Standard Protocols and Tools: Reaping the Benefits

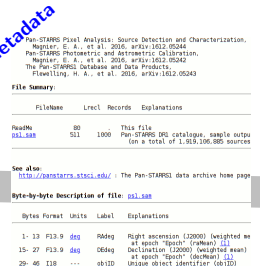
Markus Demleitner
Universität Heidelberg, Astronomisches Rechen-Institut, Mönchhofstraße 12-14, Germany
msdemlei@ari.uni-heidelberg.de

Once standards for data discovery and data dissemination are defined, standard software solutions on both the server and the client side become possible. This, in turn, facilitates rich data services accessible through a multitude of tools as part of a comprehensive data ecosystem with a very moderate effort on the publishers' side (and zero effort on the users' side). This poster illustrates this by drawing a cartoon of a publishing workflow using GAVO's DaCHS, an integrated Virtual Observatory publishing solution implementing most of the Virtual Observatory's protocols, starting from the data providers, going through the data center operators, and reaching its goal in the data exploitation with powerful, standards-based tools like TOPCAT, Aladin, SPLAT, VESPA's data browser, or the astropy affiliated package pyVO.

Data Provider

It all starts with the instruments that take the data. We gloss over data reduction here; our publishing workflow mostly starts after that.

Metadata



Data providers deliver metadata typically in unstructured (papers) or semistructured (readme files, FITS headers) form. The example shows a "byte-by-byte" description that the VizieR archive service has made a quasi-standard for tabular data in astronomy.

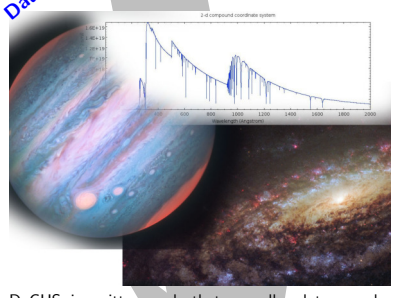
Operator

In DaCHS lingo, "operators" are the persons that turn the data into well-described data collections exposed in standard formats through standard protocols. The "users" in DaCHS are the consumers of the operators' services.

Review

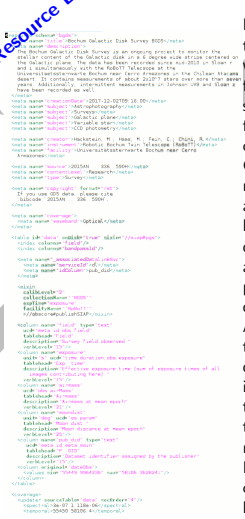
In all publishing efforts we had so far, at least one iteration of providing a prototype service and gather feedback from the data providers was necessary to straighten out kinks in the data, misinterpretations on the side of the operators, or simply usability points raised by the data providers.

Data



DaCHS is written such that, usually, data can be taken directly from the data providers. Standards like Datalink facilitate decoupling the data providers' science products from, potentially more standards-compliant, data sets delivered to users if so desired, while still offering access to the original, upstream data.

Resource Descriptor



For the operator, "publishing" in the DaCHS world mainly means to write a resource descriptor, or RD for short. The image on the left shows the beginning of the the RD for a live resource at GAVO's data center. At the top, there is Dublin Core-style metadata. Next, the main metadata table for this image collection is defined, re-using built-in definitions for a VO standard. In this way, the declarations are reasonably compact while still producing rich and complete metadata on output.
At the bottom end you can see the beginning of coverage metadata, i.e., the location of the data in space, time, and spectrum. Further items in the RD include ingestion rules, service declarations, and regression tests.
RDs in use at the Heidelberg data center are publicly available (see below).

DaCHS

Virtual Observatory Publishing

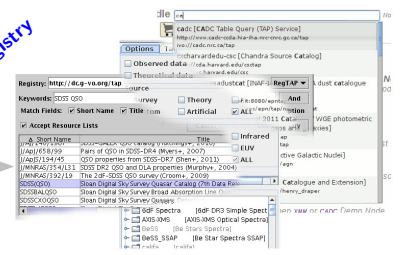
DaCHS, the Data Center Helper Suite, is an integrated publishing package. To operators, it presents an abstraction of the publishing process and spares them the implementation effort for the various protocols and standards involved in VO publishing (OAI-PMH, VOTable, UCD, ADQL, TAP, SIAP, SSAP, Obscore, SDM, and many more). In particular, its integrated metadata management means that a single piece of metadata needs only be declared once and is then re-used in ingestion, registration, querying, and result delivery. To users, it presents both protocol and web interfaces.

