



Barcamp dataESR - scanR

Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation

Vendredi 27 avril 2018

Barcamp



- ▶ Premier **Barcamp** organisé par l'équipe **dataESR** du Ministère de l'Enseignement supérieur de la Recherche et de l'Innovation
- ▶ Consacré à **scanR**, moteur de la recherche et de l'innovation ▶ **refonte** après 18 mois de vie : réflexions sur l'ergonomie, les fonctionnalités, les données
- ▶ **Concept** : un BarCamp est une rencontre, une non-conférence ouverte, qui prend la forme d'**ateliers-événements participatifs** où le contenu est fourni par des participants qui doivent tous, à un titre ou à un autre, apporter quelque chose au Barcamp. C'est le principe « pas de spectateur », « tous participants ». (source : wikipédia)

Programme



- ▶ Accueil
- ▶ **Plénière n°1 :**
 - ▶ Point sur scanR et retour sur les premiers travaux de refonte
 - ▶ Questions SMS (numéro d'appel spécial dédié à l'événement)
- ▶ **Ateliers :** identification de nouvelles **sources** d'information et proposition de nouvelles **fonctionnalités/idées**
- ▶ **Plénière n°2 :** restitution des ateliers

Plénière n°1



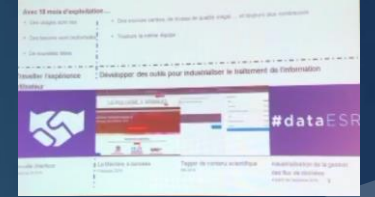
- ▶ Présentation de la manifestation (Emmanuel Weisenburger)
 - ▶ Les participants : une cinquantaine de personnes, **grande diversité**, ESR, monde de la donnée
 - ▶ L'équipe scanR :
 - ▶ Une **dizaine de personnes**
 - ▶ Sous-direction des Systèmes d'information et études statistiques du MESRI
 - ▶ Statisticiens, développeurs, économistes, cartographes
 - ▶ Mission : **ouvrir les données**, produire des services, offre entièrement ouverte

Point sur scanR



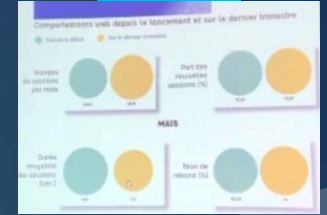
- ▶ scanR : service de **profilage des acteurs de la recherche publique et privée**
 - ▶ 30 000 **entreprises**, 5 000 **institutions** publiques et 3 700 **laboratoires**
 - ▶ Objectifs :
 - ▶ **Aide à la décision** pour le ministère
 - ▶ **Valoriser** les acteurs publics de la recherche
 - ▶ Contribuer à la **mise en relation**
 - ▶ Inciter à l'**amélioration** de la qualité de l'information
 - ▶ **Utilisation de données structurées** de bases de données, de référentiels (RNSR, HAL, SIRENE, ...), ainsi que **données crawlées**
 - ▶ Derrière l'appli : l'équipe scanR, et un partenaire privé Sword (conseil en intégration de données)

Refonte : Back office



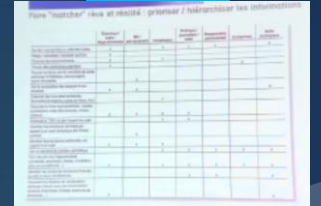
- ▶ Bilan après 18 mois d'usage
 - ▶ **Utilisateurs divers et usages divers** ▶ Comment faire pour que l'outil réponde le mieux aux besoins
 - ▶ **3 chantiers** sur les sources :
 - ▶ **Industrialiser le traitement** sur les diverses sources, rapprocher les données des sources externes et référentiels internes (à l'aide d'une « **machine à données** »)
 - ▶ **Industrialiser les flux de données** (dédoublonnage, etc.)
 - ▶ **Caractériser** les structures de recherche : à l'aide d'un **tagger de contenu scientifique** pour repérer des mots-clés et domaines de recherche, et les associer aux structures

Refonte : Front office



- ▶ Projet Front mené par la société **WeDoData** (Karen Bastien)
 - ▶ « **Raconter** » la **donnée** (par ex. l'open data) : la comprendre, à qui raconter, et comment
 - ▶ Analyse des connexions (**analytics**) ▶ connaître les utilisations de scanR :
 - ▶ Consultation régulière en semaine ▶ **utilisateur professionnel**, expert
 - ▶ Nouvelles sessions mais augmentation du **taux de rebond** (fin de vie de la version bêta)
 - ▶ **40 % de recherches avancées**, filtres sur la géographie, le type d'entités, les tutelles
 - ▶ **Taux de partage important** des urls les plus vues ▶ potentiel de viralité

Refonte : Front office



- ▶ Comment faire vivre cette **masse d'informations**
 - ▶ **Brainstorming** au sein de l'équipe scanR : des dizaines de **vues de données** sont proposées
 - ▶ Définition des **publics cibles** : labo, chercheur, étudiant, investisseur, journaliste, responsable politique,... ▶ **mise en correspondance des vues**
 - ▶ **Sondage** lancé auprès des utilisateurs avec 160 réponses ▶ **Hiérarchiser, prioriser** les vues possibles
 - ▶ ScanR : rechercher mais aussi **propulser de l'information**, produire des focus, des graphiques (ex. : les 10 thèmes les plus actifs dans la recherche française) ▶ A tester

