



Printemps de la donnée – Curathon : dev-outils pour la modération de l’entrepôt Nakala

20-05-2025

Animation : David Chesnet (MSHS, Univ-Poitiers) & Michael Nauge (DRInnov, Univ-Poitiers)

Guest : Sara Tandar (Huma-Num)



www.univ-poitiers.fr



Séance 1

<https://scienceouverte.univ-poitiers.fr/>



Programme

9h - 9h30 - accueil / café

9h30 - Présentation de la journée (contexte)

9h35 Sara Tandar (Huma-Num) : 5 minutes pour présenter le point de vue de l'IR*

9h40 Questions à Huma-Num et aux participants

10h15 Quelques exemples de données à discuter

mise en contexte avant d'aborder la grille de modération

10h30 Lecture de la grille de modération: partie obligatoire, partie optionnelle

Les questions qui ont été posé précédemment entrent-elles dans le cadre de la grille ?

Tâches:

Identifier les entrées qui font consensus et facilement automatisables

Identifier les entrées partiellement automatisables

Identifier les entrées « à débat »

Identifier les entrées « impossibles » (à réserver aux humains)

Objectif:

Sélectionner les parties qui pourraient être automatisées pour simplifier le diagnostic

11h45 Classement via wooclap

12h Repas

14h-18h Curathon-Python: implémenter l'automatisation de parties du diagnostic

www.univ-poitiers.fr



Sara Tandar (Huma-Num)



Séance 1

www.univ-poitiers.fr



Questions à Huma-Num et aux participants



Séance 1

www.univ-poitiers.fr



Programme

9h - 9h30 - accueil / café

9h30 - Présentation de la journée (contexte)

9h35 Sara Tandar (Huma-Num) : 5 minutes pour présenter le point de vue de l'IR*

9h40 Questions à Huma-Num et aux participants

10h15 Quelques exemples de données à discuter

mise en contexte avant d'aborder la grille de modération

10h30 Lecture de la grille de modération: partie obligatoire, partie optionnelle

Les questions qui ont été posé précédemment entrent-elles dans le cadre de la grille ?

Tâches:

Identifier les entrées qui font consensus et facilement automatisables

Identifier les entrées partiellement automatisables

Identifier les entrées « à débat »

Identifier les entrées « impossibles » (à réserver aux humains)

Objectif:

Sélectionner les parties qui pourraient être automatisées pour simplifier le diagnostic

11h45 Classement via wooclap

12h Repas

14h-18h Curathon-Python: implémenter l'automatisation de parties du diagnostic

www.univ-poitiers.fr



Quelques exemples de données à discuter

Rechercher dans NAKALA...



UN SI JOLI VILLAGE VILLAGE. Copie sur bande

Contactez le gestionnaire

DOI : 10.34847/nkl.f55a7m15

Publique

Auteur : Paul Misraki

Musique enregistrée sur bande pour le film réalisé par Étienne Périer : Un si joli village (1979). Editions musicales : Hortensia

| Fichiers | Visualisation |
|---|--|
|  BM_24_01.mp3 15.2 Mo |  |
|  BM_24_01.wav 269.4 Mo | |

<https://nakala.fr/10.34847/nkl.f55a7m15>

www.univ-poitiers.fr



| | | |
|-----------------------------|----------------|--|
| nakala:title | Français | UN SI JOLI VILLAGE VILLAGE. Copie sur bande |
| nakala:creator | | Paul Misraki |
| nakala:created | | 1979 |
| nakala:type | dcterms:URI | Son |
| nakala:license | | Creative Commons Attribution Non Commercial Share Alike 4.0 International (CC-BY-NC-SA-4.0) |
| dcterms:contributor | | Compositeur : Misraki, Paul |
| | | Réalisateur du film : Périer, Etienne |
| dcterms:creator | Français | Numérisation : Arcafilms (2022) |
| dcterms:description | Français | Musique enregistrée sur bande pour le film réalisé par Étienne Périer : Un si joli village (1979). Editions musicales : Hortensia |
| dcterms:extent | | 00:16:21 |
| dcterms:format | Français | MP3 et Wav 48khz-24bit |
| dcterms:identifiant | | MISRAKI_COLL_079 BM_24_01 |
| dcterms:issued | dcterms:Period | start=1978; end=1979; scheme=W3C-DTF |
| dcterms:medium | Français | Bande magnétique 1/4", 18 cm / 19 cm/s ; stéréo-mono |
| dcterms:publisher | Français | Ces archives sont mises à disposition de la recherche sous la responsabilité scientifique de Julien Ferrando (Université Aix-Marseille), Jérôme Rossi (Université de Lyon) et Joséphine Simonnot (CNRS), avec la participation de l'UMR PRISM, l'UMR IDEMEC, le CAPHI/Passages 20-21, avec le soutien des DRAC PACA et Pays de Loire et en collaboration avec la famille Misrachi. Les archives sont conservées au CNSMDP. |
| dcterms:rightsHolder | Français | Tout droits réservés à la famille Misrachi jusqu'en 2068 |
| dcterms:spatial | Français | France |
| dcterms:subject | Français | Musique de film |
| dcterms:temporal | Français | XXème siècle |



