PITHIA-NRF e-Science Centre

Prof. Tamas Kiss – University of Westminster

31.03.2022

PITHIA e-Science Centre - as described in the Grant Agreement

- central **integration** tool for data, models and scientific services
- **integration** and **more efficient** utilisation of currently existing, heterogeneous and non-interoperable tools and services
- accessible via high-level adaptable **graphical user interfaces**
- supports the **learning process**
- exploits the power of **cloud-based** computational and data resources
- **functionalities**:
  - registration, cleaning and processing of Datasets
  - registration of Models acting on these Datasets
  - utilisation of applications and workflows implementing these Models

PITHIA e-Science Centre – the planned roadmap

- Initial requirements collection – **30 November 2021**
  - Collection of user requirements and analysis of state-of-the-art technologies – **D5.1**
- Proof of concept – **31 March 2022**
  - Deploy the selected candidate tools and demonstrate them in the EGI cloud testbed (without further integration or development) – **D5.2**
- First release – 31 March 2023
  - First prototype implementation with documentation and user feedback – D5.3
- Second Release – 31 March 2024
  - Second prototype implementation with documentation and user feedback – D5.4
- Final Release – 31 March 2025
  - Final implementation with documentation, user feedback and plans for future sustainability – D5.5

Where are we now?

- Initial requirements have been collected via detailed questionnaires and feedback to the community was given
- Implemented a proof of concept where a model was executed on EGI cloud resources
- Defined the generic set of services to be supported by the e-Science Centre
- Defined the desired functionalities of the first protype
- Modified the ESPAS metadata structure and ontology according to PITHIA requirements
- Started the implementation of the first prototype

Requirements collection

- Long and complex questionnaire designed and distributed to the community
  - 166 questions
  - not everybody answers every question
  - but one person may complete the same section multiple times - answers are required per Dataset/Model
- 59 answers received
  - 19 scientific users (from 5 partners only)
  - 25 dataset owners (from 10 partners)
  - 15 model owners (from 8 partners)
- Results
  - Summarised in D5.1
  - Presented to the whole project in a workshop in January
  - Forms the basis of our understanding regarding current state and practices related to Data Collections and Models

Inventory of Data Collections and Models

- 38 Data Collections and 12 Models identified
- Living documents, constantly updated
- Available at :

https://docs.google.com/spreadsheets/d/1RkeeKmkD_cyE8bwtW6xTR9Br38RCJwaI/edit?usp=sharing&ouid=116347269825035096423&rtpof=true&sd=true

E-Science Centre Services (to be further refined)

Information and Community Services

- Home
- News
- Blogs
- Collaboration
- …

E-Learning Services

- Register Learning / Scientific Material
- Search Learning / Scientific Material
- Webinars
- …

Scientific Services

User Support Services

- Helpdesk
- Ticketing system
- …

Data Collections

- Register Data Coll.
- Search for Data Coll.
- Access to Data Coll.

Models

- Register Model
- Search for Model
- Execute Model

Cloud-based services

- Register cloud-based service
- Search cloud-based service
- Execute cloud-based service
- Create workflow
- …

User Account Management

Security Services

- Authentication
- Authorisation
- Data protection
- Privacy Management
- ….

Model execution mechanisms

- Models always need to be registered with the e-Science Centre first
- Once the user finds the "right" Model, it can be executed in various ways:

1. **Executed on PITHIA site/node**:
   - The model runs/executes on the PITHIA node providing the Model
   - Two options:
     a) user interface is a statically built webform (can be embedded into the e-Science Centre)
     b) user interface is dynamically built webform that communicates with the Model via API
2. **Executed in the EGI cloud:**
   - e-Science Centre "magically" deploys and executes the model on EGI cloud computing resources
   - this means creating a "copy" of the model, deploying it in the cloud and enabling the user to execute this private "copy"
   - Once the user is finished, the "copy" is killed (☹) and the cloud resources are freed (☺)
3. **Download and execute Model on local computer**
   - user downloads a "copy" of the model from the PITHIA node and executes on local computer

What have we done so far regarding practical implementation?

