



Needs Assessment Report: Gap Analysis

D1.2.1

| | |
|----------------------------------|--|
| Author(s) | Joy Davidson with content contributions from Sarah Jones, Jonathan Rans, and Angus Whyte from the Digital Curation Centre (DCC). |
| DOI | 10.5281/zenodo.801808 |
| Document version and date | v.01, 30/05/2017 |
| Dissemination level | Public |



Introduction

The Research Output Management through Open Access Institutional Repositories in Palestinian Higher Education Institutions (ROMOR) project aims to improve the management, visibility, and accessibility of scientific research outputs in Palestinian HEIs by establishing new or enhancing existing Open Access Institutional Repositories (OAIRs), improving institutional capacity for the management and sharing of research outputs held within the repositories, and developing and/or refining curricula to ensure that emerging researchers are better able to manage their work across the entire research lifecycle.

Planning for Open Access Institutional Repositories (OAIRs) requires identifying key stakeholders who would support us in realizing the objectives of the project. There are so many layers of work to manage institutional repositories. Universities libraries, research offices, information technology departments, academic departments, university administration work together side by side to make sure that the OAIR process proceeds smoothly and sustainably over time. They share the responsibility of capturing the research output, organizing it and ensuring that long term availability and preservation are maintained, educating the scholars and the researchers about their privileges and rights as the authors of those works, and helping them understand the larger information policy and the copyright issues that touch their works.

In the first four months of the project, the Palestinian partners conducted two surveys. The first aimed to assess researchers' current practices, and the second explored institutional support staff capacity. The four participating institutions include:

- The Islamic University of Gaza (IUG)
- Al-Quds Open University (QOU)
- Birzeit University (BZU)
- Palestine Technical University-Kadoori (KAD)

The EU partners have conducted a survey based on a survey carried out by DCC in 2015, to assess the current practice of Open Access and Research Data Management (RDM) in the UK, Europe, Australia and USA. The four participating EU institutions include:

- Vienna University of Technology (TUWIEN)
- The University of Parma (PARMA)
- The University of Brighton (BU)
- The University of Glasgow (GLA)

A workshop bringing together the key stakeholders of the ROMOR project was held on April 3 2017. Participants were invited at an early stage of the project, before the actual planning of OAIR models and before implementing or refining them, as we believe participants' opinions will shape the whole process. Our stakeholders' views will be incorporated into all aspects of the project, as the work progresses.

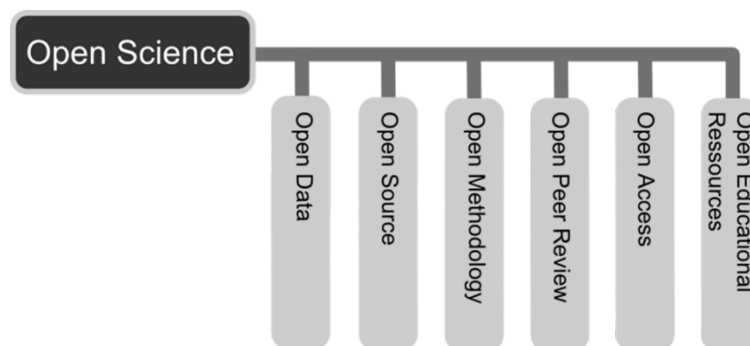
The main objectives of this workshop were:

- To present the findings of a needs assessment survey study that was carried out with researchers and support staff in four Palestinian Higher Education Institutions (PS HEIs) between December 2016 and February 2017
- To present good practice in establishing RDM and OA services from the European partners universities
- To engage the stakeholders in discussions to identify their requirements, interests or concerns regarding the OAIRs
- To setup requirements for doing the gap analysis and the road map for Research Data Management in PS partners through OAIRs.

The following report discusses a number of gaps that were identified and presented at the Needs Assessment workshop.

What is Open Science?

Open Science is more than just open access publications – although this is a key aspect. It involves a range of activities including making methodologies more transparent, providing access to research data underpinning findings, and employing open source technologies wherever possible.



