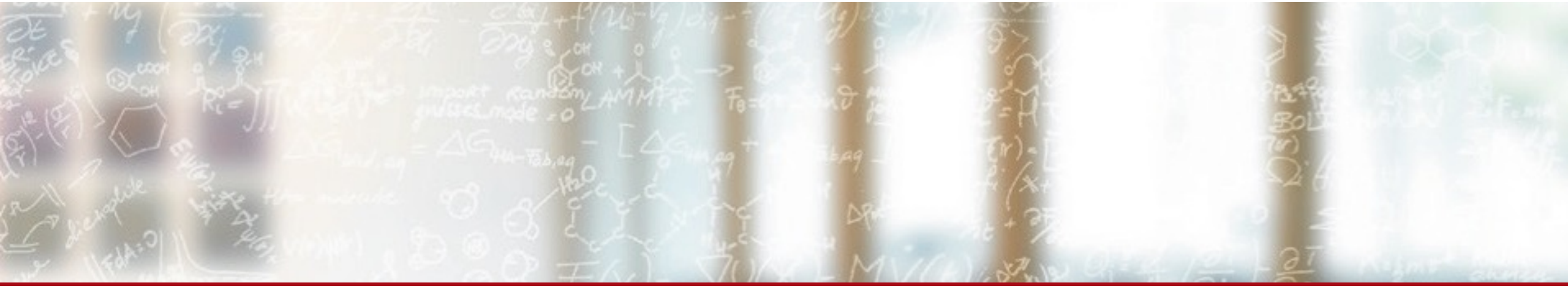
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich



# Long Term Storage webinar

23th June 2021
Webinar
Giuseppe Lo Re

# Background

- Started in Nov 2019
- Available for early users from May 2020
- Generally available from Jan 2021
- Team
  - Giuseppe Lo Re: coordination, development, CICD
  - Andrea Ceriani: API design, development
  - Alessandro Prato: development
  - Stefano Schuppli: message broker, development




# Pricing

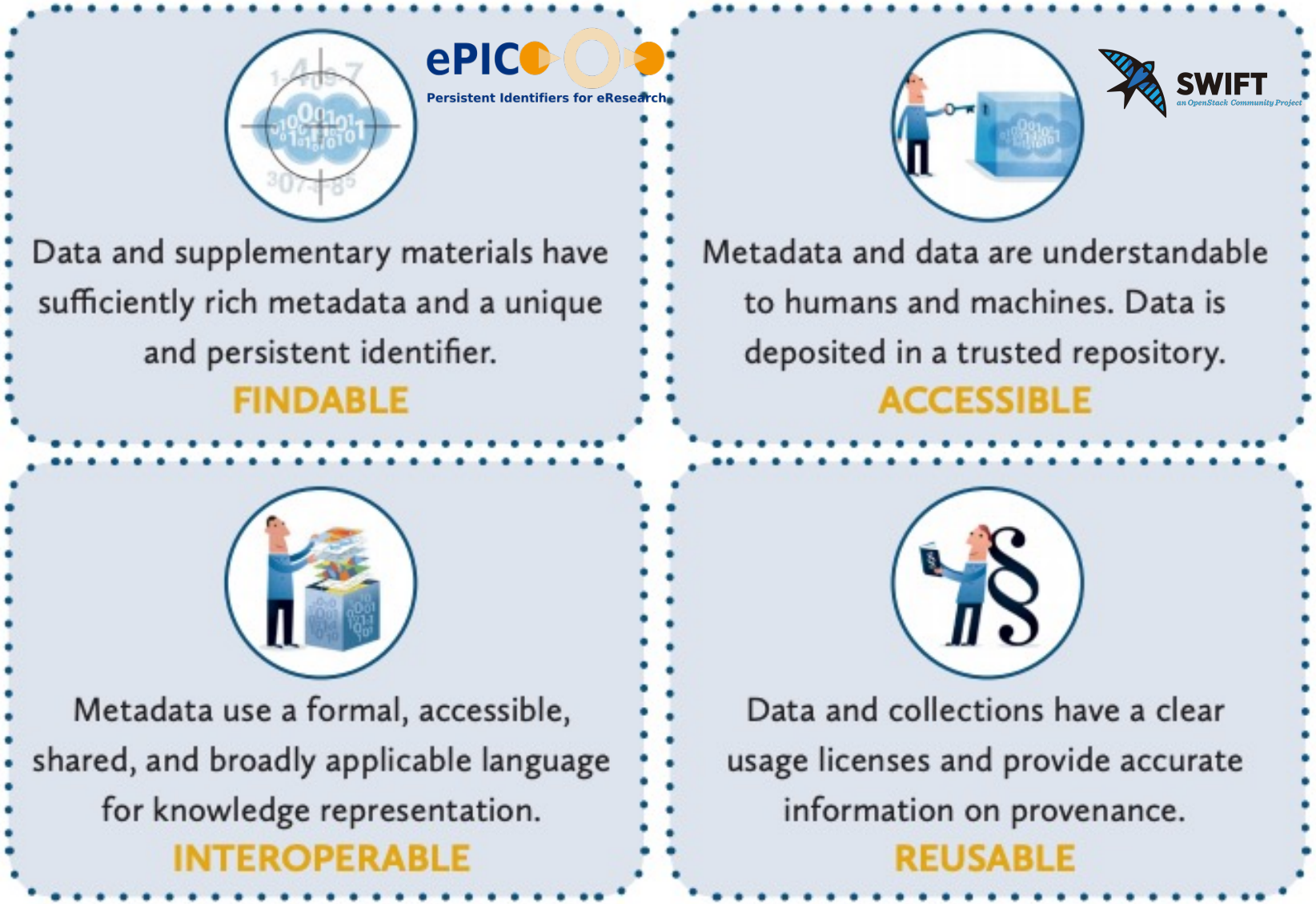
As of 2021:

- Users from the free User Lab program are entitled to use 2 TB of LTS storage quota (for 10 years) free of charge **per project**
- **Currently** additional space can be purchased for CHF 600.- for each Terabyte (for 10 years)

# FAIR and CSCS data services

There are 2 easy translations from FAIR principles to existing CSCS services:

- Accessible ➔ Object Store Service (Swift)
- Findable ➔ PID Service
- Reusable ➔ started to address it with creative commons licenses



# Requirements

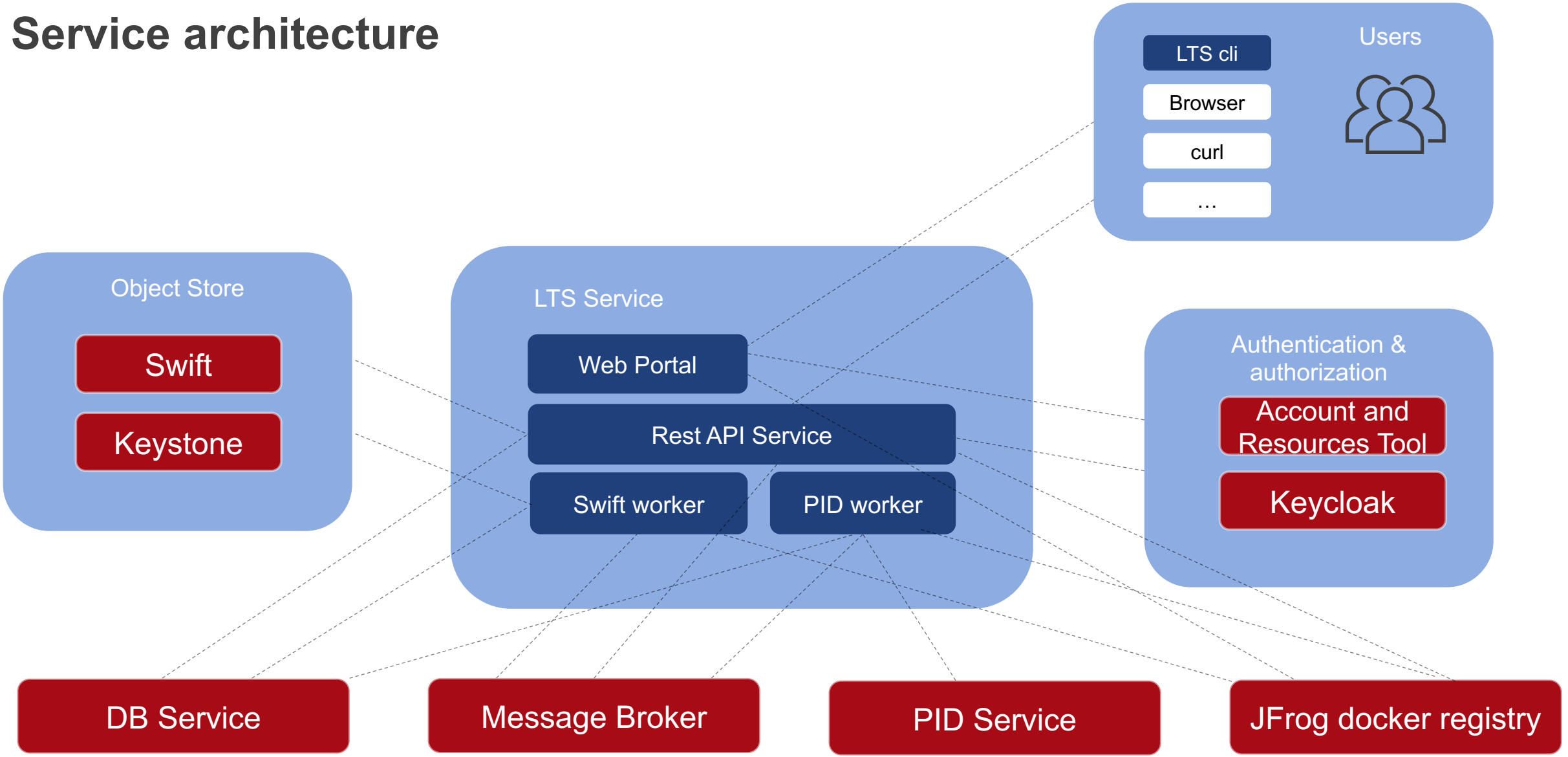
The Long Term Storage (LTS) service will provide:

- Storage repository with long term retention capabilities (10 years)
- Persistent Identifiers
- Ability to allow public access to data
- Data accessible via http
- Rest API interface to make easy the integration with third party applications and portals
- Web Portal to provide a friendly GUI
- Scalability, to cope with large volumes of data
- Data protection to ensure resiliency against hardware/software failures

# LTS concepts

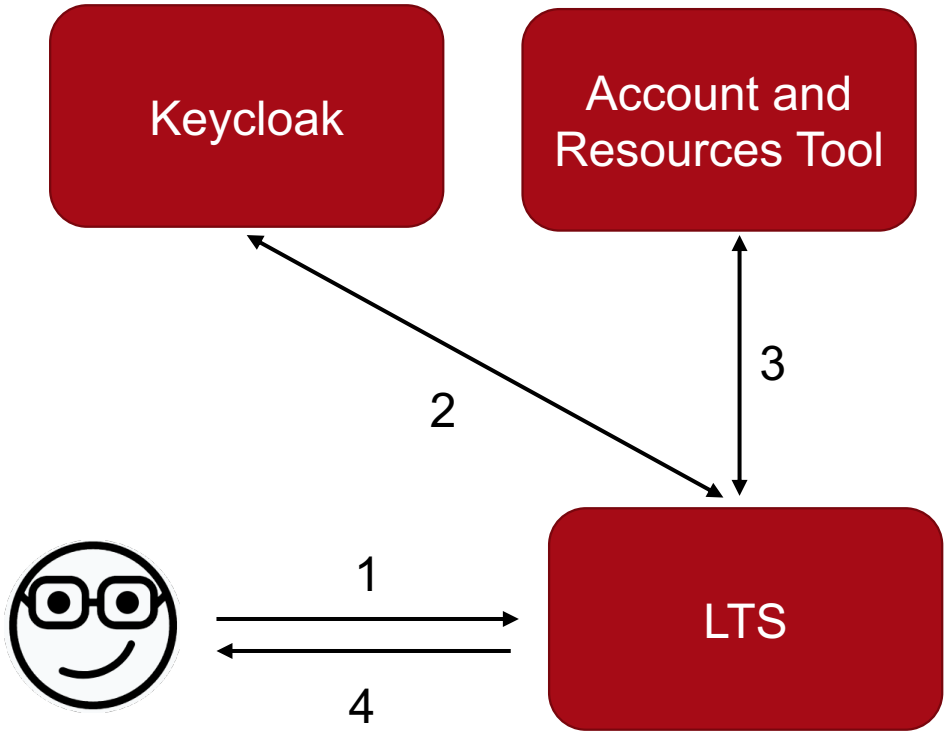
- The storage unit is a data collection
  - A data collection is composed by one or more files
- The data collection is immutable
  - Objects in the data collection are read only
- A single PID handle is used to reference the collection
  - The handle contains a list of URLs to reference multiple objects
  - The handle contains arbitrary meta data attributes to describe/enrich the data objects
- CSCS must guarantee that the PID content is consistent with the data collection
  - No direct data access for the owner, all changes will go through the LTS API
  - CSCS will perform all the needed updates to the PID to ensure consistency in case of data migration
  - LTS will provide only CSCS PID handles (handles from PID prefix managed by other sites are not allowed)

