



European Cancer Images Federation Software Architecture

Ignacio Blanquer (UPV), Miguel Caballer (UPV), Esther Bron (Erasmus MC), Marcel Koek (Erasmus MC),
Luis Martí-Bonmatí (HULAFE)

- EUCAIM in a nutshell
- User Roles
- Architecture of EUCAIM
 - Federated Catalogue
 - Federated Query
 - AAI
 - Access Negotiation
 - Distributed Processing
 - Central Storage
 - Other Components
- Conclusions

Outline

①

②

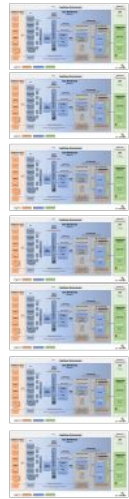
③



Hybrid Platform

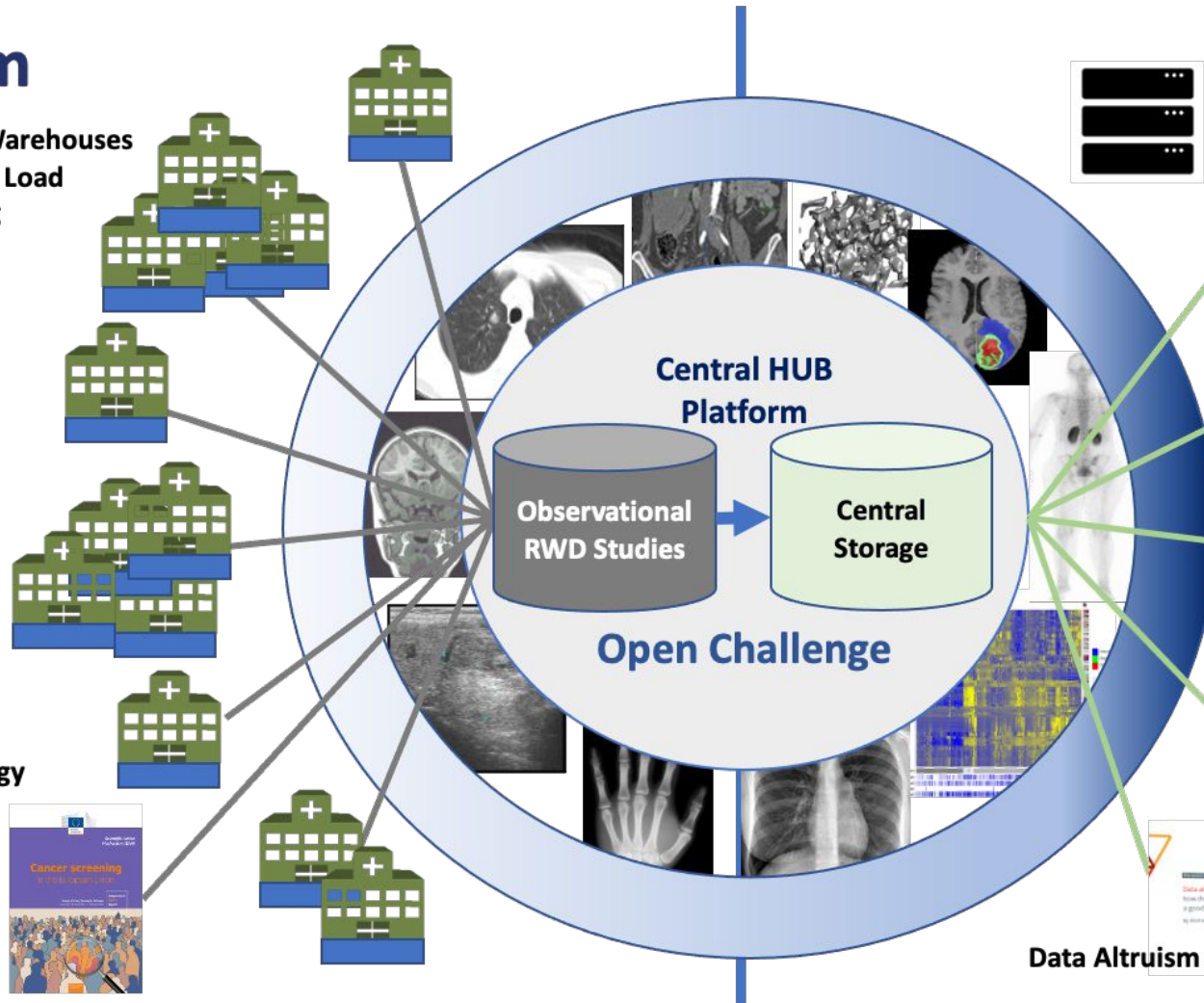
Distributed RW Data Warehouses
Extract, Transform and Load
ML Federated Learning

Primary & Secondary Used Area



CDM hyper-ontology

Cancer Screening Programs



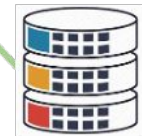
- Metadata Catalogue
- Annotated Structured Data
- AI Experimentation Platform



Atlas of Cancer Images



Secondary Use Area



DICOM-MIABIS



Data Altruism



Data Provider/Data Holder/Data Controller

Definition: Any natural or legal person, including entities, bodies, and research organisations in the health or care sectors, as well as European Union institutions, bodies, offices, and agencies, who has the right, obligation, or capability to make certain data available, including registering, providing, restricting access, or exchanging the data.



Two options for joining the federation:

- Become a federated node
- Upload anonymised data to the central storage.

Tool Provider

Definition: Entity (startups, enterprises, research institutions, government agencies, non-profit organisations) that would like to contribute with processing tools, services, or applications they have developed to the EUCAIM's marketplace for use in the federated processing module of the platform.



