

***PROSPECT* REPOSITORY TEMPLATE**

INSTRUCTIONS TO FILL IN THE PRODUCT TEMPLATE FORM

Version 0.91, 14 April 2022

Author(s): Marco Molinaro, Dario Del Moro

Contributor(s): Monica Laurenza, Rossana De Marco, Valerio Formato, Carmelo Magnafico

Introduction

A metadata schema has been defined to help standardise the description of the various products that will be included in the CAESAR ASPIS archive prototype. Those metadata need to be filled in by the science working groups. As a help in doing so, a web-based form solution has been prepared, named ***ProSpecT*** as ***Product Specification Template***. This document provides a quick overview of the web form and its usage and a guide to help filling in the required metadata elements.

Contents

Introduction	1
Web Form overview	2
Web form usage.....	2
Metadata Groups (Tabs)	3
Product.....	3
Curation	3
Content & Policy	3
Data Description	4
Software Description	6
Functionalities.....	6
F.A.Q.	8

Web Form overview

The ProSpecT web form (available at <http://prospect-caesar.ssdsc.asi.it/>) it's composed of a set of tabs, each one taking care of a sub-group of metadata elements. The form shows four common tabs (from left to right):

- Product
- Curation
- Content & Policy
- Functionalities

A fifth one appears between Content & Policy and Functionalities depending on the selection of the Product.Type from its drop-down list; it can be one of:

- Data Description
- Software Description

Product Type selection criteria is:

- Data Description: applies for a collection of datasets, a (set of) table(s) or catalogue(s) or a model lookout (the selected Product Type is data or lookup-table)
- Software Description: applies to a software product or a computational model (data visualisation software, data analysis software, or software to model physical processes)

While the tabs are independent and no special order is required in filling them in, it is suggested to use a left to right approach to the various tabs.

Note: the latest version of this document is linked in the header of the ProSpecT web form alongside a NODE2000 contact point for enquiries and issues in filling in the product's metadata elements.

Web form usage

Anytime, while filling in the web form, input values can be:

- Cleared (CLEAR DATA button; beware, there's no undo option for this)
- Saved (DOWNLOAD button)
- Uploaded/restored (UPLOAD button)

The latter two functionalities are particularly useful if there's the need to stop working on the form and restart later without losing information, that is without the need to fill in again all the form starting from an empty one. The save option prompts the download of a JSON file with the current form content, that can then be used in upload to restart.

This functionality of the web form can also be useful in re-using general metadata for multiple products, saving time in filling in multiple times the same general content.

At the end of the filling in process, when a product is considered to be completely described, the DOWNLOAD Button is the one to be used to save the product definition to be sent to NODE2000 (potentially with accompanying sample files of data products or other useful auxiliary files).

Note: please rename the downloaded JSON in some useful way, maybe referring to the [CAESAR proposal's product list](#).

Metadata Groups (Tabs)

Here follows a description of some specific elements that might be useful in understanding the goal of some metadata and clarify what to enter at specific elements in the form. Description is subdivided by tab grouping of the metadata. Some tooltips are directly available on the web form itself, as labels for the fields or as contextual helpers that show up when clicking, e.g., on text fields.

Product

The Product tab simple requires to fill in:

- Title for the product (it should be quite complete and give it uniqueness, no character limits are foreseen, with obvious precaution)
 - Note: For CAESAR project products, the Title should be given after the [product name reported](#) in the proposal.
- Shortname, to be considered a quick mnemonic handle, it shouldn't have spaces in it, be limited to [a-zA-Z0-9_-] characters and not overflow 16 in length. Ideally it could be used to build the unique identifier;
- Unique identifier: suggest a product unique identifier. The tentative form should be *aspis://publisher/product* where *product* possibly is the above shortname and *publisher* should be similar to the publisher ID reported in the Curation section. The goal is to uniquely identify the product within CAESAR/ASPIS. Please note that NODE2000 could modify the proposed identifier for technical reasons.
 - Alternate identifiers are also possible, in case the product already has a DOI or other PID that describes, links or resolves to the product itself;
- Type of the product allows for a choice among: data, lookup-table, and software. The first two will enable the Data Description tab, the latter the Software Description one. In general terms the first two identify a product whose content is describable in terms of data resources, the other refers to codes/models that have to be run to produce an actual result.
- The Created and Updated timestamps and Status of the product are roughly self-describable and, even if set as required, at this initial step will possibly be modified and made consistent by NODE2000. They refer to the product templated description, non the product content.

Curation

Curation metadata consist of people, team, or other groups responsible of having created, contributed, published the product itself, plus contact points to interact with them about availability or issues with the product itself.

One note is worth about the Publisher ID, that might be used in composing the Product.Unique Identifier itself. In general, all IDs should follow the rules of construction of URIs. At this stage of the prototype no strict rules will be applied.

Publisher is the entity (usually an organisation, institute or research infrastructure, or a group) responsible for letting the product be publicly available (could simply be the one that proposed the product to CAESAR). Creator(s) and Contributor(s) are usually individuals. For the Contact(s) it is warmly recommended to avoid personal email addresses to avoid issues in case the individual contact point changes.

Content & Policy

This tab serves two goals.

On one side, content metadata is requested.

Subject(s) are taken from the CAESAR proposal and multiple can apply to one Product.

A textual description of the product itself is welcome for human readability (Note: For CAESAR project products, the [product short description](#) can work) and a ReferenceURL is highly encouraged to provide a finer, richer and more manageable description (with respect to updating this template). The ReferenceURL can be a web page, the access page to the product itself (better if it contains some product description), a bibliographic reference,

Content includes also a way to provide relationship links to other CAESAR/ASPIS products, like a specific software product related to a data collection or model. Related product names and IDs are welcome.

