Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

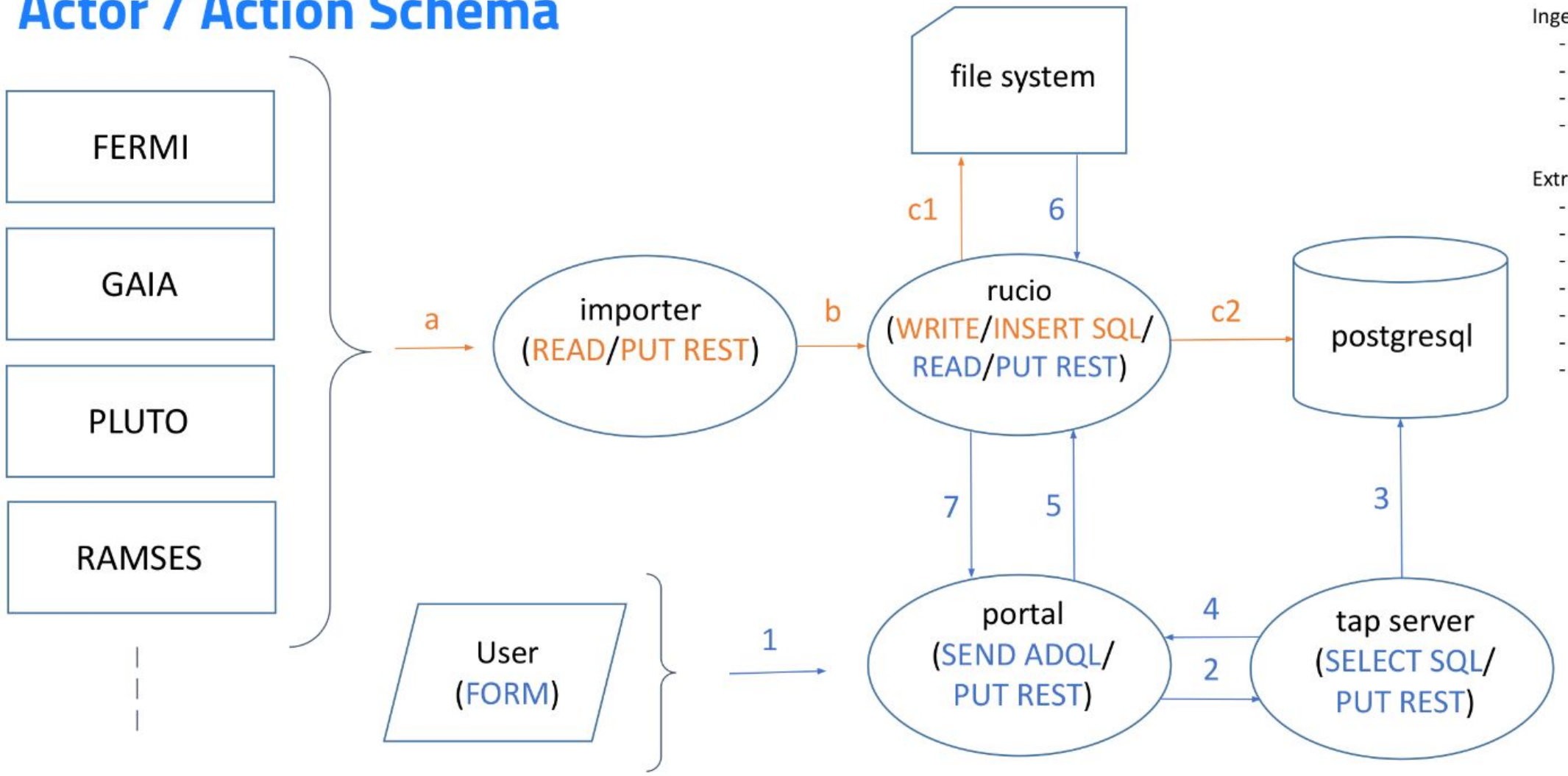
ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Introduction to metadata specification, data policy, archive interface presentation

