Lecture 22 – Client's perspective

Open Data Management & the Cloud

(Data Science & Scientific Computing / UniTS – DMG)

Lecture summary

- What do we work for
- How to collect data: ESCAPE demo – IVOA SAMP
- CDS portal
- SIMBAD, VizieR, Aladin

- Final project hints

What do we work for

- "Open science is the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process."

  Michael Nielsen
  OpenScience ASAP

- Open, FAIR, linked, interoperable, … data

- Formats, models, metadata, identifiers, vocabularies, … for resources
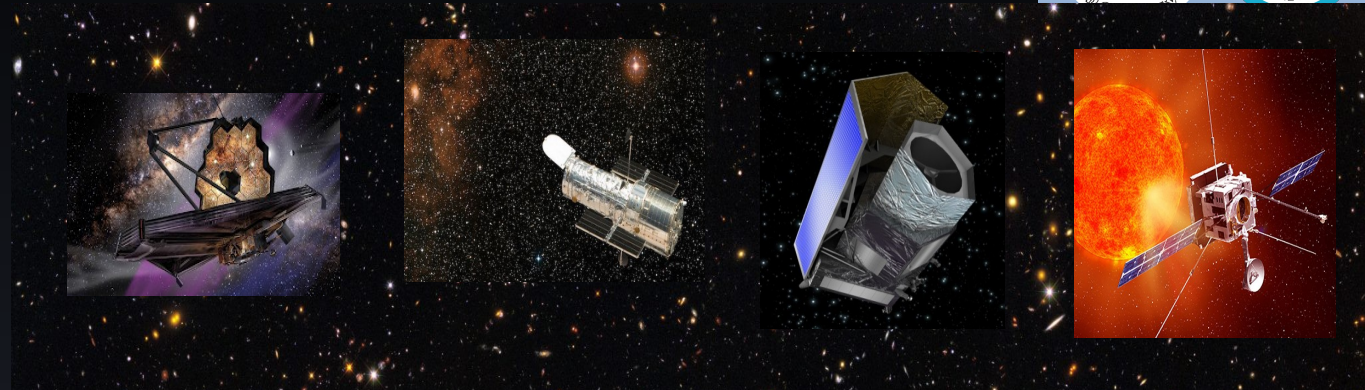
- Data management: the data providers side …

  Let's have a look at the consumer's end …

Through the Virtual Observatory interrogate multiple geographycally distributed data centers in a seamless and transparent way,like observing with a single instrument

VO

IVOA

IVOA: International Virtual Observatory Alliance (see Lecture 4 on FAIR Principles for an introduction)
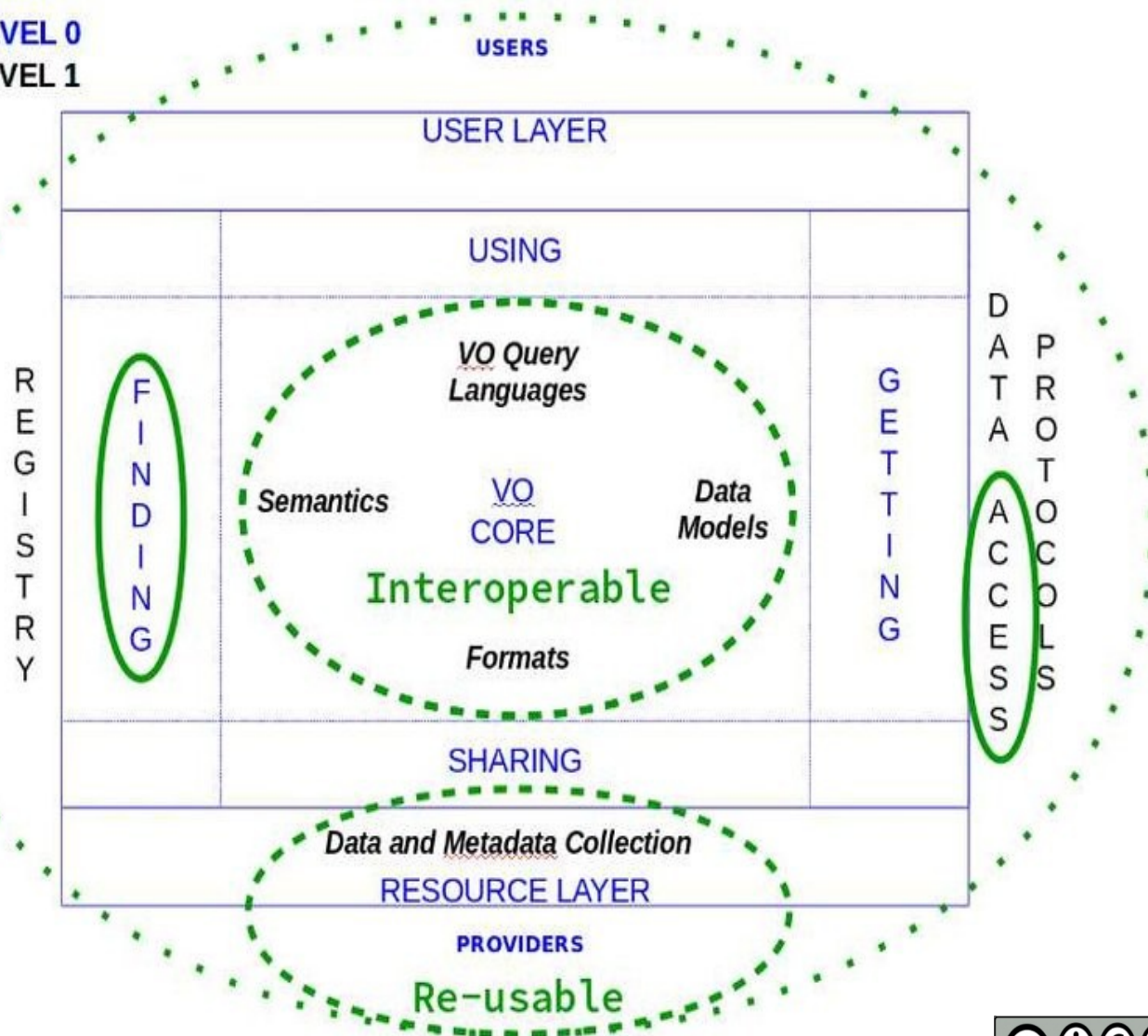
https://www.ivoa.net/documents/SAMP/20120411/REC-SAMP-1.3-20120411.html