CC-BY Andreas Neuhold https://commons.wikimedia.org/wiki/File:Open_Science_-_Prinzipien.png

Intelligent Access

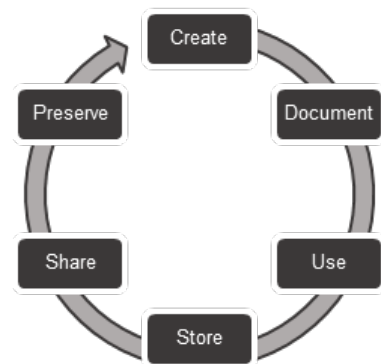
Open Science requires more than simply making your outputs available on the internet. To be truly useful, additional effort is necessary. What we are really seeking is to provide ‘intelligent access’ so that people can understand the research whether they are validating published findings or building upon them (reuse). Tim Berners Lee developed a 5 Star model that helps to clarify various levels of openness. The level of openness selected for different data outputs will be impacted by many factors including their importance, the funds available, and human resources. Good research data management over the entire research lifecycle is crucial for enabling intelligent access.

- ★ make your stuff available on the Web (whatever format) under an open licence
- ★★ make it available as structured data (e.g. Excel instead of a scan of a table)
- ★★★ use non-proprietary formats (e.g. CSV instead of Excel)
- ★★★★ use URIs to denote things, so that people can point at your stuff
- ★★★★★ link your data to other data to provide context

Tim Berners-Lee's proposal for five star open data - <http://5stardata.info>

What is research data management (RDM)?

Research data management and curation is “the active management and appraisal of data over the lifecycle of scholarly and scientific interest¹”. Managing research data helps researchers to be confident of their findings and to be certain that their work will stand up to scrutiny. Essentially, RDM is just part of good research practice.



Research data lifecycle

What sorts of data need to be managed?

It is important to be aware that research data is not limited to structured datasets. A huge variety of formats constitute research data, including but not limited to ‘raw data captured from instruments, derived data, documents, spreadsheets and databases, laboratory notebooks, visualisations, models, software, images, measurements and numbers.’²

Why bother?

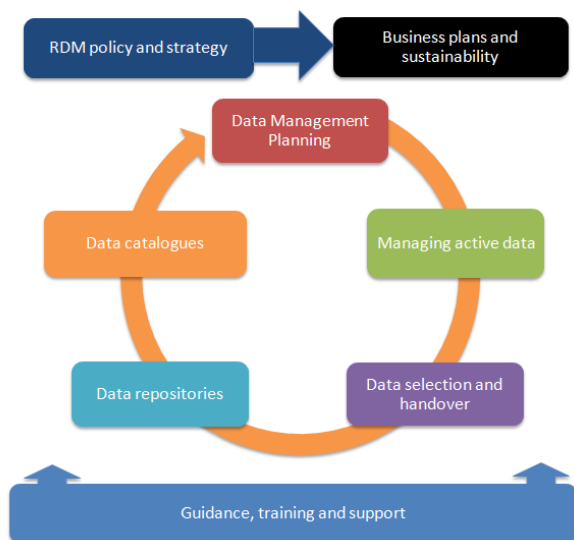
While compliance with funders’ mandates is the most obvious driver motivating HEIs in Europe to develop and implement a suite of RDM support services and infrastructure, good RDM infrastructure can help to realise many other benefits including:

- Better research practice (easier to validate and/or reproduce findings)
- Increased impact through better visibility of research outputs
- Increased awareness of researchers’ needs leads to better investment into research infrastructure (i.e., selected systems are valued and used)

¹ As defined by the Digital Curation Centre (DCC) <http://www.dcc.ac.uk>

² University of Glasgow Good Management of Research Data Policy http://www.gla.ac.uk/media/media_435489_en.pdf

- Easier reporting for internal and external assessment (e.g., REF2020)



As a reference point, the Digital Curation Centre (DCC) introduced a research data management service model which helps break down RDM infrastructure into specific activities to assist with planning and implementation³. ‘Soft’ infrastructure aspects, including policies, business planning and training, underpin the more technical infrastructure requirements in the centre which is based around the data lifecycle. It is important to note here that this is an idealised view of an RDM service model. In reality, most institutions will have some services in place as well as many gaps.

The observations presented below reflect the gaps identified through analysis of the ROMOR academic staff and managerial staff surveys carried out as part of WP1 in December 2016 and January 2017. The findings are mapped to the DCC’s RDM Service model and results are compared to the RDM landscape in the UK. The UK landscape has been drawn from the results of the 2015 RDM survey⁴ carried out by the Digital Curation Centre and also by experiences gained through delivering DCC’s programme of institutional support⁵.

