DATACTIVIST

# FICHE BRAINSTORM

## Nom du challenge

**DÉFI 2 : OPEN EDITION - ORGANISMES NON-ACADÉMIQUES**

## Les points de contexte à retenir

1. Contexte d'augmentation de l'accès à la science ouverte depuis plusieurs années, d'où l'émergence d'un nouveau public (non-académique).
2. OpenEdition est une infrastructure d'édition électronique au service de la communication scientifique en sciences humaines et sociales. Il s'agit d'une plateforme d'édition en accès libre (science ouverte).
3. En 2018 : 64 millions de visites venant du monde entier
4. Afin de cerner l'usage des contenus, OpenEdition a développé deux détecteurs de lecteurs : Umberto et Elinor, qui analysent les fréquentations des contenus.

## Les enjeux principaux du challenge

1. Lister les principales organisations qui consultent les contenus
2. Mener un « sondage » à partir de la base de données à propos des usages des visiteurs institutionnels non-académiques dans une approche de sociologie de la lecture
3. Comprendre les liens entre les publications consultées et les activités de l'institution
4. Analyser l'usage que font les organisations non-académiques des contenus

5. Construire un plaidoyer pour la science ouverte : prouver que l'accès libre aux contenus scientifiques touche une multitude de secteurs institutionnels et que ces contenus se révèlent utiles à leur secteur d'activité.

## Quel est le problème à résoudre ? Le(s) point(s) de douleur & frustrations ?

Le problème à résoudre est le manque de connaissances concernant les usages de la recherche scientifique par les organismes non-académiques.

La résolution de ce problème nous permettrait de répondre aux problématiques posées par le challenge, entre autres : qui/quels sont les nouveaux lecteurs qui consultent les publications scientifiques ouvertes au delà du monde académique? Quelles informations peut-on avoir sur ce public à partir des données statistiques?

## Qui sont les parties prenantes dans ce challenge (acteurs concernés) ? Quels seront les utilisateurs de votre réponse ?

OpenEdition est l'acteur majeur et utilisera notre réponse.

## Quelles sont les données concernées ) ?

*(en synthèse - vous pouvez aussi remplir des fiches exploration des data pour répondre à cette question)*

Il s'agit de la base de données Umberto, et plus particulièrement de l'outil Elinor, le détecteur de lecteurs d'OpenEdition (2151 cas à analyser).

## Quelles autres sources / catégories de data peut-on envisager pour répondre au challenge ?

*(cf vous pouvez aussi remplir des fiches exploration des data pour répondre à cette question)*

Sites officiels des organismes non-académiques afin d'établir des liens entre fréquentation et usages des contenus.

## Synthèse de votre idée de projet / vision :

*Vous avez brainstormé sur des solutions possibles en réponse au challenge et vous avez voté pour une idée. Décrivez cette solution en quelques mots. Vous pouvez utiliser le template suivant :*

- Renforcer les fiches de connaissances existantes qui synthétisent l'impact des articles d'OpenEdition pour les organismes non-académiques
- Rédiger de nouvelles fiches de connaissances sur les champs n'ayant pas encore été analysés (organisations internationales, européennes et régionales, associations, hôpitaux, musées, hôtellerie, bibliothèques)
- Réaliser une synthèse générale au format créatif afin de résumer nos résultats, dans une optique de plaidoyer au service de la science ouverte

Aujourd'hui (CONTEXTE)

OpenEdition offre un accès libre à du contenu scientifique (science libre) et dispose d'outils de détection des lecteurs, qui analysent les fréquentations.

Grâce à notre projet, notre utilisateur (ACTEURS CONCERNES)

OpenEdition

Pourra (PROBLEME RESOLU)

étayer sa connaissance du profil des lecteurs et des usages de la recherche scientifique par les organismes non-académiques

Ainsi nous aurons accompli notre objectif de (ENJEUX)

compréhension du choix des contenus par les organismes non-académiques et ce qu'elles en retirent. Cela nous permettra de formuler notre proposition de plaidoyer.

DATACTIVIST

# FICHE EXPLORATION DES DATA

Inspiré de : “Datasheets for datasets” - arXiv:1803.09010

## Nom du challenge

**DÉFI 2 : OPEN EDITION - ORGANISMES NON-ACADÉMIQUES**

## Nom du jeu de données :

Elinor

## Les points de contexte à retenir

1. Elinor est le détecteur de lecteurs d’OpenEdition concernant les organismes institutionnels.
2. Cet outil est une sous-section de la base de données Umberto.
3. L’outil Elinor recense 1200 visiteurs institutionnels pour les articles scientifiques d’OpenEdition. Ce recensement est effectué
4. via les adresses IP collectives consultant le plus fréquemment les contenus scientifiques.

## Les enjeux principaux du challenge

1. Déterminer l’origine des fréquentations d’OpenEdition.
2. Déterminer si les contenus consultés sont en lien avec les activités des institutions.
3. Déterminer comment les visiteurs utilisent ces contenus (citations, consultations récréatives, apprentissage, conseil...)
4. Construire un plaidoyer à partir des résultats obtenus de l’analyse des données.

## Raisons de la production de ces données / producteur des données

Volonté de la part d’OpenEdition de construire une base de données afin de recenser différents aspects de consultation du site OpenEdition.
Il s’agit d’entreprendre une sociologie de la lecture avec du big data, afin de définir les profils des nouveaux lecteurs non académiques.

## Qui ces données concernent-elles ?

Les données contenues dans Elinor concernent tous les organismes qui consultent les contenus de OpenEdition. Nous nous concentrons sur les organismes non-académiques.

## Contenu des données

- Fréquentation en provenance de l’organisme : ces données décrivent quelles publications sont les plus consultées par cet organisme, quelle en est l’évolution.
- Principaux articles visités par l’organisme : donne le détail de la fréquentation pour les principaux articles (au moins 4 visites sur toute la période 2017-2019). Le tableau vise plutôt à donner un aperçu qualitatif aux préférences de lecture de chaque institution.
- Occurrences des articles visités : donne un aperçu des mots les plus spécifiques dans le corpus d'articles visités par l’institution.
- Visites par disciplines : permet d'explorer les principales disciplines caractérisant la consultation par l'institution.
- Evolution de la fréquentation en provenance du visiteur institutionnel : donne l'évolution de la fréquentation en provenance du visiteur institutionnel.

## Type de données

- Fréquentation en provenance de l’organisme : données de log traitées par Matomo et analysées par l’outil Elinor de détection de lecteurs institutionnels (mesure : nombre de visites par publications) et évolution de la fréquentation en provenance du visiteur institutionnel (mesure : nombre de sessions)

- Principaux articles visités par l’organisme (mesure : nombre de visites par articles)

- Occurrences des articles visités : à partir du corpus textuel complet des articles.
(mesure : nombre de fois qu’apparaît le mot)

- Visites par disciplines : à partir d’un modèle de classification automatisé du corpus textuel complet
(mesure : nombre de visites par disciplines)

## Considérations légales et éthiques