

# VOTable, petit lexique

François Ochsenbein  
CDS, Strasbourg

# C'est quoi, VOTable ?

- ▶ VOTable est un format XML pour *échanger* des données tabulaires, dans le cadre de l'Observatoire Virtuel
  - doit permettre l'échange de **grosses** tables (plusieurs millions de tuples) ⇒ une certaine "factorisation" de la description
- ▶ VOTable doit rester compatible avec les tables binaires (*accès au contenu des tables FITS sans réécriture de données*)

# C'est quoi, une table ?

- ▶ Structure de données universelle (bases relationnelles), bien adaptée aux catalogues et gros ensembles de données, qui comprend 2 parties:
  - une description de la *structure* de la table (vue en tant que classe d'objet, colonnes = attributs), qu'on appelle aussi **métadonnées**
  - une collection (importante) de ces objets (tuples) qui partagent cette même structure.

# Quel intérêt du XML?

- ▶ **L'Interopérabilité** nécessite l'utilisation de *structures de données* standardisées et une *description standardisées du contenu* (metadata)
  - Problème: FITS est bien une structure de données standardisée, mais les métadonnées sont hétérogènes
- ▶ **XML built-ins:**
  - tous les logiciels de vérification, validation...
  - transformations et visualisation des documents XML (XSLT)

# Pourquoi pas du "pur XML" ?

- ▶ **Accès aux FITS** existants qui sont des fichiers binaires (ne peuvent être inclus dans du XML)
- ▶ **Performances** (CPU, mémoire, réseaux): pour des catalogues de plusieurs millions de lignes...

Exemple de ce que pourrait être un pur XML

# VOTable Data Model

- ▶ **VOTable** = hiérarchie de **Resource**
- ▶ **Resource** = ensemble de **Tables**
- ▶ **Table** = **Metadata** + **Data**
- ▶ **Metadata** = **Infos** + **Descriptions** + **Parameters** + **Links** + **Fields**
- ▶ **Data** = suite de **Rows** ou **STREAM** (remote))
- ▶ **Row** = liste de **Cells**
- ▶ **Cell** = **Primitive**  
ou bien liste (longueur fixe ou variable) de **Primitives**  
ou bien ensemble multidimensionnel de **Primitives**
- ▶ **Primitive** = integer, character, float, floatComplex ,...

# Structure d'une VOTable

contient les éléments suivants:

- ▶ **DEFINITIONS** (système de coordonnées)
- ▶ **RESOURCE** contient une **DESCRIPTION** et une liste de tables, éventuellement d'autres **RESOURCES**
- ▶ **TABLE** contient:
  - une **DESCRIPTION** de la table
  - une liste de **FIELD** (colonnes) qui décrivent le schéma de la table + des **PARAMETER** pour précision éventuelle de *constantes*
  - la partie **DATA** qui peut être incluse dans le document, ou renvoyer à des fichiers FITS, du binaire, une URL...
  - des liens **LINK** pour récupérer des détails, des données annexes,  
...
- ▶ **Example**

# <TABLE>

- ▶ Contient **DESCRIPTION** + une collection d'éléments **FIELD** + **PARAMETER** éventuels + **LINK** = *métadonnées*; suivi par le bloc **DATA**
- ▶ **FIELD** décrit une colonne de la table
- ▶ **DATA** démarre la partie données, qui contient les valeurs de chaque colonne dans le même ordre que leur déclaration, groupées en lignes (tuples)

# <FIELD> en détail

- ▶ Possède les sous-éléments **DESCRIPTION**, **LINK**, et éventuellement **VALUES** pour préciser le domaine
- ▶ Rôle = définir aussi précisément que possible le contenu de la colonne qui se trouve dans la partie **DATA**
- ▶ L'identification d'un **FIELD** se fait par **name**, et par **ID**
  - **name** est destiné par ex. à l'affichage
  - **ID** est un identifiant XML (formé d'un jeu de caractères restreint, doit être unique dans tout le document XML)
- ▶ **FIELD** doit définir un **datatype**
- ▶ Chaque colonne peut contenir un vecteur multidimensionnel du type précisé par **arraysize**, qui peut avoir la dernière dimension indéfinie (**64x64x\***)

# Datatypes

<code>datatype</code>	Meaning	FITS	Bytes
<code>"boolean"</code>	Logical	"L"	1
<code>"bit"</code>	Bit	"X"	*
<code>"unsignedByte"</code>	Byte (0 to 255)	"B"	1
<code>"short"</code>	Short Integer	"I"	2
<code>"int"</code>	Integer	"J"	4
<code>"long"</code>	Long integer	"K"	8
<code>"char"</code>	ASCII Character	"A"	1
<code>"unicodeChar"</code>	Unicode Character		2
<code>"float"</code>	Floating point	"E"	4
<code>"double"</code>	Double	"D"	8
<code>"floatComplex"</code>	Float Complex	"C"	8
<code>"doubleComplex"</code>	Double Complex	"M"	16

# Field cont..

- ▶ `unit` précise l'unité (vocabulaire contrôlé)
- ▶ `ucd` (Unified Content Descriptors)
  - donne un "type sémantique" de la donnée contenue dans la colonne, sous forme d'un texte standardisé en cours de discussion intense...
- ▶ `precision`, `width` donnent des éléments pour une représentation éditée d'une valeur
- ▶ **VALUES**
  - précise les informations relatives au *domaine* : min, max, null, valeurs permises.
- ▶ **LINK**
  - sert de pointeur vers d'autres documents ou serveurs par le biais d'UR[IL]
  - peut être typé ( "doc" ...)

# <PARAM>

- ▶ **PARAM** est très semblable à **FIELD** : peut contenir une **DESCRIPTION**, possède des attributs `unit`, `ucd`, `name`, `ID`,... plus l'attribut obligatoire `value`
- ▶ peut être interprété comme une colonne *constante*
- ▶ Utilisation typique:
  - fréquence ou  $\lambda$  utilisé
  - erreur statistique moyenne

# <DATA>

- ▶ Un seul **DATA** par table.
- ▶ Plusieurs façons d'accéder aux données, dans plusieurs formats:
  - **TABLEDATA** lorsque les données sont en XML
  - **FITS** permet d'accéder à un fichier FITS externe
  - **BINARY** pour accéder à des données binaires (pour plus d'efficacité)
- ▶ **STREAM** pour accéder à des données non XML
  - Les données peuvent être dans le flux XML (éventuellement encodées), ou stockées séparément
- ▶ **Exemple**

# Evolution de VOTable

Quelques nouveautés pour VOTable1.1:

- ▶ **GROUP** permet de définir une structuration des colonnes – par exemple, association d'une valeur avec son erreur, sa source, etc
- ▶ **utype** est un typage précis en relation avec un "data model" externe (défini par l'application)
- ▶ **encoding** pour une cellule quelconque – permet de mélanger du XML et binaire

# Quelques utilisations VOTable

- ▶ Cone Search – voir  
<http://voservices.org/cone/register/showlist.asp>
- ▶ SIAP (interrogation serveur d'image)  
retourne du VOTable
- ▶ etc...

# Quelques outils

- ▶ En java, ingestion de VOTable (savot, javot, ...)
- ▶ Conversion FITS->VOTable (toujours possible)
- ▶ VOPlot (à partir de VizierR)
- ▶ Mirage <http://www.bell-labs.com/project/mirage/>



