DAL future

current trends and issues

F. Bonnarel (CDS/CNRS)
M. Molinaro (INAF-OATs)

DAL future … in the past

Several talks or sessions already

… but too early

- Banff: recalling the future plans in AccessData session
- Sesto: WCS
- Sydney: miscellaneous

DAL landscape & data avalanche

- DAL landscape is complex
  - 14 different protocols in use or close to recommendation
    - including major version changes
- Data volume increase is going faster than our recommendation process
  - Resources / demands balance?
- Two needs
  - Understand the logic of the DAL landscape
  - Prepare evolution
  - -> ADASS XXVI poster (Molinaro and Bonnarel)

DAL protocol properties

- Data type of relevance
  - Catalogues/tables
  - Images/cubes
  - Spectra/time series
  - Theoretical data
  - Spectral lines
  - Raw or «low level» data (event lists, visibility)
- User-oriented functionalities
  - Discovery
  - Description
  - Basic access
  - Extended access
  - Link
- Interface and software design
  - Sync/async
  - dali compliance
  - Adql
  - parameter language
  - ...

DAL protocol prototypes

	Design	Functionalities					
Data type		ConeSearch-Discovery	Multi-dimensional Discovery	Description (NB: SIA1 is very different from others)	Simple-Access	Access-processing	Link
Catalogues/ tables	Sync	TAP,CS,ObsTAP	TAP,ObsTAP		TAP,CS		DataLink
	Async	TAP,ObsTAP	TAP,ObsTAP		TAP		
	ADQL	TAP,OBsTAP	TAP,ObsTAP		TAP		
	PBL	CS			CS		DataLink
	DALI	TAP,ObsTAP	TAP, ObsTAP		TAP		DataLink
	No-DALI	CS			CS		
Spectra / timeseries	Sync	SSA,ObsTAP	SSA,ObsTAP	SSA,ObsTAP	SSA	SSA	DataLink
	Async	ObsTAP	ObsTAP	ObsTAP			
	ADQL	ObsTAP	ObsTAP	ObsTAP			
	PBL	SSA	SSA	SSA	SSA	SSA	DataLink
	DALI	ObsTAP	ObsTAP	ObsTAP			
	No-DALI	SSA	SSA	SSA	SSA	SSA	DataLink
Images/ cubes	Sync	SIA1,SIA2,ObsTAP	SIA2,ObsTAP	SIA1,SIA2,ObsTAP	SIA1,SODA1.0	SIA1,SODA1.1	DataLink
	Async	SIA2,ObsTAP	SIA2,ObsTAP	SIA2,ObsTAP	SODA1.0	SODA1.1	
	ADQL	ObsTAP	ObsTAP	ObsTAP			
	PBL	SIA1,SIA2	SIA2	SIA1,SIA2	SIA1,SODA1.0	SIA1,SODA1.1	DataLink
	DALI	SIA2,ObsTAP	SIA2,ObsTAP	SIA2,ObsTAP	SODA1.0	SODA1.1	DataLink
	No-DALI	SIA1		SIA1	SIA1	SIA1	

	Design	Functionalities					
Data type		ConeSearch-Discovery	Multi-dimensional Discovery	Description (NB: SIA1 is very different from others)	Simple-Access	Access-processing	Link
Raw data/ Event list/ Visibility	sync	ObsTAP	ObsTAP	ObsTAP			DataLink
	async	ObsTAP	ObsTAP	ObsTAP			
	ADQL	ObsTAP	ObsTAP	ObsTAP			
	PBL						DataLink
	DALI	ObsTAP	ObsTAP	ObsTAP			
	No-DALI						
Spectral Lines	sync	SLA	SLA	SLA			
	async						
	ADQL						
	PBL	SLA	SLA	SLA			
	DALI						
	No-DALI	SLA	SLA	SLA			
Theory data	sync		SimDAL	SimDAL	SimDAL	SimDAL	SimDAL
	async						
	ADQL						
	PBL						
	DALI						
	No-DALI		SimDAL	SimDAL	SimDAL	SimDAL	SimDAL

Where are we?

- We have
  - a subset based on DALI with sync/async ADQL and PQL
  - This makes a multi-D data recent backbone
  - and we have peripheric protocols
- We had planned evolutions for the future
  - SIA-2.1
    - extended metadata, discovery of virtual data
  - SODA-1.1
    - extended access (regridding, axis reduction)
  - TAP-2.0 & ADQL-3.0
    - for features requiring further work or are backward incompatible
    - but also TAP-1.2 & ADQL-2.2 have potential content

Evolution trends

- Data type
  - Time domain:
    - Use Cube data model ? Which protocol
    - -LSST use case: filtering relevant events. Evolution of VoEvents
- Functionalities:
  - Extended metadata
    - two features which are in DataSet Metadata/Cube DM and not in ObsCore:
      - Composed datasets
      - World coordinates mapping
        - Mapping this in new tables extending ObsCore schema?
        - VO-DML mapping ?
        - Standard XML ?
  - Extended server side data processing for better access
    - regridding, axis reduction, denoising, deconvolution, …

Evolution trends

- Interface design :
  - Define standard extended operations
    - regridding, axis reduction, pixel cutout
  - Define standard access to data provider custom services
    - The services are customized to the data but the query mode should be standardized
  - Define how to push code to the data
    - GWS WG working on standardizing this
      - via VOSpace
      - Interface to solutions such as Docker and Ipython?
      - Using SODA? Or dedicated protocol?

Merging functionalities: HiPS and HiPS-like solutions

- HiPS is altogether a discovery/access/visualisation functionality
  - Well adapted to progressive access to data of interest.
  - Visualisation is multi-D
  - Discovery/access is fast an easy but only spatial
  - HiPS remain re-processed data
    - when is that insufficient for doing science?
- Evolutions
  - Integrate HiPS mode as part of extended access data technology for large datasets
  - Use HiPS as discovery mode for original data using DataLink embedded in HEALPix cells
  - Fine tune access on other axes using other multi-scale technology?

Table Access

- TAP & ADQL are relational
  - Add NO-SQL solution?
    - OO database scenario
      - would it help model mapping?
  - Relax ADQL support?
    - What query language could jump in?

Formats and languages

- Json integration
  - how and to which extent?
  - Querying by Json files POST
  - Other formats? (YAML, ...)
- ADQL replacement for TAP-2.0?
- 3 factor semantics PQL
  - rules to define new custom or standard parameters
- PDL driven PQL
  - is it feasible?
- Base Language on datamodel
  - forcing the generation of virtual datasets by using virtual metadata

How to proceed ?

- All this stuff coming out of use cases gathered in the past.
- Adaptation of 2013 Multi-d roadmap to current realities
- New protocols Implementation feedback (only starting)

- Create DAL WG pages on these topics
  - Use cases for new functionalities
  - Implementation feedback
  - Proposed solutions

Potential target protocol versions

- Extended metadata, virtual data discovery, DM-language
  - SIA-2.1
- Custom services interface, 3 factor semantics
  - DataLink-1.1
- Standard extended data access functionalities
  - SODA-1.1
- Code to Data
  - SODA-1.1
  - SODA-2.0 + GWS protocols
- HiPS access
  - SODA1.1
- DataLink In HiPS : DataLink 1.1
- Non relational DB / ADQL relax
  - TAP-2.0