

# The HerMES archive: a data provider's perspective



*Evanthia Hatziminaoglou*

*EURO-VO Facility Center Astronomer  
(HerMES Archiving & VO WG lead)*

# The HerMES archive: a data provider's perspective

*(“the road to hell is paved with fair decisions”)*



*Evanthia Hatziminaoglou*

*EURO-VO Facility Center Astronomer  
(HerMES Archiving & VO WG lead)*

**Famous quotes ...**

# Famous quotes ...

- *I/OA docs not good for the casual user*

# Famous quotes ...

- *IVOA docs not good for the casual user*
- *take-up of standards by astronomers*

# Famous quotes ...

- *IVOA docs not good for the casual user*
- *take-up of standards by astronomers*
- *a normal astronomer should never read this ... [oeuvre d'art]*

# Famous quotes ...

- *IVOA docs not good for the casual user*
- *take-up of standards by astronomers*
- *a normal astronomer should never read this ... [oeuvre d'art]*
- *there are some things you can do first and then proceed step-by-step*

# Famous quotes ...

- *IVOA docs not good for the casual user*
- *take-up of standards by astronomers*
- *a normal astronomer should never read this ... [oeuvre d'art]*
- *there are some things you can do first and then proceed step-by-step*
- *we must do everything we can to facilitate implementation*



# Famous quotes ...

- *IVOA docs not good for the casual user*
- *take-up of standards by astronomers*
- *a normal astronomer should never read this ... [oeuvre d'art]*
- *there are some things you can do first and then proceed step-by-step*
- *we must do everything we can to facilitate implementation*
- *everybody that wants to publish in the VO has attended a workshop*

# HerMES: the Herschel Multi-tiered extragalactic Survey

- *Herschel-SPIRE GT project, ~900h, PI S. Oliver, J. Bock*
- *6 levels (top level: deep & small; bottom level wide and shallow)*
- *includes everybody's favourite fields (CDFS, HDFN, Lockman, ELAIS N1, COSMOS, ...) ⇒ **LOTS** of data*

# HerMES: the Herschel Multi-tiered extragalactic Survey

- *will have a centralised archive for all Herschel and ancillary data, in different fields, different depths and different instruments/wavelengths, including images, spectra and catalogues*
- *HerMES is to have a VO compliant archive from its very beginning*

***but:*** *not everybody in the team found this necessary or even useful*

# Herschel data

- *processed by different data centres (PACS, SPIRE, PACS+SPIRE, XID)*
- *SPIRE & PACS images + catalogues*

# Ancillary data collection

- *dedicated groups per field, per wavelength (with cross-talks)*
- *dedicated groups for spectroscopic and photometric redshifts*
- *groups interested in specific types of objects (e.g. AGN, high-z objects, submm sources etc)*
- *each group combining heterogeneous data collections in a variety of ways and formats (images, catalogues and spectra); let's not talk about data quality issues*

# Preparation of the data & upload

- *list of mandatory keywords for data uniformity (e.g. provider, reference, contact point)*
- *VO requirements for SIA, SSA and TAP (first links to SIA & SSA docs; then list of mandatory keywords)*
- *some small scripts for people to:*
  - *convert tables between various formats (DB only ingests TSV)*
  - *extract the metadata in "standardised" readme files*
  - *add or modify existing metadata (e.g. UCDs)*
- *upload the data via a dedicated interface*

# Preparation of the data & upload

- *list of mandatory keywords for data uniformity (e.g. provider, reference, contact point)*
- *VO requirements for SIA, SSA and TAP (first links to SIA & SSA docs; then list of mandatory keywords)* (Not without: B. Rino & A. Micol)
- *some small scripts for people to:*
  - *convert tables between various formats (DB only ingests TSV)*
  - *extract the metadata in "standardised" readme files*
  - *add or modify existing metadata (e.g. UCDs)*
- *upload the data via a dedicated interface*

# Preparation of the data & upload

- *list of mandatory keywords for data uniformity (e.g. provider, reference, contact point)*
- *VO requirements for SIA, SSA and TAP (first links to SIA & SSA docs; then list of mandatory keywords)* (Not without: B. Rino & A. Micol)
- *some small scripts for people to:*
  - *convert tables between various formats (DB only ingests TSV)*
  - *extract the metadata in "standardised" readme files*
  - *add or modify existing metadata (e.g. UCDs)* (Many thanks to M. Taylor)
- *upload the data via a dedicated interface*



# Issues along the way (I): data collection and collation

- *the knowledge was fragmented, people mostly knew about their own data or data from collaborations they were participating*
- *few people knew how to retrieve data from VO interfaces (I even got a request on how to extract a bunch of tables from published papers)*
- *putting everything together was very difficult for those that did not have dedicated tools; we will end up with a long collection of ascii, vot, fits etc files, often per field, per wavelength (and a lot of e-mails in my mailbox)*

# Issues along the way (II): VO compliance

- *nobody knew about VO standards (VO astronomer is now leading the Archiving & VO effort but this can not be the case for all projects)*
- *even after links to the VO standards were put in the project's twiki, people would not read them*
- *getting the lists of mandatory & recommended keywords from the SIA and SSA documents was not straight forward (and for TAP they were not even there) + there were a lot of inconsistencies (e.g. UCD I, UCD I+; utypes - no utypes)*
- *when the documents were distilled down to lists of mandatory keywords, people still did not know (and should not have to learn) what UCDs were, some thought these keywords should be in the image headers etc (and even more e-mails ...)*

# What makes this case a “best-case” scenario

- *the decision to go VO was made at the beginning*
- *the central database will be run by an institute that has already published data in the VO (so they have the expertise)*
- *one of the projects that requested support from the second AIDA research initiative came from a HerMES group that wants to centralise part of the data set as a parallel effort*
- *there are a couple of VO-compliant people onboard*

## To summarise ...

- *the project has the correct attitude (going VO from the beginning)*
- *it has some VO insiders*
- *some request for support has been made at the level of data collection*
- *The VO concept is not clear*
- *VO tools and services are not well known*
- *the projects are overwhelmed by the amounts of data and work and they do not want to make this extra effort*
- *the IVOA documents are not written for astronomers (and it is wrong to assume that computer scientists are available for each such project)*