PITHIA-NRF

Plasmasphere Ionosphere Thermosphere Integrated Research Environment and Access services: a Network of Research Facilities

Static Datasets Registration

at PITHIA e-Science Centre

User Guide

Version 1.2

October 20, 2025

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

Table of Contents

1. Terminology and Abbreviations

DOI	Digital Object Identifier
DQ	Data Quality
DQF	Data Quality Flag
eSC	e-Science Centre
URL	Uniform Resource Locator
Data Level	Level of information processing ranging from Level 0 (unprocessed) to Level 4 (derived by secondary analysis of lower-level data or by modelling).
Data Quality Flag	DQ flag describes measures taken to clean and validate the data, as well as characterise the residual data noise. Commonly, Data Level 1 refers to observed properties of the instrument probing signal while Data Level 2 corresponds to the derived geophysical properties of the Feature of Interest.
Data Resource	Single data service item and its associated metadata, available through the PITHIA-NRF system.
Dataset	Pre-computed or pre-processed data resource available for download.
Data Subset	A portion of a Data Collection for registration as Static Dataset.
Features of Interest	[*standard ISO vocabulary*]: Real-world object that carries the property which is observed or modelled to produce a Data Collection
ISO	International Standards Organisation
Metadata Model	[*science-neutral*]: Specification of different documents and their contents that are required for registration of data resources
Ontology	[*science-specific*]: A set of standard vocabularies for the selected domain of science
Phenomenon	[*standard ISO vocabulary*]: A physical observable (a.k.a. “Mother Nature”). Not to confuse with events; phenomena are not defined in time or space. The top-level phenomenon categories are Field, Particle, and Wave.
Static Datasets	[*standard PITHIA vocabulary*]: A listing of events or investigations assembled to aid users in locating data of interest. Each Entry in a Static Dataset has distinct begin and end times and a list of

registered Data Subsets with optional DOIs to their persistent storage.

2. Introduction

This document provides detailed instructions on how users can register and publish Static Datasets at the PITHIA e-Science Centre (eSC).

As mentioned in the related interface in the eSC, a Static Dataset Entry is "*A listing of events or investigations assembled to aid users in locating data of interest. Each Entry in a Static Dataset has distinct begin and end times and a list of registered Data Subsets with optional DOIs to their persistent storage.*".

Four categories of Static Datasets are published at the eSC:

- **Datasets Used in an Academic Publication**: Datasets placed in a static storage for a publication in an academic journal that required persistent access to the source data.
- **Training Datasets for a Machine Learning Model**: Training dataset for a Machine Learning model allocated by the model designer and stored in a persistent repository for posterity. Catalog Entry registers the model, and Data Subset registers the training dataset. Retraining of the model results in registration of another Data Subset document for the same Catalog Entry.
- **Data Pertaining to an Event**: Subset of unpublished data (graphical, numerical) with distinct start and stop times that correspond to a registered event in the Event Catalogue.
- **Data from an Experiment Campaign**: Data collection during planned experiments and observation campaigns such as the World Day.

3. Publishing Static Datasets

After logging in, registered users are presented with a landing page similar to the one shown in Figure 1. At the bottom of Figure 1, the user can see the institution they are a member of and select to manage their registrations (Figure 2), i.e. register new Data Collections, Workflows, and Static Datasets and/or edit the existing ones.

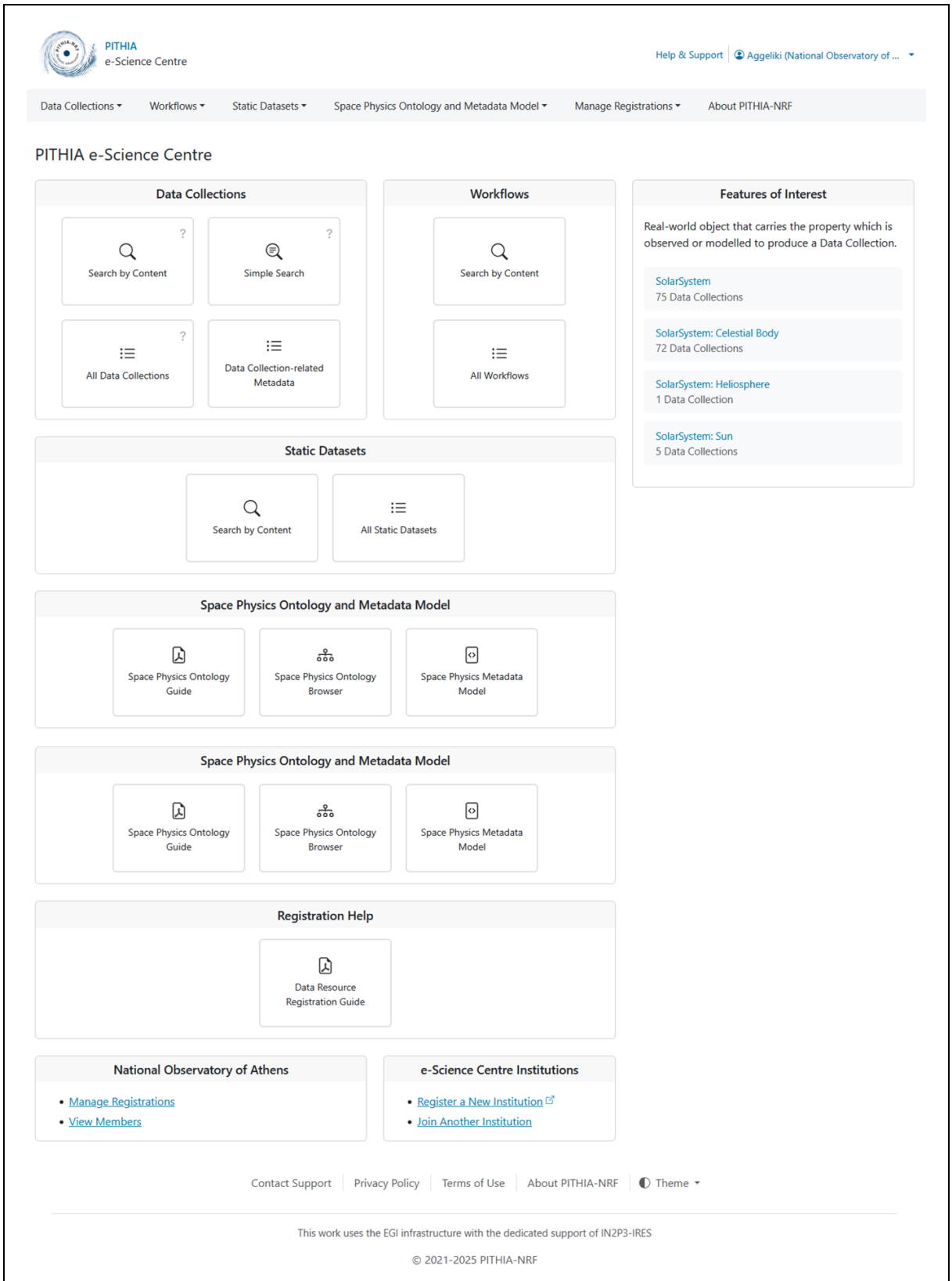

Figure 1 – Landing page of a registered user with an active membership in an institution

Figure 2 – Manage registrations option for a specific institution

For users to publish Static Datasets, they must choose the "Manage Registrations" option, which directs them to the corresponding page, as shown in Figure 3.

