

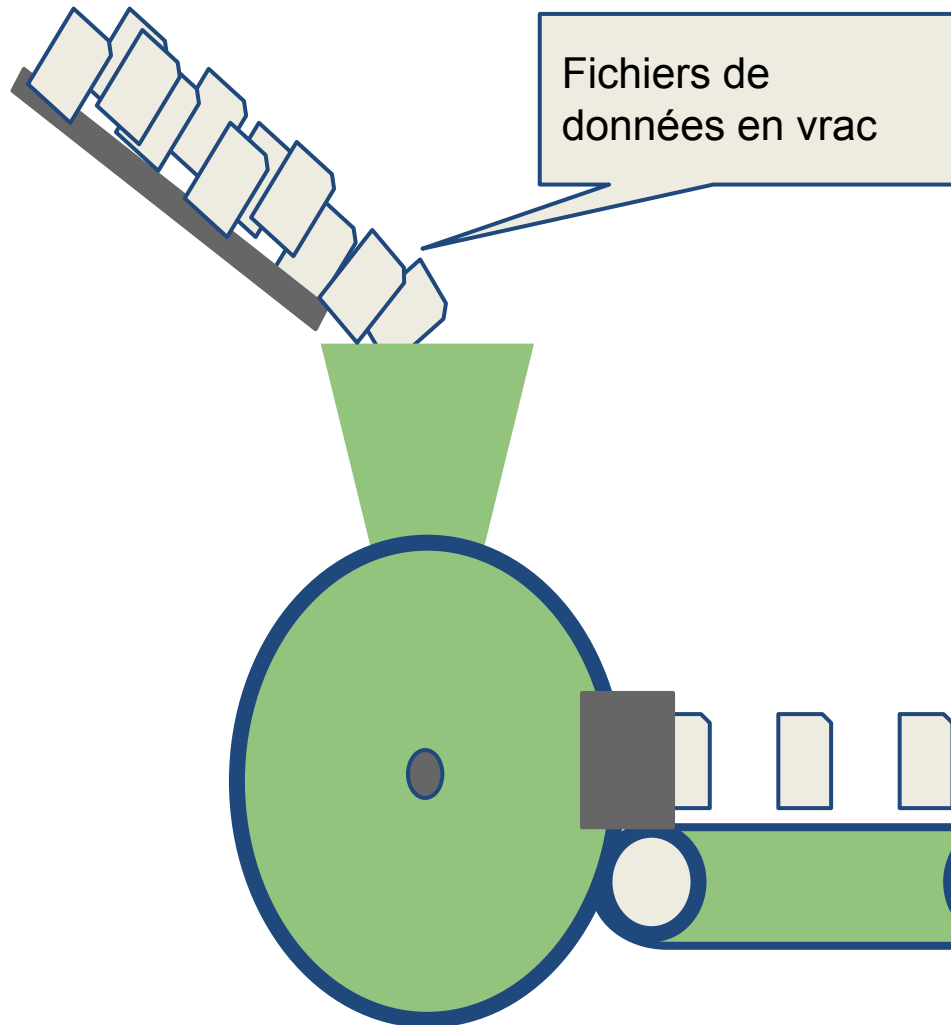


<http://saada.unistra.fr>

laurent.michel@astro.unistra.fr



Une machine à ranger et à indexer des données astro



Fichiers de données rangés.

- Ils sont **consultables en ligne** et accessibles par **téléchargement**.
- ils peuvent être **retrouvé par leur contenu**.



Saada en quelques mots

- ✓ **Construction automatique d'une base de données à partir de fichiers de données (FITS, VOTable ...).**
 - **Pas de code** à écrire
 - Archivage de **données hétérogènes**
 - Hébergement de **collections multiples**
- ✓ **Extraction et archivage des données**
 - Extraction des mots clés lus dans les fichiers chargés
 - Rangement des données de manière à faciliter la réponse aux requêtes des utilisateurs
 - Le fichier d'origine reste accessible
- ✓ **Outils de gestion et de publication des données**
 - Possibilité d'établir des **liens persistants** entre les données.
 - Marquage à la main des méta-données (UCD, unités...)
 - Accès par le biais de l'interface Web ou de protocoles VO



UNIVERSITÉ DE STRASBOURG

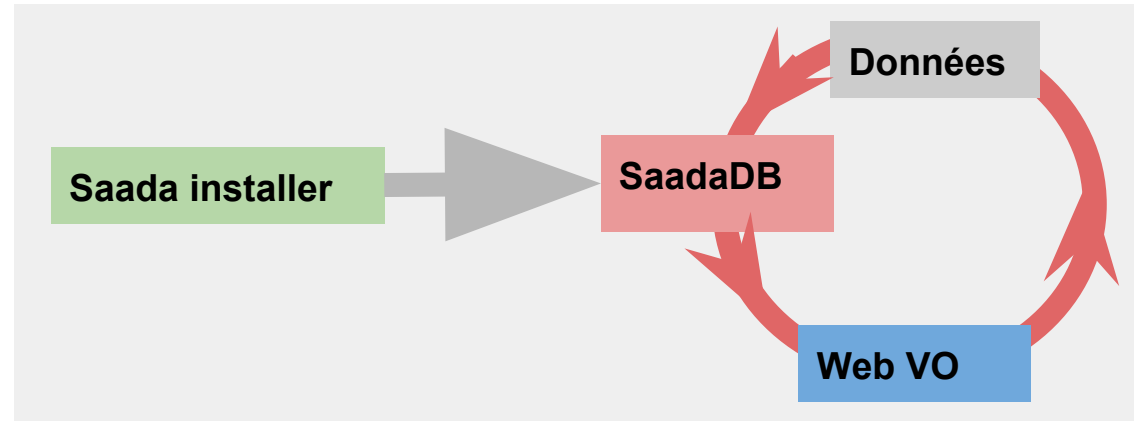




Saada en quelques mots

✓ Le principe

- Un installateur met en place un canevas de base de données, la SaadaDB) dans lequel l'utilisateur organise le stockage des données.