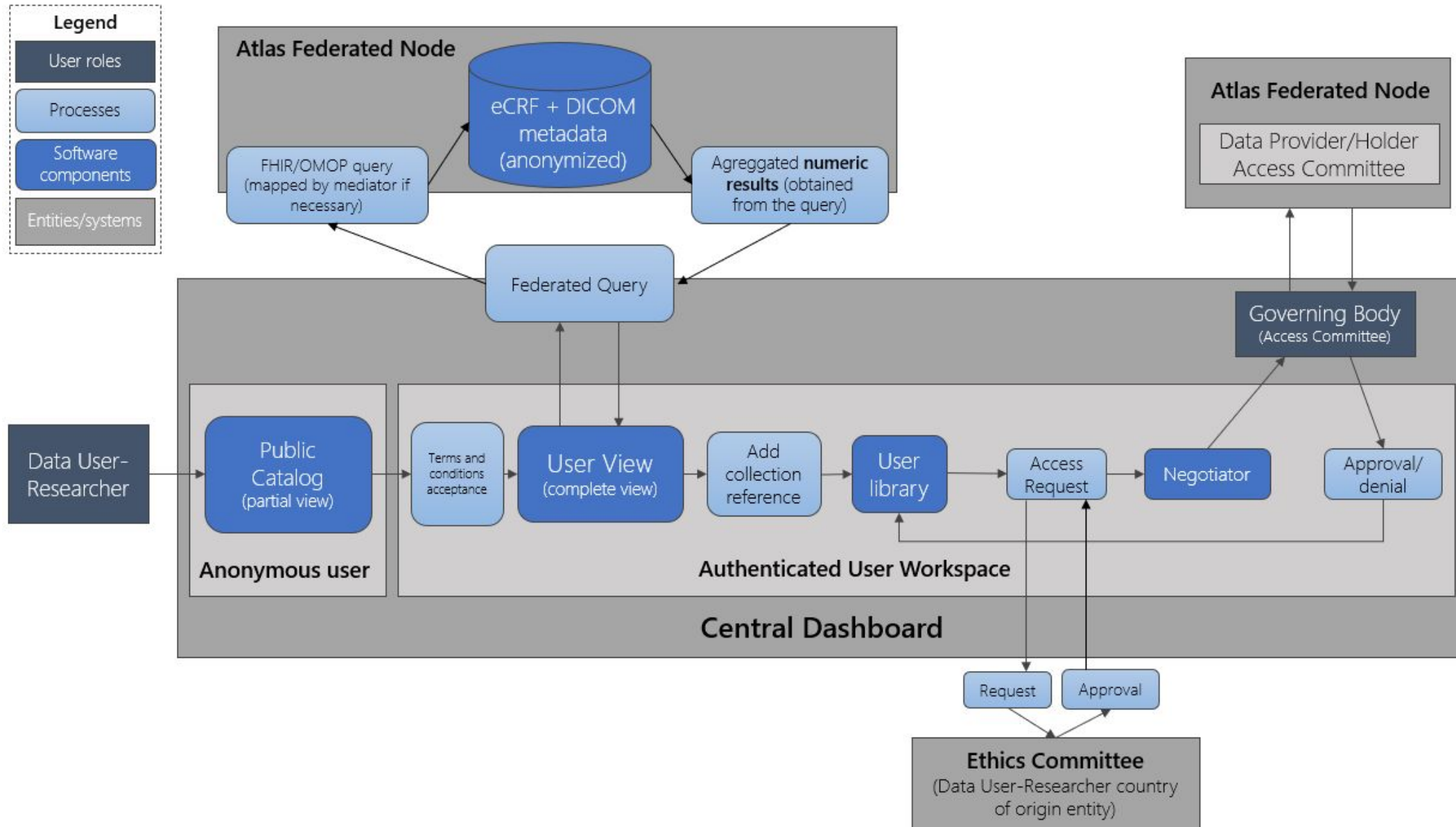
Both batch and interactive applications, following well-defined rules for participation and technical compliance guidelines.

Data User-Researcher

Definition: A person or entity that wants to explore the public catalogue and eventually request access to data and process them using either the tools available in the platform or their own AI tools to conduct studies, research, or analysis with the intention of generating new knowledge in the field of medicine and publishing the findings.

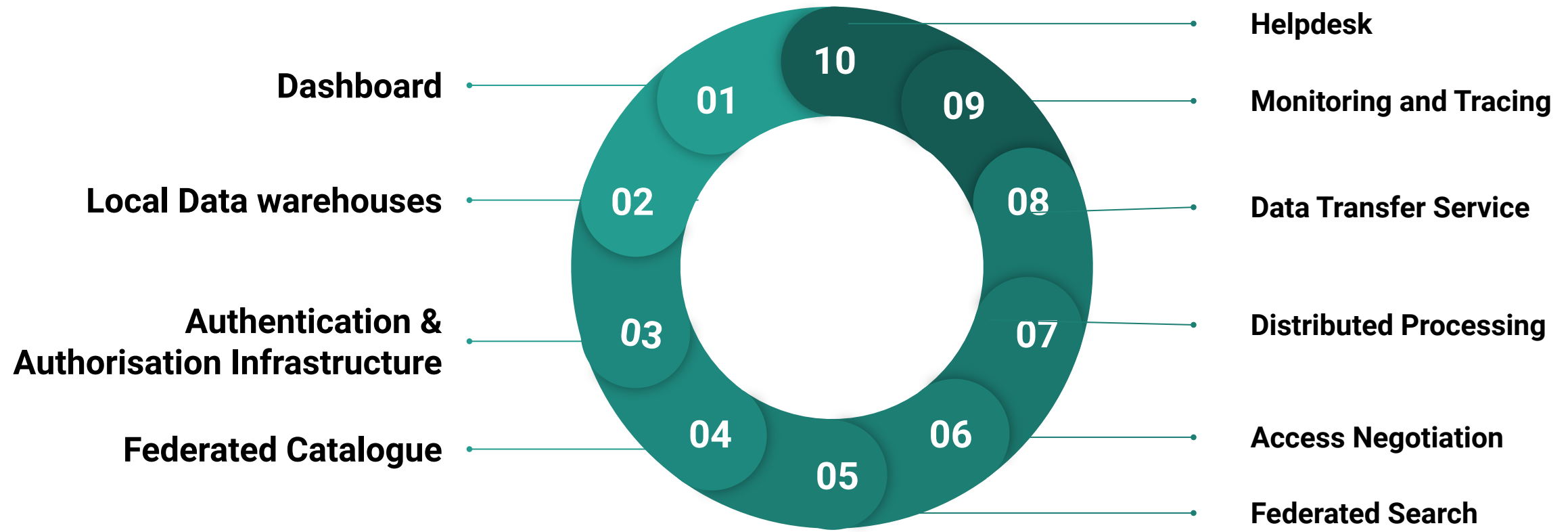


A data access request should be made through a Research and Development (R&D) project that will be evaluated by the Access Committee.