Massimo Costantini

Spoke 3 General Meeting, Elba 5-9 / 05, 2024

Actor / Action Schema

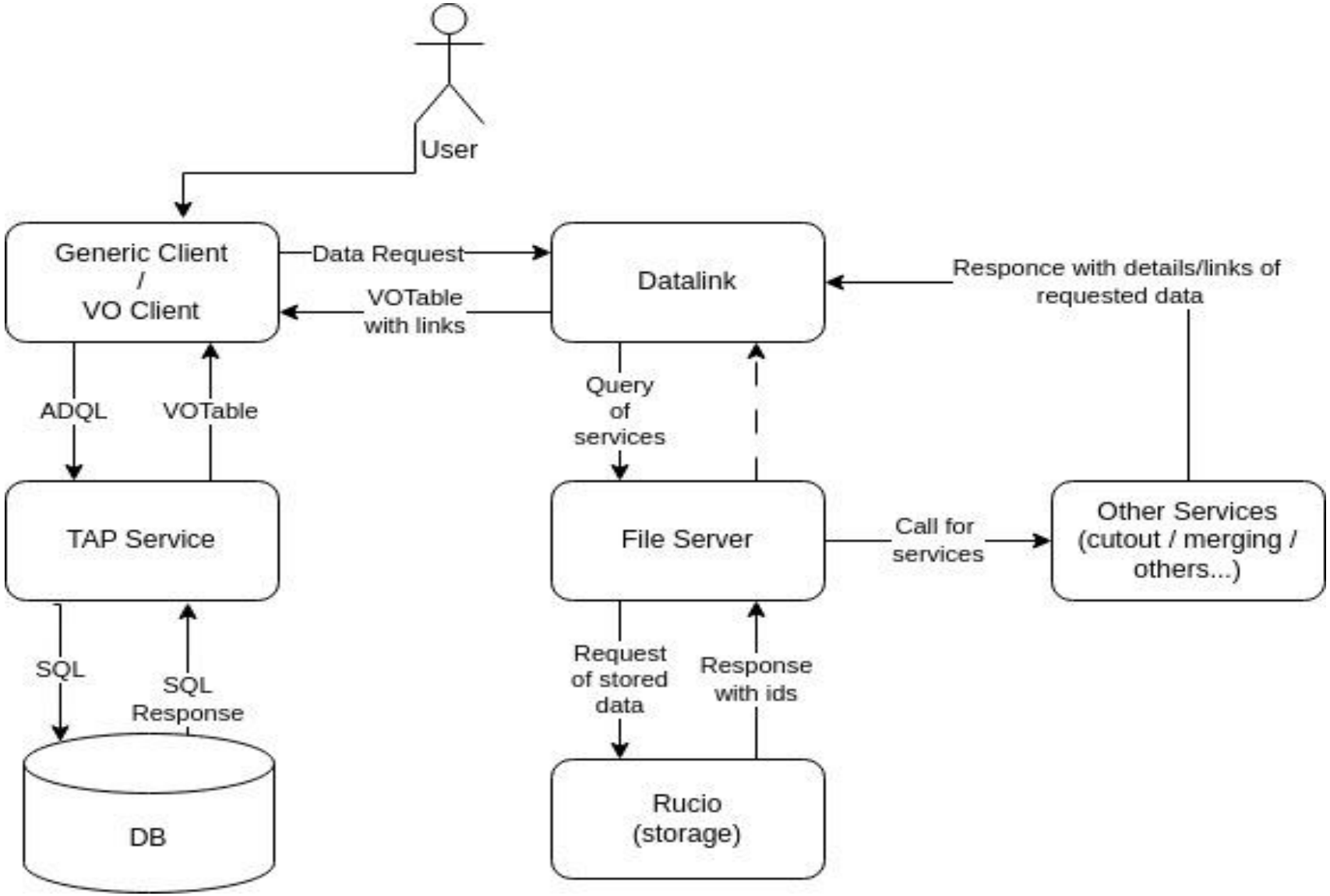


Ingestion:
- a (READ)
- b (PUT REST)
- c1 (WRITE)
- c2 (INSERT SQL)

Extraction:
- 1 (FORM)
- 2 (SEND ADQL)
- 3 (SELECT SQL)
- 4 (PUT REST)
- 5 (PUT REST)
- 6 (READ)
- 7 (PUT REST)

Actor / Action Schema

This second point of view takes into account that the user could run the query via a generic client (Python code for example).



Metadata

Data sources have to provide the list of metadata to be labelled, imported, filtered, queried and viewed as result.

