

Le plan de gestion de données du projet ERC-DHARMA

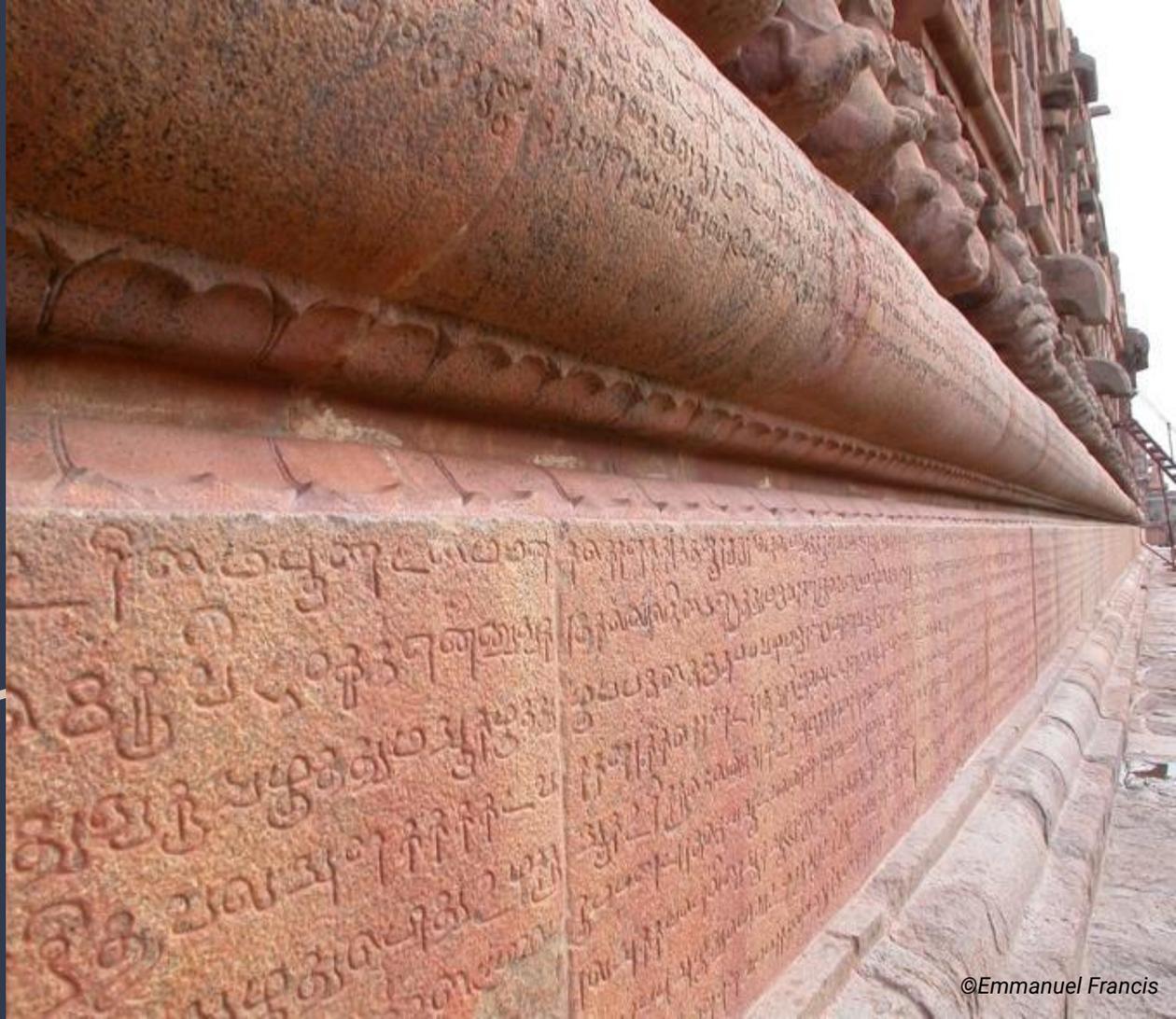
séminaire sur les PGD, le 11 février 2021, organisé par la MSHS-Poitiers

Axelle Janiak (CEIAS, CNRS-EHESS)
& Adeline Levivier (EFEO)



Pour nous contacter: axelle.janiak@ehess.fr & adeline.levivier@efeo.net

Domestication of
Hindu
Asceticism & the
Religious
Making of South
and Southeast
Asia



