

Promesses et limites d'un séminaire interne de laboratoire SHS sur la gestion des données de la recherche

Emmanuelle Morlock
IE CNRS, UMR 5189 HiSoMA

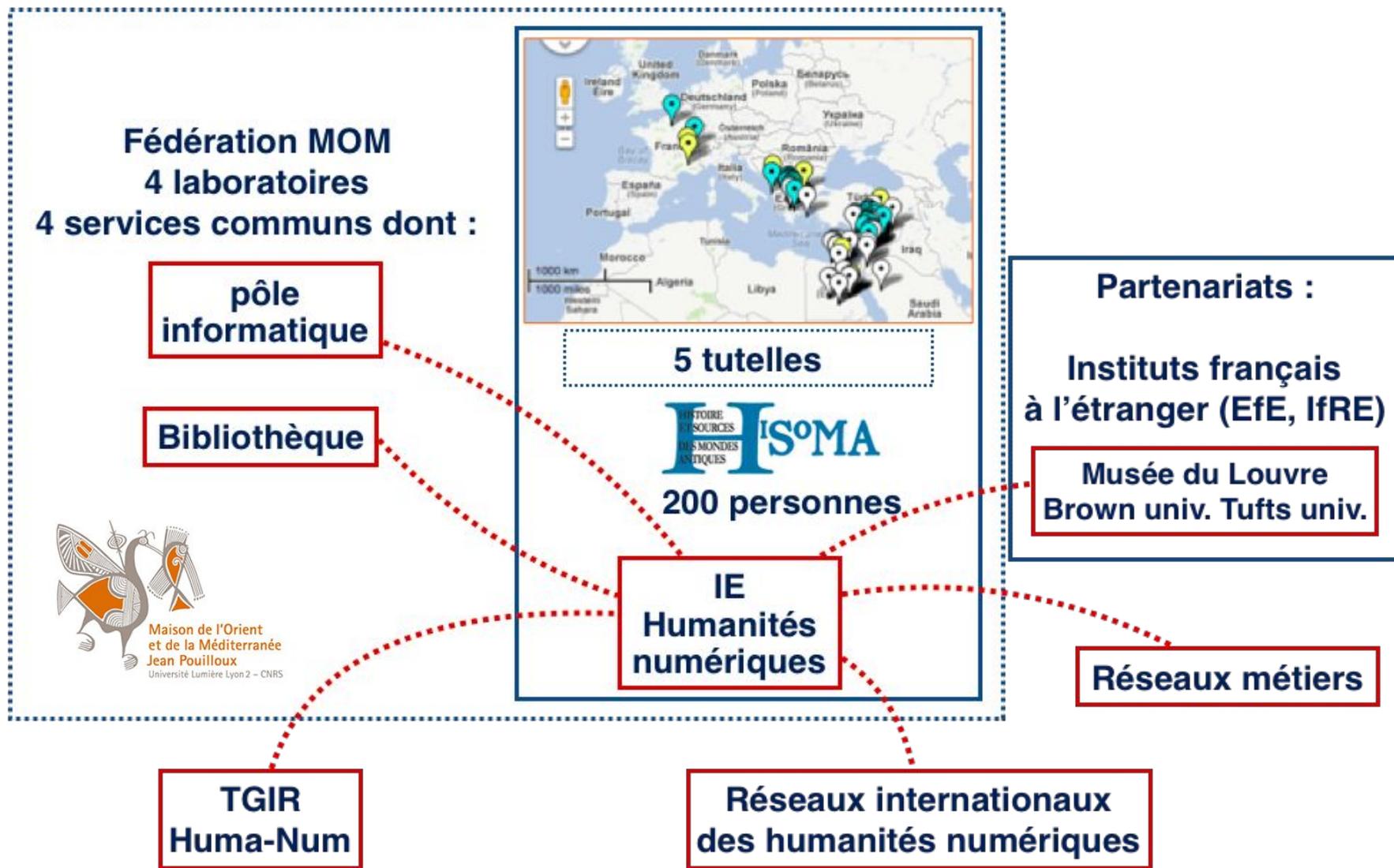
Fredocs 2018, 4/10/2018, Albi



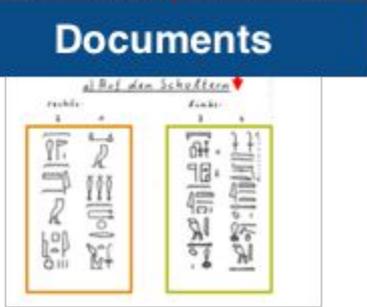
Qui parle ?

