StrongLink HSM system archival, retrieval, metadata

DKRZ User Workshop

13.10.2022

Andrej Fast & Hadieh Monajemi

Deutsches Klimarechenzentrum (DKRZ)

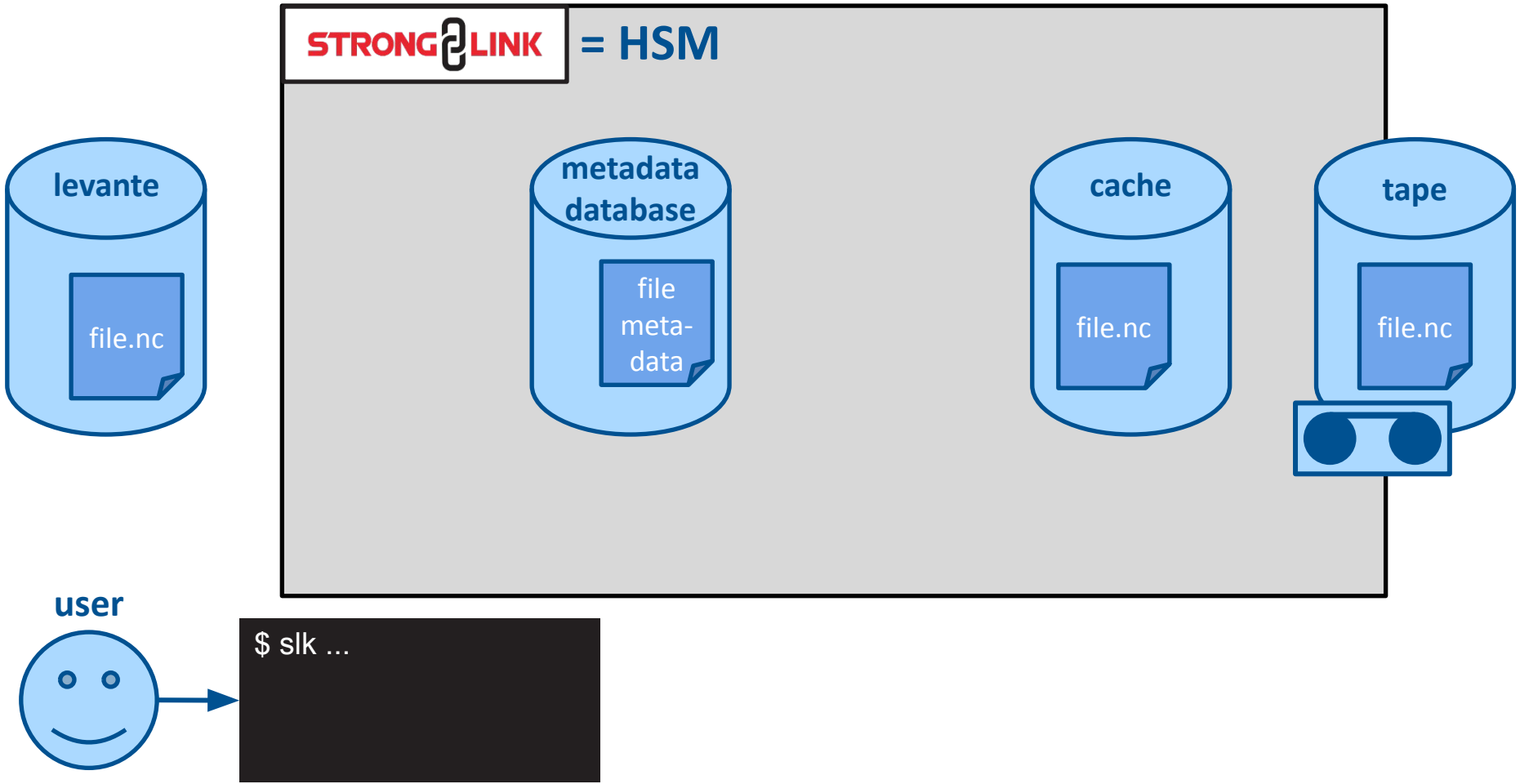
agenda

- Presentation I: Basic introduction
- Hands-On I: archival and retrieval
- Presentation II: Metadata in StrongLink
- Hands-On II: search and metadata
- Q&A

