

Tools integration challenges & possible solutions

Recording, processing, copying data from a location to the next

⇒ Potentially many operations involving a non-negligible number of different tools

⇒ Time-consuming, risk of errors (deviations compared to what one wants to be as reproducible as possible)

Integration of the individual tools into one “user-friendly” system can streamline the process (and hopefully reduce the user’s workload)

eln.epfl.ch approach:

- Collect data as early as possible (if possible, directly from the instruments)
- Store everything in a database in a way that makes the data easy to retrieve for future use
- Provide modules to perform many different analyses, adding more upon request – normally using the user’s browser environment for execution (i.e. they use your machine’s processing power)

Tools integration : the RENKU approach

<https://datascience.ch/renku/>

Developed by the Swiss Data Science Center, RENKU is a web-based environment designed to work with data and source code:

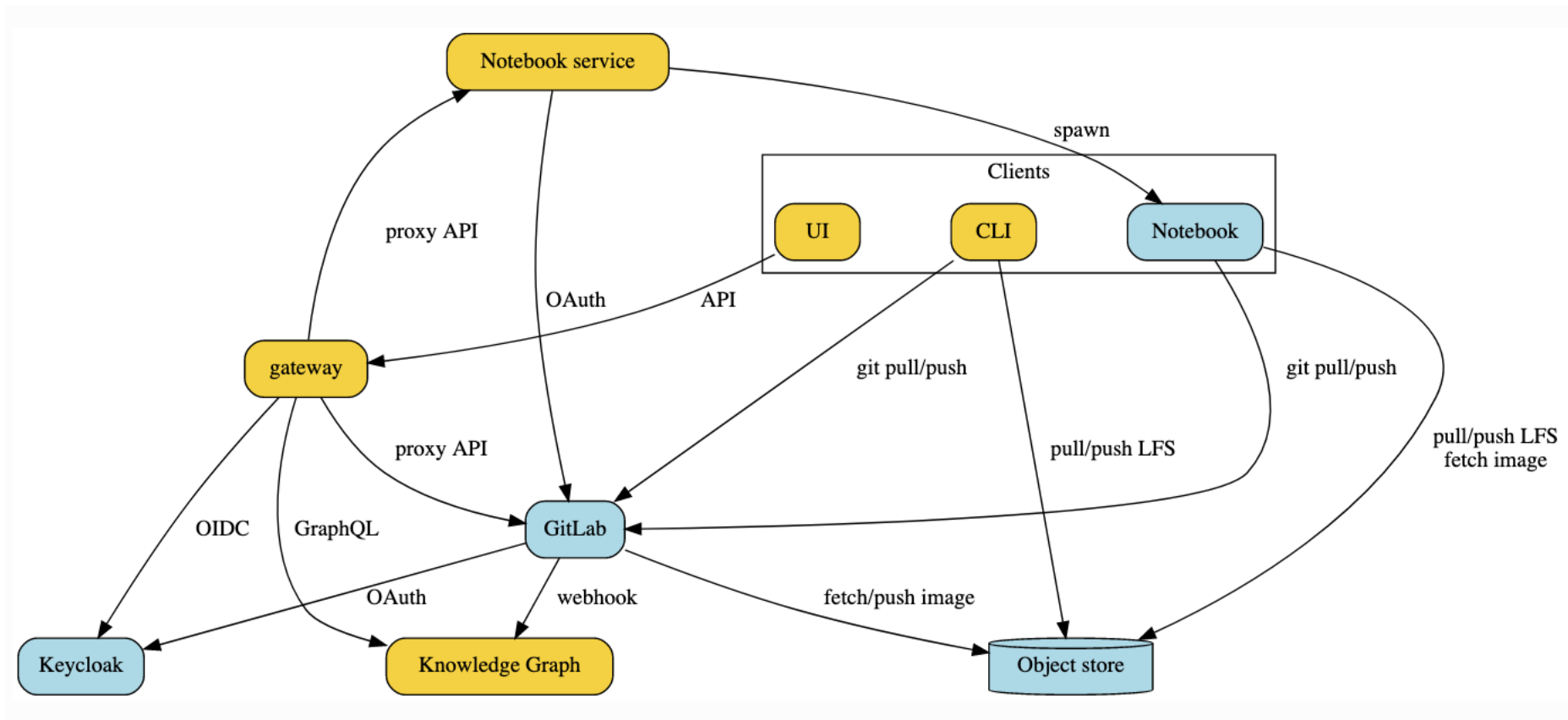
- Retrieve data hosted somewhere else
- Run the software in a controlled environment
- Save the results
- Keep track of evolving versions of the code and the data

It is similar to noto.epfl.ch (seen in the Tools 1 exercise) but more powerful (designed for professional research, not education)