- Depuis 16 ans dans l'ESR dont 10 au CNRS
11 ans au sein d'équipes mutualisées / 5 ans au sein d'une UMR
- IE BAP F, plusieurs emplois-types successifs
conceptrice/rédactrice de sites web
chargée de ressources documentaires
chargée de systèmes d'information documentaire
chargée du traitement des données scientifiques...
- Mission principale à HiSoMA
accompagnement de projets d'édition numérique en épigraphie
(inscriptions antiques) basés sur le standard , en tant que membre de
l'équipe projet...
- "Isolement" compensé par le travail en réseau (MOM /
communautés)

Contexte : un laboratoire au sein d'une fédération de recherche



Projets d'édition critique numérique de sources anciennes

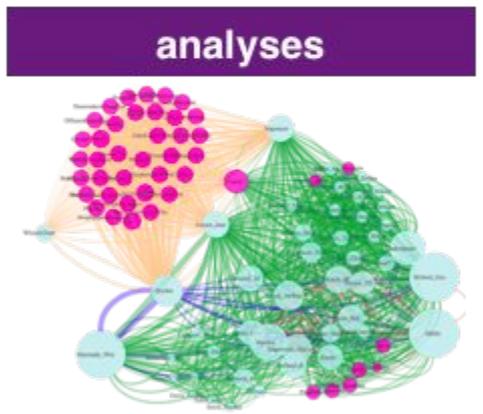


```
Upper side
/TEI/teiHeader/fileDesc/sourceDesc/maDesc/
physDesc/objectDesc/supportDesc/support/list
<item xmi:id="ckf8-upperSide" ana="upper-side">Face supérieure de la
statue</item>

Text layouts
/TEI/teiHeader/fileDesc/sourceDesc/maDesc/
physDesc/objectDesc/layoutDesc
<layout xmi:id="ckf8-rightShoulder" columns="2" corresp="#ckf8-upperSide">
<desc>Deux colonnes de texte sur l'épaule droite</desc></layout>
<layout xmi:id="ckf8-leftShoulder" columns="2" corresp="#ckf8-upperSide">
<desc>Deux colonnes de texte sur l'épaule gauche</desc></layout>

Abstract text(s)
/TEI/teiHeader/fileDesc/sourceDesc/maDesc/maContexts/maBem
<maBem xmi:id="ckf8-text3" ana="meta-egypt-cha-royal-donation-formula">
<title type="modern">Formule de donation royale</title>
<textLang mainLang="egy-a-egt"/>
</maBem>

Text edition
/TEI/text/body/tbvt[@type="edition"]
<div xmi:id="ckf8-shouldersTranlit" type="textpart"
corresp="#ckf8-rightShoulder #ckf8-leftShoulder">
<ab corresp="#ckf8-text3">
<milestone ana="layout" corresp="#ckf8-rightShoulder"/>
<lb n="1" send="verti-right-to-left"/> dy m lawt n,4 jr nawt
<lb n="2" send="verti-right-to-left"/> a hm-egr n lmn m lp-t-ew t
<milestone ana="layout" corresp="#ckf8-leftShoulder"/>
<lb n="3" send="verti-left-to-right"/> bry stw t-er n pe lmn ltr
my' jr w s5 mi n
<lb n="4" send="verti-left-to-right"/> Ny-ee-a-lmn my' jr w s5 mi
nw ltr my' jr w
</ab>
</div>
```



```
extraction
for $location in //Location
let $name := $location/@name
where starts-with($name, "E")
return $location
```

Innover ?



Plan

1. *Rappels rapides sur le PGD / DMP*
2. Innover c'est produire quelque chose de **nouveau...** oui mais comment ?
3. Innover, c'est apporter de la **valeur...** de quel type ?
4. Peut-on innover sans **controverses** ?
Est-ce toujours un **progrès** ?

Plan de gestion de données

Data Management Plan

Qu'est-ce qu'un plan de gestion de données (PGD)

Un PGD est un document

qui explique la manière dont les données sont :

- produites
- gérées
- partagées

ainsi que leur devenir dans l'après-projet.

ARCHAIDE

GRANT AGREEMENT NUMBER: 693548
PROJECT ACRONYM: ArchAIDE
PROJECT TITLE: Archaeological Automatic Interpretation and Documentation of eCeramics
FUNDING SCHEME: H2020-REFLECTIVE-6-2015
PROJECT COORDINATOR: Prof Maria Letizia Gualandri, UNIPi
TEL: +39 05022 15817
E-MAIL: maria.letizia.gualandri@unipi.it

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N.693548

Doc Title: ORDP: Open Research Data Pilot
D N° 10.2
Version: 0.2
Revision: Final Version

Work Package: 10
Lead Author (Org): Tin Evans (UoP)
Contributing Author(s) (Org): ME
Data Date: 30th November 2016

Mobilizing the Past for a Digital Future: The Potential of Digital Archaeology
Data Management Plan

Types of Data
The data produced by this proposed project will include presentations given by participants and data generated from moderated discussions in the proposed workshop. Presentations will consist of image presentations, video recording, and text files documenting a multidisciplinary two-day workshop on establishing best practices and protocols for mobile computing in digital archaeology. Dialogue and discussion generated during and after the workshop is an additional source of data. A website will be generated by the project team to publicize the event; post-text, videos, and graphic images from the event; post best practice and protocol documents generated after the event; and continue to post discussion generated after the workshop. A project white paper and final report will be produced for the NBN.