PATRONAGE
Inscriptions

Tamil South India
Deccan to Arakan
Cambodia & Campā
Indonesia

MATERIAL CULTURE
Sites & Artefacts

Temples (India)
Mahasthan (Bangladesh)
Hermitages (Cambodia)
Bumiayu (Indonesia)

INSTITUTIONALISATION
of
“HINDUISM”

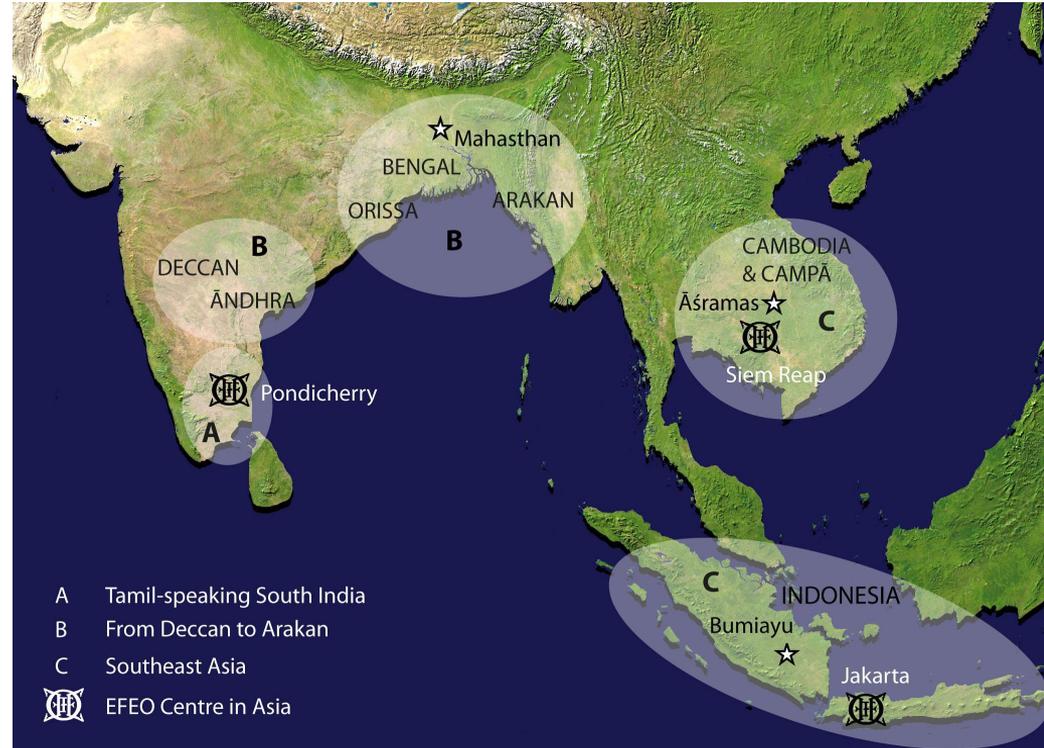
Śivadharma
Tantra
Śāsana

PRESCRIPTIONS
REPRESENTATIONS
Manuscripts

Literature
(Sanskrit, Tamil,
Old Javanese)

Organisation des Task Forces

- TFA dirigée par Emmanuel Francis (CEIAS - Paris)
- TFB dirigée par Annette Schmiedchen (UBER - Berlin)
- TBC dirigée par Arlo Griffiths (EFEO - Paris)
- TFD dirigée par Arlo Griffiths en partenariat avec Śivadharma (ERC Starting Grant) géré par Florinda de Simini (UNO - Naples) et Dominic Goodall (EFEO - Pondichéry)



