INRA
SCIENCE & IMPACT

BIBS
ANALYSES MULTI-ECHELLES DE BIOPOLYMERES

Mise en place de l'ECM Alfresco dans une unité de recherche INRA

Retour d'expérience

Stéphane BANSARD / JoSy GED

Qui je suis

IE département CEPIA / Centre Angers-Nantes

Unité BIA (Biopolymères Interactions Assemblages)

Plate-forme instrumentale BIBS / bioinformatique depuis 2013

Responsable de la gestion des “données” de la plate-forme

“Intégrateur” d’une solution Alfresco depuis 2007

Origine du projet - Contexte

- Initialement (2005), il s'agissait d'un projet de capitalisation de données scientifiques orienté métadonnées.
  - → stockage ? métadonnées avec/sans données/fichiers (BDD, Web Sémantique ?)
  - → découverte de différents outils dont DMS (=GED) vs CMS vs Librairie Numérique vs ECM…
  - → contraintes sur l'effort de saisie, interfaçage sauvegarde/archivage

- Le périmètre a évolué :
  - Suite à un audit, gestion des documents qualité (corpus de test).
  - Gestion des données scientifiques (mais mise en place de certification)
  - Remplacement de serveurs (dossiers partagés → GED)

Choix des outils

2007

- Dans quoi s'embarque-t-on ? Pour combien de temps ? Volumétrie ? Scalabilité ?

- outils INRA ? outils pérennes ?
- outils Open Source ? Nuxeo en cours de réécriture.
- Changement licence Alfresco-2.0 → GPL
  - Tests → satisfaction
- La DSI INRA s'oriente vers Silverpeas comme outil collaboratif.

Type de projet / Ressources

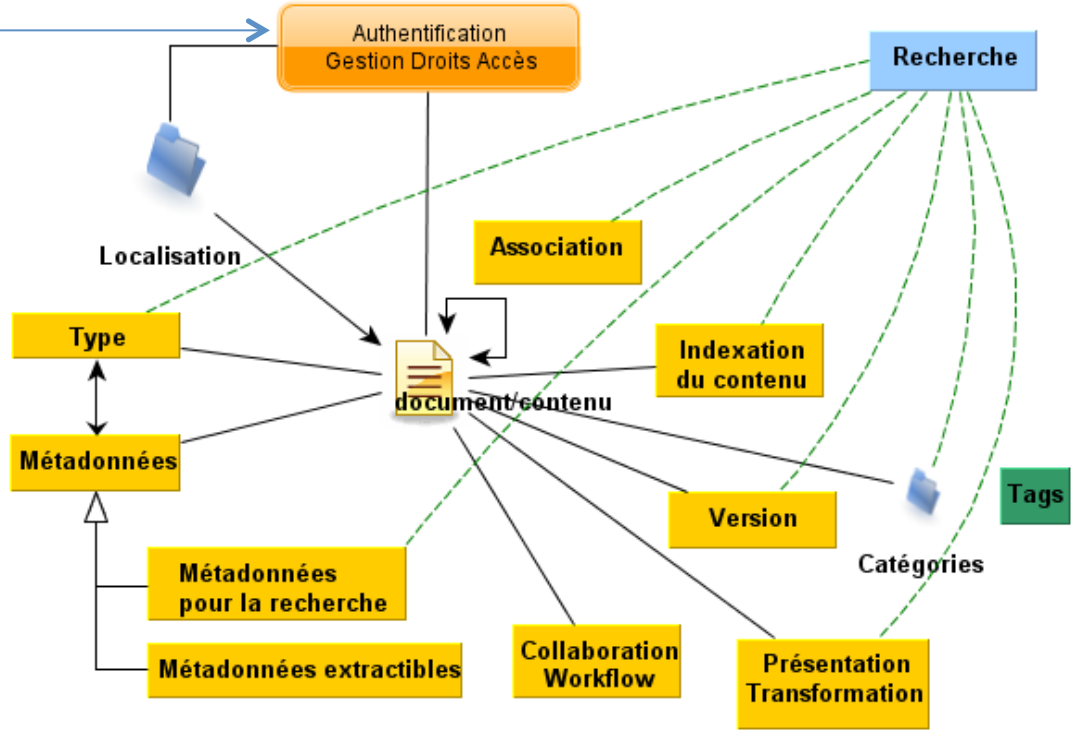
Projet 1 personne

- Rôles :
  - Installation et Maintenance de l'application (sauvegarde, mise à jour)
  - Gestion de l'entrepôt
  - Formation et support utilisateurs
  - Gestion des droits (LDAP, comptes locaux)
  - DBA
  - Tests, développement, référent Office/métadonnées/trame
  - aussi utilisateur

