



HAL
open science

Retour d'expériences : publication de données dans l'entrepôt de données SHS Nakala

Louise-Anne Charles, Françoise Acquier

► To cite this version:

Louise-Anne Charles, Françoise Acquier. Retour d'expériences : publication de données dans l'entrepôt de données SHS Nakala. 2022. hal-04292102

HAL Id: hal-04292102

<https://hal.science/hal-04292102>

Submitted on 17 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

Retour d'expériences : publication de données dans l'entrepôt de données SHS Nakala

 labedoc.hypotheses.org/10424

Louise-Anne Charles

L'année scolaire 2021-2022 du centre de documentation du CRESSON a été dédiée à **la gestion des données**, en s'appuyant sur deux projets ANR. Pour ce type de projet, la gestion des données est devenue obligatoire, notamment à travers la rédaction du plan de gestion de données attendue par l'ANR. Mais que faire pour les projets dont cet aspect n'est pas obligatoire ou ceux qui sont terminés ?

C'est grâce aux ateliers du réseau Scripto du Réseau National des Maisons des sciences de l'homme, et à une présentation de leur travail sur les **plans de gestion de données rétrospectifs**, que nous avons commencé à nous intéresser aux projets qui sont terminés. Cela permet d'avoir accès à des données stabilisées, documentées, qui peuvent faire l'objet d'un partage.

En octobre 2021, lors d'une journée d'études (que vous pouvez découvrir ici), nous avons été marquées par la qualité esthétique et scientifique des Itinéraires et des cartes produites dans le cadre du projet de recherche-crédation Les Ondes de l'Eau : mémoires des lieux et du travail dans la vallée de la Romanche (AAU-CRESSON, Regards des Lieux).

Par exemple, cette carte des mémoires sensibles de la Romanche **longue de 6 mètres** ! Ou le travail d'enquête sous forme d'itinéraires formalisés **en livrets** qui retracent un parcours avec un habitant ou un professionnel:



Image de la journée d'études du projet les Ondes de l'Eau, Sylvie Laroche, ENSAG, 1/10/2021

Ces données ont résonné avec **une formation suivie en septembre 2021 sur l'entrepôt de données Nakala développé par Huma-Num et dédié aux SHS**. Les atouts de cet entrepôt par rapport à ces données sont :

- Une **visionneuse** qui permet de zoomer en conservant une grande qualité de l'image ;
- L'**attribution de DOI** aux données que l'on peut lier à un dépôt d'article sur HAL ;
- Le **partage des données** dans des publications web comme les billets de blog.

Un espace numérique qui semblait pouvoir faire la part belle aux aspects visuels des données tout en gardant leur qualité.

Nous avons alors contacté l'équipe du projet représentée par Laure Brayer, la coordinatrice et Ryma Hadbi, doctorante et membre active du groupe et nous leur avons proposé de les accompagner, en nous appuyant sur l'expérience acquise avec les projets ANR. Cela s'est fait selon plusieurs axes :

- **Regrouper les données du projet sur un serveur sécurisé** du laboratoire et les trier. En effet, plusieurs espaces de stockage avaient été utilisés et des doublons de fichiers existaient ;
- **Valoriser** les itinéraires et les cartes produites sur Nakala ;
- Contacter les archives départementales dans l'idée d'**archiver des données de recherche**.

Ce billet est dédié à **la publication des données sur Nakala** (sans dépôt par lot) : documentation, préparation, publication et valorisation. Il s'agissait de prendre en main la plateforme avec un nombre assez réduit de données (5 itinéraires et 3 cartes) pour explorer ce qui était possible de faire. L'idée est maintenant de capitaliser ce qui a été appris, sous la forme d'un guide et de ce présent billet, pour que d'autres projets de recherche du laboratoire puissent investir la plateforme et que les réponses que nous avons eues puissent profiter à tous et toutes.

Ce travail de publication des données et de documentation s'est fait grâce à l'aide de **Sara Tandar qui fait partie du pôle « DONNÉES et coordination de l'accompagnement utilisateur » d'Huma-Num** et qui a été très disponible ces derniers mois pour nous répondre. **Un grand merci à elle !**

Sommaire

I. Processus

II. Les choix éditoriaux

III. Le travail sur la documentation

IV. Le dépôt et la publication

V. Nakala et son lien avec d'autres plateformes

VI. Les problèmes rencontrés et les solutions

VII. Conclusion

I. Processus

Le choix des données et leur organisation en collection ont été déterminés lors de plusieurs réunions entre documentalistes et chercheurs. La documentation des données s'est faite sur un tableur en suivant l'ordre des métadonnées de Nakala et a servi de base de discussion.