Round tables are planned to fill out the Excel sheet together.

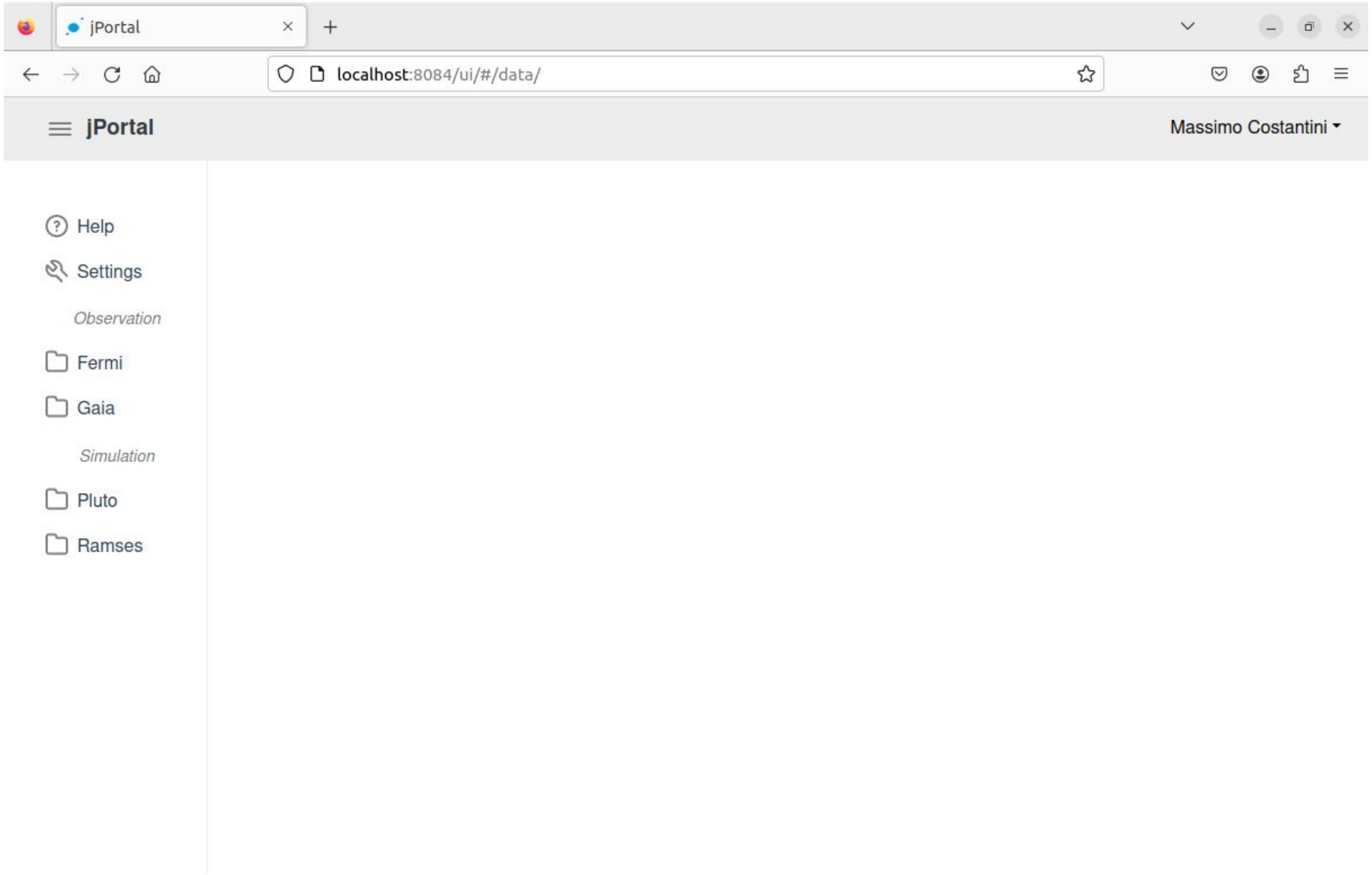
	A	B	C	D	E	F
1	NAME	LABEL	TO BE IMPORTED	TO BE FILTERED	TO BE QUERIED	IN RESULTS VIEW (DEFAULT)
2	NAXIS_HDU0		Y	N	N	N
3	EXTEND		Y	N	N	N
4	CHECKSUM_HDU0		Y	N	N	N
5	DATASUM_HDU0		Y	N	N	N
6	TELESCOP		Y	N	Y	Y
7	INSTRUME		Y	N	Y	Y
8	EQUINOX		Y	N	Y	Y
9	RADECSYS		Y	N	N	N
10	DATE		Y	N	Y	Y
11	DATE_OBS	Observation start date	Y	Y	Y	Y
12	DATE_END	Observation end date	Y	Y	Y	Y
13	TSTART	Start MET (s)	Y	Y	Y	Y
14	TSTOP	End MET (s)	Y	Y	Y	Y
15	TIMESYS		Y	N	N	N
16	TIMEUNIT		Y	N	N	N
17	GPS_OUT		Y	N	N	N
18	MJDREFI		Y	N	N	N
19	MJDREFF		Y	N	N	N
20	OBSERVER		Y	N	Y	Y
21	FILENAME		Y	N	Y	Y
22	ORIGIN		Y	N	N	N
23	CREATOR		Y	N	N	N

