

Journées Démocrite
08 – 09 avril 2013
CPPM

Recensement de la production scientifique du CPPM

R. Le Gac
CPPM, CNRS/IN2P3

Avec le soutien de D. Cristofol, M. Damoiseaux, P.-Y. Duval et T. Mouthuy

Plan

- ▶ Introduction
- ▶ L'outil développé au CPPM
- ▶ Conclusions
- ▶ Démonstration

Introduction

Les besoins

- ▶ Pour les évaluations du laboratoire et de son personnel, nous devons :
 - Établir des listes de publications ordonnées et formatées
 - Compter les publications par catégorie
- ▶ Le classement et le comptage dépendent du contexte :
 - par catégorie
 - par projet et par équipe
 - par thématique scientifique et par année.
- ▶ L'ordonnancement des listes et métriques reflète l'organisation du laboratoire.
- ▶ Produire ces listes et métriques est devenu de plus en plus difficile au fil des années.

L'existant : les entrepôts

- *cds.cern.ch* et *inspirehep.net* mais aussi *ADS*, *HAL*, ...
 - Stockage numérique de documents
 - Organisé en collections
 - Moteur de recherche
 - Extraction de documents
 - Extraction de méta-données dans différent formats
- ▶ Les entrepôts sont bien adapté pour la gestion d'un fonds documentaire et pour la recherche de documents.
 - ▶ Ils permettent d'établir une liste de publications, au sein d'une collection, pour un individu, pour une collaboration.
 - ▶ Ne sont pas adapté pour établir des listes ordonnées et formatées pour un laboratoire.

L'outil développé au CPPM

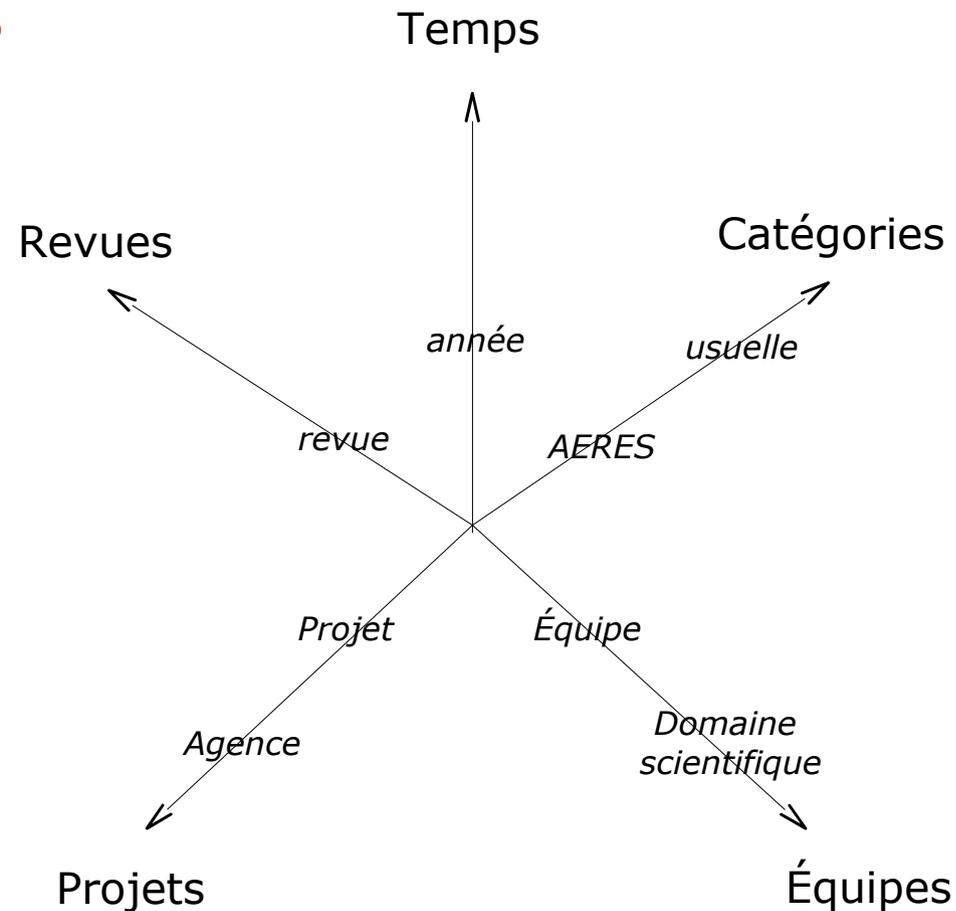
Généralités

- ▶ Mise en place d'un outil dédié pour :
 - Recenser les publications du laboratoire
 - Extraire des listes et des métriques (unité, équipe, individu)
 - Simplifier au maximum les phases de saisie et de validation
 - Indépendant par rapport aux systèmes d'exploitation (linux, window, ...) et aux différentes bases de données (sqlite, mysql, ...)
 - Paramétrable de A à Z par un(e) documentaliste
- ▶ Stocke uniquement les méta-données