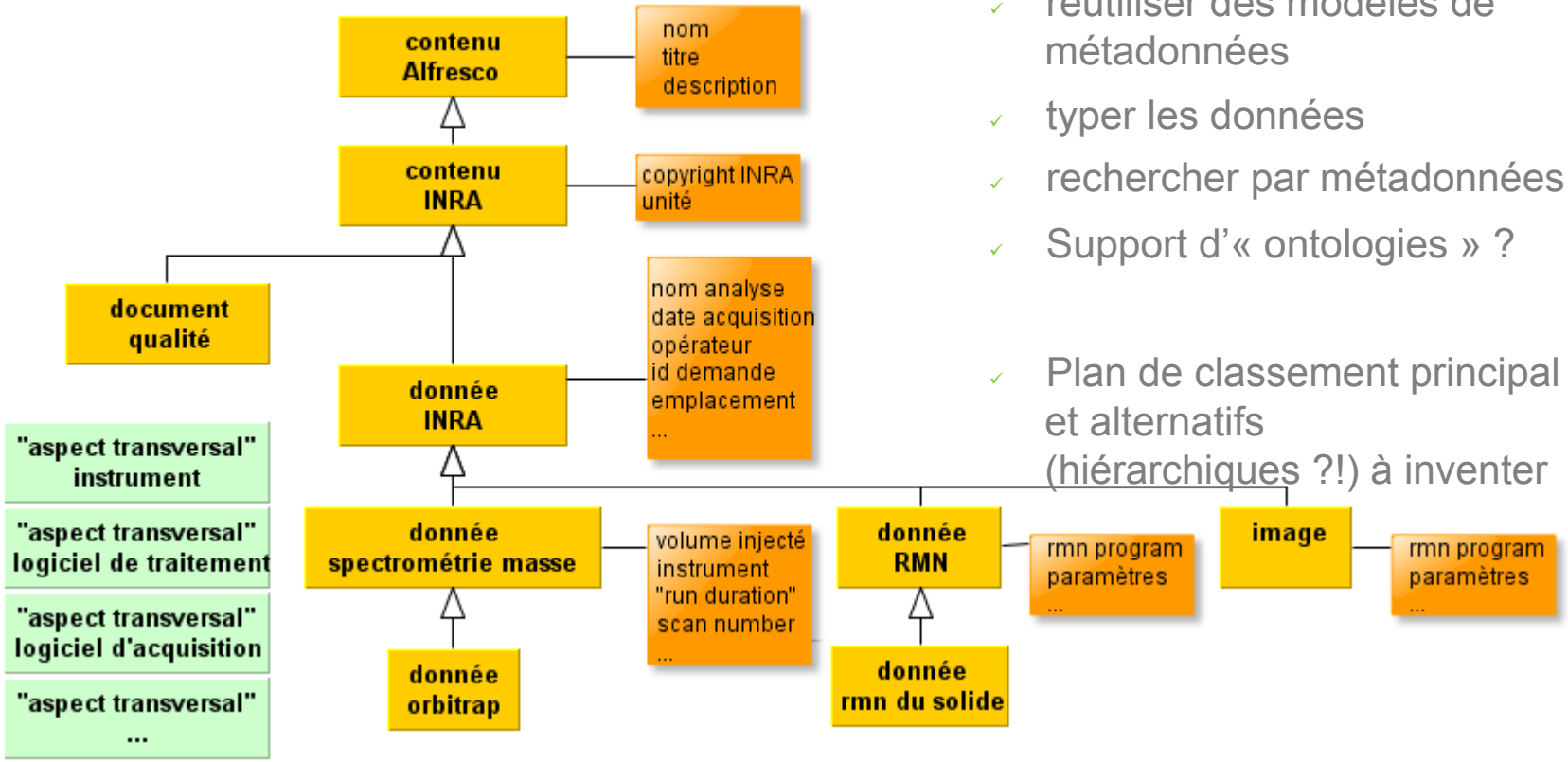
+ ressource : Administrateur Système (Serveur / OS/Linux RedHat)

Fonctionnalités attendues : exploiter la GED !