Photons | Spacecraft

jPortal

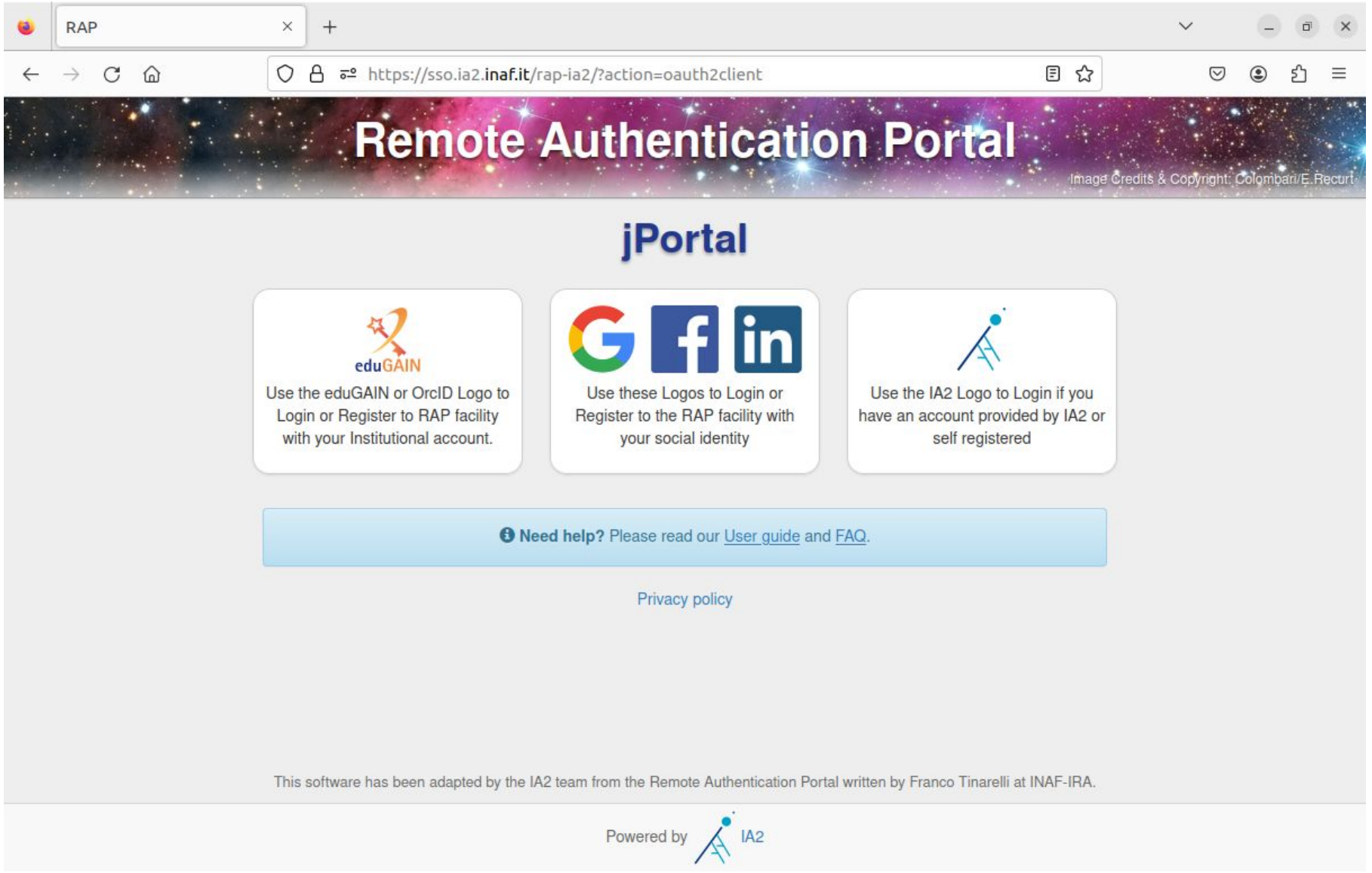
New portal with all aggregated data sources.

