ESCAPE Extended Discussion Day

WP5 + WP4 Vision for data access

Yan Grange, with mostly input from Dave Morris, Hendrik Heinl and Nico Vermaas
28 September 2021

VO introduction (thanks Hendrik!)

The Virtual Observatory is a set of standards and protocols to enable user in astronomy to Find, access and (re)use dataproducts. In modern terms, this would be covered by the FAIR principles.

Data shall be:
Findable
Accessible
Interopable
Reusable

Several VO standards are used to enable data publishers and user to achieve this goals.

VO introduction: Findable

To find data, the **VO registry** is the entry point. Here, services register their data with rich meta data in a way that a user can find the data by "asking" question to the VO registry like:

Where do I find images specialized in radio ?

What data sets exists which provide x-ray fluxes and proper motions ?

What services provide data associated to a paper ?

VO introduction: Accessible

The VO-protocols depend on the kind of data you want to access. Some of these are: **Simple Cone Search Protocol** (SCS), **Simple Image Acces Protocol** (SIAP), **Simple Spectral Access Protocol** (SIAP) and the **Table Access Protocol** (TAP/ADQL).

Some of these standards are more intuively learned, like the the typed interfaces of **SCS, SIAP** and **SSAP**, others require some efforts to master (**ADQL**, **VO-Table**).

Interoperability demands well defined standards of data formats so that clients are able to read the data, and protocols to move data between machines or software. Within the VO the standard format is the meta data rich **VO-Table**.

A special protocol for interoperability between VO-clients is the **Simple Application Messaging Protocol** (SAMP). It enables you to easily sent data between applications, but also let's you make a selection in one application that immediately is visible in another. Check it out! It's very useful, and a bit fun to figure out.

VO introduction: Reusable

To make data reusable, a surprising lot of things has to be taken into account. Crucial is rich meta data to describe what is found in the actual data.

A lot of the VO-standards deal with this kind of meta data and to keep it close to the data (e.g. **VO-Table**, **utypes**, **UCDs**, **VOUnits**).

Quite a bit of effort of the recent years was spent to develop a **VO data provenance model**, and it is still improving.

Goals of VO implementation

- Provide users the ability to query astronomical data archives world-wide and provide access to them in the processing (e.g. Jupyter notebook) through the shopping basket.
- Different data formats should be accessible:
  - Images
  - Spectra
  - Cubes
  - But also: catalogues, other tabular data
- The VO is all about metadata. This means that even when querying for a data set, the resulting metadata should all be kept together.

SAMP

- Currently we have SAMP implemented.
  - Query using 'your favourite VO tool' (Aladin, TOPCAT, etc) and broadcasting result to ESAP -> shopping basket
- In principle, this could be just it.
  - Certainly works for smaller queries
  - Not sure if this works for queries with many results -> huge tables
  - Also if a user just wants to do a cone search for data, asking them to do it in an external tool may be overhead (isn't that the added value of ESAP to begin with?)
- So the access protocols would be useful to implement.

Direct protocol access

- Minimal implementation: cone search for inexperienced user. ADQL is fairly simple to implement (single box) and helps advanced users.
- There is an original implementation of TAP/ADQL.
  - Using pyvo
  - Currently only the ASTRON-VO version works.
  - Implementing SIAP, SSAP, SLAP would be fairly straightforward.
- Query of the registry for services offering specific data, then querying the services for location
  - Returning all results for a cone.

The shopping basket

- Expectation management is an important component here
  - Do we expect users to be able to have milions of rows in the shopping basket? Wheter intentional or by accident.
  - If so: does the current design support this
  - If not: how do we prevent VO queries to become too large (there may be a point to be made for ESAP in general here)?
- ESAP-DB may be a way of supporting large queries.
- What size would anyway be useful to support in a notebook?
- So question boils down to: is this a problem at all?

