



Une préoccupation partagée : plan de gestion des données et projets Horizon 2020

Genevieve.Romier@idgrilles.fr

Plan de la présentation

- Définitions et contexte
- Pourquoi des plans de gestion de données dans les projets H2020 ?
- Ce qui est attendu
- Ouvertures

Plan de gestion de données

- **Plan de gestion de données ou « data management plan », DMP :**

Formalisation de la façon de gérer les données liées à un projet (de recherche) :

- description des données (au sens large), métadonnées, formats...
- description du cycle de vie des données y compris après le projet (préservation long terme)
- description de la politique associée aux données (accès, diffusion, propriété intellectuelle, confidentialité...)
- aspects budgétaires...

participe à la sécurisation et à la pérennisation des données

Plan de gestion de données

- **Doit être préparé AVANT le début du projet**
 - avant que les premières données soient produites et mises à jour régulièrement.
- **De plus en plus nécessaire**
 - demandé par les organismes de recherche et les agences de financement de la recherche dès la proposition et dans le suivi et l'évaluation des projets.
- **Les grandes communautés internationales ont ouvert la voie**
 - infrastructures de recherches, projets internationaux structurés autour d'instruments ou communautés produisant et utilisant de grandes quantités de données.
- **Un DMP (et son application !) est considéré comme une bonne pratique de recherche.**

Quel lien avec H2020 ? un peu d'histoire



COMMISSION EUROPÉENNE

Bruxelles, le 17.7.2012
COM(2012) 401 final

**COMMUNICATION DE LA COMMISSION AU PARLEMENT EUROPÉEN, AU
CONSEIL, AU COMITÉ ÉCONOMIQUE ET SOCIAL EUROPÉEN ET AU COMITÉ
DES RÉGIONS**

**Pour un meilleur accès aux informations scientifiques:
dynamiser les avantages des investissements publics dans le domaine de la recherche**

Annonce les mesures prises dans H2020

Les objectifs de l'Union Européenne

- Accélérer l'innovation, permettre et accélérer la mise sur le marché des résultats de la recherche,
- permettre de nouvelles formes de recherche sur les données (data-intensive),
- inciter aux collaborations, éviter les doublons, améliorer l'efficacité,
- améliorer la qualité et la transparence,
- améliorer le retour sur investissement de la R&D.

Horizon 2020

<http://www.horizon2020.gouv.fr/>

- Le programme-cadre européen actuel de financement de la recherche et de l'innovation.
 - 79 milliards d'euros sur la période 2014-2020.
 - Appels à projets (calls) vers scientifiques et industriels de l'Union Européenne.
- contient des dispositifs pour favoriser
 - l'accès aux résultats des projets, aux publications (Open Access)
 - la ré-utilisation des données de recherche (Open Data)
 - la publication des codes permettant la production et/ou le traitement de ces données (Open Source).

A noter : démarche similaire aux USA (NSF) et dans différents pays d'Europe (Pays-Bas, Royaume Uni...)

Horizon 2020, une action pilote « Open Research Data »

« Open Research Data » s'applique de façon systématique pour 2014-2015 à certains programmes (liste précise, par exemple Research infrastructures – partie e-Infrastructures), et aux projets volontaires !

"A **Data Management Plan** to be developed enabling data preservation, on-line discoverability, authorisation and re-use of data"

N.B. :

- Il est possible de se soustraire à l'action pilote (argumentation nécessaire).
- Les coûts associés au DMP sont éligibles dans le cadre du projet.
- Les projets financés hors « Open Research Data » peuvent produire un plan de gestion des données.

Projet H2020 et DMP, en pratique

- Première version du DMP : livrable dans les six premiers mois du projet
- Au moins 2 mises à jour (livrables complémentaires, plus précis et plus complets)
- Formalisme d'un livrable (écrit en anglais, partenaire responsable, revue interne, traçabilité du document, soumis à l'acceptation par la commission ...)
- Liste d'éléments de contenu spécifiques requis
 - Guide disponible (avec un « modèle » en annexe1) : Guidelines on Data Management in Horizon 2020
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
 - Traduction en français de l'INIST :
http://www.donneesdelarecherche.fr/IMG/pdf/lignes_directrices_pgd_horizon_2020_tr_fr.pdf

Qu'y a t'il dans un DMP pour un projet H2020 ?

