

Vous avez dit
« CURATION » ?
SOS
DOCUMENTALISTES
vous répond !

De la **PIERRE** au **JOYAU** :
les **MÉTADONNÉES** au service
de la **QUALITÉ** des **DONNÉES**

couperin.org

Consortium Unifié des Établissements Universitaires et
de Recherche pour l'Accès aux Publications Numériques

01 décembre 2022
GTSO données

Hélène TISSERAND
et Rachel Prost
Documentalistes Atelier de la donnée dat@UBFC



- 1 Le métier de **DOCUMENTALISTE** dans le contexte de la **SCIENCE OUVERTE** et de l'**OPEN DATA**
- 2 La **GESTION** des **MÉTADONNÉES** à l'**ATELIER** de la **DONNÉE** dat@UBFC
- 3 Des **EXEMPLES** concrets de l'**IMPORTANCE** du travail de **CURATION**
- 4 En guise de **CONCLUSION** : **VALEUR AJOUTÉE** et **LIMITES** de la **CURATION** manuelle

1

Le métier de **DOCUMENTALISTE** dans le contexte de la **SCIENCE OUVERTE** et de l'**OPEN DATA**

Le **DOCUMENTALISTE** est mort, vive le **DATA LIBRARIAN ??**

DOCUMENTALISTE au sein de l'**ATELIER** de la **DONNÉE** dat@UBFC

1 Le DOCUMENTALISTE est mort : vive le DATA LIBRARIAN ??



Petit coup d'œil dans le rétroviseur

- **Métier récent** : le terme « **documentaliste** » apparaît en **1932** ([voir le TLFi](#))
- Les ***buts de la documentation*** selon **Paul Otlet** en **1934** :
« Les Buts de la Documentation organisée consistent à pouvoir offrir sur tout ordre de fait et de connaissance des informations documentées : 1° universelles quant à leur objet ; 2° sûres et vraies ; 3° complètes ; 4° rapides ; 5° à jour ; 6° faciles à obtenir ; 7° réunies d'avance et prêtes à être communiquées ; 8° mises à la disposition du plus grand nombre¹ »
- **Profession « ringardisée »** dans les années 1990-2000 avec **le développement du Web**

¹ OTLET, Paul, *Traité de documentation. Le livre sur le livre*. Théorie et pratique, Editions Mundanéum, Palais mondial, 1934.



Aujourd'hui, dans le contexte Open Science/Open Data

- Le.la **documentaliste** reste un.e **généraliste** : c'est « **l'Encyclopaedia Minimalis du savoir²** »
- **Lien fort** entre les **professionnel.le.s** de l'IST et les **données de la recherche**
- **Enjeux importants** autour des **données** produites par la recherche → **besoin de structurer / documenter**
- **Évolution** des **pratiques** des **chercheur.se.s** et de la **recherche** → **besoin d'accompagnement**

² SIC : Hélène Tisserand ;-)

1 DOCUMENTALISTE au sein de l'ATELIER de la DONNÉE dat@UBFC



SENSIBILISER, DIFFUSER

les bonnes pratiques

- Beaucoup d'**interventions** dans les équipes, laboratoires et établissements
- Participation à des **séminaires**, des **colloques**
- Participation au développement des **politiques « Science ouverte »**



FORMER

la communauté

- **3 formations** proposées sur **3 sites**
- Formation des intervenant.e.s de **l'Atelier de la donnée**



(Beaucoup) ACCOMPAGNER

- Évolution des pratiques : RDM, DMP, **saisie de métadonnées sur le portail**



FAIRE ÉVOLUER

le portail dat@UBFC

- **Catalogue de métadonnées** (existe depuis 2016 > dat@OSU)
- Faire **converger** les attentes du.de la **chercheur.se**, les besoins en termes de **documentation** des données et les possibilités de **développement informatique**
- Assurer la **cohérence** et la **complétude** des **métadonnées** saisies



**ÊTRE
CURATEUR.RICE** de
MÉTADONNÉES

2

La **GESTION** des **MÉTADONNÉES** à l'**ATELIER** de la **DONNÉE** dat@UBFC

CURATION et **STRUCTURATION** des **MÉTADONNÉES** :
l'importance des **RÉFÉRENTIELS** et des **STANDARDS**

