

Journée thématique OV France

Publication dans l'OV et protocole TAP

« Table Access Protocol »

André Schaaff, CDS

17 Septembre 2014



Observatoire de Paris, 17 septembre 2014





Introduction

- Le but de la présentation est de donner un aperçu du standard IVOA TAP
- La partie « plus pratique » sera abordée dans les présentations suivantes avec des exemples, des retours d'expérience, des outils, etc.
- Pourquoi s'orienter vers TAP ?
 - Pour offrir une visibilité et un accès (VO) standardisé à ses données
 - On peut également choisir de rendre ses données accessibles via TAP (en mettant en ligne son propre service ou en faisant ingérer ses données par un autre service) afin de profiter des outils de traitements qui acceptent le format en entrée



Documents IVOA

- La recommandation IVOA du 27 mars 2010 (WG Data Access Layer)
 - <http://www.ivoa.net/documents/TAP/>
- Autres
 - Une recommandation « TAPRegExt » du 27 août 2012
 - <http://ivoa.net/documents/TAPRegExt/20120827/index.html>
 - Une note « TAP implementation Notes » du 13 décembre 2013
 - <http://ivoa.net/documents/Notes/TAPNotes/index.html>
 - Une recommandation « Registry Relational Schema » du 27 juin 2014
 - <http://ivoa.net/documents/RegTAP/20140627>



Avant d'aller plus loin, quelques acronymes IVOA...

- TAP utilise d'autres standards IVOA
 - ADQL (2.0): Astronomical Data Query Language, cf. suite
 - UWS (1.0): « Universal Working Service », utilisé dans les implémentations de TAP pour la gestion des travaux asynchrones
 - VOSI (1.0): « the Virtual Observatory Support Interface » définit un ensemble d'interfaces de base communes à tous les services VO
 - availability ivo://ivoa.net/std/VOSI#availability
 - capabilities ivo://ivoa.net/std/VOSI#capabilities
 - tables ivo://ivoa.net/std/VOSI#tables
 - VOTable (1.2+): VO Table, un formalisme XML pour représenter des tables (métadonnées et données)



Remarque concernant le document de recommandation TAP

- Il comporte beaucoup de paragraphes « informatifs » en plus de ceux qui sont « normatifs »
- La version actuelle est « 1.0 » donc pas encore de révision
 - Une version « 1.* » doit rester compatible pour ne pas chambouler les services et les outils, les modifications/évolutions sont donc limitées, on parle plutôt de versions de « maintenance »
 - Une version « 2.0 » introduirait sans doute des changements intéressants mais nécessiterait des efforts côté développements/implémentations
- Pas vraiment de travaux en cours ce qui permet de stabiliser le standard et les développements liés



TAP, c'est quoi ?