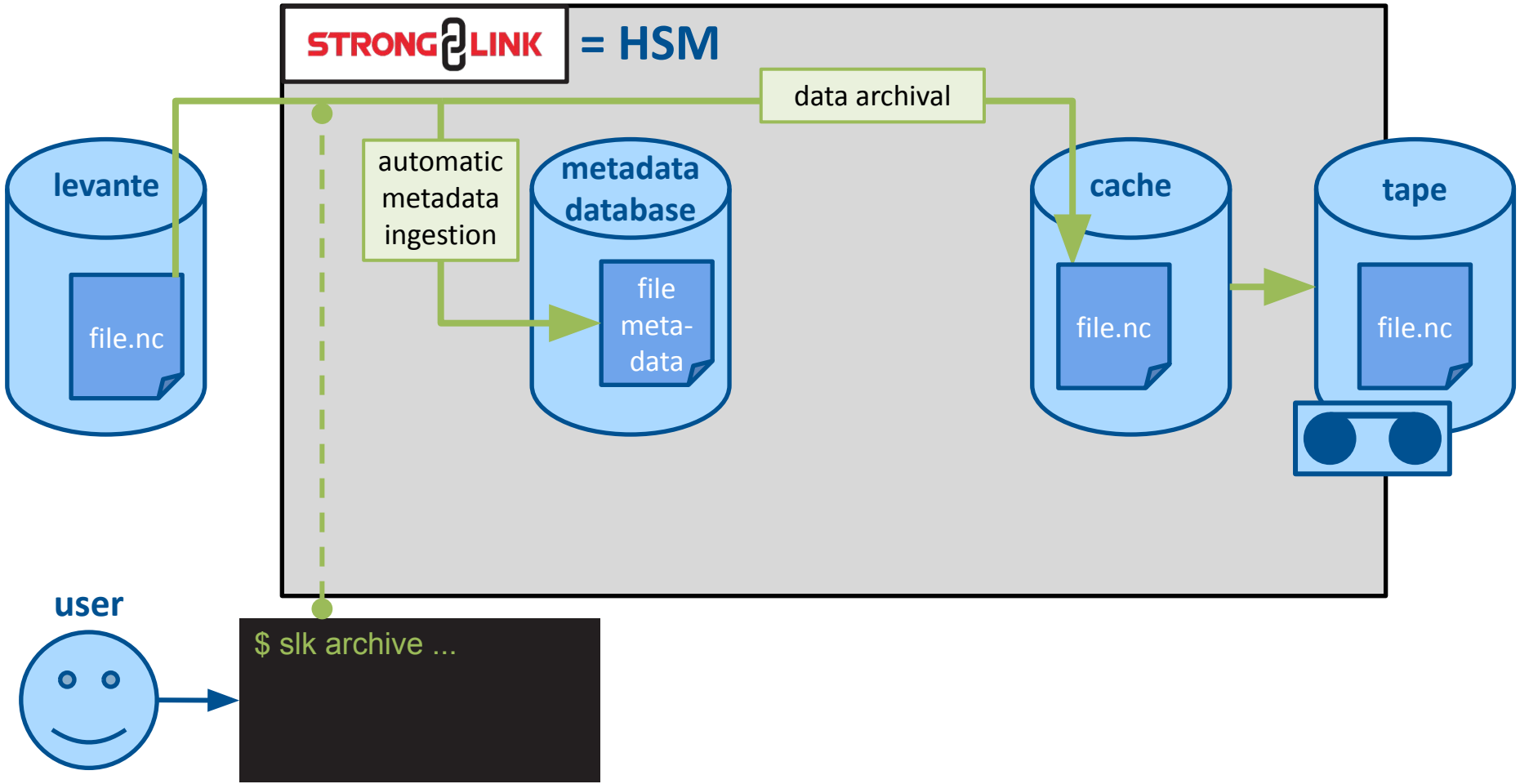
Presentation I

archival, retrieval and search

