

## Deliverable D1.1 Best Practices 1v0

<b>Project Title (Grant agreement no.):</b>	ELIXIR-CONVERGE: Connect and align ELIXIR Nodes to deliver sustainable FAIR life-science data management services (871075)		
<b>Project Acronym (EC Call):</b>	ELIXIR-CONVERGE (H2020-INFRADEV-2018-2020)		
<b>WP No &amp; Title:</b>	WP1 Expert Network		
<b>WP leader(s):</b>	Bengt Persson (UU)		
<b>Deliverable Lead Beneficiary:</b>	26 - UU		
<b>Contractual delivery date:</b>	31/01/2021	<b>Actual delivery date:</b>	04/02/2021
<b>Delayed:</b>	No		
<b>Partner(s) contributing to this deliverable:</b>	UU, ELIXIR/EMBL-EBI, VIB, SIB, UCY, CING, UOCHB, HITS, DTU, UTARTU, BSC, CSC, CNRS, ATHENA RC, TTK, UCD, WEIZMANN, CNR, UNILU, DTL PROJECTS, UiB, INESC ID, UL, UNIMAN		
<p><b>Authors:</b> Niclas Jareborg (UU/SU), Flora D'Anna (VIB), Rob Hooft (DTL PROJECTS), Korbinian Bösl (UiB), Yvonne Kallberg (UU/SU), Daniel Faria (INESC ID), Victoria Dominguez Del Angel (CNRS), Federico Bianchini (UiB/UiO), Elin Kronander (UU), Markus Englund (UU), Vasilis Promponas (UCY), Thanasis Vergoulis (ATHENA RC), Grégoire Rossier (SIB), Paulette Lieby (CNRS), Priit Adler (UTARTU), Siiri Fuchs (CSC), Minna Ahokas (CSC), Bengt Persson (UU)</p> <p><b>Contributors:</b> The participants of the ELIXIR Data Management Network<sup>1</sup></p> <p><b>Acknowledgments (not grant participants):</b> N/A</p>			
<b>Reviewers:</b>	ELIXIR-CONVERGE Management Board (MB) members.		

### Log of changes

DATE	Mvm	Who	Description
20/01/2021	0v1	Niclas Jareborg (UU/SU)	Initial version
25/01/2021	0v2	Niclas Jareborg (UU/SU)	Sent to PMU after incorporating internal WP feedback
25/01/2021	0v3	Nikki Coutts (ELIXIR Hub)	Circulated to the MB for final review before submission
02/02/2021	0v4	Niclas Jareborg (UU/SU)	MB comments addressed
04/02/2021	1v0	Nikki Coutts	Final version to be uploaded into EC Portal

<sup>1</sup><https://elixir-europe.org/intranet/data-management-network/members>

		(ELIXIR Hub)	
--	--	--------------	--

## Table of contents

Executive Summary	1
2. Contribution toward project objectives	2
3. Introduction	4
4. Methodology	4
5. Results	5
6. Conclusions	8
7. Impact	9
8. Next Steps	9
9. Deviation from Description of Action	9

## 1. Executive Summary

The aim of this deliverable is to *start* capturing, collating, and disseminating the collective data management expertise that resides across the ELIXIR member Nodes. This first version of the guidelines aims to promote the knowledge sharing and capacity building among the member Nodes when it comes to data management, as well as harmonising the guidance across Nodes, while at the same time taking the local circumstances into account. It should also help in identifying and prioritising data management training needs and the development of training material. Furthermore, this resource should highlight how ELIXIR services can solve data management issues for the research community. The work of establishing best practice guidelines to a great extent interacts with the work done in *WP2 Training*, *WP3 Toolkit*, and *WP5 Demonstrators*.

The work has been coordinated by a subgroup of the Data Management Coordinators group. It was decided that the primary focus audience for the guideline content should be the professionals in a *research-supporting Data Steward* role. The emerging “RDMKit” web resource designed and established by WP3 was selected as the medium for disseminating the best practice guidelines content.

