

The ADP: enabling access and exploitation of radio data collections through the IVOA

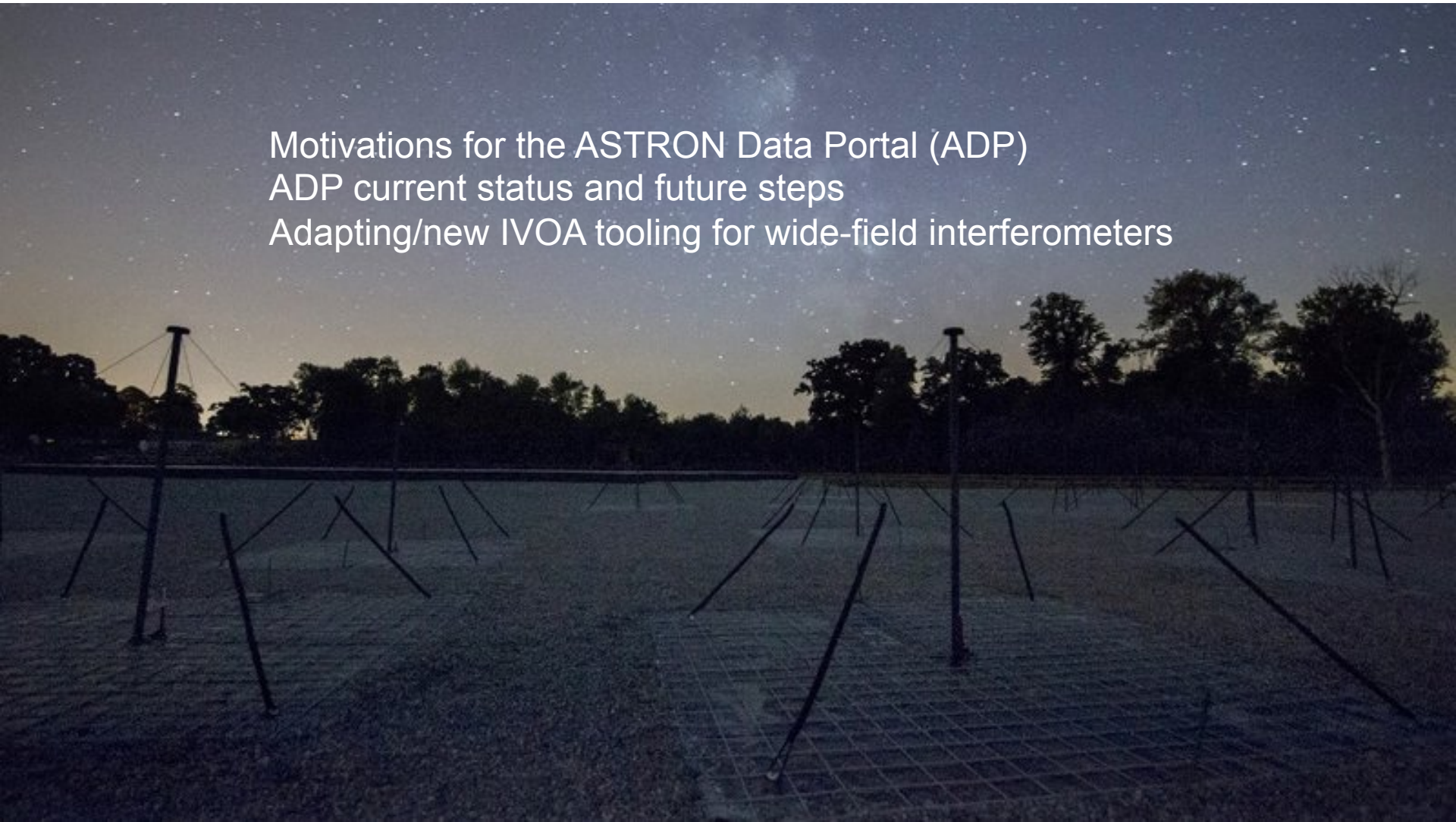
Marco Iacobelli
LOFAR telescope scientist

Asterics European Data Provider Forum and Training Event
Heidelberg, June 28th 2018

Outline



Motivations for the ASTRON Data Portal (ADP)
ADP current status and future steps
Adapting/new IVOA tooling for wide-field interferometers



ASTRON data collections



LOFAR

ASTRON

A large and heterogeneous collection of scientific data product from different radio interferometers:



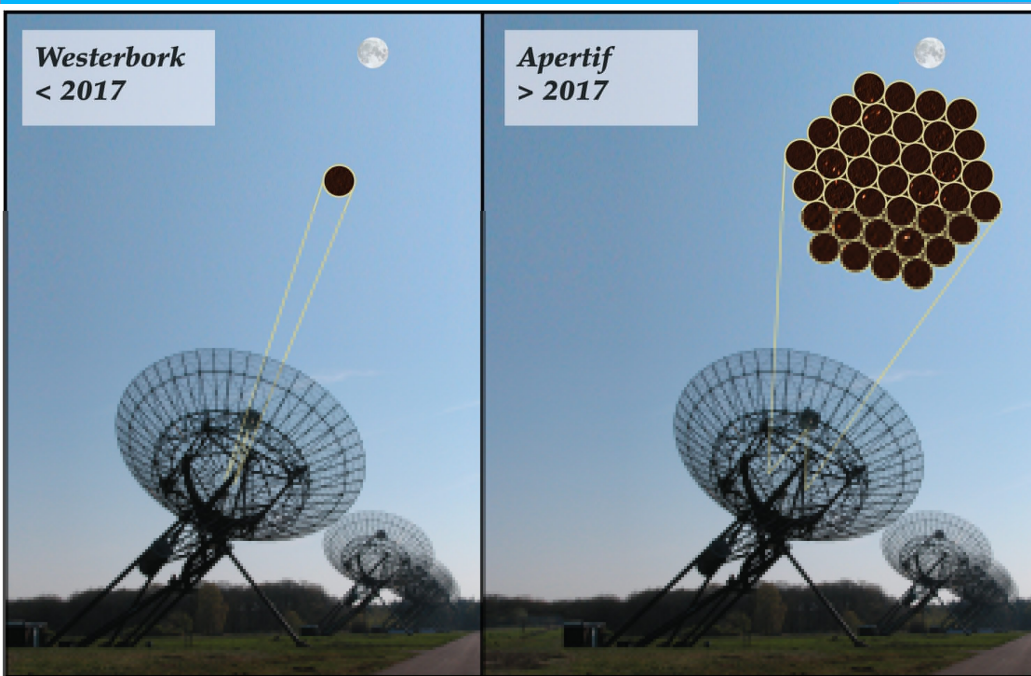
International LOFAR Telescope: 51 stations → 38 NL+13 Europe

Data transport rates over the array of order 150 Gbit/s:

- new generation wide-field interferometer (SKA pathfinder)
- data storage challenges (35 Tb/h of raw, correlated visibilities)

ASTRON data collections

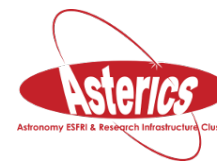
A large and heterogeneous collection of scientific data product from different radio interferometers:



WSRT Telescope / APERTIF:

- Telescope operational since 1970
- 2nd life as wide-field interferometer due to new generation of receiver array

