

Statistiques sur Oskar-Bordeaux : note de synthèse.

Thomas Blicck

2022/02/12

Résumé

Présentation des données statistiques disponibles sur Oskar-Bordeaux complétée par un focus sur les référentiels techniques respectés par Oskar-Bordeaux pour ses échanges avec l'écosystème de la science ouverte, tant national (HAL) qu'international (OpenAIRE, Baselab,...).

Table des matières

Auteur	Thomas Blicck
Date	Février 2022
Version	V1

Objet de la note :

Présenter les statistiques et indicateurs fournis par **Oskar-Bordeaux** et les connexions avec d'autres plateformes (**Hal**, **OpenAIRE**).

Quatre aspects seront vus sur le principe d'une analyse SWOT :

1. les forces
2. les faiblesses
3. les opportunités
4. Les menaces

Les Forces

Respect des normes techniques internationales

Le logiciel **Dspace** (v6.3) respecte les normes internationales recommandées pour les archives ouvertes institutionnelles. **Oskar-Bordeaux** est donc nativement compatibles avec les opérateurs internationaux.

Les données statistiques de base sont nombreuses :

Le logiciel *Dspace* fournit une palette riche de données, il n'a pas été nécessaire de demander des développements spécifiques.

Statistiques en mode anonyme :

Il existe deux moyens d'obtenir des données statistiques :

- par le bloc “Statistiques” dans le bandeau de droite
- dans certains autres blocs du bandeau de droite

Données fournies par le bloc “Statistiques” dans le bandeau de droite :

- Trois catégories sont proposées :
 - Documents les plus consultés
 - Statistiques par pays
 - Auteurs les plus consultés

Elles sont affichées dans chaque cas sous forme d'un tableau qui fournit les données suivantes :

- Multiples options de tris :
 - Nombre de résultats
 - Choix de la zone géographique
 - Choix de la période prise en compte : mois passé, 3 derniers mois, 6 derniers mois, année passée, sur mesure

Pour chaque publication, la liste des pays /téléchargements pour les 6 derniers mois.

En bas de tableau, le nombre total de téléchargements de fichiers et de consultations est affiché.

Données fournies par les autres blocs (bandeau de droite) :

- Bloc découvrir
 - Avec fichier(s)
 - Auteurs
 - Date de publications
 - Type de document
 - Unité de recherche
 - Tutelle
 - Discipline

Les données affichées ici viennent des index internes (auteurs, titres, ...) et sont mises à jour automatiquement.

Les statistiques en mode connecté (administrateur) :

Dans chaque cas l'interrogation d'**Oskar-Bordeaux** est dynamique, les résultats sont établis au moment de l'interrogation.

Les résultats sont fournis sous forme de tableau :

Tableau avec données chiffrées :

Possibilité d'appliquer des filtres chronologiques :

* mois passé, 3 derniers mois, 6 derniers mois, année passé, sur mesure

Formats d'export proposés :

* Format xls uniquement

* Envoi d'un rapport par mail selon une périodicité choisie (jour semaine,mois, trimestre, année)

Tableau sous forme de graphique

Possibilité d'appliquer des filtres chronologiques :

* mois passé, 3 derniers mois, 6 derniers mois, année passé, sur mesure

Formats proposés : * 11 formats de graphiques sont disponibles :

Barres, Barres 3D, Barres empilées, Camembert, Camembert 3D, Ligne, Points, Lignes 1 Points, Secteurs, Secteurs empilés

Formats d'export proposés :

* Export des tableaux en tant que fichier xls * Envoi d'un rapport par mail selon une périodicité choisie (jour semaine,mois, trimestre, année)

— Bandeau de droite, bloc "statistiques" :

Mêmes critères de choix et données que pour le mode "anonyme" (*Documents les plus consultés, statistiques par pays, auteurs les plus consultés*) complétés des éléments suivants :

Statistiques d'usage

Critères de choix sont multiples : 1) Par collections 2) Par publications 3) Par fichiers

Statistiques disponibles :

— Informations sur les consultations de notices et téléchargements de fichiers

— Recherches par collections, publications, fichiers, métadonnées (Langue, auteur, sujet)

Possibilité d'exclure les recherches par **bots** et les recherches internes.

Statistiques de contenu

Recherches croisées multicritères

- Recherche par jeux de données primaires :
 - Date de publication,
 - type de document,
 - langue,
 - auteur,
 - discipline Possibilité de croiser 2 items de cette liste.

Statistiques de recherche

Données sur les recherches réalisées par les utilisateurs d'**Oskar-Bordeaux** Il n'est pas possible de distinguer les utilisateurs internes des connexions anonymes. Possibilité d'appliquer des filtres de recherche : - Chronologique : dernier mois, 3 dernier mois, 6 derniers mois, année passée, sur mesure - Périmètre de recherche : toutes les collections, une collection particulière

Statistiques de stockage

Données à usage interne, à vocation de pilotage et de suivi technique d'**Oskar-Bordeaux**.

