



Managing and analyzing analytical chemistry data sets

Sana Tfaili



Sana Tfaili¹, Karima Rafes², Valérie Cantonny³, Diem Bui Thi¹, Ali Tfayli¹, Cécile Germain^{4,5} and Pierre Chaminade¹

¹Lip(Sys)²- Chimie Analytique Pharmaceutique^s, Univ. Paris-Sud, Université Paris-Saclay, F-92290 Châtenay-Malabry, France.
²FKA EA4041 Groupe de Chimie Analytique de Paris-Sud

³Bordercloud 245, quai de la Bataille de Stalingrad, F-92130 Issy-les-Moulineaux, France.

³IPS², CNRS, UMR 9213, Sciences du végétal : du gène à l'écosystème

⁴LRI- Laboratoire de Recherche en Informatique, Bât 650, Rue Noetzlin, F-91190 Gif-sur-Yvette, France.

⁵LAL- Laboratoire de l'Accélérateur Linéaire, Bât 200, voie de la faculté, F-91440 Bures-sur-Yvette, France.

- 1. Lip(Sys)² data sets.**
- 2. Participation to CDS 1.0: objective, accomplishments and obstacles.**
- 3. What are we seeking for in CDS 2.0?**
- 4. Importance and issues for project continuity.**

1. Lip(Sys)² data sets

Different Analytical measurement techniques

Data processing

Domain scientists

□ Separation techniques,
mainly chromatographic

□ Data (pre)-
processing

□ Data providers

□ Coupled mass
spectrometry techniques
(LC, GC, GCxGC/MS)

□ Multivariate analysis

□ Chemometric
techniques

□ Lipid analysis

□ Vibrational
spectroscopies (infrared,
near infrared and Raman)

2. Participation to CDS 1.0: objective, accomplishments and obstacles.

Database creation and management



Improving data analysis: computing power (Cloud)

ERM CLOUD@VIRTUALDATA

Management of Data for the Analysis of Lipids, Metabolites and Isotopes 'MODALMT'

- Ontology, metadata, common format for data fusion (ex.:aia, .cdf)**
- Automate data conversion and pre-treatment, open access**

✓ 1 year engineer (plugins for data conversion)

✓ Support from

BorderCloud (Linked data technologies)

Building DAAP Platform

✓ Working on a user friendly interface (VRAC)

✓ 40 Tb storage (ERM, AAP attractivité)

✓ Ramp on drug cancer for drug identification and quantification

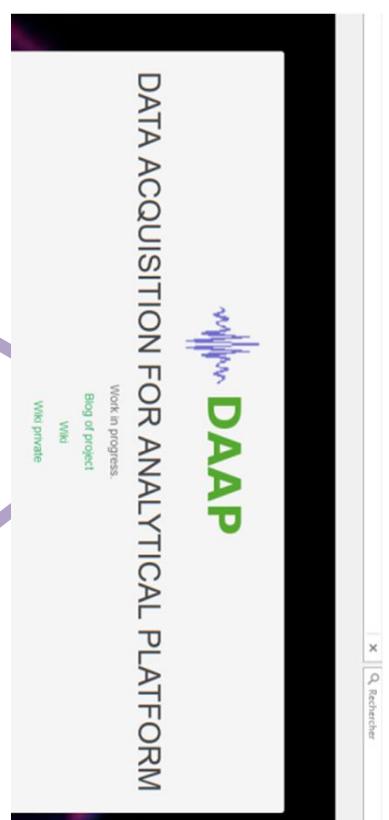
DAAP: DATA ACQUISITION FOR ANALYTICAL PLATFORM



✓ Automating scientific workflows and building an open database platform for chemical analysis metadata

Defining DAAP domain

<http://daap.eu/>



Public Wiki Private Wiki

Workflow steps

Step 1: Ontology defining and Data description in WikiDAAP pages.

The screenshot shows three main sections of the WikiDAAP platform:

- Main Page:** Displays a search bar, a sidebar with links like "DAAP", "Main Page", "Recent changes", and "Log in", and a list of recent edits.
- Connection:** A form for logging in with fields for "User name" and "Password". A red arrow points to the "Sign up" button below it.
- RamanEvolution:** A detailed page for a "Device" named "RamanEvolution". It includes a large image of the device, a green header, and several tabs: "Pilot process", "Keywords", "Organisation", "Latitude", and "Longitude". A red arrow points to the "Infobox" section at the bottom right.

