

SAADA

Fabriquer une base
De données
Astronomique
Sans programmer

Team: Laurent MICHEL
Christian MOTCH
Ngoc Hoan NGUYEN
F. Xavier PINEAU
M. ABOUCHACRA

Funded by

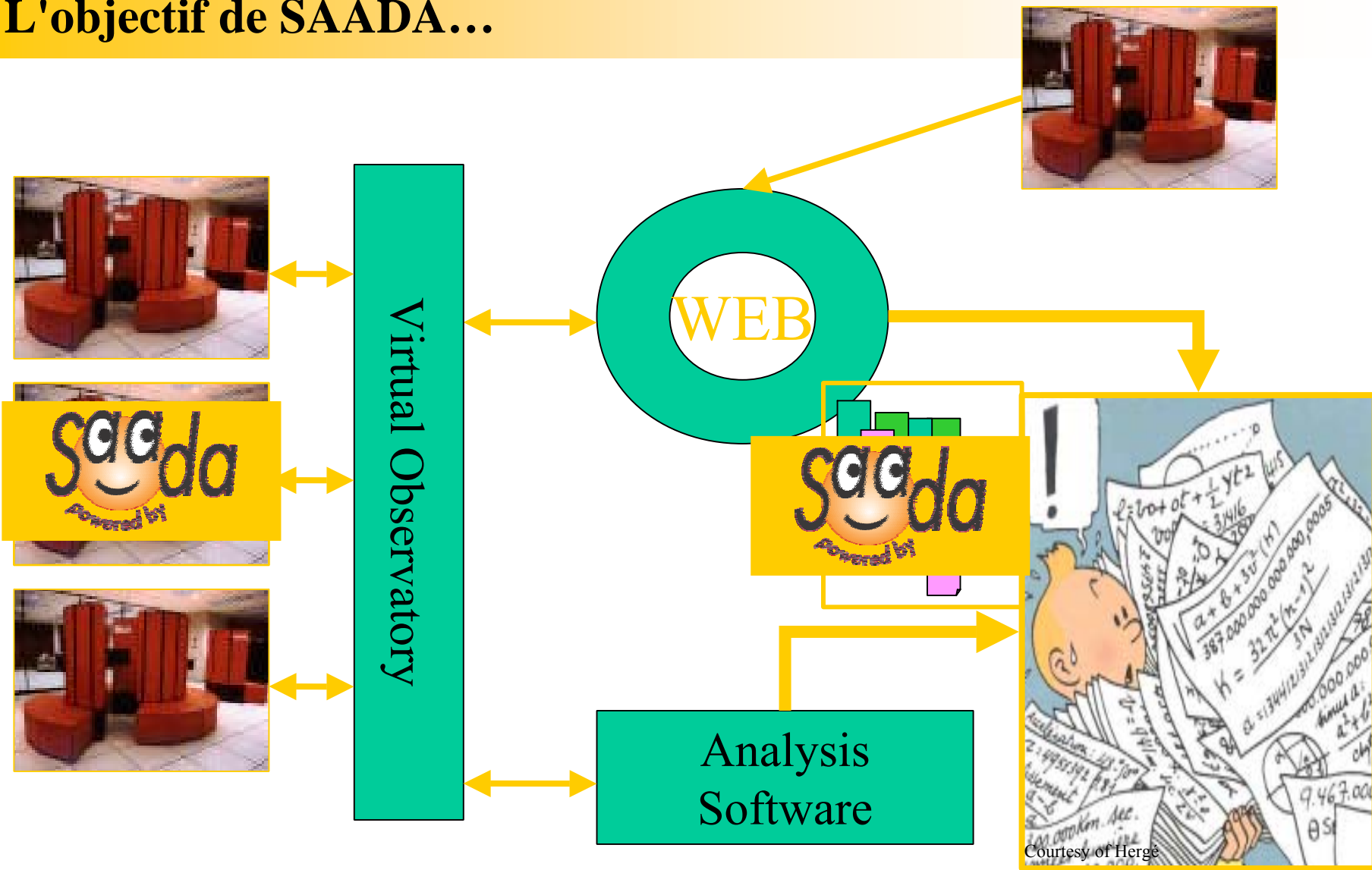


Supported by

XMM-Newton
SURVEY SCIENCE CENTRE



L'objectif de SAADA...