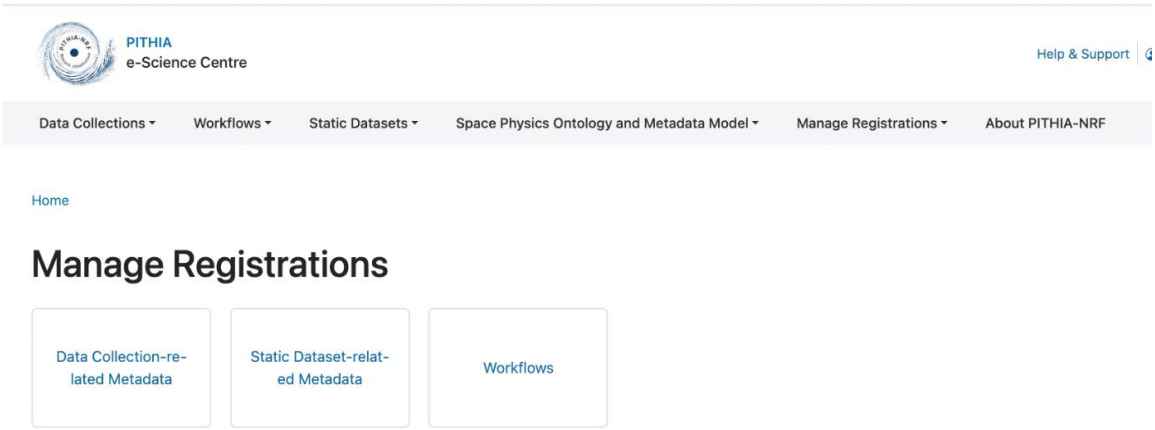

Figure 3 – Manage Registrations Page for a Specific Institution

As this guide focuses only on Static Datasets, the user must choose the option “Static Dataset-related Metadata”. This action directs users to the related page, where they must first register any required “Static Dataset Entries” and then “Data Subsets”, as shown in Figure 4.

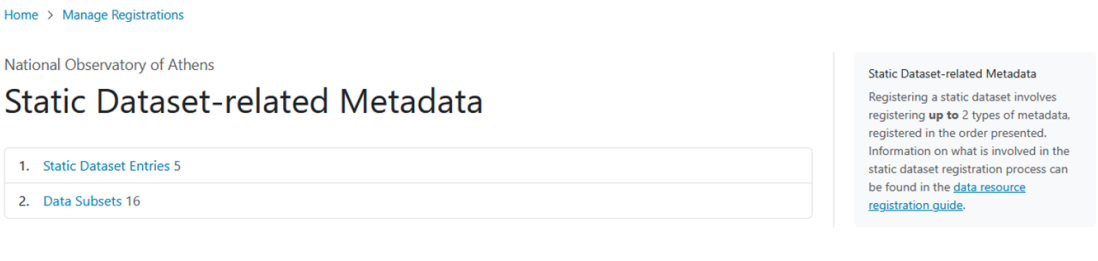

Figure 4 – Manage Page for Static Datasets-related Metadata

The subsequent section provides the instructions for registering both types of Static Dataset-related Metadata.

3.1. Publishing Static Dataset Entries

Figure 5 displays the Static Dataset Entries management page, where a user can register a new Static Dataset Entry or update an existing one. This guide focuses on new registrations, specifically through the "Register with Wizard" option, which offers a graphical user interface.

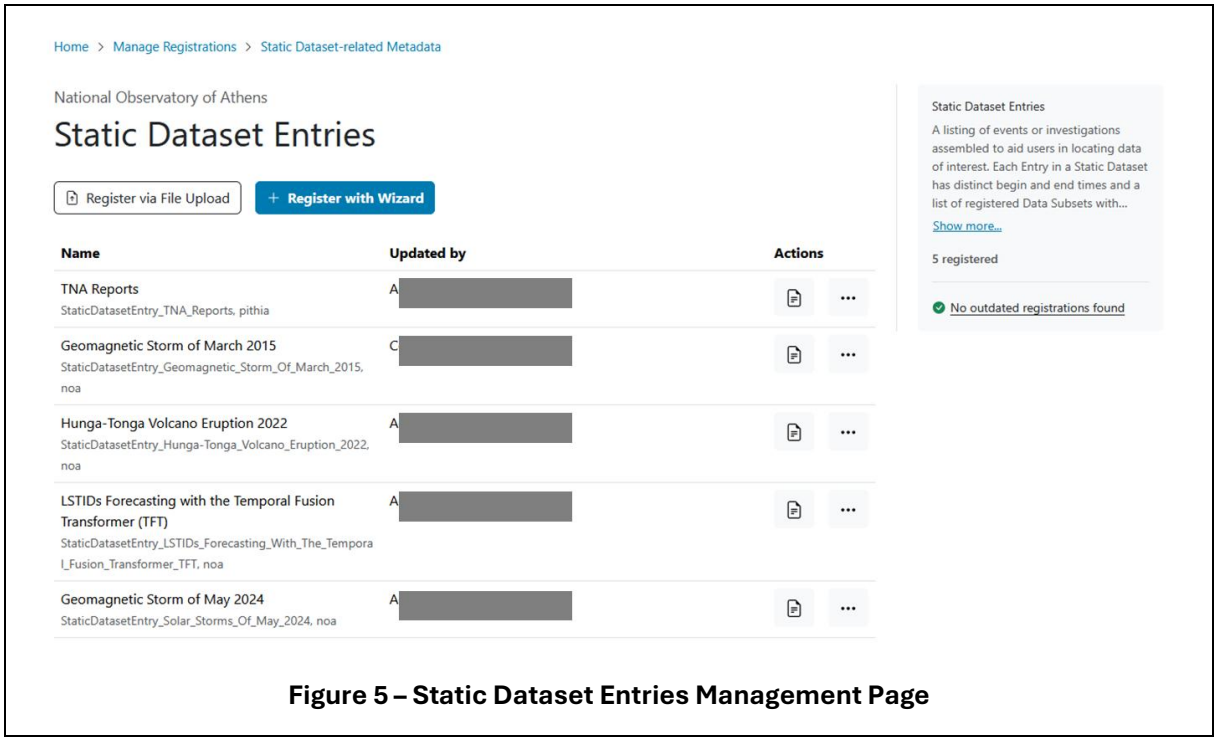

Figure 5 – Static Dataset Entries Management Page

Having selected the "Register with Wizard" option (Figure 5), users see a form where they must enter all related information. The mandatory fields are indicated by a red asterisk.