II. Les choix éditoriaux

1) La construction des collections

Deux possibilités s'offraient à l'équipe de recherche, comme il ne s'agissait que d'un nombre réduit de données :

- Soit regrouper dans une seule et unique collection toutes les données du projet ;
- Soit distinguer les itinéraires des cartes en les mettant dans deux collections différentes, en plus de faire la collection dédiée à toutes les données du projet.

Cette dernière solution a finalement été privilégiée par l'équipe car, même si une collection ne bénéficie pas d'un DOI, il est toujours possible d'envoyer son URL. Cela leur permet de renvoyer vers un ensemble de données. De plus, nous avons pu ainsi décrire avec plus de

finesse chacune des collections.

2) Le choix de la licence des données

Ces données ont déjà été montrées plusieurs fois (lors de manifestations) et l'objectif était de faciliter leur visualisation et leur communication. Cependant, elles n'ont pas encore fait l'objet de publication. L'équipe a alors décidé d'opter pour la publication des données, sans embargo, mais sous une licence « tous droits réservés ».

L'idée étant de **faire évoluer ce choix vers une licence plus ouverte au réusage lorsque leur travail d'analyse sera finalisé** dans des publications.

III. Le travail sur la documentation

1) De l'exposition aux métadonnées

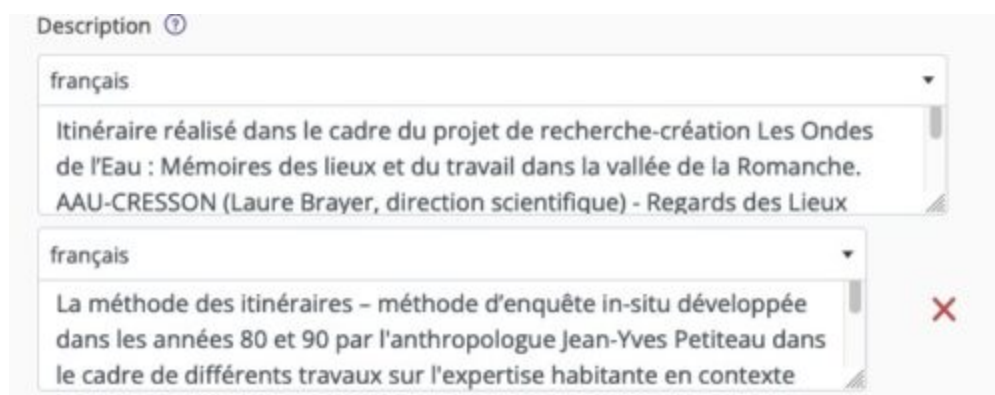
Le projet Les Ondes de l'Eau a fait l'objet d'une exposition « Cartographier les paysages sensibles » à l'École Nationale supérieure d'architecture de Grenoble du 1^{er} au 15 octobre 2021 accompagné de travaux d'étudiants. Dans ce cadre-là, des panneaux ont été réalisés pour décrire les supports affichés. Certains de ces textes ont ainsi pu alimenter les descriptions des données sur Nakala. Les livrets contenaient également des informations réutilisables.

2) La description d'une donnée en plusieurs étapes

Trois types d'information ont été mis en description de chaque donnée :

- Contexte du projet ;
- Résumé de la donnée ;
- Méthode de création de la donnée.

Nous avons choisi de dupliquer le champ description pour **mettre en valeur les méthodes** spécifiques utilisées.



Capture d'\u00e9cran des champs "description" dupliqu\u00e9s sur Nakala

3) Le problème du type de donnée

Pour rappel, l'un des types de données que nous souhaitons publier sur Nakala est un livret qui donne à voir un itinéraire avec un habitant ou un professionnel. Il y a à la fois des photos, une carte du trajet, des verbatims de l'enquêté.

Nakala propose de choisir dans une liste contrôlée le type de la donnée. Or aucun ne semblait correspondre, étant soit réducteur soit inapproprié.