✓ Une fois créée, la SaadaDB est une entité autonome

- Plus de lien avec l'installateur
- Pas de connexion avec un serveur central

✓ Constituants de la SaadaDB

- Le système de stockage (PostgreSQL, MySQL ou SQLite)
- Le data loader (Interface graphique scriptable)
- L'interface Web (RIA basée sur JQuery)
- Le repository (répertoire de travail et d'archivage de fichiers)



UNIVERSITÉ DE STRASBOURG





Saada en quelques mots

– Saada

l'installateur s'appelle Saada comme le projet (choix malheureux).

– SaadaDB

La base de données construite à partir du canevas a son propre nom. Elle est toutefois référencée dans les documents comme une SaadaDB.

– Produit

Ensemble de mots clés/valeurs

- Ensemble vide: fichier plat (flatfile)
- Ensemble non vide: fichier structuré (FITS ou VOTable)

– Format d'un produit

Ensemble des mots clés qui lui sont associés (typage inclus)

– Collection hétérogène

Ensemble de produits dont au moins un a un format différent d'un autre



UNIVERSITÉ DE STRASBOURG





Saada en quelques mots

✓ Une SaadaDB peut être vue comme:

- Une **base relationnelle** dont les tables contiennent des données extraites de fichiers par le data-loader. Cette base peut être utilisée par n'importe quelle application ayant un connectivité avec le monde des SGBDRs.
- Une (multi) **ressource OV** permettant d'accéder à des données via SIA, SSA, CS ou TAP.
- Une **application Web**

✓ Les données d'une SaadaDB peuvent être sélectionnées par

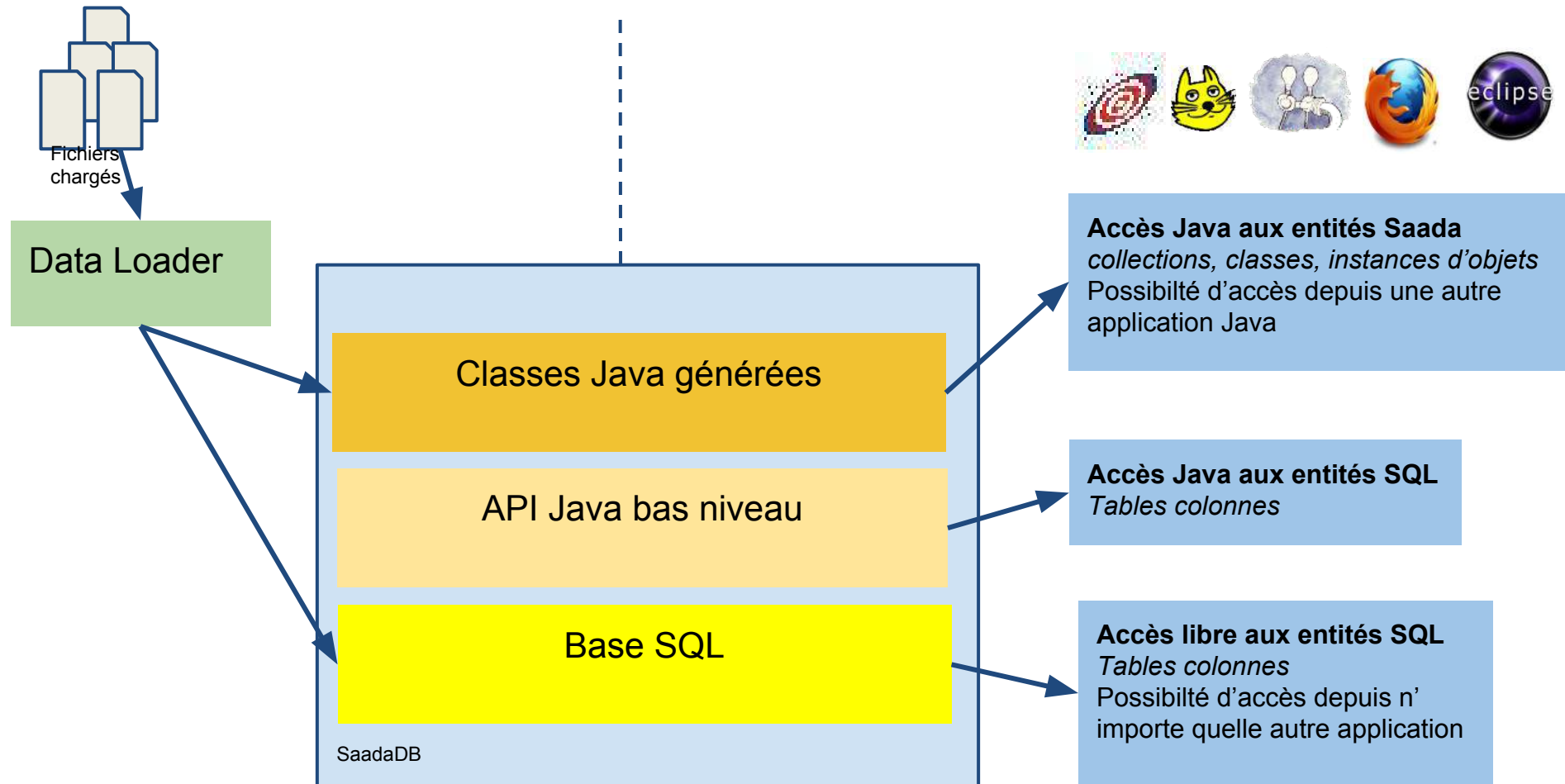
- Des requêtes simples: position, mots clés
- Des requête ayant une saveur *fouille de données*

```
Select all XMM detections
located near ABEL 426
    and having a flux greater than 1e-13
    and matching at least one Simbad source
        having an obj_type containing the string `Radio'
```



