for sharing, activity, and comments.

<https://apps.nextcloud.com/apps/notes>

<https://apps.nextcloud.com/apps/quicknotes>

NB: Full functionality (e.g. attachments) not yet available in the current version of BNL Box!

Features: Managing, Sharing and Archiving Scientific Data at BNL

In many cases, e.g. storing and sharing output generated by jobs running on the batch queues, users may want to transfer data to/from BNL Box in a scripted manner from the Unix command line

- Secure access provided via configurable WebDAV app
- A number of standard CL tools, such as curl, are available
- A custom ruby script, developed by H. Ito, provides a flexible, feature-rich copy tool that can handle bulk directory copies and wildcarding

For detailed documentation, see: [How To Access BNL Box From The Command Line](#)

```
[rind@rcas2501 ~]$ ./copy_bnl_box.rb -h
Usage: copy_bnl_box.rb [options] source destination
NOTE: to specify the bnlbox storage, you need to prepend 'box:' to your path seen in your bnl box service
-h, --help                Display help message
-s, --https_proxy ProxyHost Set https proxy
-r, --url URL              Set none-default url
    --password password   Set User Password
-u, --username username   Set User Name
-f, --file filename       Configuration File
-v, --verbose              Verbose messages
-d, --debug                Debug option
-p, --progress             Show progress
[rind@rcas2501 ~]$
[rind@rcas2501 ~]$ ./copy_bnl_box.rb -p box:/Photos/*.jpg .
Downloading Coast.jpg: 100% of total 819766 bytes      Done
Downloading Hummingbird.jpg: 100% of total 585219 bytes      Done
Downloading Nextcloud%20Community.jpg: 100% of total 797325 bytes      Done
Downloading Nut.jpg: 100% of total 955026 bytes      Done
Downloading Rainbow4862.jpg: 100% of total 1787072 bytes      Done
Downloading Pan_8478-81.jpg: 100% of total 11600097 bytes      Done
[rind@rcas2501 ~]$
```


Provide users with independent entry point to sharable filesystems from linux farm, non-SDCC storage, experiment online storage, etc.

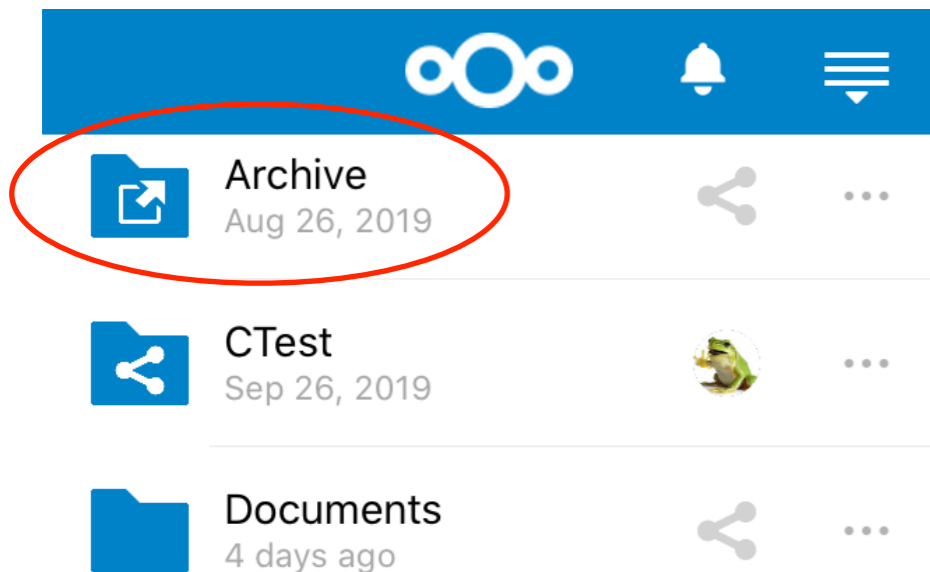
External storages *i*

External storage enables you to mount external storage services and devices as secondary Nextcloud storage devices. You may also allow users to mount their own external storage services.

| Folder name | External storage | Authentication | Configuration | Available for |
|----------------------------------------------|------------------|----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|
| <input checked="" type="checkbox"/> Archive | Local | None | /sdcc/lustre01/BNLBc | All users. Type to sele |
| <input checked="" type="checkbox"/> EIC_Data | SFTP | RSA public key | <input type="text" value="localhost"/> <input type="text" value="/gpfs02/eic/DATA"/> <input type="text" value="eickolja"/> <input type="text" value="ssh-rsa AAAAB3NzaC"/> <input type="text" value="1024"/> <input type="button" value="Generate keys"/> | Kolja Kauder ✕ |

NB: Files on external storage do not count against Nextcloud user's quota

- External Storage app enables mounting of shareable volumes via numerous protocols, e.g. SFTP, WebDAV, etc.
- Admin-created mounts are *read-only* and linked to Nextcloud user via PKI
- Files on external storage owned by original user, who grants account access to the Nextcloud user
- Nextcloud user may share/access external files in same manner as primary storage



- Provide long-term archival storage of large shared data sets
- Free up disk space with option of retrieval in (distant) future.
 - Archived files should be retrievable (with some latency) in transparent fashion, while not counting against user's storage quota.