La solution proposée par Sara Tandar a été d'**opter pour le type de document « autre » et d'utiliser les métadonnées supplémentaires pour préciser le type qui nous semblait correspondre**, et ne perdre ainsi aucune information. Nous avons utilisé la propriété Dcterms:subject et dupliqué autant de fois que nécessaire.



Capture d'écran des métadonnées supplémentaires "dcterms:subject" sur Nakala

4) Les mots-clés

Pour les mots-clés, nous nous sommes appuyés sur deux référentiels, [Archirès](#) et Rameau grâce à l'interface [IDRef](#), pour compléter ce champ.

Astuce : pour pouvoir rentrer une série de mots-clés en une fois, il suffit de séparer les mots-clés par des ; !

5) Préciser le rôle des auteurs et ajouter le nom des personnes enquêtées

Lors de l'ajout des auteurs sur Nakala, **il n'est pas possible de préciser leur fonction** dans la production de la donnée.

Nous avons alors utilisé le type **Dcterms:creator** sous la forme : nom (fonction).

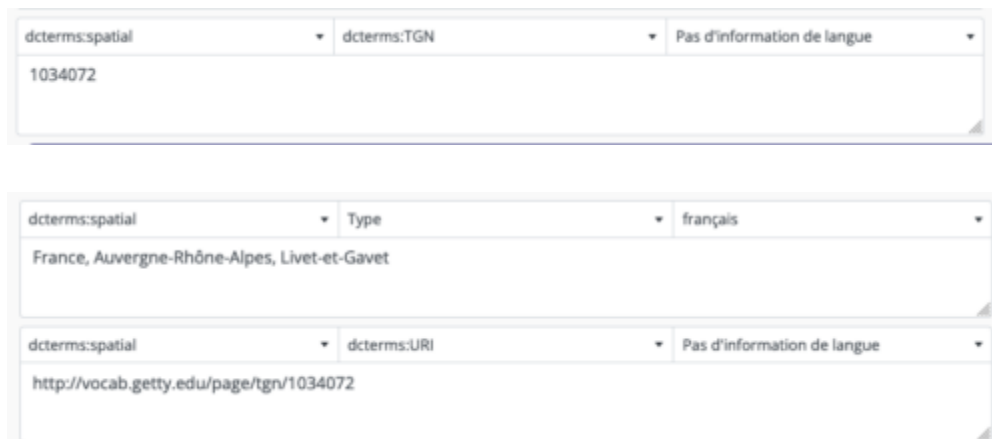
Le type Dcterms :contributor a lui servi pour mentionner les personnes enquêtées qui acceptaient de ne pas être anonymes.

6) Géolocalisation : utiliser le référentiel TGN

La dimension de l'espace est importante dans ce projet puisqu'il s'agit d'un travail sur un territoire, la basse vallée de La Romanche (38). Nous avons, pour l'exprimer sous forme de métadonnées, utilisé **le référentiel TGN (Getty Thesaurus of Geographic Names)**.

Le référentiel TGN permet à la fois d'avoir : **un identifiant, un lieu exprimé sous forme hiérarchique et une URI**. Nous avons conservé ces trois informations en dupliquant une propriété **dcterms:spatial**, où le type devait être précisé pour l'ID et l'URI.

Voici un exemple pour le lieu Livet-et-Gavet :



Captures d'écran des métadonnées supplémentaires "dcterms:spatial" sur Nakala

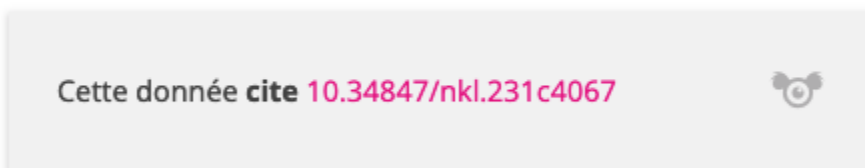
7) Autres métadonnées supplémentaires

Nous avons également précisé **le format des données** (jpeg, pdf) dans des métadonnées supplémentaires de propriété **dcterms:format**.

8) Relations entre les données

L'option de créer des relations entre les données s'est révélée intéressante : nous avons pu créer une relation entre deux cartes et une carte qui cite le travail fait pour un livret. Ces informations apparaissent sur la page de la donnée, de manière plus transparente. Une fois qu'une relation était créée sur une donnée, il n'y avait pas besoin de modifier l'autre donnée concernée.