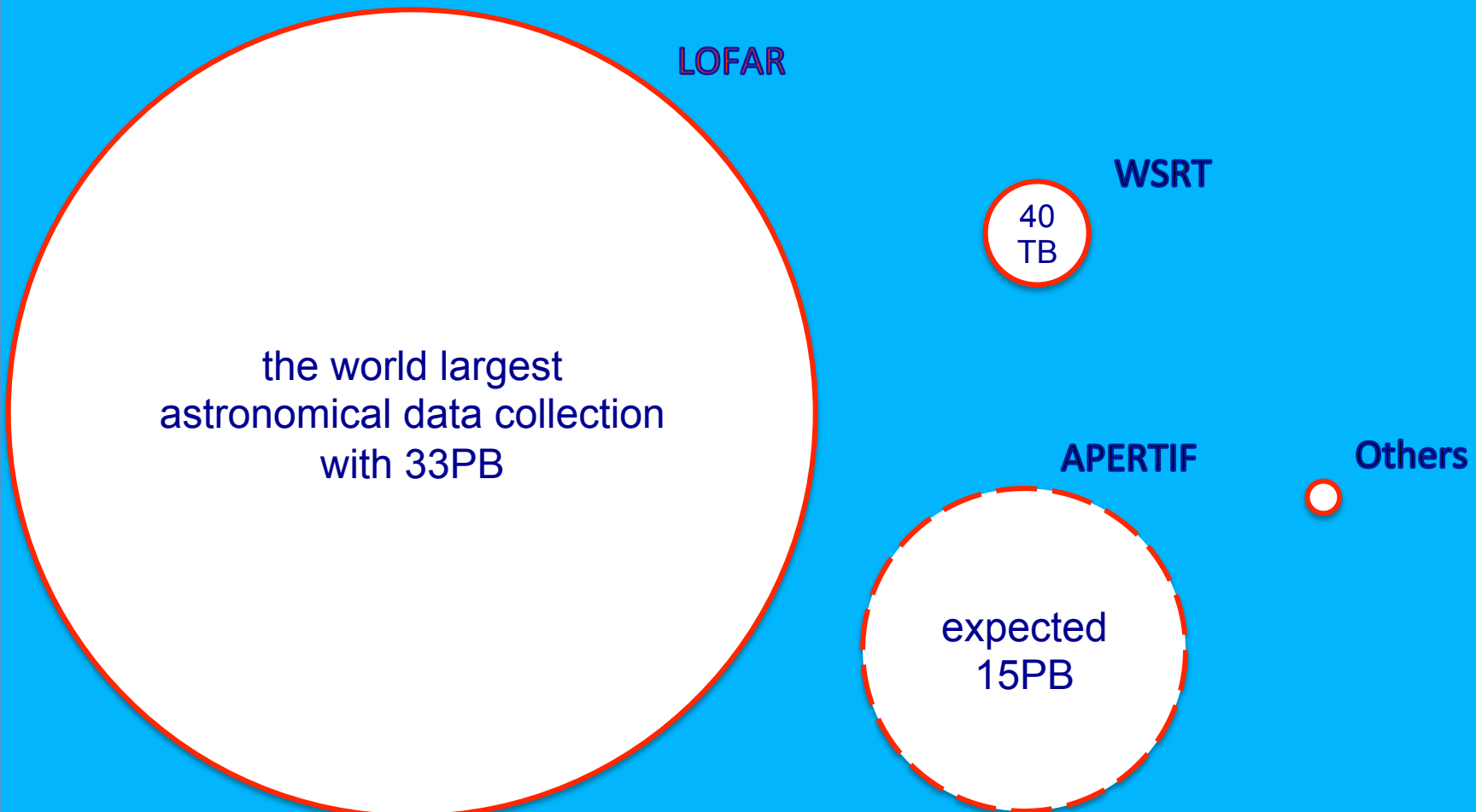
ASTRON data collections



LOFAR

ASTRON

A large and heterogeneous collection of scientific data product from different radio interferometers:



ASTRON data collections



LOFAR

ASTRON

A large and heterogeneous collection of scientific data product from different radio interferometers:

- Different data types (level 0 to 3)
- Different sizes 10^{-4} – 10^2 TB

Domain	Data Type	NASA/EOS Terminology	Description
Instrument	Device monitoring data	N/A (Level-1)	Produced by instrumentation, typically not preprocessed, and typically not stored as part of raw data products. Useful for preparing trending data to detect emerging instrument failures, and providing operational responses to other failures to dynamically improve the potential for producing quality data products.
Instrument	Raw data	Level 0	Produced by instrumentation. May be subject to limited preprocessing in firmware (for example, autocorrelation spectrometers).
Instrument	Calibrated data	Level 1	Produced by removing instrumental and environmental effects. EOS breaks this down into Levels 1A (raw data appended with annotations and calibration information) and Level 1B (raw data processed to calibrated data).
Science	Derived data	Level 2	Produced by combining calibrated data with other calibrated data, or with other derived data, according to processes, techniques, or algorithms. Scientific analysis can take place at this level or any higher level.
Science	Assimilated data	Level 3	Produced by gridding, resampling, and/or changing the frame of reference for derived data.
Science	Model data	Level 4	Produced by applying one or more mathematical, physical, or stochastic models to collections of assimilated and derived data products.