Step 2: Data hosting. **ONTOLOGY Hosting** via SPARQL access point within Paris Sud University

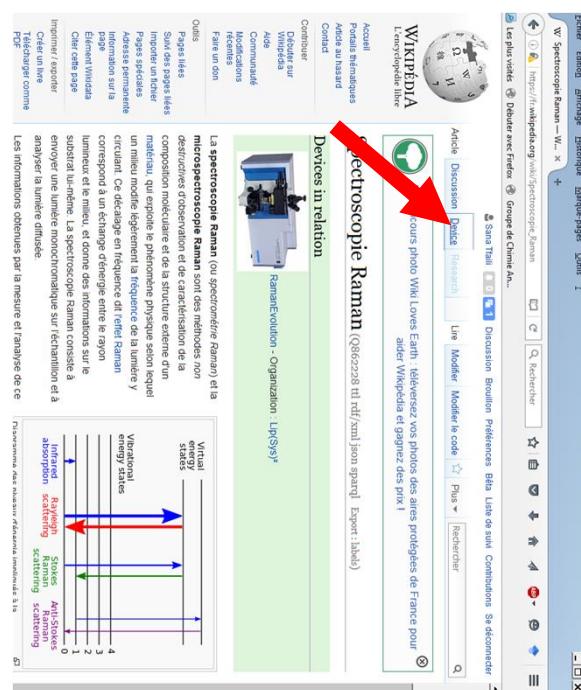


The infobox in the wiki compares information between the formal ontology defined with TopBraid Composer and the data in the Infobox.

- IRI were used to describe the IDs of all concepts.
- The IRI links on wiki public pages.
- RDF files that describe the ontology hosted in:
<http://daap.eu/ontology/2015>

Workflow steps

Step 3: linking the ontology to linked data technologies.



A screenshot of a web-based data science project page titled "Paris Saclay". The page features a large map of the Paris-Saclay area, overlaid with various data layers such as roads, green spaces, and buildings. To the left of the map, there is a sidebar with several tabs: "About" (selected), "Data", "Model", "Predict", "Report", and "Help". Below the tabs, there are sections for "Project Overview" and "Data Sources". The "Data Sources" section lists "Paris Saclay" and "OpenStreetMap" with small icons. The overall interface is clean and modern, designed for data exploration and analysis.

Perform queries

Definition the ontology of available instrument of measurements. Wikipedia keywords permitted to link each measurement technique by using the same keyword in the infobox of wikiDAAP.

Virtual Research Environment for Analytical Chemistry

VRAC

Main Window

DAAP endpoint: <https://opendata1.opendata.u-psud.fr/sparql>

Lab endpoint: ???

Charger les paramètres par défaut Sauver ces paramètres

Organisation

- Lip(Sys)² ▾ Rafræichii
- Büchi NirFlex N500 ▾ Rafræichii
- Phospholipids analysis by Raman vibrational spectroscopy ▾ Rafræichii
- Protocol for the characterization of phospholipid classes by ▾ Rafræichii
- Campagne test Campaign 1 ▾ Rafræichii
- Destination des fichiers (i) file:///filer.ups.u-psud.fr/archivagesciences/Chimie%20Analytique/projets/f... ▾ Rafræichii
- Destination des fichiers (u) Z:\Chimie Analytique\projets\PLs_standards_Raman\Protoc... Ouvrir

Liste des fichiers

Ajouter des fichiers Effacer

Archivage

archivagesciences\$ (\\filer.ups.u-psud.fr) (Z:)