The best practice guidelines content has been developed in a crowdsourcing manner by the members of WP1 and the broader Data Management Network<sup>2</sup>, in large parts through the “Hackathons” and “Contentathons” organised by WP3. The work of WP1, WP2, WP3 and WP5 to produce this resource has been deeply integrated with each other, as many of the network members

<sup>2</sup><https://elixir-europe.org/about-us/how-funded/eu-projects/converge/wp1/dm-network>



are involved in several different WPs of the project, and because the subject matter covered in the different WPs overlap.

Hence:

- A first version of best practice guidelines, relevant to the life science domain, has been developed. The content of the guidelines focuses on the different data life cycle stages, as well as key data management issues related to them. The content is targeted to research-supporting Data Stewards.
- The best practice guidelines content is being made available through the RDMKit resource, that has been designed and established by WP3.
- Contributions of content have come from a large number of Nodes and experts, including members not only in WP1, but also from WP2, WP3 and WP5.
- Content has been generated in a crowdsourcing manner; much of it during hackathons organised by WP3. This process has sparked discussions that in themselves are an important part of the knowledge-sharing and capacity building process.

## 2. Contribution toward project objectives

With this deliverable, the project has reached or **the deliverable has contributed** to the following objectives/key results:

Objective no. / Key Result no. Description	Contributed to:
<b>Objective 1:</b> Develop a sustainable and scalable operating model for transnational life-science data management support by leveraging national capabilities ( <b>WP1, WP5</b> )	
<b>Key Result 1.1:</b> Established European expert network of data stewards that connect national data centres and similar infrastructures and drive the development of interoperable solutions following international best practice, including national interpretations of the General Data Protection Regulation (GDPR)	<b>Yes</b>
<b>Key Result 1.2:</b> Development of joint guidelines and common toolkit that are adopted into funder recommendations, with support available nationally and in local languages	<b>Yes</b>
<b>Key Result 1.3:</b> The catalogue of successful national business models incorporated into national strategies	<b>No</b>
<b>Key Result 1.4:</b> The developed “sustainable and scalable operating model for transnational life-science data management support” is adopted into national ELIXIR Node	<b>No</b>
<b>Objective 2:</b> Strengthen Europe’s data management capacity through a comprehensive training programme delivered throughout the European Research Area ( <b>WP2, WP6</b> )	



<p><b>Key Result 2.1:</b> A comprehensive ELIXIR Training and Capacity building programme in Data Management, directed at both data managers and ELIXIR users, and connected to the national training programmes in Data Management in the ELIXIR Nodes and prospective ELIXIR Member countries.</p>	<p><b>Yes</b></p>
<p><b>Key Result 2.2:</b> Development of a collective group of trainers that support scalable deployment of Data Management training across ELIXIR Nodes.</p>	<p><b>No</b></p>
<p><b>Key Result 2.3:</b> A substantial cohort of data managers, Node coordinators and researchers with specific data management skills, business planning and knowledge of transnational operations across the ELIXIR Nodes</p>	<p><b>No</b></p>
<p><b>Objective 3:</b> Align national data management standards and services through a sustainable, scalable and cost-effective data management toolkit (<b>WP2, WP3, WP5</b>)</p>	
<p><b>Key Result 3.1:</b> Assemble a full-stack harmonised common toolkit comprising all aspects of data management: from data capture, annotation, and sharing; to integration with analysis platforms and making the data publicly available according to international standards.</p>	<p><b>Yes</b></p>
<p><b>Key Result 3.2:</b> Provide exemplar toolkit configurations for prioritised demonstrators to serve as templates for future use.</p>	<p><b>No</b></p>
<p><b>Key Result 3.3:</b> Establish national capacity in using as well as updating, extending and sustaining the toolkit across the ERA.</p>	<p><b>No</b></p>
<p><b>Key Result 3.4:</b> Enable 'FAIR at source' practice for data generation, and analytical process pipeline implementation by flexible deployment of the toolkit in national operations</p>	<p><b>No</b></p>
<p><b>Objective 4:</b> Align national investments to drive local impact and global influence of ELIXIR (<b>WP4,WP6</b>)</p>	
<p><b>Key Result 4.1:</b> Development of a Node Impact Assessment Toolkit based on RI-PATHS methodology.</p>	<p><b>No</b></p>
<p><b>Key Result 4.2:</b> Adoption of Impact assessment in ELIXIR Nodes, supported by Node coordinators network and feedback on applicability from dialogues with national funders.</p>	<p><b>No</b></p>
<p><b>Key Result 4.3:</b> Creation of national public-private partnerships and industry outreach where open life-science data and services stimulate local bioeconomy</p>	<p><b>No</b></p>
<p><b>Key Result 4.4:</b> Growth in reach, impact and engagement of stakeholder communication assessed by established ELIXIR Communications metrics</p>	<p><b>No</b></p>
<p><b>Key Result 4.5:</b> Initiating and advancing discussions on Membership (EU and international) or strategic partnerships (international countries) following ELIXIR-CONVERGE workshops.</p>	<p><b>No</b></p>