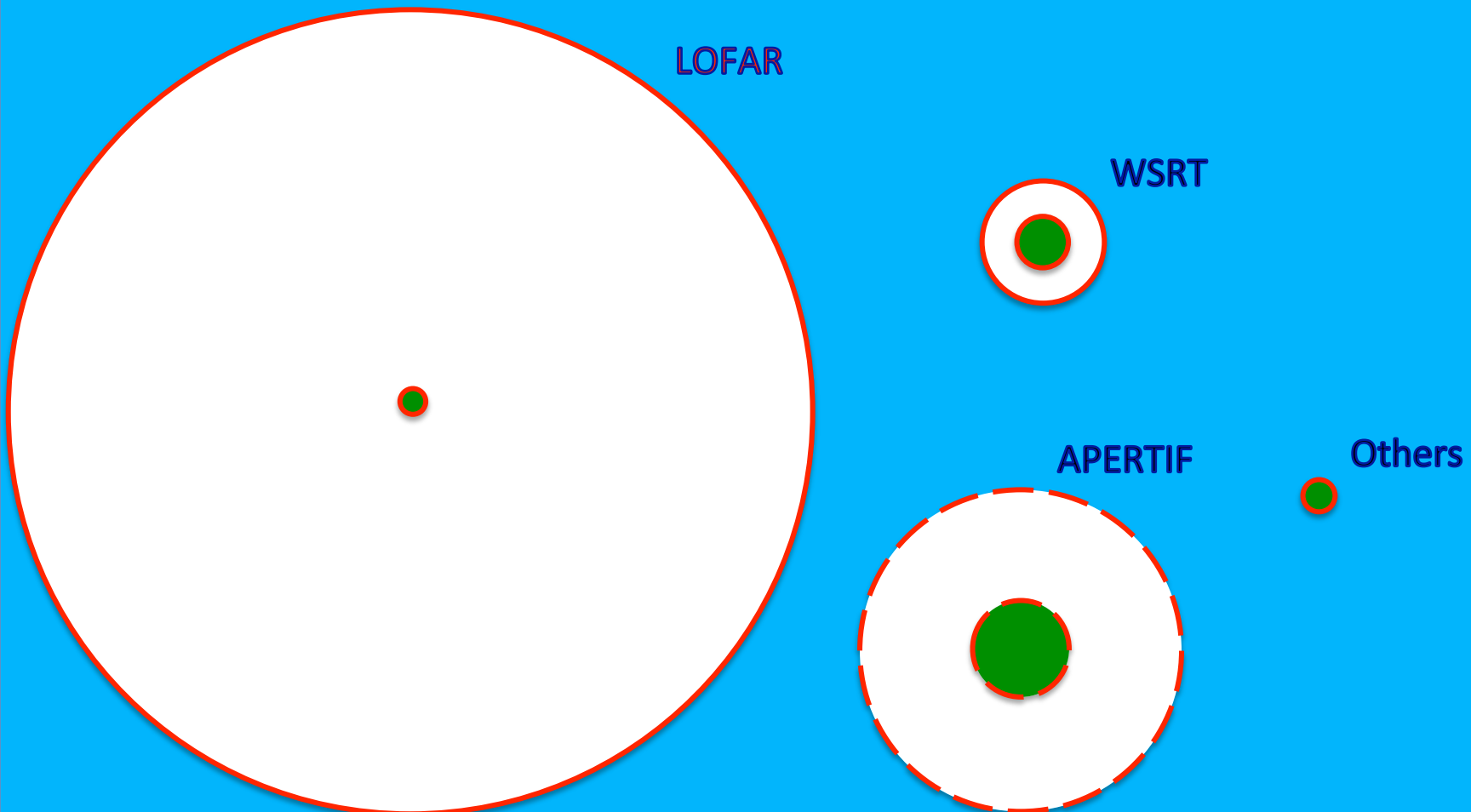
ASTRON data collections



LOFAR

ASTRON

A large and heterogeneous collection of scientific data product from different radio interferometers:



ASTRON data collections



LOFAR

ASTRON

A large and heterogeneous collection of scientific data product from different radio interferometers:

- Different data types (level 0 to 3)
- Different sizes 10^{-4} – 10^2 TB
- **Data sharing impacted by the network limitations**, especially for level 0 data products.
- **Processing requires large computing power**, which can not be available to a generic user.
- RO and KSPs are exploring the option to move the processing directly to the archiving sites. A successful pilot project is ongoing at SARA. Currently being replicated at Jülich.

Domain	Data Type	NASA/EOS Terminology	Description
Instrument	Device monitoring data	N/A (Level-1)	Produced by instrumentation, typically not preprocessed, and typically not stored as part of raw data products. Useful for preparing trending data to detect emerging instrument failures, and providing operational responses to other failures to dynamically improve the potential for producing quality data products.
Instrument	Raw data	Level 0	Produced by instrumentation. May be subject to limited preprocessing in firmware (for example, autocorrelation spectrometers).
Instrument	Calibrated data	Level 1	Produced by removing instrumental and environmental effects. EOS breaks this down into Levels 1A (raw data appended with annotations and calibration information) and Level 1B (raw data processed to calibrated data).
Science	Derived data	Level 2	Produced by combining calibrated data with other calibrated data, or with other derived data, according to processes, techniques, or algorithms. Scientific analysis can take place at this level or any higher level.
Science	Assimilated data	Level 3	Produced by gridding, resampling, and/or changing the frame of reference for derived data.
Science	Model data	Level 4	Produced by applying one or more mathematical, physical, or stochastic models to collections of assimilated and derived data products.

ASTRON data collections



A large and heterogeneous collection of scientific data product from different radio interferometers:

- LOFAR: ASTROWISE → need for improvements: chance for new algorithms
- WSRT: de-commissioned + WENSS
- APERTIF: ALTA → under development
- Others (MSSS, LBCS, LOTSS, TGSS) → VO@ASTRON

The ADP scopes



In view of an European Science and Data Centre, ASTRON has the ambition to set up a science data portal to serve a broad astronomical community to:

- support the exchange of data by scientists
- maximise scientific return from existing data collections.

Main goals:

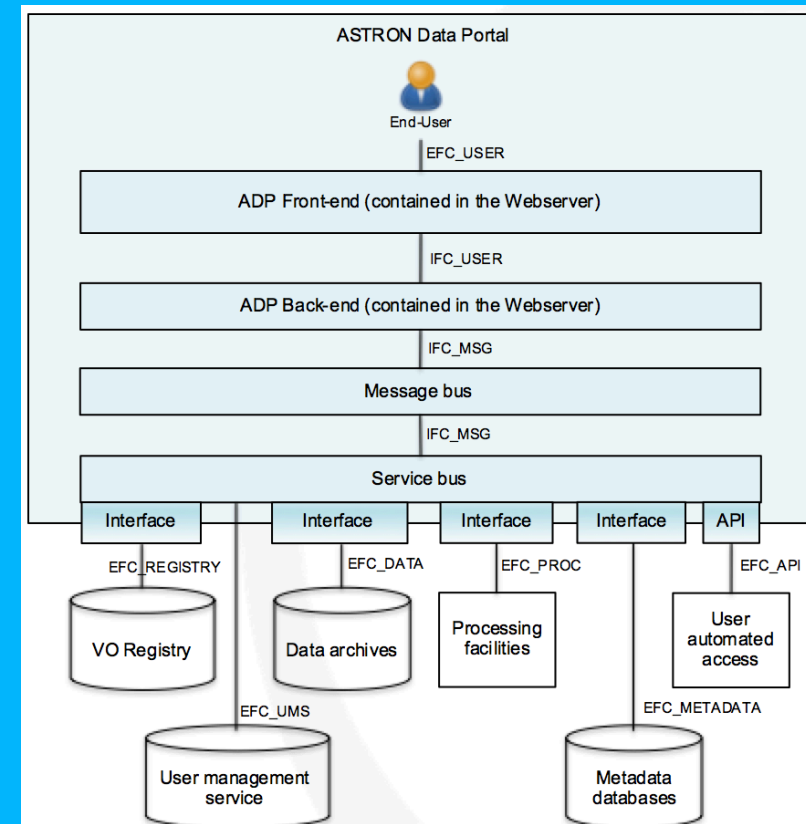
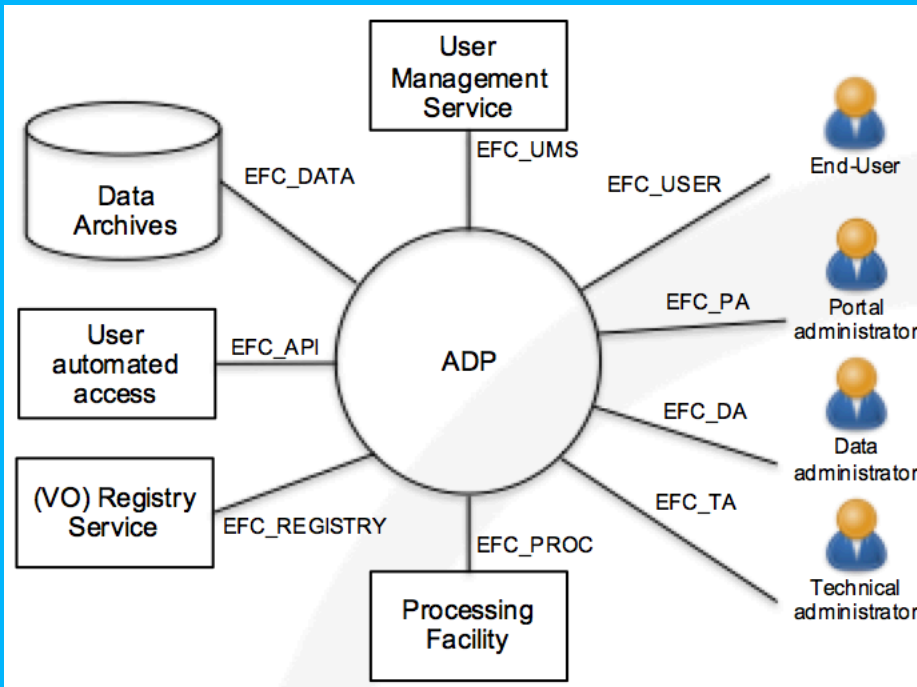
- exposing LOFAR, WSRT and APERTIF data collections, following principles of Findability, Accessibility, Interoperability, and Reusability
- added value services → pipelines, analytics and visualization

The portal can grow wider by including data collections of future instruments (MeerKat, NCLE, SKA).