Vision 2008



Modéliser nos données (modèle de contenu)



- réutiliser des modèles de métadonnées
- typer les données
- rechercher par métadonnées
- Support d'« ontologies » ?
- Plan de classement principal et alternatifs (hiérarchiques ?!) à inventer

Bref historique Alfresco (et l'impact sur le projet)

- 2005 : version 1 Explorer JSF, Lucene, CIFS, FTP/WebDAV, JCR
- 2007 : version 2 (CE → GPL)
- 2009 : version 3

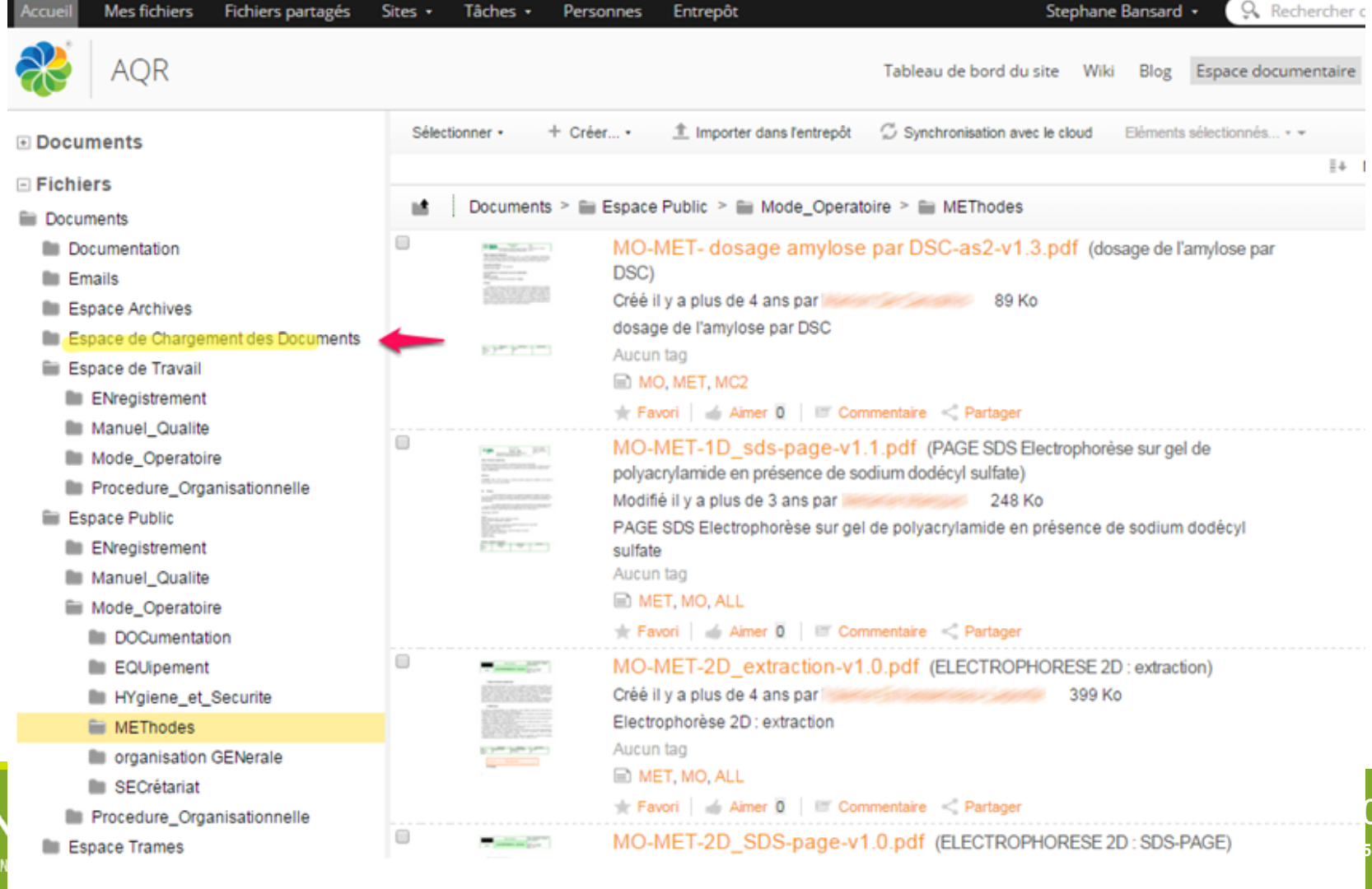
  Nouvelle interface (Share vs Explorer), Ajax

  + SharePoint, IMAP, Web Scripts (JS/XML/FTL), CMIS
- 2011 : version 4 SOLR, Activiti, drag'n drop, Tika, Cloud
- 2015 : version 5 HTML5 preview, SOLR4 (exit lucene, Explorer, JCR)

UC1 : la gestion de la qualité d'une unité

Pour la qualité :