UNIVERSITÉ DE STRASBOURG





✓ Les collections

- Les données sont réparties dans des collections définies par l'utilisateur
- Une collection est un conteneur abstrait doté d'un nom
- L'affectation d'une collection à un ensemble de fichiers à charger est entièrement libre

✓ 5/6 catégories de données

- Les données sont organisées en 5 catégories qui correspondent aux grandes classes de données astronomiques

<i>catégorie</i>	<i>Type de fichiers</i>	<i>Type d'extension</i>
FLATFILE	Any	no
MISC	FITS, VOTable	Any
SPECTRUM	FITS, VOTable	Table or image
IMAGE	FITS	Image
TABLE + ENTRY	FITS, VOTable	Table

Saada 2 niveaux d'archivage

✓ Stockage des données natives

- Les valeurs extraites des fichiers sont **stockées telles quelles**
- Elle restent accessibles en lecture ou utilisables comme critères de sélection

✓ Création d'une interface commune

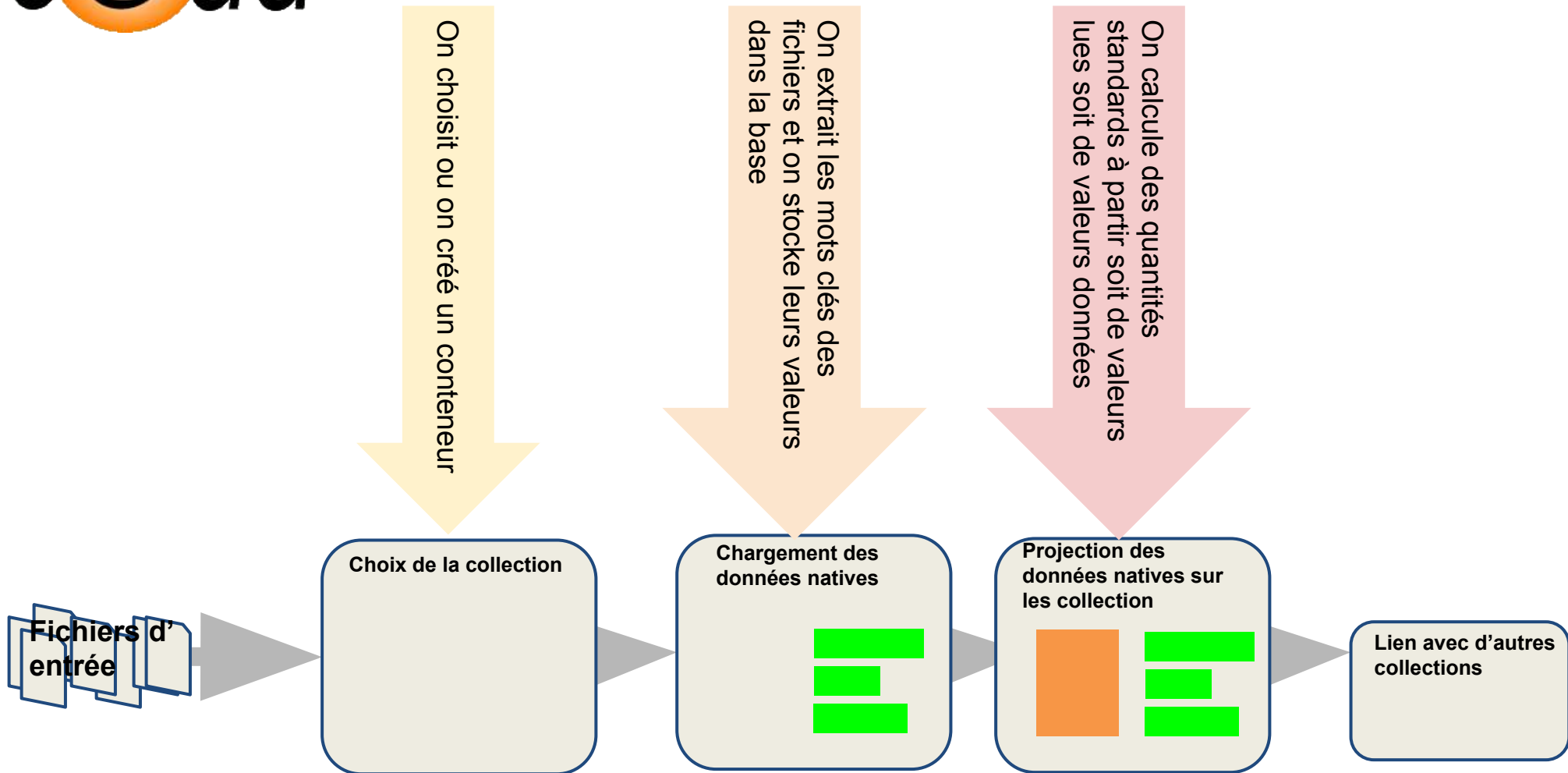
- Saada calcule un certain nombre de valeurs caractéristiques qui sont également exposées
- Le format de ces valeurs est **commun à toute la base** (nom, type unité...).
- Elles couvrent les axes de caractérisation (espace, temps, énergie, observation)
- Elles permettent de faire des **requêtes** sur des collections de données **initialement hétérogènes**
- L'utilisateur peut ajouter ses propres colonnes à cette interface commune
- Ces colonnes sont calculées automatiquement, mais l'utilisateur peut fournir des règles
- Cette interface commune est nommée **données de niveau collection**

Rupture entre
1.8 et 2





UNIVERSITÉ DE STRASBOURG





Saada Démo 1

Data Loader

◀ Home  Foo  TABLE
Filter: Default

Filter Chooser

Default	Description of selected filter
Galaxy	ArgsParser(-category=table)

Repository parameters

Copy Input Files into the Repository
 Use Input Files as Repository
 Move Input Files to the Repository
 Do not rebuild indexes after loading

Data Files Selector

3 file(s) selected.