Data Standards and Capture
The Creighton Digital Repository (CDR) has committed to providing a stable and secure digital archive indefinitely in order to store, preserve, and make accessible the data generated from the proposed workshop using ObjDex. Data for this project consists of documents related to the organization, running, and follow up on the two-day workshop. Project directors in coordination with the CDR will upload event promotional material, and practical information for attendees and participants, as well as the presentations (text and PowerPoint), video, and discussion from the workshop itself. After the workshop, continuing discussion, and best practice and protocol documents will be uploaded by project directors and participants. Whenever possible, Technology has committed to providing audio and video recording of workshop events; these will be captured as uncompressed audio or video but will be converted and maintained in the form of widely-supported open formats such as MP4 that are suitable for events will be live-streamed (if not) will be made to use which are likely to remain be performed to normalize

Items with the Creighton and Dublin Core as its

All property rights in material for the conference under to both Creighton and participants and the workshop under a similar agreement. Video of presentations to be captured for this project.

University Libraries and the led by Creighton University

18.pdf 20

Figure 4.4. GRIDC Data Portal conceptual use-case.

5. DATA PROVENANCE AND METADATA

The advent of the internet and ability of researchers to move data easily, and facilities to extract data from documents or databases, has introduced a new problem when defining the origin of the data. In a multi-disciplinary/multi-institutional scenario, it is often difficult to determine the origins of the data collected, compiled and archived in a central repository. To facilitate the process of tracing the origin of data, sensors or process used and transformations (if any) and allow the validation of scientific databases, the use of metadata standards will be imposed. Data provenance is central in the validation of data and as such, no data will be archived in RDB and served to the public unless data provenance is addressed. The RDB will maintain its integrity and reliability when the process used to create them are reproducible and analyzable for defects.

Page | 16

Pas une contrainte “administrative”... Une aide aux projets

Avant : un outil d'aide à la décision

- vue d'ensemble des choix techniques, juridiques, organisationnels
- oblige au réalisme
- permet d'anticiper et d'optimiser les coûts de gestion / documentation des données

Pendant : un support pour collaborer efficacement

- exprime une vision commune sur les données
- chaque métier peut y projeter son expertise et comprendre le rôle des autres intervenants

Après : une garantie de ré-utilisabilité

- enregistre “l'histoire des données” dans un document unique
- facilite la rédaction des articles et des datapapers
- permet la réutilisation scientifique par d'autres équipes

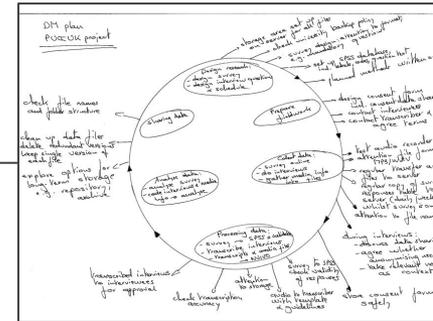
Workflow

Inventaires questions audits

Politiques

Consultations d'experts IST / informaticiens

Data Stage	Output	# of Files / Typical Size	Format	Other / Notes
Primary Data				
Raw				
Processed				
Analyzed				
Finalized				
Ancillary Data				
Ancillary Data #1				
Ancillary Data #2				



Choix d'un modèle

DCC, ERC, Univ. paris diderot, Inrap...

Saisie



Création d'outils

Rendu final

Validation technique

Révisions



**Innover c'est produire quelque
chose de **nouveau...**
oui mais comment ?**

Pas de demande : comment sensibiliser sans se “griller.” ...

(...) non seulement il faut **rejeter tout discours d'injonction, de pression moralisatrice ou culpabilisante sur la question du partage des données**, et opter résolument pour la sensibilisation progressive, mais il faut de plus **adapter les discours de sensibilisation** à la complexité du paysage des données, tenir compte à la fois de la **valeur affective** attachée aux données de recherche par leurs producteurs, des nombreuses craintes et interrogations manifestées (sur le plagiat, etc.), **des réticences, souvent légitimes, à un partage trop systématique**, et enfin il conviendrait de cibler les priorités sur les besoins les plus urgents.

Historique

2013	Découverte des DMP (Digital Curation Conference - Fredocs) puis arrivée à HiSoMA
fin 2015	Nouvelle mission responsable d'un programme du quinquennal à venir: "réflexion sur les pratiques d'édition et d'archivage"
2016-2017	Incubation <ul style="list-style-type: none">● idée d'utiliser le plan de gestion de données comme thème fédérateur● rédaction d'un projet● formations (3)● 2 actions de sensibilisation● formateur-expert dans une masterclass Dariah● discussions à HiSoMA
2018	Lancement <ul style="list-style-type: none">● Labellisation par le nouvel axe stratégique de la MOM● 4 séances

Programme scientifique
laboratoire HiSoMA
Quinquennal 2016-2020
Edition, Archives,
Humanités numériques
(AXE T)

**REFLEXION SUR
LES PRATIQUES
D'ÉDITION
ET D'ARCHIVAGE**

Séminaire

1e séance
**Judi
29 mars
2018**
de 12h30
à 14h

MILC
Salle 410
35 rue
Raulin
Lyon 7e

Introduction
au Plan
de Gestion
de Données (PGD)

Un outil pour faciliter
la gestion de ses données
numériques.

Présentation générale
et exemples dans
le domaine
de l'archéologie et
du patrimoine culturel.