| | | |
|-----------------------------|----------------|--|
| nakala:title | Français | UN SI JOLI VILLAGE VILLAGE. Copie sur bande |
| nakala:creator | | Paul Misraki |
| nakala:created | | 1979 |
| nakala:type | dcterms:URI | Son |
| nakala:license | | Creative Commons Attribution Non Commercial Share Alike 4.0 International (CC-BY-NC-SA-4.0) |
| dcterms:contributor | | Compositeur : Misraki, Paul |
| | | Réalisateur du film : Périer, Etienne |
| dcterms:creator | Français | Numérisation : Arcafilms (2022) |
| dcterms:description | Français | Musique enregistrée sur bande pour le film réalisé par Étienne Périer : Un si joli village (1979). Editions musicales : Hortensia |
| dcterms:extent | | 00:16:21 |
| dcterms:format | Français | MP3 et Wav 48khz-24bit |
| dcterms:identifiant | | MISRAKI_COLL_079 BM_24_01 |
| dcterms:issued | dcterms:Period | start=1978; end=1979; scheme=W3C-DTF |
| dcterms:medium | Français | Bande magnétique 1/4", 18 cm / 19 cm/s ; stéréo-mono |
| dcterms:publisher | Français | Ces archives sont mises à disposition de la recherche sous la responsabilité scientifique de Julien Ferrando (Université Aix-Marseille), Jérôme Rossi (Université de Lyon) et Joséphine Simonnot (CNRS), avec la participation de l'UMR PRISM, l'UMR IDEMEC, le CAPHI/Passages 20-21, avec le soutien des DRAC PACA et Pays de Loire et en collaboration avec la famille Misrachi. Les archives sont conservées au CNSMDP. |
| dcterms:rightsHolder | Français | Tout droits réservés à la famille Misrachi jusqu'en 2068 |
| dcterms:spatial | Français | France |
| dcterms:subject | Français | Musique de film |
| dcterms:temporal | Français | XXème siècle |



Search

ADVANCED SEARCH

Showing 50 of 42119 results.

Items per page: 50 Page 1 of 843

| ORCID ID | First Name | Last Name | Other Names | Affiliations |
|-------------------------------------|---------------|-----------|-------------|---|
| 0000-0002-3304-1613 | Jean | Michaud | | Université Laval, Université de Montréal |
| 0000-0002-7184-0336 | Jean | Michaud | | |
| 0000-0003-3883-4484 | Jean | Michaud | | |
| 0000-0002-0799-017X | Jean-Philippe | Michaud | | CHU de Québec-Université Laval, Canadian Society of Transplantation, Université Laval |

Langue ?

Un titre ?

Déposée par CESCМ-CIFM Corpus des inscriptions de la France médiévale le 23/10/2024

| | | |
|-------------------------|-------------|--|
| nakala:title | | CIFM 10, pl. LII, fig. 110 |
| nakala:creator | | Jean Michaud CESCМ-CIFM |
| nakala:created | | 1985 |
| nakala:type | dcterms:URI | Image |
| nakala:license | | Creative Commons Attribution Non Commercial No Derivatives 4.0 International (CC-BY-NC-ND-4.0) |
| dcterms:relation | | 123. Époque romane — Église Saint-Jean (pl. LII, fig. 110) |