The information is divided into the following categories:

1. **New Static Dataset Category** (Figure 6), where users must:
    a. Choose from drop-down menu the category of the Static Dataset; the available categories are:
        - Datasets Used in an Academic Publication.
        - Training Datasets for a Machine Learning Model.
        - Data Pertaining to an Event.
        - Data from an Experiment Campaign.

b. Provide the name of the Static Dataset.

c. Select from the drop-down list the associated organisation.

2. **Identifier** (Figure 7), which includes dynamically generated information specifically for the eSC. Here, the users can only modify the metadata version.

3. **Description** (Figure 8), where the user describes the registration that will be shown when someone finds and reads the specific registration-related information.

4. **Features of Interest (Named Regions)** (Figure 9), offers a drop-down list where the users must select the related feature of interest. Details on the vocabulary are available at https://esc.pithia.eu/ontology/categories/featureOfInterest/. Further information on the PITHIA Space Physics Ontology can be found on the relevant guide at https://esc.pithia.eu/ontology/guide/.

5. **Phenomenon Time** (Figure 10), where the user can optionally add the phenomenon's begin and end times.

6. **Validate and Register** (Figure 11), once all the above options are complete the user clicks the "Validate and Register" button, and the eSC completes the registration providing a confirmation message.

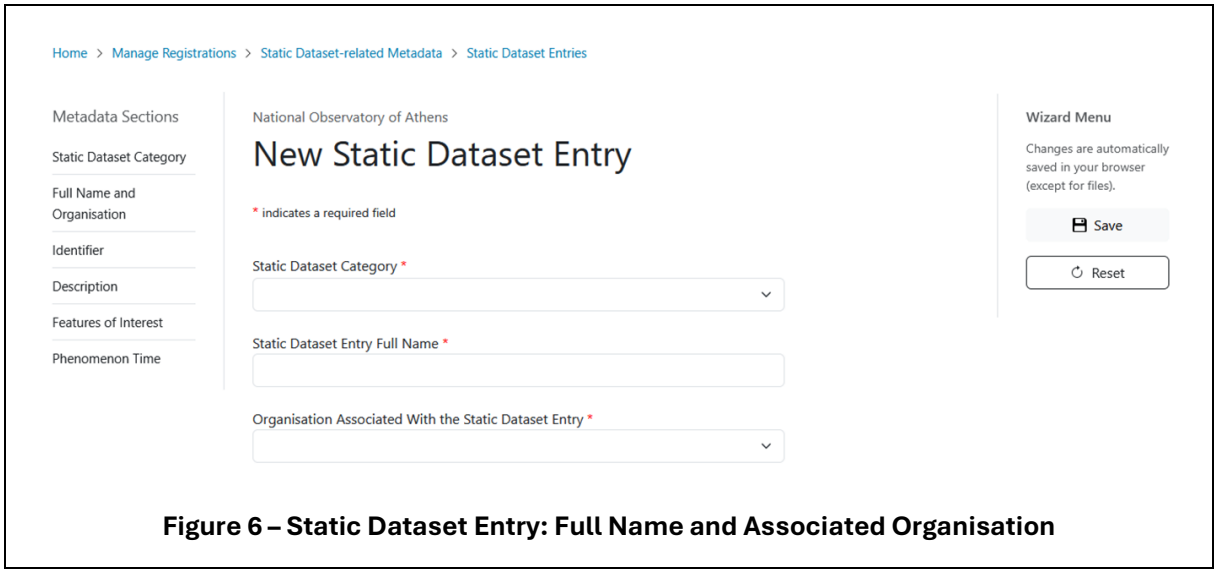

Figure 6 – Static Dataset Entry: Full Name and Associated Organisation

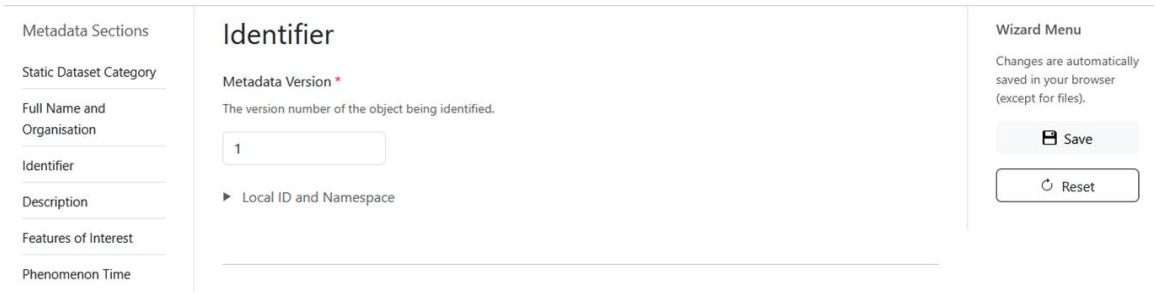

Figure 7 – Static Dataset Entry: Identifier

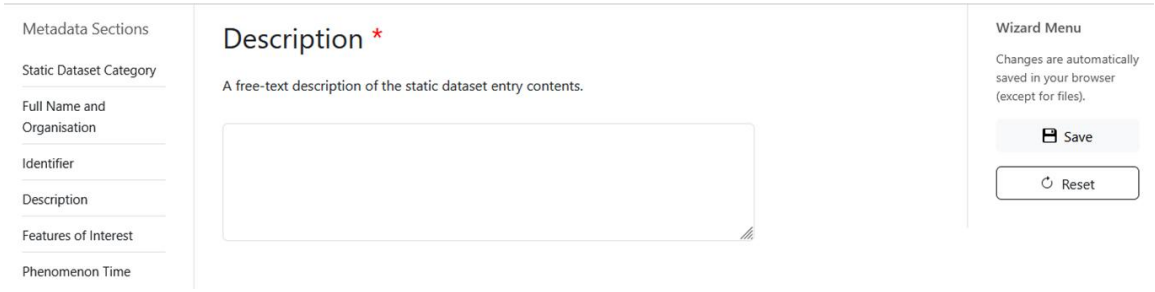

Figure 8 – Static Dataset Entry: Description

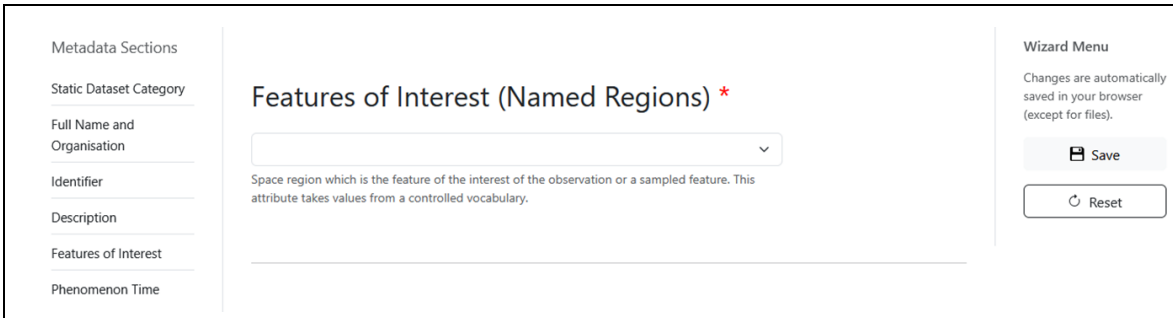

Figure 9 – Static Dataset Entry: Features of Interest

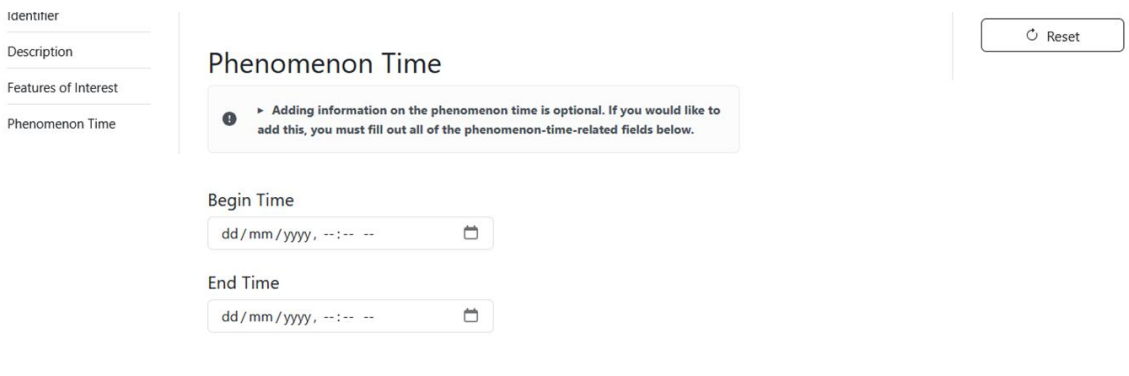

Figure 10 – Static Dataset Entry: Phenomenon Time

Figure 11 – Static Dataset Entry: Validate and Register

3.2. Publishing Data Subsets

Figure 12 displays the Data Subsets management page, where a user can register a new Data Subset or update an existing one. This guide focuses on new registrations, specifically through the “Register with Wizard” option, which offers a graphical user interface.

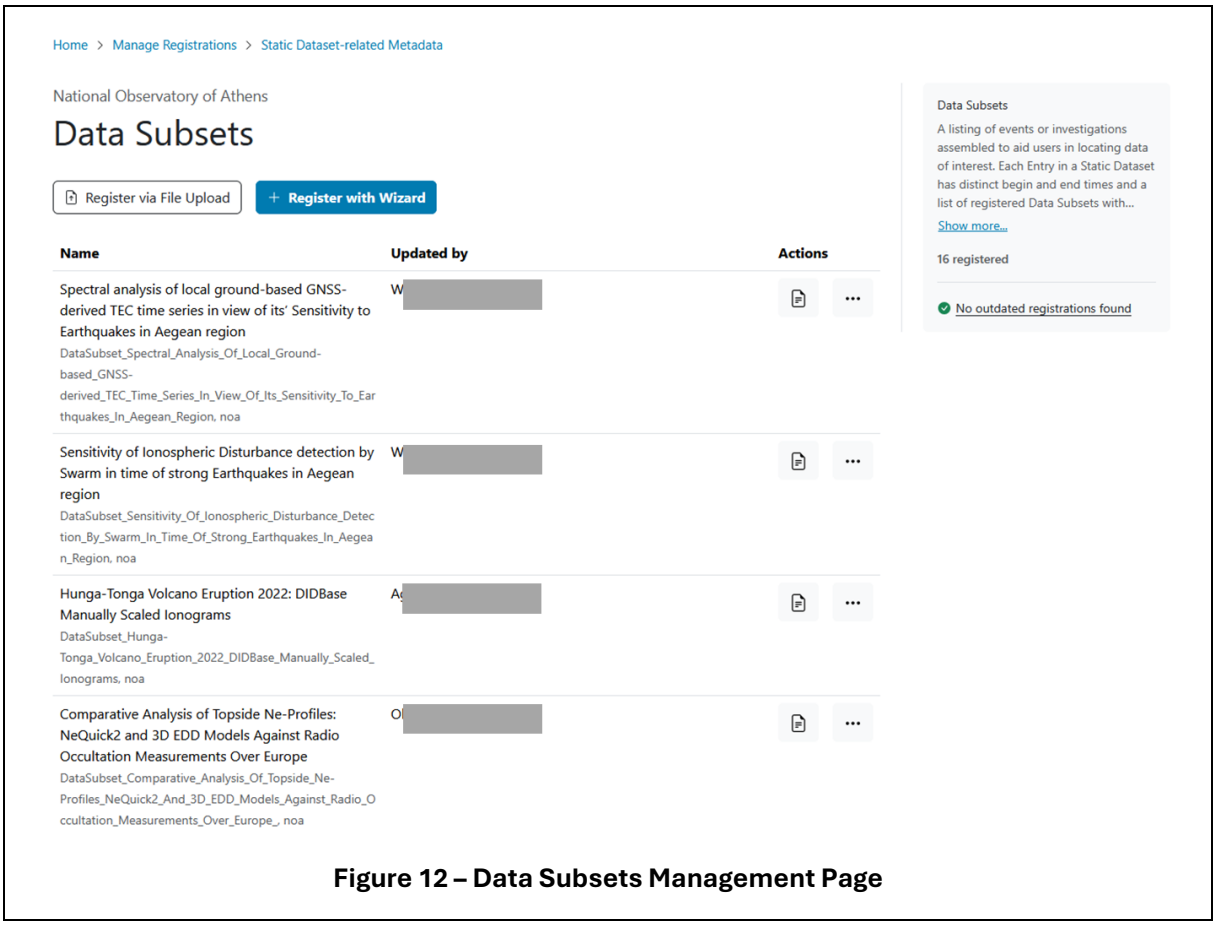

Figure 12 – Data Subsets Management Page

Having selected the "Register with Wizard" option (Figure 12), users see a form where they must enter all related information.