HiSoMA



www.cnrs.fr

Les enjeux du « Data Management Plan » pour les chercheurs

Emmanuelle Morlock IE CNRS, Laboratoire HiSoMA (UMR 5189)
Geneviève Romier, IR CNRS, Institut des grilles et du Cloud



Pérennisation et communication de
l'information numérique
19-23 septembre 2016

Programme



URFIST de Lyon



Les données de la recherche dans les thèses de doctorat

Joachim Schöpfel, GERiICO/ANRT
Université de Lille 3



FILM MASTER CLASSE DE

Master class franco-allemande en humanités numériques I.
Utiliser/penser les données de la recherche

Deutsch-Französische Master Class in Digital Humanities I.
Forschungsdaten in den Geisteswissenschaften:
Nutzung und Konzepte

Formule

- Séances ouvertes à toute la MOM (et au-delà) pour les chercheurs et enseignants-chercheurs, les ingénieurs, les doctorants et masterants...
- de 12h30 à 14h, un jeudi par mois
- séances thématiques
- appel à contribution ouvert...

“Chaque séance s’appuiera sur des études de cas concrets issus de projets numériques en cours de développement ou de montage. Si un thème de séance vous intéresse et si vous avez envie d’y contribuer avec une « étude de cas », un ensemble de questionnements ou une présentation, merci de remplir ce [court questionnaire](#).”

Programme : séances thématiques

29 mars 2018

Introduction au PGD

26 avril 2018

Edition numérique

24 mai 2018

Bases de données - SEANCE REPORTEE

21 juin 2018

~~Conférence invitée : Marie Puren
(Parthenos)~~ Discussion ouverte

20 septembre 2018

Principes FAIR

25 octobre 2018

Aspects éthiques et questions juridiques

22 novembre 2018

Stratégies de publication de données

jeudi 20 décembre 2018

Bilan - Co-rédaction d'un rapport d'étape, proposition de modèles de plans de gestion de données.

affiches (C. Develay, chargée de com. de la MOM)

Programme scientifique
laboratoire HiSoMA
Quinquennal 2016-2020
Edition, Archives,
Humanités numériques
(AXE T)

REFLEXION SUR LES PRATIQUES D'ÉDITION ET D'ARCHIVAGE

Introduction au Plan de Gestion de Données (PGD)

Un outil pour faciliter
la gestion de ses données
numériques.

Présentation générale
et exemples dans
le domaine
de l'archéologie et
du patrimoine culturel.