Relations



Capture d'écran sur la relation à une autre donnée sur Nakala

IV. Le dépôt et la publication

D'après notre expérience, le plus long est le travail sur la documentation des données, alors que la création de listes, de collections et le dépôt de chaque donnée sont relativement rapides.

Nous avons suivi l'ordre conseillé par Nakala :

1. Créer son HumanID et demander un accès à Nakala ;
2. Créer les listes d'utilisateurs ;
3. Créer des collections et attribuer des droits ;
4. Déposer les données et publier.

V. Nakala et son lien avec d'autres plateformes

1) Visualiser sur Hypothèses

Pour visualiser une donnée publiée sur Nakala, il faut passer par un bloc « code court » (et non « contenu embarqué » ni « PDF viewer ») lors de la rédaction du billet. La procédure a été trouvée sur une [présentation produite par Nakala](#).

Nous voulions également centrer la donnée dans le billet, ce qui a donné le code suivant :

```
<p align="center"><iframe  
src="https://api.nakala.fr/embed/10.34847/nkl.2bad8uj6/23bd487bbee67766c37fcb183f9c161  
width="677" height="507"></iframe></p>
```

Brayer, Laure, Hadbi, Ryma, McOisans, juL, Séraudie, Christophe, Laroche, Sylvie, Thomas, Rachel. (2022) «Carte des mémoires sensibles de la Romanche – version finale» [Map] NAKALA. <https://doi.org/10.34847/nkl.2bad8uj6>

Le lien API est l'URL d'intégration qui se trouve sous la visualisation de la donnée. Ce code embarque également la visionneuse de Nakala, contrairement à ce code-là qui permet de n'avoir qu'**une image fixe** :

```

```



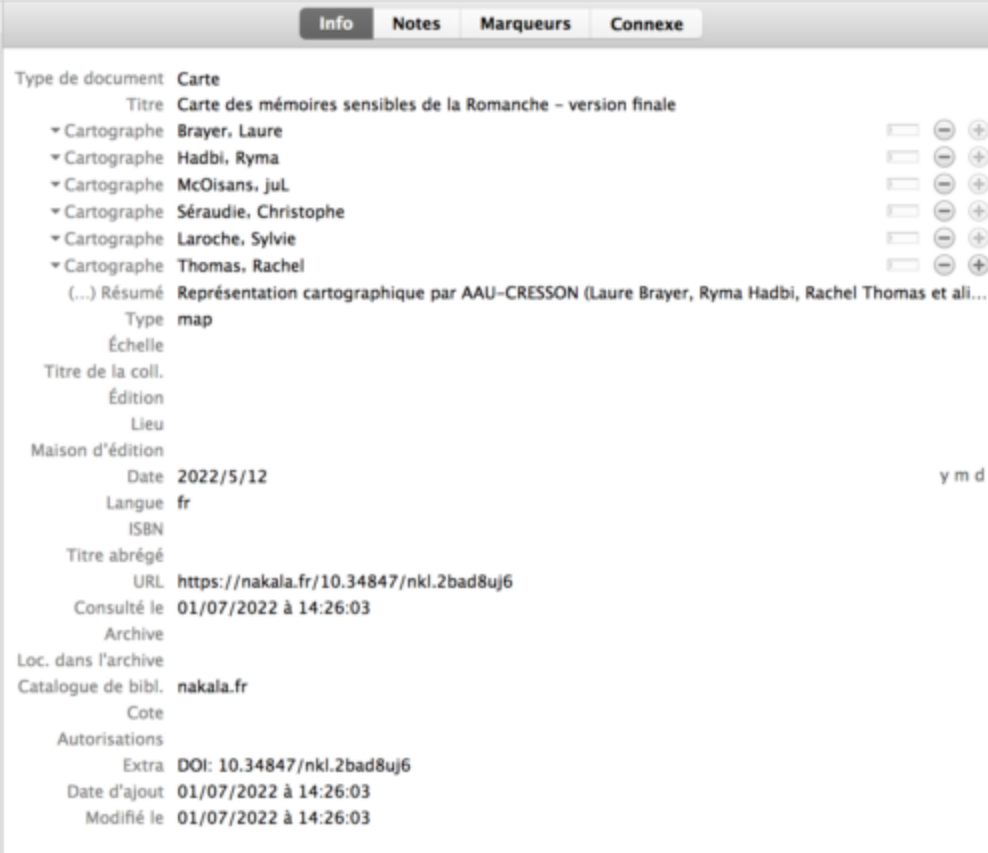
Brayer, Laure, Hadbi, Ryma, McOisans, juL, Séraudie, Christophe, Laroche, Sylvie, Thomas, Rachel. (2022) «Carte des mémoires sensibles de la Romanche – version finale» [Map] NAKALA. <https://doi.org/10.34847/nkl.2bad8uj6>