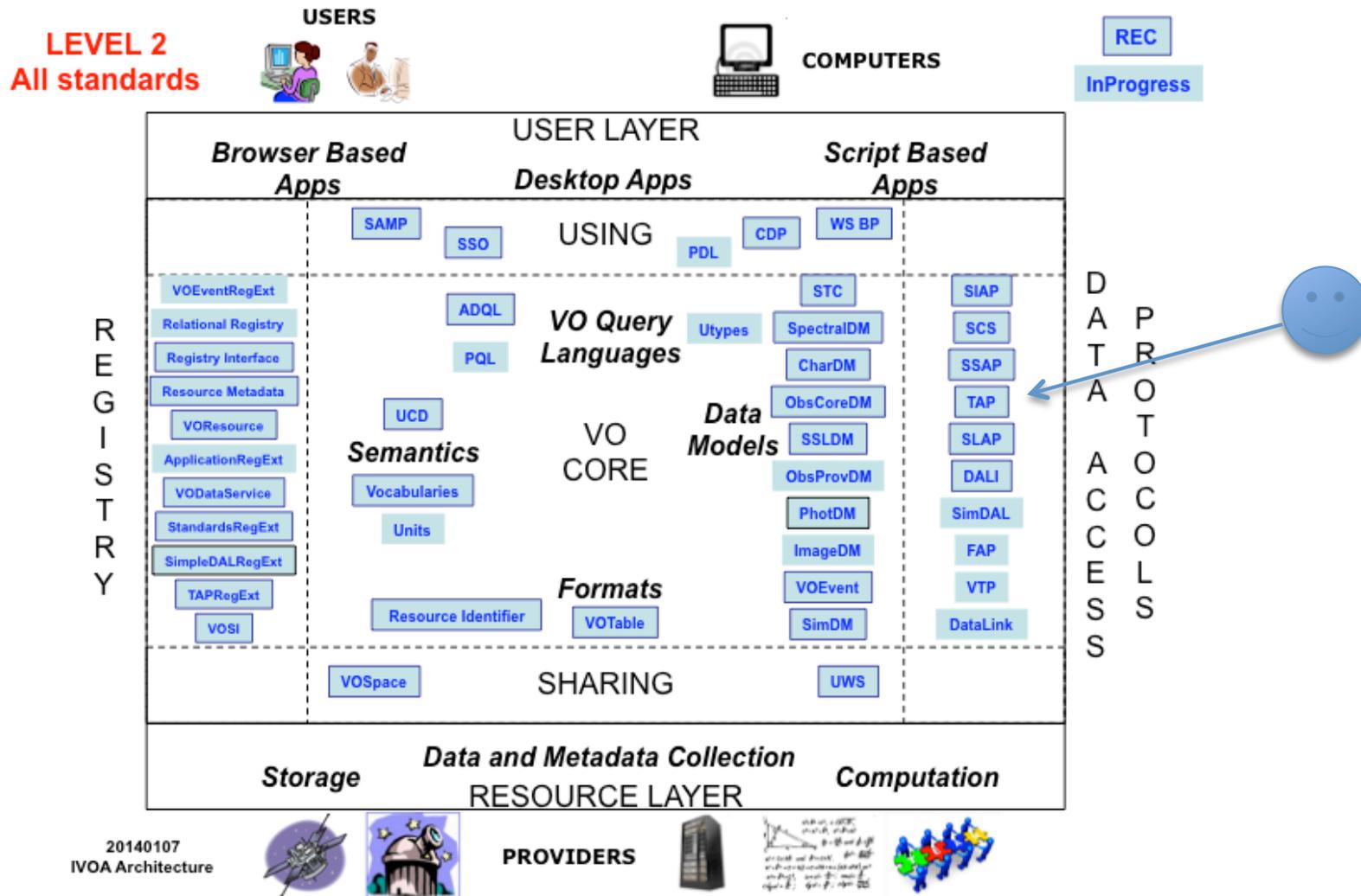
- Il existe des protocoles IVOA pour les images (SIAP), les spectres (SSAP), etc. donc spécialisés
- TAP est un protocole de service Web plus générique qui donne accès à des collections de données tabulaires issues de bases de données
- Basé sur REST (Representational State Transfer)



TAP, c'est quoi ? (2)

- Un service TAP accepte des requêtes concernant les données disponibles et, en réponse, retourne une table
 - En proposant différents langages pour soumettre une requête: **ADQL** à minima, SQL, etc.)
 - En proposant différents formats pour la table retournée: **VOTable** à minima, JSON, etc.
- Remarque:
 - Un service TAP ne proposera pas forcément toutes les données réellement présentes dans la base de données
 - Un utilisateur pourra charger temporairement ses propres données, ce qui lui permettra par exemple de les croiser avec les données du service

TAP dans l'architecture IVOA





Un service TAP accepte donc des requêtes...

- ... de 3 types
 - « Data Queries »
 - Accès au « contenu astronomique » via l'un des langages de requête supporté par le service
 - « Metadata Queries »
 - Mêmes langages de requête que précédemment mais pour des tables particulières (cf. plus loin TAP_SCHEMA)
 - VOSI, VO Support Interface
 - Pour récupérer par exemple des métadonnées concernant la disponibilité du service TAP



Un langage de requête en particulier...

