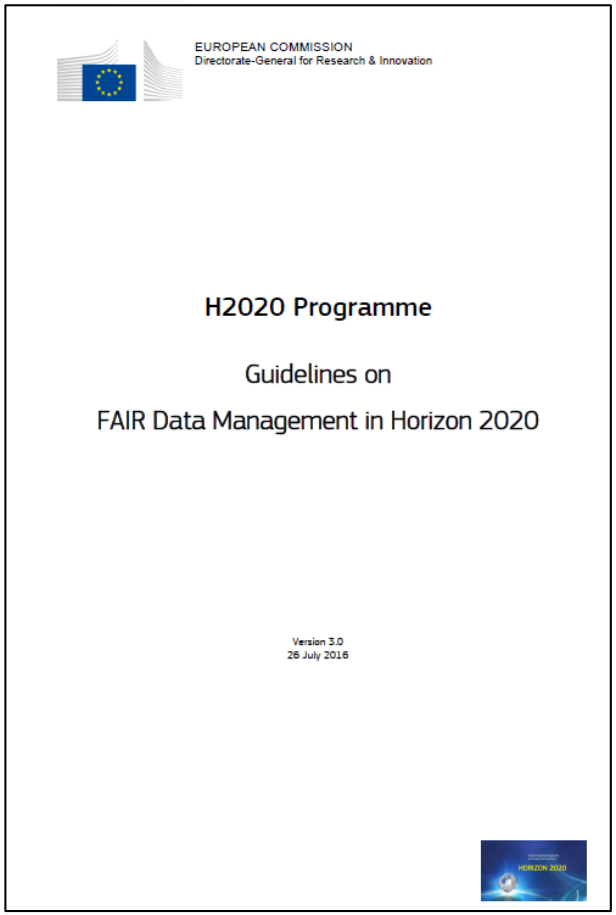
Retour d’expérience, évaluation de plans de gestion de données

O. Hologne, déléguée à l’IST

Guide H2020 « Data management »


EUROPEAN COMMISSION
Directorate-General for Research & Innovation

H2020 Programme

Guidelines on

FAIR Data Management in Horizon 2020

Version 3.0
26 July 2016

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

Guidelines

DMP component	Issues to be addressed
1. Data summary	• State the purpose of the data collection/generation • Explain the relation to the objectives of the project • Specify the types and formats of data generated/collected • Specify if existing data is being re-used (if any) • Specify the origin of the data • State the expected size of the data (if known) • Outline the data utility: to whom will it be useful
2. FAIR Data 2.1. Making data findable, including provisions for metadata	• Outline the discoverability of data (metadata provision) • Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? • Outline naming conventions used • Outline the approach towards search keyword • Outline the approach for clear versioning • Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how

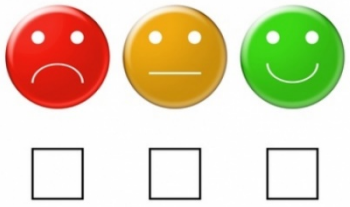
2.2 Making data openly accessible	• Specify which data will be made openly available? If some data is kept closed provide rationale for doing so • Specify how the data will be made available • Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? • Specify where the data and associated metadata, documentation and code are deposited • Specify how access will be provided in case there are any restrictions
2.3. Making data interoperable	• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. • Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
2.4. Increase data re-use (through clarifying licences)	• Specify how the data will be licenced to permit the widest reuse possible • Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed • Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why • Describe data quality assurance processes • Specify the length of time for which the data will remain re-usable
3. Allocation of resources	• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs • Clearly identify responsibilities for data management in your project • Describe costs and potential value of long term preservation

4. Data security	• Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	• To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

PGD : Evaluation

Expérience de reviewing

15 PGD / Commission Européenne



- Grille d'évaluation alignée sur la trame

- Environ 4 heures par PGD

  - Comprendre le projet
  - Comprendre si le sujet « data » a été pris au sérieux
  - Faire des recommandations pertinentes

Interpretation of the scores

0 — The proposal **fails to address the criterion** or cannot be assessed due to **missing or incomplete** information.

1 — **Poor.** The criterion is inadequately addressed, or there are serious inherent **weaknesses**.