- Executed a real PITHIA model in the EGI cloud (option 2 in previous slide)
  - we call this a proof of concept
  - used some cutting edge technologies developed by our group to “make the magic happen”
  - published a scientific paper about this (collaboration between UoW and NOA)

- Started designing and implementing the first e-Science Centre prototype
  - the focus is on publication and search for Data Collections and Models
  - Nothing “fancy”, just find the Data Collection or Model using scientific terms (ontology) and then return the URL where the user can interact with it (return a simple link)

Proof-of-concept model execution in EGI

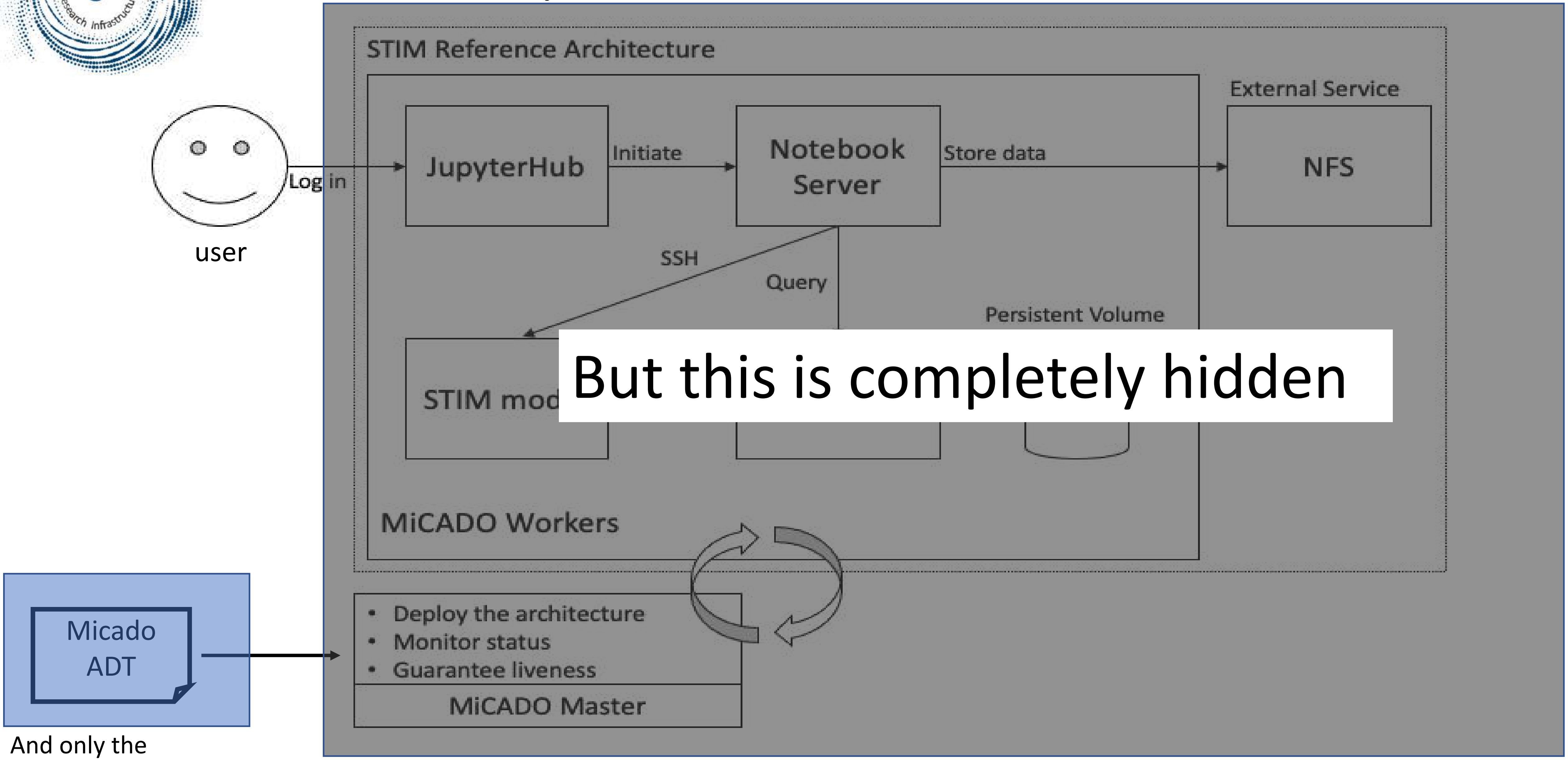
Which model did we execute? (this is where I need help)

- SWIF: Topside Solar Wind driven autoregressive model for Ionospheric short-term Forecast
- STIM: Storm-Time Ionospheric Model (STIM)

What does it do? (and even more help here)

- solar wind-driven empirical model for the middle latitude ionospheric storm-time response
- forecasts ionospheric storm effects at middle latitudes triggered by solar wind disturbances
- uses near-real-time datasets describing the solar wind conditions in the Earth's vicinity
- homogenises and resamples data at a standard temporal resolution
- analyses temporal variations of the interplanetary magnetic field parameters
- detects intervals of ionospheric storms
- stores the relevant data in an appropriate relational database schema

Proof-of-concept model execution in EGI



And only the developer sees this

Towards a first e-Science Centre Prototype

What do we want to achieve first?