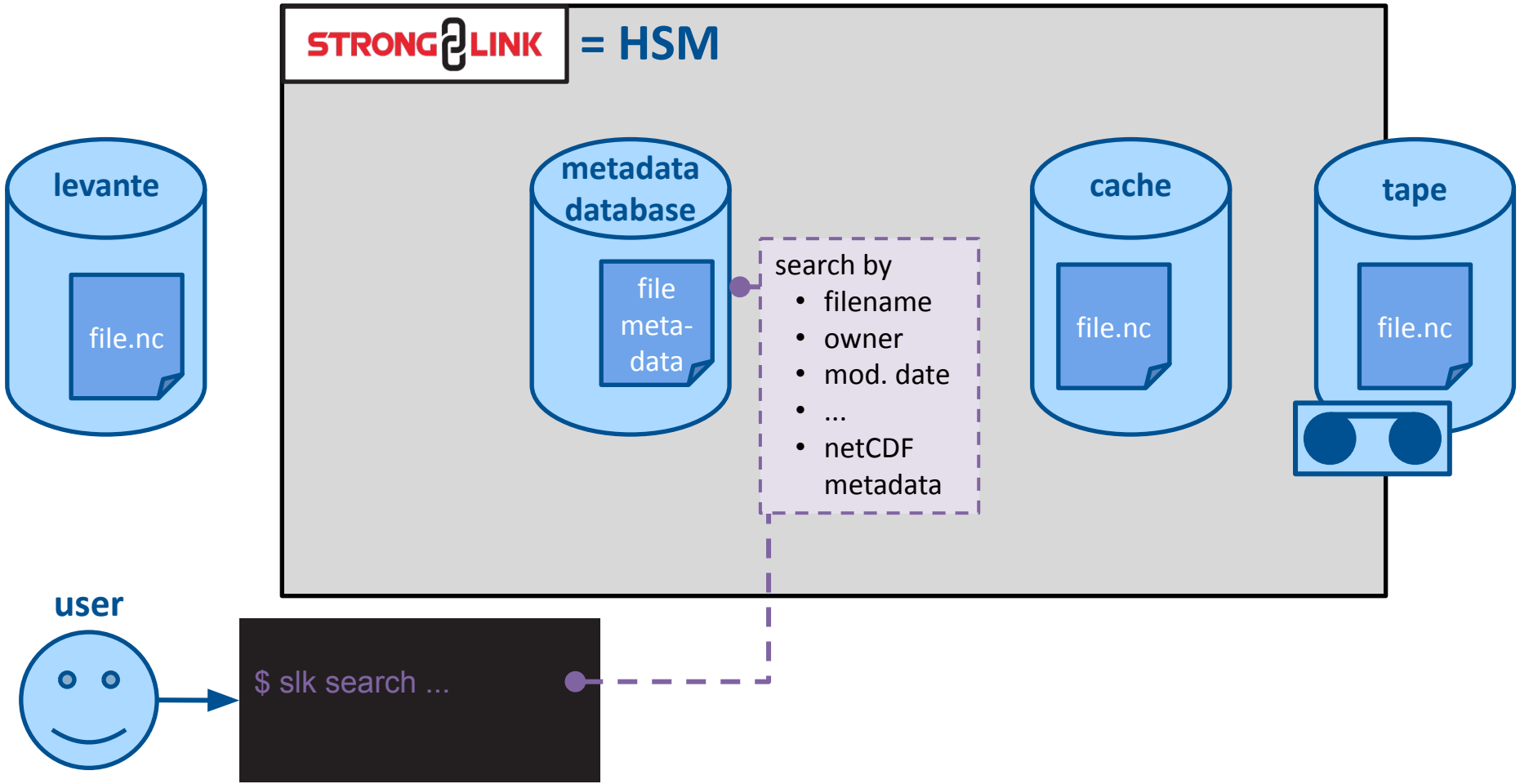
Structure HSM and slk command line interface



