



Daniel Spichtinger

Supervision: Gerda McNeill

Horizon 2020 DMPs

what beneficiaries think and what we can learn from their experience

(“DMP Use Case Project”)

1. Introduction

Context

Data are increasingly conceptualized as inherently valuable products of scientific research and research funders both on the international and on the national level increasingly require open data and good data management in the projects they fund.

In *Horizon 2020*, the European Commission ran an initial research data pilot scheme (ORD Pilot) which was extended to the whole programme as of the work programme 2017. A key component is the obligation to create a Data Management Plan (DMP). In recent years, the objective to make data not only open but FAIR (findable, accessible, interoperable and reusable), has been gaining prominence as important principles for such DMPs.

The project

The aim of this project is to analyze DMP Use cases to identify good practices but also common challenges and mistakes amongst a number of use cases across different disciplines. The goal is to use these examples to support researchers with their DMP obligations throughout their own European projects. The results will also be reported in a publication.

The project is part of the OpenAIRE-Advance Projects RDM Task Force Group.

I am grateful to the IT team of the University Library of Vienna for their support.

This study has been supervised by Gerda McNeill (University of Vienna Library).

This presentation reflects only the author's view and the Research Executive Agency and/or the Commission and/or OpenAIRE are not responsible for any use that may be made of the information it contains.



2. Methodology

Methodology

two main components to the project

- **qualitative part:**
 - qualitative analysis of 6 DMPs and interviews with 6 cases studies
- **quantitative part**
 - manual and automated screening process to establish a whitelist of DMPs
 - survey of the DMP experiences of H2020 projects.

a) Qualitative part –Selection criteria

- Balanced thematic representation (ERC classification)
- Availability of more than 1 project DMP: (check for updates & progress)
- Geographic balance (interviews)
- Gender balance (interviews)

a) Qualitative part –thematic areas

Establishment of a short list and final selection (+ back ups) based on ERC classification covering

- **Social Sciences and Humanities (SH):** Education: systems and institutions, teaching and learning (SH 4_11), Linguistics: formal, cognitive, functional and computational linguistics (SH4_6)
- **Physical Sciences and Engineering (PE)** Web and information systems, database systems, information retrieval and digital libraries, data fusion PE 6_10 (2x)
- **Life Sciences (LS)** LS7_8 (Health services, health care research) LS7_2 Diagnostic tools (e.g. genetic, imaging)