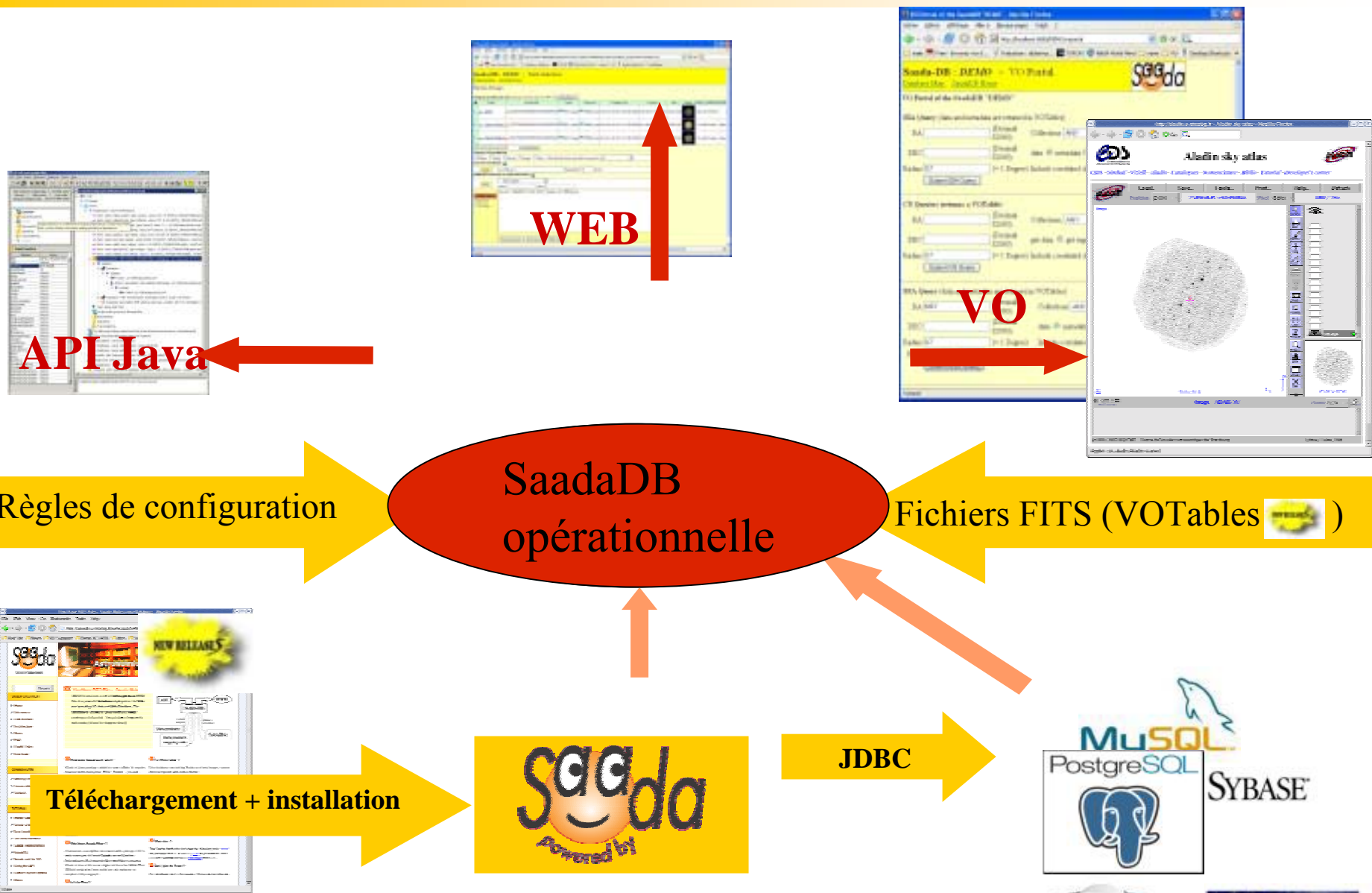


Courtesy of Herge



- **SAADA se propose de donner aux astronomes le moyen de mettre leurs données individuelles dans de vraies bases de données: les SaadaDBs.**
 - Les SaadaDBs sont conçues pour les données astronomiques
 - Les SaadaDBs sont dédiées aux astronomes
 - Les SaadaDBs sont auto-configurables – Pas de code à écrire -
- **SAADA est un outil pour faire de la Science.**
 - Des données sont archivées et sélectionnées en suivant des considérations scientifiques.
 - Les données peuvent être associées entre elles par des liens permanents.
 - Les données peuvent être manipulées par une API Java.
- **Saada est un outil pour exposer ses données.**
 - Interface Web
 - Publication dans l'Observatoire Virtuel