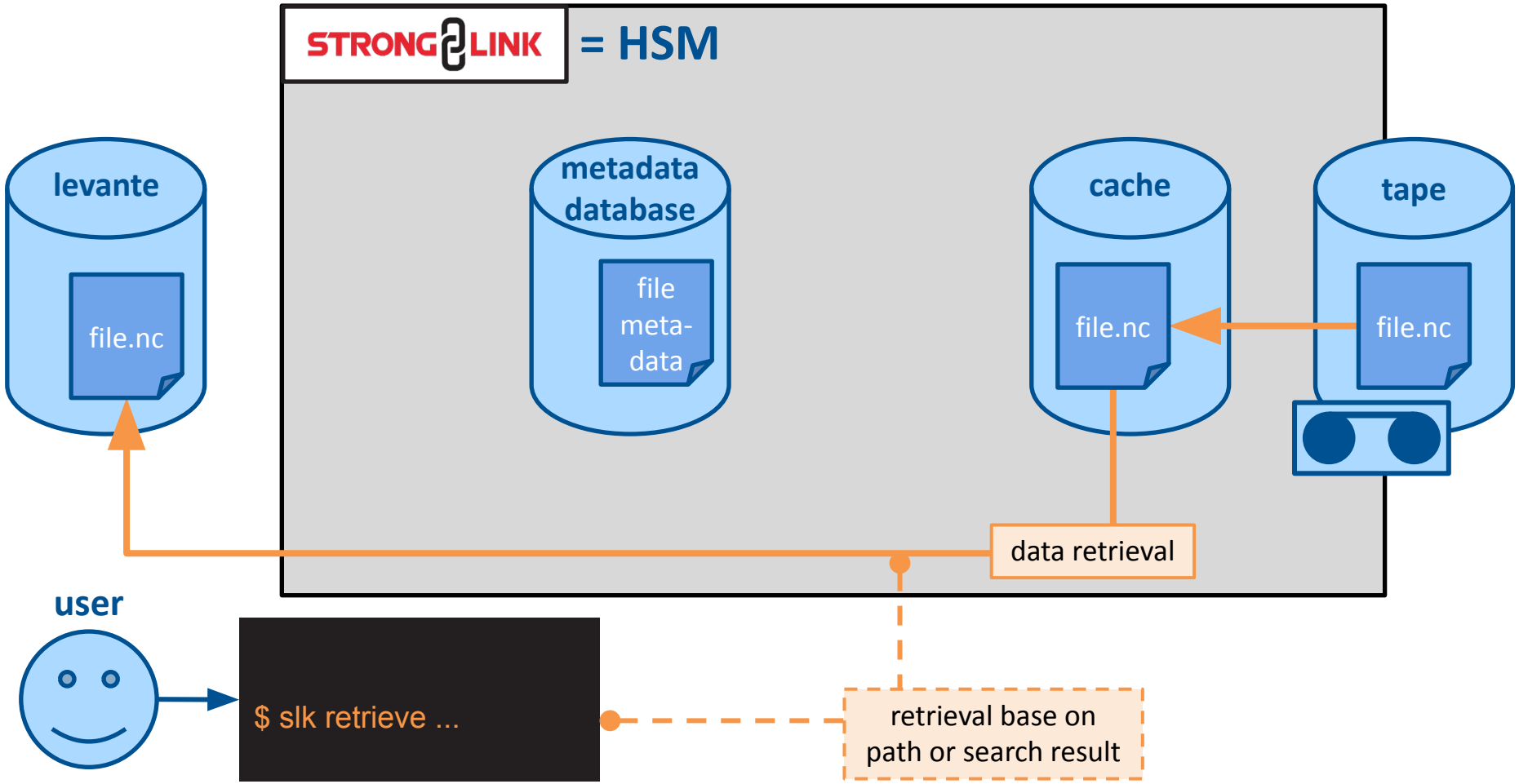
Structure HSM and slk command line interface