a) Qualitative part - process

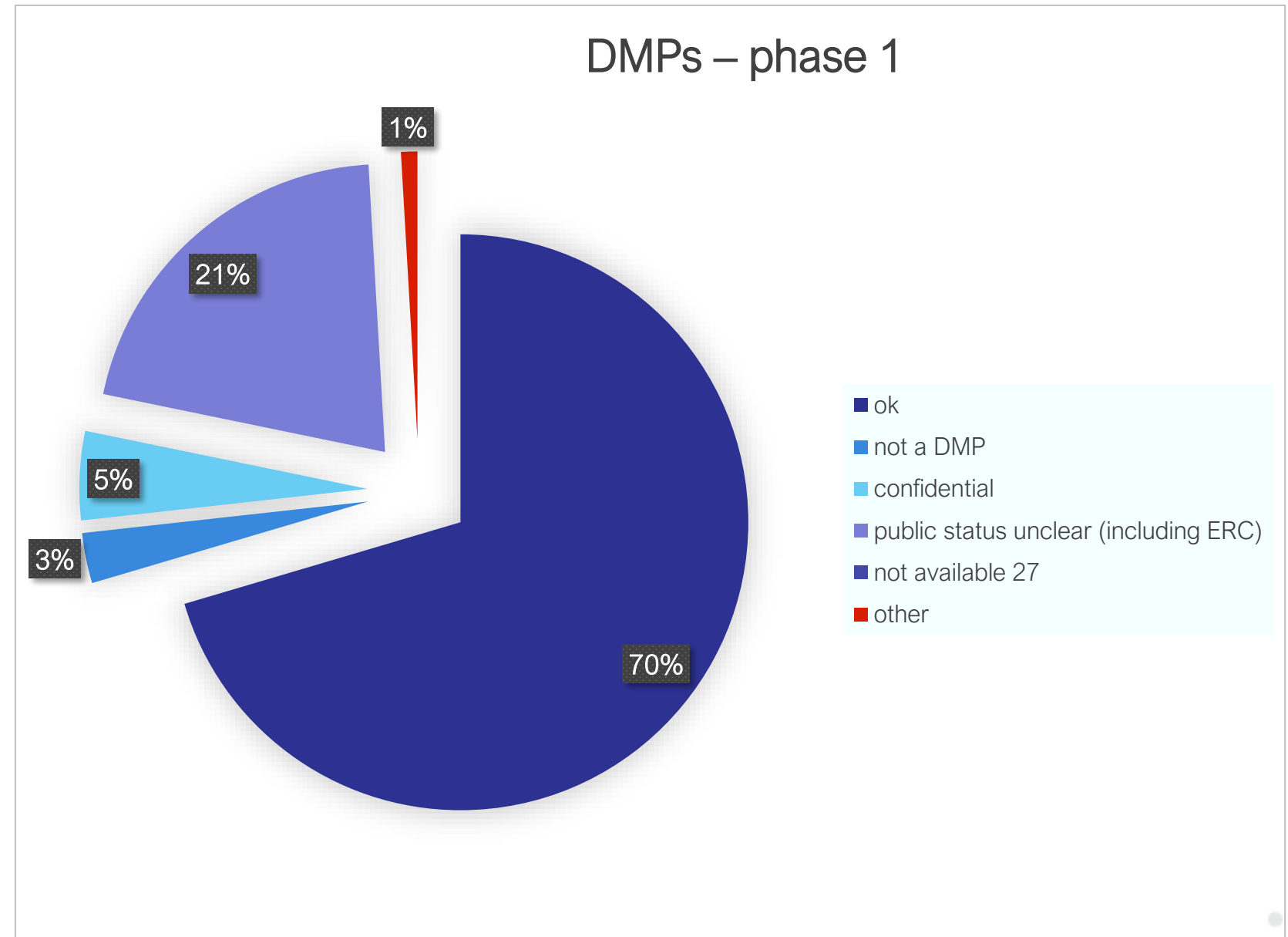
- Qualitative screening of project DMP, based on University of Ghent matrix
- Development of an interview guide
- In 2 cases interview refused, and backup candidates used
- 6 Interviews, ca 30 min each – 5 interviews in English, 1 interview in German
- Summary documents for each interview (not full transcript)
- List of interviewed projects in annex 2

b) Quantitative part

- **White List**
 - Step 1: Manual vetting: with the help of volunteers from the RDM task force to establish a white list – check: is it a DMP, is it a public document (PU deliverable)
 - Step 2: Automated vetting: remove DMPs with copyright DMP
 - White list of 840 DMPs
- **Survey of 108 projects**

b) Quantitative part – white list

- initial list of 1552 DMPs downloaded from CORDIS
- Step 1 (manual): checking document nature (DMP or not) and status (PU or not)
- 1053 DMPs that passed the screening (see pie chart). ERC DMPs not included



Examples of documents flagged for removal

Deliverable Report			
Deliverable No.	D8.1		
Deliverable Name	Data Management Plan		
WP no. and title	WP8	Dissemination and Exploitation	
Task no. and title	T8.6	Implementation of Data Management Plan	
Dissemination level: ¹	CO	Type: ² ORDP	
Due delivery date	31.12.20	Actual delivery date	30.06.17 (first release at M6, updates every 6 months)