Chargement images par défaut

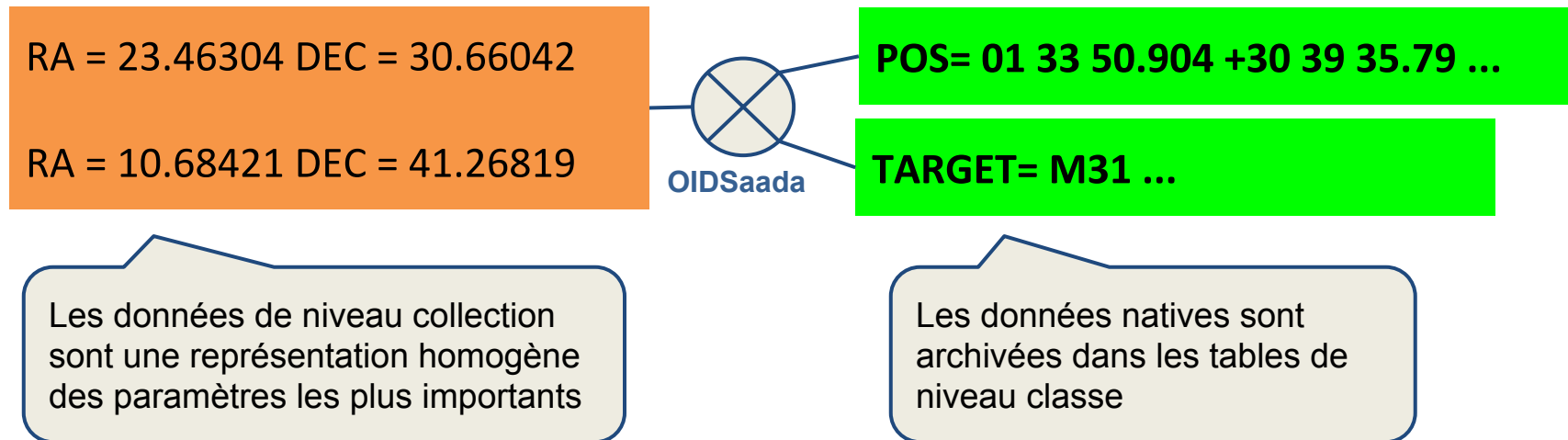


- ✓ **Les données natives sont regroupées dans des classes**
 - Une classe contient un ensemble d'enregistrements de **même format**
 - Elle est implémentée à la fois sous la forme d'une **classe Java** et d'une **table SQL**
 - Elle est attachée à une catégorie dans une collection
 - Les classes sont utilisées pour délimiter le périmètre des requêtes

- ✓ **Deux modes de classification**
 - Le mode **classifier** (mode par défaut): tous les fichiers ayant le même format (même jeu de mots clés) sont rangés dans une même classe.
 - Le mode **fusion**: tous les fichiers sont rangés dans la même classe quelque soit leurs formats.
 - ✓ Les colonnes non affectées restent nulles
 - ✓ Les conflits de types sont résolus par transtypage (boolean-> string)
 - Le choix du mode de classification relève de l'utilisateur

✓ Structure interne d'une collection

- Une collection est découpée en 6 conteneurs: un par catégorie + ENTRY
- Pour chaque catégorie:
 - ✓ Une **table** par pour les données de **niveau collection**
 - ✓ Une **table** par classe de **données natives**
- Les tables de niveau **classes** et **collections** sont **jointes** par un identifiant universel de Saada (OIDSaada)



saa da Calcul des données collection

✓ 2 modes d'affectation:

- **Détection automatique** par le data loader: recherche basée sur les noms de colonnes et les UCDs
- **Règles utilisateur**: pour chaque colonne l'utilisateur peut fournir soit le mot-clé natif correspondant soit une valeur constante.

✓ 3 modes de priorité:

- L'utilisation des deux modes est régulé par un ordre de priorité défini par l'utilisateur
 - ✓ ONLY: on applique uniquement la règle utilisateur
 - ✓ FIRST: (default) on applique d'abord la règle utilisateur, puis la recherche automatique en cas d'échec.
 - ✓ LAST: on applique d'abord la recherche automatique puis la règle utilisateur en cas d'échec.

-position='M33' : La valeur de la position (niveau collection) pour tous les produits appliquant cette règle sera la position de M33 .

-position=RA,DEC: Tous les produits pour lesquels cette règle est appliquée auront pour position (niveau collection) les valeurs des mots clés RA et DEC





UNIVERSITÉ DE STRASBOURG



Saada Démo 2

Data Loader

◀ Home  Foo  TABLE
Filter: Default

Filter Chooser

Default	Description of selected filter
Galaxy	ArgsParser(-category=table)