### 3. Introduction

The aim of this deliverable is to *start* capturing, collating, and disseminating the collective data management expertise that resides across the ELIXIR member Nodes. Since the growth and diversity of molecular datasets and platforms have steadily increased in the last decade, the need for the relevant data to be better managed becomes highly important. Advancing the understanding of life and translating open data resources into societal benefits and economic growth in the long run requires that all components follow the FAIR principles. However, there is a large unmet need for Data Management capacity building, guidance and harmonization in the life-sciences to achieve this. In this context, ELIXIR is uniquely placed to coordinate national-level expertise in terms of life-science data management in Europe. However, it should be pointed out that the level of expertise and capacity is not uniformly spread across the Nodes, nor is the coverage of expertise in the various areas of data management. There is a need for capacity building, whereby all Nodes contribute part of the knowledge, as well as learn from others. The ELIXIR-CONVERGE project has established the ELIXIR Data Management Network that gathers the data management expertise existing in the member Nodes. This deliverable is the first version in the effort of collating the key aspects of the data management practice guidelines expertise that reside within the Data Management Network. The guidelines aim to promote the knowledge sharing and capacity building among the member Nodes when it comes to data management, as well as harmonizing the guidance across Nodes, while at the same time taking the local circumstances into account. It should also help in identifying and prioritizing data management training needs and the development of training material. This resource should also highlight how ELIXIR services can solve data management issues for the research community.

The work to establish best practice guidelines to a great extent interacts with the work performed in WP2 Training, WP3 Toolkit, and WP5 Demonstrators, both on the subject matter covered, and in the actual persons involved. Expert knowledge from WP1, WP2 and WP5 was brought together to establish the *best practice guidelines content*. The WP3 RDMKit web resource offers a *structure* and a *platform* to disseminate the content of these guidelines.

### 4. Methodology

A *Best practice guidelines working group* was formed by a subset of members from the Data Management Coordinators group<sup>3</sup>, consisting of representatives from ELIXIR-BE, ELIXIR-FR, ELIXIR-NL, ELIXIR-NO, ELIXIR-PT, and ELIXIR-SE, to coordinate the work of establishing a first version of the Best practice guidelines. The group has had regular meetings throughout the first year of the project. The initial work was focused on defining the scope, target audience and format of dissemination for this first version of the Best practice guidelines. Given that the finer details of data management guidelines are highly dependent on the local situation, it was decided that the scope should be limited to general principles that would be applicable regardless of the environment in which a potential reader would be working. The research-supporting *Data Steward* role was selected as the primary focus audience for the guideline content. As WP3 rapidly moved forward in the work

---

<sup>3</sup><https://elixir-europe.org/about-us/how-funded/eu-projects/converge/wp1/dm-coordinators>



around a Research Data Management Toolkit web resource, and because good practice recommendations are so tightly connected to the tools used for data management, it was decided to use this emerging “RDMKit” web resource as the medium for disseminating the best practice guidelines content.