Not public

Legal and Ethics Policy Paper

Update August 2018

Deliverable Report
D1.5

WP1 - Project Coordination and Management

Deadline: August 2017

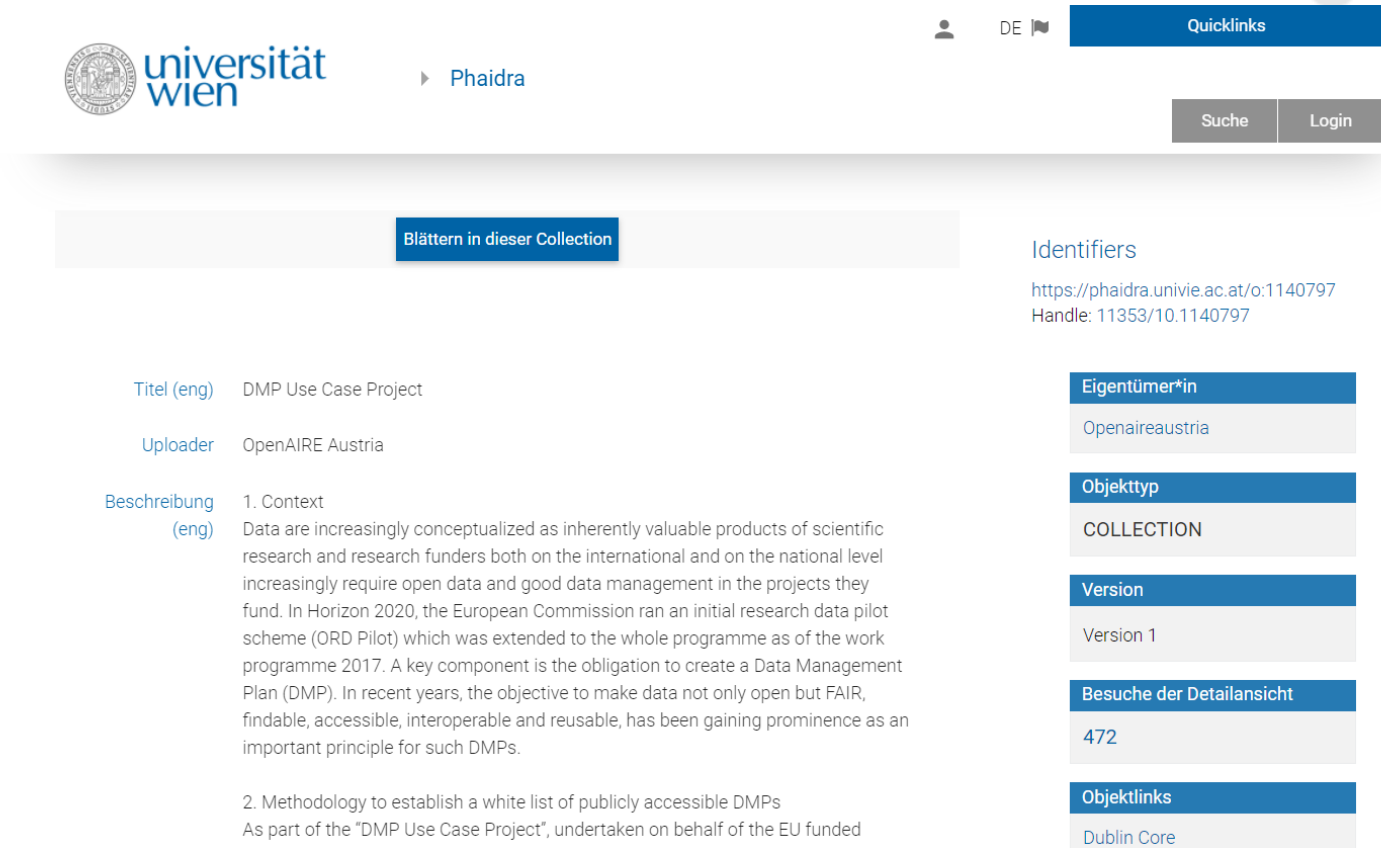
Upload by Coordinator: 21 September 2018