Repository parameters

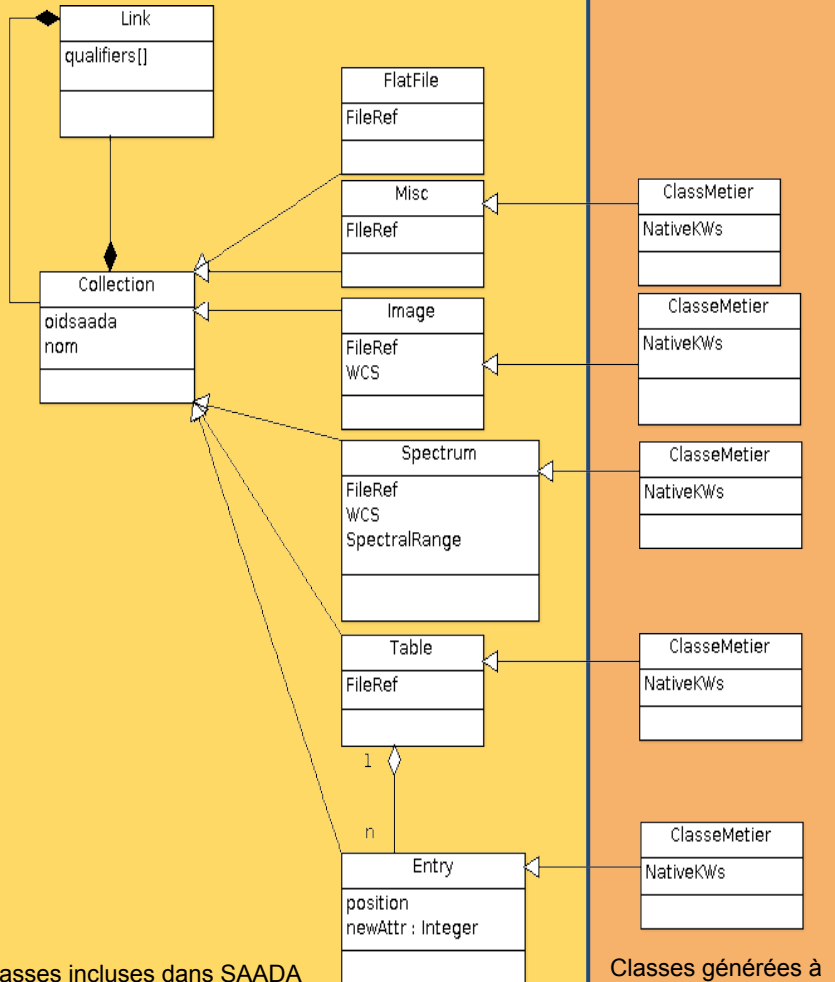
- Copy Input Files into the Repository
- Use Input Files as Repository
- Move Input Files to the Repository
- Do not rebuild indexes after loading

Data Files Selector

3 file(s) selected.

Chargement images une classe une position

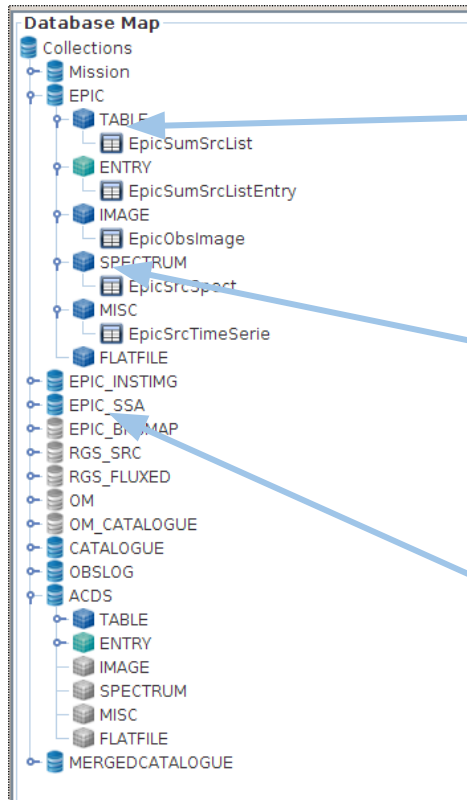




Classes incluses dans SAADA
Des attributs peuvent être ajoutés lors de la création de la base

Classes générées à la volée lors du chargement

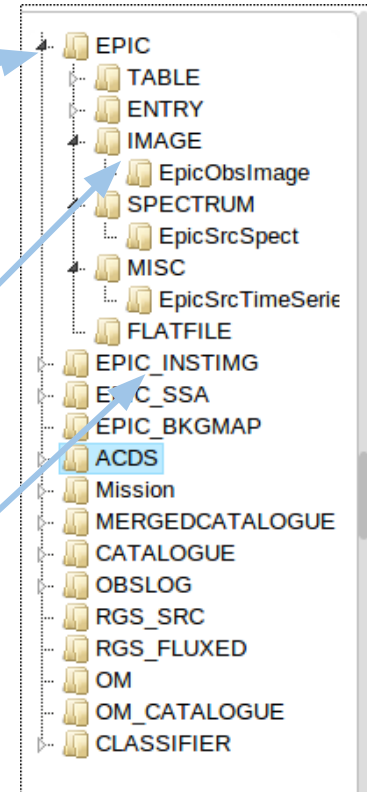




collection
Conteneur de données

catégorie
Données calculées par Saada
suivant un modèle commun à l'
ensemble de la base

classe
Données natives des
produits chargés



Depuis l'outil d'
administration

Depuis la page Web



Gérer une SaadaDB

The screenshot displays the Saada 1.8.0.admin tool interface. On the left, an Ant script editor shows the following XML configuration:

```
<?xml version="1.0" encoding="null"?>
<project default="user.task" name="userTask">
  <property file="http://192.168.0.37:8888/tapdemo/bin/saadadb.properties"/>
  <property name="jvm_initial_size" value="-Xms64m"/>
  <property name="jvm_max_size" value="-Xmx1024m"/>
  <!-- This classpath is used by all java calls. Classes or jar files specifiv for an
  <path id="saadadb.classpath">
    <pathelement location="${SAADA_DB_HOME}/class_mapping"/>
    <fileset dir="${SAADA_DB_HOME}/lib">
      <include name="**/*.jar"/>
    </fileset>
    <fileset dir="${SAADA_DB_HOME}/jtools">
      <include name="**/*.jar"/>
    </fileset>
  </path>
  <target name="user.task">
    <java classname="saadadb.dataloader.Loader" failonerror="true" fork="true">
      <classpath refid="saadadb.classpath"/>
      <arg value="-collection=Foo"/>
      <arg value="-repository=no"/>
      <arg value="-noindex"/>
      <arg value="-classfusion=EPICSpectra"/>
      <arg value="-category=spectrum"/>
      <arg value="-spcmapping=only"/>
      <arg value="-spccolumn='0,2 12'"/>
      <arg value="-spcunit=keV"/>
      <arg value="-tapdemo"/>
    </java>
  </target>
</project>
```

The main interface, titled "Saada 1.8.0.build5 - Admintool for the tapdemo database", features a "Root Panel" with a "Database Map" on the left and a "Data Management" section with icons for "Create Collection", "Load Data", "Explore Data", "Manage Data", "Manage Meta Data", and "Manage Relationships". A "Data publication" section at the bottom includes a "VO Publishing" icon.