Cette fois, c'est l'URL de téléchargement qui est utilisée.

2) Ce qui est récupéré sur Zotero

En ajoutant une collection de données Nakala sur Zotero, Zotero le reconnaît comme article de revue et récupère le titre, le résumé, la date de mise en ligne.

Si nous essayons d'ajouter une donnée carte à Zotero, il le reconnaît comme « carte ». Il récupère les auteurs, le résumé, la date de mise en ligne sur Nakala (et non celle mise dans les métadonnées de la carte), la langue, l'URL. Le DOI est mis en extra.



The screenshot shows the Zotero 'Info' tab for a document. The document type is 'Carte'. The title is 'Carte des mémoires sensibles de la Romanche - version finale'. The authors listed are Brayer, Laure; Hadbi, Ryma; McOisans, juL; Séraudie, Christophe; Laroche, Sylvie; and Thomas, Rachel. The summary is 'Représentation cartographique par AAU-CRESSON (Laure Brayer, Ryma Hadbi, Rachel Thomas et ali...)'. The type is 'map'. The date is '2022/5/12'. The language is 'fr'. The URL is 'https://nakala.fr/10.34847/nkl.2bad8uj6'. The document was consulted on '01/07/2022 à 14:26:03'. The archive location is 'nakala.fr'. The extra field 'DOI' is '10.34847/nkl.2bad8uj6'. The document was added and modified on '01/07/2022 à 14:26:03'.

Capture d'écran de la notice d'une donnée sur Zotero

Pour un itinéraire, le type est « document » comme nous avons mis « autre ». Zotero récupère le titre, le résumé, l'auteur, la date de mise en ligne comme pour la carte. Par contre, il récupère à la fois les auteurs et les métadonnées dcterms :creator (sous la forme nom (fonction)).

Toutes les métadonnées de lieu et les mots-clés ne sont pas récupérés.

3) Relier la notice d'HAL aux données

Il est possible de lier la publication avec les données de recherche entreposées dans Nakala : il faut utiliser le champ « données associées ».



Capture d'écran des "données associées" sur une notice HAL

L'affichage est pour l'instant sobre mais une prévisualisation des données de Nakala dans HAL serait envisagé.

En revanche lors de la prochaine version de Hal à venir en septembre 2022, il n'y aura pas de type de document « jeu de données » dans HAL (source Agnès Magron CCSD).

4) ORCID

Sur ORCID, il est possible d'ajouter une collection sous forme de data set (jeu de données) en utilisant l'URL et en remplissant soi-même les métadonnées.

□ **Projet de recherche ODE - Les Ondes de l'Eau : Mémoires des lieux et du travail dans la vallée de la Romanche**

AAU-CRESSON; Région & DRAC Auvergne-Rhône-Alpes - Programme "Mémoires du XXe et XXIIe siècles"
2022-05-16 | Data set Show less detail

Language
French

Subtitle
Cartes et itinéraires

URL
<https://nakala.fr/collection/10.34847/nkl.2329tw44>

Citation (formatted-apa)
[Show citation](#)

Description

La vallée de la Romanche fait actuellement l'objet de profondes mutations. Depuis 2010, EDF a creusé la roche pour y installer une nouvelle centrale hydroélectrique et procède aujourd'hui au démantèlement des anciens barrages et usines édifiés à au fil du siècle passé. Au-delà du patrimoine matériel, qu'en est-il des mémoires collectives : comment celles et ceux qui vivent et travaillent dans la vallée peuvent-ils être témoins, acteurs et passeurs de la mémoire des lieux ? Comment les transformations actuelles viennent-elles questionner et ré-agencer leurs relations au paysage ? Quelles expériences sensibles passées, présentes et potentielles la Romanche offre-t-elle, complexifiant son image de simple ressource à exploiter ?