33 corpora



Tiruppattūr
©Emmanuel Francis



Dharma Members at work
©Arlo Griffiths



©EFEO
P. 116
R. 3049
KANT



©Emmanuel Francis



©Emmanuel Francis



©Emmanuel Francis

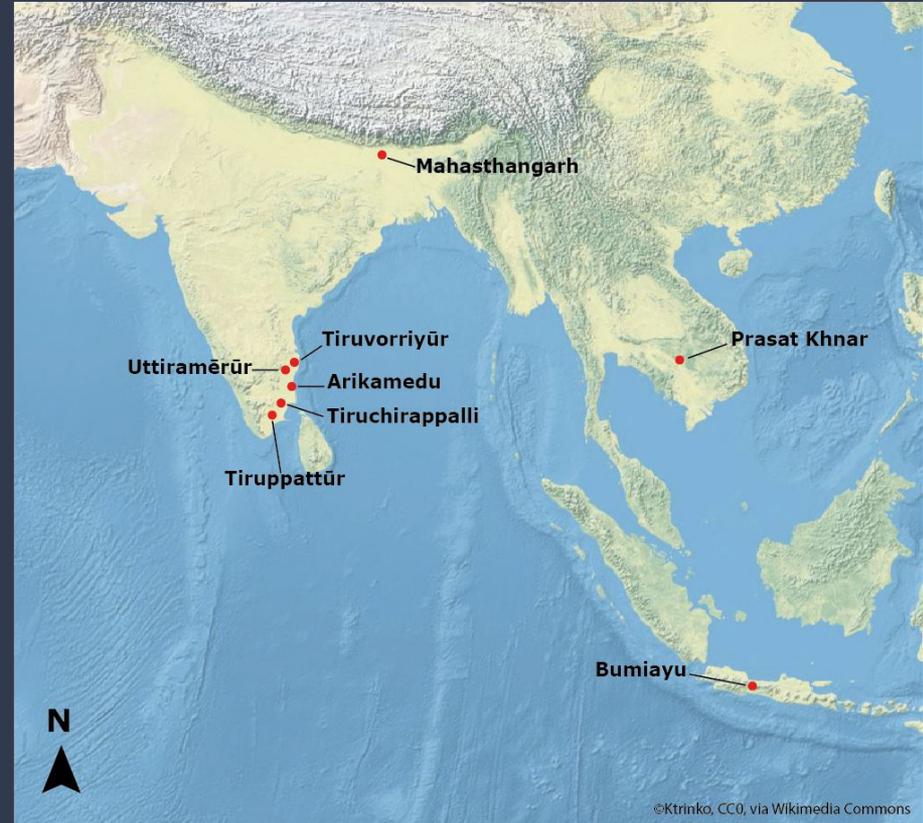


©Emmanuel Francis

Sites archéologiques



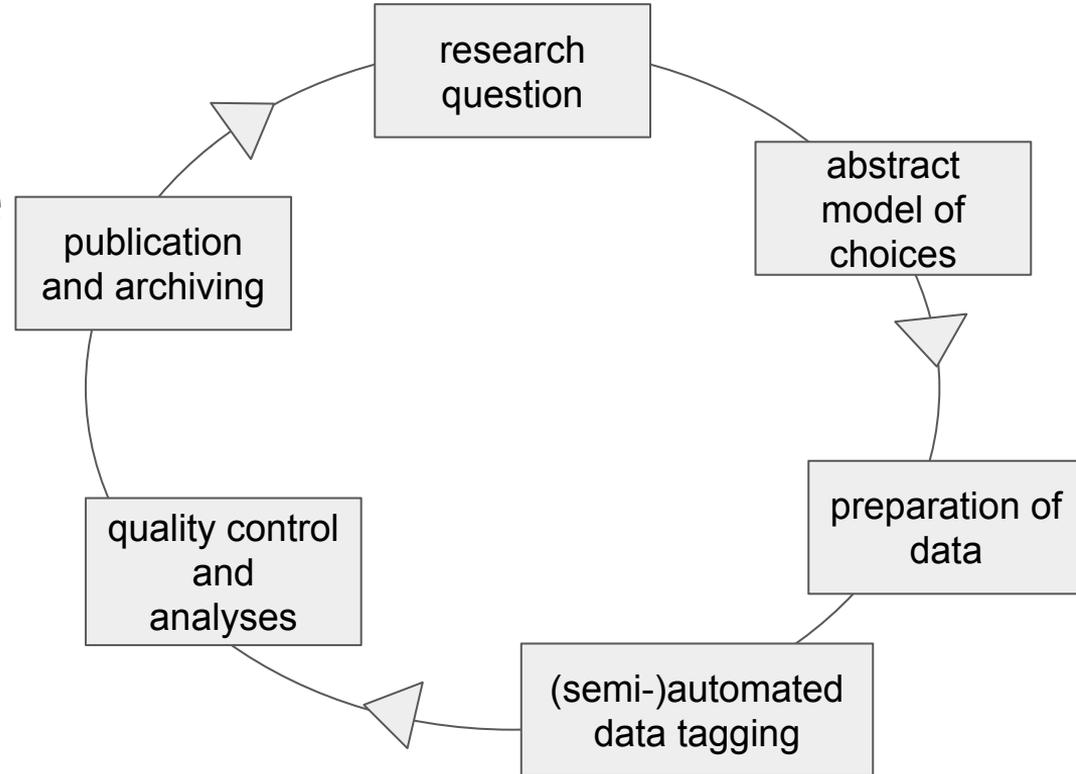
Citadelle de Mahasthan, Bangladesh
©MAFBM



