



OpenEdition

Des données et des pratiques

Sandra Guignon, *Directrice adjointe éditoriale, OpenEdition*





Plan

- Introduction
 - Présentation générale
 - Quatre plateformes au service de la communauté des SHS
- Les données d'OpenEdition
 - La lente émergence d'un besoin...
 - ... jusqu'au projet COMMONS
- Quelles données pour quelles publications ?
 - État des lieux sur les quatre plateformes
 - Exemples d'usages
- Perspectives

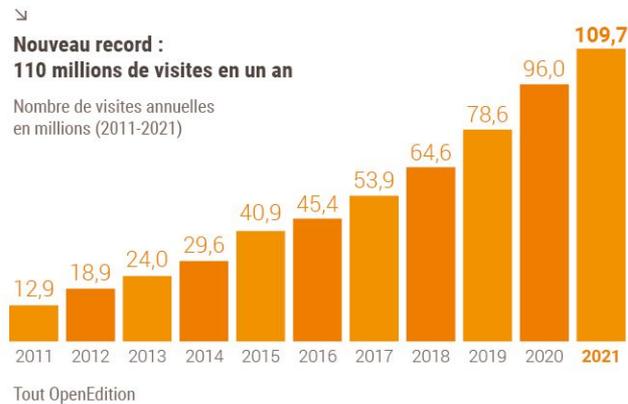


Présentation générale

- Infrastructure nationale de recherche depuis 2016
- Soutenue par le CNRS, Aix-Marseille Université, l'EHESS, l'Université d'Avignon, le MESR et le programme Investissements d'avenir
- **Promouvoir la publication et la communication scientifiques ouvertes dans les sciences humaines et sociales**
- Inscrit son action dans le cadre du [Plan national pour la science ouverte](#).
- Ressources et services numériques : OpenEdition Journals, OpenEdition Books, Calenda, Hypotheses

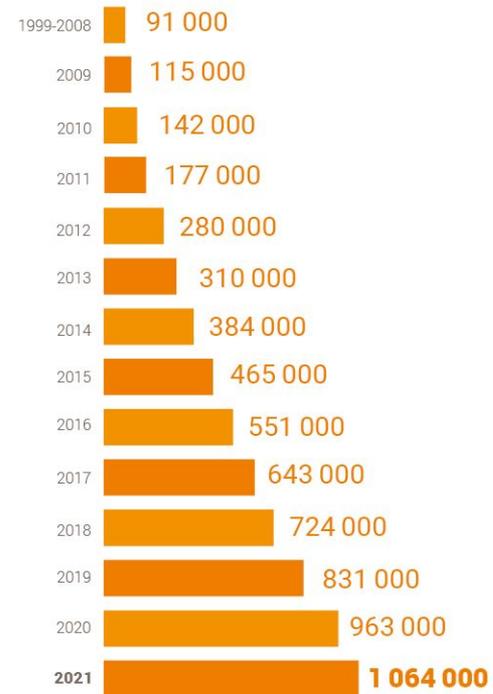


Chiffres clés 2021



**Plus d'un million
de documents publiés**

Nombre de documents publiés
sur OpenEdition (1999-2021)





Chiffres clés. La fréquentation

Plus de 110 millions de visites (9 millions/mois)

- OE Journals: 71.3 (61.7 en 2020 | + 17 %)
- OE Books: 21.4 (16.3 en 2020 | + 32 %)
- Hypothèses: 14.8 (17 en 2020 | - 13 %)
- Calenda: 1.3 (1.04 en 2020 | + 25 %)





Chiffres clés. Les contenus

+1 million de documents, dont 96 % en accès ouvert



43 revues candidates (en 2021)
572 revues en ligne
81,5 % des revues en accès ouvert



200 soumissions par mois (en 2021)
45 892 Événements scientifiques en ligne
100 % des contenus en accès ouvert



18 candidatures par semaine (en 2021)
7 040 Blogs scientifiques 4 226 au catalogue
100 % des contenus en accès ouvert



65 collections soumises, 1 462 livres numérisés (en 2021)
12 026 Livres, 117 éditeurs, 846 collections
82 % des contenus en accès ouvert



Émergence de la question des données

Des logs, des indicateurs, des corpus

- décrits par des jeux de métadonnées (Dublin Core, catégories), typés.
- *Exemple.* Convergence des types de documents vers la norme Z 39-98 <http://www.daisy.org/z3998/2012/vocab/periodicals/>