Toutes les commandes peuvent s'effectuer soit depuis l'interface graphique, soit par script *ant*



saada L'interface WEB

✓ Une application JEE

- Servlet + JQuery
- Déploiement sur Tomcat depuis l'outil d'administration
- Doit être redéployée après chaque modification des classes de données

✓ Principales fonctions

- Navigation dans le contenu de la base
- Éditeur de requêtes SaadQL
- Éditeur de requêtes S*P
- Connexion SAMP
- Téléchargement par lots via un Caddie



UNIVERSITÉ DE STRASBOURG





L'interface WEB

Localisation des données affichées

Gestion du Caddie

Connecteur SAMP

Sélection de données

Arbre du contenu de la base

Accès aux détail des données

Editeur de requêtes

The screenshot displays the 'ThreeXMM CATALOGUE>ENTRY' page. On the left is a tree view of the database content. The main area shows a table of entries with columns for Access, Position, Error (arcsec), Name, and various relationship links. Below the table is a query editor with fields for Position, Const on Keywords, Relationship(s), and Counterpart Class.



Access	Position	Error (arcsec)	Name	Rel : CatSrcToEpicimg	Rel : CatSrcToObsimg	Rel : O
	03:18:23.22+41:09:48.2 (s)	±1.4114	3XMM J031823.2+410949	3 links		No link
	03:18:38.50-66:23:10.5 (s)	±1.4502	3XMM J031838.4-662310	3 links		No prev
	03:18:46.49-62:30:17.4 (s)	±0.8681	3XMM J031846.4-623016	3 links		No link

Query Editor details:
Relationship(s): CatSrcToArchSrc
Qualifier identifi_proba
Qualifier identifi_params
Qualifier local density
Counterpart Class: arch_0121TEntry












saa da Démo 3

saa da ThreeXMM CATALOGUE>ENTRY

about  

Previous Next Show 10 entries Showing 1 to 10 of 100 entries

Access	Position	Error (arcsec)	Name	Rel : CatSrcToEpicimg	Rel : CatSrcToObsimg	Rel : O
	03:18:23.22+41:09:48.2 (s)	±1.4114	3XMM J031823.2+410949	3 links 		No link
	03:18:38.50-66:23:10.5 (s)	±1.4502	3XMM J031838.4-662310	3 links 		No prev
	03:18:46.49-62:30:17.4 (s)	±0.8681	3XMM J031846.4-623016	3 links 		No link

SUBMIT

Position Const on Keywords UCD based Const Pattern Plain Text Query

Relationship(s): CatSrcToArchSrc

Qualifier identifi_proba
Qualifier identifi_params
Qualifier local_density

Counterpart Class arch_0121TEntry

Pattern under Construction

Active Patterns
CatSrcToArchSrc

Result Limit: 100
1/1

Query Mode



Les relations Saada

✓ Des liens permanents qualifiés

- Saada peut gérer des liens persistants reliant des données de deux catégories/collection
- Ces liens peuvent être qualifiés par des valeurs numériques (distance p.e.)
- Les relations ont un nom utilisable dans des requêtes

✓ Utilisation des relations

- Intégration des données associées sur la page Web

Access	Position	Error (arcsec)	Name	Rel : CatSrcToEpicimg	Rel : CatSrcTo...
	03:18:23.22+41:09:48.2 (s)	±1.4114	3XMM J031823.2+410949	3 links	



données associées

```
Select ENTRY From CatalogueEntry In CATALOGUE
Where Relation {
  matchPattern { CatSrcToArchSrc,
  AssObjClass{arch_0003AEntry},
  Qualifier{epic_cat_dist < 3}}
}
```



associées dans le Caddie

Keep/Discard	Data Source	Include Linked Data	Resource Name
<input checked="" type="checkbox"/>	QUERY_RESULT	<input checked="" type="checkbox"/>	query_0

Processing status

Current Job Status nojob

Manage Content



Manage Job

Get the Result



saa da Démo 4

Data Loader

◀ Home  Foo  TABLE
Filter: Default

Filter Chooser

Default	Description of selected filter
Galaxy	ArgsParser(-category=table)

Repository parameters

Copy Input Files into the Repository
 Use Input Files as Repository
 Move Input Files to the Repository
 Do not rebuild indexes after loading

Data Files Selector

3 file(s) selected.

✓ Saada a son propre langage de requêtes: SaadaQL

- Les requêtes endossent les périmètres tracés par le modèle de donnée de Saada collection/catégorie/classe
- Les requêtes ne peuvent retourner que des OIDSaada. C'est à l'API de rechercher le contenu des objets

```
Select CATEGORY From CLASSES In COLLECTION
```

✓ 4 types de contraintes

WherePosition	Liste des positions de recherche
WhereAttributeSaada	Filtre SQL s'appliquant au données de niveau collection et aux données de niveau classe si la portée de la requête se limite à une seule classe
WhereRelation	Liste de filtres appliqués aux vecteurs formés par les liens partant de la collection sur laquelle porte le requête
WhereUCD	Filtre exprimé en utilisant les UCDs



UNe requête complexe

```
Select ENTRY From CatalogueEntry In CATALOGUE
WherePosition {
    isInCircle("M33",1, J2000, FK5)
}
WhereAttributeSaada {
    XCAT_OBSERVER = 'Isaac Newton'
and namesaada = 'NGC598' and
_obs_id > 123000
}
WhereUCD {
    [phot.flux] < 1e-14 [erg/s/cm2]
}
WhereRelation {
    matchPattern { CatSrcToArchSr
, AssObjClass{ Arch_8034AEntry }
, AssObjAttSaada{ _bj < 20 }
, Qualifier{ epic_cat_dist < 1)
}
    matchPattern {CatSrcToArchSrc
, Cardinality > 1
, AssUCD{ [src.class] = 'QSO [none]}
, AssObjAttSaada{ namesaada like '%q%' }
}
}
```