The ADP overview

The ADP will consist of:

- a graphical user interface (frontend), publicly available on the internet.
- interfaces to the existing data archives that ASTRON uses (backend), to query the different data archives and to translate the results back to users.



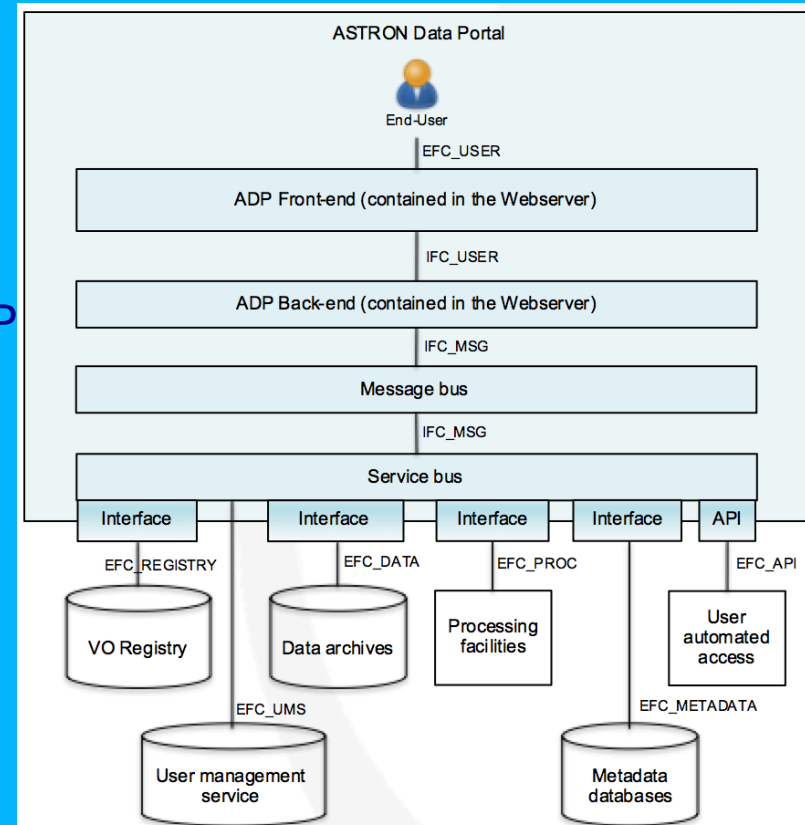
The ADP overview

The ADP will consist of:

- a graphical user interface (frontend), publicly available on the internet.
- interfaces to the existing data archives that ASTRON uses (backend), to query the different data archives and to translate the results back to users.

The ADP will adopt IVOA standards for, e.g.:

- visualization → Aladin / HiPS
- protocols → ADQL for queries / TAP / SAMP
- data model → ObsCore / Datalink



The ADP roadmap



Current strategy:

- long term goal: a platform for scientists to do science with ASTRON provided data collections, including services to process and analyse data
- short term goal: a simple but usable release of the ADP

Project status: concept system design / system requirements phase finalized

- the ADP will be developed iterative and scalable
- phase 1 → development & implementation in 1 year
 - search & retrieve interface
 - user management (& user area)

The ADP design



LOFAR

ASTRON

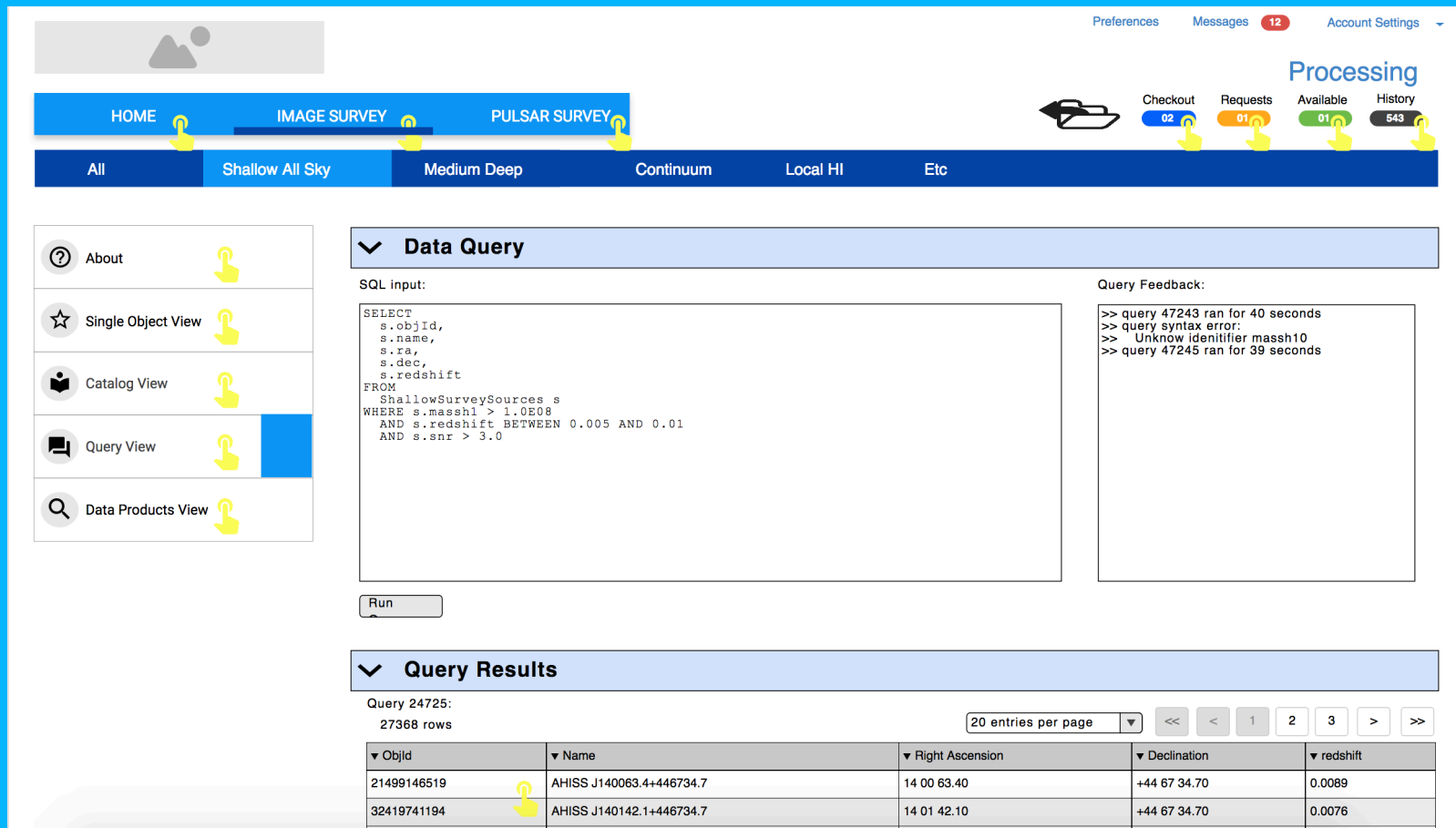
Design inspired by the ALTA (in progress) and ESA Sky 2.x portals