Awareness of existing RDM policies and strategies in Palestinian State (PS) HEIs:

- **65% of researchers think they do not have an RDM policy at their institution.**
- **Most Palestinian HEIs do have ad-hoc policies currently in place**

Lack of awareness of institutional policies is not unusual. In fact, it is a common problem in most EU HEIs. Researchers can feel somewhat bombarded with the numerous policies and strategies governing their working practices – both those from the institution they work in but increasingly from external sources such as funding bodies, publishers and government. In many cases, the policies themselves tend to be lengthy and full of jargon which can be very off-putting to researchers who may struggle to find adequate time to digest the contents.

It is a good idea to develop policies with the involvement of researchers to ensure that they are realistically implementable. It is also a good idea to trial the policy widely before ratifying it. It is important for HEIs to be aware of how much work might be needed across the institution to educate stakeholders about the policy content, to monitor compliance and to handle cases of non-compliance. Monitoring and ‘policing’ compliance can be very costly in terms of staff time and financial resource. It may in many cases be better to offer incentives for complying with

³ RDM Service Model <http://www.dcc.ac.uk/resources/developing-rdm-services>

⁴ Results from DCC RDM survey which was circulated in 2015 to assess institutional progress towards EPSRC Policy Framework on Research Data compliance deadline .

The 2015 DCC survey collected responses from 60 institutions

⁵ <http://www.dcc.ac.uk/tailored-support/institutional-engagements>

policies rather than punishing those that do not comply (e.g., linking to career development and promotion).

General aspects of good practice that have been identified in the UK:

- Establishing an institutional working group that is comprised of all relevant stakeholders (service and infrastructure providers, researchers, ethics, administration, finance).
- Having at least one senior management representative chair the working group
- Emphasizing RDM as a core aspect of good research practice rather than linking too closely to institutional policies and/or funder mandates. Good research practice tends to motivate researchers more than compliance does.
- Building in regular review periods. This approach will allow HEIs to develop and adapt RDM related policies. More importantly, regular reviews and updates will allow the RDM working group to consider the feasibility and sustainability of policy based on experiences in rolling out and testing pilot infrastructure and support services.
- RDM activities should be proportionate to the value of research being undertaken. It is important to state here that we do not mean just the financial value of the research being undertaken but potential value in terms of impact of the research as well. Many researchers feel threatened by the amount of effort that might be needed so it is crucial to ensure researchers that RDM effort should be fit for the research being undertaken. It is important to avoid gold plating data management plans – instead they should be realistic.
- The policy should be explicit about who the policy applies to and what is expected.
- The location of the RDM pages in a ‘Research’ section of the University website rather than the Library pages. It is important to make sure that support and guidance are visible and easy to find (i.e., are located somewhere that researchers would use regularly).

Possible implementation challenges:

- Try to have a clear idea about the amount of all research activity taking place across the HEI. Without a fairly good estimate, it will be difficult to pitch the policy to the right level. Consider the scope of the policy. It is unlikely that an HEI will be able to develop the policy covering all research activity in the first attempt. Consider piloting an RDM policy with a small group such as those in receipt of external funding (e.g., H2020).
- Measure the effort and support required to implement the policy. Be sure that the institution has sufficient resources to support the implementation before the policy is ratified.
- Consider how all staff will be made aware of the policy and what is expected of them.
- Consider how the HEI will be able to measure compliance and how non-compliance will be handled.

Support for Data Management Planning in PS HEIs

- **5% of researchers have access to dedicated data management planning support**
- **Most HEIs indicated that this as an area they were keen to address**

In the UK, most funding bodies require researchers applying for funding to submit a data management plan (DMP). DMPs ‘typically state what data will be created and how, and outline

the plans for sharing and preservation, noting what is appropriate given the nature of the data and any restrictions that may need to be applied'⁶. In response, many HEIs have developed policies that require researchers to produce a data management plan as part of their grant application workflows. Some HEIs have developed comprehensive policies applying to all research conducted by staff and postgraduate research students at the University, regardless of whether it is funded or unfunded. In other HEIs, only those seeking external funding are required to develop a data management plan. In either case, it has become clear that researchers require assistance and support when completing their DMP. There are some questions in DMPs that only the researcher can answer, such as the nature of the work to be carried out and aims for sharing and reuse. Other questions such as those relating to backup of data and long term preservation are generally aspects that researchers will need assistance with.