- Register a Data Collection or Model with the e-Science Centre
  - we need metadata to describe the scientific meaning/properties – ESPAS (see Ivan’s presentation)
  - performed by the Data Collection/Model owner
- Search for a Data Collection or Model using scientific terms and keywords
  - use the terms of the ontology to find the “right” Data Collection or Model (e.g. searching for Data Collections where the Observed property is “Inner Heliosphere”)
  - performed by the scientific user
- Interact with the found Data Collection or Model
  - Retrieve the necessary data from the Data Collection
  - Execute the Model to get the desired results
  - performed by the Scientific User

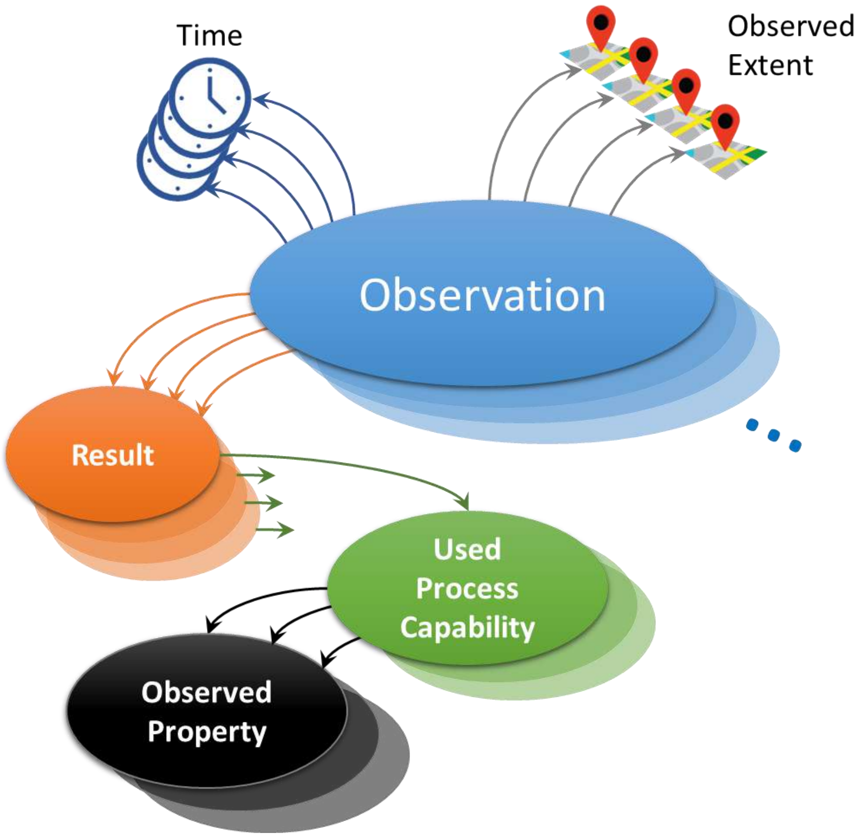
e-Science Centre prototype

Registration of models and datasets

- **Preparation of metadata in a standard format**
  - Utilises data model and space physics ontology developed in ESPAS project
  - Complies with the ISO 19156 standard on Observations and Measurements (O&M)
  - Enhanced specifically for space physics
  - The same language and schema are used for Data and Models
- **Significant simplification:** only data collections are registered (not individual records)
- **Registration process:**
  - a set of descriptive XML documents have to be prepared and uploaded by the provider (or created automatically through web forms)
  - registration includes a number of standard terms from the ESPAS domain ontology

