REGISTRE FIAS DES SPÉCIFICATIONS DE RECHERCHE

Profil minimal de transcription des étiquettes d’herbier pour des missions de science participative

IDENTIFIANT DU REGISTRE
FOAR-CON2X-F8XPQCG7Y

VERSION
1.0

TYPE D’OBJET
Profil d’une norme externe

DOMAINE
Biologie et biodiversité

MATURITÉ DU DOCUMENT
Stable

SOUMISE
12 avril 2018, 12:00 UTC

PUBLIÉE
28 juin 2018, 12:00 UTC

LICENCE
Creative Commons Attribution 4.0

VERSION FAISANT AUTORITÉ
L’enregistrement consigne le résultat de l’examen du Registre pour cette version : la soumission a été suffisamment identifiée, documentée et déterminée en interne pour la procédure indiquée. La portée, l’adoption et les conventions concurrentes sont documentées séparément lorsqu’il y a lieu.

STATUT AU REGISTRE

Acceptée pour enregistrement

28 juin 2018, 12:00 UTC

Description principale

TYPE DE SOUMISSION

Nouveau concept de registre

TITRE COMPLET

Profil minimal de transcription des étiquettes d’herbier pour des missions de science participative

TITRE COURT OU ACRONYME

PMTEH 1.0

TYPE D’OBJET

Profil d’une norme externe

STATUT REVENDIQUÉ PAR LE DÉPOSANT

Norme de projet

MATURITÉ DE CETTE VERSION

Stable

VERSION

1.0

DOMAINE PRINCIPAL

Biologie et biodiversité

MOTS-CLÉS

- herbier
- spécimen préservé
- transcription participative
- Darwin Core
- provenance
- Les Herbonautes

PROBLÈME, PORTÉE ET EXCLUSIONS

Le profil prescrit les champs minimaux et les états de validation d'une transcription participative d'étiquette de planche de plante vasculaire. Il couvre le collecteur, son numéro, la date, la localité, l'identification et le lien à l'image. Son périmètre se limite à la transcription et exclut nouvelle détermination taxonomique, géolocalisation inférée, correction silencieuse de l'étiquette historique et modification de l'objet physique.

UTILISATEURS ET CONTEXTES VISÉS

- Contributeurs de missions de transcription d'herbiers numérisés
- Conservateurs contrôlant les conflits de lecture
- Agrégateurs exportant des occurrences en Darwin Core

RELATION AVEC LA PRATIQUE EXISTANTE

Profile une norme existante

CONTEXTE SCIENTIFIQUE, SOURCES ET ÉCHANGES SCIENTIFIQUES

Cette section décrit les fondements scientifiques, les travaux sous-jacents et les sources de la spécification, ainsi que les groupes de recherche et les cadres scientifiques dans lesquels la solution proposée ou sa formalisation a été présentée ou discutée. Elle documente le contexte et la nature de ces échanges scientifiques.

La numérisation des herbiers ouvre l'étude des collections à de nouveaux usages, à condition que les informations des étiquettes restent reliées aux images et aux spécimens. G. Le Bras et al. (2017) décrivent l'organisation du jeu de données de l'Herbier national de Paris en distinguant image de la planche, données minimales d'occurrence, informations de collecte et enrichissements issus de la transcription participative. Les termes Darwin Core fournissent le vocabulaire d'échange pour ces données.

Le programme Les Herbonautes organise la lecture collective des étiquettes en soumettant une même information à plusieurs contributeurs et en établissant une validation par consensus. La publication de 2017, associant la Direction des collections du MNHN et l'ISYEB, présente ce travail commun de gestion de collection, de systématique et de participation du public. Les contributions de G. Le Bras, M. Pignal, C. Aupic et E. Pérez Pimparé s'inscrivent dans cette mise à disposition scientifique des collections botaniques.

Le profil formalise les informations minimales à transcrire : collecteur, numéro de collecte, date, localité, identification et lien à l'image. Il conserve le texte de l'étiquette avec les états de validation, les désaccords et les éventuelles normalisations. Des valeurs distinctes rendent explicites les informations inconnues et les champs encore non traités. Cette structure permet de reprendre une transcription, d'examiner les décisions collectives et de projeter le résultat vers un jeu de données d'occurrences Darwin Core.