- Le seul qui soit obligatoire: ADQL
 - Permet d'effectuer des requêtes sur une ou plusieurs tables (y compris de métadonnées) du service TAP
 - En général, on effectue une requête pour accéder aux métadonnées pour connaître les noms des tables et des colonnes puis on peut effectuer des requêtes sur les données
 - Les requêtes sont dépendantes du service sauf si l'on reste très générique



ADQL: un peu plus de détail

- La plupart des caractéristiques de SQL
- Support de fonctions de recherche spatiale
 - Utile pour le XMatch positionnel de tables
- Prise en compte de tables temporaires chargées par l'utilisateur
- Pas de support des procédures stockées
- Il n'autorise pas la modification de la base de données...



Exemple de requête ADQL dans TAP VizieR

- Un exemple simple très « SQL like »...

SELECT *

FROM tables

WHERE description **LIKE** '%white dwarf%'

```
schema_name,description,table_name,table_type,utype  
"viz1","Complete M dwarf-white dwarf wide binary target list","J/AJ/129/2428/table1","table", ""  
"viz2","All white dwarfs detected in the ROSAT PSPC Survey","J/A+A/316/147/table1","table", ""  
"viz4","Kinematic parameters of the white dwarfs","J/A+A/400/877/table2","table", ""  
"viz4","Photometric data for known white dwarfs in NGC 2516","J/AJ/123/290/table4","table", " »  
....
```



Exemple de requête ADQL dans TAP VizieR (2)

- Un exemple plus étoffé

```
--xmatch_wise_hipparcos
```

```
-- xmatch between wise (WISE) and I/239/hip_main (hipparcos) around the lmc center within 1 deg
```

```
SELECT TOP 100 "II/311/wise".RAJ2000, "II/311/wise".DEJ2000, "I/239/hip_main".HIP, "I/239/  
hip_main"."RA(ICRS)", "I/239/hip_main"."DE(ICRS)", "II/311/wise".WISE
```

```
FROM "II/311/wise", "I/239/hip_main"
```

```
WHERE 1=CONTAINS(POINT('ICRS', "I/239/hip_main"."RA(ICRS)", "I/239/hip_main"."DE(ICRS)"), CIRCLE('ICRS',  
80.89417, -69.75611, 1.))
```

```
AND 1=CONTAINS(POINT('ICRS', "I/239/hip_main"."RA(ICRS)", "I/239/hip_main"."DE(ICRS)"), CIRCLE('ICRS', "II/  
311/wise".RAJ2000, "II/311/wise".DEJ2000, 10/3600.))
```

```
RAJ2000,DEJ2000,HIP,RA(ICRS),DE(ICRS),WISE
```

```
81.660548,-70.063505,25452,81.66074576,-70.06368899,"J052638.53-700348.6"
```

```
82.708637,-69.524824,25822,82.70866058,-69.52483500,"J053050.07-693129.3"
```

```
81.409941,-69.595237,25369,81.41017980,-69.59551844,"J052538.38-693542.8"
```

```
79.443100,-69.849222,24694,79.44322532,-69.84923159,"J051746.34-695057.1"
```

```
78.378178,-69.539904,24347,78.37823715,-69.53990311,"J051330.76-693223.6"
```

```
81.645502,-68.868693,25448,81.64533902,-68.86870606,"J052634.92-685207.2"
```



En HTTP

En synchrone

```
HTTP POST http://example.com/tap/sync
REQUEST=doQuery
LANG=ADQL
QUERY=SELECT * FROM magnitudes as m where m.r>=10 and m.r<=16
```

En asynchrone

```
HTTP POST http://example.com/tap/async
REQUEST=doQuery
LANG=ADQL
QUERY=SELECT * FROM magnitudes AS m WHERE m.r>=10 AND m.r<=16
```



Remarque concernant les langages de requêtes

- Il est possible de proposer d'autres langages de requêtes comme SQL et dans le futur peut-être d'autres langages de requêtes définis au niveau du VO...



Mais comment décrit-on un service TAP ?