VO Resource Record

VOResource: an XML Encoding Schema for Resource Metadata

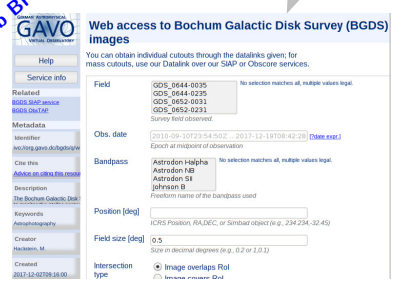
DaCHS can generate documents in the Virtual Observatory's resource metadata schema, VOResource, and will exchange them with the VO infrastructure using standard OAI-PMH, giving operators minimal-hassle registration.

VO Registry



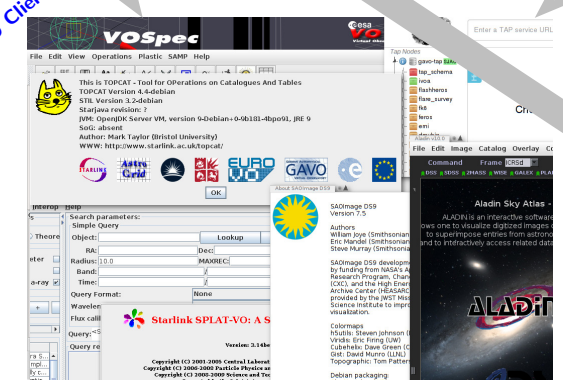
The Virtual Observatory registry is a set of about 20000 metadata records of data services and ancillary resources in astronomy. Both protocol (RegTAP) and Web interfaces are available.

Web Browser



DaCHS has facilities for enabling access to services through web browsers at essentially no additional cost, and lets operators define advanced web browser-based interactions. We are convinced, however, that web pages do not scale to a world of diverse and massive data; thus, DaCHS always encourages operators to think protocol access first.

VO Client



Currently, the most common mode of accessing VO facilities is through desktop client programs. Many astronomical programs are VO-enabled. The illustration has pixels from TOPCAT, Aladin, Splat, ds9, TAPHandle, and VOSpec.

User Program

```
def search_one_service(svc):
  print("...now querying ", svc.res_title)
  svc = sia.SIAService(svc.access_url)
  for pos in [
      (10, 20)]:
    images = svc.search(pos, 0.5, verbosity=2)
    for rec_no, match in enumerate(images):
      if not DATE_MIN<match.dateobs<DATE_MAX:
        continue
      match.cachedataset()

def main():
  for svc in registry.search(servicetype="image"):
    search_one_service(svc)
```

One of the great things about standards-compliant services is that it people can program against them without having to resort to provider-specific screen scraping. A particularly straightforward way to access both VO services and the Registry is the astropy-affiliated package pyVO.

Advantages of publishing to the Virtual Observatory with DaCHS as illustrated in the figure include

- Software installation is as easy as setting up a Debian machine and apt-get installing.
- Proven components allow ingestion and service interfaces with no or very little custom code
- Built-in service registration with negligible overhead
- Rich ecosystem of client software giving users rich functionality at no implemenation or support cost to data providers or service operators.

Further Reading

- DaCHS downloads and documentation: http://soft.g-vo.org/dachs
- For a more thorough discussion of DaCHS, see Demleitner, M., Neves, M.C, Rothmaier, F., Wambsganss, J, 2014: "Virtual Observatory Publishing with DaCHS", Astronomy and Computing **7**, 27
- RDs of services published in Heidelberg: http://svn.ari.uni-heidelberg.de/svn/gavo/hdinputs/
- More on the VO protocols mentioned: http://ivoa.net/documents
- VO client programs: http://ivoa.net/astronomers/applications.html
- Programmatic access with pyVO: http://github.com/pyvirtobs
- VO tutorials and use cases: http://dc.g-vo.org/VOTT

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

This work was supported by BMBF grant 05A17VH2 ("e-inf-astro") and various previous BMBF grants.