WORKFLOW de **VALIDATION** d'une fiche de
MÉTADONNÉES

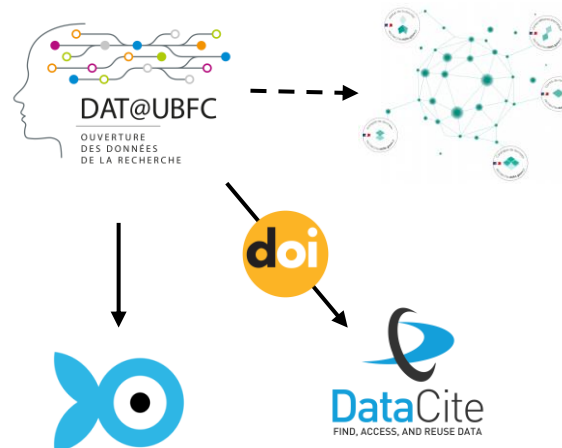
EN PRATIQUE : un **ENVIRONNEMENT MOUVANT** qui
complique notre affaire

DÉFINITION du PROFIL de METADONNÉES

- Schéma de **métadonnées** riche
- Respect des **standards** Dublin Core et DataCite



INTEROPÉRABILITÉ du PORTAIL

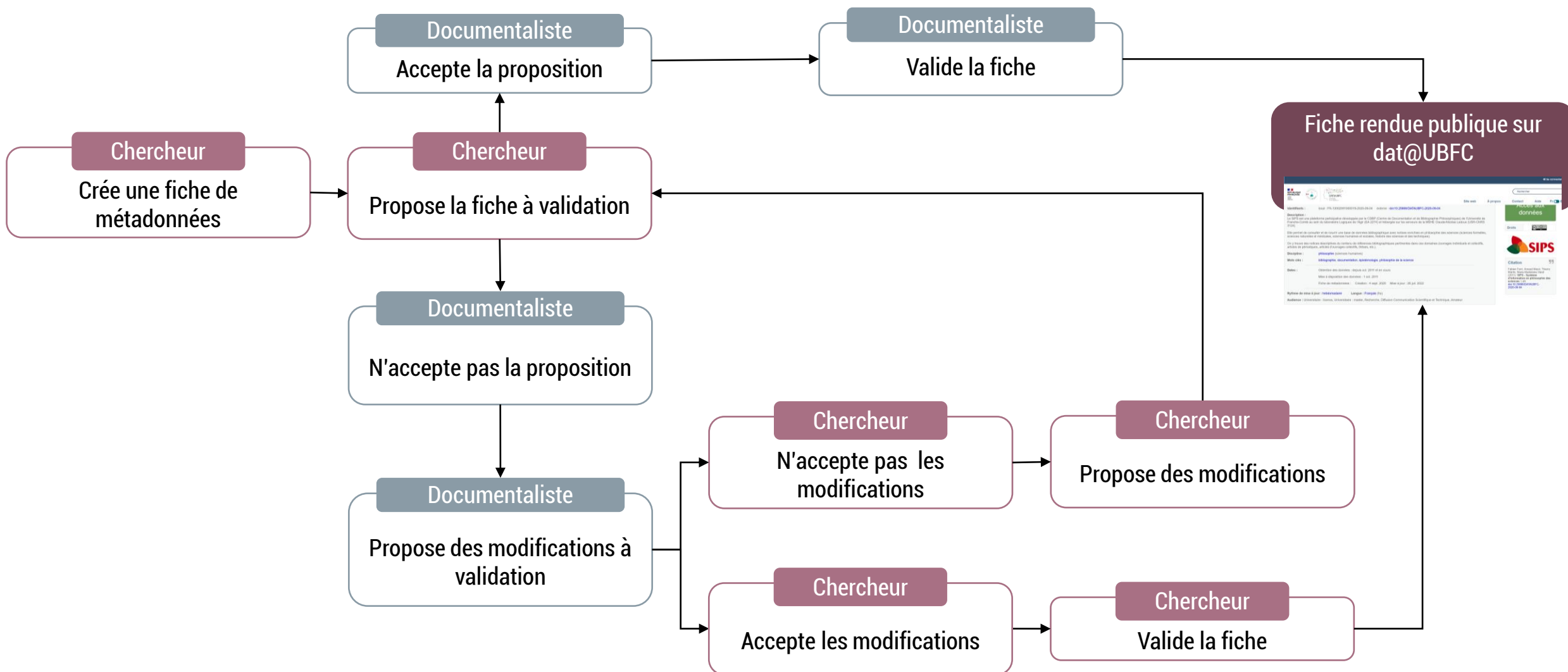


→ Travail de saisie simplifié

CURATION des METADONNÉES

- **Uniformisation, correction, complétion, nettoyage**
- Mise en lien avec des **référentiels reconnus par la communauté**







ÉVOLUTION des RÉFÉRENTIELS

- **Référentiels externes variables** pour les structures
- **Nouveaux référentiels** associés au portail

CHOIX des IDENTIFIANTS

- **Nombreux identifiants** disponibles (structure, personnes)
- Quelle **qualité des identifiants** proposés à la communauté ?
- **Quelles recommandations ?**

HARMONISATION des FINANCEMENTS

- **Diversité des informations** disponibles sur chaque **financement**
- **Quelles recommandations ?**

FACILITATION du TRAVAIL du CHERCHEUR

- **Pré-saisie d'informations** (financements, personnes...)