Principe de création d'une SaadaDB



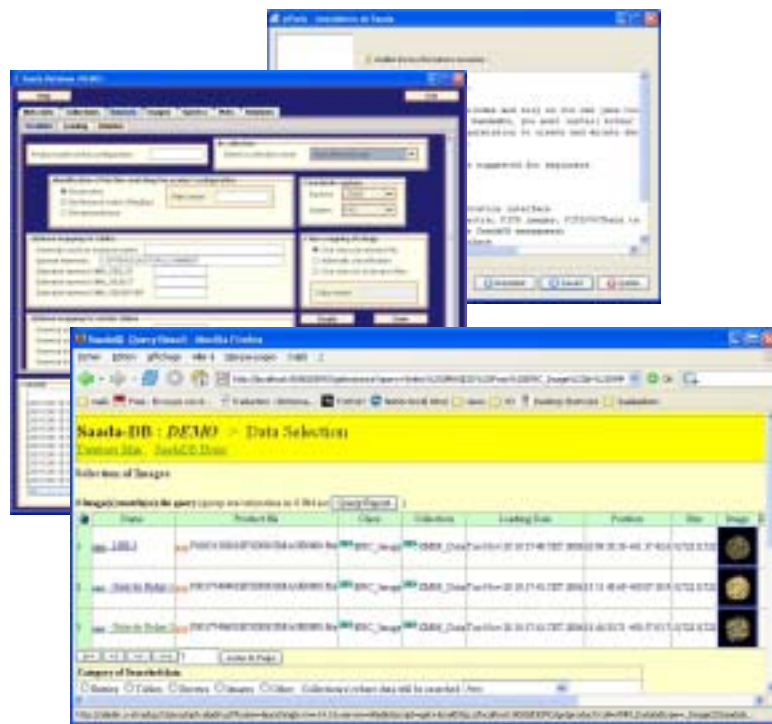
Le principe de base: Un générateur de base de données

- **Saada est un générateur de base de données**

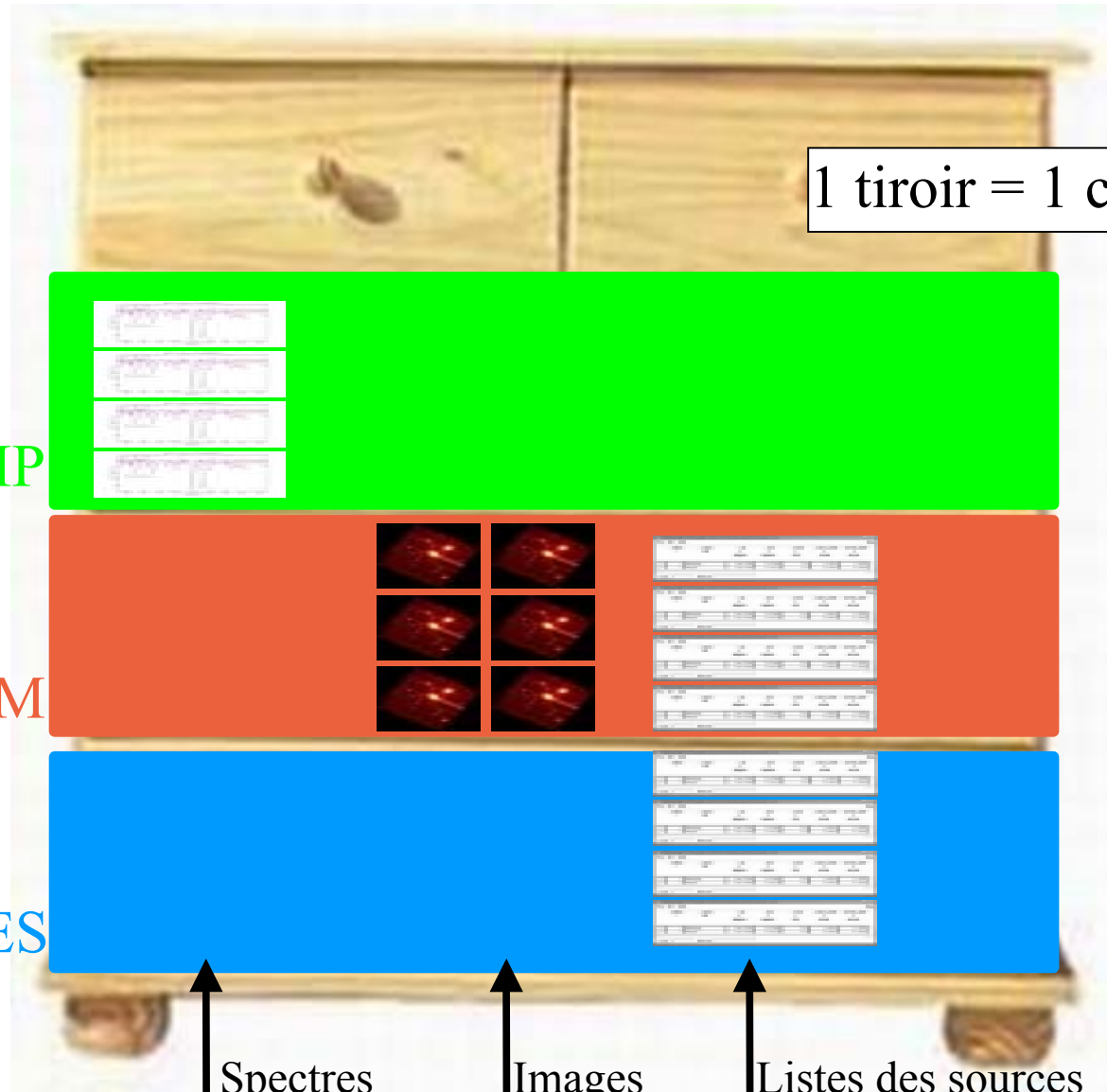
- Saada crée une base de données (une SaadaDB) vide suivant un canevas prédéfini
- Une fois la SaadaDB créée, Saada ne sert plus à rien sinon à créer une autre SaadaDB.