Références:

- Le Bras, G. et al. (2017), The French Muséum national d'histoire naturelle vascular plant herbarium collection dataset, Scientific Data 4: 170016. DOI: https://doi.org/10.1038/sdata.2017.16
- Muséum national d'Histoire naturelle, The vascular plants collection (P) at the Herbarium of the Muséum, jeu de données d'occurrences. DOI: https://doi.org/10.15468/nc6rxy
- Darwin Core Task Group (2009), Darwin Core, standard technique TDWG 450, ratifié le 9 octobre 2009. Texte: https://www.tdwg.org/standards/dwc/

STATUT DE NORMALISATION EXTERNE

Aucune prévue

LICENCE DE LA SPÉCIFICATION SOUMISE

Creative Commons Attribution 4.0

DÉCLARATION DE PUBLICATION

Oui

Profil d’une norme externe

NORME DE BASE, VERSION ET RÉFÉRENCE STABLE

Darwin Core (TDWG Standard 450, ratifié le 9 octobre 2009), notamment occurrenceID, institutionCode, collectionCode, catalogNumber, basisOfRecord, recordedBy, recordNumber, verbatimEventDate, eventDate, verbatimLocality, locality, countryCode, scientificName, identifiedBy, dateIdentified et associatedMedia. Le jeu de données de l'Herbier P (10.15468/nc6rxy) sert d'exemple d'archive d'occurrences.

QUELLE EST LA CIBLE DE CONFORMITÉ ?

Jeu de données ou collection d’enregistrements

FONCTIONNALITÉS REQUISES SÉLECTIONNÉES DANS LA NORME DE BASE

- occurrenceID et catalogNumber identifiant sans ambiguïté la planche
- basisOfRecord avec la valeur PreservedSpecimen
- recordedBy et recordNumber quand ils sont lisibles sur l'étiquette
- verbatimEventDate et eventDate, sans normalisation silencieuse
- verbatimLocality, locality et countryCode avec provenance de la normalisation
- texte d'identification visible, scientificName, identifiedBy et dateIdentified
- associatedMedia résolvant vers l'image utilisée pour la transcription

FONCTIONNALITÉS EXCLUES OU INTERDITES

- Valeurs géographiques déduites sans indice visible ni autorité citée
- Nouvelle détermination taxonomique par le simple acte de transcription
- Texte OCR brut présenté comme transcription validée
- Écrasement d'une détermination historique visible par un nom normalisé
- Valeur vide employée pour signifier à la fois absent, illisible et non traité

EXTENSIONS OU CONTRAINTES LOCALES

Chaque champ transcrit ajoute transcriptionStatus parmi unread, single, consensus et adjudicated; contributorCount; conflictFlag; sourceRegion décrivant la zone de l'image; et transcriptionNote pour les lectures incertaines. Une valeur normalisée conserve toujours la valeur verbatim et l'identifiant de l'autorité ou de la règle de normalisation. Les identités publiques des contributeurs restent facultatives.

COMPATIBILITÉ AVEC LES IMPLÉMENTATIONS CONFORMES À LA NORME DE BASE

Le profil étend la norme de base

EXPLICATION DE LA COMPATIBILITÉ

Les champs d'occurrence exportés restent des termes Darwin Core. Les états locaux de transcription et de consensus sont des extensions de provenance; un consommateur

Darwin Core peut les ignorer sans perdre la valeur publiée, mais perd alors l'information sur son mode de validation.

TEST DE CONFORMITÉ DU PROFIL

Pour chaque ligne, occurrenceID doit résoudre vers la planche et associatedMedia vers l'image effectivement lue. Toute valeur normalisée doit conserver son équivalent verbatim. Deux lectures divergentes doivent rester conflictFlag=true jusqu'à consensus ou adjudication. Un validateur vérifie le vocabulaire des états, les cardinalités, l'absence de confusion entre inconnu et vide et la traçabilité de chaque correction.

Historique d’examen du Registre

Soumission reçue
12 avril 2018, 12:00 UTC

Soumission reçue.

Acceptée pour enregistrement
28 juin 2018, 12:00 UTC

Le profil est étroit, testable et conserve la distinction entre ce que dit l'étiquette et une normalisation ultérieure. Le consensus et l'adjudication sont explicites, les valeurs inconnues et les champs non traités sont distingués, et la sortie se projette vers l'archive d'occurrences Darwin Core décrite par Le Bras et al. (2017).