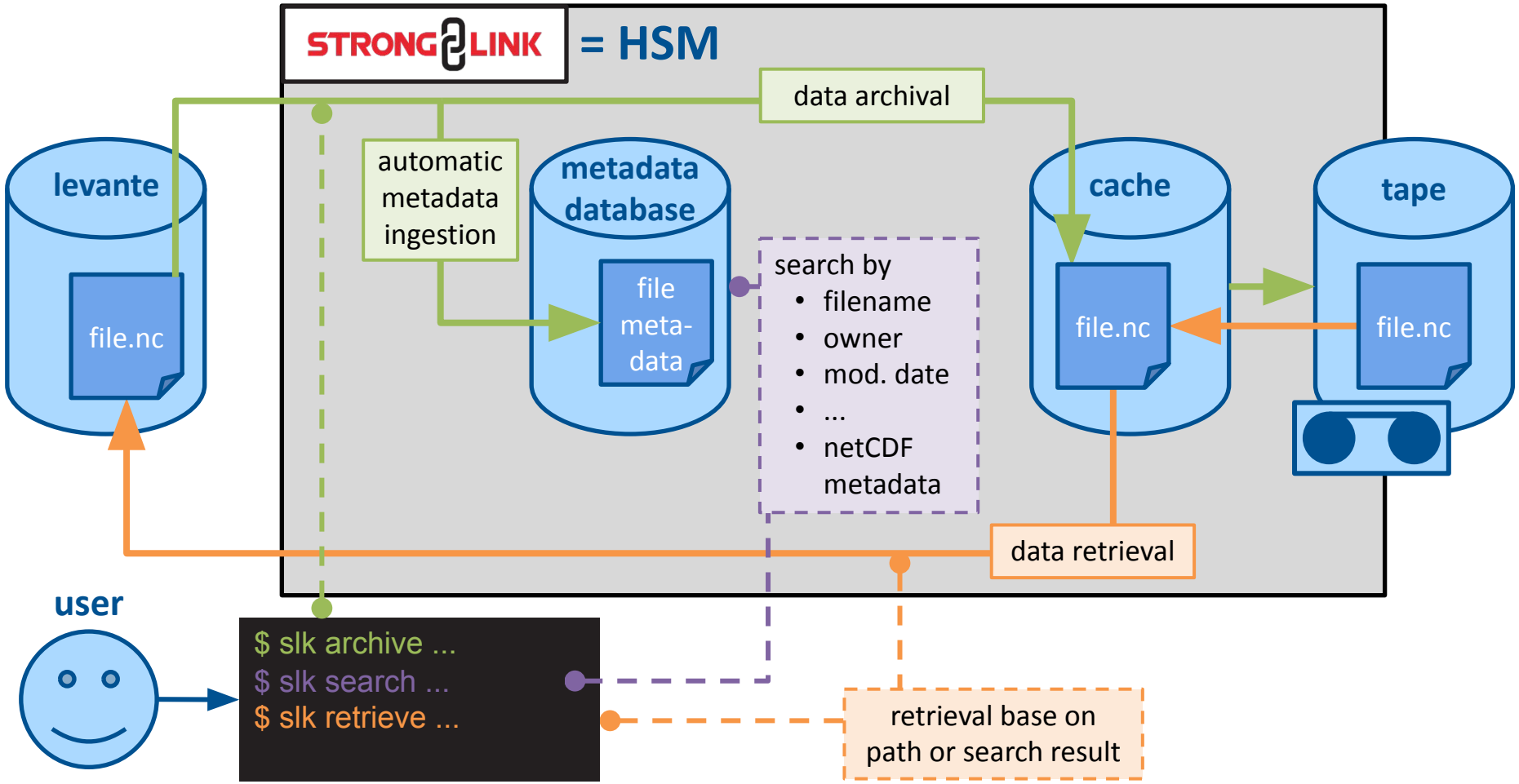
Structure HSM and slk command line interface