Users logged in get a token with a time expiration.



RAP

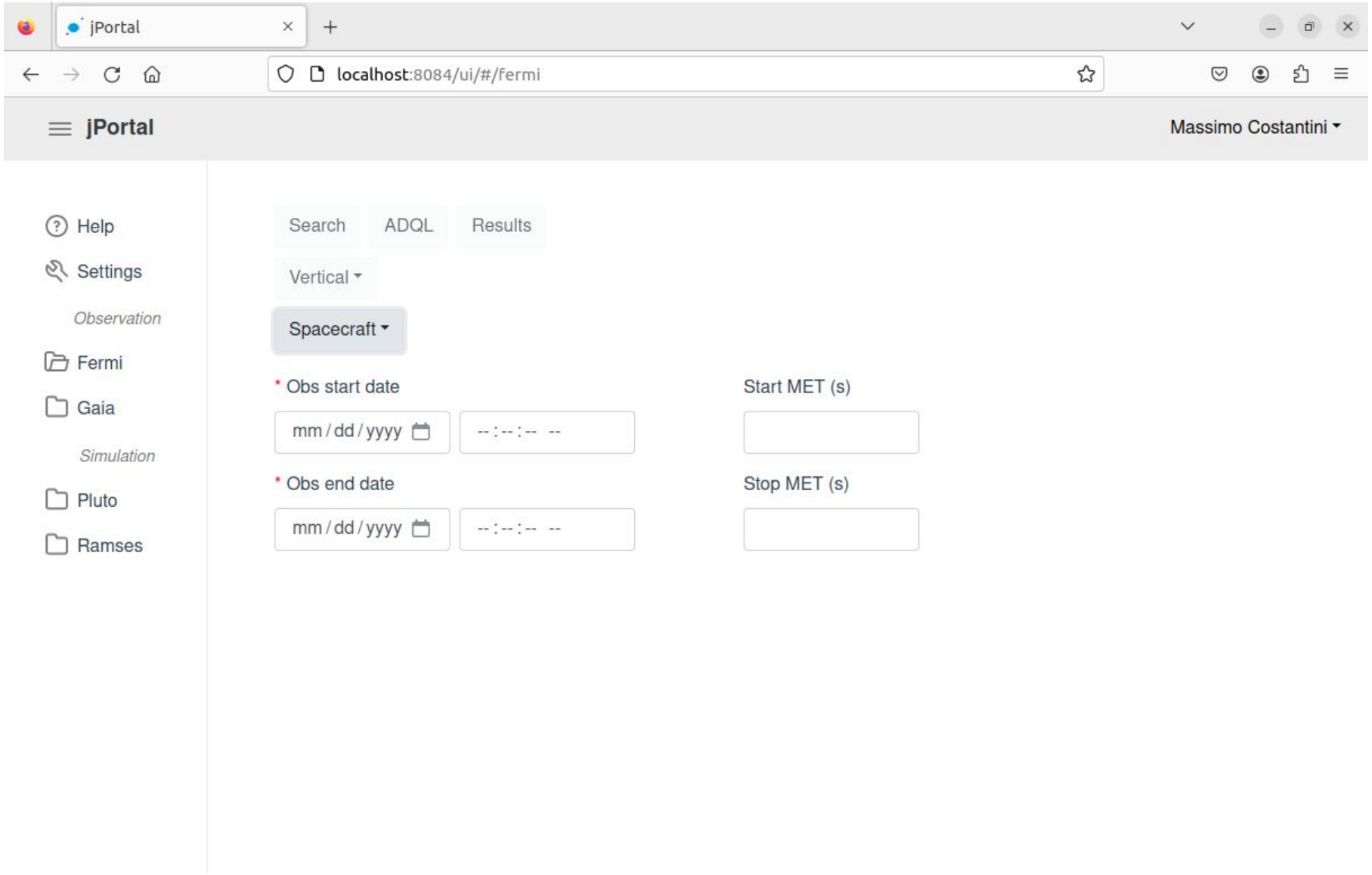
Integration with the Remote Authentication Portal permits to grant or revoke access to private data (Gaia).