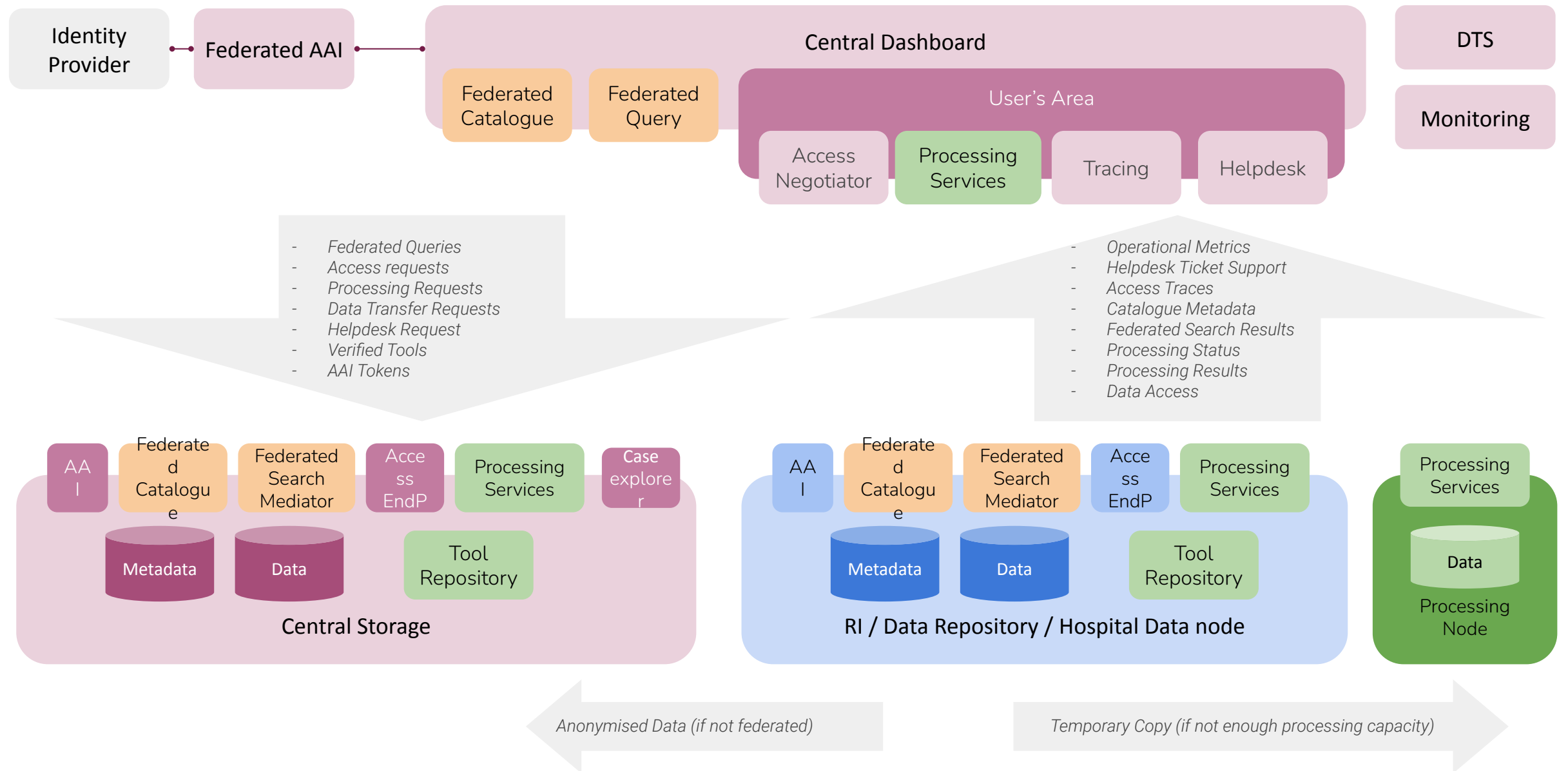


- Federated
 - It should provide services that discover, request, access and process data without requiring transferring the data out of the provider's borders.
- Extensible
 - It should be possible to extend to new providers at a reasonable effort.
- Open and Standardised
 - It should follow open standards for the specification of APIs, data and protocols.
- Efficient
 - It should be able to integrate high-performance resources for storage and processing
- Secure
 - Privacy, security and auditability by design.

