



Les différentes facettes du lien entre les données et les publications

De l'étude des pratiques à la mise en oeuvre

Elodie Faath, chargée de projet COMMONS

Sandra Guigonis, directrice adjointe éditoriale OpenEdition

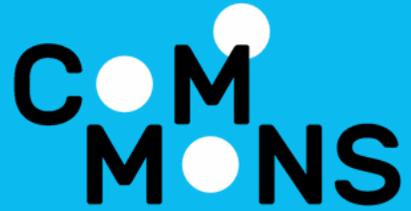
Nicolas Larrousse, directeur adjoint Huma-Num

Journées annuelles du réseau Mate-SHS, 15 et 16 mai 2025



Le projet COMMONS 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





Introduction et contexte

ÉquipEx+ COMMONS – Consortium de moyens mutualisés pour des services et données ouvertes en SHS

COMMONS est un équipement qui construit un **environnement numérique unifié** accessible au travers d'une offre de **services liant données et publications**. Il s'inscrit pleinement dans les logiques et dans les objectifs de mise en œuvre des principes de la science ouverte dans le domaine des SHS

3 infrastructures de recherche nationales



Les cibles de COMMONS :

- les communautés scientifiques
- les éditeurs et des producteurs de données
- les utilisateurs académiques
- les instances de financement de la recherche

5 établissements

- AMU
- AU
- CNRS
- EHESS
- Université de Caen

- les instances de pilotage des politiques scientifiques
- les médiateurs vers la société
- les acteurs socio-économiques et les citoyens

ÉquipEx+ COMMONS – COnsortium of Mutualised Means for Open data & Services for SSH

Le projet s'articule autour de 4 éléments :

- Mettre en place une architecture logicielle modulaire mutualisée :
 - Définition de l'architecture générale
 - Mise en place de l'authentification commune (Single Sign On)
 - Développement des composants (Components)
- Structurer et héberger de manière innovante :
 - Sécurisation des données – Mise en place de l'infrastructure sécurisée (Data security)
 - Mise en œuvre de la standardisation et de l'interopérabilité (Standard and interoperability)
- **Lier données et publications :**
 - Etablir les liens données-publications (Citation and crosslinking)
 - Intégration des données dans les publications (Representation and integration)
 - Cas d'usages – publications « complexes » (Use cases - Scholarly publishing)
- Amplifier et fluidifier l'accès aux services :
 - Mise en place côté utilisateurs des interfaces communes – (Definition of COMMONS users experience and deployment of seamless user interfaces)
 - Formation des acteurs/opérateurs et des utilisateurs (Training and Networking)
 - Retour sur les usages (Usage Observatory)

EquipEx+ COMMONS et HALiance

EquipEx + HALiance : refonte des services cœur de HAL et un alignement sur les recommandations et les critères d'excellence définis par l'ESR français et dans le cadre des initiatives internationales en faveur de la science ouverte.

Partenariat COMMONS /HALiance établi dès la première année de l'EquipEx+ :

- participation réciproque aux instances de gouvernance des équipements
- pratiques et savoir-faire
- mutualisation de développements de logiciels

État des lieux des pratiques & usages

Observations depuis les plateformes d'OpenEdition

+1,3 million de documents, dont près de 98 % en accès ouvert



29 revues candidates (en 2024)
649 revues en ligne, + 350 000 unités éditoriales
87 % des revues et **98 %** des contenus en accès ouvert



282 soumissions par mois (en 2024)
57 058 événements scientifiques en ligne
100 % des annonces en accès ouvert



16 candidatures par semaine (en 2024)
8 118 blogs, **5 038** au catalogue, + 640 000 billets
100 % des billets en accès ouvert



39 collections soumises (en 2024)
134 éditeurs, + 280 000 unités éditoriales
86 % des livres en accès ouvert

- **Nature du corpus :** textuel
- **Volume :** 1 314 729 documents (31.12.2024)
- **Couverture spatiale :** monde
- **Couverture temporelle :** 1999 – 2025 (numérique) | 1900-2025 (imprimé)
- **Formats :** HTML (en ligne) | Dublin Core, METS, XML-TEI (dépôt OAI-PMH)
- **Langues :** français (69 %), anglais (11.6 %), allemand (10.1 %), espagnol (4.2 %), portugais (2.2 %), italien (1.9 %), autres langues (0.9 %)
- **Stockage :** BDD au Centre de Calcul de l'In2p3, dépôt OAI-PMH
- **Accès :** Navigation avec Search + métadonnées moissonnables
- **Licence :** CCO pour les métadonnées + politiques de diffusion & droit d'auteur