Registration metadata and ontology

ESPAS



PITHIA

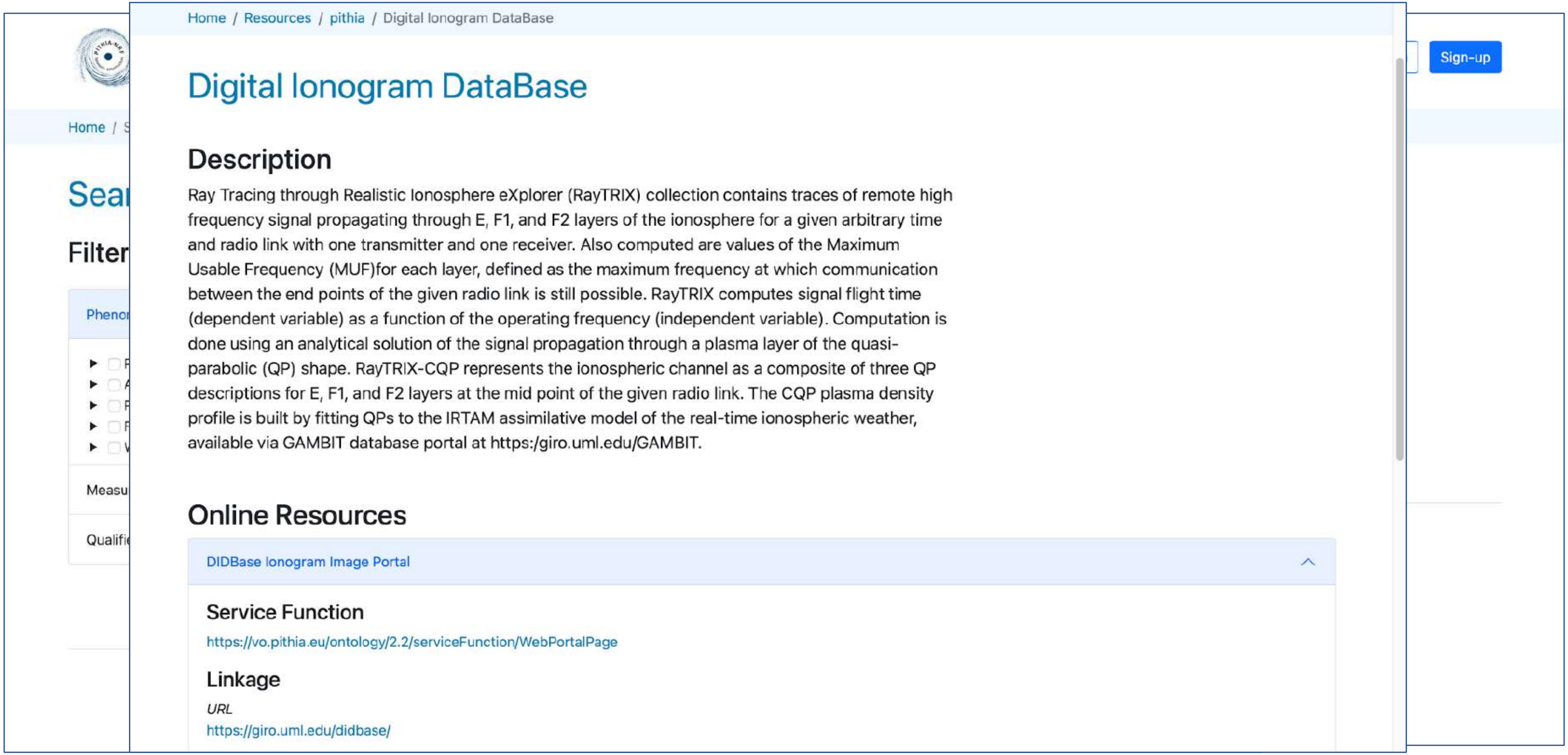
Registration metadata and ontology

5. Registration metadata and ontology

- **Three example observation collections have been prepared:**
  - **DIDBase (Digital Ionogram Database)**: a dataset from ionosonde observatories;
  - **IRTAM 3D (IRI-based Real-Time Assimilative Model):** a weather nowcast of the global electron density in the ionosphere computed using DIDBase collection; and
  - **RayTRIX-CQP (RayTracing through Realistic Ionosphere Explorer, Composite Quasi-Parabolic representation):** a model computation of the high-frequency signal propagation through E, F1, and F2 layers of the ionosphere, presented as an oblique ionogram

- 35 XML files altogether

A very first sneak-peak into the first prototype



Thank you for your attention!
Any questions?

The PITHIA-NRF project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101007599