- **La SaadaDB est autonome**

- Elle est installée localement
- Elle possède ses propres outils.
- Elle possède ses propres bibliothèques.
- Elle possède sa propre base SQL
- Elle possède sa propre URL d'accès



Organiser la SaadaDB suivant des critères scientifiques



1 tiroir = 1 collection

OHP

XMM

ARCHIVES

↑ Spectres

↑ Images

↑ Listes des sources

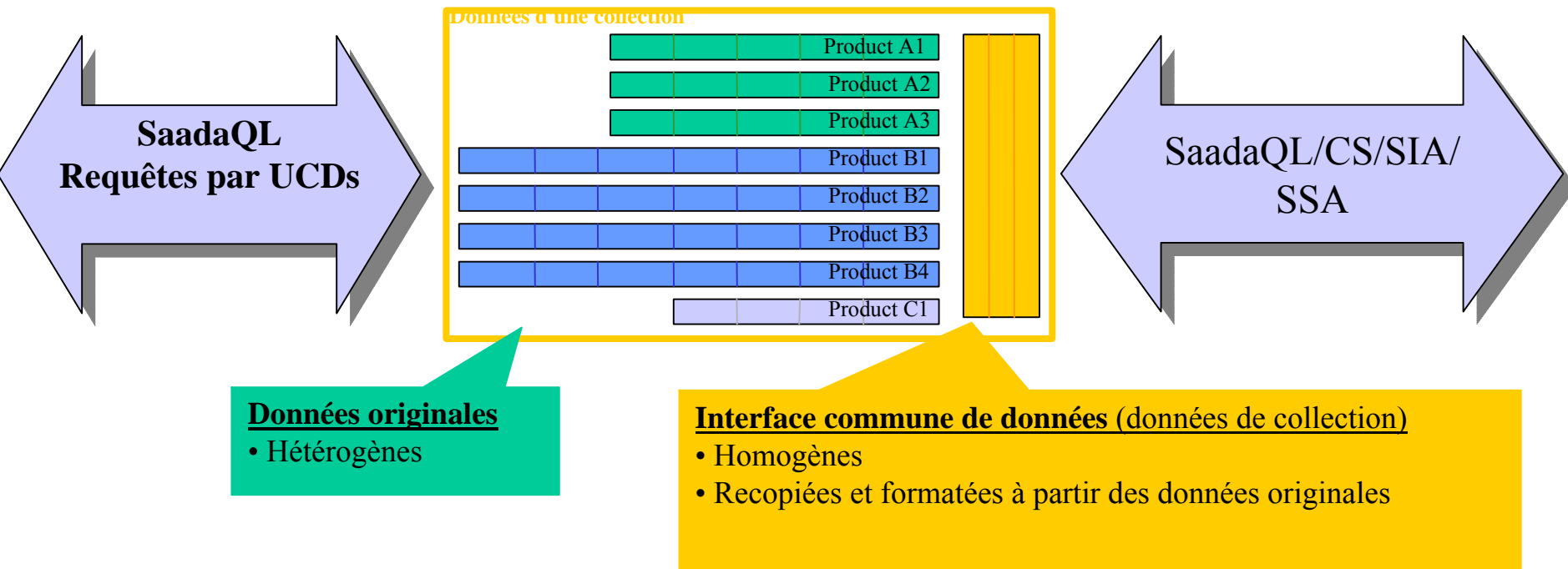


Sélectionner des données hétérogènes dans une collection

Les données présentent deux interfaces pour les requêtes

1. Une **Interface commune** pour les sélections simples
2. Les **données natives** pour les requêtes plus sélectives

Les deux interfaces peuvent être utilisées dans la même requête



Les relations dans Saada

OBSERVED

spectrum

entry

images

Link qualified with δv

$$\delta v = \sqrt{(\delta v_x^2 + \delta v_y^2 + \delta v_z^2)}$$

SIMULATED

entries

Image

Link qualified with pixel position

Pix_X = ... Pix_Y = ...