- Côté Alfresco-2.1 (Explorer, pas Share, pas de site)
  - Type « Qualité » avec champs obligatoires (XML/code, équipe, catégories + aspects)
  - Règles appliquées aux documents (respects de convention de nommage p.ex)
  - Logique par scripts (classement automatique, transformation)
  - Droits d'accès (lecteurs, groupes)
- Côté documents :
  - 1000 documents bureautiques dans différentes équipes à intégrer
  - « Normalisation » : nom, doublons, écriture des codes, formats
- Transfert par FTP → déclenchement des règles → scripts « one shot »
- Migration à une date donnée


Accueil
Mes fichiers
Fichiers partagés
Sites
Tâches
Personnes
Entrepôt
Stephane Bansard
Rechercher
AQR
Tableau de bord du site
Wiki
Blog
Espace documentaire
Documents
Fichiers
Documents
Documentation
Emails
Espace Archives
Espace de Chargement des Documents
Espace de Travail
ENregistrement
Manuel_Qualite
Mode_Operatoire
Procedure_Organisationnelle
Espace Public
ENregistrement
Manuel_Qualite
Mode_Operatoire
DOCumentation
EQUipement
HYgiene_et_Securite
METhodes
organisation GENerale
SECrétariat
Procedure_Organisationnelle
Espace Trames
Sélectionner
Créer...
Importer dans l'entrepôt
Synchronisation avec le cloud
Eléments sélectionnés...
Documents > Espace Public > Mode_Operatoire > METhodes
MO-MET- dosage amylose par DSC-as2-v1.3.pdf (dosage de l'amylose par DSC)
Créé il y a plus de 4 ans par
89 Ko
dosage de l'amylose par DSC
Aucun tag
MO, MET, MC2
Favori
Aimer 0
Commentaire
Partager
MO-MET-1D_sds-page-v1.1.pdf (PAGE SDS Electrophorèse sur gel de polyacrylamide en présence de sodium dodécyl sulfate)
Modifié il y a plus de 3 ans par
248 Ko
PAGE SDS Electrophorèse sur gel de polyacrylamide en présence de sodium dodécyl sulfate
Aucun tag
MET, MO, ALL
Favori
Aimer 0
Commentaire
Partager
MO-MET-2D_extraction-v1.0.pdf (ELECTROPHORESE 2D : extraction)
Créé il y a plus de 4 ans par
399 Ko
Electrophorèse 2D : extraction
Aucun tag
MET, MO, ALL
Favori
Aimer 0
Commentaire
Partager
MO-MET-2D_SDS-page-v1.0.pdf (ELECTROPHORESE 2D : SDS-PAGE)

UC1 : reprise de l'existant

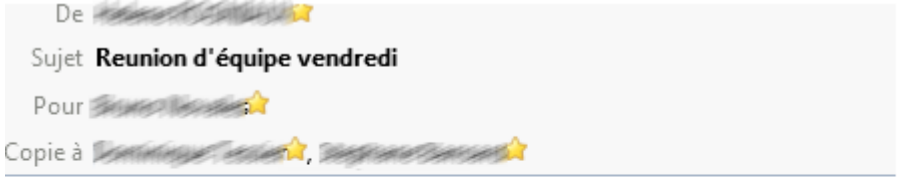
- Découverte de *problèmes*
  - Transformations PDF via OpenOffice pas toujours satisfaisantes
    - en-têtes, nouveaux formats (OpenXML)
  - Recherche par métadonnées/contenu/nom du document… étrange
  - métadonnées extraites imparfaites (encodage)
- Utilisateurs :
  - Changements d'habitude (→ lourdeur ?)
  - Finalement assez complexe (comprendre ce qui se passe, versionner « check in/check out », publier, archiver)
- Administrateur
  - Scripts, listings dynamiques manquants, infos dans les documents, migration vers Share
  - Gestion des groupes (très mouvant comme stagiaires, OpenLDAP et CIFS !)
  - Erreurs liées à « Qualité » (enregistrements, versions)
- Bilan + : pas de dérive dans le classement, évolution de l'unité bien prise en compte (restructuration → changement de métadonnées), mais obsolescence (Alfresco Explorer). Recherche s'est améliorée.