*Grants: F.Pasian, M.Molinaro*

IVOA-SAMP implementation demo

SAMP: Simple Application Messaging Protocol

is a messaging protocol that enables astronomy software tools to interoperate and communicate. It supports communication between applications on the desktop and in web browsers.

https://www.ivoa.net/documents/SAMP/20120411/REC-SAMP-1.3-20120411.html

IVOA-SAMP for data collection

SAMP has a hub-based architecture.
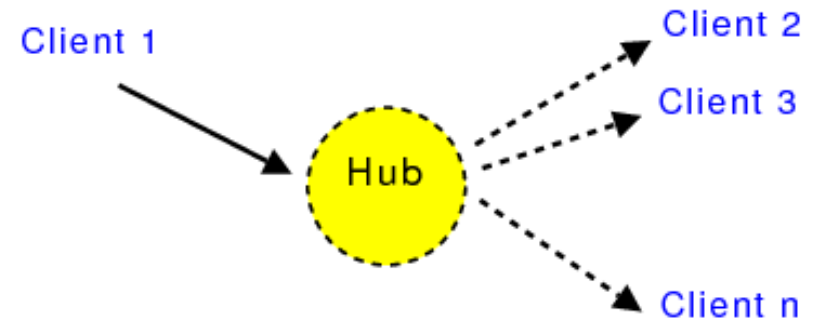
The hub is a single service used to route all messages between clients.

This makes application discovery more straightforward in that each client only needs to locate the hub, and the services provided by the hub are intended to simplify the actions of the client.

A disadvantage of this architecture is that the hub may be a message bottleneck and potential single point of failure.

SAMP may not be suitable for extremely high throughput requirements;

may be mitigated by an appropriate strategy for hub restart if failure is likely.

https://www.ivoa.net/documents/SAMP/20120411/REC-SAMP-1.3-20120411.html

IVOA-SAMP implementation example

ESCAPE EU Project   https://projectescape.eu/

European Science Cluster of Astronomy \& Particle physics ESFRI research infrastructures

brings together the astronomy, astroparticle, and particle physics communities to address fundamental challenges in data-driven research, inspired by the goals and needs of major European research infrastructures, or ESFRIs

ESAP   https://sdc-dev.astron.nl/esap-gui/

https://projectescape.eu/services/esfris-science-analysis-platform-esap

ESFRI Science Analysis Platform

is a platform-service gateway with the capability to access and combine data from multiple collections and stage for subsequent processing and analysis. It allows data discovery and handling of large and distributed data collections. It is a flexible science platform for the analysis of open access data available through EOSC

IVOA-SAMP implementation demo

Open Topcat

java -jar /home/bertocco/work/VO_School_feb2021/topcat-full.jar

Open esap-samp

https://sdc-dev.astron.nl/esap-gui/samp

Hit register

Follow the first example of the tutorial:

~/work/VO_School_feb2021/slides/topcat$ evince topcat_tutorial.pdf

What is TOPCAT?