Émergence de la question des données

- structurés en XML-TEI pour les livres et les revues

```
<item>Ports</item>
<item>Globalization</item>
<item>Privatization</item>
<item>Reform</item>
<item>Hinterland</item>
</list>
</keywords>
- <keywords scheme="personcited">
  <list/>
</keywords>
</textClass>
</profileDesc>
</teiHeader>
- <text>
  - <front>
    - <div xml:lang="fr" type="abstract">
      - <p xml:id="otx_8">
        L'essor contemporain des ports indiens s'inscrit dans la trajectoire de l'ouverture au libéralisme
        modernisation de ces ports ne se réalise pas dans le sous-continent de manière homogène, car l
        autorise une diversité de partis politiques à la tête des États, et donc des gouvernements plus ou
        secteur privé dans les ports. De là, des divergences peuvent surgir entre les divers acteurs, publ
        conflits d'intérêts. L'Union indienne est reconnue aujourd'hui comme un pays émergent doté d'u
        celle-ci ne peut être soutenue sur le long terme, sans un secteur portuaire moderne en mesure c
        notamment chinois. En lançant de nouvelles réformes destinées à la modernisation des infrastru
        <!-- <pb/> -->
        Modi entend rattraper le grave retard accumulé par les ports à l'échelle nationale, et ce grâce à
        projets de grande envergure promus, celui de Sagarmala aurait pour principale fonction de redy
        arrière-pays.
      </p>
    </div>
    - <div xml:lang="en" type="abstract">
      - <p xml:lang="en" xml:id="otx_11">
        The contemporary rise of the Indian ports is in line with the opening to liberalism starting in 19
        these ports is not realized evenly in the subcontinent because of the federal and democratic stru
        political parties in power in the different states. This is why the individual state's governments a
        promotion of the private port sector. Hence, differences may arise between the public and privat
        lead to conflicts of interest. The Indian Union is recognized as an emerging country with strong
        be sustained in the long run without a modern port sector which is able to compete with neighb
```



Émergence de la question des données

- disponibles dans un dépôt OAI-PMH aux formats Dublin Core (DC) et METS (Metadata Encoding and Transmission Standard) > <http://oai.openedition.org/?verb=ListMetadataFormats>

```
-<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2018-10-01T05:05:33Z</responseDate>
  <request verb="ListMetadataFormats">http://oai.openedition.org/</request>
  -<ListMetadataFormats>
    -<metadataFormat>
      <metadataPrefix>oai_dc</metadataPrefix>
      <schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd</schema>
      <metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/</metadataNamespace>
    </metadataFormat>
    -<metadataFormat>
      <metadataPrefix>qdc</metadataPrefix>
      -<schema>
        http://dublincore.org/schemas/xmls/qdc/2008/02/11/qualifieddc.xsd
      </schema>
      <metadataNamespace>http://purl.org/dc/terms/</metadataNamespace>
    </metadataFormat>
    -<metadataFormat>
      <metadataPrefix>mets</metadataPrefix>
      <schema>http://www.loc.gov/standards/mets/mets.xsd</schema>
      <metadataNamespace>http://www.loc.gov/METS/</metadataNamespace>
    </metadataFormat>
    -<metadataFormat>
      <metadataPrefix>tei</metadataPrefix>
      <schema>http://lodel.org/ns/tei.openedition.1.2.xsd</schema>
      <metadataNamespace>http://www.tei-c.org/ns/1.0</metadataNamespace>
    </metadataFormat>
    -<metadataFormat>
      <metadataPrefix>basictci</metadataPrefix>
      <schema>http://lodel.org/ns/tei.openedition.1.2.xsd</schema>
      <metadataNamespace>http://www.tei-c.org/ns/1.0</metadataNamespace>
    </metadataFormat>
  </ListMetadataFormats>
</OAI-PMH>
```



Émergence de la question des données

Des corpus exploitables à des fins de recherche

Des programmes de recherche avec des chercheurs en informatique
(Laboratoire d'Informatique & Systèmes, ex-LSIS)

- Qui visent à lier des contenus entre eux à partir des données structurées dont on dispose (ex. Bilbo)
- Ou qui visent à savoir comment les gens s'en saisissent (ex. Appropriation des savoirs ouverts)



Émergence de la question des données

Des corpus exploitables à des fins de recherche

Des programmes externes de recherche menés sur nos corpus de données

- Une thèse sur l'épistémologie des savoirs géographiques à travers l'analyse des principales revues françaises (Lyon)
- Une thèse portant sur les conditions de production et de circulation des savoirs au sein des plateformes numériques en sciences humaines (Liège)
- Une recherche sur les pratiques d'écriture des doctorants dans les carnets de recherche (URFIST)



Émergence de la question des données

Du côté des pratiques sur les plateformes ?