Aspects of good practice:

- Make sure researchers know what their funders expect.
- Provide institution-specific guidance to help researchers answer questions in DMPs.
- Grants and Contracts staff have been briefed to ensure that RDM is part of standard operating practice, at both pre- and post-award stage.
- Research web pages should include a distilled version of the RDM policy and be clear about what is expected of researchers.
- Provide links to related resources, such as Ethics and Data Access policies and identify relevant clauses to consult to reduce burden of researcher and to avoid confusion.
- Wherever possible, link to external resources being provided to fill any gaps rather than duplicating effort unnecessarily. For instance, linking to the DCC's DMPonline tool⁷ and funder policy table⁸.

Possible implementation challenges

- Providing links to relevant policies is a good first step. However, given the length and sheer number policies to be consulted, it can be helpful to draw out relevant sections or clauses to reduce the burden on researchers trying to familiarize themselves with what is expected of them.
- Develop institution-specific guidance to help researchers complete data management plans.

Active Data Storage in PS HEIs

- **3% of researchers store their data on institutional servers**
- **23% of researchers have lost data**

HEIs should provide mechanisms and services for storage, backup, registration and retention of research data during a research project and after its completion.

Aspects of good practice:

- Making sure that researchers have access to systems that back up their work

⁶ DCC definition <http://www.dcc.ac.uk/resources/data-management-plans>

⁷ <http://www.dcc.ac.uk/dmponline>

⁸ <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>

- Helping researchers to develop good practices of transferring their data to a managed environment when using external systems in the field (flash drives and laptops).
- Understanding your HEIs storage needs – ensure that preliminary investigations have been undertaken into the amount, type, access and storage types of research data practices across the University.
- Reviewing internal systems and processes to optimise RDM services and efficiencies
- Being aware of services from external providers (risks, benefits)
- Choosing flexible solutions that scale

Possible implementation challenges:

- Ability to accurately estimate short and mid-term data storage needs can be difficult
- Piloting solutions across the entire institution for all research activity could be tricky – hard to find something that works equally well for all disciplines

Data Selection and Handover in PS HEIs

- **HEIs with repositories find it difficult to motivate researchers to deposit materials**

In the UK, data that is required to underpin published findings should be retained in an appropriate archive or database for a minimum of 10 years. such as project websites. In addition, the policy outlines that where possible and applicable, research data that is not in a digital format be digitised to facilitate access.

Aspects of good practice:

- Being clear on what data might need to be retained (reproducibility, validation)
- Being clear on who to contact to help with selection and appraisal of research data
- Being clear about how long the data needs to be retained
- Helping researchers to be clear on any restrictions on access
- Educating researchers to avoid putting their data onto their own project websites to provide longer term access. Most project websites are not sufficiently managed and most will not be retained for the minimum timeframes required by external funders.

Possible implementation challenges:

- Availability of advice and guidance on selecting data that should be retained – how to get this to everyone who may need it?
- Availability of guidance and support in selecting non-digital data that should be digitised
- Providing sufficient guidance on assessing low, moderate and high risk projects may be needed

Data Repositories in PS HEIs

- **85% of researchers feel they are responsible for archiving data**
- **72% of researchers are not aware of discipline specific repositories**
- **Most PS HEIs will provide storage on request**

In the UK, Research Councils UK's Open Access policy states that published papers should include a statement describing how, and on what terms, supporting research data may be accessed. In the UK, one of the key benefits that many HEIs are using to encourage researchers to deposit their data, is by assigning a DOI that can be cited.

Aspects of good practice:

- Clear policy on acquisition (limits on size, formats, sensitivity)
- Be clear about any normalisation processes to preferred formats that may occur upon ingest (e.g., Excel to PDF). Normalisation can greatly affect usability of deposited data
- Provide advice on how to link to related publications in data repository – links from items in IR to related items in data repository and vice versa.
- Provide guidance on how to cite data that has been deposited (generated automatically upon deposit where feasible)

Possible implementation challenges:

- Ensuring that researchers are aware of acquisition policies for data to be deposited and what other options are available externally (are there preferred external repositories that could be recommended for each subject area?).

Data Catalogues in PS HEIs

- **64% of researchers do describe their data**
- **Some HEIs use Dublin Core as their metadata standard**
- **Most HEIs do not use any formal metadata standards**

It is good practice to ensure that the HEI can keep a record of all the data outputs it staff have produced even if the data are hosted elsewhere. Researchers should be encouraged to provide sufficient metadata and explanatory documentation about their research data to ensure that data are discoverable, understandable and re-useable. Most UK funders expect that such metadata will be published by the University, normally within 12 months of the data being generated.