The screenshot displays the ADP web interface. At the top right, there are links for 'Preferences', 'Messages' (with a red badge showing '12'), and 'Account Settings'. Below this is a 'Processing' section with a folder icon and four status indicators: 'Checkout' (02), 'Requests' (01), 'Available' (01), and 'History' (543). The main navigation bar includes 'HOME', 'IMAGE SURVEY', and 'PULSAR SURVEY'. Below this is a secondary bar with 'All', 'Shallow All Sky', 'Medium Deep', 'Continuum', 'Local HI', and 'Etc'. On the left sidebar, there are menu items: 'About', 'Single Object View', 'Catalog View', 'Query View', and 'Data Products View'. The main content area features a 'Navigation' section with a star map and a 'Position on map' box containing the following data: r.a.: 11:44:01.27, dec.: +19:57:16.97, fov x: 12.71 arcmin, fov y: 7.36 arcmin. Below the map is a search bar with 'NGC' selected and a 'Go!' button. Further down are input fields for 'r.a.', 'dec.', and 'radius' (set to 100 arcsec), with 'Apply' and 'Reset' buttons. At the bottom, there is a 'Source List (summary)' section with a table showing 20 entries per page. The table has columns for 'Selected', 'Catalog Name', 'Resolved Name', 'Right Ascension', 'Declination', 'z', 'M HI [10^8 M_sun]', 'Flux', and 'SNR'. The first entry is: Selected: , Catalog Name: AHISS J140063.4+446734.7, Resolved Name: CGC 1245, Right Ascension: 14 00 63.40, Declination: +44 67 34.70, z: 0.0089, M HI: 8, Flux: 1.8, SNR: 3.6.

Selected	Catalog Name	Resolved Name	Right Ascension	Declination	z	M HI [10 ⁸ M _{sun}]	Flux	SNR
<input type="checkbox"/>	AHISS J140063.4+446734.7	CGC 1245	14 00 63.40	+44 67 34.70	0.0089	8	1.8	3.6

The ADP design

Design inspired by the ALTA (in progress) and ESA Sky 2.x portals



The screenshot displays the ADP web interface. At the top right, there are links for 'Preferences', 'Messages' (with a red badge showing '12'), and 'Account Settings'. Below this is a 'Processing' status bar with 'Checkout' (02), 'Requests' (01), 'Available' (01), and 'History' (543). The main navigation bar includes 'HOME', 'IMAGE SURVEY', and 'PULSAR SURVEY'. Below this is a secondary navigation bar with 'All', 'Shallow All Sky', 'Medium Deep', 'Continuum', 'Local HI', and 'Etc'. On the left, a sidebar contains 'About', 'Single Object View', 'Catalog View', 'Query View', and 'Data Products View'. The central area features a 'Data Query' section with an SQL input field containing a query for shallow survey sources, a 'Run' button, and a 'Query Feedback' section showing execution logs. Below the query is a 'Query Results' section for 'Query 24725' showing 27368 rows and a table of results.

Preferences Messages 12 Account Settings

Processing

Checkout 02 Requests 01 Available 01 History 543

HOME IMAGE SURVEY PULSAR SURVEY

All Shallow All Sky Medium Deep Continuum Local HI Etc

About Single Object View Catalog View Query View Data Products View

Data Query

SQL input:

```
SELECT
  s.objId,
  s.name,
  s.ra,
  s.dec,
  s.redshift
FROM
  ShallowSurveySources s
WHERE s.masshl > 1.0E08
      AND s.redshift BETWEEN 0.005 AND 0.01
      AND s.snr > 3.0
```

Run

Query Feedback:

```
>> query 47243 ran for 40 seconds
>> query syntax error:
>>   Unknow identifier massh10
>> query 47245 ran for 39 seconds
```

Query Results

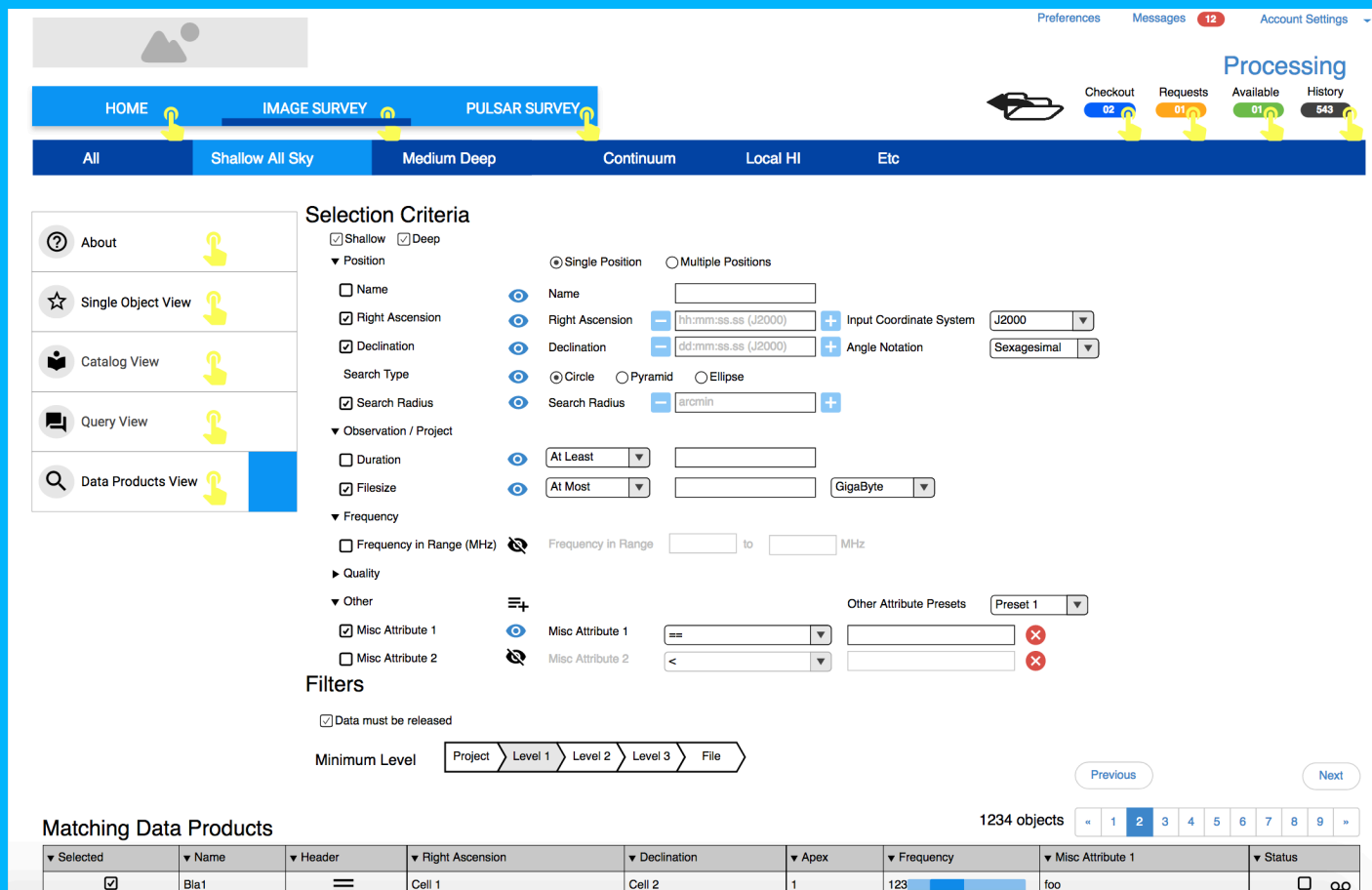
Query 24725:
27368 rows

20 entries per page << < 1 2 3 > >>

ObjId	Name	Right Ascension	Declination	redshift
21499146519	AHISS J140063.4+446734.7	14 00 63.40	+44 67 34.70	0.0089
32419741194	AHISS J140142.1+446734.7	14 01 42.10	+44 67 34.70	0.0076