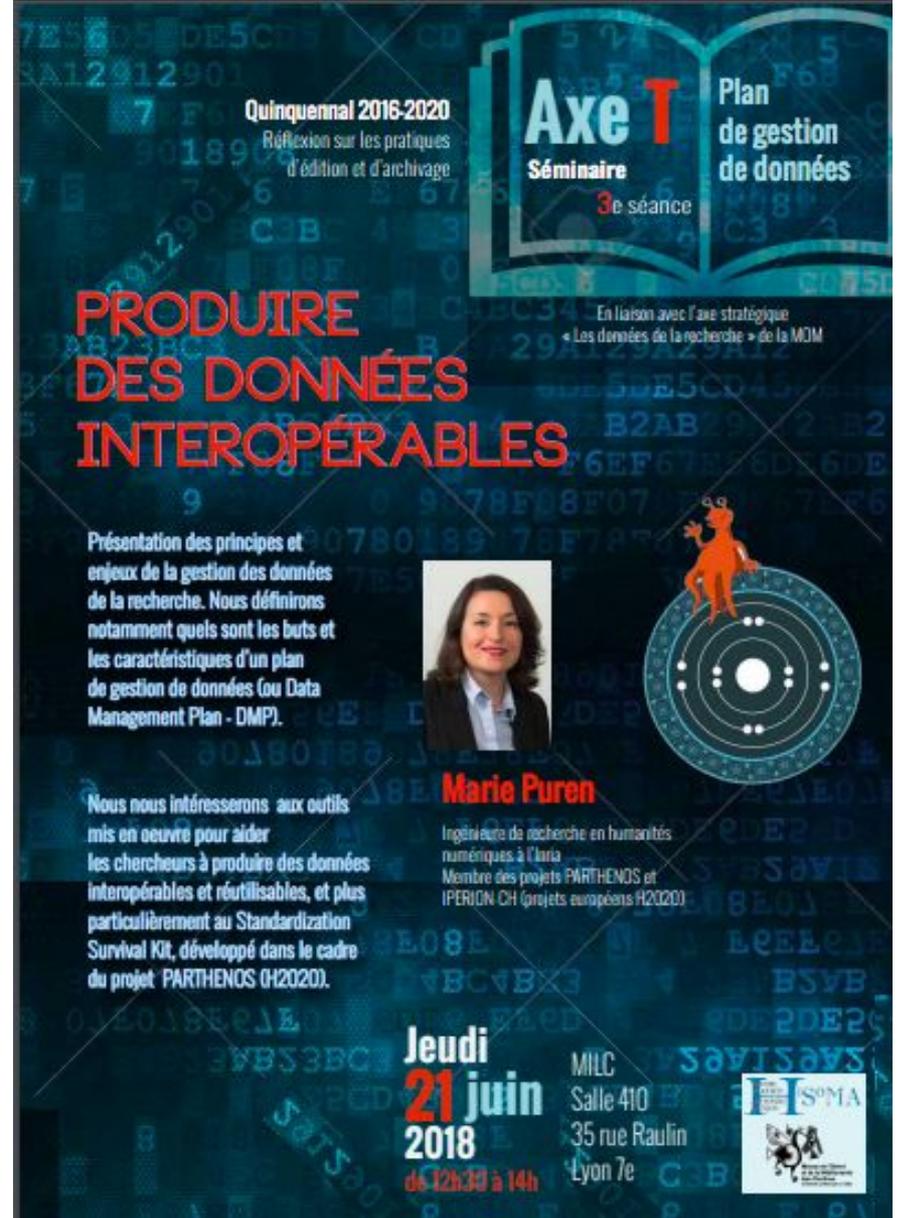
Séminaire

1e séance

**Jeudi
29 mars
2018**
de 12h30
à 14h

MILC
Salle 410
35 rue
Raulin
Lyon 7e

HiSoMA



Quinquennal 2016-2020
Réflexion sur les pratiques
d'édition et d'archivage

Axe T Plan
de gestion
de données
Séminaire
3e séance

En liaison avec l'axe stratégique
« Les données de la recherche » de la MOM

PRODUIRE DES DONNÉES INTEROPÉRABLES

Présentation des principes et
enjeux de la gestion des données
de la recherche. Nous définirons
notamment quels sont les buts et
les caractéristiques d'un plan
de gestion de données (ou Data
Management Plan - DMP).



Marie Puren
Ingénieure de recherche en humanités
numériques à l'Iria
Membre des projets PARTHENOS et
IPERION.CH (projets européens H2020)

Nous nous intéresserons aux outils
mis en œuvre pour aider
les chercheurs à produire des données
interopérables et réutilisables, et plus
particulièrement au Standardization
Survival Kit, développé dans le cadre
du projet PARTHENOS (H2020).