The information is divided into the following categories:

1. **New Data Subset** (Figure 13), where users must:
    a. Provide the name of the Data Subset.
    b. Select from the drop-down list the associated organization.
2. **Identifier** (Figure 14), which includes dynamically generated information in the eSC. Here, the users can only modify the metadata version.
3. **Description** (Figure 15), where the user provides a description of the Data Subset.
4. **Static Dataset Entries, Features of Interest and Data Collections** (Figure 16), which offers three drop-down lists where users must select the related types.
    a. **Static Dataset Entry**: Select the appropriate Static Dataset Entry, the one to which the Data Subset belongs.
    b. **Features of Interest (Names Regions)**: Select one or more features of interest of the observation or a sampled feature related to the Data Subset. Details on the vocabulary are available at

https://esc.pithia.eu/ontology/categories/featureOfInterest/. Further information on the PITHIA Space Physics Ontology can be found on the relevant guide at https://esc.pithia.eu/ontology/guide/.

c. **Data Collections**: Select one or more (if any) Data Collections related to the Data Subset. A full list of the eSC registered Data Collections can be found at https://esc.pithia.eu/data-collections/.

5. **Result time** (Figure 17), one or more time intervals included in the Data Subset.
6. **Online Resources**, where the user selects to update the data related to the Data Subset to the eSC (*toggle on*) (Figure 18), or to provide URLs (*toggle off*) (Figure 19). More than one online resources can be added. Each online resource is fully described by providing details about its:

    a. **Service Function** (drop-down list): The function performed by the online resource. The "Service Function" vocabulary is available at https://esc.pithia.eu/ontology/categories/serviceFunction/.

    b. The **resource itself** by providing either "File" (Figure 18) or "Link to Online Resource" (Figure 19).

    c. **Name**: The name of the online resource.

    d. **Protocol**: The connection protocol (e.g. http, ftp, file).

    e. **Description**: A text description of what the online resource is/does.

    f. **Data Formats** (drop-down list): The format of the stored result accessible from the online resource. The "Data Formats" vocabulary is available at https://esc.pithia.eu/ontology/categories/resultDataFormat/.

7. **Data Level** (Figure 20), the user selects the data level of the Data Subset. The "Data Level" vocabulary is available at https://esc.pithia.eu/ontology/categories/dataLevel/.
8. **Quality Assessment** (Figure 21), where the user selects from the two drop-down lists the related:

    a. Data quality flags. The "Data Quality Flags" vocabulary is available at https://esc.pithia.eu/ontology/categories/dataQualityFlag/.

    b. Metadata quality flags. The "Metadata Quality Flags" vocabulary is available at https://esc.pithia.eu/ontology/categories/metadataQualityFlag/.

9. **DOI** (Figure 22), where the user selects to generate a DOI related to the data subset (optional but strongly recommended).
10. **Validate and Register** (Figure 23), once all the above options are complete, the user clicks the "Validate and Register" button, and the eSC completes the

registration, providing a confirmation message, along with the DOI (if selected the generation of a DOI in the previous step 9).