Organiser ▾ Inclure dans la bibliothèque ▾ Graver Nouveau dossier

Favoris Bureau Dropbox Emplacements récents Téléchargements

Bibliothèques

Chimie Analytique Dossier de fichiers

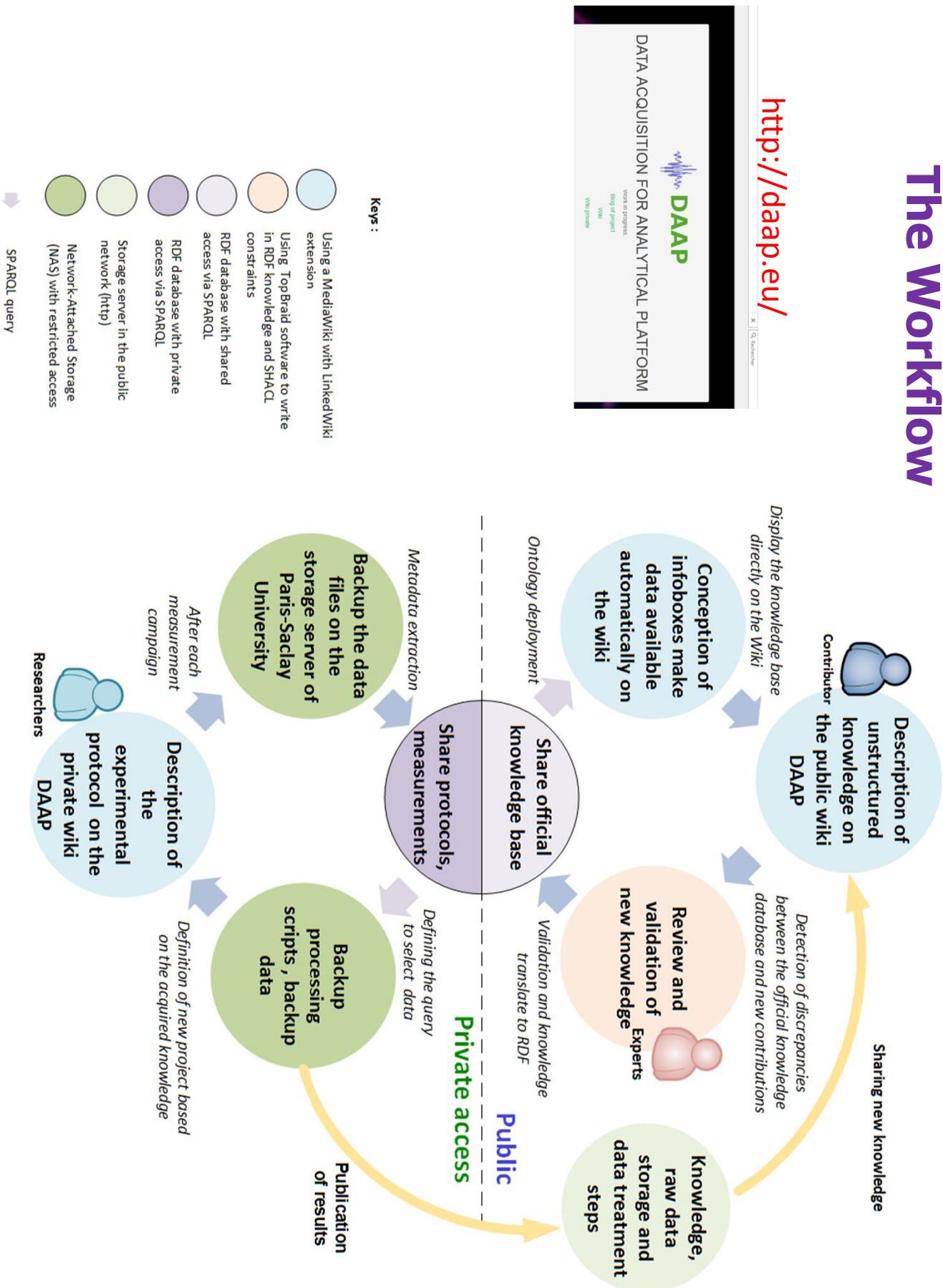
Campagne001

Organizer ▾ Indire dans la bibliothèque ▾ Synchroniser ▾ Graver Traveller hors connexion Nouveau dossier Rechercher dans : Campagne001

Téléchargements Emplacements récents Bibliothèques Documents Images Musique

Dossier de fichiers artefacts Dossier de fichiers Dossier de fichiers Dossier de fichiers script scriptR

The Workflow



CDS 2.0 project

3. What are we seeking for in CDS 2.0?

- Enrich our database, define the ontology for the projects and data sets
- Optimize some plugin for data conversion
- user friendly software VRAC
- Analyze conjointly data sets of various origins
- new kinds of computational analyses

Data scientists □
support.

Training (ontology, SPARQL queries and Wikimedia).

CDS 2.0 project

3. What are we seeking for in CDS 2.0?

- Qualitative identification and quantitative prediction (*a RAMP was organized around cancer drugs, Laetitia Le*)
- Correction of the response of two universal detectors used in chromatography techniques (*peak estimation and modeling*)
- Data fusion between mass spectra obtained in positive and negative ionization mode
- Analysis of spectral images obtained by vibrational spectroscopy (*Raman and Infrared spectroscopy*)

CDS 2.0 project

4. Importance and issues for project continuity

- *Data providers, Open access data sets (correctly described)*
- *Workflow is transposable to other research structures and could be interesting for industries in pharmaceutical domain.*
- *get external funding by combining in the same project between the defined workflow and health-related research topics*

Thank you!