Fermi

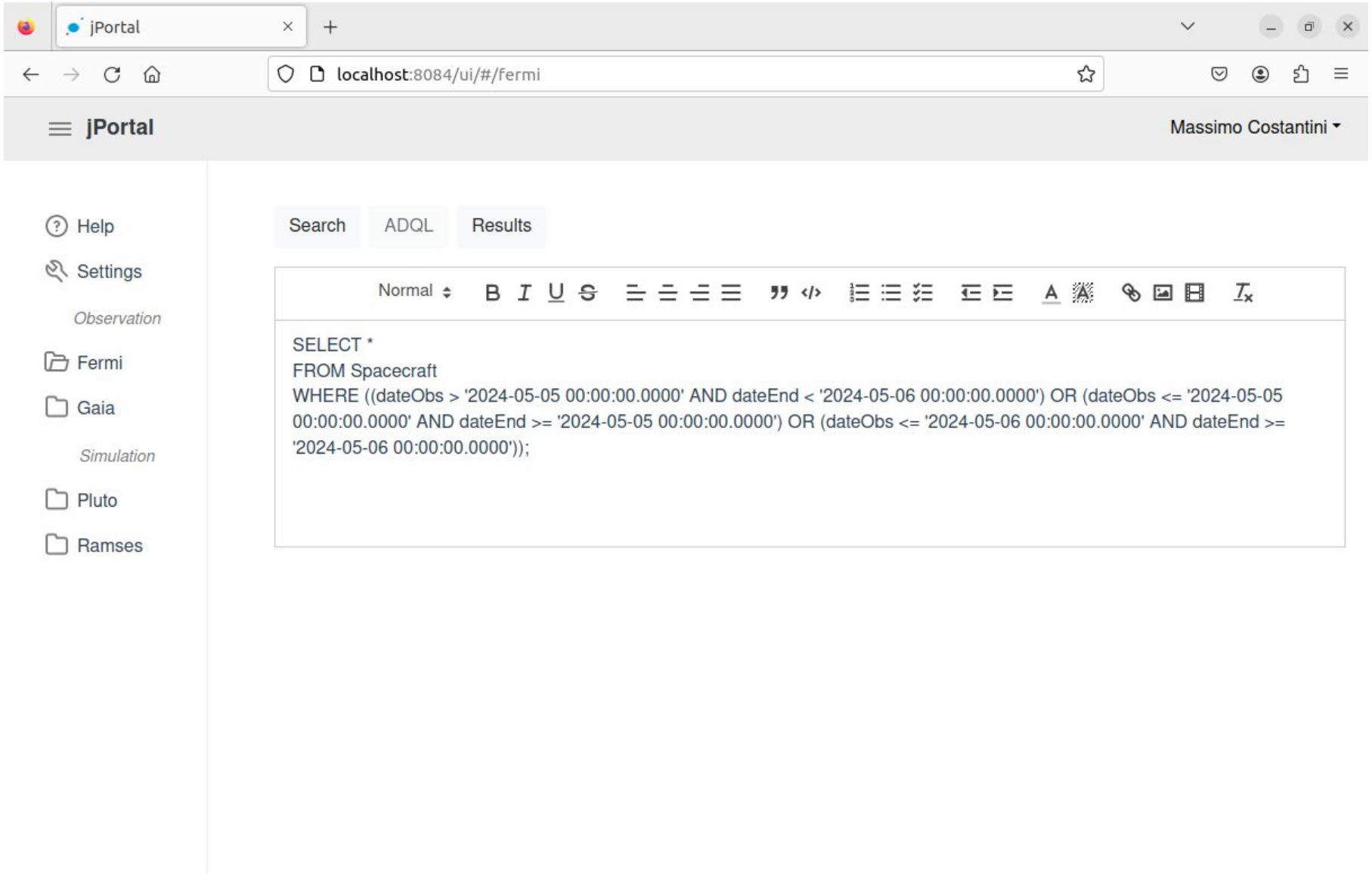
Example of mandatory fields component.