Une partie consacrée à **chaque** jeu de données (Data set) comprenant (entre autres) :

- référence, description, standards et métadonnées, politique d'accès, info techniques d'accès, archivage, préservation,
 - décrit en annexe 1 du guide
- la réponse à quelques questions essentielles sur la possibilité de ré-utilisation des données : les trouver (DOI), y accéder (licence), les comprendre, les utiliser, les échanger (interopérabilité).
 - décrit en annexe 2 du guide

L'annexe 1 du guide

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

- **Data set reference and name**

Identifier for the data set to be produced.

identifiant

description des données produites ou collectées

- **Data set description**

Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.

existence de données similaires

possibilités d'intégration et de réutilisations

Standards existants à défaut métadonnées prévues

- **Standards and metadata**

Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.

raisons éventuelles aux restrictions de partage des données

- **Data sharing**

Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).

partage des données : procédures, outils nécessaires, restrictions éventuelles, entrepôt...

préservation long terme, durée, volume, coût et financement

In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

- **Archiving and preservation (including storage and backup)**

Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

PENSEZ A L'ANNEXE 2

Quelles données rendre accessibles et quand ?

- 1) Les données (et métadonnées) nécessaires à la validation des résultats présents dans les publications dès que possible,
- 2) Les autres données et métadonnées en concordance avec le DMP.

Deux étapes :

- déposer les données dans un entrepôt de données de recherche.
- ouvrir l'accès aux données (licence adaptée) et les documenter (y compris avec les outils de traitement).

l'accès doit être gratuit pour l'utilisateur

Est-ce spécifique à H2020 ? Que font les « autres » ?

- Les bases sont les mêmes, les détails et le niveau de précision peuvent varier selon les projets, les financeurs, les disciplines....
- Le DMP fait généralement partie du contrat de financement («grant conditions» ou «Grant Agreement»), certains proposent des modèles plus ou moins précis.
- Domaine encore neuf qui va évoluer.

D'autres exemples de DMP

- [UK Data Archive](#) (Royaume Uni)
- [Inter-university Consortium for Political and Social Research](#) (USA)
- [Data Management Plan for Scientific Research - april 2013 \(.pdf\)](#) (Pays-Bas)
- En France, initiatives de certains établissements qui proposent des guides, modèles :
 - [CIRAD](#), [INIST](#), universités

Difficultés

- **Domaine encore nouveau, tout n'est pas prêt !**
 - il manque encore des "data archives" (mais la commission prévoit des services pour les projets H2020)
 - fiabilité des "data archives" (certification, audits...)
- **Comment assumer le coût ? en particulier après la fin du projet.**
- **Nouvelles compétences, nouveaux métiers.**
- **Qui va vérifier l'application des DMP ?**

Et vous ?

- Un plan de gestion de données est utile même hors du cadre H2020.
- Votre projet a démarré. Vous n'avez pas de DMP.
 - Il est encore temps !
- Un DMP est vivant ! Il doit évoluer.
- Vous ne savez pas comment faire ?
 - Demandez de l'aide aux spécialistes,
 - inspirez-vous de l'expérience des autres.

Références utiles

- Module INIST
[http://www.inist.fr/donnees/co/module Donnees recherche.html](http://www.inist.fr/donnees/co/module_Donnees_recherche.html)
- CIRAD, Découvrir des plans de gestion de données de la recherche en 4 points
<http://url.cirad.fr/ist/donnees-de-la-recherche>
- Workshop Aparsen-EGI :
<https://indico.egi.eu/indico/conferenceTimeTable.py?confId=2052#20140304.detailed>
- Projet (livrables)
<http://www.sim4rdm.eu/>



Parce que vos données sont précieuses !