| | | |
|----------------------------|-----------------|--|
| nakala:title | Français | Carte internationale du Monde : Grande Comore : SC-38 au 1:1.000.000 |
| nakala:creator | | Anonyme |
| nakala:created | | 1965 |
| nakala:type | dcterms:URI | Image |
| nakala:license | | Creative Commons Attribution Non Commercial Share Alike 4.0 International (CC-BY-NC-SA-4.0) |
| dcterms:coverage | Français | Océan Indien |
| | Français | Aldabra Island |
| | Français | Grande Comore |
| | Français | Moroni |
| | Français | Îles glorieuses |
| dcterms:creator | | France, Ministère des travaux publics et des transports |
| | | Service géographique de Madagascar |
| | | Institut Géographique National (Madagascar) |
| dcterms:extent | | 1:1.000.000 |
| dcterms:format | | 62 x 89 |
| dcterms:identifiant | | FR_R_MADA_F01_13_00004 |
| dcterms:language | dcterms:URI | fr |
| | dcterms:RFC5646 | français (fr) |
| | dcterms:RFC5646 | français (fr) |
| dcterms:license | dcterms:URI | https://creativecommons.org/licenses/by-nc-sa/4.0/ |
| dcterms:medium | Français | Papier |
| | Français | Polychrome |
| | Français | Méridien de Greenwich |
| | Français | Projection cônique conforme de Lambert |
| dcterms:spatial | dcterms:Point | east=45.145; north=-9.865;name=Point central de la carte |
| | dcterms:Point | east=43.244591; north=-11.699912;name=Moroni |
| | dcterms:Box | northlimit=-7.73; southlimit=-12; westlimit=42; eastlimit=48.29 |
| | dcterms:Point | east=43.335571; north=-11.452838;name=Grande Comore |
| dcterms:subject | | 1:1.000.000 |
| | Français | hypsométrie |
| | Français | transport ferroviaire |
| | Français | fleuve |
| | Français | infrastructure |
| | Français | ville |
| | Français | route |
| | Français | hydrographie |
| dcterms:type | dcterms:URI | http://purl.org/dc/dcmitype/StillImage |
| | Français | Carte topographique |



| | | | |
|-----------------------------|--|---|-----------|
| nakala:title | <input type="text"/> | POITIERS_Baptistère_Sarcophages_n87_CESCM-CROSLAND_2022 | → Titre ? |
| nakala:creator | Lisa-Oriane Crosland | | |
| nakala:created | 2022 | | |
| nakala:type | dcterms:URI | Image | |
| nakala:license | Creative Commons Attribution Non Commercial No Derivatives 4.0 International (CC-BY-NC-ND-4.0) | | |
| dcterms:contributor | Aurélie Devillechaise | | |
| dcterms:format | JPEG | | |
| dcterms:rightsHolder | CESCM/L-O Crosland, 2022 | | |
| dcterms:source | CESCM Poitiers | | |

Déposée par CESCM Photothèque le 01/06/2021

| | | |
|-----------------------|---|--------------------|
| nakala:title | Français | IMAGES BASE ROMANE |
| nakala:creator | null null | |
| nakala:created | 2021 | |
| nakala:type | dcterms:URI | Image |
| nakala:license | Creative Commons Attribution Non Commercial Share Alike 4.0 International (CC-BY-NC-SA-4.0) | |



Programme

9h - 9h30 - accueil / café

9h30 - Présentation de la journée (contexte)

9h35 Sara Tandar (Huma-Num) : 5 minutes pour présenter le point de vue de l'IR*

9h40 Questions à Huma-Num et aux participants

10h15 Quelques exemples de données à discuter

mise en contexte avant d'aborder la grille de modération

10h30 Lecture de la grille de modération: partie obligatoire, partie optionnelle

Les questions qui ont été posé précédemment entrent-elles dans le cadre de la grille ?

Tâches:

Identifier les entrées qui font consensus et facilement automatisables

Identifier les entrées partiellement automatisables

Identifier les entrées « à débat »

Identifier les entrées « impossibles » (à réserver aux humains)

Objectif:

Sélectionner les parties qui pourraient être automatisées pour simplifier le diagnostic