2 — **Fair.** The proposal broadly addresses the criterion, but there are significant **weaknesses**.

3 — **Good.** The proposal addresses the criterion well, but a number of **shortcomings** are present.

4 — **Very Good.** The proposal addresses the criterion very well, but a small number of **shortcomings** are present.

5 — **Excellent.** The proposal successfully addresses all relevant aspects of the criterion. Any **shortcomings** are minor.

Half scores given

Grille d'évaluation

1. DATA SUMMARY

1.a Is header information fully provided (action ID, acronym, DMP version and date, name of the DMP responsible)?
Yes ☐ Partially ☐ No ☐

1.b Are the purpose of the data collection and its relation to project objectives explained and comprehensible?
Yes ☐ Partially ☐ No ☐

1.c Are data types and formats accurately listed? Yes ☐ Partially ☐ No ☐

1.d Is the expected volume of the data foreseen and estimated? Yes ☐ Partially ☐ No ☐

1.e Is reuse of pre-existing data described including its origin? Yes ☐ Partially ☐ No ☐

1.f Is data utility outlined? Yes ☐ Partially ☐ No ☐

Comments:

Recommendations:

---

2. FAIR DATA

2.1. Making data findable, including provisions for metadata

2.1.a Will the data be assigned a unique and persistent identifier? Yes ☐ Partially ☐ No ☐

2.1.b Will the data be registered in a searchable resource? Yes ☐ Partially ☐ No ☐

2.1.c Are data naming conventions described? Yes ☐ Partially ☐ No ☐

2.1.d Will the data be described with rich metadata (following standard practises in the field)?
Yes ☐ Partially ☐ No ☐

Comments:

Recommendations:

2.2. Making data openly accessible

2.2.a Are open access to the data and reasons for access restrictions (if any) described?
Yes ☐ Partially ☐ No ☐

2.2.b Are there instructions on how to gain access to restricted data? Yes ☐ Partially ☐ No ☐

2.2.c Is it specified where the data and associated metadata, documentation and code are deposited?

Yes ☐ Partially ☐ No ☐

2.2.d Is it described how (e.g. methods and software) the data can be accessed including relevant documentation?
Yes ☐ Partially ☐ No ☐

Comments:

Recommendations:

2.3. Making data interoperable

2.3.a Is it described how data interoperability will be facilitated, e.g. through use of data and metadata vocabularies, standards or methodologies? Yes ☐ Partially ☐ No ☐

Comments:

Recommendations:

2.4. Making data re-useable

2.4.a Is data licensing and its role in facilitating re-use described? Yes ☐ Partially ☐ No ☐

2.4.b Are data quality assurance processes described? Yes ☐ Partially ☐ No ☐

2.4.c Are procedures to decide what data will remain re-usable outlined?
Yes ☐ Partially ☐ No ☐

Comments:

Recommendations:

3. ALLOCATION OF RESOURCES

3.a Are the costs for data management estimated? Yes ☐ Partially ☐ No ☐

3.b Are data management responsibilities described Yes ☐ Partially ☐ No ☐

3.c Are the costs precisely foreseen and described Yes ☐ Partially ☐ No ☐

Comments:

Recommendations:

4. DATA SECURITY

4.a Are procedures for data backup and recovery, as well as for transfer of sensitive data (if any)-specified?
Yes ☐ Partially ☐ No ☐

4.b Are procedures for secure storage and archiving foreseen?
Yes ☐ Partially ☐ No ☐

Comments:

Recommendations:

5. ETHICAL ASPECTS

5.a Are there references to the relevant ethics aspects described in the GA or ethics deliverables?
Yes ☐ Partially ☐ No ☐

5.b Is complementary information provided, e.g. on consent for preservation and sharing of personal data as well as data anonymisation? Yes ☐ Partially ☐ No ☐

Comments:

Recommendations:

https://osf.io/kh2y6/ **DART project : Grille d'analyse et d'évaluation des DMP commune**