UC2 : Documents d'une plate-forme

Nos « documents » (données) scientifiques

La gestion des fichiers bureautiques et de leurs versions

→ Alfresco devient le « SI » de la plate-forme (« C'est dans Alfresco ! »)



→ « définition JoSy »

Ensemble des documents d'une entité…

documents électroniques de nature administrative et scientifique…

…démarche qualité…

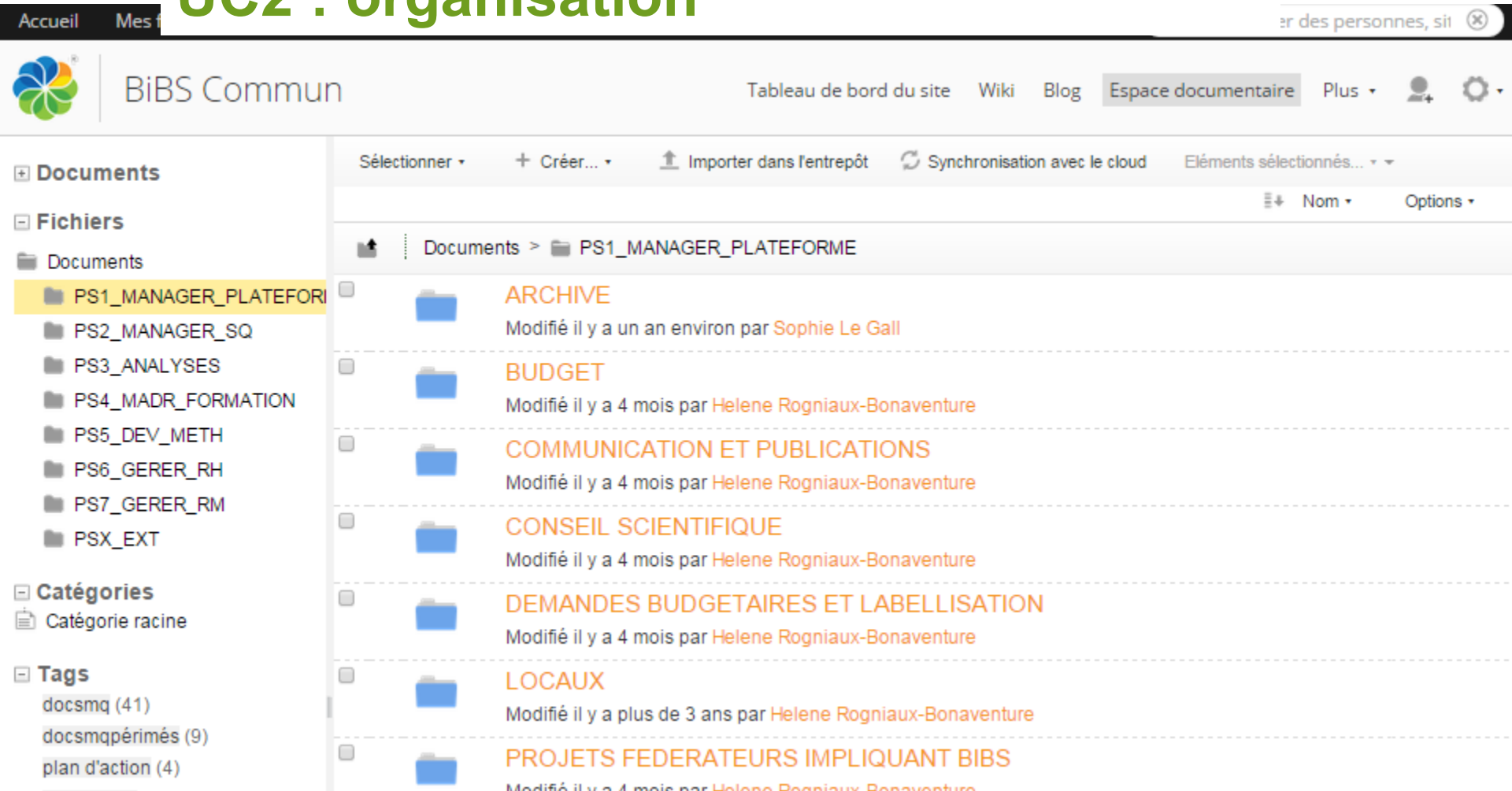
UC2 - La gestion des documents

Contexte ISO 9001
Organisation
en Processus

- Audits internes et certification / renouvellement
- Nombreux documents à bien identifier (non conformité ?)
- Certains documents +/- confidentiels (hors logique Site)
- Mise à jour parfois de manière concurrente, à faire valider
- Accéder aux dernières versions
- Accessibilité depuis tous les postes de travail
- Logique ISO 9001 à assimiler pour tous les membres de la plate-forme (permanents et temporaires)

Les dossiers partagés devenaient délicats pour la gestion des droits d'accès, la transformation des documents, les dernières versions, la recherche…