11h45 Classement via wooclap

12h Repas

14h-18h Curathon-Python: implémenter l'automatisation de parties du diagnostic

www.univ-poitiers.fr



Lecture de la grille de modération: partie obligatoire, partie optionnelle

| A | B | C | D | E | F |
|-----------------------------|--|--|----------------------------|---|---------------|
| | Les métadonnées | | | | |
| metaldMod | | critères obligatoires pour valider une donnée modérée | Oui / non / non applicable | Commentaires et propositions d'amélioration | automatisable |
| check_type-consistency | Type de dépôt | Le type sélectionné est en cohérence avec le fichier déposé. Si le type est incohérent: ne pas valider Remarque : dans le cas d'une donnée multifichiers, le type sélectionné rend compte du type principal de la donnée. | | Une image de correspondance, c'est une correspondance. Mais: https://nakala.fr/10.34847/nkl.62da51u0 ! | Un peu |
| check_title-consistency | Titre | Un titre explicite et signifiant (comme le serait un titre d'article) est renseigné. S'il n'y a pas au moins un titre descriptif: ne pas valider. | | nakala:title | Oui |
| check_lang-str | | Si la langue du ou des titres n'est pas renseignée: ne pas valider. | | On peut tenter une détection de la langue et proposer une réponse dans le format attendu | Oui |
| check_author-orcid | Auteurs | Si un identifiant ORCID existe pour un auteur mais n'est pas renseigné, ne pas valider. | | [API ORCID] | Oui |
| check_nklCreated-unknown | Date de création | Si une date est connue mais n'est pas renseignée: ne pas valider. Si « Inconnue » a été coché, une date sous forme libre est renseignée dans dcterms:date ou dcterms:created. | | Obligation de vérifier si le déposant n'a pas menti ? | Un peu |
| check_licence-justification | Licence | Si la licence n'est pas en cohérence avec la donnée (par exemple, la licence renseignée est libre alors que les données sont sous droit): ne pas valider. | | Obligation de vérifier auprès du déposant, PGD ? | Non |
| | | La licence est libre ou, si la licence n'est pas libre, le déposant peut justifier des contraintes d'accessibilité et de réutilisation. | | Obligation de vérifier auprès du déposant, Comment justifier: ajouter un champs DCTerms ? | Non |
| check_description | Description | Une description qui précise le contexte scientifique de production ou de collecte de la donnée est renseignée dans le champ «Description». Si aucune description n'est renseignée (ni dans champs Description, ni dans dcterms:abstract, ni dans dcterms:tableOfContents): ne pas valider | | | Un peu |
| check_lang-str | | La langue de chaque description, du dcterms:abstract et du dcterms:tableOfContent est renseignée. Si la langue de la description n'est pas renseignée: ne pas valider. | | | Oui |
| check_keywords-numbers | Mots-clés | Si ne sont pas renseignés au moins deux mots-clés ou si les mots-clés sont incohérents avec le contenu intellectuel de la donnée: ne pas valider. | | Vérifier présence: oui. | Un peu |
| check_lang-str | | Si la langue des mots-clés n'est pas renseignée: ne pas valider. | | | Oui |
| check_lang-file | Langues | Pour des fichiers vidéo, son ou texte, la ou les langue du contenu est renseignée. Si la ou les langues ne sont pas renseignées ou en incohérence avec le contenu: ne pas valider. | | on ne peut pas valider l'exactitude de la langue par prog (ChatGPT peux-tu reconnaître du document ?) dcterms:ISO639-2 ; dcterms:ISO639-3 ; dcterms:RFC-5646 pour dcterms: subject et dcterms:language | Un peu |
| check_bicCitation | Référence bibliographique complète (dcterms:bibliographicCitation) | Si une publication est associée aux données, la référence bibliographique complète de la publication est renseignée dans un dcterms:bibliographicCitation. Si une publication associée à la donnée existe et que cette métadonnée n'est pas renseignée: ne pas valider | | ISIDORE: IA de détection et d'analyse des liens : Une IA de détection, d'analyse et d'exposition des liens entre données et publications a été intégrée pour enrichir les résultats de recherche. | Non |
| check_tvdeUri | URI | Si une métadonnée contenant uniquement une URI est renseignée mais que le type « dcterms:URI » n'est pas sélectionné: ne pas valider | | | Oui |

<https://cloud.univ-poitiers.fr/s/kr7PctYcTd4rRAB>

www.univ-poitiers.fr



Construction des priorités

wooclap.com

****  Join a Wooclap event

Distribuer le plus grand nombre de points aux propositions qui vous semblent utile d'intégrer en priorité dans un outil de génération de pré-rapport de modération automatisé.

Distribuez 100 points entre les différentes propositions **100**

vérifier que le titre est explicite et signifiant

- +

vérifier la cohérence d'un texte (titre, description, etc) avec la "lang" choisie (et signaler si la langue est absente)

- +

vérifier les conformités aux attendues du CINES (format de fichier ouvert, fichier bien formé...)

- +

vérifier la cohérence et l'intelligibilité des noms des fichiers

- +

www.univ-poitiers.fr



Programme

9h - 9h30 - accueil / café

9h30 - Présentation de la journée (contexte)

9h35 Sara Tandar (Huma-Num) : 5 minutes pour présenter le point de vue de l'IR*

9h40 Questions à Huma-Num et aux participants

10h15 Quelques exemples de données à discuter

mise en contexte avant d'aborder la grille de modération

10h30 Lecture de la grille de modération: partie obligatoire, partie optionnelle

(Les questions qui ont été posé précédemment entrent-elles dans le cadre de la grille ?)