Entity	Name of person responsible	Short name institution	Date (Received)
--------	----------------------------	------------------------	-----------------

Not a DMP

b) Quantitative part – white list

- Step 2: automated): removal of copyright DMPs
- Final number of vetted DMPs: 840
- Transfer of whitelist to Phaedra repository (University of Vienna):
<https://phaidra.univie.ac.at/detail/o:1140797>
- OpenAIRE Blog:
<https://www.openaire.eu/blogs/establishing-a-collection-of-841-horizon-2020-data-management-plans>



The screenshot shows the Phaidra repository interface. At the top, there is a header with the University of Vienna logo, the text 'universität wien', and a 'Phaidra' link. To the right of the header are links for 'DE', 'Quicklinks', 'Suche', and 'Login'. Below the header, there is a section titled 'Blättern in dieser Collection'. The main content area displays the details of a collection titled 'DMP Use Case Project'. The 'Uploader' is listed as 'OpenAIRE Austria'. The 'Beschreibung (eng)' section contains two paragraphs: '1. Context' and '2. Methodology to establish a white list of publicly accessible DMPs'. The 'Identifiers' section shows the URL 'https://phaidra.univie.ac.at/o:1140797' and the handle 'Handle: 11353/10.1140797'. The 'Eigentümer*in' section lists 'Openaireaustria'. The 'Objektyp' section lists 'COLLECTION'. The 'Version' section lists 'Version 1'. The 'Besuche der Detailsicht' section lists '472'. The 'Objektlinks' section lists 'Dublin Core'.

universität wien

Phaidra

Quicklinks

Suche Login

Blättern in dieser Collection

Identifiers

<https://phaidra.univie.ac.at/o:1140797>
Handle: 11353/10.1140797

Eigentümer*in

Openaireaustria

Objektyp

COLLECTION

Version

Version 1

Besuche der Detailsicht

472

Objektlinks

Dublin Core

Titel (eng) DMP Use Case Project

Uploader OpenAIRE Austria

Beschreibung (eng)

1. Context
Data are increasingly conceptualized as inherently valuable products of scientific research and research funders both on the international and on the national level increasingly require open data and good data management in the projects they fund. In Horizon 2020, the European Commission ran an initial research data pilot scheme (ORD Pilot) which was extended to the whole programme as of the work programme 2017. A key component is the obligation to create a Data Management Plan (DMP). In recent years, the objective to make data not only open but FAIR, findable, accessible, interoperable and reusable, has been gaining prominence as an important principle for such DMPs.

2. Methodology to establish a white list of publicly accessible DMPs
As part of the "DMP Use Case Project", undertaken on behalf of the EU funded

b) Quantitative part

Survey

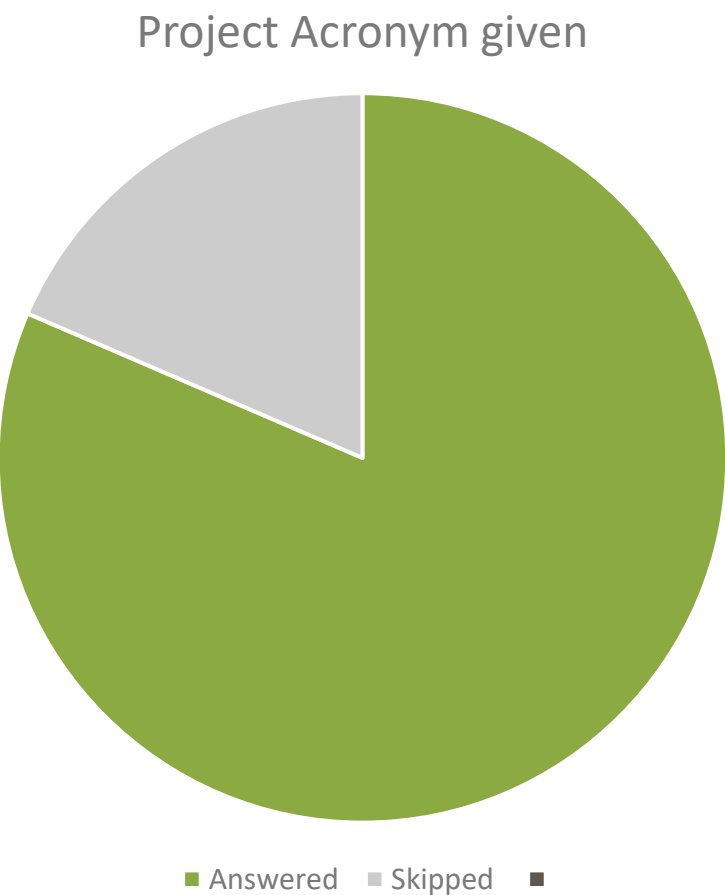
- based on interview guide for qualitative part (see part a)
- Sent to contact people of white listed DMPs
- 108 responses (Survey Monkey)