Des pratiques *ad hoc* et des objets insuffisamment traités :

- des représentations visuelles de données au format image : schémas, cartographies, graphes
- des contenus multimédia peu structurés

Des demandes portant sur la restitution ou la visualisation plutôt que sur la structuration

- peu d'alignements avec des référentiels
- quelques pratiques de data papers



Émergence de la question des données

Du côté des pratiques sur les plateformes ?

- Exemple pour les études théâtrales >

<https://journals.openedition.org/skenographie/218>

Cahier photographique

Angélica Liddell, Klaus Lefebvre, Brigitte Enguérand, Mickaël Troivaux, Elizabeth Carecchio, Claude Bricage, Didier Grappe, Michèle Laurent et Christophe Raynaud de Lage
Des écritures et des plateaux : photographies [Texte intégral]

Carnet de la création

Traduction

Elena Balzamo
Un suédois nous parle de la guerre [Texte intégral]

Hjalmar Söderberg
Basculement [Texte intégral]
Pièce en trois actes, 1922. Extrait de l'acte II, La décision de l'Empereur

Inédit

Pascale Anin
Le Spasme du Sanglot [Texte intégral]
Texte inédit, 2013. Extrait « Jeu d'Ombres »



Émergence de la question des données

Des demandes qui se structurent peu à peu

Exemple en archéologie > <https://journals.openedition.org/archeomed/24044>

The screenshot shows the website for the journal 'Archéologie médiévale'. At the top, there is a navigation bar with links for 'ACCUEIL', 'CATALOGUE DES 595 REVUES', and 'OPENEDITION SEARCH Q'. The journal title 'Archéologie médiévale' is prominently displayed in a red serif font. Below the title, the logo for the 'Centre Michel de Bouard CRAHAM' is visible. A search bar is located on the left side of the page. The main content area features the article title 'Peuples – Cultures protohistoriques' with its URL: <https://ark.frantiq.fr/ark:/26678/pctqUi1NA6MpA>. The article is categorized as a 'Notice archéologique'. The page lists the responsible editor as Madeleine Châtelet, the notice was co-edited with Boris Dottori, and it was published in 'Archéologie médiévale', volume 51, in 2021. The article title is 'Algolsheim (Haut-Rhin), Kirchmatten [Texte intégral]'. Another responsible editor mentioned is Caroline Touquet Laporte-Cassagne, and the notice was co-edited with Pierrick Tigreat. The article title is 'Frépillon (Val-d'Oise), Zone d'aménagement concerté des Épineaux 3, secteur A [Texte intégral]', published in 'Archéologie médiévale', volume 51, in 2021. The responsible editor for this notice is Gaëlle Pertuisot.



Émergence de la question des données

Des demandes qui se structurent peu à peu

Exemple en géographie > <https://journals.openedition.org/cybergegeo/28545>

The screenshot displays the website for the journal 'Cybergegeo', which is the 'European Journal of Geography' and 'revue européenne de géographie'. The page is titled 'Data papers' and features a navigation menu on the left with categories such as 'Index', 'Auteurs', 'Mots-clés', 'Index géographique de référence', 'Années', 'Langues', 'Anniversaire', and 'Rubriques'. The main content area includes a search bar, a language selector (English, Français, Español), and a paragraph of text discussing the journal's commitment to open science and data. Below the text, there is a 'Lire la suite' link and a list of years from 2022 to 2017. At the bottom, the author 'Gabriel Bideau et Ronan Ysebaert' is mentioned, along with the title of the article: 'Data paper : Les communes nouvelles françaises (2012-2022) : une méthode pour l'analyse de données à l'échelon municipal selon des limites évolutives'.



Lier des données et des publications

Horizon 2030 - Mise en œuvre de l'EquipEx+ COMMONS

Les principes **FAIR** et la collaboration avec les **infrastructures Métopes et Huma-Num** sont des points centraux de ce projet.