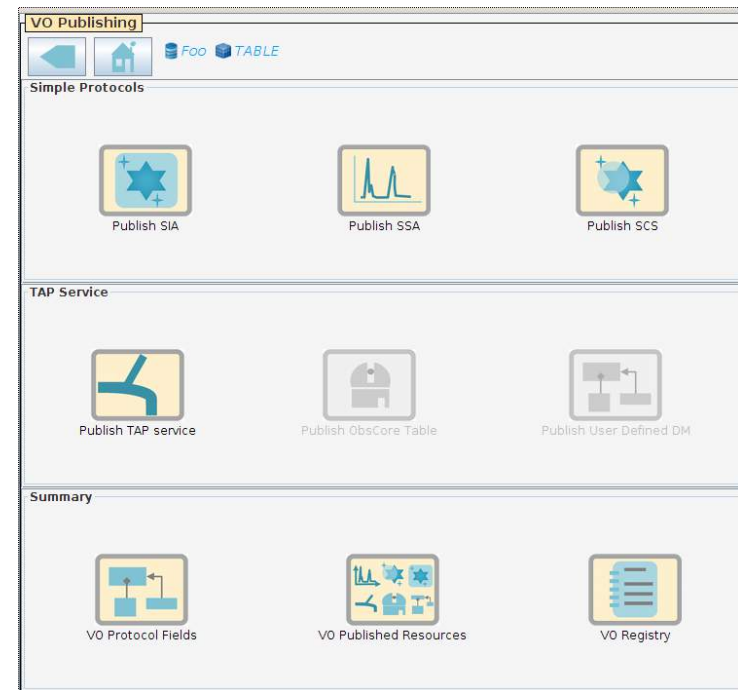
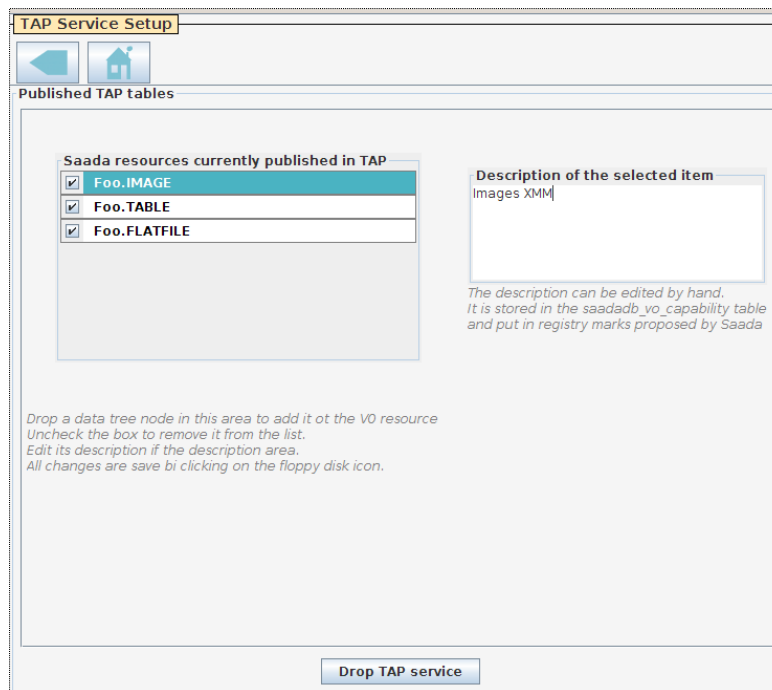


UNIVERSITÉ DE STRASBOURG



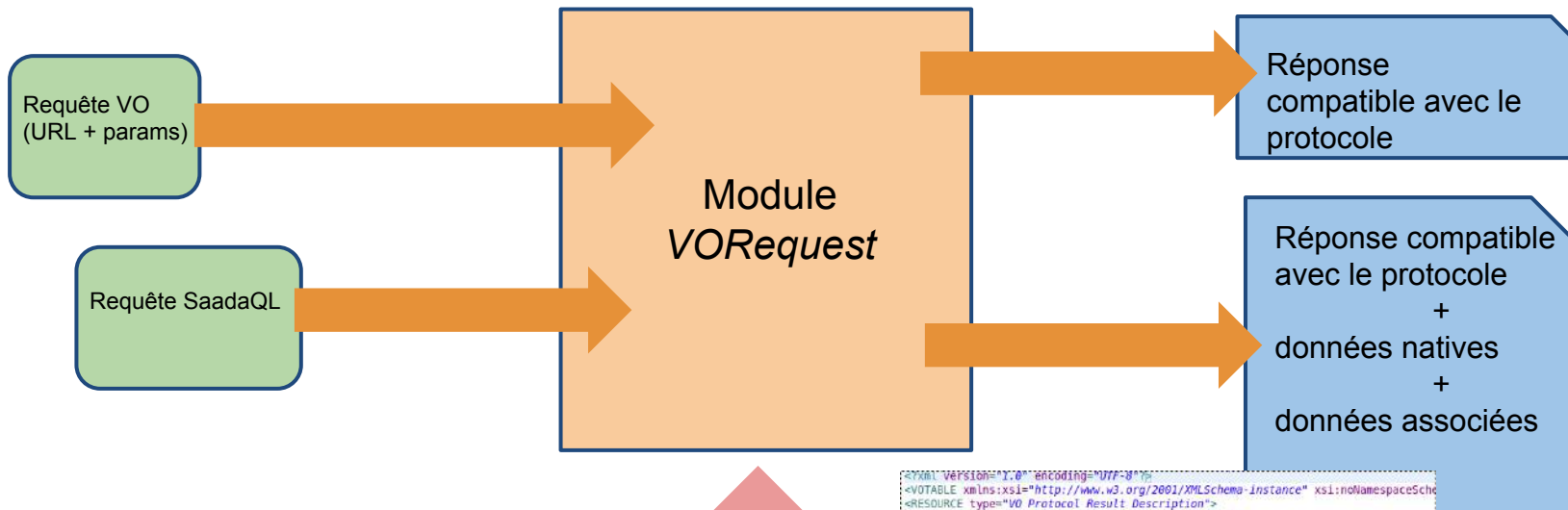
✓ Publication des ressources

- Les ressources S*P n'ont pas besoin d'être publiées pour être accédées via une requête OV. La fonction de publication de la SaadaDB permet juste de créer une description pour le registry.
- Les tables visibles par TAP doivent par contre être explicitement déclarées



✓ Flexibilité des réponses

- Le contenu des réponses S*P est contenu dans une VOTable. Il peut être adapté en éditant ce fichier.