3. Results

of survey and qualitative interviews

Q1 For which Horizon 2020 project were/are you involved in the Data Management Plan? If you were involved in more than one project, choose the one you know best (optional question, skip if you do not want to answer)

Answered: 87 Skipped: 21 – projects answering the survey in annex 1, projects interviewed in annex 2



Q2 Survey : Is your Horizon 2020 project still ongoing?

Answered: 108 Skipped: 0

ANSWER CHOICES	RESPONSES	
Yes	62.96%	68
No	37.04%	40
TOTAL		108

Qualitative interview:

Completed 4 / Ongoing 2

– in one case data management activities are still continuing despite the completion of the project

Q3: Did you know about Data Management Plans before your Horizon 2020 project?

Answered: 108 Skipped: 0

ANSWER CHOICES	RESPONSES	
Yes	54.63%	59
No	45.37%	49
TOTAL		108

Qualitative interview

In most of the 6 interviews, participants were already familiar with data management in general before but in several cases the Horizon 2020 project was the first time they actually had to write a data management plan.

One project did not participate in the ORD pilot but nevertheless volunteered to do a data management plan because they thought it would be positively evaluated.

Several partners indicate that since their initial involvement, their knowledge about DM and DMPs has increased significantly (along the lines of “I would be doing it differently, if I would be doing it today”). In at least one case, the initial DMP was upgraded as the project developed.

Q4: Which work package was/is the DMP part of?

Answered: 107 Skipped: 1

ANSWER CHOICES	RESPONSES	
part of the project management work package	51.40%	55
part of the dissemination work package	21.50%	23
own work package for data management	17.76%	19
other, namely	9.35%	10
TOTAL		107

Qualitative interview

In most of the qualitative interviews DM was dealt with in the management work package. There is an indication that this is becoming more standard practice, as DMPs are becoming more widespread.

This said in one project, DM was formally part of the management WP, but informally spread over three WPs related to data acquisition and analysis.

In one case DM was split between the management and the dissemination WP

Q5: How easy/difficult was it to obtain feedback from the partners for the DMP? (1-very easy, 10 very difficult)

Answered: 105 Skipped: 3

ANSWER CHOICES	RESPONSES	
1	1.90%	2
2	3.81%	4
3	10.48%	11
4	6.67%	7
5	15.24%	16
6	12.38%	13
7	17.14%	18
8	13.33%	14
9	7.62%	8
10	1.90%	2
N/A - did not consult with partners	9.52%	10
Total Respondents: 105		

Qualitative interview

Challenges encountered referred to personal data and GDPR. The amount of time and resources was also mentioned as a challenge by the same project.

Coordination among geographically distant partners was mentioned as a challenge in one interview – though this is not necessarily limited to DM

One interviewee stated that easy and quality of feedback depended on the type of data. In this project there was at least one person per partner involved

In one project user agreements were signed with the researchers – the data belongs to them but the project has limited usage rights

Data management does not necessarily mean open – some data was opened for scientific conferences (e.g. deposited on zenodo)

Q7: Did you use a template or online tool when creating the data management plan for your project?