Tâches:

Identifier les entrées qui font consensus et facilement automatisables

Identifier les entrées partiellement automatisables

Identifier les entrées « à débat »

Identifier les entrées « impossibles » (à réserver aux humains)

Objectif:

Sélectionner les parties qui pourraient être automatisées pour simplifier le diagnostic

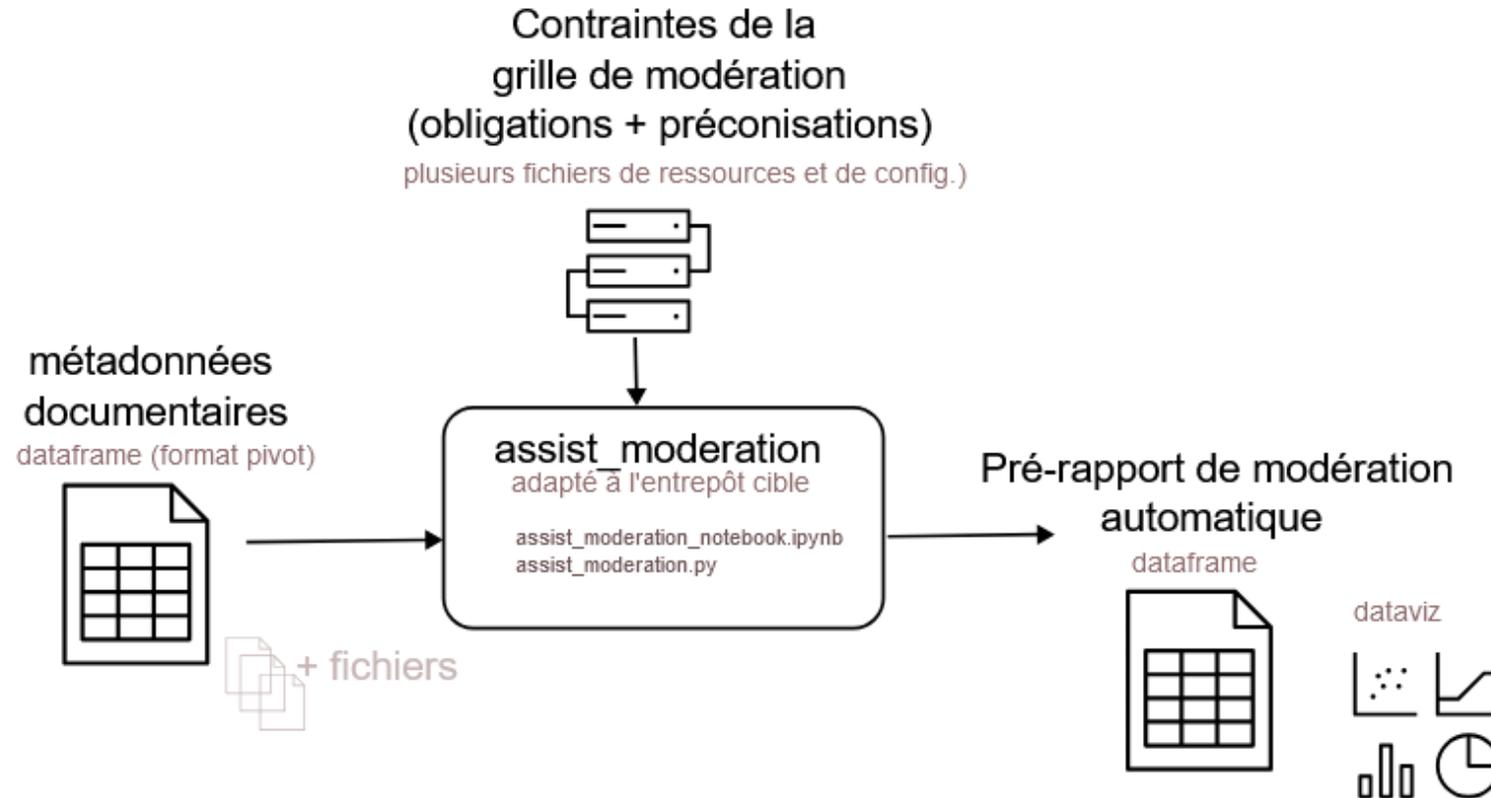
12h Repas

14h-18h Curathon-Python: implémenter l'automatisation de parties du diagnostic



Outil d'assistance à la modération

Vue simplifié du fonctionnement



https://gitlab.huma-num.fr/mshs-poitiers/plateforme/assist_moderation

www.univ-poitiers.fr