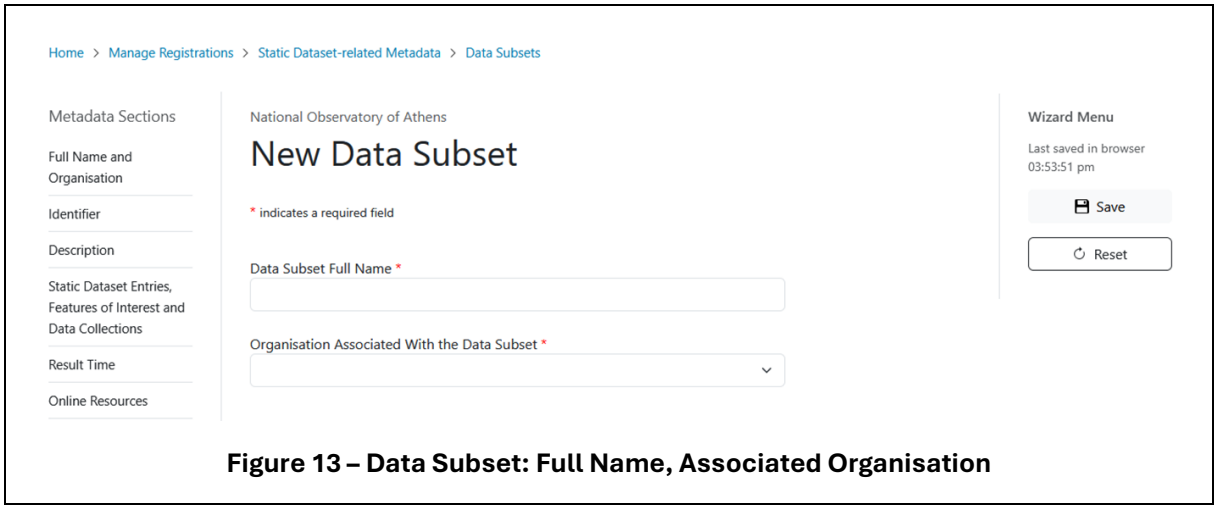

Figure 13 – Data Subset: Full Name, Associated Organisation

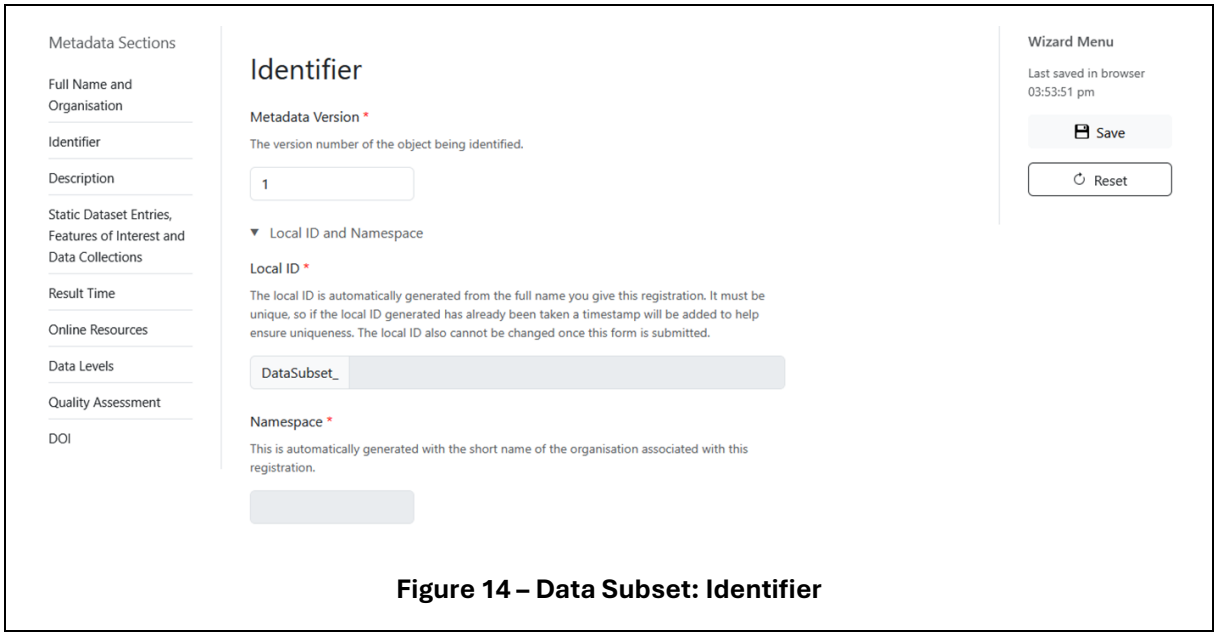

Figure 14 – Data Subset: Identifier

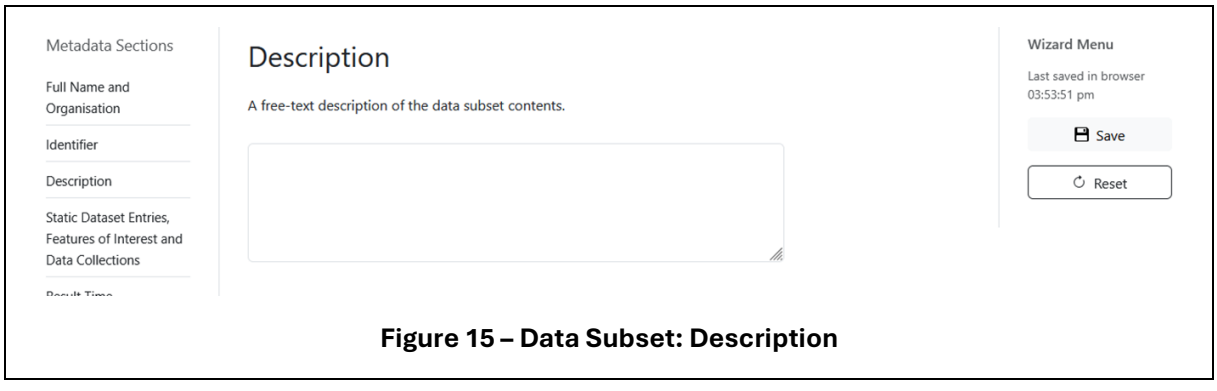

Figure 15 – Data Subset: Description

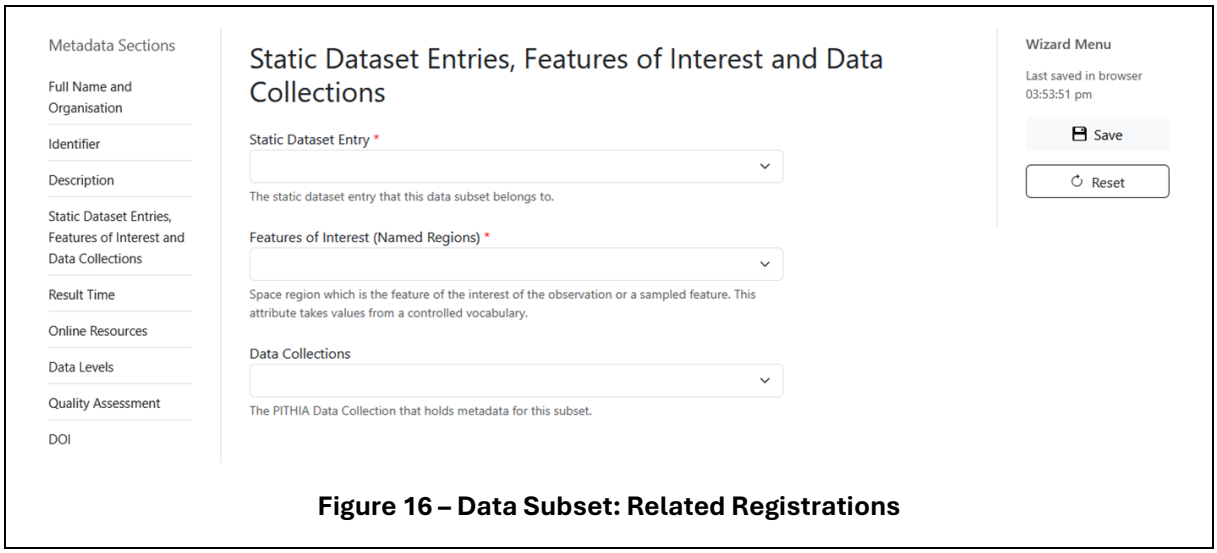

Figure 16 – Data Subset: Related Registrations

