

Deliverable D3.1 Assembled starter toolkit

Project Title (Grant agreement no.):	ELIXIR-CONVERGE: Connect and align ELIXIR Nodes to deliver sustainable FAIR life-science data management services (871075)		
Project Acronym (EC Call):	ELIXIR-CONVERGE (H2020-INFRADEV-2018-2020)		
WP No & Title:	WP3 Common Data Management Toolkit		
WP leader(s):	Frederik Coppens (VIB), Carole Goble (UNIMAN), Inge Jonassen (UiB)		
Deliverable Lead Beneficiary:	28 - UNIMAN		
Contractual delivery date:	31/01/2021	Actual delivery date:	4/2/2021
Delayed:	Yes		
Partner(s) contributing to this deliverable:	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>Authors: Carole Goble (UNIMAN), Pinar Alper (UNILU), Bert Droesbeke (VIB), Flora D'Anna (VIB), Munazah Andrabi (UNIMAN), Mijke Jetten (DTL PROJECTS), Frederik Coppens (VIB), Rob Hoofst (DTL PROJECTS), Martin Cook (ELIXIR-Hub), Celia van Gelder (DTL PROJECTS)</p> <p>Contributors: Daniel Faria (INESC-ID), Niclas Jareborg (UU/SU), Markus Englund (UU), Yvonne Kallberg (UU/SU), Ulrike Wittig (HITS), Korbinian Bösl (UiB), Inge Jonassen (UiB), Chris Child (UNIMAN), Olivier Collin (CNRS), Federico Bianchini (UiB/UiO), Karel Berka (UPOL), Xenia Perez Sitja (ELIXIR-Hub), Grégoire Rossier (SIB), Danny Ben-Avraham (WEIZMANN), Robert Pergl (UOCHB), Laura Portell (BSC), Pier Luigi Martelli (Bologna), Siiri Fuchs (CSC), and all the hackathon and contentathon participants.</p> <p>Acknowledgments (not grant participants): Peter McQuilton (Oxford University), Philippe Rocca-Serra, (Oxford University), Anil Wipat (Newcastle University), Nicola Soranzo (Earlham Institute), Robert Andrews (Cardiff University), Krzysztof Poterlowicz (UBRADFORD), Adam Hospital (IRB), all the hackathon and contentathon participants.</p>			
Reviewers:	ELIXIR-CONVERGE Management Board (MB) members.		

Log of changes

DATE	Mvm	Who	Description
25/01/2021	0v1	Carole Goble (UNIMAN)	Initial version
25/01/2021	0v2	Carole Goble (UNIMAN)	Sent to PMU after incorporating internal WP feedback
25/01/2021	0v3	Nikki Couttts (ELIXIR Hub)	Circulated to the MB for final review before submission

04/02/2021	1v0	Nikki Coutts (ELIXIR Hub)	Final version to be uploaded into EC Portal
------------	-----	------------------------------	---

Table of contents

Executive Summary	3
2. Contribution toward project objectives	4
3. Introduction	5
3.1 Scope of Deliverable	6
3.2 Relationship with other WPs	7
4. Description of work accomplished	7
4.1 RDMKit Design Methodology	7
4.1.1 The RDMKit Team and Management	8
4.1.2 Research Data Lifecycle	8
4.1.3 Tool and Resource Landscaping	9
4.1.4 RDMKit Users Stories and Persona	10
4.1.5 RDMKit Storyboarding and Prototyping	11
4.2 RDMKit Implementation	12
4.2.1 Content Structure and Navigation	12
4.2.2 Web Development Best Practices	13
4.2.3 Design Challenges	14
4.3 Content processes	14
4.3.1 Editorial Processes	14
4.3.2 Hackathons and Contentathons	16
4.3.3 Community Development	17
4.4 Workshops, Dissemination and External Engagement	17
5. Results	17
5.1 Landing Page	18
5.2 RDM Lifecycle	18
5.3 Your Role	19
5.4 Your Problem	22
5.6 Your Domain	23
5.7 Tool Assemblies	24
6. Conclusions	25
7. Impact	26
8. Next Steps	26



8.1 RDMKit website release, promotion, extension and continual development	26
8.2 Tool Assemblies from WP5 and pattern development	27
8.3 Data Stewardship Wizard	28
8.4 Training and capacity building	28
9. Deviation from Description of Action	28
10. Appendix	28
10.1 Editorial Board	28

1. Executive Summary

An ELIXIR Common Data Management Toolkit is intended to support researchers and data stewards to enable lifecycle management for their research data, according to international standards, and integrated in the national context. The goal is to support FAIR by Design: from the first steps of producing FAIR data and Data Management Planning to the final steps of depositing data in public archives through data brokerage pathways from data management systems (Task 1.2) and enabling its reuse.

The vision of the Toolkit has matured into three, interlinked, streams:

1. **The RDMKit**, a website-based toolkit designed to guide life scientists and data stewards in their efforts to better manage their research data.
2. **Tool Assemblies**, which are examples of combining tools to cover stages of the RDM Lifecycle and assembly patterns to report how tools can be combined and replaced by local specialisations.
3. **Data Stewardship Wizard**, which leads data stewards through decision trees to choose the right tools, resources and practices when planning their data management.