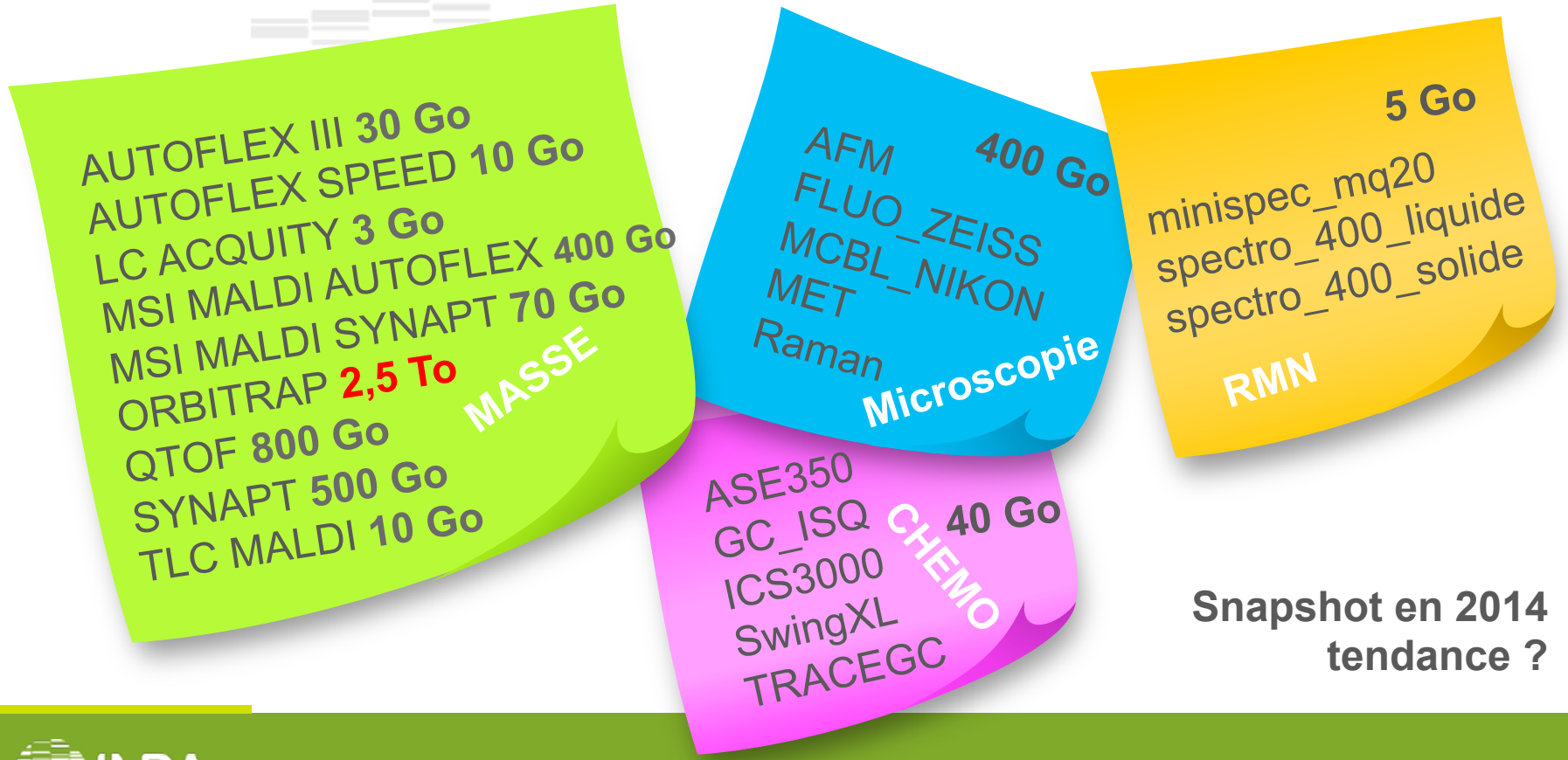
UC2 : organisation



Les boîtes noires qu'on appelle données