- C'est plutôt souple, il n'est absolument pas obligatoire de décrire toutes les tables de la base de données, de dévoiler le schéma réel de celle-ci
- Le service doit fournir le « TAP_SCHEMA » qui décrit les tables exposées dans le service
 - TAP_SCHEMA = Ensemble de 5 tables qui contiennent une description des schémas, tables et colonnes. Elles sont stockées comme des tables de données et sont donc accessibles via les mêmes outils (ADQL par exemple)



Mais comment décrit-on un service TAP ? (2)

- La description comporte des métadonnées (nom, type de données et unité) et des données sémantiques (description textuelle, un UCD pour la quantité décrite ou un UTYPE pointant vers une partie d'un modèle de données)
- TAP_SCHEMA contient également une description de la façon dont les tables peuvent être jointées
- Il faut par ailleurs tenir compte de VOSI et implémenter les « capabilities », le contenu publié et la disponibilité (« availability » dans VOSI) du services
 - La partie « capabilities » est obligatoire (très utile pour le Registry)



Remarque

- Dans l'absolu il n'est pas obligatoire d'avoir une base de données relationnelles derrière un service TAP ou que le « TAP_SCHEMA » décrive des tables existantes
- La seule obligation est de fournir le « TAP_SCHEMA », d'accepter des requêtes ADQL et de fournir en sortie des données au format VOTable. Le reste peut être simulé par le biais de l'implémentation.



- Ce que doit/peut proposer un service TAP

```
HTTP POST http://example.com/tap/sync
REQUEST=doQuery
LANG=ADQL
QUERY=SELECT * FROM magnitudes as m where m.r>=10 and m.r<=16
```

Feature	Resource	Resource support	Parameters	Parameter support
synchronous query execution	/sync	must		
asynchronous query execution	/async	must		
availability (VOSI)	/availability	should		
capabilities (VOSI)	/capabilities	must	REQUEST=getCapabilities	must
table metadata (VOSI)	/tables	should		
ADQL queries			REQUEST=doQuery LANG=ADQL	must
PQL queries			REQUEST=doQuery LANG=PQL	may
other query languages			REQUEST=doQuery LANG=<other>	may
table upload			UPLOAD	may
VOTable output			FORMAT	must
other formats			FORMAT	should
limiting output			MAXREC	must
logging			RUNID	should



Le résultat d'une requête

- Les formats de sortie conseillés, seul VOTable étant obligatoire

table type	MIME type(s)	short form
VOTable	application/x-votable+xml text/xml	votable
comma separated values	text/csv	csv
tab separated values	text/tab-separated-values	tsv
FITS binary table	application/fits	fits
pretty-printed text	text/plain	text
pretty-printed Web page	text/html	html

Chargement de tables propres à l'utilisateur (Table Upload)

- Le service doit accepter les tables au format VOTable, les tables auront un nom préfixé par le schema TAP_UPLOAD donc TAP_UPLOAD.<nom de la table>
- Les noms de colonnes sont récupérés à partir des noms d'attribut des éléments FIELD et PARAM du fichier VOTable
- Les types de données ? Cf. tableau

VOTable: datatype	VOTable: arraysize	VOTable: xtype TAP_SCHEMA.columns: datatype	database column type
boolean	[1]		Not supported
short	[1]		SMALLINT
int	[1]		INTEGER
long	[1]		BIGINT
float	[1]		REAL
double	[1]		DOUBLE
<numeric type>	> 1		VARBINARY
char	[1]		CHAR(1)
char			VARCHAR ³
char	n*		VARCHAR(n)
char	n		CHAR(n)
unsignedByte			VARBINARY ⁴
unsignedByte	n*		VARBINARY(n)
unsignedByte	n		BINARY(n)
unsignedByte	n, *, n*	adql:BLOB	BLOB
char	n, *, n*	adql:CLOB	CLOB
char	n, *, n*	adql:TIMESTAMP	TIMESTAMP
char	n, *, n*	adql:POINT	POINT
char	n, *, n*	adql:REGION	REGION



TAP_SCHEMA, un peu de détail

TAP_SCHEMA.schemas

Column name	datatype	
schema_name	vvarchar	schema name, possibly qualified
description	vvarchar	brief description of schema
utype	vvarchar	UTYPE if schema corresponds to a data model