- vodm.CS.xml 343
- vodm.ObsCore.xml 671
- vodm.SIA.xml 803
- vodm.SSA.xml 129

Modèles inclus dans la distribution

Description du modèle (VOTable)


```
<?xml version="1.0" encoding="UTF-8"?>
<VOTABLE xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="VOTable.xsd">
  <RESOURCE type="VO Protocol Result Description">
    <DESCRIPTION>Saada VO service mapping</DESCRIPTION>
    <INFO name="Cone Search default" value="OK"/>
    <TABLE name="Result">
      <PARAM name="da" value="Cone Search default" datatype="char" arraysize="*" />
      <FIELD name="group" datatype="char" arraysize="*" />
      <FIELD name="name" datatype="char" arraysize="*" />
      <FIELD name="ucd" datatype="char" arraysize="*" />
      <FIELD name="utype" datatype="char" arraysize="*" />
      <FIELD name="type" datatype="char" arraysize="*" />
      <FIELD name="arraysize" datatype="int" />
      <FIELD name="unit" datatype="char" arraysize="*" />
      <FIELD name="hidden" datatype="boolean" />
      <FIELD name="default value" datatype="char" arraysize="*" />
      <FIELD name="description" datatype="char" arraysize="*" />
      <FIELD name="requ level" datatype="char" arraysize="*" />
    <DESCRIPTION>MAN=1 REC=2 OPT=3</DESCRIPTION>
  </FIELD>
  <DATA>
    <TABLEDATA>
      <TR>
        <TD>DataID</TD>
        <TD>unique_id</TD>
      </TR>
    </TABLEDATA>
  </DATA>
</VOTABLE>
```



UNIVERSITÉ DE STRASBOURG






Saada Démo 4



TAPHandle

[Node Selector @ b](#)

SJA xcatdb>OBSLOG>ObsLogEntry>45_OBSLOG

Tap Nodes

- └ xcatdb
 - └ CATALOGUE
 - └ MERGEDCATALOGI
 - └ OBSLOG
 - └ OBSLOG.ObsLog
 - └ EPIC
 - └ ivoa
 - └ TAP_SCHEMA
- └ Goodies

Show entries
Search:

oidsaada	namesaada	md5keysaada	_recno	
580121945635291137	XTE J0421+560	0a176433f2d7fcc0ae5717c617d5983b	1	00
580121945635291138	OBSLOG-ObsLogEntry04 19 49.4+56 00 27	a65a1d1f0b388daaf9834624566ebc34	2	00
580121945635291139	HD159176	0b299b33562929d5d48166ba12fbec85	3	00
580121945635291140	HD159176	fd5c6a5db5a47def9ec1b0f8d068b28e	4	00
580121945635291141	HD159176	7ea1c8ba5a5dea2003e6481f6eb2c6e4	5	00
580121945635291142	HD47129	899eed2da2fc2374ee6e463f2017b728	6	00
580121945635291143	HD47129	4a68365ded4a31ad669bd662985c1257	7	00
580121945635291144	IRAS F00235+1024	b2afed4cc7b83ba3a13a601728830c78	8	00
580121945635291145	IRAS F12514+1027	f3c70761bd20e1baf72b602b52e0de65	9	00
580121945635291146	CFHT-PI-12	b54896155bb291b3403a85545d4ec793	10	00

Showing 1 to 10 of 100 entries ◀ Previous Next ▶



Évolution de Saada

✓ Saada au CDS

- Fin 2013, le CDS propose d'utiliser Saada pour gérer les données attachées à Vizier
 - ✓ Stockage
 - ✓ Interface Web
 - ✓ Services OV
- Version 2.0Beta attendue pour la fin de l'année (intégration en cours)

✓ Adaptations nécessaires

- Utilisation de ObsCore pour les données de niveau collection.
- Outil interactif d'aide au mapping des données de niveau collection
- Utilisation d'expression pour le mapping:

em_max	NOT SET	NOT_SET	
em_min	NOT SET	NOT_SET	
em_res_power	NOT SET	NOT_SET	
facility_name	XMM	BY_KEYWORD	kw TELESCOP detected by name
healpix_csa	9223372036854775807	BY_SAADA	
instrument_name	EPIC	BY_KEYWORD	kw INSTRUME detected by name
naxis1	648	BY_SAADA	
naxis2	648	BY_SAADA	
o_calib_status	3	BY_VALUE	Infered from detected units Infered from detected units

$$\text{em_max} = \text{RESTFREQ} + \text{CRVAL1} + (\text{NAXIS1} - \text{CRPIX1}) * \text{CDELTA1}$$



UNIVERSITÉ DE STRASBOURG



Saada Évolutions de Saada

✓ Chargement des données

- Règles de mapping incluant des expressions arithmétiques
- Pré-visualisation du mapping des données de niveau collection
- Amélioration des performances

✓ Requêtes

- Support des requêtes par régions
- Éditeur graphique de régions

✓ Interfacec OV (en cours)

- Mise à niveau des protocoles simples
- TAP Upload
- SIAP V2
- DataLink (transposition OV des relations Saada dans les VOTables réponses)