Proposed design

- Ideally 1 query field (cone search) to query the VO for data.
  - Selection on wavalength, other parameters?
- Selection box for type of data (spectra, image, line, table)
- ADQL -> single box
- Should all meta data appear in the shopping basket overview? Can immagine we want all the metadata to go to the notebook, but that would make the shopping basket very hard to parse for a human brain.
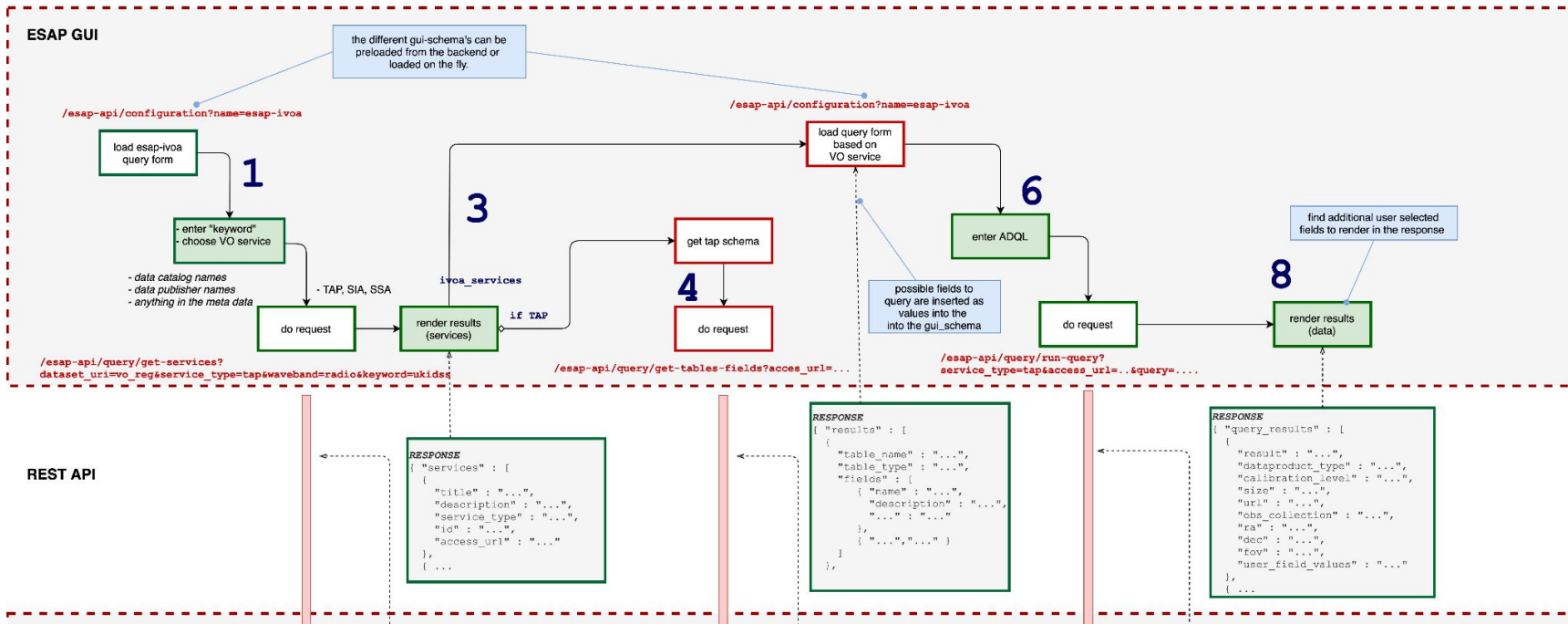
Original implementation (top half)