Interrogation par collection ou par nom d'utilisateur

Pour chaque collection, indication du nombre de fichiers et de la taille mémoire utilisée. Pour les usagers, mêmes données que pour les collections complétées par le groupe d'appartenance dans le workflow de saisie.

Statistiques de workflow

Données statistiques pour chaque étape du workflow : * dépôt, *reviewing*, modifications, édition finale.

Granularité des données disponibles :

- Tout le site
- Par collection
- Par groupe de workflow ; déposants, valideurs, gestionnaires, administrateurs,
- Par compte utilisateurs

Les Faiblesses

Impacts des personnalisations de *Dspace*

De nombreuses personnalisations de la version standard de *Dspace* ont été réalisées pour les besoins spécifiques d'**Oskar-Bordeaux** (cf. cahier des charges). A l'usage, plusieurs bugs sont apparus dans les résultats statistiques fournis :

- Ticket 674 : indexation des publications
- Ticket 1218 : Problèmes avec les statistiques des pages most-popular
- Ticket 1494 : Affichage des publications : tri par ordre croissant ou décroissant de date
- ticket 2065 : Navigation par facette : bug “date de publication”

Ces bugs portaient sur des fonctions standards de *Dspace*. Ils n’ont pas été bloquant. La parfaite maîtrise par notre prestataire de ce logiciel lui a toujours permis d’apporter des solutions aux problèmes signalés.

Module statistiques

Oskar-Bordeaux est également doté d’un module statistique complémentaire (*Plugin*) développé par le prestataire¹.

Contrairement au point précédent, ce module ne fait pas partie de l’“offre standard” du logiciel *Dspace*.

C’est un outil à usage interne, réservé aux administrateurs fonctionnels. Il apporte une interface dédiée pour l’exploitation des données statistiques. voir ci dessus 1.1.2. *Statistiques en mode connecté (administrateur)*

- 2) Les fonctionnalités techniques proposées sont “figées” :
 - Format d’export limité au seul format **xls**. Pas de possibilité d’export en **CSV** ou **json** pour une exploitation de grands volumes de données avec des outils externes dédiés (exemple environnement *Python*).
 - L’ergonomie du plugin n’est pas “user friendly”
 - Certaines statistiques fournies (exemple les statistiques de recherche) sont peu exploitables et gagneraient à être améliorées.

Les autres utilisateurs de ce module (les administrateurs des archives ouvertes **LilloA**, **SAM**) partagent cette impression de “boîte noire”.

Les Opportunités

- **Oskar-Bordeaux** répondre au “cahier technique” du *Baromètre de la Science Ouverte* proposé par le **Ministère de l’Enseignement Supérieur, de la Recherche et de l’Innovation**² :

Extraits des recommandations techniques du *Baromètre de la Science Ouverte* :

1.1 Perimeter definition

1.1.1 French Open Science Monitor (...)

1. ???

2. <https://barometredelascienceouverte.esr.gouv.fr/a-propos/methodologie>

only publications with at least an author who has a French affiliation are considered (...)

only the publications with a Crossref DOI are considered (...)

All genres of publications are considered (journal articles, proceedings, books ...) as long as the publication is associated to a *Crossref DOI*.

Le choix de *Dspace* comme infrastructure technique permet à **Oskar-Bordeaux** d'être conforme aux normes techniques internationales d'échanges de données pour une communication "machine à machine".

Exemple du référencement d'**Oskar-Bordeaux** par OpenDOAR :

Les données peuvent donc être mises à disposition des autres plateformes selon différents modes.

Modes disponibles sur Oskar-Bordeaux :

HTML :

Consultation des pages HTML et de leurs contenus par les moteurs de recherche comme **Google Scholar**. Les publications d'**Oskar-Bordeaux** sont automatiquement indexées et disponibles sur ces outils.

API disponibles :

Utilisation de différentes API (*Application Programming Interface*) pour les interactions avec l'écosystème de la Science Ouverte nationale et internationale :

- API Sword (version 2.0) : utilisée pour "pousser" les publications vers l'archive ouverte **HAL**
- API OAI-PMH (version 2.0) : permet l'exposition et l'échange des contenus d'**Oskar-Bordeaux** avec d'autres plateformes. Exemple de sets de données disponibles sur **Oskar-Bordeaux**

Ce protocole est utilisé dans de nombreuses interactions :

- **HAL** (moissonnage des collections)
- **Star** (moissonnage des thèses)
- **OpenAIRE**, agrégateur européen de ressources dans le secteur de la Science Ouverte (archives ouvertes institutionnelles, revues en libre accès, serveurs de données)
- API REST : autre format d'exposition et d'échange des contenus d'**Oskar-Bordeaux** basé sur le protocole REST.