Un état des lieux des liaisons entre OpenEdition et Nakala (actualisation 2025)

Tableau des résultats de l'analyse

Mention Nakala sur OpenEdition

Plateformes OEJ, OEB, Hypothèses

Plateforme	Nov 2022	Sept 2024	Différence
OEJ	134	283	149
OEB	71	90	19
Hypothèses	360	428	68
Total	565	801	236
			+42%

Embed sur OpenEdition

Éléments distants intégrés par embed

Plateforme	Nov 2022	Sept 2024	Différence
OEJ	2087	2274	187
OEB	38	43	5
Total	2125	2317	197

Un état des lieux des liaisons entre OpenEdition et Nakala (actualisation 2025)

Tableau des résultats de l'analyse

A	B	C	D	E	F
Titre revue	URL OEJ	Nombre	URL Nakala	Type de données	Format des données
	Recomendaciones para los autores	1	néant	texte 'El repositorio predeterminado es Nak	chaîne de caractères
Histoire & mesure	Recomendaciones aux auteurs	1	néant	texte 'par ex. Nakala en sciences humaines	chaîne de caractères
	À l'épreuve du temps	4	https://nakala.fr/10.34847/nkl.1430qth3#898ca	dataset	csv + png
	Les topic models au service de l'histoire d'un	10	https://nakala.fr/10.34847/nkl.b1ecn1u5#968cf	dataset ?	png + pdf
Journal de la société des américanistes	Broken words, furious wasps. How should we	2	https://api.nakala.fr/data/11280%2F4f602cec/e	son	mp3
La Révolution française	La médiation des dubashes	13	néant	texte 'Kanakalatha'	chaîne de caractères
Mappemonde	Diffusion globale et fonctionnements locaux	1	https://www.nakala.fr/data/11280/7239e0dd	dataset	?? (exploitable par un script R)
Médiévales	L'édition numérique du Livre de Justice et de pl	2	http://josticeetplet.huma-num.fr/	texte 'extensions NAKALA import (...) NAK/	base de données prosopographiques
Nouvelles de l'archéologie	https://journals.openedition.org/nda/3800	?			
	E-STAMPAGES : archivage et publication en li	1	néant	texte 'entrepôt OAI PMH par le biais de Nak	chaîne de caractères
Nuevo Mundo	El Imperio del Gaviotín Chico. Política y natura	15	https://nakala.fr/10.34847/nkl.c25fj58m	vidéo (2) + photo (2)	mp4 + jpeg
Préhistoires méditerranéennes	Recomendaciones aux auteurs	1	néant	texte 'la revue privilégie des entrepôts instil	chaîne de caractères
	Modélisation de l'extinction de Néandertal	1	https://nakala.fr/10.34847/nkl.fce42542	dataset MeN-Script R	text
	Modelling Neanderthal Extinction - abridged	1	https://nakala.fr/10.34847/nkl.fce42542	dataset MeN-Script R	text
	Raccomandazioni agli autori	1	néant	texte 'la rivista privilegia archivi istituzional	chaîne de caractères
	Recomendaciones a los autores	1	néant	texte 'la revista favorece los repositorios in	chaîne de caractères
	Photographic coverage of the Trou de la Fécla	1	https://nakala.fr/10.34847/nkl.f7ab196j	texte rapport d'expertise	texte
Humanités numériques	D'un projet d'observatoire démographique à l'é	5	https://www.nakala.fr/	texte 'nous mobilisons les outils de la TGIR	chaîne de caractères
	Thesaurus Exemplorum Medii Aevi : une base	8	néant	texte 'Nakala, fournit « un entrepôt sécurisé	chaîne de caractères
RFSIC	Corpus d'enquêtes sur les pratiques d'informa	5	néant	texte 'Collection Omeka / Exposition des dx	chaîne de caractères
Sciences de la société	Donner à voir et à penser les contenus autopo	1	néant	texte 'Pensées pour être hébergées par NaI	chaîne de caractères
Sociologie du travail	Coopérer avec les éléphants dans le Nord-Est	4	néant	texte 'Nakalan'	chaîne de caractères
Tracés	Ouvrir les données de la recherche ?	1	néant	texte 'service Nakala mis en place par la TC	chaîne de caractères

