Simulation Runs Discovery Service
Spoke3 WP4

Hendrik Heinl, INAF-OATS

May 26, 2025

SPOKE 3
Astrophysics
& Cosmos Observations

Agenda

Data Publishing

SimDM

Where are we ?

References

Why bother ?

In observational Astronomy it's commonly accepted best practice to publish data, and ADS and Vizier are used – or at least known. For Simulation Data unfortunately not so much yet. The reasons are manyfold. There are not many published data, and where there is not much data, researchers don't search for it (and demand it), thus there is no established sharing culture as for the observational community.

Why bother ?

Of course there are some published simulation data, like the millennium database. And of course when mentioning the observational astronomy, one should admit that there are communities that are more or less closed, which is changing slowly, like radio astronomy or high energy physics.

But – why me ?

- following good practice
- peer pressure
- because editors from Vizier ask for it
- enabling others to find the treasures in data thus...
- improving science by selfless good practice to ...
- gain reputation (citations)
- because people paid for it with their taxes.
- somebody wrote in a funding proposal to do so.

The VO approach – data discovery

The VO is a set of standards and protocols to enable data discovery, data access and interoperability.
Standards and protocols enable you to use a client software of your choice to ask question to a registry ('search engine') like

- Which service offers GaiaDR3 time series to X-match with SDSS and 2MASS ?
- Are there data sets in radio data available for this specific position on the sky (ra, dec, radius) ?
- Are spectral data available in IR for my local set of transient events within the correlating time intervals ($< starttime >, < endtime >$) ?

The VO approach – metadata schema

To enable searches like these examples the data needs to be

- rich of annotated metadata
- registered in a searchable registry
- delivered through standardized protocols and
- in standardized data formats

The IVOA looks back to almost 25 of experience of developing these standards and protocols.

Interdisciplinary: 'FAIR'

- **F**indable
- **A**ccesible
- **I**nteroperable
- **R**eusable

VO for Simulations

Also, the VO provides a data model and a metadata schema for simulation models: SimDM.

Unfortunately implementations are scarse so it's fair to say, there is no experience out there when it comes to standardized simulation data publishing.

Spoke3 challenge:

- 3 Groups with different demands
- No expert for publishing Simulations via SimDM available
- Finding a common language between data stewards and scientists
- Adapting the SimDM metadata schema
- Customizing where necessary
- Defining features of the data access service
- publishing to the VO registry

What we have

- a metadata schema to annotate metadat and its relations to the actual data
- baby steps towards a common language
- a definition of what a minimal service should provide, and therefore...
- a definition of the minimal metadata requirements for each of these services ...
- to publish them FAIR through the VO registry.

What's up next ?

- More interaction with the 3 groups to improve and adapt the meta data schema
- find a place for the data
- implement the data access services
- publishing to the VO registry

But...

as so often the devil is in the details. From past experience I don't expect that what we have now will be the final implementation of the meta data schema. Eventually we are aiming to give feedback with our experience to the IVOA and Improve SimDM and recommendations to simulation publishing.

FAQ

We already had a few questions coming up, which we'd like to address:

- Once we have the schema for our group – can we then simply upload our simulation data to your service and will be published ? [No. At least not simply]
- Can we use this to publish Simulation data not origining from Spoke3 ? [That's the goal. But please have patience.]

Thank you! ... and references

- https://ivoa.net/
- https://www.go-fair.org/
- https://www.ivoa.net/documents/SimDM/20120503/REC-SimulationDataModel-1.00-20120503.htm