

## Nom du challenge : OpenEdition : Contenus atypiques

### Les points de **contexte** à retenir

- 1 . OpenEdition est une plateforme de contenus en libre accès qui vise à rendre accessible les recherches scientifiques
- 2 . Umberto analyse les épisodes de consultations atypiques de littérature scientifique
- 3 . Les consultations atypiques sont les moments où des articles dormants connaissent un pic de consultation
- 4 . Des analyses quantitatives ont déjà été entreprises sur un échantillon réduit (environ 10 cas)

### Les **enjeux** principaux du challenge

- 1 . Identifier les événements, les causes qui ont donné lieu à un pic de consultation
- 2 . Identifier les parcours des utilisateurs
- 3 . Identifier les profils types de lecteurs atypiques
- 4 . Rédiger un plaidoyer en faveur de l'open science

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

Déterminer les situations / événements qui ont conduit à un pic de consultation atypique

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

- OpenEdition
- Plus largement pour le plaidoyer : communauté scientifique, les éditeurs, les institutions publiques, MESRI

Quelles sont les données concernées ) ?

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

Les données d'Open Edition retraitées par l'application Umberto créée par Pierre Car...

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)*

Ensemble des contenus en ligne qui ont donné lieu à cette consultation atypique.

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 :*

Aujourd'hui (**CONTEXTE**) \_on observe des pics de consultations par des utilisateurs atypiques qu'on ne sait pas expliquer.

Grâce à notre projet, notre utilisateur (**ACTEURS CONCERNES**) OpenEdition

Pourra (**PROBLEME RESOLU**) identifier les causes de ces consultations atypiques à partir de cas particuliers pertinents

Ainsi nous aurons accompli notre objectif de (**ENJEUX**) : expliquer la nécessité de rendre davantage accessible les recherches scientifiques

**Nom du challenge****Open Edition #1 Contenus atypiques****Nom du jeu de données :**

Umberto : le détecteur de lecteurs d'Open Edition

**Les points de contexte à retenir**

- 1 . OpenEdition est une plateforme de contenus en libre accès qui vise à vulgariser et à rendre accessible les recherches scientifiques
- 2 . Umberto analyse les épisodes de consultation atypiques de littérature scientifique
- 3 . Les consultations atypiques sont les moments où des articles dormants connaissent un pic de consultation
- 4 . Des analyses quantitatives ont déjà été entreprises sur un échantillon réduit

**Les enjeux principaux du challenge**

- 1 . Identifier les évènements, les causes qui ont donné lieu à un pic de consultation
- 2 . Identifier les parcours des utilisateurs
- 3 . Identifier les profils types de lecteurs atypiques
- 4 . Rédiger un plaidoyer en faveur de l'open science

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

Open Edition a développé Umberto pour analyser la consultation atypique des contenus scientifiques avec une fréquentation inhabituelle et soudaine et expliquer leurs causes. Cette collecte de données permet de mettre en exergue la nécessité de démocratiser les contenus scientifiques en ligne. Pour le moment, une dizaine de cas a été analysé. Nous avons utilisé les données d'Umberto retraitées par Pierre-Carl afin de les filtrer.

**Qui ces données concernent-elles ?**

Population : tous les utilisateurs d'Open Edition

Territoire : mondial

Date : de janvier 2017 à juillet 2019 (220 millions de logs d'origine d'où ont été extrait 5000 anomalies)

**Contenu des données**

L'article et son titre, le nombre de visiteurs, l'intensité, la période, le site qui l'a relayé (source), le type de site, la concentration temporelle, la répartition géographique

**Type de données**

Nature : principalement des données quantitatives et certaines qualitatives (zones géographiques) : base de données chiffrées de la communauté d'Open Edition

**Considérations légales et éthiques**

Les personnes sont informées de la collecte de données par le bandeau d'autorisation d'utilisation des cookies et ils ont la possibilité de consulter la politique de confidentialité. Ils peuvent ne pas être suivis en activant dans son navigateur l'option « Do not track ». Ils ont un droit d'effacement, un droit d'opposition pour des informations obtenues sans consentement. Les données ne sont pas sensibles car il n'y a pas d'obligation de se créer un compte pour y accéder. Elles ne sont donc pas identifiantes. De plus, elles ont pour vocation à rester au service d'OpenEdition et ne sont pas commercialisées.