Le projet bénéficie d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du Programme d'Investissements d'Avenir portant la référence ANR-21-ESRE-0045





Lier des données et des publications

Horizon 2030 - Mise en œuvre de l'EquipEx+ COMMONS

L'environnement intégré de COMMONS couvre l'ensemble de la chaîne de production des connaissances, de la constitution des données à leur diffusion dans la sphère publique. En assurant à la fois l'accès aux publications et aux données, et la liaison entre publications et données, le projet COMMONS apporte une contribution décisive à l'amélioration des conditions de production de la recherche en SHS et favorise la discussion et la comparaison des résultats de recherche, garantissant ainsi le respect de l'intégrité scientifique dans la production et le traitement des données.

Dans cette perspective, les infrastructures OpenEdition, Huma-Num et Métopes se sont alliées afin de faire progresser radicalement la dynamique de la science ouverte autour des trois axes suivants :

- La FAIRisation des données
- L'ouverture des publications et des données
- La montée en compétences des producteurs et des utilisateurs



Lier des données et des publications

COMMONS. E3. Liaison données-publications

(données-média-documents-dispositifs)

Articulé en trois phases de déploiement

1. Citer et lier

Il s'agit de traiter toutes les modalités de citation et de crosslinking

2. Représenter

- a. Représentation à plat
- b. Représentation dynamique

3. Use Cases. Édition érudite

Les avancées des phases 1 et 2 permettront de proposer des prototypes d'objets éditoriaux complexes pour dégager des modèles stabilisés de publication par exemple pour l'édition savante ou les data journals.



COMMONS. E3. Liaison données-publications

Exemple 1. Citation simple (corpus linguistiques)

Kirill Maslinsky, « Positional skipgrams for Bambara: a resource for corpus-based studies », *Mandenkan* [En ligne], 62 | 2019, mis en ligne le 07 mai 2020, consulté le 26 mai 2020. URL : <http://journals.openedition.org/mandenkan/2119> ; DOI : <https://doi.org/10.4000/mandenkan.2119>

Dans la note 1, les auteurs proposent un lien vers

- le corpus, hébergé par Huma-Num > <http://cormand.huma-num.fr/>
- le dataset, hébergé également par Huma-Num > <http://cormand.huma-num.fr/ngrams/>.



Lier des données et des publications

COMMONS. E3. Liaison données-publications

Exemple 2. Représentation à plat de sources distantes. *Bibnum* (sources)

Jérôme Gavin et Alain Schärli, « Luca Pacioli : la multiplication des multiplications », *Bibnum* [En ligne], Calcul et informatique, mis en ligne le 01 mars 2016, consulté le 26 mai 2020. URL :

<http://journals.openedition.org/bibnum/524>

La revue dépose les sources numérisées de textes scientifiques sur Internet Archive cf. <https://archive.org/details/@bibnum>

Chaque source y est décrite et dispose d'identifiants, notamment ark, et de métadonnées

La source est intégrée dans l'article via un iframe, les fonctionnalités proposées par la plateforme sont disponibles pour feuilleter la source depuis le site, et au lien on est dirigé sur la plateforme d'hébergement, à la page consultée.



COMMONS. E3. Liaison données-publications

Exemple 3. Des Data Papers *ad hoc*. *Cybergeo* (Géographie)

Marion Maisonobe, Laurent Jégou et Denis Eckert, « Delineating urban agglomerations across the world: a dataset for studying the spatial distribution of academic research at city level », *Cybergeo : European Journal of Geography* [En ligne], Data papers, document 871, mis en ligne le 12 novembre 2018, consulté le 27 mai 2020. URL : <http://journals.openedition.org/cybergeo/29637> ; DOI : <https://doi.org/10.4000/cybergeo.29637>

La revue propose une section de Data papers depuis 2017

Les documents comprennent des éléments textuels présentant les objectifs et la méthodologie de recueil des données, une partie discussion ainsi qu'une partie Dataset description, qui est mise en forme en s'appuyant sur les possibilités de structuration d'un document de type 'article'.