3

Des **EXEMPLES** concrets de l'**IMPORTANTANCE** du travail de **CURATION**

VALIDATION de **TAXONS**

TRAVAIL de l'**OMBRE** : la **CURATION AU LONG COURT**

L'IMPORTANCE de la CURATION : exemple de la VALIDATION de TAXONS



Le CONTEXTE :

- Fiche de **métadonnées** concernant des données **d'observation d'oiseaux**
- **Référentiels taxonomiques** utilisés pour catégoriser les espèces vivantes : [Catalogue of Life](#), [Mammal Species of the World](#) et [TaxRef](#) (Inventaire National du Patrimoine Naturel)
- **170 taxons à créer** sur cette fiche → **import** à partir d'un **csv** plutôt que saisie manuelle
- Avant l'import, **vérification de la liste csv de taxons** fournie par le chercheur → travail en **binôme** chercheur / documentaliste

	A	B	C	D	E	F	G
1	Rang tax.	Nom latin	Nom anglais	Nom français	Référentiel tax.	ID dans la class.	Commentaire
2	Species	<i>Accipiter badius</i>	Shikra	Épervier shikra	Catalogue of life	64FV5	
3	Species	<i>Amadina fasciata</i>	Cut-throat Finch	Amadine cou-coupé	Catalogue of life	CD8Y	
4	Species	<i>Anthreptes platura</i>					Trouvé dans aucune classification. Autres noms latins potentiels : <i>Anthreptes platurus</i> -> syn. <i>Hedydipna platura</i> (COL) --> est-ce que cela conviendrait ?
5	Species	<i>Anthus campestris</i>	Tawny Pipit	Pipit rousseline	Catalogue of life	67KDW	
6	Species	<i>Aquila rapax</i>	Tawny Eagle	Aigle ravisseur	Catalogue of life	FZLX	
7	Species	<i>Batis senegalensis</i>	Senegal Batis	Pririt du Sénégal	Catalogue of life	68G5N	
8	Species	<i>Bubalornis albirostris</i>	White-billed Buffalo Weaver	Alecto à bec blanc	Catalogue of life	NKLJ	
9	Species	<i>Bubo africanus</i>	Spotted Eagle-Owl	Grand-duc africain	Catalogue of life	NKS5	
10	Species	<i>Buphagus africanus</i>	Yellow-billed Oxpecker	Piquebœuf à bec jaune	Catalogue of life	5WTPV	
11	Species	<i>Burhynus capensis</i>	Spotted Thick-knee	Œdicnème tachard	Catalogue of life	NXG4	trouvé mais sous la forme de " <i>Burhinus capensis</i> " --> OK ?
12	Species	<i>Butastur rufipennis</i>	Grasshopper Buzzard	Busautour des sauterelles	Catalogue of life	NYTJ	
13	Species	<i>Camaroptera brachyura</i>	Green-backed Camaroptera	Camaroptère à tête grise	Catalogue of life	5X8YX	
14	Species	<i>Campethera punctuligera</i>	Fine-spotted Woodpecker	Pic à taches noires	Catalogue of life	QCJ9	
15	Species	<i>Centropus senegalensis</i>	Senegal Coucal	Coucal du Sénégal	Catalogue of life	SB7V	
16	Species	<i>Cercotricas podobe</i>	Black Scrub Robin	Agrobate podobé	Catalogue of life	69LLR	trouvé mais sous la forme de " <i>Cercotrichas podobe</i> " --> OK ?