Both fields have to be filled in to enable the search button.



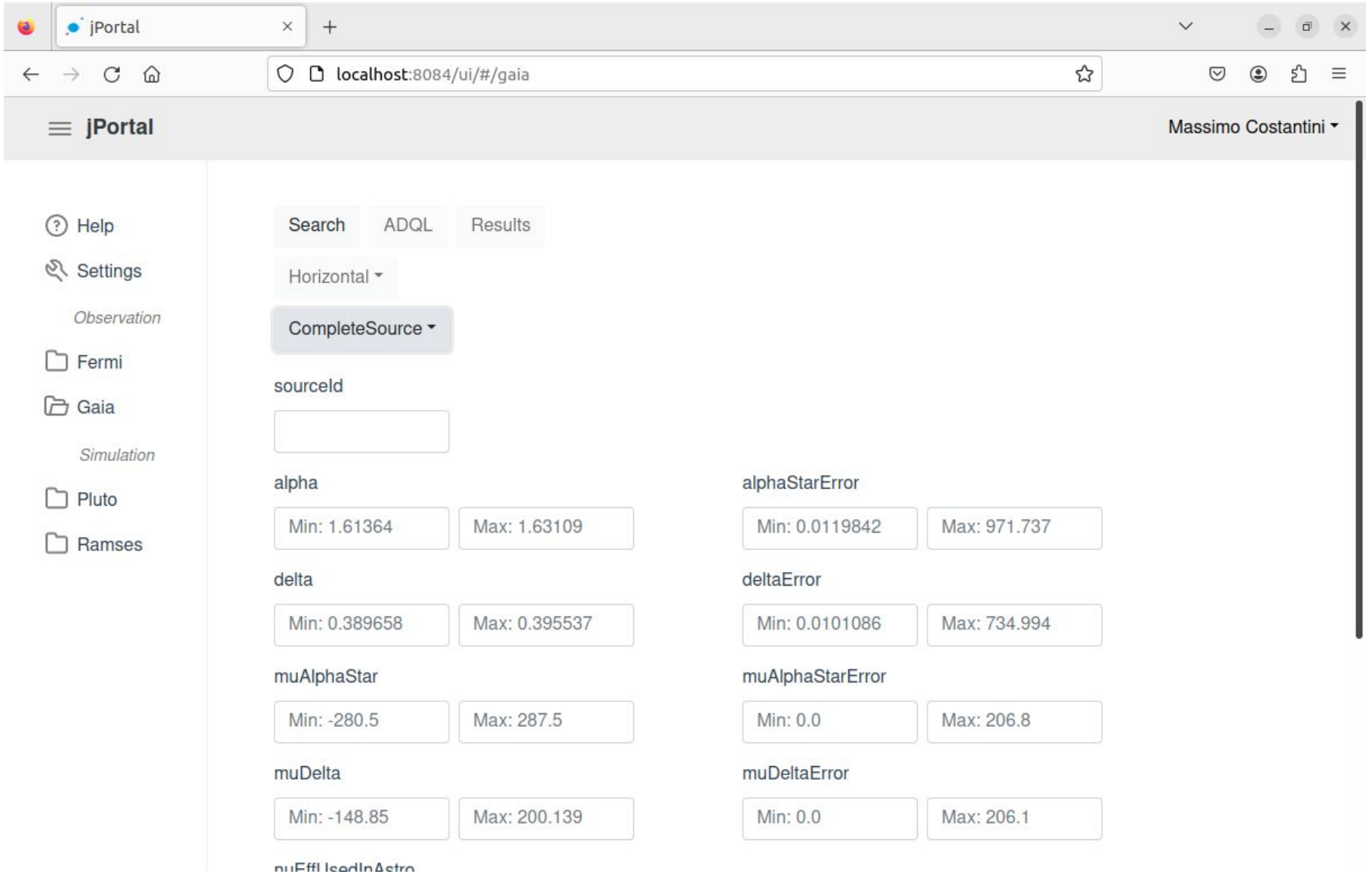
ADQL

Editor to modify and resubmit the query.