Example of a constraint on a correlation pattern:

```
matchPattern{« ObsEntrImgCorr »
```

```
, Cardinality(« > », 0, 0)
```

```
, Qualifier(« Pix_X », « <> », 123, 208)
```

```
, Qualifier(« Pix_Y », « <> », 67, 87)
```

```
, AssObjClass(« OHP_Image »)
```

```
, AssObjAtt(« name = '*K12' »)
```

```
}
```

Relation name

Constraint on the cardinality

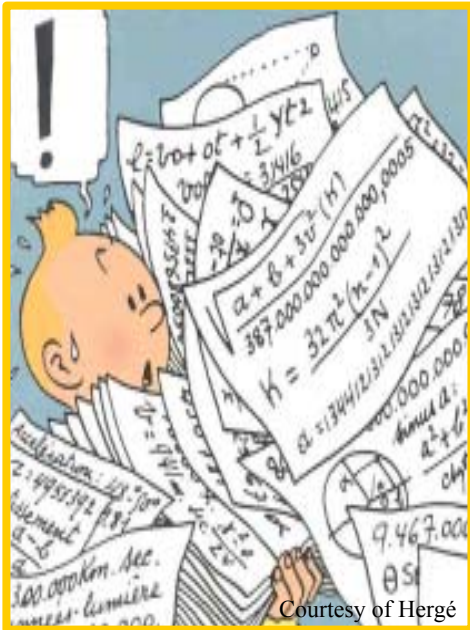
Constraints on qualifiers

Constraint on class of counterparts

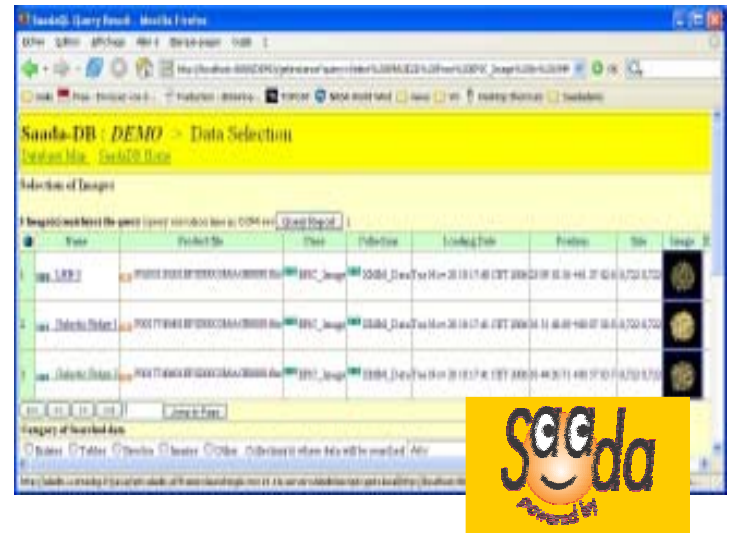
Constraint on counterparts attributes

Cas d'utilisation de Saada

- **Création d'une archive locale**
 - Possibilité d'accéder à ses données
 - par requête
 - Par les outils de l'OV