©Ktrinko, CC0, via Wikimedia Commons

Engagements du projet

1. Une base de données épigraphique
2. Glossaire de l'épigraphie de l'Asie du Sud et du Sud-Est
3. Publications en *Open-access* (15 monographies, 10 volumes de journées d'études, rapports archéologiques et articles)



Vers la rédaction d'un PGD: les grandes étapes

1. Comprendre le contexte de rédaction
2. Identifier les jeux de données
3. Visualiser les chaînes de travail
4. Établir sa bonne gestion des données
5. Formaliser l'ensemble avec la rédaction du PGD
6. Et après?

Participation à Open Research Data Pilot

Open Research Data Pilot oblige à:

- Rédiger un plan de gestion de données
- Fournir en *Open-access* les données de recherche (lorsque cela est possible)

Théoriquement, cela n'implique que les données et les métadonnées nécessaires à la validation des publications scientifiques.

Calendrier de soumission pour H2020

- A l'étape de la proposition : fournir une courte description de la politique de gestion des données.
- Deux versions obligatoires: première version dans les 6 premiers mois du projet et une version mise à jour en fin de projet (**Final review DMP**)
- H2020 incite à une publication d'une mise à jour au milieu du projet (**Mid-term review DMP**) pour assurer que les changements sont reportés dans le document.

Rédiger un PGD pour l'ERC: comprendre les attentes du financier

“[...] a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR)”

[H2020 Programme Guidelines on FAIR Data Management in Horizon 2020](#), v3.0.

Les exigences minimales de l'ERC pour la première version

- Une description des données qui seront générées ou collectées
 - Les standards et les métadonnées qui seront utilisés
 - Le partage des données
 - L'archivage et la préservation
- Principes FAIR et *Open Science*

Identifier les jeux de données

Cela suppose de:

- Identifier et sélectionner les données
- Organiser les données en jeux cohérents
- Décrire les jeux de données avec des standards
- Définir les durées de conservation
- Définir les responsabilités de chacun
- Définir les infrastructures et les coûts pour le stockage des données

L'organisation de nos jeux de données

Les grandes catégories de données du projet DHARMA

- Données textuelles
- Données visuelles
- Données archéologiques
- Données techniques
- Données scientifiques

Division au sein des *task forces*, puis subdivision en grands *corpora* et/ou selon les terrains archéologiques

Deux exemples de jeux de données

- Données textuelles

- Formats collectés: .txt, .rtf, .doc, .docx, .tex et .xml
- Métadonnées
- Enjeux de traçabilité des origines des fichiers et des modifications apportées

- Données visuelles

- Collecte et création de photographies numériques
- Variété de formats (particulièrement pour la collecte)
- Formats 3D et RTI
- Question des métadonnées

Chaînes de travail et gestion des données

Définition des chaînes de travail:

- Adaptées à chaque type de donnée
- De la création/collection à l'exposition
- Préparation de l'archivage pérenne
- Préparation du cadre juridique et éthique