**Jeudi
21 juin
2018**
de 12h30 à 14h

MILC
Salle 410
35 rue Raulin
Lyon 7e

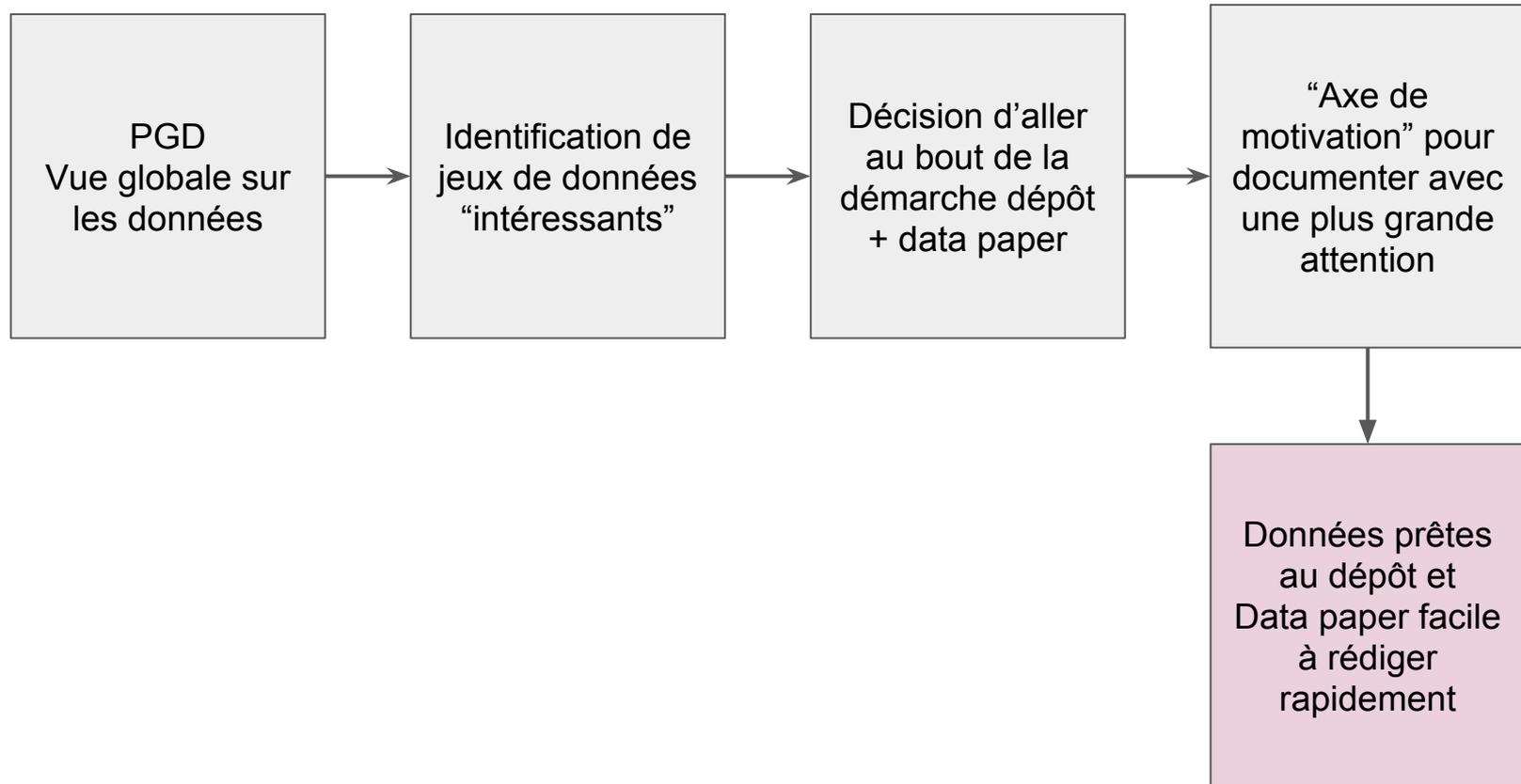
HiSoMA

Objectifs et ambitions

- aborder les multiples aspects de la gestion des données de la recherche en suivant un “fil rouge” (thème = appât)
- avoir un temps de rencontre entre chercheur/personnels IST/ informaticiens pour développer un dialogue et une “culture” commune de la donnée en interne à la MOM
- préparer le terrain d’une politique de laboratoire par le développement d’un consensus, sur une cible commune d’objectifs a minima (standards, degré de documentation)
- Développer une pratique du PGD comme “outil stratégique” pour l’aide à la décision
- Déboucher sur des produits concrets : comptes rendus, guides, vademecum, checklists...

idée force

Faire des choix “stratégiques”
car la rendre réutilisable = documenter et expliciter
= **temps chercheur** incompressible



des premiers résultats encourageants

- Fréquentation : de 12 à 25 personnes
- Mixité ? oui : ingénieurs du labo, quelques chercheurs (engagés dans des projets ERC), un chargé de projet, des doctorants, les documentalistes de la bibliothèque... les contractuels du service informatique chargés de travailler sur les archives papier et numériques, quelques personnels de soutien externes...
- Succès d'estime
- *Désormais je suis sollicitée au moment de la réponse aux appels... (avant, non !)*

La démarche : aides et freins

- Soutien de la direction et du conseil de laboratoire (budget de 400 euros)
- Encouragement à l'autonomie
- Contexte favorable ("On en parle")
 - Doranum et autres travaux de l'INIST
 - Modèles et tutoriels de Paris Diderot
 - Plan pour la Science ouverte
 - Très riche littérature (en anglais)
- 5-6 "fidèles"
- Soutien des réseaux (Renatis, MASA, Atelier inter-réseau de la MITI)
- Formule coûteuse en temps de préparation (pas vraiment de propositions de contribution spontanées)
- Sentiment de manque de compétences, d'expérience concrète (légitimité)
- Difficulté à passer directement à la pratique avec un public hétérogène et peu averti
- Le pb des compte-rendus
- Pas de perception d'un véritable besoin, sauf chez les porteurs de projets européens
- Résistances ("on le fait déjà !")

Clarifier !



Accueil

Présentation

Recherche et activités

Production scientifique

Formations

Contacts

Accueil > Recherche et activités > Séminaires de recherche > Plan de gestion des données

A⁺ A⁻ Imprimer Partager

Recherche et activités

Présentation générale
Quinquennal 2016-2020
Quinquennal 2011-2015
Quadriennal 2007 / 2010
Activités de terrain
Programmes cofinancés
Séminaires de recherche
Rencontres scientifiques
Thèses
On en parle...
Zoom sur

Plan de gestion de données 2018

Axe Transversal - Edition, archives, humanités numériques
Programme 3 - Réflexions sur les pratiques d'édition et d'archivage
Responsable : [Emmanuelle Morlock](#)

Ce séminaire sur la gestion des données de la recherche est organisé dans le cadre du [programme de réflexion sur les pratiques d'édition et d'archivage](#) de l'axe transversal Editions, archives, humanités numériques d'HiSoMA et s'inscrit également dans l'axe stratégique "Gestion des données de la recherche" de la MOM.



Séminaires annuels

Quinquennal 2016-2020

Année 2018-2019

- Formes de transmission du savoir - Axe A
- Nommer les hommes dans les mondes anciens - Axe A
- L'atelier des poètes - Axe B
- Sociétés en mutation - Axe C
- Épigraphie grecque et latine - Axe Transversal

Les plans de gestions de données (à ne pas confondre avec le plan de classement des fichiers sur un serveur...) sont des **documents synthétiques**, qui **expliquent comment sont gérées les données**, depuis l'étape de création ou de collecte jusqu'au dépôt dans des plateformes de partage et d'archivage. Ils offrent ainsi une vue d'ensemble des actions à mettre en oeuvre par

partage et d'archivage. Ils offrent ainsi une vue d'ensemble des actions à mettre en oeuvre par les chercheurs et les personnels de soutien. Il sont aussi utiles pour identifier des manques, faciliter la recherche de solutions et de collaborer efficacement. Il fournissent aussi une aide à la décision pour le développement de stratégies de dissémination/valorisation des données

- Nommer les hommes - Axe A
- Patristique - Axe A
- L'atelier des poètes - Axe B
- Constitutions mixtes - Axe C

Comment ?