Un état des lieux des liaisons entre OpenEdition et Nakala (actualisation 2025)

Tableau des résultats de l'analyse

Titre revue / livre	Nombre de données identifiées dans NAKALA	Lien donnée Nakala -> OE : champ utilisé	Lien donnée Nakala -> OE : forme utilisée	Lien collection -> OE : champ utilisé	Lien collection -> OE : forme utilisée
Sociologie	25	dcterms:description	Texte et url : "... , Sociologie, 2024-1 (vol. 15), https://www.cairn.info/revue-sociologie.htm "	Titre et dcterms:description	texte et URL
Anthropologie & Santé	40	pas de lien	/	Titre et dcterms:description	texte et URL
Tipa : travaux interdisciplinaires sur la parole et le langage	49	pas de lien	/	Titre et dcterms:description	Texte
Ateliers d'anthropologie	67	dcterms:description	Texte : "... numéro 54-55 des Ateliers d'anthropologie"	Titre et dcterms:description	Texte

Une grille de description

Proposition de grille pour décrire les cas d'usages

Intitulé du cas d'usage			
URL Publication			
URL Données			
Champs disciplinaires couverts			
Données			
Description des données			
Métadonnées associées			
Volumétrie			
Formats			
Publication			
Type de publication			
Format de publication			
Structuration de la publication			
Modalités			
Modalités de liaison			
Modalités de visualisation			
Modalités d'usage(s)			
Pratiques / Mise en oeuvre			
Limites			
Fonctions/outils souhaitables			

Cas avancé avec données présentes dans Nakala

Intitulé du cas d'usage	Une revue expose des données tabulaires		
URL Publication	https://journals.openedition.org/histoiremesure/14969		
URL Données	https://nakala.fr/10.34847/nki.1430qth3		
Champs disciplinaires couverts	Études des sciences, Histoire des sciences, Histoire, Méthodes quantitatives		
Données			
Description des données	Collection "Séries textuelles chronologiques" > Dépôt "Textométrie de la correspondance de Charles VIII" > 15 fichiers. Citation Deruelle, Benjamin ; Lamassé, Stéphane (2021) «Textométrie de la correspondance de Charles VIII» [Data paper] NAKALA. https://doi.org/10.34847/nki.1430qth3		
Métadonnées associées	Titre, Auteurs, Date de création, Licence, Type, Mots-clés, Langue Licence (CC-BY-4.0)		
Volumétrie	15 fichiers, poids total ?'		
Formats	11 .csv + 4 .png		
Publication			
Type de publication	Article de revue		
Format de publication	HTML	Fac-similé PDF	ePub non généré
Structuration de la publication	XML-TEI natif (chaîne Métopes)		
Licence	Tous droits réservés		
Modalités			
Modalités de liaison	<ul style="list-style-type: none"> 2 mentions textuelles sur OEJ dont 1 lien invalide vers Nakala (404) Pas de mention de l'article sur Nakala 		
Modalités de visualisation	<ul style="list-style-type: none"> 6 tableaux texte + 4 images dans le document chargé sur OEJ Tableaux manipulables + images visualisables sur 		

Pratiques / Mise en oeuvre / Observations	<ul style="list-style-type: none"> Depuis OEJ <ul style="list-style-type: none"> pas d'accès direct aux données présentes sur Nakala, informations incomplètes (absentes sur images) traitement assuré par l'éditeur typé comme Article (de recherche) Depuis Nakala <ul style="list-style-type: none"> pas d'information sur la publication liée granularité des informations dépôt assuré par les auteurs Signalé comme Data paper Contradiction possible sur les droits Cas d'une revue tripartite Cairn https://www.cairn.info/revue-histoire-et-mesure-2021-2-page-163.htm
--	---

Perspectives pour le traitement du lien données-publications dans COMMONS

- Partir d'exemples
- Qualifier et quantifier les besoins, notamment disciplinaires
- Décrire les cas d'usages
- Identifier les standards et les workflows
- Élaborer des prototypes
- Dimension partenariale autour des pratiques disciplinaires