COMMONS. E3. Liaison données-publications

Exemple 4. De l'édition savante *ad hoc*, L'agenda de Marcel Proust (BnF)

MAURIAC DYER, Nathalie ; et al. Marcel Proust, L'Agenda 1906 : Édition génétique et critique. Nouvelle édition [en ligne]. Paris : Éditions de la Bibliothèque nationale de France, 2015 (généralisé le 28 octobre 2020). Disponible sur Internet : <<http://books.openedition.org/editionsbnf/1457>>. ISBN : 9782717726947. DOI : <https://doi.org/10.4000/books.editionsbnf.1457>.

La mise en ligne de l'agenda a rencontré un écho certain, au-delà même de la communauté académique, mais la plateforme en l'état, ne permet pas de traiter l'objet en structurant ce qui constitue le cœur de l'ouvrage. Voir par exemple <https://books.openedition.org/editionsbnf/1468>



Lier des données et des publications

Une commande du Collège Données du COSO

- <https://www.ouvrirlascience.fr/recommandations-aux-revues-souhaitant-definir-une-politique-de-donnees-liees-aux-publications/>

The screenshot shows the website 'Ouvrir la science!' with a navigation menu: LA SCIENCE OUVERTE, NOS ACTIONS, RESSOURCES, ACTUALITES, QUI SOMMES NOUS?. The main content area features a yellow document cover titled 'Recommandations aux revues pour définir une politique de données' (2021) with the 'Ouvrir la science!' logo. A teal circle highlights the cover. To the right, a text block explains the document's purpose for editorial committees.

Ouvrir la science! LA SCIENCE OUVERTE NOS ACTIONS RESSOURCES ACTUALITES QUI SOMMES NOUS ?

En pratique

Recommandations aux revues pour définir une politique de données

2021 **Ouvrir la science!**

Le document est destiné à des comités éditoriaux qui souhaitent mettre en place, pour leurs revues, une politique de données. Celle-ci définit ce que la revue attend de ses auteurs et autrices en matière de gestion et de diffusion des données liées à ses publications. Les recommandations portent sur sept éléments à prendre en compte dans la politique.



Lier des données et des publications

État des lieux des pratiques

A	B	D	E
Titre revue	URL OEJ	URL Nakala	Type de données
Études africaines	Sorcellerie et contre-sorcellerie : un réajustement	néant	texte "nakalam"
	Possession par esprit tromba et reconfiguration	néant	texte "zanakalahy"
Études rurales	Les voix des ancêtres et les voies du développement	néant	texte "zanakalahy"
	La fabrique des mondes insulaires	néant	texte "zanakalahy"
Gallia préhistoire	Recommandations aux auteurs	néant	texte "L'entrepôt par défaut est Nakala"
	Raccomandazioni per gli autori	néant	texte "L'archivio dati predefinito è Nakala"
	Recomendaciones para los autores	néant	texte "El repositorio predeterminado es Nakala"
Histoire & mesure	Recommandations aux auteurs	néant	texte "par ex. Nakala en sciences humaines"
	À l'épreuve du temps	https://nakala.fr/10.34847/nkl.1430qth3#898ca	dataset
	Les topic models au service de l'histoire d'un continent	https://nakala.fr/10.34847/nkl.b1ecn1u5#968c5	dataset ?
Journal de la société des américanistes	Broken words, furious wasps. How should we	https://api.nakala.fr/data/11280%2F4f602cec/e	son
La Révolution française	La médiation des dubashes	néant	texte "Kanakalatha"
Mappemonde	Diffusion globale et fonctionnements locaux : le cas de la France	https://www.nakala.fr/data/11280/7239e0dd	dataset
Médiévales	L'édition numérique du Livre de justice et de plénitude	http://iosticeetplet.huma-num.fr/	texte "extensions NAKALA import (...) NAKALA"
Nouvelles de l'archéologie	https://journals.openedition.org/nda/3800		
	E-STAMPAGES : archivage et publication en ligne	néant	texte "entrepôt OAI PMH par le biais de Nakala"
Nuevo Mundo	El imperio del Gaviotín Chico. Política y naturaleza	https://nakala.fr/10.34847/nkl.c25fj58m	vidéo (2) + photo (2)
Préhistoires méditerranéennes	Recommandations aux auteurs	néant	texte "a revue privilégie des entrepôts instit"
	Modélisation de l'extinction de Néandertal	https://nakala.fr/10.34847/nkl.fce42542	dataset MeN-Script R



Merci de votre attention

- Sandra Guignonis, Directrice adjointe éditoriale
- OpenEdition <https://www.openedition.org/>