Gaia

Due to the huge amount of data and the number of columns, performance could be a problem.

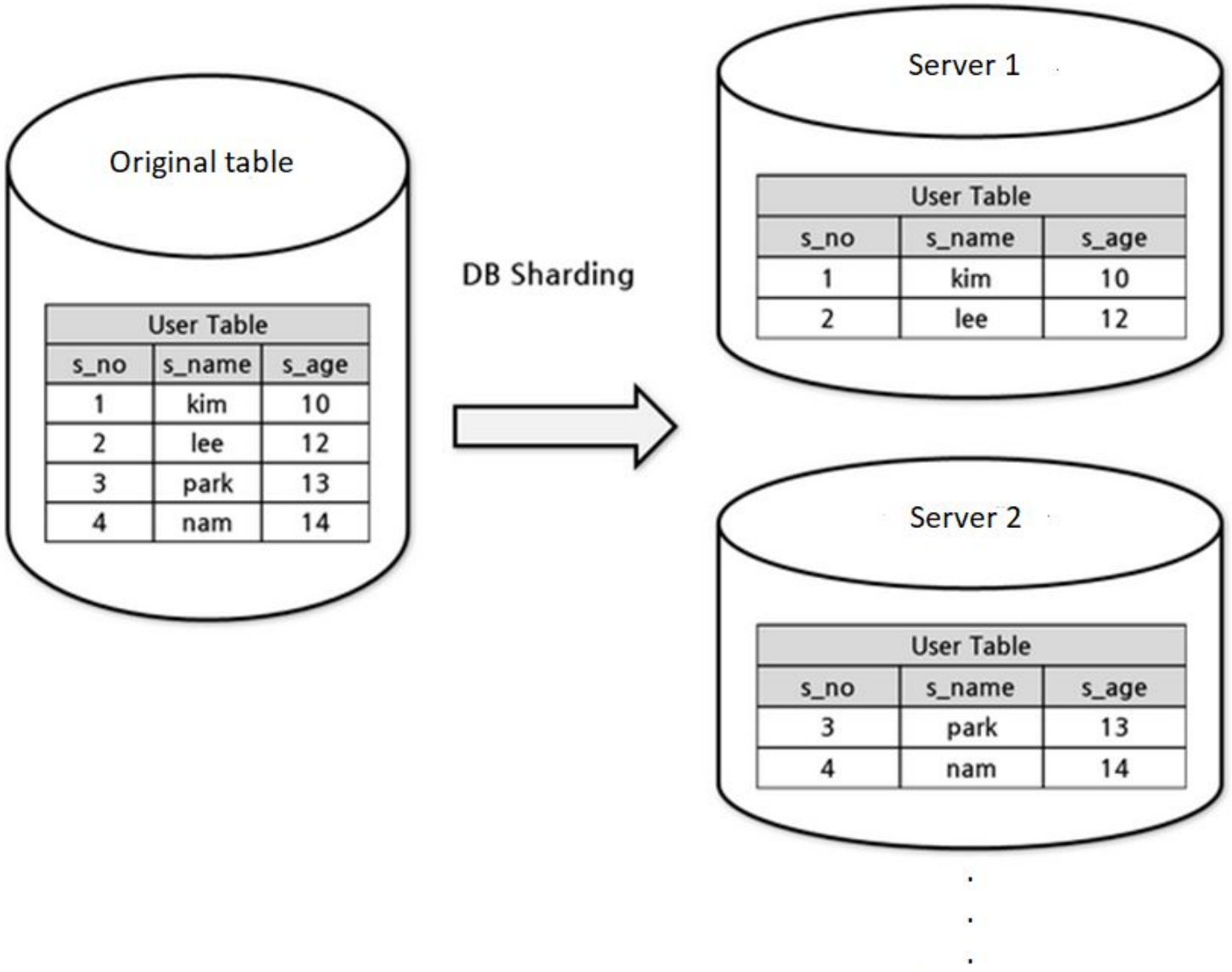


Sharding

Indexes could be a bottleneck, better partitioning tables across multiple servers.

Tables can be fragmented by rows or by hash.

PostgreSQL has extensions (pg_part and pg_shard for example) to do that job.



Thanks for your attention!