TAP_SCHEMA.tables

Column name	datatype	
schema_name	vvarchar	the schema name from TAP_SCHEMA.schemas
table_name	vvarchar	table name as it should be used in queries
table_type	vvarchar	one of: table, view
description	vvarchar	brief description of table
utype	vvarchar	UTYPE if table corresponds to a data model



TAP_SCHEMA, un peu de détail (2)

TAP_SCHEMA.columns

Column name	datatype	
table_name	varchar	table name from TAP_SCHEMA.tables
column_name	varchar	column name
description	varchar	brief description of column
unit	varchar	unit in VO standard format ⁵
ucd	varchar	UCD of column if any
utype	varchar	UTYPE of column if any
datatype	varchar	ADQL datatype as in section 2.5
size	integer	length of variable length datatypes ⁶
principal	integer	a principal column; 1 means true, 0 means false
indexed	integer	an indexed column; 1 means true, 0 means false
std	integer	a standard column; 1 means true, 0 means false



TAP_SCHEMA, un peu de détail (3)

TAP_SCHEMA.keys

Column name	datatype	
key_id	varchar	key identifier from the TAP_SCHEMA.keys
from_column	varchar	key column name in the <from_table>
target_column	varchar	key column in the <target_table>

TAP_SCHEMA.keys_columns

Column name	datatype	
key_id	varchar	unique key identifier
from_table	varchar	fully qualified table name
target_table	varchar	fully qualified table name
description	varchar	description of this key
utype	varchar	utype of this key

Cela fait donc bien 5 tables

Des outils pour publier ses données et les rendre accessibles via TAP (mais pas seulement...)

Publication Tools	Description
Saada http://saada.unistra.fr Strasbourg Observatory (Fr)	Integrated tool including the database creation and management, a Web interface and the VO services. Use the CDS/ARI library.
DaCHS http://vo.ari.uni-heidelberg.de/soft/dachs GAVO (Ge)	DaCHS is a multi-protocol toolkit containing a full TAP stack including geometries and upload
OPENCADC https://code.google.com/p/opencadc/ CADC (Ca)	Java toolkit for deploying a TAP service. Open-source branch of the software developed by the CADC for its services
VODance http://ia2.oats.inaf.it/index.php/vodance INAF (It)	VODance is a tool for rapid Virtual Observatory compliant services deployment.
CDS library https://github.com/gmantele/taplib CDS (Fr) + ARI (Ge)	A pack of 3 Java libraries (ADQL, UWS and TAP) helping to build a TAP layer on the top of an existing database.
TAPtoolkit http://dev.usvao.org/vao/wiki/Products VAO (US)	To be released (Its to help data provider their own data without coding all TAP service themselves, it helps as a guideline.)

Des clients pour accéder aux services TAP

Generic Clients	Description
STILTS tapquery http://www.star.bristol.ac.uk/~mbt/stilts/sun256/tapquery.html M Taylor - Bristol University (UK)	Java library: Stilts is a set of command-line tools based on the Starlink Tables Infrastructure Library. It deals with the processing of tabular data.
TOPCAT TAP client http://www.star.bristol.ac.uk/~mbt/topcat/sun253/TapTableLoadDialog.html M Taylor - Bristol University (UK)	JAVA Standalone client: TOPCAT is an interactive graphical viewer and editor for tabular data
TAPHandle http://saada.unistra.fr/taphandle/ Strasbourg Observatory (Fr)	Universal Web portal for TAP services
GAVO DC Virtual Observatory Libraries http://soft.g-vo.org/subpkgs GAVO (Ge)	Python library: Some parts of DaCHS useful for other VO software are available as separately installable libraries.
TAP shell http://vo.ari.uni-heidelberg.de/soft/tapsh University of Heidelberg (Ge)	The TAP shell provides a command line interface to querying TAP servers, complete with metadata management and command line completion
VESPA http://voparis-europlanet-new.obspm.fr/planetary/data/epr/query/all/ VO PARIS (Fr)	Web client for solar and planetary data access
jc client http://jvo.nao.ac.jp/jcclient NAO (Jp)	JVO command line tool
VODb https://sites.google.com/site/virtualobservatorydatabase/credits University Observatory Munich (Ge)	VODb is a Java desktop application that simplifies the process of extracting data from astronomical databases
Seleste http://cda.cfa.harvard.edu/seleste/CFA	Seleste is a Java tool that uses VO standards and protocols to access archives that expose tabular data and query them using a uniform interface.
Astrotaverna http://amiga.iaa.es/p/290-astrotaverna.htm Instituto de Astrofísica de Andaluca (Es)	Taverna (workflow management system) plugin integrating discovery and access to TAP services.