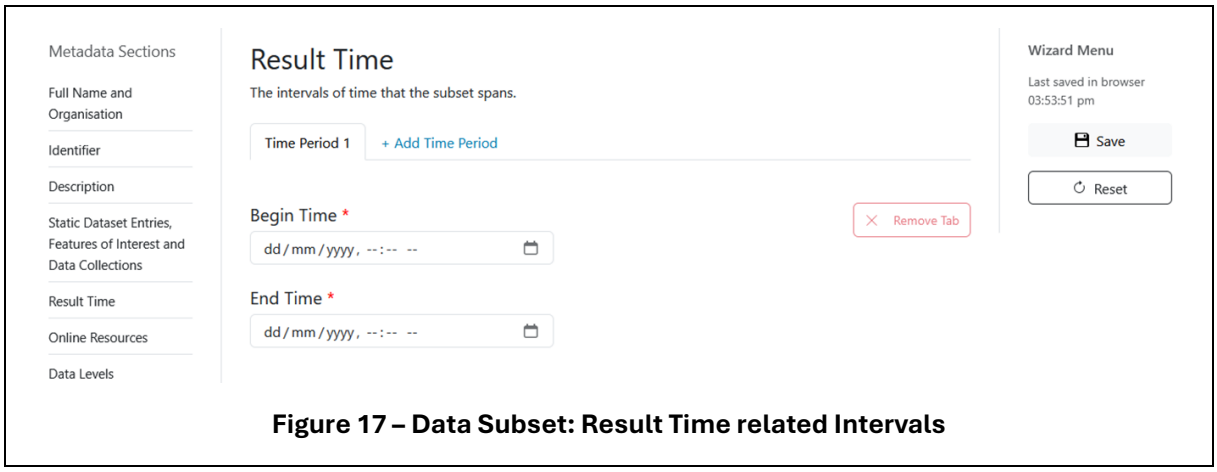

Figure 17 – Data Subset: Result Time related Intervals

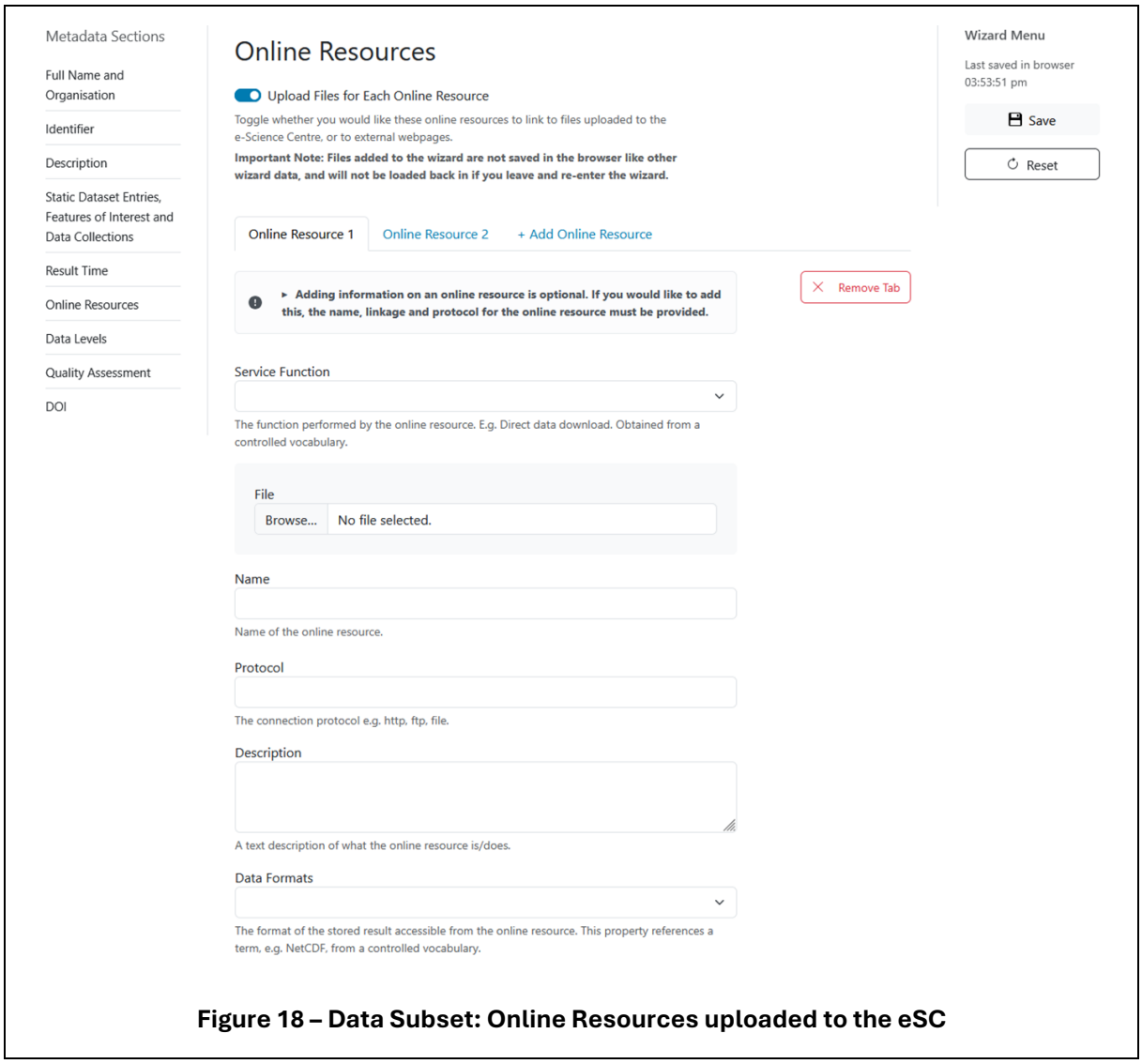

Figure 18 – Data Subset: Online Resources uploaded to the eSC

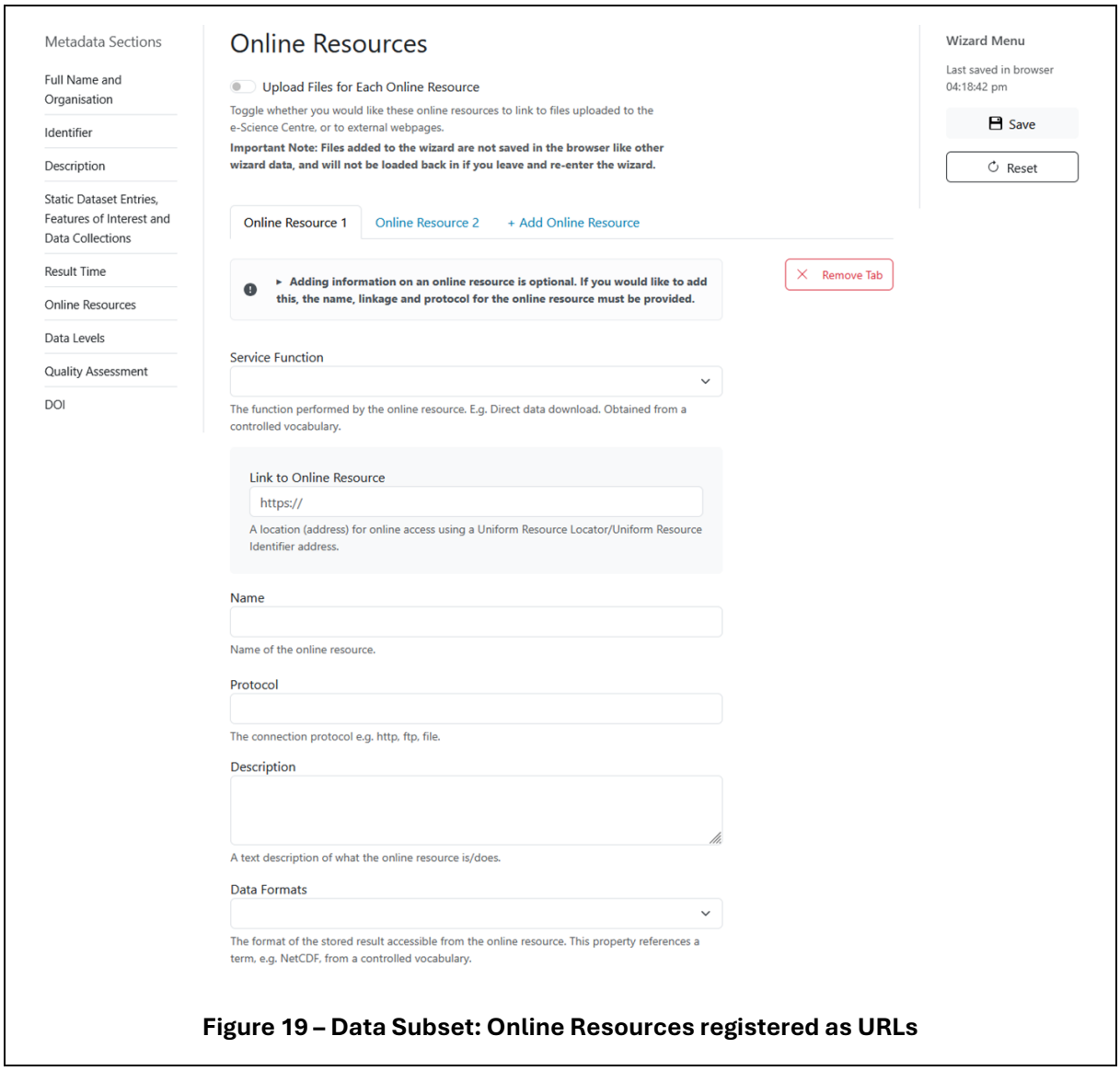

Figure 19 – Data Subset: Online Resources registered as URLs