Best practice guidelines content has been collected in a crowdsourcing manner by the members of WP1 and the broader Data Management Network<sup>4</sup>, in large parts through the “Hackathons” and “Contentathons” (events for groups of people to meet up and collaboratively create code and content) organised by WP3. The work of WP1, WP2, WP3 and WP5 to produce this resource has been deeply integrated with each other, as many people are involved in several WPs.

We especially want to point out that the work led by WP3 to design and establish the RDMKit web resource has been crucial in providing an effective dissemination platform for the Data Management best practice guidelines.

## 5. Results

In close collaboration with WP2, WP3 and WP5, members of WP1 and the broader Data Management Network have contributed with analysis of the types of practices that need to be captured in the RDMKit resource and other content to be covered, such as ELIXIR services that have a role in data management, as well as contributed to templates to be used when presenting the different aspects of data management.

Content has been provided by many partners in most of the Nodes, and through the establishment of the Data Management Network, even more than those that are participating in the CONVERGE project. Updates on progress have been reported on the regular monthly meetings of the Data Management Network.

The RDMKit offers several orthogonal entryways to the data management content – the *Data Life Cycle*, *Your Problem*, *Your Domain*, *Your Role* and *Tools Assembly*. The RDMKit is described in more detail in D3.1 Assembled starter toolkit<sup>5</sup>. In this first version of the best practice guidelines the focus has been on the content for the *Data Life Cycle* and *Your Problem* aspects, and to some extent, also on parts of the *Your Domain*, which are driven by the demonstrator projects of WP5.

The *Data Life Cycle* pages describe a high-level description of the importance of, and considerations for, good data management practices during the different phases of the data life cycle. These pages then refer to relevant *Your Problem* pages for more in-depth information.

The *Your Problem* pages describe particular data management issues in more detail, and also refer to relevant tools and resources to address the problem. These recommendations highlight existing high-quality resources, such as e.g. those assigned to the ELIXIR Deposition databases list, as well as to established ELIXIR services.

An overview of the content for the *Data Life Cycle* and *Your Problem* pages is summarized in *Table 1*. A screenshot of an example *Data Life Cycle* page is shown in *Figure 1*.

An important aspect of the work done, that might not be apparent from the output generated, but merits mention, is the value of the process of the collective discussions generated by the various viewpoints and backgrounds of the participants. It is this rich and open exchange that eventually led

---

<sup>4</sup><https://elixir-europe.org/about-us/how-funded/eu-projects/converge/wp1/dm-network>

<sup>5</sup>[https://zenodo.org/record/4501634#.YBweuo\\_7T0o](https://zenodo.org/record/4501634#.YBweuo_7T0o)



to the generation of the content presented here. . The discussions themselves are an important part of working towards a shared language and knowledge base of harmonized best practice guidelines for the ELIXIR Data Management Network.

**Table 1.** Overview summary of page contents

Data Life Cycle	
Planning <sup>6</sup>	Describes the aims and benefits of data management planning, and Data Management Plans (DMPs).
Collecting <sup>7</sup>	Describes the importance of practices to ensure data quality and capturing provenance.
Processing <sup>8</sup>	Describes practices for preparing data for analysis, such as harmonizing data formats, file structure, and applying documentation and metadata standards.
Analysing <sup>9</sup>	Focuses on structuring data, and documenting data and analysis.
Preserving <sup>10</sup>	Outlines practices and considerations for preserving the data long term.
Sharing <sup>11</sup>	Describes the reasons and considerations to be addressed to make the data known and available to the research community and society at large, according to the FAIR principles.
Reusing <sup>12</sup>	Describe practices to employ to ensure quality and interoperability when reusing data.
Your problem	
Compliance monitoring & measurement <sup>13</sup>	Focuses on how to document and monitor that legal obligations and policies are being adhered to.
Data analysis <sup>14</sup>	Describes how to ensure reproducibility of analyses performed.
Data classification <sup>15</sup>	Guidance on determining if data is sensitive (with focus on personal data).
Data management plan <sup>16</sup>	Recommendations for DMP content, templates and solutions.
Data organisation <sup>17</sup>	Describes good practices for file naming, file and folder organisation, and version control.
Data protection <sup>18</sup>	Describes the issues to be considered when handling personal data under the GDPR.