Answered: 105 Skipped: 3

ANSWER CHOICES	RESPONSES	
YES- template from the European Commission/European Research Council	40.00%	42
YES - template from another organisation	17.14%	18
YES - online tool	8.57%	9
NO	25.71%	27
not applicable	8.57%	9
TOTAL	105	

Most commonly named tool in comments: dmp online

Qualitative interview

Several projects indicated that the EC template was not yet available when their project started, or they were not aware of it. Most of the project did their own research, or based it on previous knowledge. Some partially used the template and augmented it with information from other sources & their communities; in one case the library was involved.

In one case each partner contributed their own part – based on experience in current project this has been changed to a more unified approach

Q8: Did you receive support when creating your data management plan?

Answered: 104 Skipped: 4

ANSWER CHOICES	RESPONSES	
YES- from other partners	39.42%	41
YES-from the library	11.54%	12
YES-from the IT departement	2.88%	3
YES- from OpenAIRE	3.85%	4
NO	26.92%	28
YES Other - please specify	15.38%	16
TOTAL		104

Qualitative interview

One project states the involvement of the university's data protection officer as well as technical supervision as regards data security

One contact states that it would be helpful to have a designated contact at the Commission to ask

One project explicitly mentions support from OpenAIRE

One project mentions help from another partner with expertise

One partner mentions the library and a data archive

Q9: Was there feedback from the European Commission / the Agency you submitted the plan to?

Answered: 104 Skipped: 4

ANSWER CHOICES	RESPONSES	
YES from the Project Officer	22.12%	23
YES from the Reviewer(s)	22.12%	23
NO	55.77%	58
TOTAL		104

Comments: most that received feedback, thought it was helpful

Qualitative interview

Nobody received feedback from the EC (or agency) PO but several received feedback from reviewers

In one case it was mentioned that the EC itself seemed unsure about the deliverable

Qualitative interview

DMP project specificities (not asked in survey)

- One DMP includes a privacy impact assessment (GDPR compliant). The data in this DMP was not in fact open: “the objective was to be accountable, not open”
- One DMP does not consider creative commons very useful – the data was not considered an original work in the sense of the German word “Urheberrecht”
- One project illustrates the progressive evolution of DMPs from one version to the next, with some questions only being able to be answered in the final iteration of the document (while in others there is little change over time)
- One project primarily used pre-existing open data. There was therefore no problem in using an open license – however, the business partners in the project were somewhat critical and saw open data more as an obstacle, rather than as an opportunity
- One project explicitly mentions that lack of community standards as a major barrier.
- One project was concerned with vulnerable groups and therefore has a strong focus on personal data consent forms, data security and ethical issues

Qualitative interview

Major challenges

- Reading all the input and turning it into one understandable document, in particular at the beginning of the project, when there was little experience
- Where to put the focus and how much details to give – internal procedures or output; also whether to tackle any data or data underlying publications (the latter strongly preferred)
- Understanding the technicalities
- How to create the DMP from scratch with zero experience
- Understanding the requirements and convincing partners to submit thorough information (done through peer pressure). Easier in newer projects
- Covering all partners, some of them in non-EU countries where different national policies apply

Q11: Do you consider the development of a data management plan useful beyond it being a requirement from the side of the European Commission

Answered: 107 Skipped: 1

ANSWER CHOICES	RESPONSES	
Yes	53.27%	57
Partially	28.97%	31
No	17.76%	19
TOTAL		107