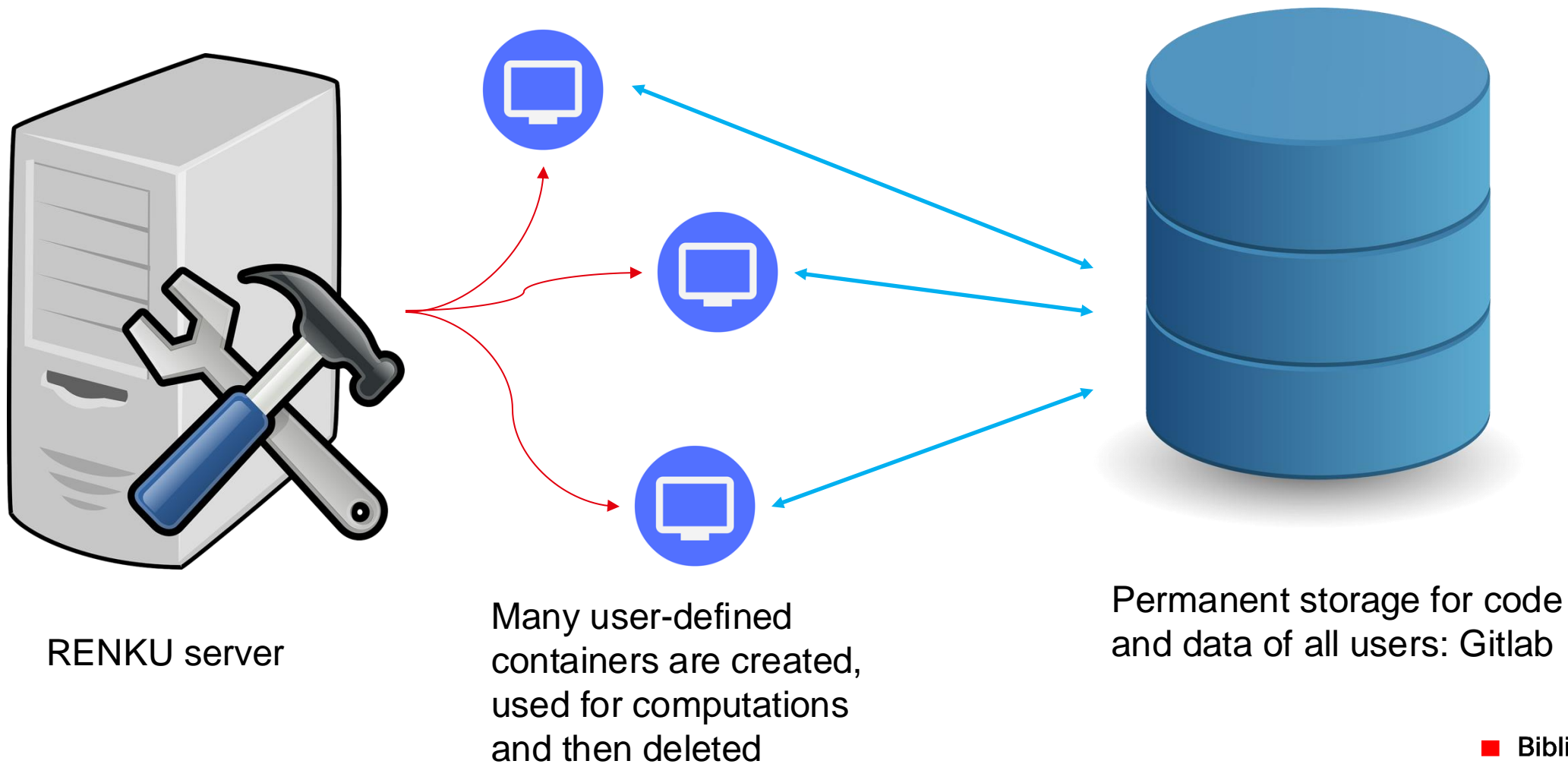
RENKU approach:

- Handle data as is, without assuming much about it
- Provide a platform where the user's software will be executed in disposable *containers* (virtual machines that only exist as long as needed and will be erased after use to release their resources to other users)
- Use the Git versioning system for code and data

Renku architecture: the official version



Renku architecture: a simplified view?



Demo process (not the most efficient but easier to understand)

1. Start the container on Renkulab.io
2. Start the Python3 launcher
3. Open a terminal and clone the data+code repository using git.
DATA IS COPIED INTO THE CONTAINER
4. Run the nmr_test notebook.
SOME DATA IS PROCESSED IN MEMORY AND DISPLAYED
5. (not shown because I didn't write the code)
THE NOTEBOOK SAVES THE PROCESSED DATA (still inside the container)

Then the processed data could be copied to another location (for example using git), or the code could be updated and a new version saved to the git repository...