

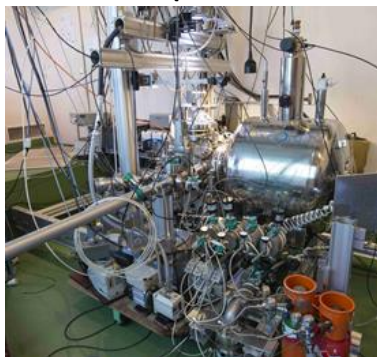
Groupe métier GT2I

Bases de données pour Nanocosmos

Odile Cœur-Joly

Données issues de 2 manip sur 2 sites différents

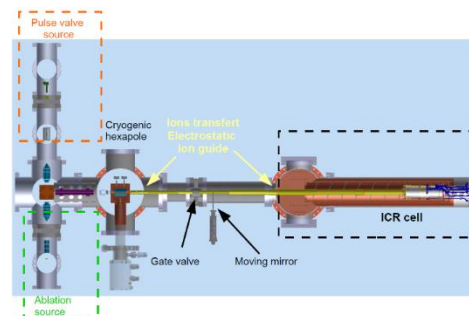
PIRENEA (FT-ICR-MS)



@IRAP



PIRENEA-2 (FT-ICR-MS ++)



@LCAR

Volumétrie, format, organisation

- Environ 20000 spectres par an et par site
- 10 à 15 ans d'exploitation
- Format = couple (binaire, XML) = (data, params)
- Même format pour 2 sites
- data = environ 100 Go par an et par site
- XML = environ 50 Mo par an et par site
- Organisation sur disque par année et date du jour
- Unicité des noms (data ou XML)

Souhaits

- Si BD : pas d'ingestion manuelle des données
- Sélection, tri, récupération sans cahier de manip
- 10 utilisateurs (@LCAR et @IRAP)
- Requêtes sur des données étalées dans le temps (10-15 ans)
- Suppression des données inutiles
- Accès via son poste de travail (Mac, PC)
- Après sélection : traitement et visu ou plus (TBD)
- Compatibilité avec autres bases de données (TBD)



Questions pour le GT2I (1/2)

- Faut-il une BD ou pas ? Une ou plusieurs BD ? Une par site ?
- Schéma des BD : conception ? outils nécessaires ? Gestion de version des schémas ? Évolution/maintenance si le schéma de la BD change ?
- Choix d'un SGBD : qu'est-ce qui est déterminant ? Volumétrie ? Populaire ? Simple ? (SQLite, NoSQL, MySQL, PostgreSQL, MariaDB...)
- Type des tables encore d'actualité ?(Mylsam, InnoDB ?)
- Ingestion automatique lors de l'acquisition ? Construction « après coup » d'une BD à partir de l'arborescence disque des fichiers XML ?

Questions pour le GT2I (2/2)

- Partage des données, des BD entre sites ? (dossiers partagés ? serveur du SI ? Cloud ?)
- BD multi-sites, multi-utilisateurs : quelle architecture ? quel langage ?
- Accès réseau ? Sécurité ? Facile ?
- Requêtes, sélection de données, récupération, traitement, visu : REX du GT2I ?
- Quels outils pour le développement : Django, Flask, ... ? REX du GT2I ?
- Mutualisation possible des dév. ou trop spécifique ?