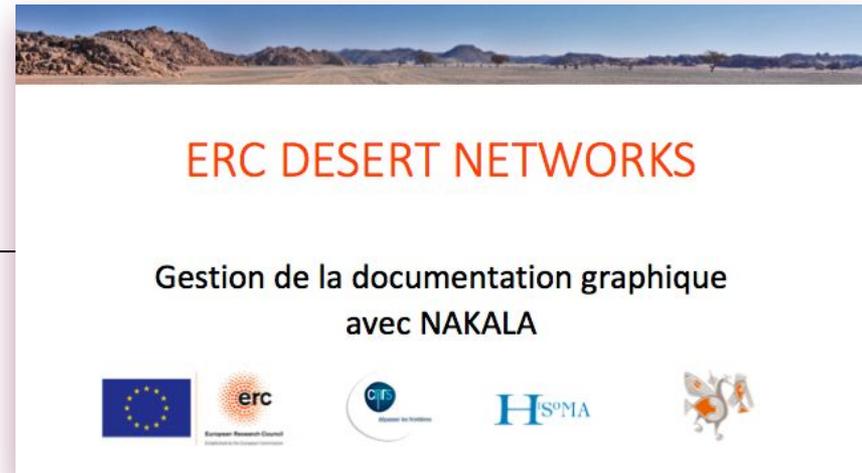
- Oser demander, proposer,
- Rédiger un projet et le faire valider (qui ne dit mot consent)
- Essayer de réinterpréter les orientations données selon un angle personnel
- Se préparer (un peu)
- Se lancer, même si on ne se sent pas prêt
- Apprendre en faisant, parfois la veille pour le lendemain...
- Créer des synergies entre activités, notamment via les réseaux (articulation local / global)
- 5-6 collègues motivés... = suffisant pour démarrer

Innover, c'est apporter de la valeur... de quel type ?

le prisme de la valeur

- économique ?
 - perdre du temps pour en gagner
 - un pis aller pour compenser le manque de moyens (vrai service de gestion de données mutualisé)
- social
 - un rendez-vous régulier pour s'approprier la démarche de planification de la gestion et du partage des données
 - des rapports plus horizontaux entre chercheurs et ingénieurs, oser afficher ses pratiques, ses questionnements, ses erreurs...
- organisationnel
 - légitime un discours sur les données (en l'absence d'une politique explicite)
 - offre des ouvertures pour trouver des points d'entrée dans les projets de recherche

Réfléchir ensemble...

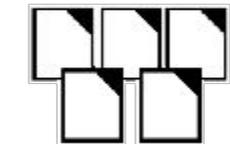


Questions

Quelles métadonnées intégrer?
(droits de réutilisation, diffusion,
destination du fichier, ...)

Métadonnées

Ouverture / fermeture des données
Comment interroger ces données à plusieurs?
Les critères pour faire des lots (sites, WP, ...)



documentation graphique

NAKALA
service HumaNum



ZENODO
projet OpenAIRE



Extraction automatique vers
Zenodo?

- dépôt d'un lot de données
archivables et diffusables (lot 1)
- dépôt d'un lot de données
archivables et diffusables (lot 2)
- dépôt d'un lot de données
archivables et diffusables (lot ...)