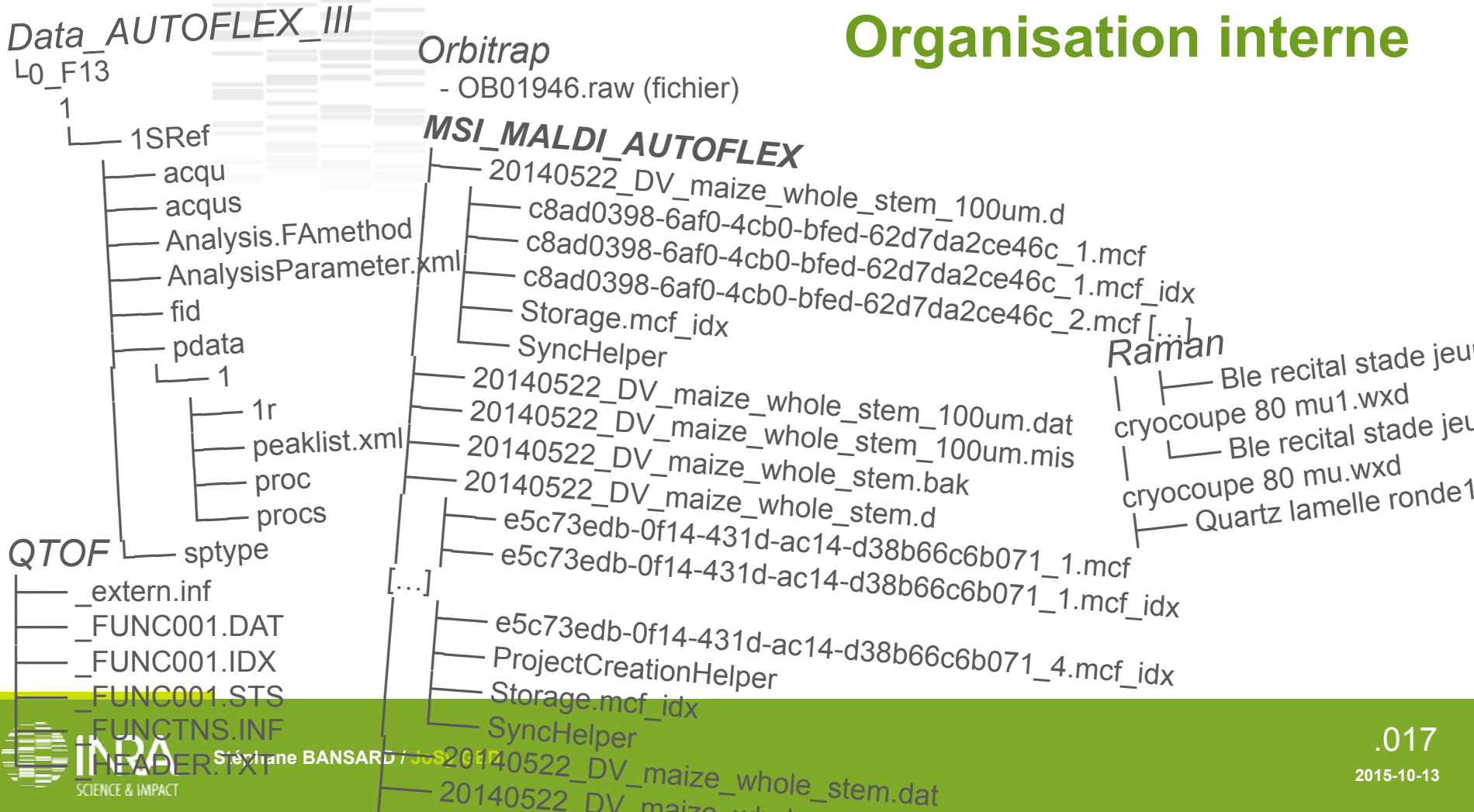
Données brutes
élaborées
Métadonnées

Volumétrie ?



Snapshot en 2014
tendance ?

