



DOI automatisé : Retour d'expérience

Mise en oeuvre dans le cadre du projet SSHADÉ

Philippe Bollard (CNRS/IPAG/OSUG)

`philippe.bollard@univ-grenoble-alpes.fr`

10/03/2020 - Hackaton ASOV 2020 - Paris

1. Contexte
2. Problématique
3. Mise en oeuvre
4. Conclusion

Contexte

<https://www.sshade.eu/>

Projet

- Entrepôt de données spectrales issues d'analyses de solides
- Données et toutes les métadonnées du contexte expérimental
- Stockage, indexation, diffusion (VO), publication (DOI)
- SNO labellisé par le CNRS/INSU

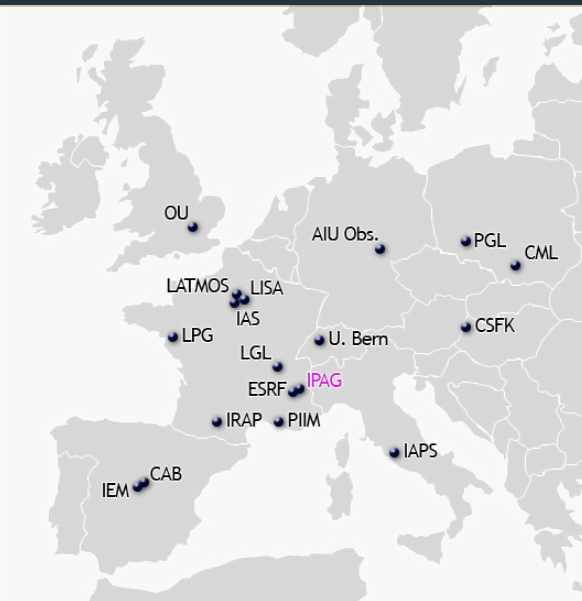
Équipe active

- Bernard Schmitt (porteur du projet scientifique)
- Philippe Bollard, Damien Albert (développement)
- L. Bonal, O. Poch (contributeurs scientifiques)



20 bases actives

- France (10)
- Pologne (2)
- Italie (2)
- Suisse (1)
- Allemagne (1)
- Espagne (1)
- Hongrie (1)
- UK/Inde (1)



Et d'autres à venir !

 SQLAlchemy

 PostgreSQL

 Pyramid™

 python™


elasticsearch

 **jQuery**
write less, do more.

 **B** Bootstrap

 plotly.js

et GAVO DaCHS pour la partie VO / EPN-TAP

Problématique

Générer un DOI automatisé

- sur chaque jeu de données ('experiment')
- sur chaque base partenaire ('database')
- sur SSHADE

Contraintes

- pas de travail suppl. pour les producteurs de données
- gestion des mises à jours des métadonnées

Responsabilité scientifique des données

- incombe à chaque partenaire pour ses données
- ne doit pas être changée au profit de l'OSUG car SSHADE héberge les données de partenaires étrangers

Visibilité de SSHADE

- dans les statistiques DataCite
- dans les matricules DOI (inclus dans les citations)

Mise en oeuvre

Entité neutre associée aux DOI

- responsabilité scientifique des données “partagée”
- responsabilité technique des données incombant à l'hébergeur

Contraintes

- Nécessité de signer un contrat individuel avec l'INIST
- SSHADE n'est pas une entité juridique : besoin de faire porter le contrat par l'OSUG (UGA)

Intégration du datamodel DOI à celui de SSHADE

- Définition d'une table de correspondance entre les datamodels
- Ajout des quelques champs nécessaires dans les XML d'import de données dans SSHADE
- Surcharge de travail limitée pour les producteurs de données :)

Nomenclature des matricules DOI

- {prefixe}/SSHADE/{identifiant donnée}
- 10.26302/SSHADE
- 10.26302/SSHADE/GHOSST
- 10.26302/SSHADE/EXPERIMENT_BS_20121213_002

Module Python 'datacite'

- <https://pypi.org/project/datacite/>
- <https://github.com/inveniosoftware/datacite>

Worklow

- Déclenchement
 - explicite : par mise en "public" d'une donnée et validation de l'utilisateur
 - implicite : par mise à jour d'une donnée "public"
- Génération d'un XML DataCite + Validation par XSD
- Contrôle éventuel par l'utilisateur
- Soumission à la plateforme DataCite
- Enregistrement du matricule DOI dans SSHADE + XML associé

Conclusion

Contrat : Lenteurs administratives (plusieurs mois !)

- compréhension de nos besoins par les différents acteurs (OSUG / INIST)
- plusieurs itérations pour obtenir une proposition de contrat satisfaisante
- signature du contrat par les différents acteurs (INIST / UGA)

Plateforme DataCite

- pas de compte de test avant signature du contrat ⇒ impossibilité de tester l'implémentation en parallèle
- évolution récente de la plateforme DataCite : 2 instances séparées (prod/test) avec des comptes et préfixes différents
- indisponibilités serveur à gérer dans le workflow de soumission

Sérieux...

- soutenu et utilisé par le CERN
- forké par Caltech (et B. Cecconi ;)

Mais pas à jour !

- Datacite schema 4.2
- dernier commit : juillet 2019
- dernière release : mars 2018 (PyPi)

Et buggé !

- XML généré par conversion d'une structure JSON (beurk!)
- Erreurs dans les enums (Organizational / Organisational) : le XSD ne valide pas la sortie de la conversion JSON... :/

Questions?