# Service architecture



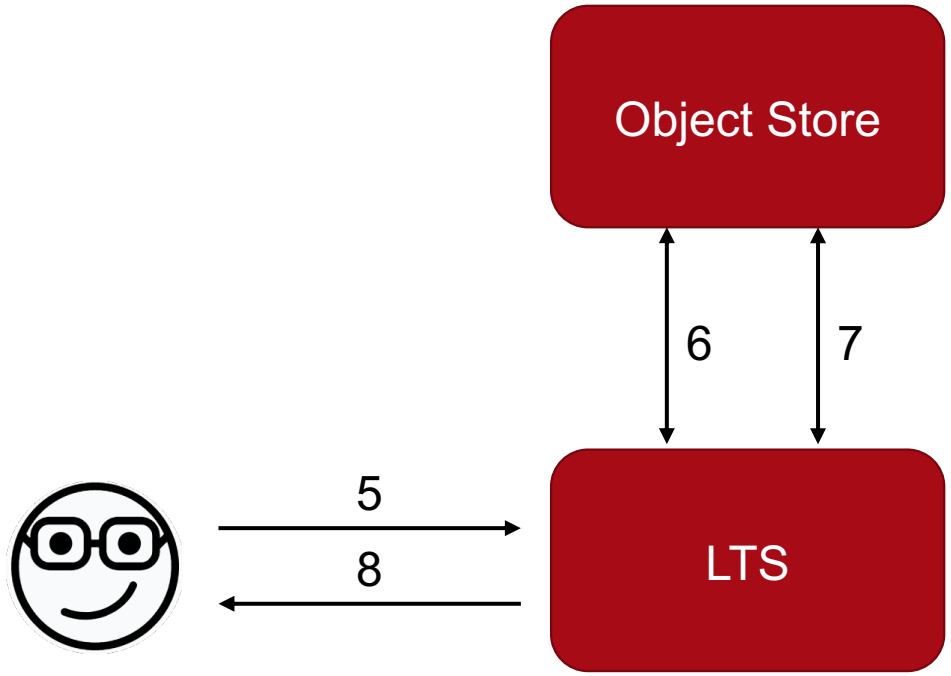
# LTS Workflow

1. The user logs in into the LTS portal
2. User credential validation
3. User authorization validation
4. The LTS service complete the its verification and the user can start using the service