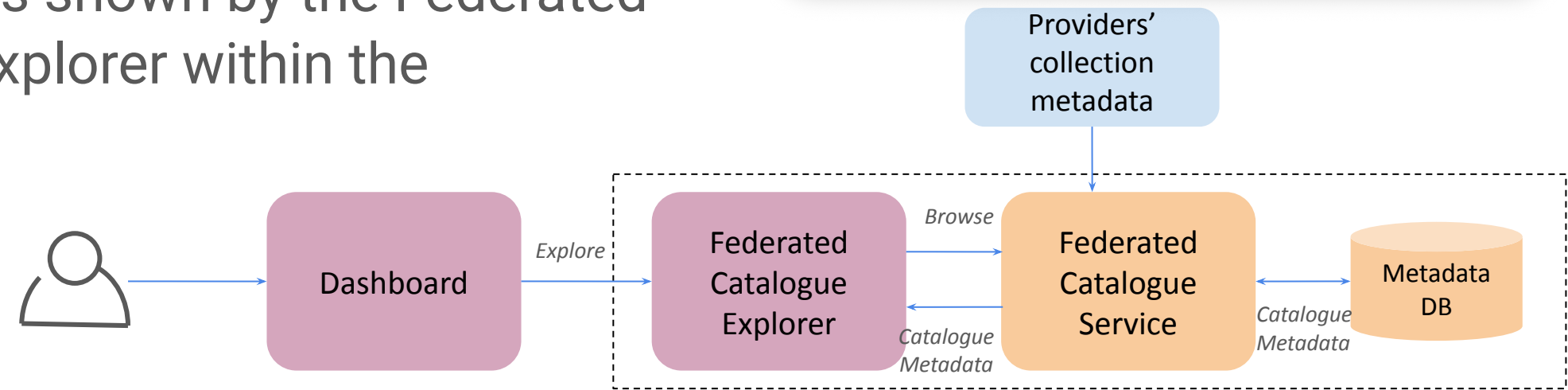
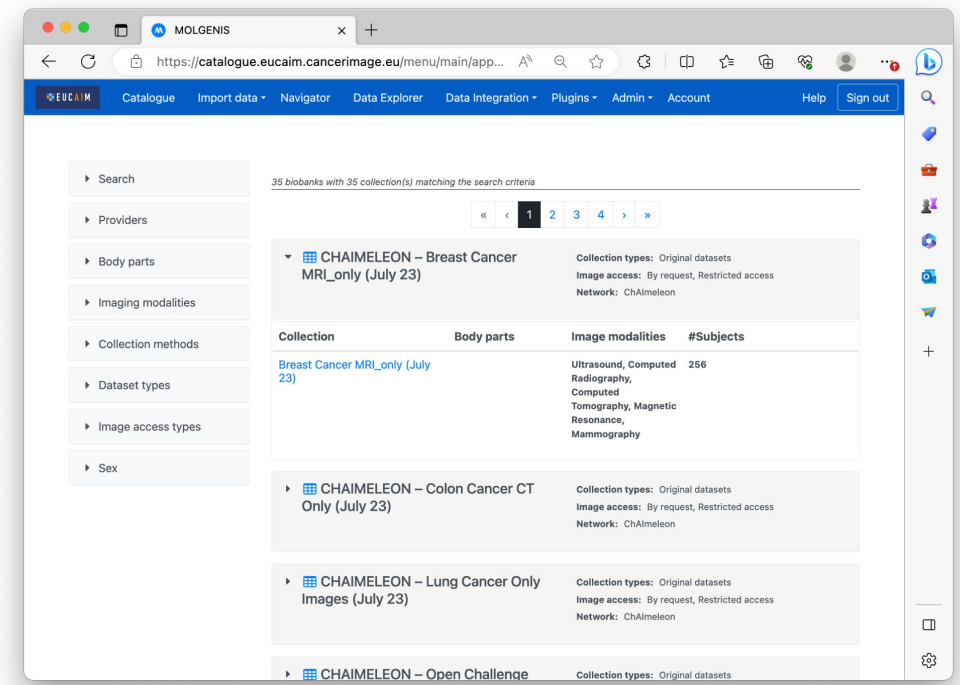