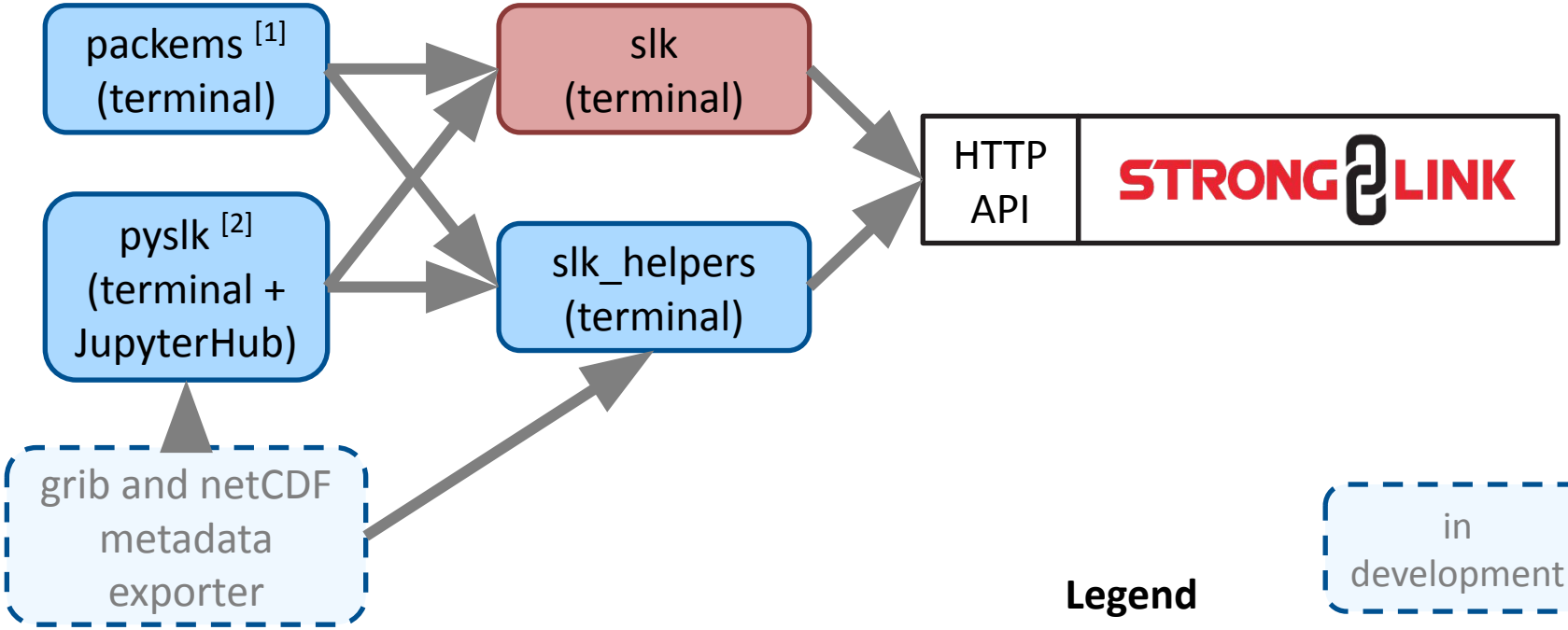
Structure HSM and slk command line interface



Structure HSM and slk command line interface



HSM tools



[1] by Kalle Wieners (MPI-M; packems) and DKRZ (listems + unpackems)
[2] DKRZ with contributions from GERICS and hereon

HSM support

- Documentation: docs.dkrz.de
- FAQ: docs.dkrz.de/doc/datastorage/hsm/hsm_faq.html
- issues / questions support@dkrz.de

Hands-On I

archival, retrieval and search

Note: Please start your Laptops and login to Levante

A few questions!

- Did you already use slk?
- Do you experience with netCDF files?
- Do you mainly use grib files?
- Which other data file formats do you use? Do you expect to archive files with this format?

Material

- We will work with slk v3.3.64 (default on levante v3.3.21) and slk_helpers v1.5.1:

```
$ module load slk/.3.3.64-gcc-11.2.0
```

- Training data (levante):

```
/work/bm0146/k204206/material_hsm_workshop
```

- Sandbox in HSM:

```
/workshop
```

Changelog:

slk: https://docs.dkrz.de/blog/2022/changelog-slk-3-3-61.html
slk_helpers: https://docs.dkrz.de/doc/datastorage/hsm/slk_helpers.html#major-changes

slk – first steps

- Login once every 30 days (currently):

```
$ module load slk/.3.3.64-gcc-11.2.0

$ slk login
Username: k204221
Password:
Login Successful

$ slk_helpers mkdir /workshop/$USER
```

Today, we use slk retrieve on the levante login nodes.
Commonly, slk retrieve is deactivated there.
Please use the interactive partition instead:
https://docs.dkrz.de/doc/datastorage/hsm/retrievals.html#interactive-retrieval