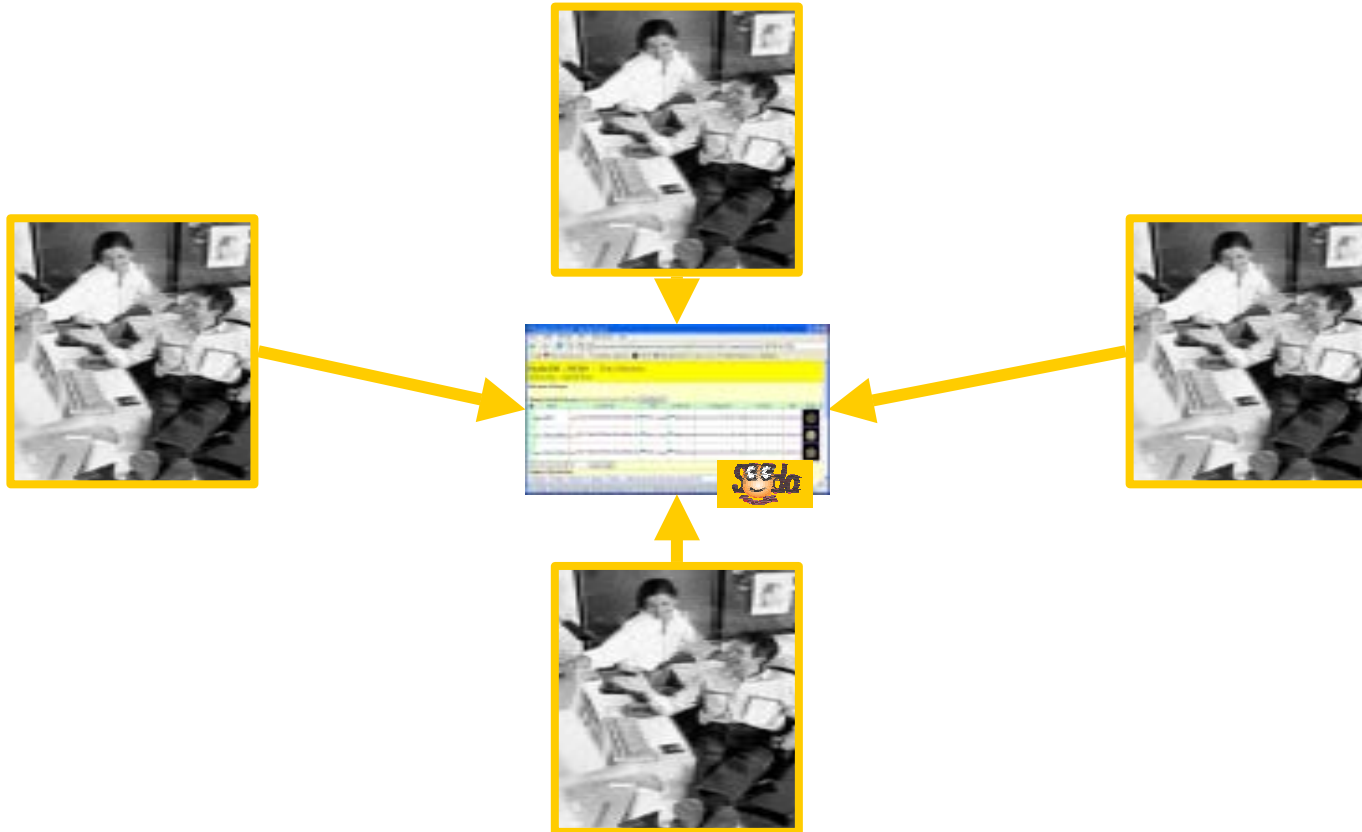
Courtesy of Hergé



Cas d'utilisation

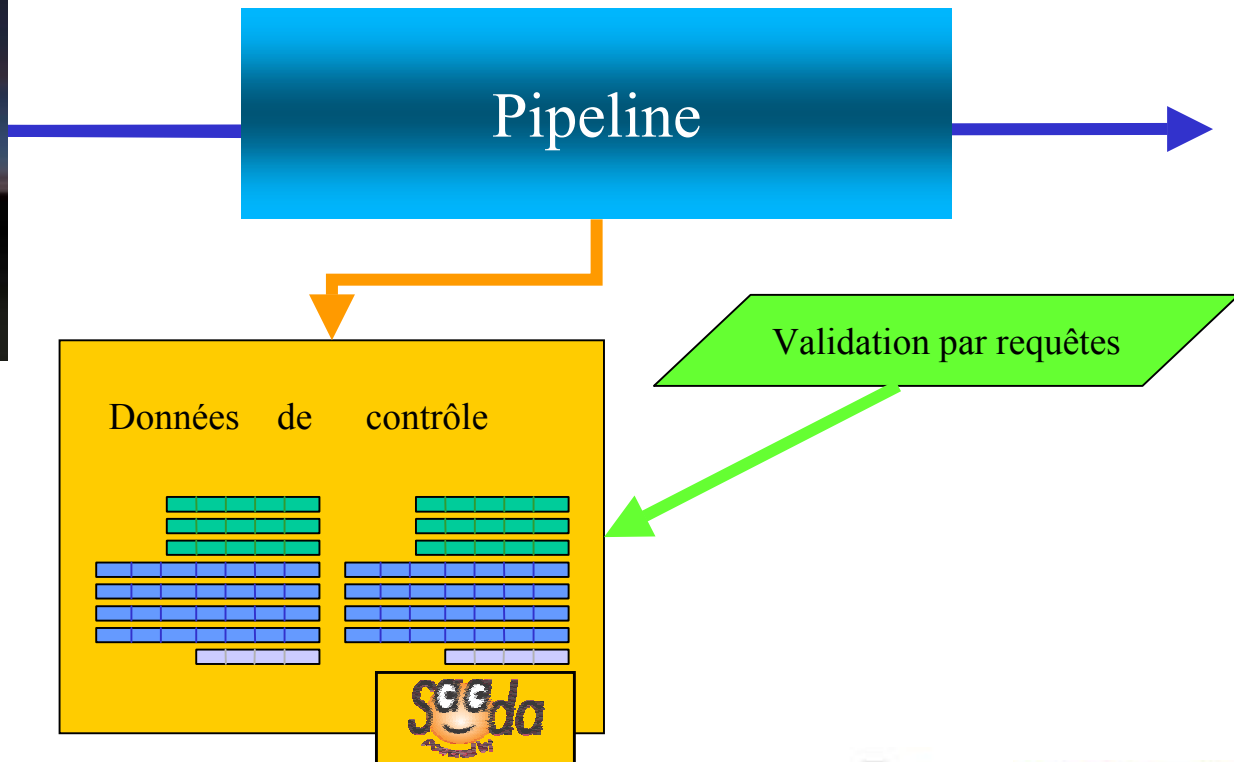
- **Partage de données**

- Mises à jour à distance par des collaborateurs
- Distribution vers la communauté