Définir les données pour les SHS

« La synthèse la plus exhaustive consiste à dire que les données sont des représentations d'observations, d'objets ou d'autres entités qui servent à mettre en évidence des phénomènes à des fins de recherche. »

Borgman, Christine L. « 2. Qu'est-ce qu'une donnée ? ». *Qu'est-ce que le travail scientifique des données ?*, traduit par Charlotte Matoussowsky, OpenEdition Press, 2020, <https://doi.org/10.4000/books.oep.14732>, traduction de : Borgman, Christine L., 2015. *Big Data, Little Data, No Data - Scholarship in the Networked World*. The MIT Press. ISBN 978-0-262-02856-1

Observer les usages

« The study found that lack of time and resources, inconsistent data practices, technical training gaps, labour intensity and difficulties in finding data were the most challenging. Participants revealed a number of enabling factors in data reuse as well, and chief among them were collaboration and autodidacticism as a feature of DH. The results indicate a gap between data reusers and data sharers - low rates of sharing reduce the amount of findable and accessible data available for reuse. »

Harper, L. M. (2023). *Data reuse among digital humanities scholars : A qualitative study of practices, challenges and opportunities* [Université d'Ottawa / University of Ottawa]. <http://hdl.handle.net/10393/45445>

Qualifier et quantifier les besoins

Le cas des Data papers / Data journals

Bizien, L., Cohoner, V., Edmond, F., Natal, A., & Peraldi-Mittelette, P. (2024). *Rapport d'Enquête. Création d'une revue d'articles sur des jeux de données Data Journal SHS* (p. 54). MSH Lorraine. <https://hal.univ-lorraine.fr/hal-04541094>

Decoupes, R. (2024). *Analyse de l'impact des data papers de l'UMR TETIS*. UMR TETIS, 500 rue Jean-François Breton, 34000 Montpellier. <https://hal.science/hal-04428092>

Gassama, M., Szabo, D., Tang, C., & Bravo, S. (2024). *Analyse de l'enquête sur les pratiques des scientifiques en matière de publication de data paper* (p. 28 p.). INRAE. <https://doi.org/10.17180/vrh7-r606>

Kindling, M., & Strecker, D. (2022). *List of data journals* (Version 1.0). Zenodo. <https://doi.org/10.5281/zenodo.7082126>

Dimension partenariale autour des pratiques disciplinaires

Traitement des images avec InVisu/INHA

- le projet PerVisum (FNSO 3)

PerVisum

Explorer les formats éditoriaux mettant l'image au cœur de la démonstration scientifique en utilisant les potentialités des technologies IIIF



Dimension partenariale autour des pratiques disciplinaires

Analyse de corpus avec *Cybergegeo*

- CybergegeoNetworks
- CybergegeoNetworks2



Kosmopoulos, C., Aveline, N., & Lavie, É. (2023). CybergegeoNetworks 2 : Une application d'apprentissage profond au service de la publication scientifique. *Cybergegeo: European Journal of Geography*. <https://journals.openedition.org/cybergegeo/40639>



Le rôle des infrastructures dans la création et la gestion du lien données publications

Le lien données-publications : plusieurs visions et utilisations

- Une référence à un jeu de données dans une publication
- Un lien entre deux objets numériques, article et jeu de données (eg. interviews filmées et transcriptions) ... par exemple en liant leurs identifiants en identifiants les versions concernées et en gérant l'évolution de la relation dans le temps
- Un lien dans la publication vers des illustrations par exemple des images, un tableau ou autre ... qui sont montrés dans la publication et constituent des composants de la publication
- Un mélange des trois visions précédentes ...

COMMONS vise à répondre à ces différents besoins

Le lien données-publications : Une simple citation dans une publication



Journal of French
Language Studies

Article contents

Résumé

References

J'avoue, j'ai un compte en Suisse: *Étude d'un marqueur parenthétique à la mode*

Published online by Cambridge University Press: 19 February 2016

JULIETTE DELAHAIE

Show author details ▾

Article Metrics

Get access

Share

Cite

Rights & Permissions

CORPUS ET BASES DE DONNÉES

Corpus d'Orléans, ESLO1 et ESLO2, Université d'Orléans: <http://eslo.humanum.fr> [Google Scholar](#)

Corpus de Français Parlé Parisien des années 2000, Université Paris 3-Sorbonne nouvelle: <http://cfpp2000.univ-paris3.fr/> [Google Scholar](#)