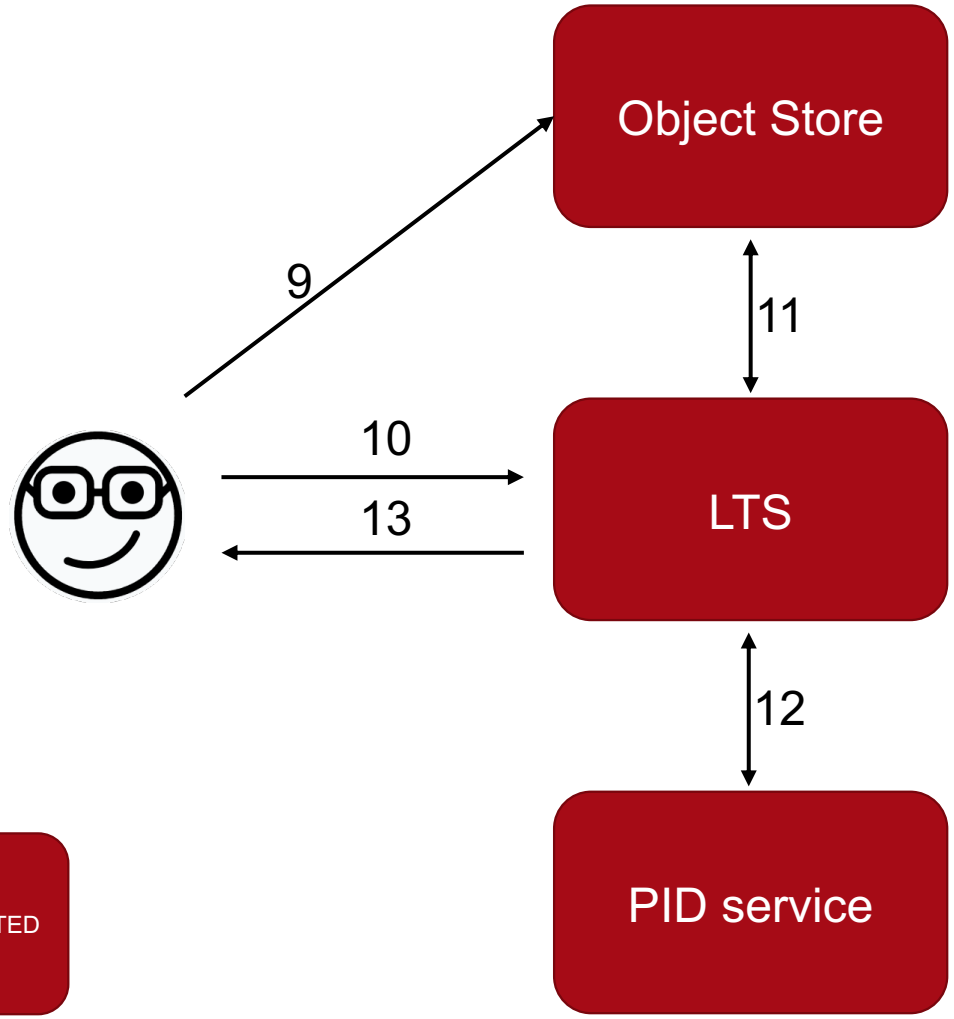

# LTS Workflow

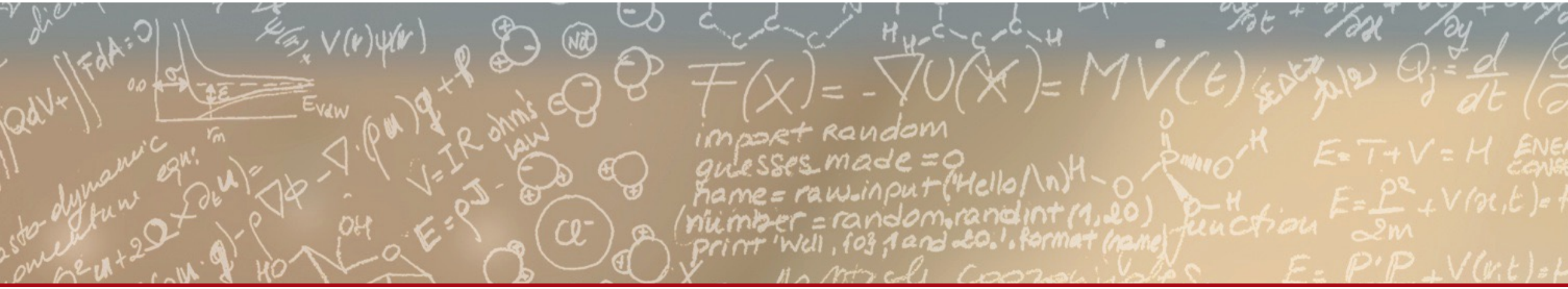
5. The users makes data collection create request specifying the following info
   - Data collection name
   - Data collection description
   - Collection metadata
   - List of objects with their checksums
6. The LTS service prepare an Object Store container for the collection
7. The LTS service generates upload URLs
8. LTS returns them to the users for the upload operation




# LTS Workflow

9. The user starts the upload operations
10. When the upload operations is done the user commits the collection
11. LTS starts the data collection validation:
    - Checksums before and after the download will have to match
12. LTS contacts the PID service and gets a new handle for the data collection.
13. The user follows the workflow polling the collection status, at the end he sees whether the collection was store successfully and its PID handle.




# Thank you for your attention.