		Performance Level			
	Performance Criteria	Complete / detailed	Addressed issue, but incomplete	Did not address issue	Directorates
General Assessment Criteria	Describes what types of data will be captured, created or collected	Clearly defines data type(s). E.g. text, spreadsheets, images, 3D models, software, audio files, video files, reports, surveys, patient records, samples, final or intermediate numerical results from theoretical calculations, etc. Also defines data as: observational, experimental, simulation, model output or assimilation	Some details about data types are included, but DMP is missing details or wouldn't be well understood by someone outside of the project	No details included, fails to adequately describe data types.	All NSF
Directorate- or division-specific assessment criteria	Describes how data will be collected, captured, or created (whether new observations, results from models, reuse of other data, etc.)	Clearly defines how data will be captured or created, including methods, instruments, software, or infrastructure where relevant.	Missing some details regarding how some of the data will be produced, makes assumptions about reviewer knowledge of methods or practices.	Does not clearly address how data will be captured or created.	GEO AGS, GEO EAR SGP, MPS AST
	Identifies how much data (volume) will be produced	Amount of expected data (MB, GB, TB, etc.) is clearly specified.	Amount of expected data (GB, TB, etc.) is vaguely specified.	Amount of expected data (GB, TB, etc.) is NOT specified.	GEO EAR SGP, GEO AGS

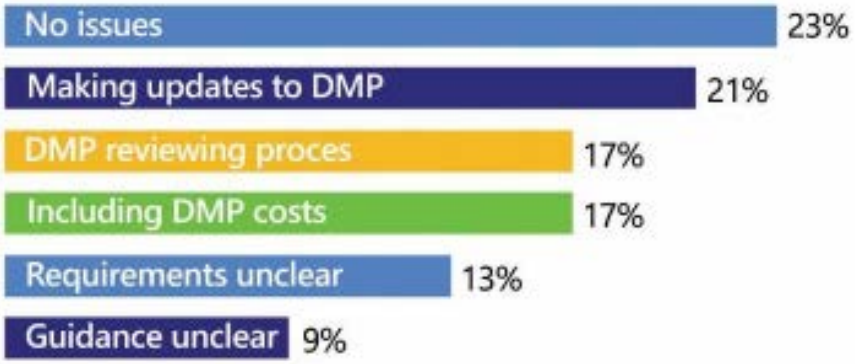
Grille	Constat	Conseils
Data Summary	• Semble parfois rempli rapidement	• Bien renseigner la version du PGD (document évolutif) • Ne pas trop fragmenter auprès des partenaires • Evaluer la volumétrie et chiffrer les coûts
F	• Métadonnées souvent mal renseignées	Dès le **stockage** : • « Rich metadata » (gestion vs publication) • règles de nommage…
A	• Fréquent choix de Zenodo	• Privilégier les entrepôts disciplinaires ou Data Inra • Exigence sur la qualité des métadonnées
I	• Peu renseigné • Un seul projet Linked Data	• Utiliser des schémas de métadonnées et vocabulaires standards
R		• Réfléchir au potentiel de réutilisation des données
Global	• Outil à utiliser dès le montage des projets, aide potentielle à la **conception** • Doit révéler la bonne prise en compte des questions **d'ouverture,** et **réutilisation** des données • Pas assez d'approche collective, patrimoniale des données	

09/01/2018

Results of the

SURVEY ON THE HORIZON 2020 DMP TEMPLATE

Difficultés rencontrées



289 réponses
50% chercheurs

5 priorités

1. Suggest relevant **standards** for my field and data type
2. **Drop-down options** based on good practice per discipline
3. Give more **examples** or suggested answers
4. Include **disciplinine specific** guidance and tailoring
5. Recommend **repositories or tools** that I can use

Evolutions à prévoir

Revise the DMP template structure

Reduce technical terminology

Provide discipline-specific guidance

Offer example DMPs and costings

Clarify DMP review processes

https://zenodo.org/record/1120245#.WlhWRzciFaQ

Ressources à consulter

- http://opidor.fr (CNRS)
- http://datapartage.inra.fr
- https://seminaire.inra.fr/data
- Couts :
  - UK Data service. Costing data management
  - OpenAIRE support
  - Data management costing tool (TU Delft)

Merci pour votre attention…

odile.hologne@versailles.inra.fr
@Holo_08

twitter