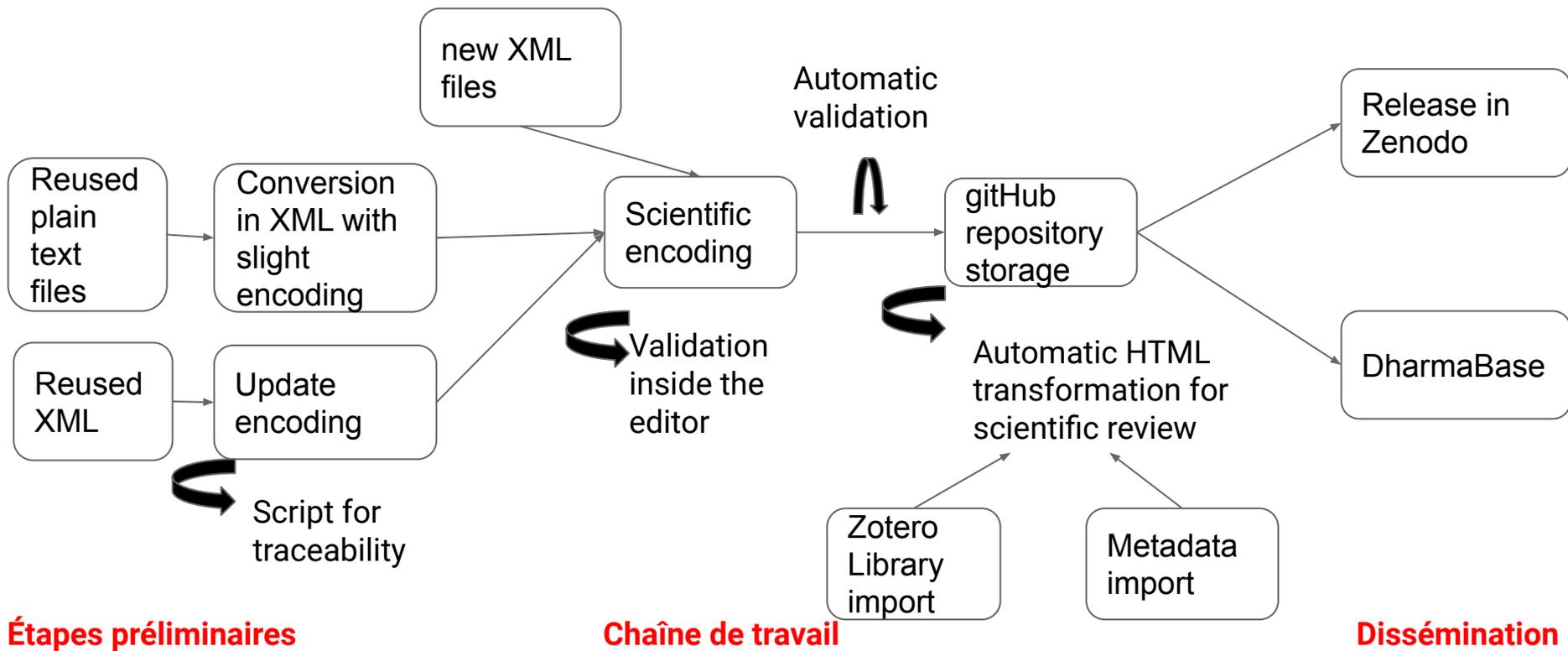
Pourquoi gérer les données scientifiques?

- Éviter/réduire la perte de données
- Améliorer la transparence du travail
- Rendre les données accessibles
- Faciliter la réutilisabilité et le partage des données