Hands-On I

- See workshop_tasks.md

Presentation II

Metadata in StrongLink

Metadata in StrongLink - Motivation

If metadata are properly filled we can ask StrongLink to provide us all files that

- were created:
  - at institute ABCD,
  - with model EFGH,
  - in project IJKL
- and contain variable MNOP.

useful for automated collection of files
e.g. via intake and Freva

Structure of metadata in StrongLink

- Metadata fields grouped into schemes
- each schema reflects metadata of similar file types:
  - *image* (JPG, PNG, …)
  - *netcdf_header* (netCDF3, netCDF4)
- one field may exist multiple times in different schemata
  - „*Title*“ in schema „*document*“ und in schema „*netcdf*“
  - If *document.title* is updated, *netcdf.title* is not automatically updated

Which metadata schemata are useful for us?

- *netcdf*: a general short metadata schema meant for different file formats (also grib) [1]
- *netcdf_header*:
  approx. 100 global attributes from netcdf files [2]
- these schemata are automatically filled when a netCDF file is archived [3]

[1] https://docs.dkrz.de/doc/datastorage/hsm/ref_metadata.html#schema-netcdf
[2] https://docs.dkrz.de/doc/datastorage/hsm/ref_metadata.html#schema-netcdf-header
[3] all global attributes with „_“ and „-“ in name are skipped currently

Metadata Tools

- slk search
  search files based on metadata
- slk tag
  set individual metadata fields
- slk tag -display
  print metadata (JSON), bug in slk 3.3.61!
- slk_helpers metadata
  print metadata (human readable)
- slk_helpers json2hsm
  write metadata from JSON files into the HSM
- slk_helpers hsm2json
  write metadata from HSM into JSON files