The ADP design

Design inspired by the ALTA (in progress) and ESA Sky 2.x portals



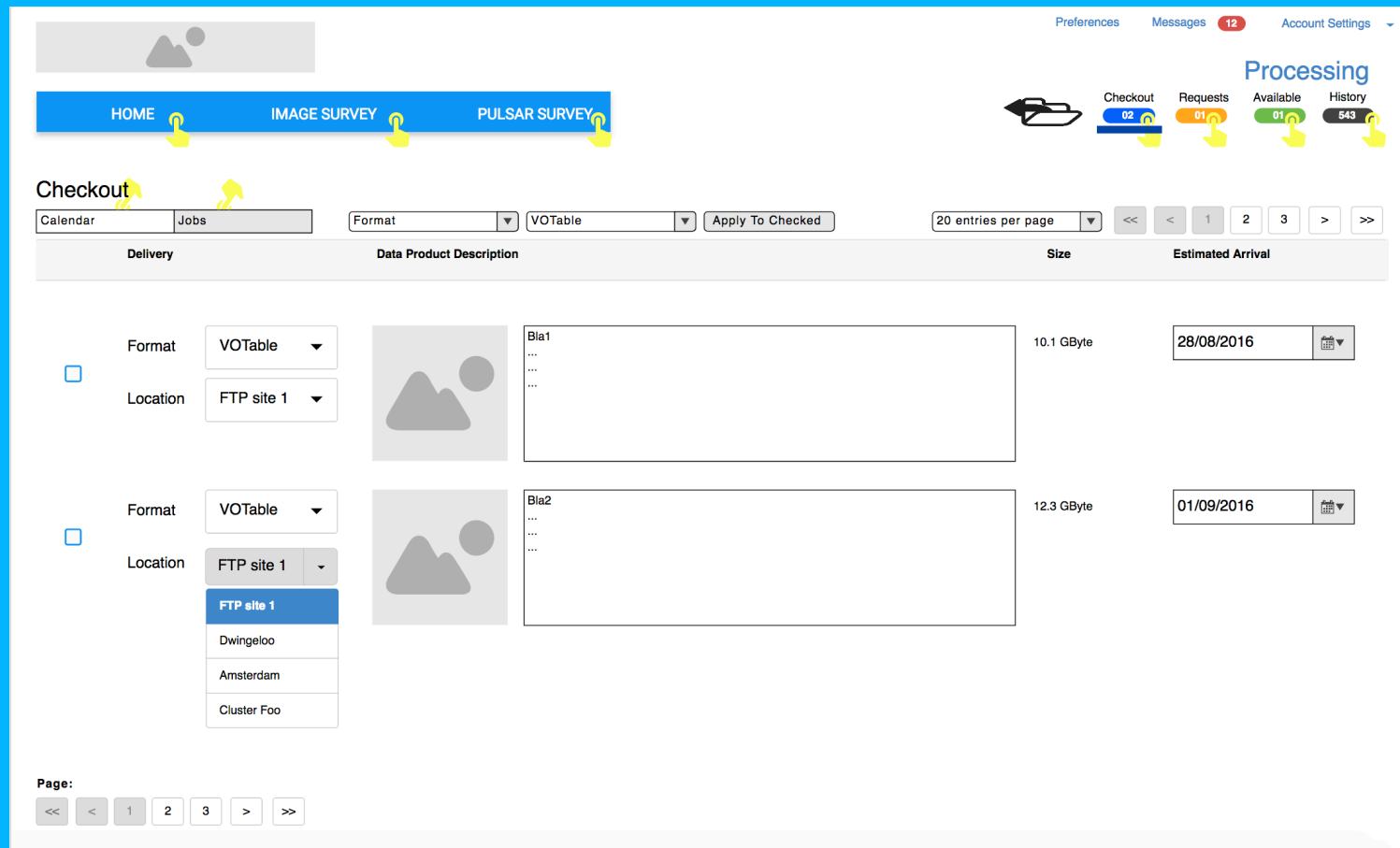
The screenshot displays the ADP interface with the following components:

- Navigation:** A top bar with 'HOME', 'IMAGE SURVEY', and 'PULSAR SURVEY' tabs. Below it, a dark blue bar contains 'All', 'Shallow All Sky', 'Medium Deep', 'Continuum', 'Local HI', and 'Etc' options.
- User Profile:** Top right corner shows 'Preferences', 'Messages' (12), and 'Account Settings'.
- Processing Status:** A 'Processing' section with a 'Checkout' (02), 'Requests' (01), 'Available' (01), and 'History' (543) indicator.
- Left Sidebar:** A menu with 'About', 'Single Object View', 'Catalog View', 'Query View', and 'Data Products View' (highlighted).
- Selection Criteria:** A central panel with sections for:
 - Position:** Radio buttons for 'Single Position' and 'Multiple Positions'. Fields for 'Name', 'Right Ascension' (hh:mm:ss.ss (J2000)), 'Declination' (dd:mm:ss.ss (J2000)), 'Input Coordinate System' (J2000), and 'Angle Notation' (Sexagesimal).
 - Search Type:** Radio buttons for 'Circle', 'Pyramid', and 'Ellipse'. A 'Search Radius' field with 'arcmin'.
 - Observation / Project:** 'Duration' (At Least, At Most) and 'Filesize' (GigaByte).
 - Frequency:** 'Frequency in Range (MHz)' and 'Frequency in Range' (to MHz).
 - Quality:** 'Other' section with 'Misc Attribute 1' and 'Misc Attribute 2'.
- Filters:** A checkbox for 'Data must be released' and a 'Minimum Level' selector (Project, Level 1, Level 2, Level 3, File).
- Matching Data Products:** A table at the bottom showing 1234 objects. The first object is selected.

Selected	Name	Header	Right Ascension	Declination	Apex	Frequency	Misc Attribute 1	Status
<input checked="" type="checkbox"/>	Bla1	☰	Cell 1	Cell 2	1	123	foo	<input type="checkbox"/>

The ADP design

Design inspired by the ALTA (in progress) and ESA Sky 2.x portals



Preferences Messages **12** Account Settings

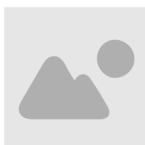

HOME IMAGE SURVEY PULSAR SURVEY

Processing

Checkout 02 Requests 01 Available 01 History 543

Checkout

Calendar Jobs Format VOTable Apply To Checked 20 entries per page

Delivery	Data Product Description	Size	Estimated Arrival
<input type="checkbox"/> Format: VOTable Location: FTP site 1	 Bla1	10.1 GByte	28/08/2016
<input type="checkbox"/> Format: VOTable Location: FTP site 1 Dwingeloo Amsterdam Cluster Foo	 Bla2	12.3 GByte	01/09/2016

Page: << < 1 2 3 > >>

The ADP challenges



Users management & private area (authentication)

- authentication vs federation → federated identity management

Data modeling

- see A. Renting talk

(Meta-)Data quality

- which standards and visualization / flagging tools ?