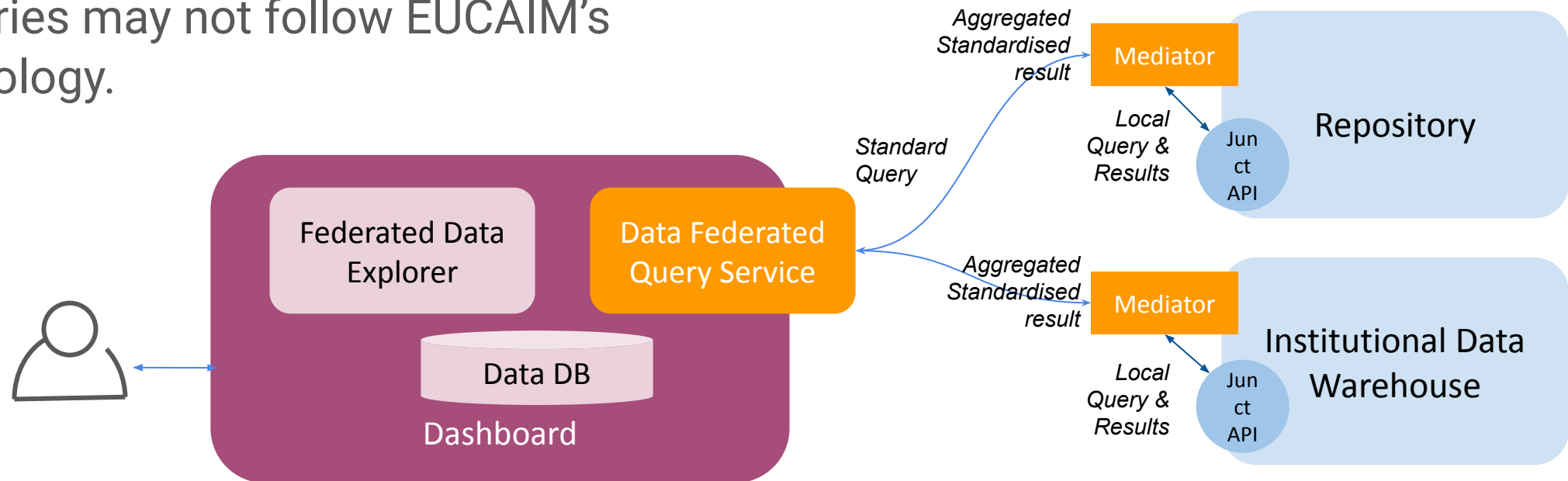
Overview of the Architecture



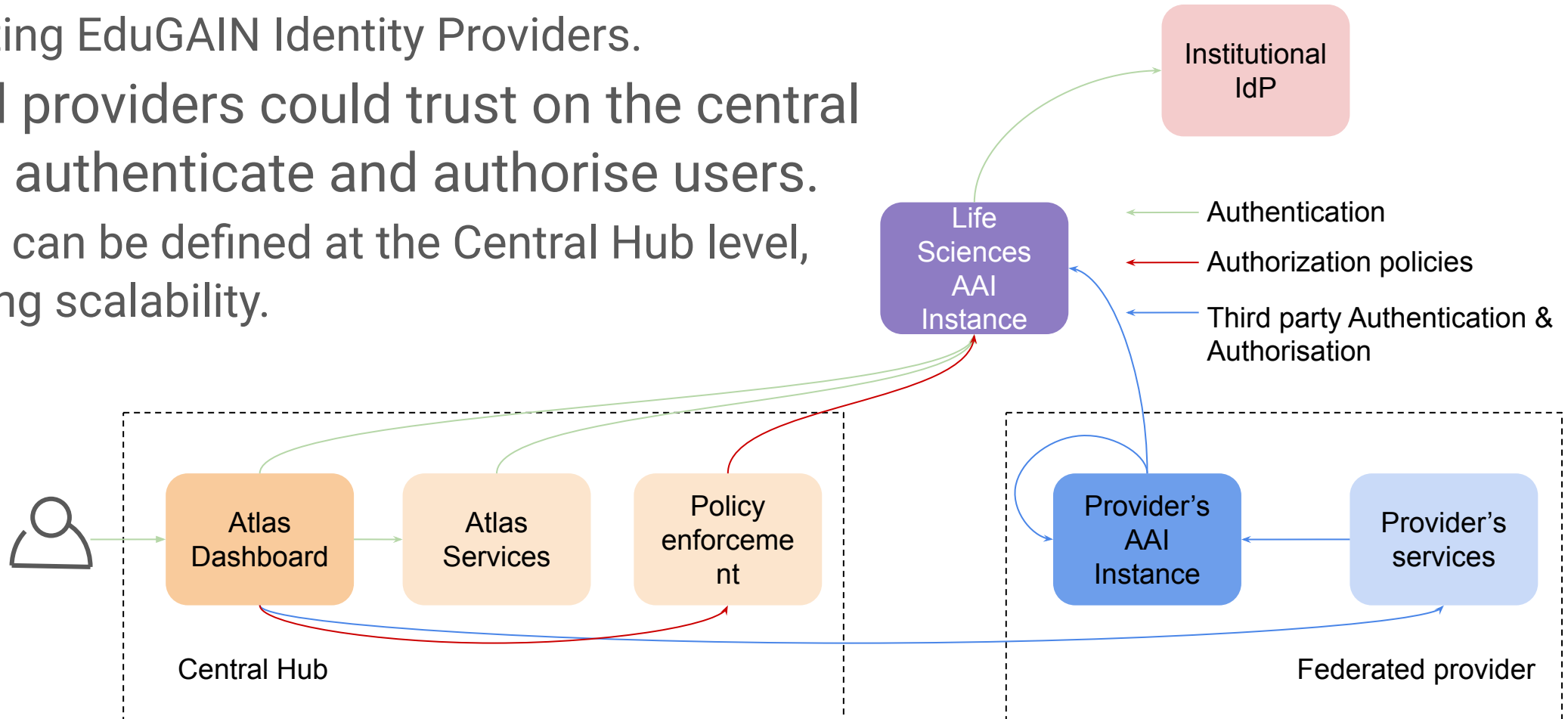
- A federated catalogue indexes the collections from the providers of the federation and the central storage.
- Collections' metadata is registered in the Federated Catalogue Service
 - Collections metadata follow a common metadata model.
- These data is shown by the Federated Catalogue explorer within the Dashboard.



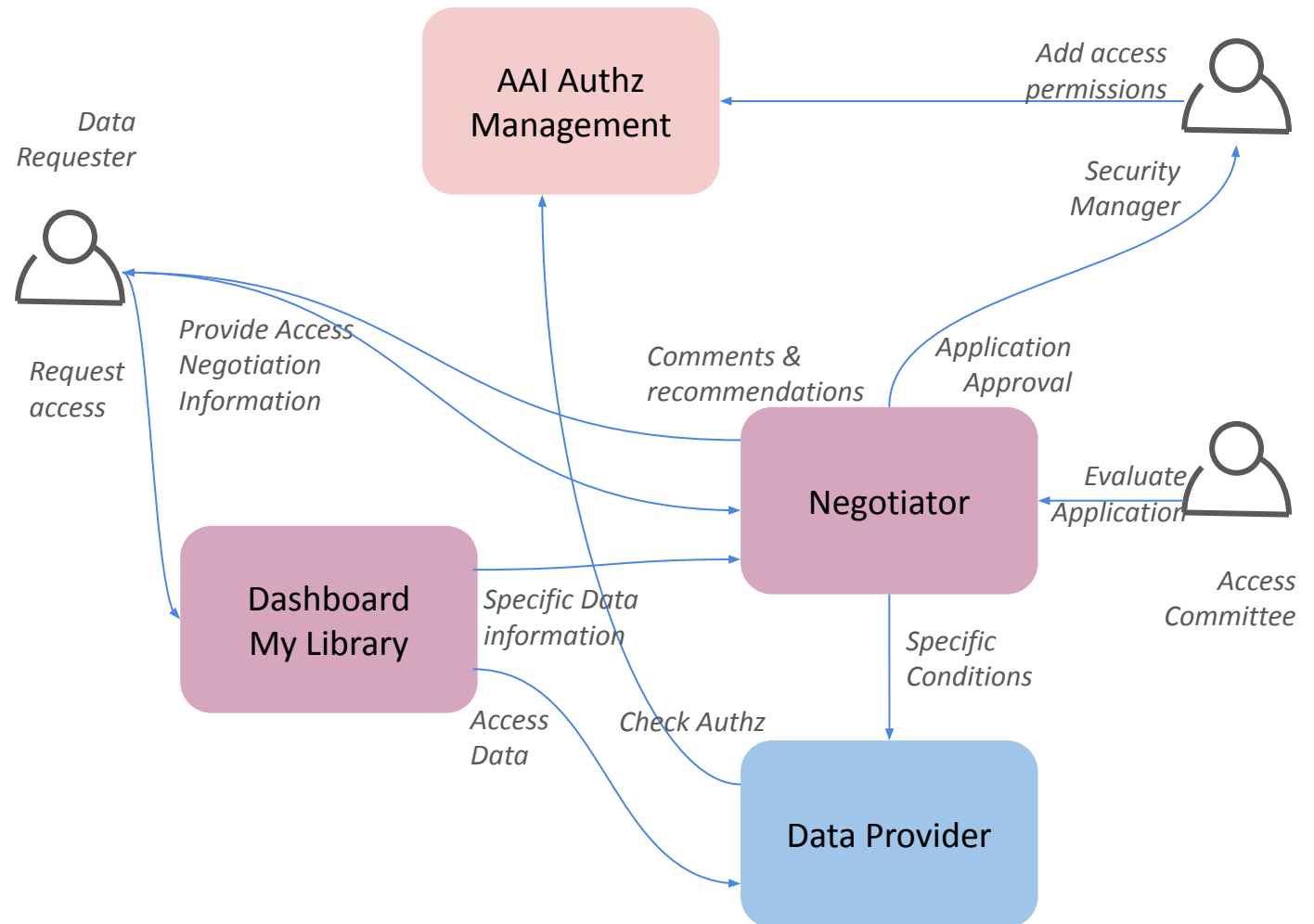
- A Federated Query service can retrieve the number of cases which fulfil a searching criterion
 - Searching criteria based on selected fields of the images metadata defined by the hyperontology (E.g. Body Part, Gender, Age range, ...)
- This will require that providers have a mediator service that adapts the query to the specific format of the provider
 - Repositories may not follow EUCAIM's hyperontology.