Chronologie

- ▶ A l'initiative de la direction du laboratoire, mise en place d'un groupe de travail en octobre 2010, mené par M. Damoiseaux.
- ▶ Cahier des charges rédigé en février 2011.
- ▶ En novembre 2012, un prototype est opérationnel. Validé par D. Cristofol, il reproduit la liste EAOM 2012.
- ▶ En décembre 2012, malgré des ressources très limitées, nous décidons de continuer l'expérimentation pour la période 2009–13.
- ▶ En mars 2013, mise en place des correspondants dans chaque équipe et organisation de la collecte 2012–13.

Les cinq axes



- ▶ La classification des données et leurs comptage se font par rapport à ces axes
- ▶ L'organigramme du laboratoire est projeté sur les axes *équipes* et *projets*

Moissonneurs

- Automate logiciel paramétrable
 - Recherche des documents dans un entrepôt
 - Extrait les méta-données dans un format déterminé et les décode
 - Enregistre les méta-données, dans notre base de données
-
- ▶ Ingrédient logiciel essentiel pour simplifier la saisie
 - ▶ La version actuelle :
 - moissonne les entrepôts CDS et INSPIRE
(basé sur le logiciel *invenio* développé au CERN)
 - décode les méta-données au format MARC
 - Paramétré pour trouver les articles et les actes de conférences pour Antares, Atlas, BOSS, D0, H1, KM3NeT, LHCb, SNLS

Les rôles

▶ Commun des mortels

- Extraction de listes et de métriques prédéfinies

▶ Correspondants d'une équipe

- Modifie le contenu de la table des publications
- Fait tourner les moissonneurs en fonction des besoins
- Aide à la paramétrisation des moissonneurs
- Interlocuteur(trice) privilégié(e) de notre documentaliste

▶ Documentaliste

- A tous les droits
- Garant de la cohérence de la base
- Gestion du paramétrage des listes, métriques et moissonneurs
- Responsable de l'extraction des listes pour les évaluations du laboratoire
- Valide le contenu de la base après consultations du laboratoire

Les difficultés

- ▶ **Prototype**
 - Déploiement incrémental des fonctionnalités
 - Parfois les interfaces utilisateur sont rudimentaires
- ▶ **La liste complète des auteurs et la liste des auteurs du CPPM**
- ▶ **Cohérence et doublons ?**
- ▶ **Relation avec HAL**
 - Base de données plus complète que HAL car très proche des équipes
 - Double saisie
 - Dans un monde idéal, HAL devrait tirer des données de notre base.
- ▶ **Absence de moyen dédié pour le suivi et le développement**

Le code

- ▶ Développement *open source*, écrit en python et qui s'appuie sur les logiciels :
 - Ext JS
 - MathJax
 - plugin_dbui
 - Python
 - web2py

Progiciel encapsulé dans une machine virtuelle (web appliance).

- ▶ Cet outil pourrait être déployé en dupliquant la machine virtuelle du CPPM ou en tant que service web.

Cependant, un déploiement nécessite d'être organisé et des soutiens pour accompagner les différents types d'utilisateurs.

Conclusions

- ▶ Un prototype fonctionnel mais perfectible :
 - Efficace dans la collection de données à partir des entrepôts
 - Grande flexibilité dans les listes et le comptage des publications
 - Possibilité d'analyse à différentes échelles
 - Déjà plus de 850 publications enregistrées
- ▶ Un outil « hors des sentiers battus » qui devrait nous aider dans notre quotidien.
- ▶ Les rôles du documentaliste et des correspondants sont essentiels pour être efficaces dans les opérations de saisie et de validation.
- ▶ Cette approche nécessite une relation étroite avec HAL.
- ▶ Affaire à suivre ...

Démonstration

- ▶ Listes et métriques
- ▶ Table des publications et son filtre
- ▶ Moissonner
- ▶ Les différents assistants
- ▶ La procédure de validation
- ▶ Gérer les droits des utilisateurs