Original implementation (bottom half)
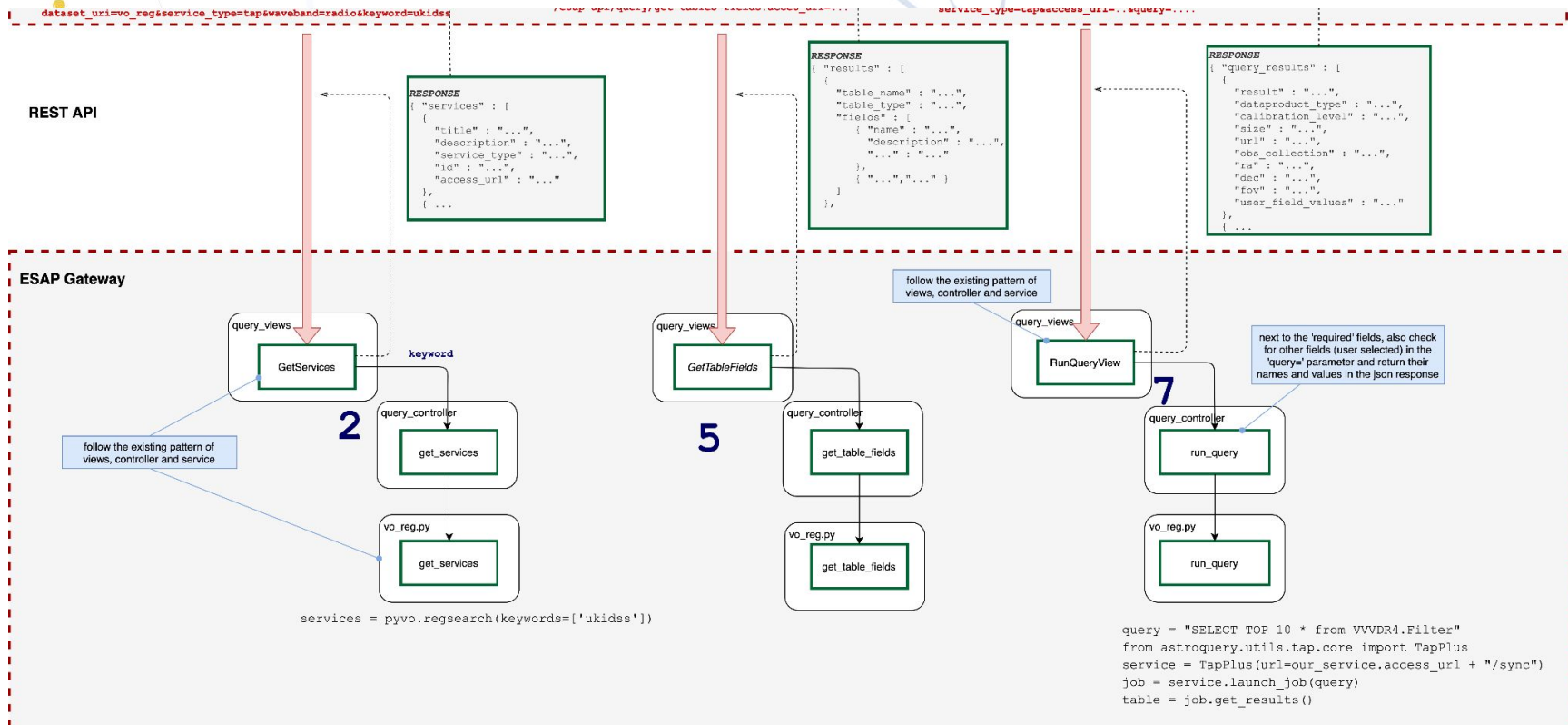
Other topics worth discussion

- UWS: Workflow execution
  https://ivoa.net/documents/UWS/
    - Abstract interface for executing asynchronous jobs
    - Used in TAP to manage asynchronous queries
    - INAF and UEDIN are working on a prototype to launch Docker containers via a REST webservice API

- VOSpaces: Distributed storage
  https://ivoa.net/documents/VOSpace/
    - VOSpace is a simple data transfer service
    - Rucio is ideal for managing data within ESCAPE domain
    - VOSpace is useful for transferring data between ESCAPE and external domains, e.g. INAF archive in Italy or CADC in Canada
    - INAF and UEDIN are working on a prototype VOSpace|Rucio bridge

Other topics worth discussion

- ProvTAP: Provenance information
  - TAP implementation of the Provenance data model
  - Access to Provenance data using standard TAP client
  - https://ivoa.net/documents/ProvenanceDM/

- Table data in the non-astro ESCAPE fields (ie particle physics)
  - This would require the IVOA to endorse additional vocabularies from outside the astronomy domain for UCDs and DataModels

Other topics worth discussion

- Bidirectional SAMP

Would it be useful to query using ESAP, or broadcast information from the processing end to 'my favourite VO tool'?

Documented view

- https://git.astron.nl/grange/esap-vo-doc