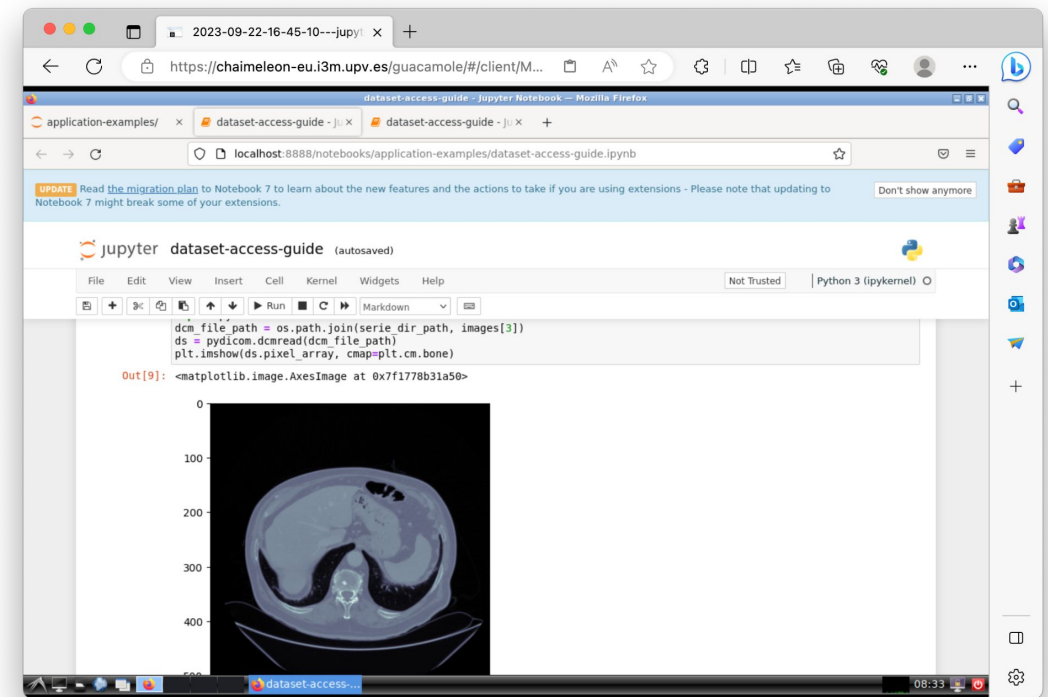
- A central AAI (Based on Life Sciences Login AAI) supports authentication & authorisation based on groups and roles.
 - Supporting EduGAIN Identity Providers.
- Federated providers could trust on the central service to authenticate and authorise users.
 - Policies can be defined at the Central Hub level, improving scalability.



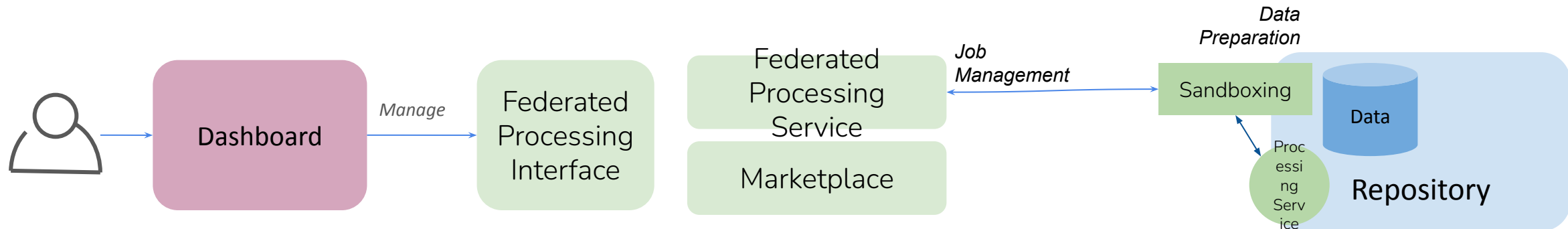
- Access to the collections will require submitting a project proposal, ethical evaluation and institutional mandates.
- Proposals will be evaluated by the Access Committee, involving the providers of the federation.
- The proposal could go through several iterations, until it gets accepted.
- Central authorisation is updated.