<sup>6</sup><https://rdm.elixir-europe.org/planning>

<sup>7</sup><https://rdm.elixir-europe.org/collecting>

<sup>8</sup><https://rdm.elixir-europe.org/processing>

<sup>9</sup><https://rdm.elixir-europe.org/analysing>

<sup>10</sup><https://rdm.elixir-europe.org/preserving>

<sup>11</sup><https://rdm.elixir-europe.org/sharing>

<sup>12</sup><https://rdm.elixir-europe.org/reusing>

<sup>13</sup>[https://rdm.elixir-europe.org/compliance\\_monitoring.html](https://rdm.elixir-europe.org/compliance_monitoring.html)

<sup>14</sup>[https://rdm.elixir-europe.org/data\\_analysis.html](https://rdm.elixir-europe.org/data_analysis.html)

<sup>15</sup>[https://rdm.elixir-europe.org/data\\_classification.html](https://rdm.elixir-europe.org/data_classification.html)

<sup>16</sup>[https://rdm.elixir-europe.org/data\\_management\\_plan.html](https://rdm.elixir-europe.org/data_management_plan.html)

<sup>17</sup>[https://rdm.elixir-europe.org/data\\_organisation.html](https://rdm.elixir-europe.org/data_organisation.html)

<sup>18</sup>[https://rdm.elixir-europe.org/data\\_protection.html](https://rdm.elixir-europe.org/data_protection.html)





Data publication <sup>19</sup>	Guidance on how to select suitable repositories for data publishing.
Data quality <sup>20</sup>	Describes factors affecting data quality, and solutions to help ensure data quality.
Data transfer <sup>21</sup>	Focuses on solutions and considerations to transfer large data files.
Licensing <sup>22</sup>	Guidance on how to select licenses for reuse of data and code, to promote reusability.
Metadata management <sup>23</sup>	Describes solutions for finding suitable metadata standards and ontologies to describe data.
Storage <sup>24</sup>	Describes different options for data storage during different stages of the data life cycle.
<b>Your Role</b>	
Data Steward Policy <sup>25</sup>	Focuses on the role of providing policy development and the implementation of research data management in their institution.
Data Steward Research <sup>26</sup>	Describes the role of Data Stewards that provide data management support and advice to researchers in their institution and ensures the quality of their data management plans.
Data Steward IT <sup>27</sup>	Focuses on the role of ensuring compliance between data infrastructure and RDM policies.
Researcher <sup>28</sup>	Presents the view of a researcher, that with the help of the Data Steward Research, focuses on best practices in RDM and in writing Data Management Plans.

<sup>19</sup>[https://rdm.elixir-europe.org/data\\_publication.html](https://rdm.elixir-europe.org/data_publication.html)

<sup>20</sup>[https://rdm.elixir-europe.org/data\\_quality.html](https://rdm.elixir-europe.org/data_quality.html)

<sup>21</sup>[https://rdm.elixir-europe.org/data\\_transfer.html](https://rdm.elixir-europe.org/data_transfer.html)

<sup>22</sup><https://rdm.elixir-europe.org/licensing.html>

<sup>23</sup>[https://rdm.elixir-europe.org/metadata\\_management.html](https://rdm.elixir-europe.org/metadata_management.html)

<sup>24</sup><https://rdm.elixir-europe.org/storage.html>

<sup>25</sup>[https://rdm.elixir-europe.org/data\\_steward\\_policy.html](https://rdm.elixir-europe.org/data_steward_policy.html)