- Solution: Archive Directory

- Independent directory on local Lustre filesystem mounted as External Storage
- File archiving via Lustre Copytool interface to HPSS tape storage API (configurable, driven by watermarks)
- Provisioned automatically upon BNL Box account creation and linked to user via Postgres DB trigger

Future

- With NC 18, Nextcloud has enhanced their integrations, especially with office applications
 - OpenOffice, Collabora Online integration -> Interactive realtime document editing
- Rebranded to Nextcloud Hub, emphasizing their service as an integrated, on-premises, content collaboration platform.
- We are continuing to explore functionality upgrades and further integrations driven by the needs of our scientific community



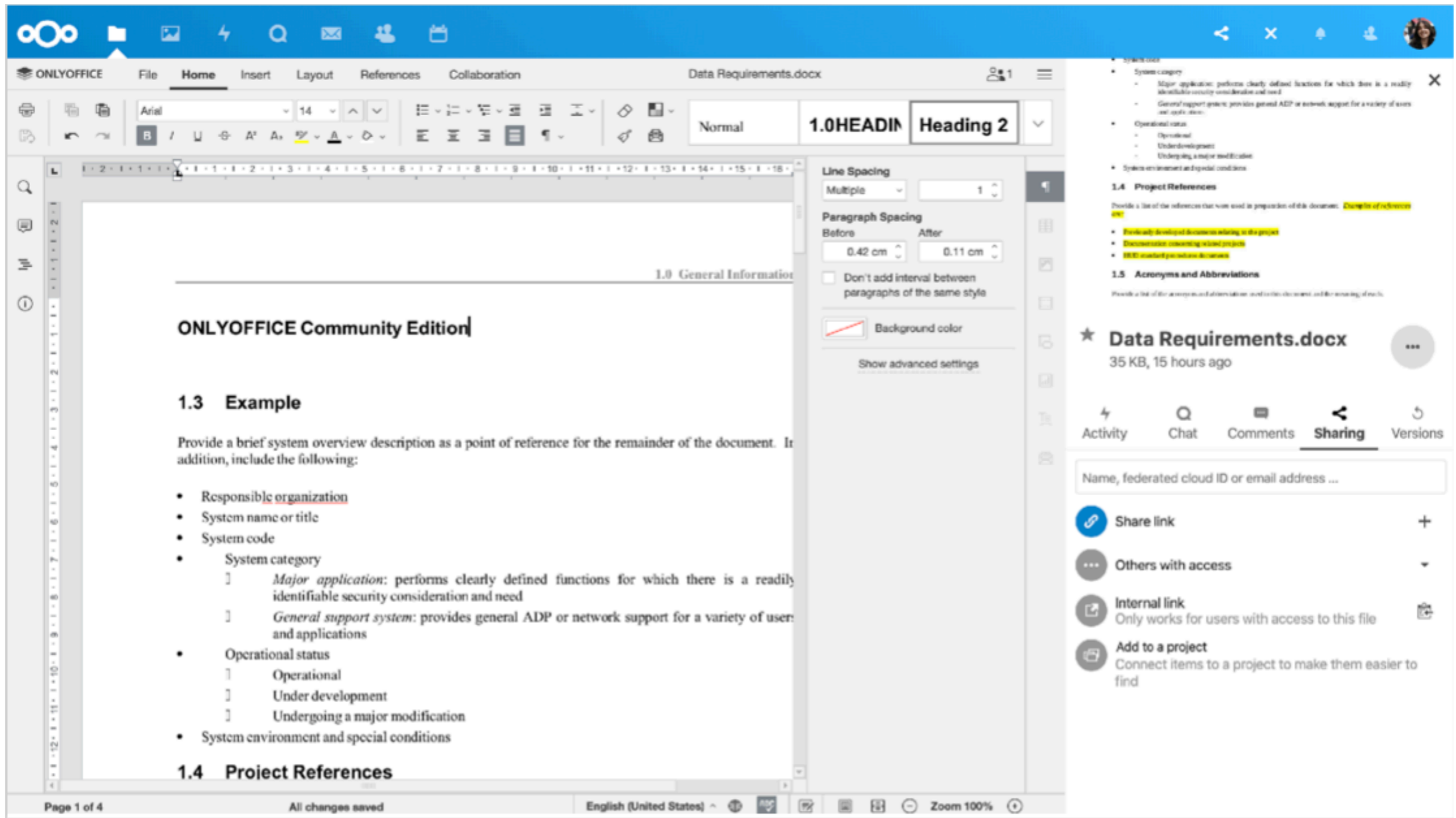


Image: F. Karlitschek

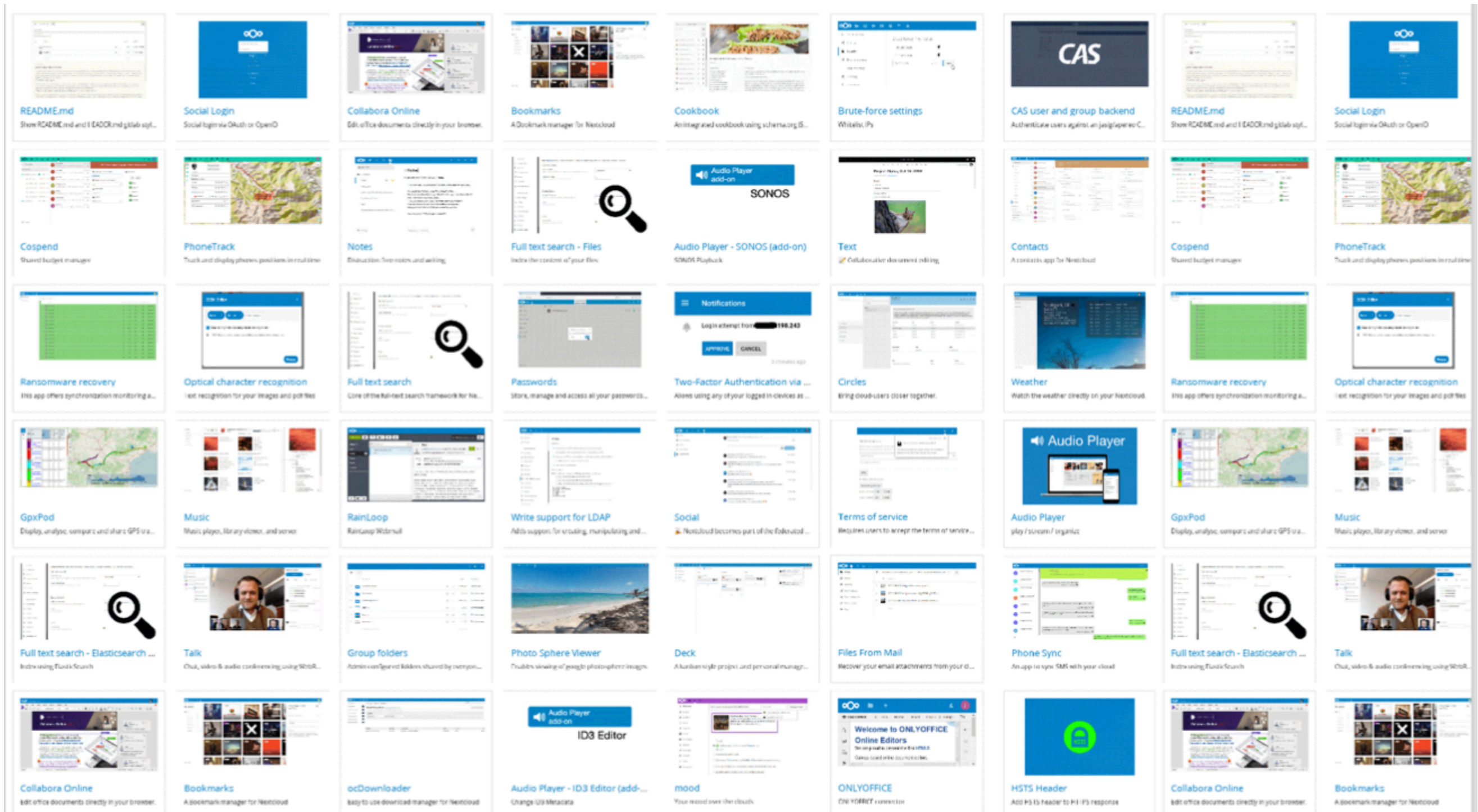


Image: F. Karlitschek

- BNL Box is a flexible, locally-hosted, secure, cloud storage solution for science, currently available to everyone with a BNL computer account
- BNL Box provides a feature-rich platform for organizing and sharing all types of data, including those generated and stored at SDCC and user facilities
- The SDCC is continuing to explore further integrations and use cases for BNL Box, driven by stakeholder needs
- If you have a BNL computer account, give it a try today!

<https://bnlbox.sdcc.bnl.gov>

Thank You

Questions?