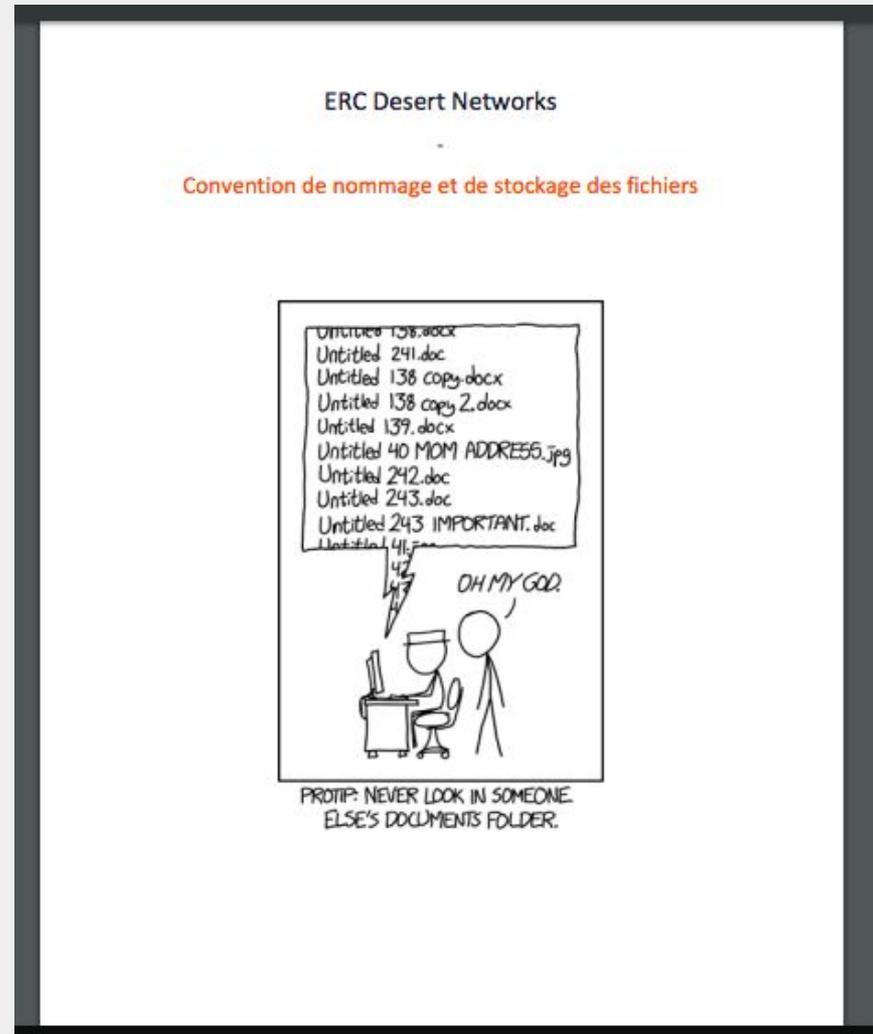


BDD Desert Networks



Prolonger par de l'accompagnement *sur mesure*

- Dans la phase de préparation du plan
 - Entretiens d'audit pour repérer des "manques"
 - relire les documents, suggérer des améliorations
 - Optimiser (moments propices, automatisation)
- Exemples
 - conventions de nommage : anticiper les opérations de tris, sélections automatisées
 - idée d'organiser des comités de validation des données à des étapes clés



**Peut-on innover sans
controverses ? est-ce toujours
un progrès ?**

Quelques pistes de réflexion

- Tout dépend du soutien du DU...
 - cela n'aurait pas existé si le conseil de laboratoire avait dit "niet"
- Objectifs non dits (tabous)
 - Une réponse au manque de moyens = conditionner le soutien des ingénieurs au respect des "bonnes pratiques" et à un engagement concret dans la "Science ouverte", objectiver par des critères clairs (*vous n'êtes pas obligés de planifier ou de partager, mais c'est obligatoire si vous voulez du temps ingénieur ou des crédits...*)
 - Valoriser les coûts dans les budgets prévisionnels insérés dans les réponses aux appels = si on pense que l'on peut apporter un soutien aux projets, ce n'est pas pour nous faire plaisir, c'est un financement en nature !
- Les personnels bap F pour la gestion des données, est-ce si évident ? Est-ce un service mutualisable ?

Concours externes CNRS 2018

N°	BAP	Corps	Emploi-type	Nbre de postes
37	D	IR-2	Ingénieur-e de recherche en production, traitement et analyse de données	1
38	D	IR-2	Ingénieur-e de recherche en analyses des sources historiques et culturelles	1
39	D	IR-2	Ingénieur-e de recherche en analyses des sources historiques et culturelles	1
40	D	IR-2	Ingénieur-e de recherche en archéométrie	1
83	D	IE	Ingénieur-e d'études en production, traitement, analyse de données et enquêtes	2
84	D	IE	Ingénieur-e d'études en sciences de l'information géographique	1
85	D	IE	Ingénieur-e d'études en analyses des sources historiques et culturelles	1

Bap D

**4 postes d'IR,
Nombreuses
missions de gestion
des données**



Bap F

**0 postes d'IR,
0 missions de gestion
des données**



N°	BAP	Corps	Emploi-type	Nbre de postes
94	F	IE	Chargé-e des ressources documentaires	1
95	F	IE	Concepteur-trice rédacteur-trice web	1
147	F	AI	Assistant-e de ressources documentaires et scientifiques	1
148	F	AI	Assistant-e des technologies de l'information et de la communication	1
184	F	T	Technicien-ne d'information documentaire et de collections patrimoniales	1
185	F	T	Technicien-ne des métiers de l'image et du son	1

Pistes d'amélioration / de développement

- **Trouver le temps de rédiger les compte-rendus**
- **Organiser des ateliers de tests du modèle de plan pour l'archéologie en développement au sein du consortium "Mémoires des archéologues" (MASA) de la TGIR Huma-Num**
 - Objectif à terme : le partager ensuite sur OpiDor...
- **Proposer des expérimentation de rétro-ingénierie de bases de données « froides » à l'aide du plan**
- **Partager l'animation avec les collègues IST de la MOM ?**
 - A minima, leur demander des interventions : pour affirmer un point de vue IST sur les projets de recherche et une offre de service sur la création de plan de gestion et de partage...
- **Faire appel à plus d'intervenants extérieurs**
 - intéressés ? venez me voir !

Pub

Enregistrer



Mission pour
les Initiatives Transverses et Interdisciplinaires



Interopérabilité et pérennisation des données de la recherche :

comment FAIR En pratique ? Retours d'expérience

27 novembre 2018, Amphithéâtre Charpak, campus de Jussieu

Merci pour votre attention !

@emma_morlock