Refonte : Front office



- ▶ Comment **optimiser** l'expérience utilisateur
 - ▶ Offrir un **système de recherche et de filtrage puissant**
 - ▶ **Visualisation interactive** : ex. carte de France à télécharger
 - ▶ **S'adapter au contenu** en constante évolution : travail sur des briques d'information
 - ▶ Maquette non figée
 - ▶ Travail sur **l'accessibilité** (compréhension des termes, sources, calculs) ▶ FAQs dynamiques
- ▶ Test de la **pré-maquette** sur une quinzaine d'utilisateurs aux profils variés



Echanges et Questions SMS

- ▶ **Qualité des données** dans scanR
 - ▶ **Correction en amont** les sources qui alimentent scanR (ex. : RNSR)
 - ▶ Etre transparent sur les **liens des sources** (identifiants HAL, IDREF, etc.)
 - ▶ Réseau de correspondants dans les établissements (en plus du RNSR)
 - ▶ RNSR : unité de base = labo (responsable des données = directeur), campagne annuelle de mise à jour par invitation de la tutelle auprès de chaque directeur ▶ MAJ sera en augmentation quand scanR sera plus visible, plus utilisé
 - ▶ Reste le problème des mises à jour des entreprises
 - ▶ En France : RNSR + SIRENE ouvert depuis 2017 = sources importantes à disposition
 - ▶ Exhaustivité impossible ▶ voir avec le futur Conditor pour une couverture plus importante



Echanges et Questions SMS

- ▶ Possibilité de **recupérer les données** de scanR
 - ▶ APIs scanR ▶ fichier json d'infos sur les structures
 - ▶ Portail Opendata : données MAJ moins souvent (mais beaucoup de sources à MAJ annuelle)
 - ▶ Possibilité de naviguer par sous-collection thématique, tag ?
 - ▶ Récupérer infos au choix, exports personnalisés, etc.
 - ▶ Souhait de récupérer ses propres données pour les réintégrer et les combiner sur son site
 - ▶ Possibilité de se connecter pour récupérer ses recherches, etc. ▶ en 2019 ?



Echanges et Questions SMS

► Collaborations internationales

- Chantier sur les données : OpenCorporates, etc. ► identifiant unique de structure étrangère (pour info : 50 % références HAL avec données étrangères)
- Brevets : pas Patstat, pas de bases internationales ; pour l'instant : sous-ensemble de brevets français
- CrossRef API : marginal, pour DOI
- Référencement de formations
- Ce qui se passe à l'étranger : équivalent de scanR : Research Apps : avec Modène, Milan

Ateliers



- ▶ Les participants sont repartis en **6 ateliers** selon la couleur du programme qu'ils ont reçu à leur arrivée
- ▶ Chaque atelier est **animé** par l'un des membres de l'équipe scanR
- ▶ Un **rapporteur** est désigné pour la restitution ultérieure
- ▶ Chaque participant se présente et évoque son **utilisation de scanR**
- ▶ Une réflexion sur de **nouvelles sources d'information** à intégrer dans scanR est menée ▶ chaque participant propose plusieurs sources et en explique l'intérêt pour scanR
- ▶ A l'issue de l'atelier, **3 idées nouvelles** doivent également être dégagées

Plénière n°2 : Restitution des ateliers



► Sources

Reflet de la typologie des acteurs – sources existantes ou souhait d'information – nouveau type d'information ou complément

- Financements : appels à projets nationaux ou européens
- Partenaires, experts, industriels : COST-EU, projets, CIC
- Technologies : Technéo (BPI), Biotech
- Valorisation : brevets (OEB – Patstat)
- Jeux de données des laboratoires
- Actualités des laboratoires : réseaux sociaux
- Science et société

Approche orientée territoire pour offrir des informations concernant la valorisation

Plénière n°2 : Restitution des ateliers



▶ Idées

▶ Qualification de l'information

- ▶ Normaliser les données pour ce qui concerne les thématiques pour pouvoir relier les différentes communautés
- ▶ Préciser l'origine des informations, les dates de mises à jour
- ▶ Pouvoir agir sur la granularité de l'entité ou de la thématique pour récupérer l'information ou produire des statistiques
- ▶ Mettre en relation les projets, les compétences

Plénière n°2 : Restitution des ateliers



▶ Idées

▶ Accès à l'information

- ▶ Améliorer la recherche à l'aide de mots-clés
- ▶ Recevoir des informations sur une thématique, un secteur d'activité
- ▶ Proposer des usages types en fonction des profils des usagers
- ▶ Pouvoir choisir les informations pour l'export des données

▶ Et ...

Pour plus d'infos

- ▶ Les vidéos de la manifestation sont disponibles sous :
<https://twitter.com/ScanrM>
- ▶ A suivre : #scanR #dataESR
- ▶ Sur Framavox : bit.ly/users_scanr