Des clients directement liés aux services

Dedicated Clients	Description
Vizier TAP interface http://tapvizier.u-strasbg.fr/adql/ CDS (Fr)	A complete web application dedicated to the Vizier TAP service. A table selector help to locate Vizier tables
Simbad TAP interface http://simbad.u-strasbg.fr/simbad/sim-tap CDS (Fr)	Complete web application dedicated to the SIMBAD TAP service. Query templates are provided.
CADC TAP interface http://www1.cadc-ccda.hia-ih.nrc-cnrc.gc.ca/tap/ CADC (Ca)	Simple page helping to discover the CADC TAP service with various examples.
GAVO TAP interface http://dc.zah.uni-heidelberg.de/system_/tap/run/tap/async GAVO (Ge)	Basic html form to submit ADQL jobs on the GAVO TAP node.



TAP et le Registry

- VOSI/capabilities, VOSI/availability, VOSI/tables
- + extensions

Les services TAP dans le Registry Euro-VO



The screenshot displays the Euro-VO Registry interface. At the top, the logo for EURO AIDA (Astronomical Infrastructure for Data Access) is visible. Below the logo, there are navigation tabs for 'The Euro-VO projects: VOTECH EuroVO-DCA EuroVO-AIDA'. A left-hand sidebar contains a list of resource types, with 'Table Access (TAP)' highlighted in green. The main content area is titled 'Search Results' and shows a list of search results. The first result is 'Two Micron All Sky Survey [TWOMASS_JVO]' with an IVOA Identifier and a description of the project. The second result is 'Subaru Suprime-Cam data service [SUBARU_SUP]' with an IVOA Identifier and a description of the camera. The third result is 'Subaru HDS Spectrum data service [SubaruHDS]' with an IVOA Identifier and a description of the data service. The fourth result is 'The SWIRE Data Service [SWIRE]' with an IVOA Identifier. Each result includes a 'Published by' line and a 'last updated' line.

.. Et côté VAO
<http://vao.stsci.edu/directory/>



NOTICE: For our long-standing users who have come to this page through our old website at <http://nvo.stsci.edu/vor10>, please note that as part of our project's administrative switch to VAO, and as part of a major hardware and software upgrade, all pages under the nvo.stsci.edu domain have been deprecated, and will not remain active indefinitely.

Please point your browsers to the same pages under <http://vao.stsci.edu/directory> for VAO registry search pages and services. We have made considerable efforts toward backward-compatibility. If you are using the new site and continue experiencing any issues, please contact the [VAO help desk](#).

Find Astronomical Data Resources

Enter terms in the text box that describe the type of data you are looking for. Results will show catalogs and data collections that have these terms as part of their descriptions.

Examples: [quasar](#), [AGN](#), [binary stars](#), [Chandra](#), [GALEX](#), [far ultraviolet](#)

Use [the VAO Data Discovery Tool](#) to search and view data for a particular object or position.