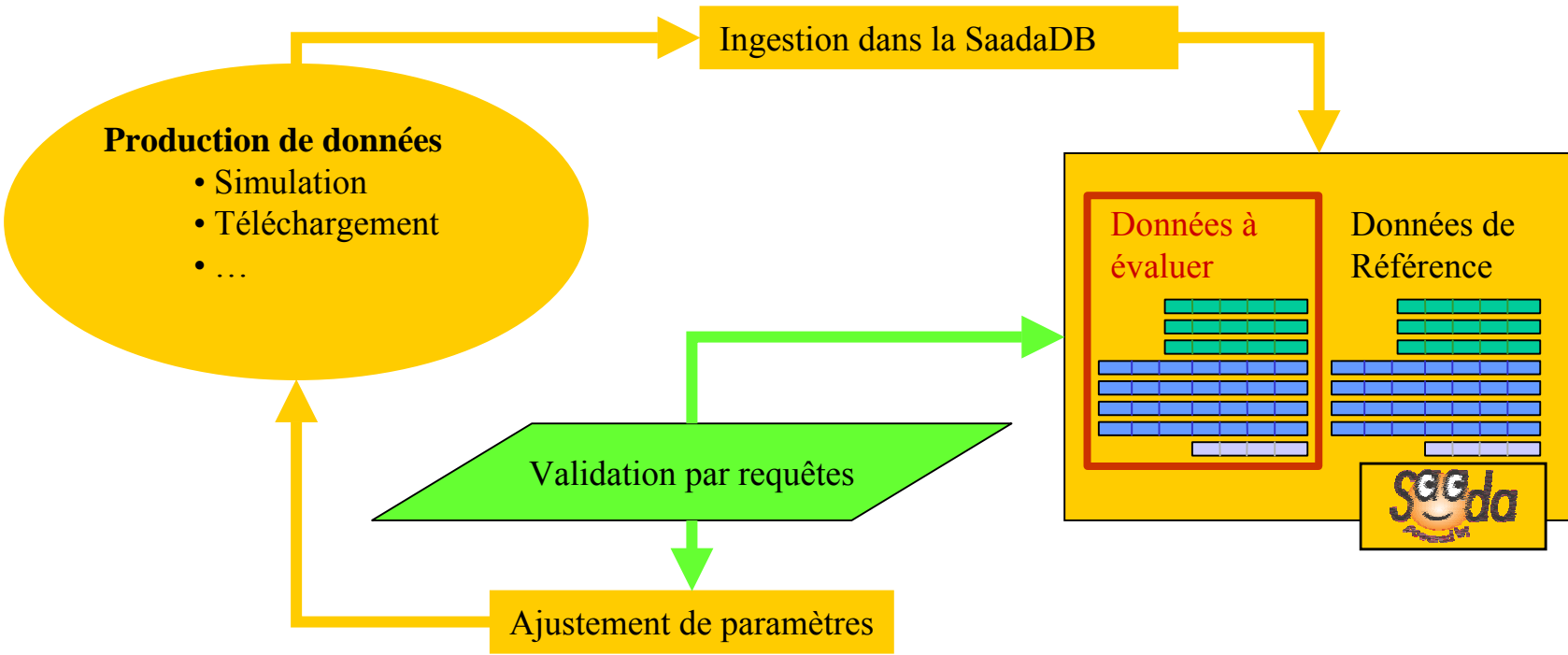
Cas d'utilisation

- **Création d'une archive en ligne**
 - Contrôle en ligne des produits
 - Vérification par requêtes