Dans le cadre du projet de recherche-crédation Les Ondes de l'Eau : Mémoires des lieux et du travail dans la vallée de la Romanche, AAU-CRESSON et Regards des Lieux ont arpenté la vallée à la rencontre de ses habitants. Cette collection retrace en partie les différentes recherches et créations liées au projet, à cette exploration plurielle de la vallée, à ces rencontres. Cette collection comporte 5 itinéraires réalisés auprès d'habitants ou de professionnels et des cartes sensibles des lieux.

Equipe
AAU-CRESSON // ENSAG : Laure Brayer (Direction scientifique), Ryma Hadbi, Sylvie Laroche, JuL McOisans, Rachel Thomas. Avec la collaboration d'Emmanuelle Pilon et de Christophe Séraudie.
Regards des Lieux : Laure Nicoladzé (Direction culturelle), Martin Debisschop, Jérémie Lamouroux.

Financement
Région & DRAC Auvergne-Rhône-Alpes - Programme "Mémoires du XXe et XXIIe siècles".

Country of publication
France

Added
2022-06-30

Last modified
2022-06-30

Capture d'écran d'un data set sur ORCID

Ou il est également possible de faire un ajout par DOI, le type de document est alors « autre ». ORCID récupère le titre, le résumé, la date (de mise en ligne), le lien et crée la citation. La citation bibtex peut être importée dans Zotero.

The screenshot shows a web interface for a document record. At the top, the title is "En remontant la Romanche. Itinéraire de Mathieu Grenier, un agent du Symbhi, le 5 janvier 2021, Les Clavaux". Below the title, there is a section for "2022 | Other" with the DOI "10.34847/NKLA5AE8Y33" and a "Show less detail" link. The "Language" is listed as "French". The "URL" is "https://nakala.fr/10.34847/nkl.a5ae8y33". The "Citation (bibtex)" section includes a detailed BibTeX entry: "@misc(https://doi.org/10.34847/nkl.a5ae8y33, doi = {10.34847/NKLA5AE8Y33}, url = {https://nakala.fr/10.34847/nkl.a5ae8y33}, author = {Brayer, Laure and McOisans, JuL and Laroche, Sylvie and {Sylvie Laroche (entretien, retranscription, édition)} and {Laure Brayer (Photographie , retranscription, édition)} and {JuL. McOisans (Retranscription)}}), keywords = {Histoires de vie, paysage de l'eau, histoire orale, Marche, Sens et sensations, Mémoires des lieux, chemin, piéton, passerelle, méthode des itinéraires, perception sensible, Romanche, Vallée de la (France), énergie hydraulique, aménagement du territoire, patrimoine industriel, tourisme, gestion du risque, Romanche, Vallée basse de la (France), matériaux de terrain éditorialisés, roman-photo, itinéraire), language = {fr}, title = {En remontant la Romanche. Itinéraire de Mathieu Grenier, un agent du Symbhi, le 5 janvier 2021, Les Clavaux}, publisher = {NAKALA - https://nakala.fr (Huma-Num - CNRS)}, year = {2022}}". Below the citation, there is a "Description" section: "Itinéraire réalisé dans le cadre du projet de recherche-crédation Les Ondes de l'Eau : Mémoires des lieux et du travail dans la vallée de la Romanche. AAU-CRESSON (Laure Brayer, direction scientifique) - Regards des Lieux (Laure Nicoladzé, direction culturelle). Après plusieurs annulations suite aux confinements, le RDV est enfin fixé. En empruntant le chemin réalisé par le Syndicat Mixte des Bassins Hydrauliques de l'Isère, de la centrale des Clavaux jusqu'à la passerelle himalayenne, Mathieu Grenier nous présente les stratégies urbaines à l'origine du projet. Aux détours de la discussion, il évoque sa perception du territoire." The "Added" date is "2022-07-01" and the "Last modified" date is "2022-07-01".