- Note: For CAESAR project products, the name should be given after the [product description reported](#) in the proposal; the ID can be left blank unless you manage the related product and already have suggested/know an ID for it.

The policy section simply asks for what kind of right the product might undergo. Public, private or mixed solutions are possible, a textual description will help understanding the details.

Data Description






This complex metadata block asks for

- The actual source of the data (Provider), where Name and Source can be considered like telescope and instrument or other two step description of the observer/data source (e.g. SOHO/LASCO, DST/IBIS, ...)
- Data Product Type, from a controlled list of Types, its level of calibration or processing (based on the [PDS4 vocabulary](#))

Plus a set of meta information about the data Format the product provides, the Bulksize and Number of artefacts (files) it is composed of, whether it is a static collection or dynamically growing/changing (Data Collection Behaviour).

A Sample Dataset (filename) for the data collection would help a lot, listed here as a filename, maybe named after the List of Products at point 16 (pages 23-50) of the CAESAR proposal, and then submitted with the form JSON download to NODE2000.

Tableset Details metadata (that shows up only if you choose tableset as the Data Product Type) is meant to describe a data collection that consists of a (limited) number of tables and columns therein. Each table can be described with its name and some textual description and with all columns where the description should contain units and other conceptual descriptors to better understand the data, as e.g. the Observable UCD (Unified Content Descriptor – see paragraph below for more info about UCD).

Table Name		
ICME_complete_dataset_v3		
Table Desc		
Table containing information about 213 ICMEs, ordered in 29 columns		
Columns		+
Column Name	Column Desc	
CME_num	index (1) (meta.code) - progressive number	
LASCO_Start	datetime (1997-01-06 16:23:38.000) (time.epoch) - Time of first detection	
Start_Date	datetime (1997-01-07 08:35:05.600) (time.epoch) - Estimated time then	
Arrival_v	km/s (450) (phys.veloc) - Measured velocity of the ICME front at L1	
Transit_time_err	hours (4.8468) (time.resolution) - Error associated with the ICME transit	

Example of Tableset description. Note that the Column Descriptors are optional. In this example, a full description has been used, with the datatype, an example, the IVOA standard UCD, and a textual description of the column content, in order.

Coverage information (spatial, temporal, spectral, messenger and observable) are also useful metadata.

- Spatial coverage is requested as one or more Spatial Location(s) based upon the proposal WP descriptions and available as a drop-down list;
- Temporal coverage is the full extension in time of the data collection (please, be aware that the form asks for a full date and time, so set time to 00:00 if only the date fits your needs);
- Spectral coverage is the range covered on the spectral/energy axis for the full collection, unit of the axis is needed to understand the range extension;
- Messenger(s) carrying the information and Observable(s) that the data collections contain are also asked for. Messenger type(s) are simple free text labels like photons, neutrons, ...;
- Observable(s) should be reported from controlled vocabularies, like [IVOA UCDs](#) (Unified Content Descriptor - use [this form](#) to search for an UCD that represents your data!) or the [Unified Astronomical Thesaurus](#) (contact NODE2000 in case you do not find any UCD suitable for describing your observable).
 - NOTE that Observable is only available for non-tableset data products (because it can be embedded, with column granularity in the columns descriptions – see example in the above figure).

Coverage

Spatial

Choose one or more locations
+

Corona
▼

Interplanetary Space (not including L1 and Planets)
▼

L1
▼

Temporal

start *
1997-01-06 00:00

stop *
2018-03-06 23:59

Example of Coverage and Temporal description. Note that more than one physical location for the product can be selected. In this example, based on an ICME information table, the locations are the Solar Corona, as the source of the CME, the interplanetary space, since the ICME travels through the IS, and Earth L1, as the arrival point.

Software Description

Software Description tab is the counterpart for describing software code or numerical model products.

It initially asks for some basic metadata about the Code itself, whether it is provided as a binary blob or source code (Type entry), its version, whether it can be already found elsewhere (Repository entry) and some documentation.

The Code Scope lets identify whether it's an analysis or visualisation software or a numerical model software.

Then it is required to enter all possible information about Input, Output & Configuration Parameters, all of them described with a Name, a Type that can be numeric, string but even file or structure, the data Domain the parameter lives in (e.g., a declination must be in the $[-90^{\circ}, 90^{\circ}]$ range) and, if available, a Default value (this could be used for internal software configuration parameters).

The Parameter Scope entry is needed to identify the parameters as being input parameters to the code/model, output of the same or internal configurations or fixed parameters.

Environment type and hardware Requirements are finally asked for.

Requirements, even if approximate, are mandatory to produce an estimate of the computing power needed to run the software and/or the typical computational time to produce the output.

Functionalities

These optional metadata are meant to connect the products to the various Functionalities the CAESAR prototype described in its proposal. See the proposal document, pages 20-21, or at this [weblink](#) for a complete list and check their relationship to products in the "functionalities" column in the [Table](#) at section 16 of the proposal itself.

Alternatively, they can provide links and description for external functionalities (like already available web based solutions or API(s) for searching and accessing the product).

Please select the expected functionalities from the dropdown and use the free text description to add information (if needed). The optional textual description is specifically important for “Other” functionalities that you may want to add.

F.A.Q.

My Product already has a consumable interface to interrogate (filter, retrieve, consume, access) it. Is there a place I can describe/alert about this feature?

The Functionalities tab is the place where, using one of the proposal-based functionalities or adding a custom one (selecting Other from the name drop down list), it is possible to describe (in free text) the existing programmatic interfaces or access points to the Product (or part of them).