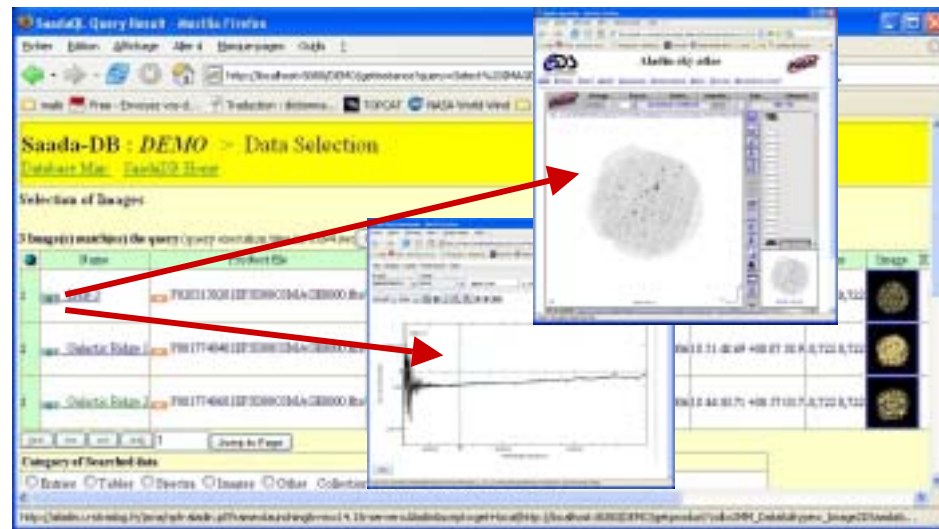
- **Validation de données**

- Utilisation de la capacité de Saada à héberger des données hétérogènes valider les données produites avec des données de référence.



- **Accès direct à des portails OV**

- Aladin, Specview
- Traitement de des données locales dans l'OV sans publication



- **Publication dans l'OV.**

- SIA
- Cone Search
- SSA
- Skynode (tests en cours)



- Utilisation des UCDs dans les requêtes

The screenshot displays the Saada web interface. On the left, there is a table titled "2579 Entry(s) match(es) the query (query execution time)". The table has columns for "Data Access", "Created RA DEC (J2000)", and "Created Dec RA DEC (J2000)". The table contains 25 rows of data. On the right, a "Query report" window is open, showing the SQL query used for the search. The query is: `select ENTRY from * in ARCH_CAT whereUCD 1 [phot.flux.density:em.radio.750-1500Hz] > 1 [mJy]`. A red oval highlights the query text in the report. Below the table, there is a "Search by coordinate" section with various search criteria. A red oval highlights the search criteria in this section.

Plan d'exécution

Requête utilisant les UCDs

Le status actuel de Saada


- **La version 1.4.1 beta est disponible pour le téléchargement**
 - <http://amwdb.u-strasbg.fr/websaada>
 - **Amélioration du déploiement**
 - Refactoring du data loader
 - Lecture des fichiers FITS à simple extension
 - Lecture des VOTables
 - Toutes les commandes utilisables soit en mode graphique soit en mode script (scripts *ant*)
 - Interface graphique d'administration
 - Interface WEB
 - Installeur (requiert Java 1.5)
 - Serveurs OV



Les axes de développements

- **Amélioration du data-loader**
 - Supporter une plus large palette de formats
 - FITS multi extensions
 - TSV
 - Autres
 - Pousser plus avant l'automatisation du chargement
 - Simplification de la configuration
 - Amélioration de la gestion des erreurs
 - Banc de test automatisé
- **Amélioration de l'interface graphique**
 - Afficher les paramètres contraints
- **Interface OV**
 - Améliorer l'implémentation des protocoles
 - Skynode
 - Associations dans les VO tables

Le développement/support

L. Michel	Chef de projet (temps partiel) Développement, Support
F.X. Pineau	Doctorant (BDI) Connexion de Saada avec des outils scientifiques
Vacataire Stagiaires	Tâches ponctuelles Gestion des exceptions => site Web
Autres	Tous les contributeurs sont les bienvenus <ul style="list-style-type: none">• Fourniture d'échantillons de données• Développement (une fois le déploiement sur SourceForge terminé. 

- **ADASS 2005**
 - Déploiement d'un SIA retournant des images avec des spectres et des sources associées.
- **IVOA 2006 (Victoria)**
 - Utilisation des UCDs dans les requêtes
- **ADASS 2006**
 - Utilisation les relations pour la recherche *d'outlayers* (données de la seconde release du catalogue XMM)



Merci



Les origines du projet

- **SSC de XMM-Newton**
 - Traitement systématique et automatique des données
 - Corrélations avec 180 catalogues d'archives
- **Besoin d'une base pour le SSC**
 - Héberger des données hétérogènes
 - **Implémenter les liens de corrélations** de manière à en favoriser l'exploitation scientifique
- **1997: Choix d'une base à objets: O2**
 - **Base de données, corrélateur, contrôle du pipeline**
- **2000: Mort de O2**
 - Mais le support continue
- **2002: Opportunité de financement d'une bourse de thèse CNES**
 - Ne pas faire du ad hoc pour remplacer O2 dans le SSC
 - Offrir à la communauté un système offrant les **caractéristiques intéressantes des bases à objets** (données hétérogènes, modèle de données interconnecté) et adapté à une large variété de produits
 - Ignorer les autres caractéristiques du modèle à objets (schéma entièrement défini pour chaque base)