Metadata for publication in GAVO's Data Center

Author: Markus Demleitner
Email: gavo@ari.uni-heidelberg.de

To make your data visible to other astronomers and the general public, the umbrella organization of the VO, the IVOA, runs a registry of services. To feed it meaningful data, and also to enable users of your service to find out whose data they are using, we ask you to provide data as specified below (and, of course, as applicable). You are most welcome to get back to us if you have questions.

The italicized text in the explanations is taken from [RMI].

Title: A short phrase (preferably less than 6 words) giving people an idea what the service is about. This will be the title and the headline on forms, etc. *Title should be an unabbreviated form (e.g., Hubble Space Telescope) rather than an acronym unless the acronym is so well known as to be part of standard usage.*

Short Name: This should be an abbreviation of less than 16 characters. If you can't think of anything sensible, don't worry, we'll make one up for you. *The ShortName will be used where brief annotations for the resource name are desired, such as in GUIs that might refer to many resources in a compact display. [...] ShortName strings are limited to a maximum of sixteen characters. Care should be taken to define illuminating ShortNames indicating either where the resource comes from or what data collection it provides. ShortNames are not required to be unique. Indeed, a resource provider may use the same ShortName for several related resources (e.g., different services that access the same collection), or the same ShortName might be used by different providers for common/mirrored resources.*

Creator Name: Well, your name and the names of your collaborators. If you have a logo, be sure to pass it to us. Some GUIs

display logos in their result listings, and that increases visiblity dramatically... *An entity primarily responsible for making the content of the resource. Examples of a Creator include a person or an organisation. Users of the resource should include Creator in subsequent credits and acknowledgments. [..] If the resource is a data collection or service accessing a collection, then Creator fields should list the scientists responsible for the original data collection. Typically, this would be list of authors associated with the defining published paper for the collection. [...] Full names should be given, not just surnames.*

Version: Optional. If you think you will make "data releases" and old "releases" might be required to stick around, giving a version is a good idea. In those cases, you should talk to us, since we might need to plan how old releases might be kept. For most cases, we believe this versioning adds largely unnecessary overhead (until a service is very successful and in some way changes behaviour, that is).

Subjects: *A list of the topics, object types, or other descriptive keywords about the resource. [...] To support keyword-based searches of registry contents, the Subject element should be as specific as possible and include as many relevant terms as possible.* If at all possible, the subjects should come from a controlled vocabulary. We recommend you first check the IVOA Thesaurus (the dictionary may be useful, or download the the turtle version and use what's under skos:definition), but both Astronomy Journal Keywords and Vizier Keywords will do fine if you already have those. In the GAVO DC the first term is *always* capitalized, successive words should be lower case if at all posssible.

Description: *Thorough text descriptions are particularly encouraged in order to make text-based searches against the registries maximally useful. Description should emphasize what the resource is about, as other matters such as who created it, when it was created, and where it is located are described elsewhere in the resource metadata.* This should not be a complete article, but *at most* something like an abstract. In general, we feel a few lines would be optimal here. For anything more, we can include a "long documentation" that will be exposed in the service info. Please write in the third person (i.e., "Fantasurvey is..." rather than "We present the fantasurvey...")

Source: Optional. Some paper describing what you did to arrive at your data, if you already have published one. Preferably, we'd like to see a bibcode here.

Reference URL: This is *usually* generated by us; it is information on the service in a human-readable from. In special cases, you may want to override that, but if you plan to do this (or are curious why you might want to), talk to us.

Facility: Optional. *The observatory or facility where the data was obtained. Some resources are likely to hold data from multiple observatories. If just a few, this could be a list; if very many, just say "many". Theoretical data will not originate with an observatory, but rather might be characterized by the computational facility used to create them (NCSA, SDSC, etc.). Facility should be used only to describe entities that specifically produce or manage data. Observatory names are the most common values.* Actually, facilities can receive identifiers that in turn makes more information about a facility accessible within the VO. GAVO can give you such an identifier if you want.

Instrument: Optional. *Can be a specific instrument name (Wide Field/Planetary Camera 2) or generic instrument type (CCD camera). Theoretical data is produced by a computer code, and the name of the code could be specified.*

Coverage: Optional. This is supposed to reflect what part of the sky, the electromagnetic spectrum, etc., the data is talking about. In general, it consists of a "waveband" (Optical, Infrared, etc) and of a more precise specification of the spatio-temporal extent of the observations. Leave the technicalities to us and just say something like "the observations have been taken between 1980-01-04 and 2010-10-30 and cover positions between 30,40, and 60, 120 degrees ICRS; we only have objects of redshifts between 0.1 and 0.2".

Copyright information: If you have any special wishes ("if you use..., please acknowledge..."), or if you want to restrict what people can do with your data, this would go here. GAVO offers "embargoed" publication where, for a limited time, only persons having a password can access data files or a service.

Service profile: Certain VO protocols have a notion of "verbosity", giving more or less fields depending on the query. For data giving enough columns, it would be nice if you gave us your assessment of the "importance" of a column, going from 1 (most important, data is worthless without this column) to 30 (only very specific and exotic applications will need this column).

Other issues

If you have data "with positions" (i.e., your data will be published using the Simple Cone Search (SCS) protocol), a unique identifier for each table row is required. While a row counter will do fine, you will probably do yourself a favour if you follow the IAU conventions ([IAU]).

If you have any wishes concerning the appearance of your service (on the Web or within the VO), just let us know. We can do a lot of customization and even if we can't, we'd still like to know about them.

Bibliography

[RMI] Hanisch, R., et al, "Resource Metadata for the Virtual Observatory", http://www.ivoa.net/Documents/latest/RM.html

[IAU] IAU Commission 5, "Specifications concerning designations for astronomical radiation sources outside the solar system", http://cdsweb.u-strasbg.fr/Dic/iau-spec.html