3

L'IMPORTANCE de la CURATION : exemple de la VALIDATION de TAXONS

101	Species	Turtur abyssinica	Black-billed Wood Dove	Tourtelette d'Abyssinie	Catalogue of life	59R5S	Trouvé mais sous la forme "Turtur abyssinicus" --> OK ?
102	Species	Upupa epops	Common Hoopoe	Huppe fasciée	Catalogue of life	7DLHC	
103	Species	Vanellus senegalus	African Wattled Lapwing	Vanneau du Sénégal	Catalogue of life	7FFJQ	Trouvé mais sous la forme "Vanellus sengallus" --> OK ?
104	Species	Vanellus tectus	Black-headed Lapwing	Vanneau à tête noire	Catalogue of life	7FFJN	
105	Species	Vidua chalybeus					Pas trouvé dans les référentiels... Dans COL, on trouve quelque chose d'approchant : Vidua chalybeata / Village Indigobird / Combassou du Sénégal (id. COL : 7FSG2) --> est-ce que ça colle ?
106	Species	Vidua orientalis	Sahel Paradise Whydah	Veuve à collier d'or	Catalogue of life	5BDB5	
119	Species	Quelea quelea	Red-billed Quelea	Travailleur à bec rouge	Catalogue of life	4R42L	
120	Species	Nectarinia senegalensis	Scarlet-chested sunbird	Souimanga à poitrine rouge	Catalogue of life		Trouvé mais synonyme de "Chalcomitra senegalensis" --> lequel convient ? + je ne suis pas certaine du nom anglais --> OK ?
121	Species	Hypolais polyglotta	Melodious Warbler	Hypolais polyglotte	Catalogue of life	3M26F	Trouvé mais sous la forme "Hippolais polyglotta" --> OK ?
122	Species	Macrodipteryx longipennis	Standard-winged Nightjar	Engoulevent à balanciers	Catalogue of life	QSP3	Synonyme de "Caprimulgus longipennis" --> doublon !! Lequel on garde ?
123	Species	Emberiza cabanisi	Cabanis's Bunting	Bruant de Cabanis	Catalogue of life	39FSP	



PROBLÈMES rencontrés

Liés au recensement « **humain** » des taxons :

- Fautes de frappe, inversion de caractères
- Doublons

Liés au caractère « **instable** » des référentiels taxonomiques :

- Taxons **absents** des référentiels
- **Synonymes, plusieurs formes** existantes → des référentiels sont regroupés, d'autres divisés
- **Évolution des nomenclatures dans le temps** donc évolution des taxons, évolution des connaissances



TRAVAIL de CURATION

- **Travail complémentaire** : compétences scientifiques + compétences documentaires
- Travail de curation nécessaire car **intervention humaine** (sinon, on ne sert à rien ;-)
- Accepter que les **connaissances** sont en **évolution**, donc les **référentiels** sont **mouvants**
- **Valeur ajoutée** par rapport à un entrepôt dont les métadonnées ne font pas l'objet d'une curation : les taxons impropres auraient été acceptés sans problème

Ajouter / mettre à jour les **liens** vers les **publications**
(y compris thèses)

Ajouter le **lien d'accès** aux **données** en cas d'embargo

Vérifier / mettre à jour le **lien d'accès** aux **données**

Dédoublonnage
(cause technique, erreur humaine, forme approchante) - principalement pour les **mots clés** et les **personnes**

Associer des **mots clés** à des **référentiels** pour ceux qui n'en ont pas (+ ajout de nouveaux référentiels)

Mise à jour des DOI (en cas de modification sur la fiche de métadonnées)

Harmonisation de la **saisie** dans les **référentiels** : structures, projets, financements, données de thèse...



En cours de réflexion :

Pré saisie des **personnes** (membres d'UBFC) et des **projets financés** (via data.gouv, ANR, etc.)

A vertical line runs down the left side of the slide. A dark grey circle is positioned on this line, containing the number '4'.

4

En guise de **CONCLUSION** : **VALEUR AJOUTÉE** et **LIMITES** de la **CURATION** manuelle



VALEUR AJOUTÉE

- Respect des standards
- Optimisation du référencement, plus de visibilité → **FINDABLE**
- Maintien de l'accès aux métadonnées et données si ouvertes → **ACCESSIBLE**
- Interopérabilité (avec Isidore par ex. et bientôt avec RDG) → **INTEROPERABLE**
- Richesse des métadonnées → **REUSABLE**



LIMITES, DIFFICULTÉS

- **Chronophage**
- Travail « **invisible** » pour l'humain, ingrat
- À reprendre sans cesse, **obsolescence** des référentiels (ex. du référentiel disciplines, de l'ajout de nouveaux référentiels mots clés et des mises à jour)



MERCI
pour votre écoute !



Nous **CONTACTER**

dataubfc-doc@ubfc.fr



Nous **SUIVRE**

@dataUBFC



Point d'entrée **UNIQUE** sur les
DONNÉES de la recherche à
UBFC

<https://data.ubfc.fr/>



CATALOGUE des **DONNÉES** de la
RECHERCHE en Bourgogne
Franche-Comté

<https://search-data.ubfc.fr/>



Mercredi **07 décembre 2022**
15h00

DATA SHS (MSH Dijon)
Gérer et ouvrir mes données de
recherche : les bonnes pratiques



Vendredi **09 décembre 2022**
10h00

CIMEOS (EA 4177)
Présentation de l'atelier de la
donnée dat@UBFC



Source des **PICTOGRAMMES**
utilisés :

<https://iconmonstr.com/>