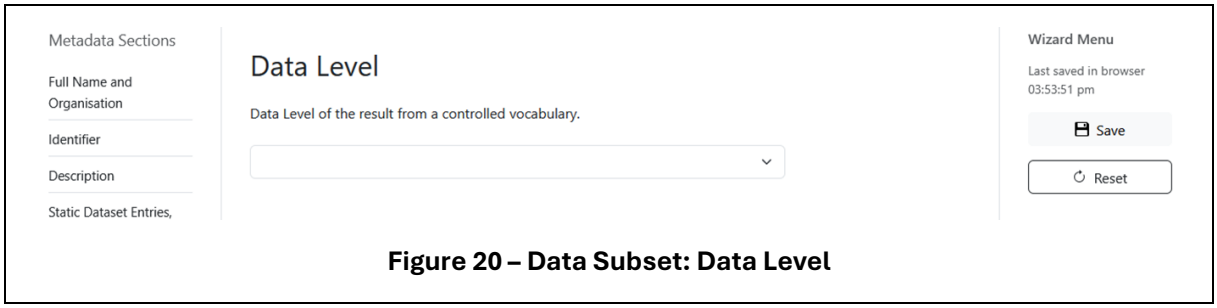

Figure 20 – Data Subset: Data Level

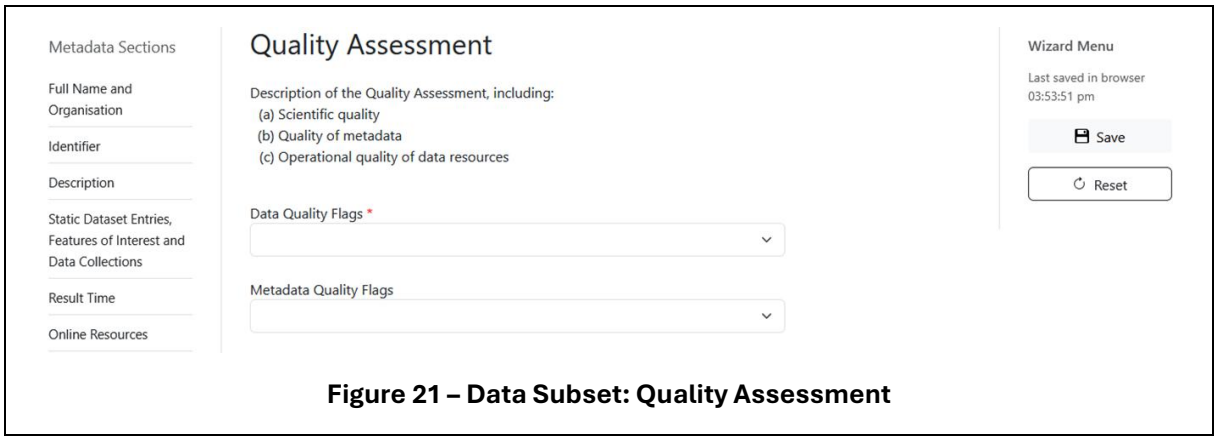

Figure 21 – Data Subset: Quality Assessment

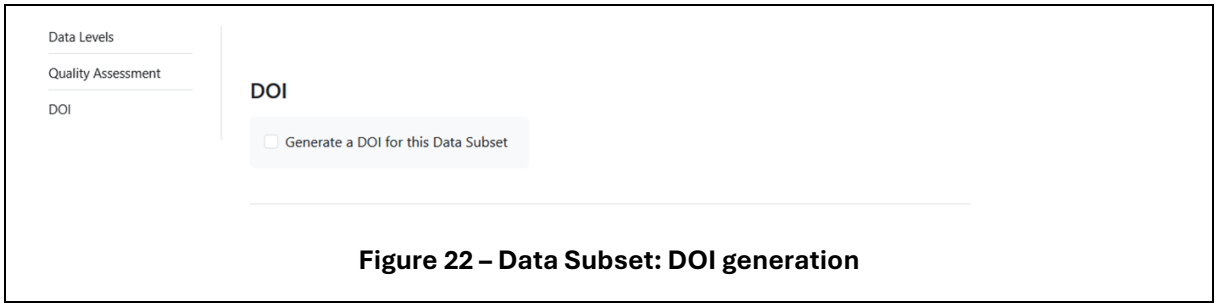

Figure 22 – Data Subset: DOI generation

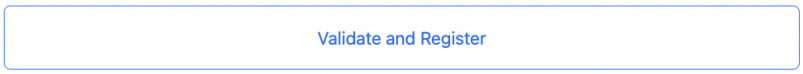

Figure 23 – Data Subset: Validate and Register