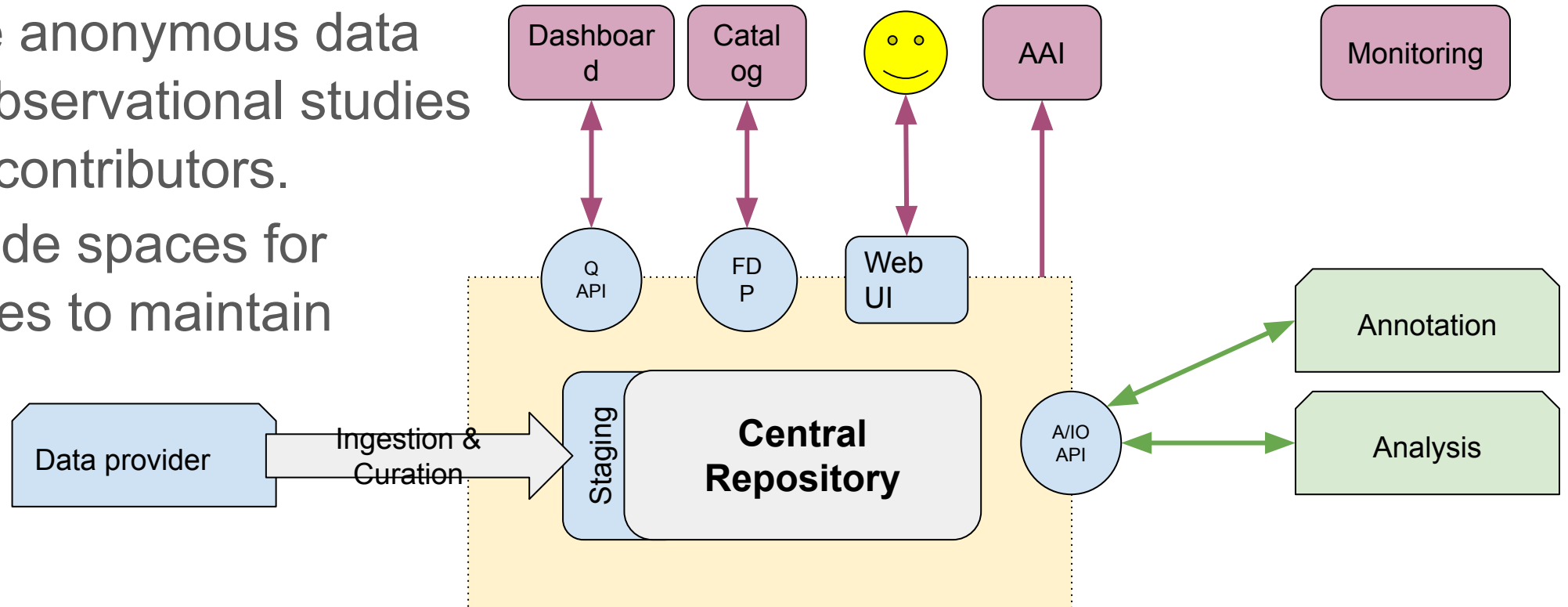
- Three access models are foreseen
 - Data can be downloadable to the requester premises
 - This is the weakest case in terms of traceability and will apply to public datasets.
 - Data cannot be downloadable but it can be accessed in a closed, trusted environment
 - The provider will expose a Virtual Research Environment to the user, who will explore, display and process the data “in-situ”.
 - Higher traceability, institutions should give permission to display data.
 - Data cannot be even visualized
 - The user can only run distributed/federated processing.
 - Higher data traceability, minimises risks by using tools from the marketplace only.



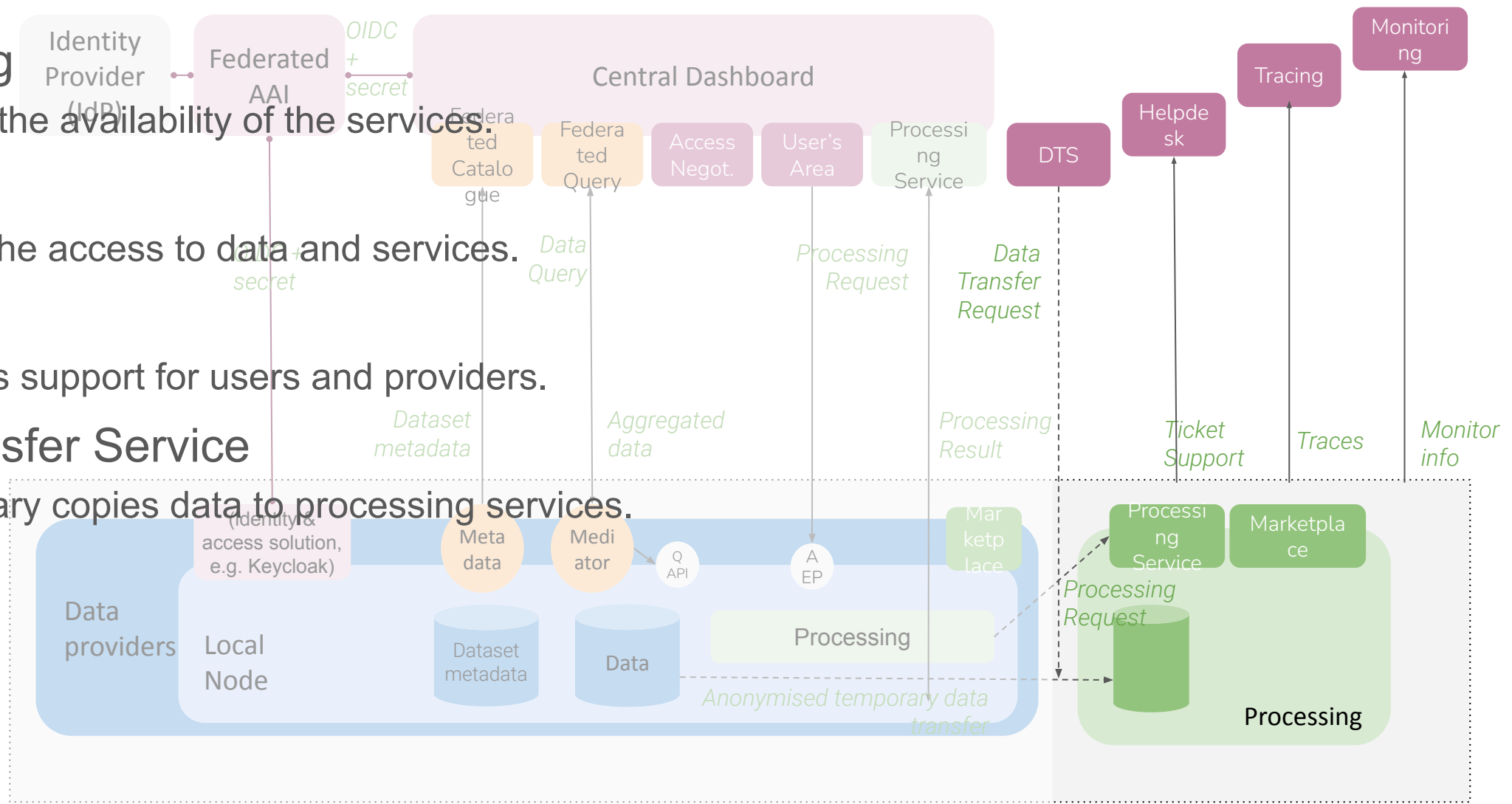
- Federated processing will trigger the execution of jobs on the providers of the federation
 - Eventually it will require temporary data transfer to a trusted processing service.
- Data must be prepared in a platform-agnostic way so processing tools can run on different providers.
 - Either copied or linked to a sandbox area.
 - Tools will be registered on a trusted central repository.