Lots of different opinions in the comments

Qualitative interview

- We turned something that was initially a chore into a Socratic work and learned a lot from it
- Not every project needs a DMP; a simple checklist would suffice
- Its challenging having to write a DMP through a pre-existing template because you need to fit your project to pre-existing guidelines. At the same time, it is also useful because during the work one can lose sight of FAIR data and the exercise reminds you to remain on track. A more advanced project might benefit from having its own template though
- The only thing it was useful for was to clarify in project meeting which datasets we were talking about. For the overall objective of the project a DMP was not very important
- Very important to be done for each project (regardless of EU funding) but needs different approaches and categories based on the size and the nature of the project (currently not much of a distinction whether it is a project with 500 partners with a lot of shared data or 5 “friends and family”)
- DMPs very useful, also for projects which deal with vulnerable and marginalized groups and long-term curation and preservation we are switching from destroying data to archiving data after project end.

Q12: Did you publish your data management plan somewhere

Answered: 105 Skipped: 3

ANSWER CHOICES	RESPONSES	
YES - in a repository	22.86%	24
YES - on the project website	30.48%	32
YES - somewhere else	8.57%	9
NO	38.10%	40
TOTAL		105

Comments mention zenodo, cordis, transfer to partners

Qualitative interview

- Many interviewees were not sure and had to look it up
- Many stated it was published on their website (some of which, however, are no longer online)
- In one instance the project was contacted by OpenAIRE and uploaded the DMP to OpenAIRE (also all other deliverables)

Quantitative & Qualitative interview

Time and resources spent for many quite difficult to estimate, very different estimates in quantitative part

Qualitative feedback:

- Easier now that there is a template available
- One or two months during the whole project duration
- Good idea to spend some time on DM when planning, then it is easier in implementation (“something well planned is half done”)- normalize and routinise DMPs
- Assistance from project unit
- 2 months for the last version, due to revisions from different partners about one week full time for the first one and then consecutively less (because framework already exists)

Qualitative interview

What kind of support is needed and who *should* provide it?

- Reference contact in the Commission to provide training and advice
- A support paper which contains the requirements from the EU as concretely as possible – a matrix when then just need to be applied. Research support organizations should execute that, single researcher should have an overview and operational support
- Sustainability questions are important, including how to pay for data management after the end of the project; what are the limits to make data FAIR but at the same time sustainable. Larger infrastructures (ERICs) can help
- The best approach is to have someone in the data community with expertise to help; the data community should be more approachable for everyone
- For bigger organizations the library can provide support (and sometimes also the data archive)

Qualitative interview

Final thoughts

- Need for awareness raising (in particular as concerns legal regulations for personal data)
- Integrating AI into data processing: taking issues connected with that into account (very new and not always included in DMPs)
- Zenodo is a useful tool and collaboration with OpenAIRE works well – we need sufficient political will to continue that
- When project ends data tends to disappear, people save data in different repositories which make it very dispersed – the best solution would be to have one repository ,although a monopoly can also pose problems. We may need a global agreement to releasing open data (COVID could be an opportunity)
- We often fail at longevity both as concerns tools and repositories (will they still be here in 5 years?) – they don't always allow you to take data out in accessible formats in an easy way
- Templates are very useful; it helps to think about data collection but also use (even after project end)

4. Conclusions & Recommendations

Personal opinion of the author
not necessarily endorsed by Uni Wien and/or OpenAIRE

Conclusions and Recommendations -1

- For a significant number of projects (49/108) Horizon 2020 was the first time they encountered a Data Management Plan. This underlines the importance of funder mandates to spread good data management practices.
- Results from the qualitative interviews indicate that knowledge increased over time and that some have significantly developed their practices since their work on the project – this points towards the fast development of the area of data management but also to corresponding increase in competences.
- The acceptance of data management plans as more than a bureaucratic exercise mandated by the EC is surprisingly high: 82% find it useful or partly useful