Capture d'écran d'un ajout de DOI sur ORCID

VI. Les problèmes rencontrés et les solutions

1) Ajouter un auteur en oubliant son ORCID lors de sa création

Nous avons ajouté un auteur sans son ORCID et nous avons voulu l'ajouter après-coup. Pour l'instant, **il n'est pas possible de modifier un auteur déjà créé.**

Le conseil qui nous a été donné est de créer une nouvelle fois cet auteur avec son ORCID. Deux formes apparaissent alors lorsque nous tapons son nom. **La deuxième forme est celle plus récente** avec l'ORCID. Nous avons repris les données créées par cet auteur pour lui attribuer la bonne forme.

Nakala a pour axe d'amélioration le développement de la base auteurs.

2) Impossible d'ajouter la donnée à une collection

Nous avons déposé les données avant de les publier afin de permettre une vérification par les membres de l'équipe projet. Or lorsque nous avons repris les données pour les publier, il n'était plus possible de les ajouter aux collections. Ces dernières n'apparaissaient plus.

Comme nous les avons préalablement déposées, il a fallu les publier avant de pouvoir les ajouter à des collections publiques.

3) Impossible de zoomer sur un PDF

Il faut publier les données pour avoir accès à toutes les possibilités de visualisation et de navigation, notamment pour les PDF. Pour avoir accès au zoom sur PDF, il suffit de cliquer sur « voir le fichier brut » (icône d'un œil).

Cependant, lorsque nous ajoutons un PDF provenant de Nakala sur Hypothèses par exemple, le zoom n'est pas disponible. Un exemple juste en dessous :

Brayer, Laure, Rachel, Thomas, et Pilon, Emmanuelle. (2022) «Sur les chemins et les ponts. Itinéraire des Jeunes Filles, le 19 juin 2019 à Rioupéroux» [Other] NAKALA.
<https://doi.org/10.34847/nkl.231c4067>

4) Contrôler l'ordre des mots-clés

Les mots-clés avaient été sélectionnés par l'équipe et un ordre d'importance, avait été privilégié.

Cependant, il est pour l'instant impossible de contrôler de manière fine cet ordre et **l'affichage des mots-clés est aléatoire.**

1^{ère} technique de contournement :

Un des moyens que nous avons utilisés pour contourner cela est de diviser en deux groupes les mots-clés : un avec ceux qui sont cruciaux et un autre avec ceux qui ne devaient pas s'afficher en premier.

En dupliquant le champ mots-clés, nous avons mis ce second groupe dans le deuxième champ. Ces deux groupes de mots-clés ne se mélangent pas. Nous avons ainsi pu mettre en avant les mots-clés les plus importants.

2^{ème} technique de contournement :

Une autre possibilité si l'ordre est crucial est de faire au compte-goutte. D'ajouter un ou deux mots-clés puis publier, modifier en ajoutant un ou deux autres et publier etc... L'ordre sera ainsi respecté mais c'est un procédé assez lent et répétitif.

VII. Conclusion

Avec ce billet dense, nous le concevons bien !, nous espérons pouvoir donner quelques clés sur la prise en main de Nakala, sur la documentation qu'il est possible de faire et sur ce que nous avons pu noter tout au long du parcours.

Il s'agissait d'un projet d'un genre nouveau pour le centre de documentation qui, même s'il a acquis de l'expérience dans le partage et l'archivage des données sonores, s'est confronté cette année à une multitude de typologie de données et à des nouveaux outils ! Ce travail a

confirmé **l'intérêt de la gestion des données pour les projets terminés** et souligné comment celle-ci pouvait **répondre à des besoins concrets**.

De plus, une réflexion nécessaire semble se dessiner sur **l'articulation à faire entre publications et données** : comment le stade de publication influence le partage mais également comment les outils qui se développent communiqueront à l'avenir, à l'instar d'HAL et de Nakala. **L'enjeu est d'également communiquer et d'accroître la visibilité de ces données pour** les chercheurs et les usagers.

Source de l'image mise en avant : gallica.bnf.fr / INSEP, "Foss bat le record du saut à la perche", Le miroir des sports, 1920 (mis en ligne en 2013). Disponible à l'adresse : <https://gallica.bnf.fr/ark:/12148/bpt6k65253135/f10.item#>

Cite this article as: Louise-Anne Charles et Françoise Acquier, "Retour d'expériences de publication de données : l'entrepôt de données SHS Nakala," in *Lab & doc*, 22/07/2022, <https://labedoc.hypotheses.org/?p=10424>.