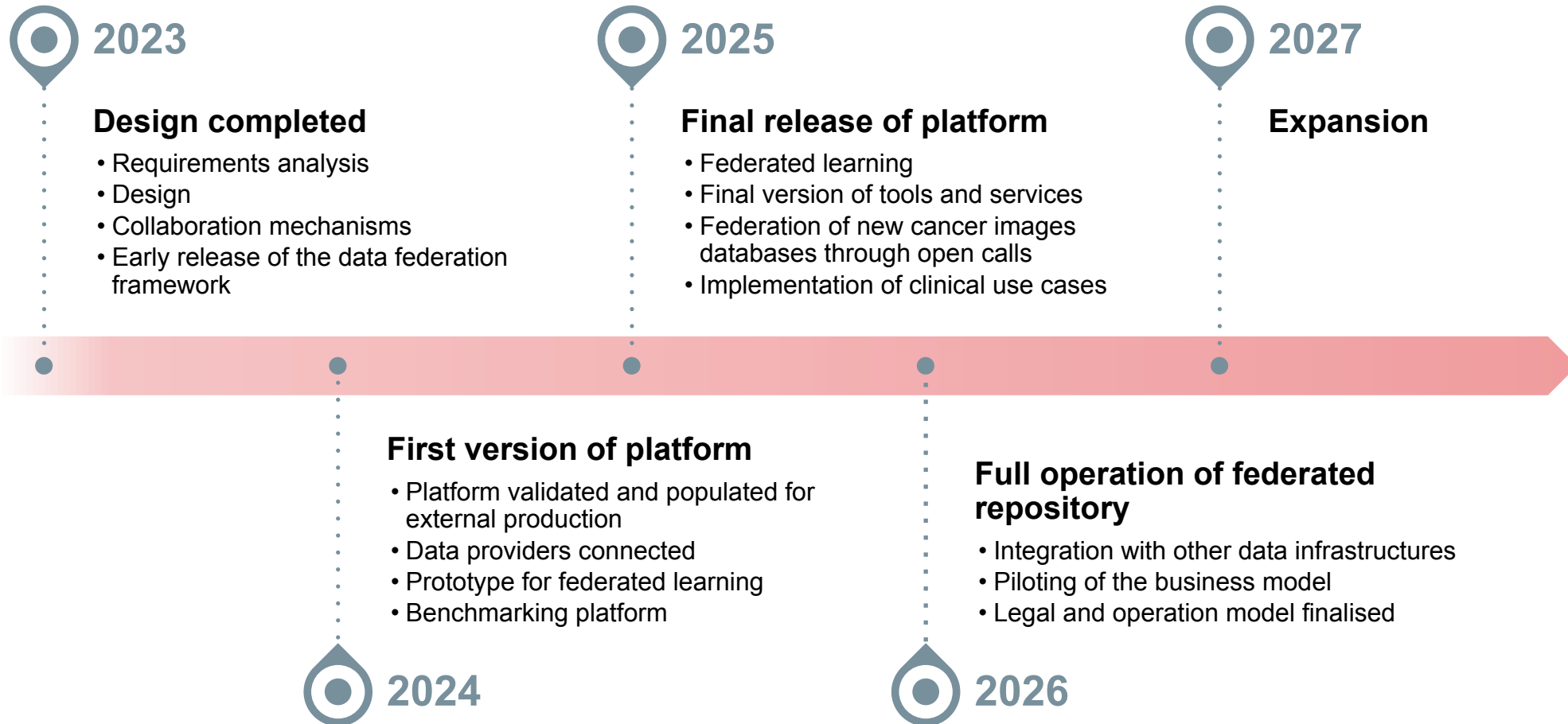
- The central storage will work technically as another node of the federation
 - It will provide storage, computing capacity and other services.
- It will store anonymous data from the observational studies and other contributors.
- It will provide spaces for communities to maintain their data.



- **Monitoring**
 - Verifies the availability of the services.
- **Tracing**
 - Tracks the access to data and services.
- **HelpDesk**
 - Provides support for users and providers.
- **Data Transfer Service**
 - Temporary copies data to processing services.



Thank you for your attention



Get in touch!