Corpus de Langues Parlées en Interaction (CLAPI), Université de Lyon 2: <http://clapi.univ-lyon2.fr/> [Google Scholar](#)

Frantext, ATILF: www.frantext.fr [Google Scholar](#)

Résumé

J'avoue apparaît dans des usages peu traditionnels en français. *Moi j'avoue l'opéra ce n'est pas ma passion*, ou bien, nota exprimer un accord : *Tu as vu il fait beau !/ J'avoue*. S'agit Au contraire, on montrera que *J'avoue* s'inscrit dans des

La citation peut être faite par un simple référence textuelle, un identifiant ou un mélange des deux
Cette citation est utilisée par des outils comme ceux de Google qui associent Data Search & Scholar
mais le résultat peut être aléatoire

Le lien données-publications entre HAL et NAKALA Une préfiguration de ce qui sera fait dans COMMONS



Suite d'un POC réalisé dans le cadre du projet Européen EOSC-PILLAR (<https://hal.science/hal-03170550>)

- Une vision de la création du lien données-publications au niveau des systèmes d'information des infrastructures entre deux objets numériques
- Choix de s'inscrire dans l'écosystème national en utilisant le protocole COAR Notify pour communiquer entre les infrastructures
- Travail en commun sur les possibilités de publication des liens en utilisant le standard Scholix
- Echanges sur les modalités de création des liens dans les deux plate-formes

La mise en oeuvre du lien entre HAL et NAKALA servira de socle à ce qui sera réalisé dans COMMONS - actuellement en pré-production



Métopes: modélisation du lien données publications en XML-TEI Commons Publishing

- Répondre aux besoins des plateformes (dans leurs échanges et dans l'objectif "Scholix") sans alourdir le modèle XML-TEI Commons Publishing
- Répondre aux besoins des éditeurs (dans l'encodage des contenus édités) en prenant en compte l'unité "Document" liée et les traitements nécessaires
- Définir les vocabulaires normés et les standards communs
- Assurer la compatibilité avec JATS

Propositions visant à la création de publications "complexes"

Des styles de liaisons qualifiées

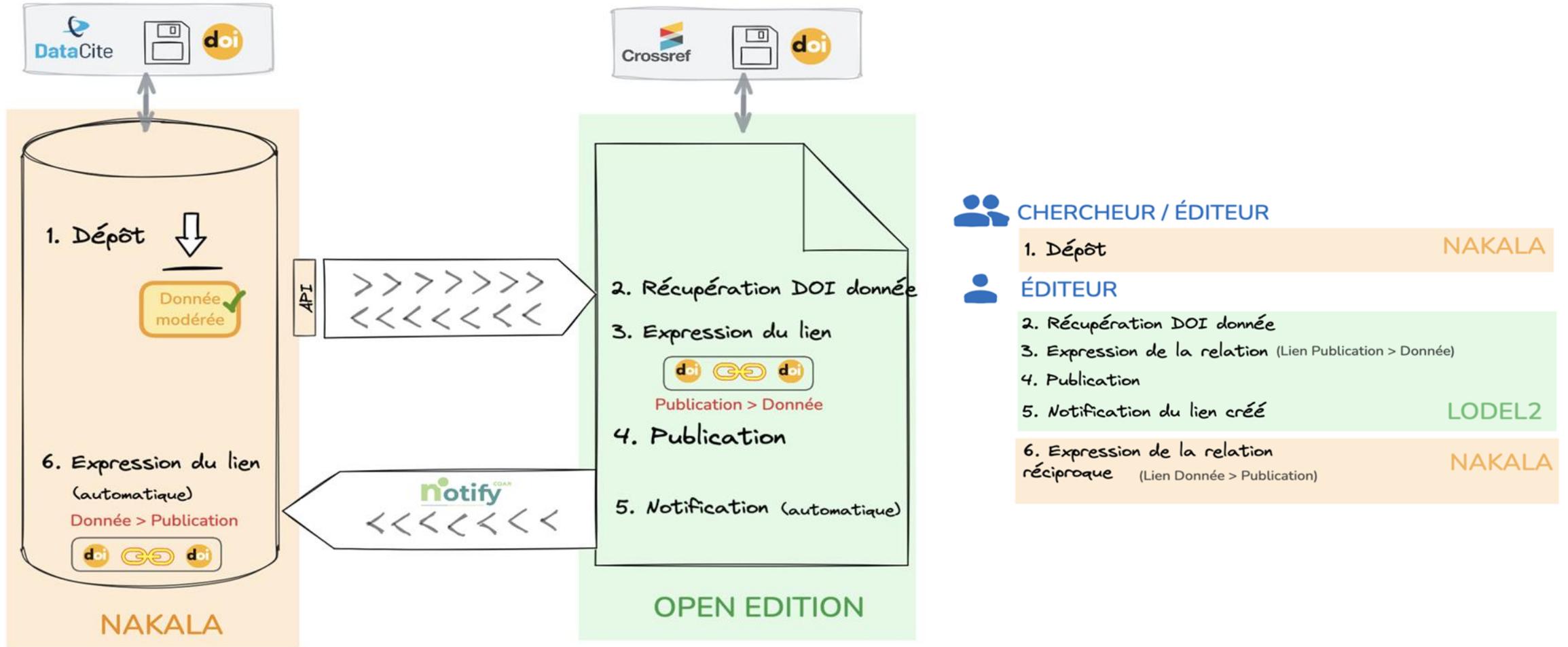
- Données associées
- Publications associées