TOPCAT is an interactive graphical viewer and editor for tabular data.

- analysis and manipulation of source catalogues and other tables

- It understands a number of different astronomically important formats (including FITS, VOTable and CDF)

- Good at interactive exploration of large (multi-million row, lots of columns) tables.

- Using a powerful and extensible Java-based expression language new columns can be defined and row subsets selected for separate analysis. Table data and metadata can be edited and the resulting modified table can be written out in a wide range of output formats.

- It is a stand-alone application which works quite happily with no network connection. However, because it uses Virtual Observatory (VO) standards, it can cooperate smoothly with other tools, services and datasets in the VO world and beyond.

CDS portal example

- CDS: Centre de Données astronomiques de Strasbourg (The Strasbourg astronomical Data Center)

  - CDS Portal

    - http://cdsportal.u-strasbg.fr/

    - Simbad, Vizier, Aladin(Lite)

    - Aladin, (TOPCAT)

Tutorials:

- https://www.asterics2020.eu/dokuwiki/lib/exe/fetch.php?media=open:wp4:school4:tutorial-cds.pdf

- "Hyantis" tutorial

  - https://github.com/hendhd/hyantis

    - See the PDF step guide there

  - TOPCAT, PyVO (python), Aladin

What is CDS

- CDS: Centre de Données astronomiques de Strasbourg (The Strasbourg astronomical Data Center)

  It is dedicated to the collection and worldwide distribution of astronomical data/objects and related information in electronic form

- Main services hosted:

  - **SIMBAD** astronomical database for the identification of astronomical objects

  - **VizieR** the catalogue service for the CDS reference collection of astronomical catalogues and tables published in academic journals

  - **Aladin** interactive software sky atlas for access, visualization and analysis of astronomical images, surveys, catalogues, databases and related data

SIMBAD

- **SIMBAD: S**et of **I**dentification, **M**easurements and **B**ibliography for **A**stronomical **D**ata

- The <u>purpose</u> of Simbad is to collect and provide information on astronomical objects of interest which have been studied in scientific articles.

- Simbad <u>is</u> a dynamic database, updated every working day.

- It provides:

  - the bibliography

  - available basic information, such as the nature of the object, its coordinates, magnitudes, proper motions and parallax, velocity/redshift, size, spectral or morphological type, and the multitude of names (identifiers) given in the literature.

    The CDS team also performs cross-identifications based on the compatibility of several parameters, in the limit of a reasonably good astrometry.

- By construction it is highly inhomogeneous as data come from any kind of instruments at all wavelenghts with any resolution and astrometry, and different names from one publication to another.

- Simbad is not a catalogue, and should not be used as a catalogue.

- Bibliography for SIMBAD:

    The SIMBAD astronomical database (https://arxiv.org/pdf/astro-ph/0002110.pdf)

VizieR

- **VizieR** is an astronomical catalog service collecting a heterogeneous set of catalogues and data and allowing uniform access to them
  - The astronomical catalogues are kept in their original form
  - homogeneous descriptions of all these data sets are provided through metadata in order to maximize their usability
  - the set of machine-readable astronomical catalogues is transformed into a set of machine-understandable data

- VizieR actually consists in an interface able to query this set of machine-understandable astronomical catalogues

- Bibliography for VizieR:

  The VizieR database of Astronomical Catalogues

  (https://arxiv.org/abs/astro-ph/0002122)

Aladin

- **Aladin** is an interactive software sky atlas

- provides simultaneous access to digitized sky photographs, catalogs and databases

- interactive access related data and information from

  - the SIMBAD database,

  - the VizieR service

  - other archives

    for all known sources in the field

- facilitate

  - visualization of digitized astronomical images,

  - visual comparison of user data with previously classified data,

  - automatic source extraction and calibration tools.

- Bibliography for Aladin:

    The ALADIN interactive sky atlas

    (https://doi.org/10.1051/aas:2000331   and   https://doi.org/10.1017/S007418090012933X)

HiPS : Hierarchical Progressive Surveys

HiPS is the hierarchical tiling mechanism which allows to access, visualize and browse seamlessly image, catalogue and cube data.

https://aladin.u-strasbg.fr/hips/

Video tutorials

- Virtual Observatory Tools for Astronomers

  https://www.youtube.com/watch?v=dDgpwtVrGbo

- Aladin SIMBAD pointer tutorial

  https://www.youtube.com/watch?v=0iIby3XXvp0

- IVOA Interoperability meeting: newcomers sessions

  https://ws.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/files/vault/pdowler/ivoa

  /virtual2021b/session-0-intro1.mp4

  https://ws.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/files/vault/pdowler/ivoa/

  virtual2021b/session-0-intro2.mp4