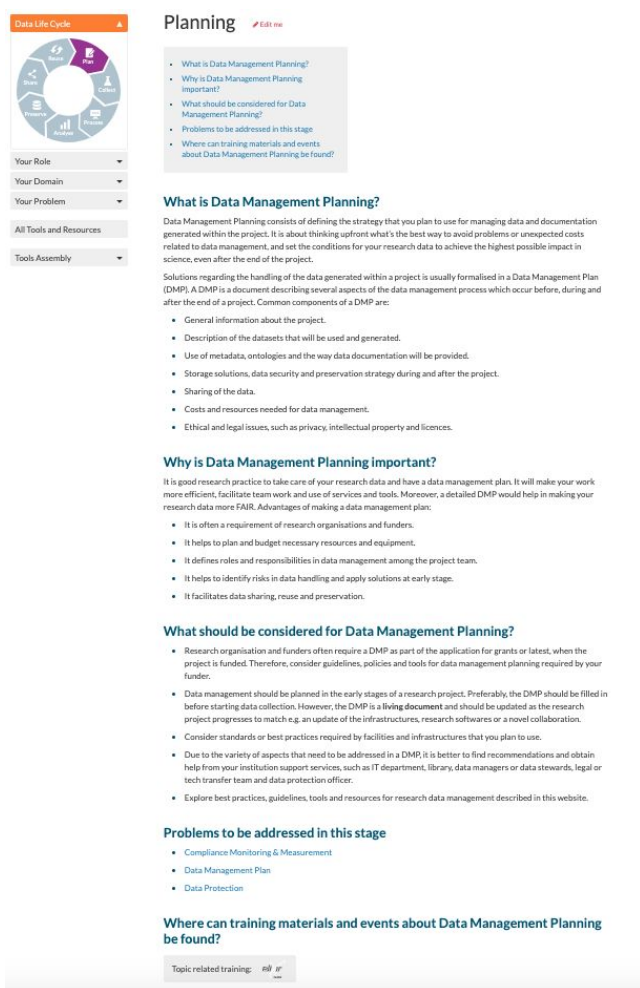
<sup>26</sup>[https://rdm.elixir-europe.org/data\\_steward\\_research.html](https://rdm.elixir-europe.org/data_steward_research.html)

<sup>27</sup>[https://rdm.elixir-europe.org/data\\_steward\\_IT.html](https://rdm.elixir-europe.org/data_steward_IT.html)

<sup>28</sup><https://rdm.elixir-europe.org/researcher.html>







The screenshot shows a web interface for the 'Data Life Cycle' project. On the left, there is a navigation menu with a circular diagram representing the data life cycle stages: Plan, Create, Share, and Reuse. Below the diagram are dropdown menus for 'Your Role', 'Your Domain', 'Your Problem', 'All Tools and Resources', and 'Tools Assembly'. The main content area is titled 'Planning' and includes several sections:

- What is Data Management Planning?**

Data Management Planning consists of defining the strategy that you plan to use for managing data and documentation generated within the project. It is about thinking upfront what's the best way to avoid problems or unexpected costs related to data management, and set the conditions for your research data to achieve the highest possible impact in science, even after the end of the project.

Solutions regarding the handling of the data generated within a project is usually formalised in a Data Management Plan (DMP). A DMP is a document describing several aspects of the data management process which occur before, during and after the end of a project. Common components of a DMP are:

  - General information about the project.
  - Description of the datasets that will be used and generated.
  - Use of metadata, ontologies and the way data documentation will be provided.
  - Storage solutions, data security and preservation strategy during and after the project.
  - Sharing of the data.
  - Costs and resources needed for data management.
  - Ethical and legal issues, such as privacy, intellectual property and licences.
- Why is Data Management Planning important?**

It is good research practice to take care of your research data and have a data management plan. It will make your work more efficient, facilitate team work and use of services and tools. Moreover, a detailed DMP would help in making your research data more FAIR. Advantages of making a data management plan:

  - It is often a requirement of research organisations and funders.
  - It helps to plan and budget necessary resources and equipment.
  - It defines roles and responsibilities in data management among the project team.
  - It helps to identify risks in data handling and apply solutions at early stage.
  - It facilitates data sharing, reuse and preservation.
- What should be considered for Data Management Planning?**
  - Research organisation and funders often require a DMP as part of the application for grants or latest, when the project is funded. Therefore, consider guidelines, policies and tools for data management planning required by your funder.
  - Data management should be planned in the early stages of a research project. Preferably, the DMP should be filled in before starting data collection. However, the DMP is a living document and should be updated as the research project progresses to match e.g. an update of the infrastructures, research softwares or a novel collaboration.
  - Consider standards or best practices required by facilities and infrastructures that you plan to use.
  - Due to the variety of aspects that need to be addressed in a DMP, it is better to find recommendations and obtain help from your institution support services, such as IT department, library, data managers or data stewards, legal or tech transfer team and data protection officer.
  - Explore best practices, guidelines, tools and resources for research data management described in this website.
- Problems to be addressed in this stage**
  - Compliance Monitoring & Measurement
  - Data Management Plan
  - Data Protection
- Where can training materials and events about Data Management Planning be found?**

Topic related training: [#18 #17](#)

Figure 1. An example of a Data Life Cycle page

## 6. Conclusions

- A first version of best practice guidelines relevant to the life science domain has been developed. The content of the guidelines focuses on the different data life cycle stages, as well as key specific data management issues related to those. The content is targeted towards research-supporting Data Stewards.
- The best practice guidelines content has been made available through the RDMKit resource that has been designed and established by WP3.
- Contributions of content have come from a large number of Nodes and experts, including members not only in WP1, but also WP2, WP3 and WP5, as well as from the broader Data Management Network.
- Content has been generated in a crowdsourcing manner; much of it during hackathons organised by WP3. This process has sparked discussions that in themselves are an important part of the knowledge-sharing and capacity building process.



## 7. Impact

At this point in time, it is too early to evaluate the impact of the work done. Nevertheless, in a meeting with the European Commission on 14 January 2021 on Data Management guidelines for the Horizon Europe programme where the RDMKit was presented, the Commission expressed their enthusiasm to be able to refer to the RDMKit in their recommendations. In any case, this work has laid the groundwork for a joint knowledge resource that will allow researchers and data management professionals to be able to learn and apply good data management practices in the life science domain. Researchers and data stewards developing DMPs will be able to build upon the best practices descriptions as background material. ELIXIR service providers will be able to ensure that their services are discoverable by data managers and researchers who can make use of them to enable and facilitate their work. The ELIXIR data managers in the Data Management Network will be able to use the RDMKit to refer to well written descriptions of best practices and tool assemblies. This contributes towards the ELIXIR-CONVERGE Key Results 1.1, 1.2 and 3.1. The work also starts scoping the knowledge body of data management practices that are relevant to the life science domain, and thus provides a base on which to establish a training and capacity building programme. In this, it also contributes to Key Result 2.1.

## 8. Next Steps

The current state is not a finished product. On the contrary, it is a starting point. A “final” version is due at M30 (*D1.4 ELIXIR Best Practices recommendations*). Immediate steps will be to continue expanding, deepening and refining the guideline content. Expansions of content should cater for more target audiences beyond the research-supporting Data Steward role, such as policy makers, IT staff and researchers themselves. Also the data management needs identified by all demonstrators should be catered for.

As the work to be done is so tightly interconnected with work within WP1, WP2, WP3 and WP5, an even more clear management process on how to progress the joint development of the RDMKit resource is desirable. It is assumed that the editorial board (that has WP1 representation) set up for the establishment of the RDMKit will continue to operate through ELIXIR-CONVERGE. One challenge to address in this work is how more experts can be engaged and feel involved in contributing to the content.

## 9. Deviation from Description of Action

Not applicable