Aspects of good practice:

- Ability to register data sets generated through research activity even when stored in external data repositories
- Ability to link to related publications
- Encouraging researchers to register for ORCIDs that can be assigned to all outputs
- Making use of internal grant IDs when describing research outputs to make them easier to track
- How to make use of data held in other systems (research outputs systems, institutional publications repository)

- Reusing any lay summaries produced for research activity. This may simply be the abstract produced for the grant application itself. Such descriptions can be very valuable in assisting potential reusers to make sense of the data.

Possible implementation challenges:

- Support to generate metadata for selected data may be time consuming and could require some subject expertise.
- There is a balance to be reached – too many mandatory metadata fields are often seen as a barrier to deposit while too little can make the outputs very difficult to assess for reusers.

Guidance, training and support in PS HEIs

- **18% of researchers wanted RDM guidance and support**
- **12% of researchers wanted RDM in the curricula**
- **Most HEIs stated that RDM training, and support were key priorities**

Most UK HEIs provide training, support, advice and guidance regarding research data management. Many tie into existing staff and student training programmes and many offer web-based guidance.

Aspects of good practice:

- Appropriate training should be provided for different stakeholder groups
- Discipline specific training may be required
- Reuse of existing freely available training materials is a good idea rather than starting from scratch

Possible implementation challenges:

- Not always easy to get researchers to participate in RDM related training. HEIs need to ensure that such training is seen as valuable and not just a waste of time.
- RDM champions can be a good draw for training courses (e.g., senior researchers who have employed RDM and been successful)

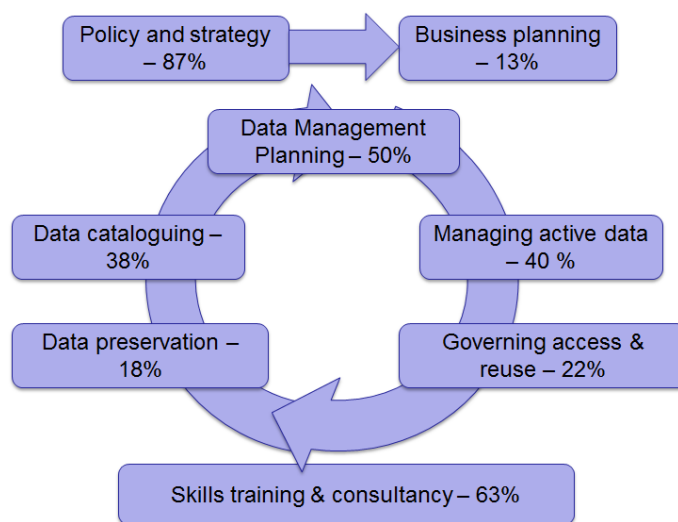
Lack of business and sustainability plans for data repositories and RDM services in PS HEIs:

The costs of supporting RDM can be very difficult to pin down and as such only a small proportion of HEIs in the UK and Europe have developed solid business cases that are sustainable. This will be an issue for the PS HEIs as well. Unfortunately, there are few concrete answers at this stage about what such services will cost to run over time as the scale of born digital research outputs increases year on year. The best bet is to be as clear as you can about what you want to support and to invest in these areas effectively.

Aspects of good practice:

- To ensure that you have a good chance of securing financial and human resources, make sure that the RDM working group is chaired by the PVC Research or equivalent. Working groups without this level of support tend to have greater difficulty in progressing - and more importantly sustaining - their emerging RDM infrastructure and support services.
- RDM needs vary across disciplines so having representation from the relevant faculties will help to highlight areas where there are shared requirements across all domains as well as differences in the level of services required. An understanding of disciplinary requirements will be very helpful as PS HEIs plan for longer term sustainability (i.e., what is provided centrally and what may need to be set up at the Faculty level; how these services will be resourced)
- Be sure to include representation from the Ethics Committee. Improved accessibility and reuse of research data needs to be balanced against ethical concerns such as Data Protection and informed consent. Researchers need to be clear about their data sharing objectives in light of any informed consent agreements that are developed for their projects.

How do UK HEIs compare?



Services that are being implemented across UK HEIs based on results of 2015 DCC survey which collected responses from 60 institutions

As can be seen in the diagram above, many UK HEIs are making good progress in developing and implementing policy and developing training and guidance. Similar to PS HEIs, longer term preservation and business case development are proving more difficult aspects of infrastructure implementation across UK HEIs.