Processing (including ingest) & analysis area

- allocation and monitoring of resources crucial for operations

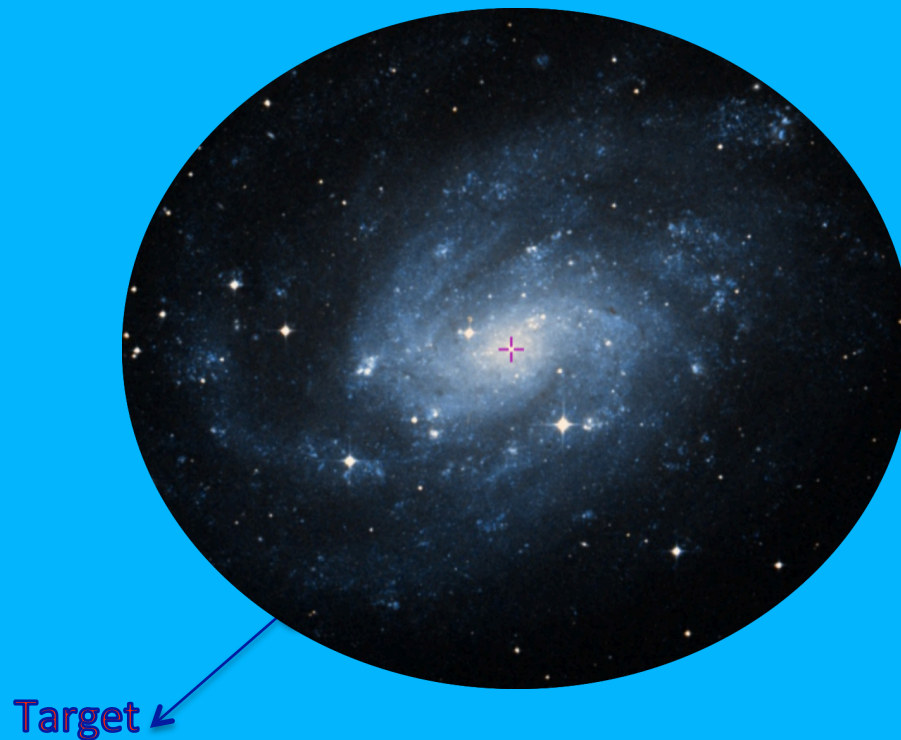
The ADP & new tooling



Characterization tools needed when dealing with level 0/1 data products from wide-field interferometers → accounting for beam-shapes of different modes for searches and selection but also for processing / analysis

The ADP & new tooling

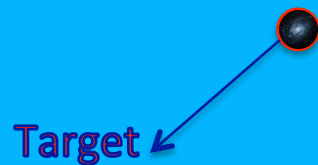
Characterization tools needed when dealing with level 0/1 data products from wide-field interferometers → accounting for beam-shapes of different modes for searches and selection but also for processing / analysis



The ADP & new tooling



Characterization tools needed when dealing with level 0/1 data products from wide-field interferometers → accounting for beam-shapes of different modes for searches and selection but also for processing / analysis



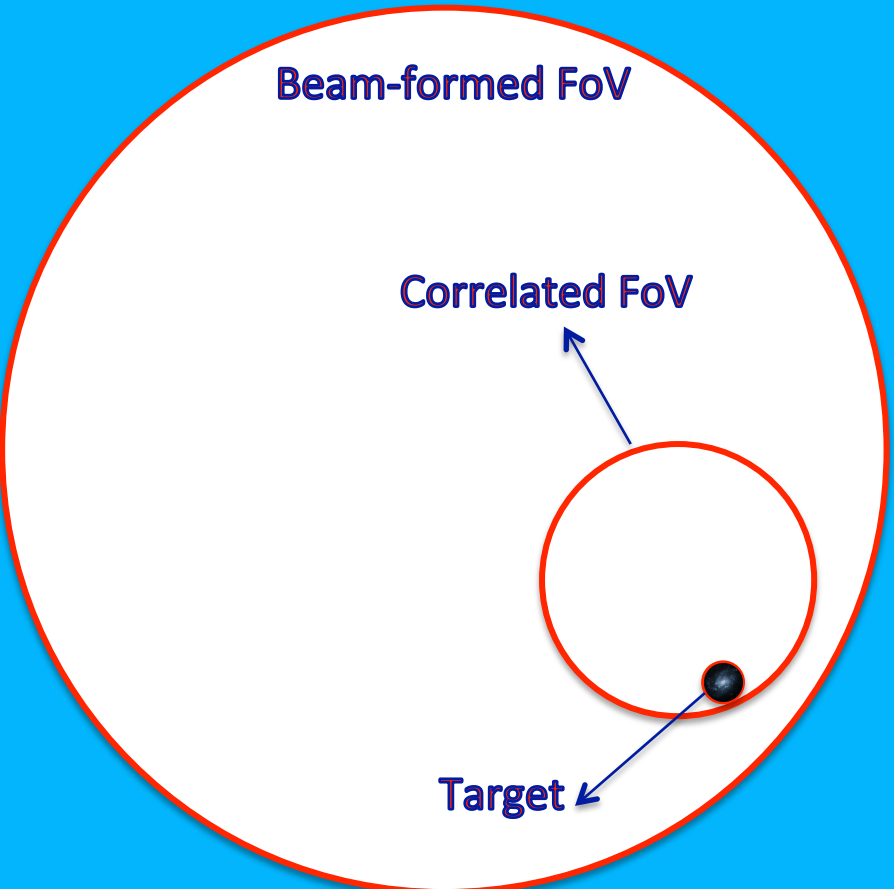
The ADP & new tooling

Characterization tools needed when dealing with level 0/1 data products from wide-field interferometers → accounting for beam-shapes of different modes for searches and selection but also for processing / analysis