Dans le but de produire des structures TEI annotables susceptibles de permettre d'alimenter des requêtes vers les entrepôts

Une approche d'éditeur

The screenshot shows a web interface for a document. At the top, there is a navigation bar with a home icon, a search icon, and links for 'PRÉSENTATION', 'NUMÉROS', 'INDEX', and 'CONTRIBUER'. Below this, a 'TEXTE INTÉGRAL' button is visible. On the left side, a table of contents is displayed with a 'Lien' column, a 'Texte' column, and a 'Classe' column. The items listed are: 'Résumé', 'Résumé lang alt', 'Notes préliminaires', 'Chapô', 'Épigraphie', 'Sources', '* sources', 'Remerciements', 'Dédicace', '* Partenaires inst.', '* Données associées', and '* Publications associées'. On the right side, there is a section titled 'Données associées à cet article :'. It contains a link to a DOI: <https://doi.org/10.34847/nkl.98773264>. Below this, there is a paragraph describing the data set: 'Le jeu de données contient 14 tableaux (chacun aux formats ODT et XLS) présentant les variables de population des 14 recensements de la population égyptienne opérés par le CAPMAS (Central Agency for Public Mobilisation and Statistics) entre 1882 et 2017, ainsi qu'un tableau présentant la liste exhaustive des variables de l'ensemble des recensements de 1882 à 2017 (aux formats ODT et CSV)'. Below this, there is another section titled 'Ressource numérique liée à cet article :'. It contains a paragraph describing the resource: 'Cedej- Capmas, application de visualisation des données socio-économiques CAPMAS (Central Agency for public mobilisation and statistics)/ CEDEJ (Centre d'études et de documentation économiques, juridiques et sociales) : <https://www.cedejcapmas.org/>.' At the bottom, there is a list of footnotes: '1 Troisième pays du continent africain après le Nigeria et l'Éthiopie et premier pays d'Afrique du Nord par son poids démographique (97 millions d'habitants en 2017), l'Égypte est l'un des rares pays au sud de la Méditerranée à disposer d'un recensement intégral et régulier de sa population depuis le XIX^e siècle (Alleaume et Fargues 1998). C'est à l'époque de Méhémet Ali Pasha que l'Égypte effectue le premier comptage de la population, plus précisément en 1846. Elle prend à partir de ce moment un élan moderniste, poursuivie par la présence britannique à partir de la fin du XIX^e siècle.' and '2 Les deux premiers recensements publiés datent de 1882¹ et 1897. Après

Lien données publications : définition d'un workflow



Vue de synthèse pour l'établissement de liens entre les plate-formes de OpenEdition et de NAKALA

Le rôle des infrastructures dans la création et la gestion du lien données publications

- Repérer les pratiques et les besoins des communautés
- Proposer des passerelles entre les plate-formes pour créer et gérer les liens "souplement"
- Assurer la bonne cohérence des liens entre les plate-formes dans le temps (suppression, changement de version etc.)
- Diffuser l'information via des standards et des pratiques reconnus (DOIs, Relations DataCite, SCHOLIX, référentiels ...)
- **Former les utilisateurs**



Merci !