TAP

Next, Send Results to

Results 1-20 of 131

 Show results per page

 Click column heading to sort list - Click checkbox to select

 Text boxes under columns select matching rows

select	browse / query	categories	shortName	title	description	publisher	waveband	Identifier
<input type="checkbox"/>	Full Record More Info	VOSI Availability VOSI Capabilities VOSI Tables Table Service	SIMBAD TAP	SIMBAD TAP query engine	This service provides TAP access to a simplified view of the SIMBAD database	CDS SIMBAD service		ivo://CDS.SIMBAD/TAP
<input type="checkbox"/>	Full Record More Info	Web Page Catalog Table Service Custom Service HTTP Request VOSI Capabilities HTTP Request VOSI Availability HTTP Request HTTP Request HTTP Request		ATNF Publishing Registry	Repository of pulsar observations made at the Parkes, Australia radio telescope	Australia Telescope National Facility		ivo://au.csiro/psrda/ATNF_Pulsa
<input type="checkbox"/>	Full Record More Info	Table Service	MACHO TAP	MACHO TAP	MACHO TAP service	National Computational Infrastructure		ivo://nci.org.au/macho/tap
<input type="checkbox"/>	Full Record More Info	Table Service	WiggleZ TAP	WiggleZ Final Data Release TAP	WiggleZ final data release TAP service	National Computational Infrastructure		ivo://nci.org.au/wigglez/tap
<input type="checkbox"/>	Full Record More Info	Table Service	HIPASS TAP	HIPASS TAP	HIPASS TAP service	National Computational Infrastructure		ivo://nci.org.au/hipass/tap
<input type="checkbox"/>	Full Record More Info	Web Page VOSI Availability VOSI Capabilities VOSI Tables	GloTS plain	Global TAP Schema Simple Search	The global TAP schema collects information on tables and columns from known TAP servers. This facilitates locating queryable data by physics (via UCED) or keywords (via description).	The GAVO DC team		ivo://org.gavo.dc/glots/q/plain
<input type="checkbox"/>	Full Record More Info	VOSI Capabilities VOSI Availability		CADC Table Query (TAP) Service		Canadian Astronomy Data Centre		ivo://cadc.nrc.ca/tap



Il existe des outils pour tester des services TAP

- Exemple: taplint
 - <http://www.star.bristol.ac.uk/~mbt/stilts/sun256/taplint.html>, M. Taylor

B.15.2 Examples

Here are some examples of taplint:

```
stilts taplint http://dc.zah.uni-heidelberg.de/__system__/tap/run/tap
```

Performs a default validation run against the TAP service based at the given URL.

```
stilts taplint tapurl=http://example.com/tap
report=EW stages='TMS UWS' truncate=80 maxrepeat=4
```

A validation run is done against the named TAP service. Only Error and Warning type messages are output, only two validation stages are performed, lines are truncated to a maximum of 80 characters, and each message is repeated a maximum of 4 times. An invocation like this may be suitable if you find the default operation too verbose.

The output of this invocation might look like this:

```
Section TMS: Check content of tables metadata from TAP_SCHEMA
E-TMS-CINT-1 Column principal in TAP_SCHEMA.columns has wrong type char not int
E-TMS-CINT-2 Column std in TAP_SCHEMA.columns has wrong type char not int
W-TMS-CLUN-1 Unused entry in TAP_SCHEMA.columns table: ivoa.obscore

Section UWS: Test asynchronous UWS/TAP behaviour
E-UWS-GMIM-1 Incorrect Content-Type text/xml != text/plain for http://exampl....
E-UWS-GMIM-2 Incorrect Content-Type text/xml != text/plain for http://exampl....
E-UWS-GMIM-3 Incorrect Content-Type text/xml != text/plain for http://exampl....
E-UWS-GMIM-4 Incorrect Content-Type text/xml != text/plain for http://exampl....
E-UWS-GMIM-x (3 more)

Totals: Errors: 9; Warnings: 1
```



Des exemples de services TAP (non exhaustif)

- <http://www1.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/tap/>
- <http://simbad.u-strasbg.fr/simbad/sim-tap>
- <http://tapvizier.u-strasbg.fr/adql/>
- Etc.

- Pour les outils, cf. les présentations à venir



Nous n'avons pas abordé...

- STC-S
- La gestion des erreurs, les overflows, le versioning,
- ObsTAP: service TAP qui retourne la description d'un ensemble de données non tabulaires
- PQL: Parameter Query Language



Conclusion

- Développer son service TAP sans recourir à une librairie ou à un « framework » existant est tout à fait faisable mais prend du temps
- TAP étant basé sur REST, le développement d'un client TAP est par contre plus simple
- Les présentations à venir vous donneront:
 - de nombreux exemples d'implémentation du standard à la fois côté serveur et client dans des « frameworks » et des librairies
 - l'occasion de revoir des exemples de requêtes ADQL