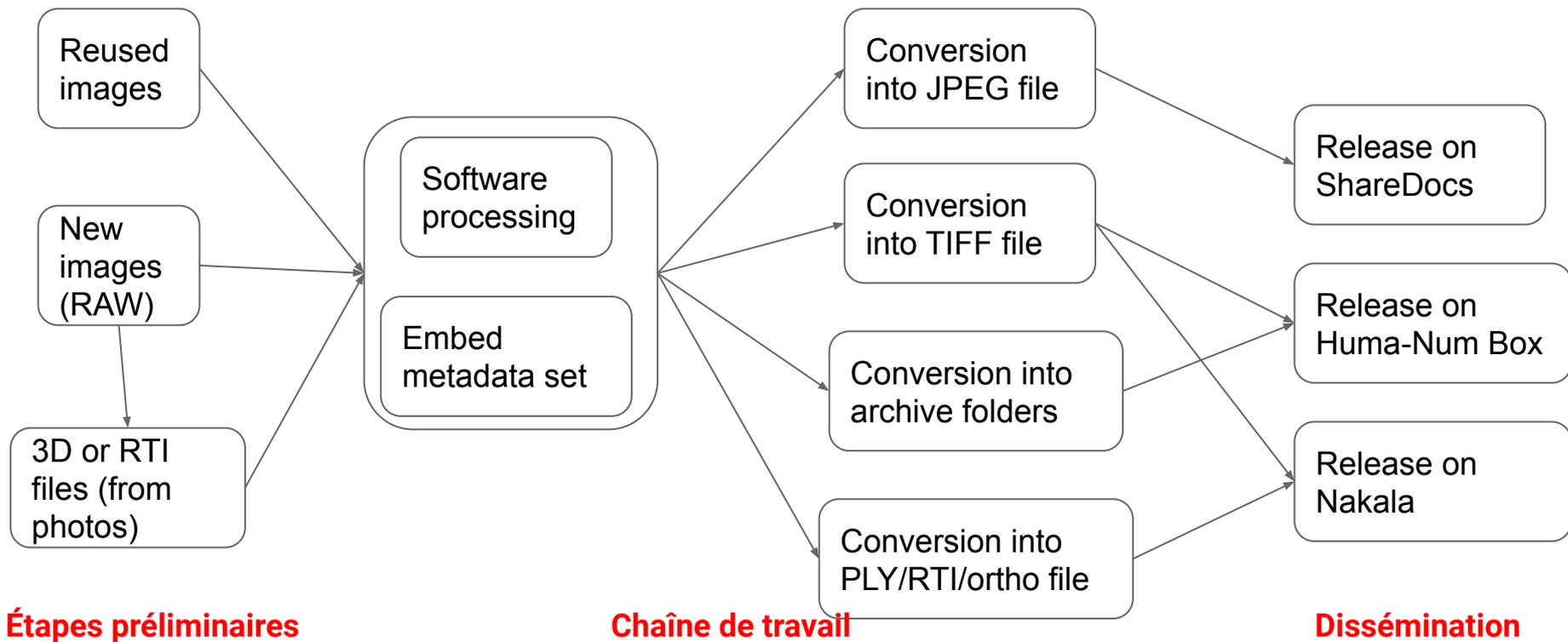
Gestion des données

- Organisation:
 - Nommage des fichiers
 - Formats des données
 - Outils et applications utilisés
 - Transfert et partage des fichiers entre membres du projet
 - Versionnage
- Stockage, dissémination et publication
- Management:
 - Sauvegarde
 - Documentation et métadonnées
 - Contrôle des accès
 - Sécurité des données
- Gestion des données en conformité avec les lois de la propriété intellectuelle et respect des principes éthiques

Exemple de la chaîne de travail des textes



Exemple de la chaîne de travail des images



La rédaction du PGD

- Document formel aux spécifications floues
 - Grande variété parmi les exemples disponibles
 - Peu de retours d'expériences dans les SHS
- Choix arbitraire et de style qui compile le travail préparatoire

“The plans are an opportunity to demonstrate your awareness of good practice and reassure funders that your proposal is in line with their data policy.”

S. Jones , “How to Develop a Data Management and Sharing Plan”, [Digital Curation Center How-to Guides](#), Edinburgh, 2011.

Plan de notre PGD

- Rédaction mixte avec prise en compte de la nature des données et des sections par domaines “techniques”
 - Présentation des données avec une sous-partie pour chaque jeu de données
 - Notre adhésion aux principes FAIR
 - Détails du budget
 - Sécurité des données
 - Aspects éthiques et juridiques

Notre tableau récapitulatif pour les données

- **Format** : Formats de collectes, de travail, de dissémination et d'archivages
- **Storage**: Stockage au cours du projet, lieu de dissémination et archivage à long terme, plus indication de volumétrie
- **Process**: Traitement des données - contenu varie selon la catégorie des données
- **Rights**: licences attribuées aux métadonnées, licences attribuées aux données elles-mêmes, indication si données sensibles et si traitements supposent des traitements différents

Formats	
Collected formats	.jpg, .raw, .tiff, .png, .ptm, .rti, .obj, .ply and .ai
Working formats	.pdf, .raw, .jpg, .tiff and .svg
Dissemination formats	.pdf, .jpg and .svg
Archiving format	.pdf, .tiff and .svg
Storage	
Current storage platforms	Siem Reap NAS server, PSIR/MOM in Lyon NAS server, own storage of the excavation project in question, and ShareDocs. Any sensitive data will be stored on a closed repository, such as Huma-Num Box.
Dissemination platforms	HAL-SHS and ARIADNEplus portal
Long-term repository	CINES through Huma-Num or Zenodo, except for the sensitive data which can not be disseminated or archived.
Volume	1-10 TB range
Rights	
Metadata rights	CC0
Resource rights	CC-BY, but as some data are sensitive, their access will be subject to agreements with the local authorities

Temps consacré au PGD

Mai à septembre 2019 pour rassembler des informations sur les jeux de données

Octobre 2019 pour la rédaction et les différentes relectures.

Et après la rédaction...?

- Retour de l'ERC?
- Formation et sensibilisation des chercheurs?
- Valorisation des PGD?
 - Aucune injonction des financeurs
 - *Data papers*

Bilan

- État des lieux des données
- Permet de définir des chaînes de travail en fonction des besoins du projet et de ses questions de recherche
- Définit les rôles et responsabilités de chacun
- Impose une ligne de conduite vis-à-vis des données
- Prise en compte dès le départ
 - de l'archivage et de la gestion après-projet
 - des questions juridiques et de licences