Organisation interne

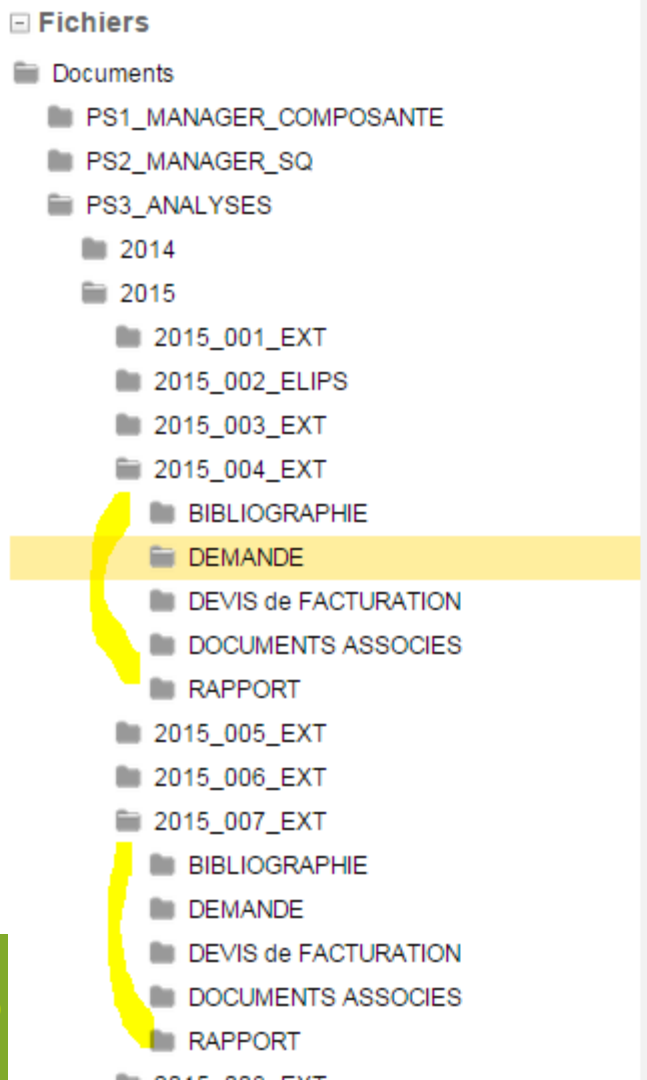


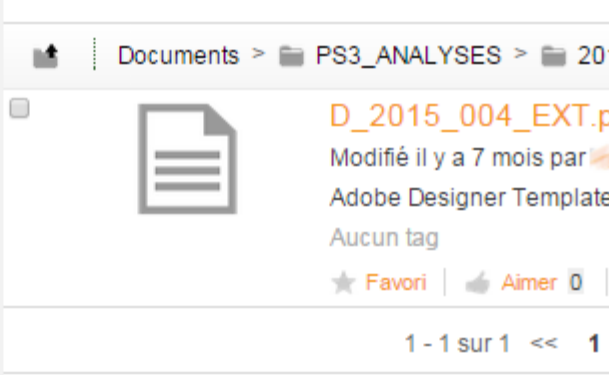
Organisation du contenu

Rester homogène !

Même organisation
Des sites,
Des espaces
documentaires

Données ?
Normalisation ! Transfert ?
Scalabilité ?





Explorés, peu ou pas utilisés

- ✓ Les workflows
- ✓ Les web scripts
- ✓ Les extensions (Add-ons)
- ✓ Le développement spécifique (SDK)

Architecture en place

- 1 serveur de production + 1 serveur de test
- 8 CPUs et 64 Go (8*8) dont 12 alloués à Java (Alfresco n'est pas si gourmand)
- OS : RHEL 5
- Base de données : MySQL-5.1 (→ 5.6)
- Moteur de recherche : lucene (→ SOLR-4)
- Accès HTTPS Tomcat
- Partition 10 To sur NAS NetApp (non dédié)
- Sauvegarde à chaud (snapshot NAS)
- Sauvegarde logique quotidienne (mysqldump)
- Mise à jour des utilisateurs quotidienne

Quelques statistiques

- Nombre de documents ~100 000
- Nombre de dossiers : 15 000
- Nombre de sites : 25
- Utilisateurs quotidiens : 50
- Volumétrie actuelle : ~100 Go

Bilan

- Côté utilisateur
  - Notions à bien faire passer
    - Environnement web down|upload
      - La connaissance des protocoles aide (FTP)
    - Regarder comment les autres organisent leur contenu
    - Les métadonnées minimales (→ Trames avec Champs)
    - Que fait Open|LibreOffice ?
  - La recherche insatisfaisante

Bilan

- Côté admin
  - Charger le système assez vite (donc, ne pas trop modéliser)
  - Basculer complètement (↔ confiance dans Alfresco)
  - Il faut des contraintes (qualité, audits)
  - meilleure image avec les versions successives d'Alfresco (Share)

- La licence Entreprise (support) aide. Pour combien de temps ?

- Personnaliser vs Exploiter le système

Différences **Inria** / INRA

- Inria: volonté de fédérer tous les documents
  produits/échangés par l'Institut (principalement côté Administratif)

- Inria: 5000 utilisateurs internes + ouverture vers l'extérieur
- Inria: Base documentaire pour d'autres services Web
  (Intranet, gestion des séminaires d'évaluation, …)
- Inria: Du spécifique qui arrive seulement maintenant

Différences Inria / **INRA** (BIA)

- INRA: mise en place locale, bottom-up

  Doute sur la scalabilité / faibles ressources.

  Spécifique vers générique !
- Besoins spécifiques (Certification)
- INRA/BIA : 200 utilisateurs internes + ouvertures vers l'extérieur

  Outil quotidien pour 25 personnes.
- Ouvrir vs personnaliser ?

MERCI

??? Questions ???