Beam-formed FoV

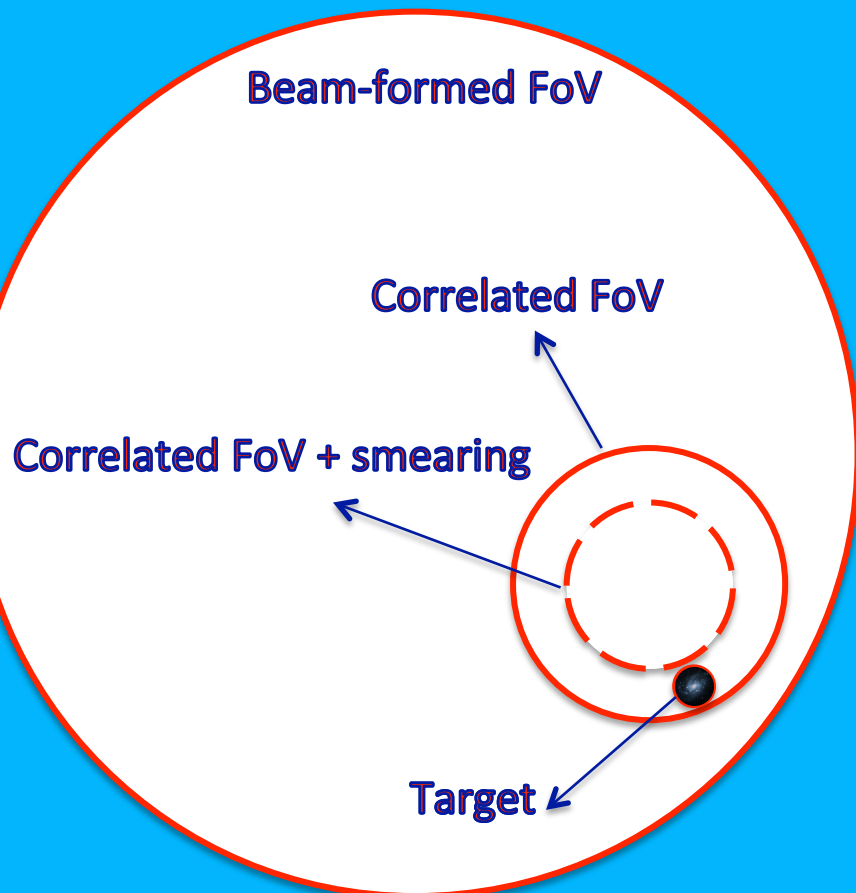
Correlated FoV

Target



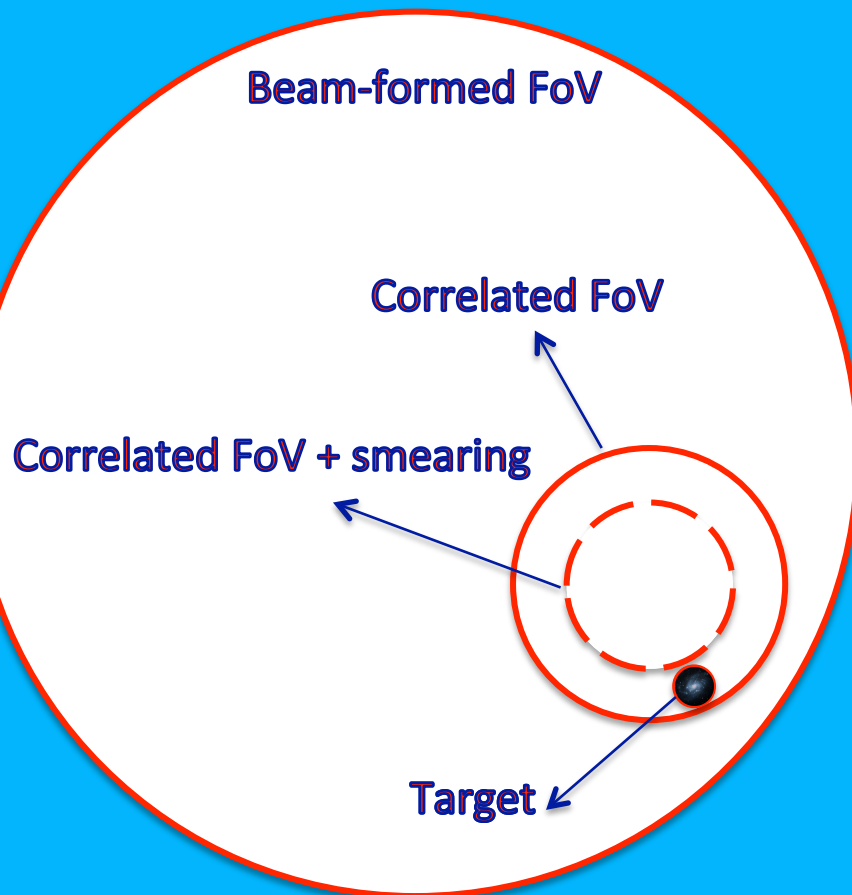
The ADP & new tooling

Characterization tools needed when dealing with level 0/1 data products from wide-field interferometers → accounting for beam-shapes of different modes for searches and selection but also for processing / analysis



The ADP & new tooling

Characterization tools needed when dealing with level 0/1 data products from wide-field interferometers → accounting for beam-shapes of different modes for searches and selection but also for processing / analysis



Observed / processed data characterization in terms of:

- uv-coverage
- beam pattern (FoV & sensitivity)

Accurate reconstruction:

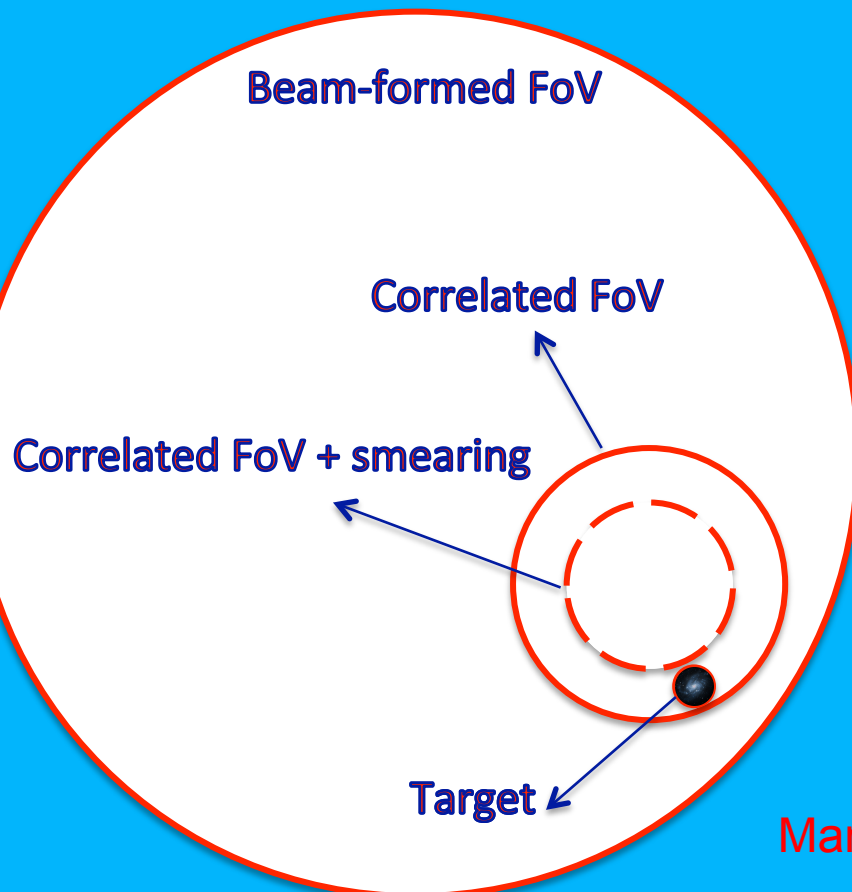
- computationally expensive, need data
→ not practical

Approximated reconstruction:

- only metadata needed → enough for data selection

The ADP & new tooling

Characterization tools needed when dealing with level 0/1 data products from wide-field interferometers → accounting for beam-shapes of different modes for searches and selection but also for processing / analysis



Observed / processed data characterization in terms of:

- uv-coverage
- beam pattern (FoV & sensitivity)

Accurate reconstruction:

- computationally expensive, need data
→ not practical

Approximated reconstruction:

- only metadata needed → enough for data selection

ObsCore modeling . . .
Managing VO standards regions for visualization . . .

Summary



Motivations for the ASTRON Data Portal (ADP)

- Exposing LOFAR, WSRT and APERTIF data collections to a broad astronomical community, following principles of Findability, Accessibility, Interoperability, and Reusability

ADP current status and future steps

- **Concept design phase final stage of review**
- **First development phase (search and retrieve)**
- **Scalable portal**

Adapting/new IVOA tooling for wide-field interferometers

- **Characterization needed for user to proper search / select data and derive suited high level products**