EC policy action / Horizon 2020 ORD mandate has had a significant impact

Conclusions and Recommendations -2

- Having a DMP as part of the Work Package on Management (as opposed to dissemination or others) seems to be become the norm, in particular for small to mid-sized projects which do not have data science as their focus. We would therefore generally recommend projects to follow this approach, if there are no good reasons to do otherwise, but to also ensure links with the dissemination work package.
- In general, having one person among each partner organization responsible for data issues is a good practice (except potentially in very small projects). There also needs to be one person that takes overall responsibility for the project DMP – a DMP should not simply be made up of the parts delivered by the partners (“Frankenstein approach”) but form an organic whole

Conclusions and Recommendations -3

- Templates are clearly important: 40% of the survey participants used the EC/ERC template. Some ask for a more tailor made approach, which could be done by providing a EC approved data management tool.
- Support was primarily received from other partners, in some cases also the library and in a minority from OpenAIRE
- In the qualitative interviews, none of the participants received content feedback on the DMP from the PO but some did receive feedback from the reviewers. In the qualitative survey, the majority (55%) received feedback from neither but those that did found it helpful
- Especially beginners report a feeling of being lost and, in particular before the template was available, had to do a significant amount of self-learning (qualitative interviews)
- A number of interviewees as for a contact at the EC to contact for help
- **I would therefore recommend to set up a “one-stop-shop for Horizon research data management”, akin to the IP helpdesk (could be done through a public procurement procedure)**

Conclusions and Recommendations -4

- Interviewees often had to check whether their DMPs were available – and if yes, most often made them available on their project website (this chimes in with the finding from the quantitative survey) – however, several websites were no longer up and running
- This points to the importance of CORDIS as a source for public DMP which, however, is also not well known as a source for DMPs and thus also for the need to raise awareness to deposit DMPs in repositories to ensure preservation.

Next steps

- Publication (probably Open Research Europe)
- Presentation at EARMA Conference
- Presentation to REA taskforce? (tbc)

Potential further work (to be funded)

- Analyzing the IP provisions of those (ca 200) DMPs that contain copyright
- Publishing the ERC DMPs if possible



universität
wien

Universitätsbibliothek

Thank you!

Daniel Spichtinger

daniel.spichtinger@univie.ac.at

daniel@spichtinger.net

Annex 1 – projects that filled in the survey

- 87 gave their acronym, 21 preferred to remain anonymous

UPTIME CENTAUR Inclusive PublicSpace ADG787258 InDivEU MILEDI PAPERCHAIN ArchAIDE ReMAP HOVER a GeoERA project EOPEN SEAFOODTOMORROW IMI ConcePTION Adult Education as a Means to Active Participatory Citizenship (EduMAP) AfricanBioServices Consortium INFACt CarE VICINITY'FormilkEOPEN scan4reco, ARCH TITANIUM 766955 CARBAFIN RAWFIE HBM4EU STARGATE AUTOPILOT <http://aida-space.eu/> eXtreme-DataCloud proDataMarket SUNSET Project My-TRAC FLORA, Transforming Transport ROSIN IIT GreenCharge GoGreen Routes NanoCommons SAFIRE ADVANCE GenTree (project 676876) Net4SocietyESSNUSB MEISTER LIQUEFACT Car-E CleanSky 2 MIDAS ERC FIDUCEO MARCONI ARCHES SafeWaterAfricaGreenCharge LANDMARK PROSEU RESOLUTE (IMI) AfricanBioServices I-Media-Cities UNEXMIN YAKSHA HIT2GAP Charisma NoAW GAIA-CLIMCUREAfricanBioServicesEarthServer2SECUReALIGNEDRETOPEABio4CompHyper 360ERC Starting GrantComfort, TriAtlasFoTRRIS, RAISD, MiCREATEELIOTMeMADFET OpenRINGOLEVEL-UPREPSSponGESAdult Education as a Means to Active Participatory Citizenship (EduMAP) (Contract No. 693388)DYNACOMP EduMAP ENGAGE

Annex 2 – projects interviewed

CareGiverPro

READ

EURHISFIRM

FREME

LUCA

EDUMAP

Projects not available for interviews:

AfriAlliance, APOLLO