Hands-On II

metadata

Note: Please start your Laptops and login to Levante

Hands-On II: Metadata

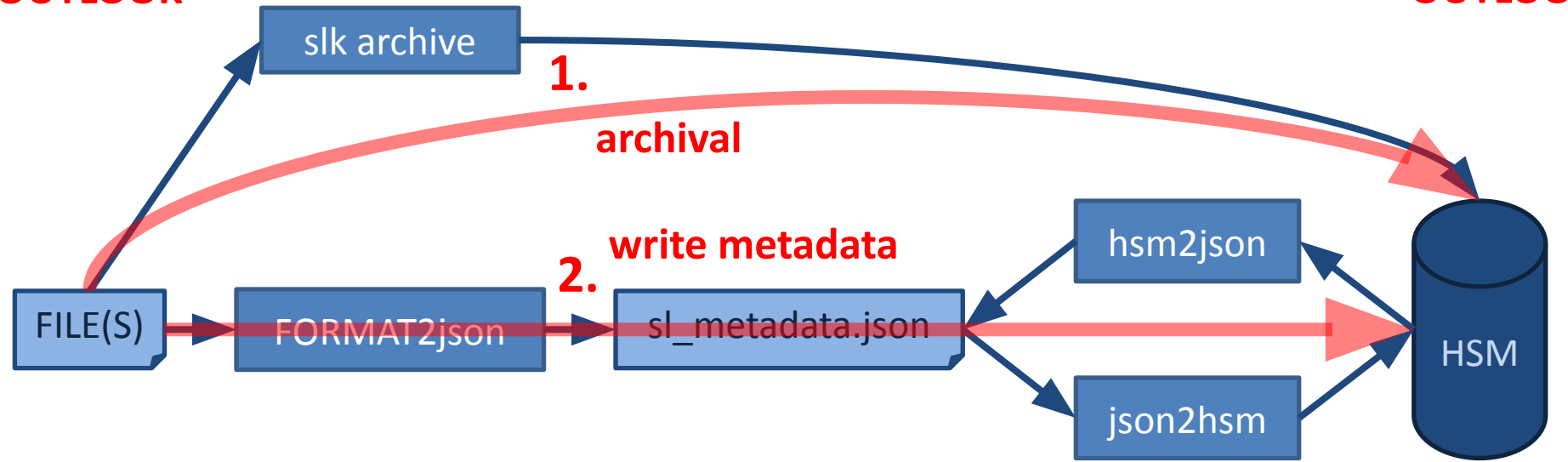
- See workshop_tasks.md

Outlook

for metadata extensions

OUTLOOK

OUTLOOK



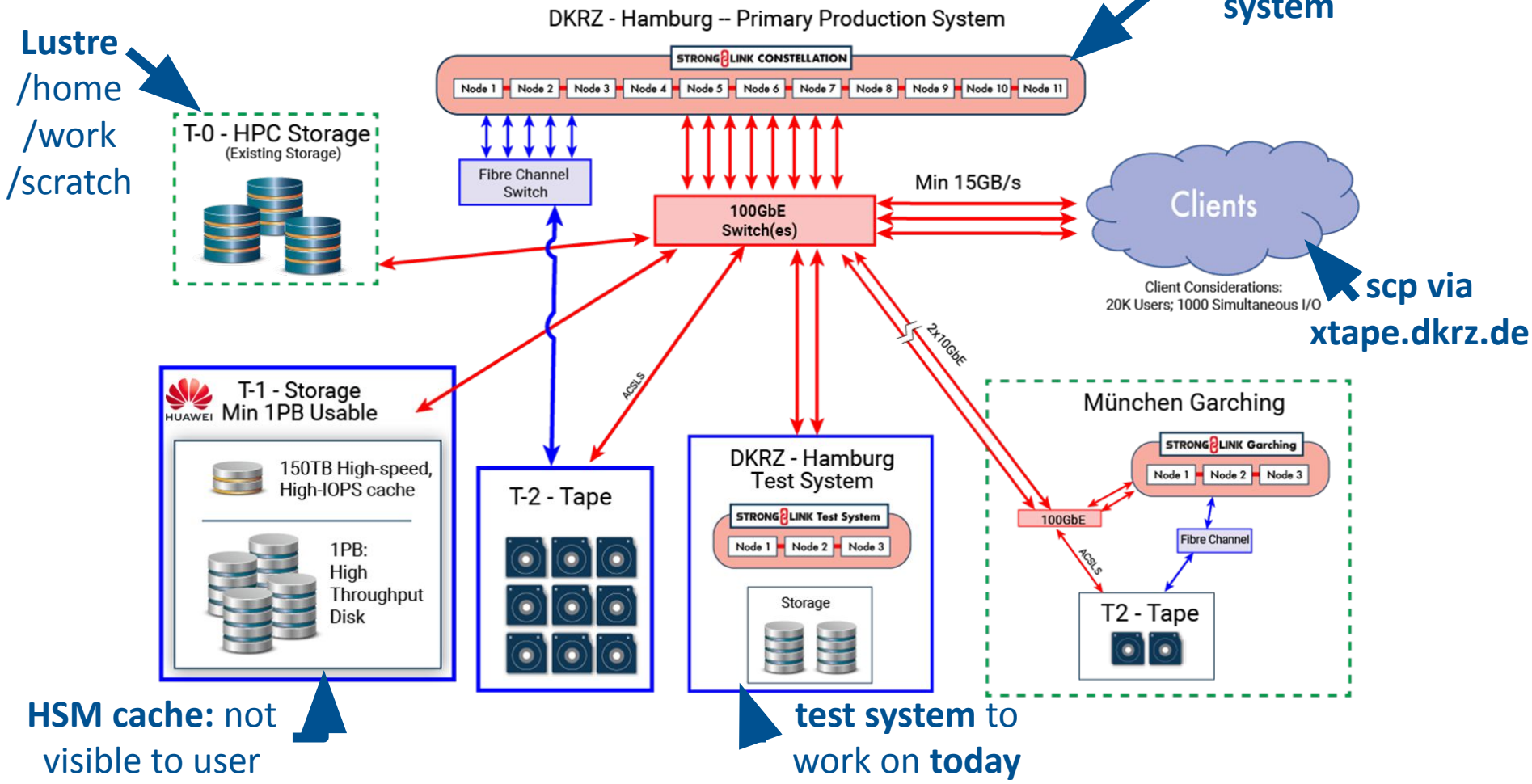
Use Case:

OUTLOOK

OUTLOOK

Q&A

The End

Backup

Detailed hardware setup