WP3 has **concentrated on the RDMKit**. This provides rapid and maximum benefit to the Node's Life Science communities and effectively coordinates and integrates the work of the CONVERGE WPs around a common goal. It also provides a timely resource for funders such as the EC's Open Science Unit during development of the HorizonEurope programme. The RDMKit is targeted at researchers, data managers and data stewards in laboratories, facilities or universities; project consortium coordinators and funding agencies and policy makers.

The RDMKit's underlying technical infrastructure has been established using GitHub (<https://github.com/elixir-europe/rdmkit>) and the contribution and editorial processes put in place. Content has been assembled through hackathons/contentathons.

This deliverable reports on the conception, design, implementation of the RDMKit, its open contribution, review and editorial processes, gives screenshot examples and lays out the next steps.



The RDMKit is available at <https://rdm.elixir-europe.org/> and <https://rdmkit.elixir-europe.org/> (aliases) and will be internally beta released for review and awareness raising on 31 January 2021 and publicly released on 31 March 2021.

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:
Objective 1: Develop a sustainable and scalable operating model for transnational life-science data management support by leveraging national capabilities (WP1, WP5)	
Key Result 1.1: 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)	Yes
Key Result 1.2: Development of joint guidelines and common toolkit that are adopted into funder recommendations, with support available nationally and in local languages	Yes
Key Result 1.3: The catalogue of successful national business models incorporated into national strategies	No
Key Result 1.4: The developed “sustainable and scalable operating model for transnational life-science data management support” is adopted into national ELIXIR Node	Yes
Objective 2: Strengthen Europe’s data management capacity through a comprehensive training programme delivered throughout the European Research Area (WP2, WP6)	
Key Result 2.1: 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.	No
Key Result 2.2: Development of a collective group of trainers that support scalable deployment of Data Management training across ELIXIR Nodes.	No
Key Result 2.3: 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	No



Objective 3: Align national data management standards and services through a sustainable, scalable and cost-effective data management toolkit (WP2, WP3, WP5)	
Key Result 3.1: 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.	Yes
Key Result 3.2: Provide exemplar toolkit configurations for prioritised demonstrators to serve as templates for future use.	Yes
Key Result 3.3: Establish national capacity in using as well as updating, extending and sustaining the toolkit across the ERA.	Yes
Key Result 3.4: Enable 'FAIR at source' practice for data generation, and analytical process pipeline implementation by flexible deployment of the toolkit in national operations	Yes
Objective 4: Align national investments to drive local impact and global influence of ELIXIR (WP4,WP6)	
Key Result 4.1: Development of a Node Impact Assessment Toolkit based on RI-PATHS methodology.	No
Key Result 4.2: Adoption of Impact assessment in ELIXIR Nodes, supported by Node coordinators network and feedback on applicability from dialogues with national funders.	No
Key Result 4.3: Creation of national public-private partnerships and industry outreach where open life-science data and services stimulate local bioeconomy	No
Key Result 4.4: Growth in reach, impact and engagement of stakeholder communication assessed by established ELIXIR Communications metrics	No
Key Result 4.5: Initiating and advancing discussions on Membership (EU and international) or strategic partnerships (international countries) following ELIXIR-CONVERGE workshops.	No

3. Introduction

At the heart of FAIR Science lies good data management practice. This is increasingly important as life science research becomes data-intensive and traditional 'wet labs' make space for 'dry (computer) labs'. The size and complexity of datasets generated in life science research is unprecedented. More than ever, scientists need relevant tools and guidance to better manage their data and hence contribute to making FAIR Science a reality as well as help researchers be more productive for themselves and their collaborators.

An ELIXIR Common Data Management Toolkit is intended to support researchers and data stewards



to enable lifecycle management for their research data, according to international standards, and integrated in the national context. The goal is to support FAIR by Design: from the first steps of producing FAIR data and Data Management Planning to the final steps of depositing data in public archives through data brokerage pathways from data management systems (Task 1.2) and enabling its reuse.

The toolkit will cover data governance, provisioning, curation, cataloging, sharing and access as well as integration with local, national and EOSC storage and analysis infrastructures. Content and components would be sourced from the WP1 Data Management Coordinators group (Deliverable D1.1) and the Node Data Management experts group, WP2 training. We also source from technical activities ongoing in Nodes, and build on established practice and emerging consensus in Communities (WP5).

The vision of the Toolkit has matured into three, interlinked, streams:

1. **The RDMKit**, a website-based toolkit designed to guide life scientists and data stewards in their efforts to better manage their research data. It is based on the various steps of the RDM Lifecycle and can be navigated via 'role', 'domain', 'problems' and 'tool assemblies'. The contents are generated and maintained by the WPs and the wider ELIXIR community. The toolkit was initially inspired by the UK's Jisc RDM Toolkit¹ (the RDMKit is more specific and tailored to the Life Science community) and by the Pistoia FAIR Toolkit².
2. **Tool Assemblies**, which are linked to from the RDMKit. Tool Assemblies are examples of combining tools to cover the key functional areas (called "problems" in the RDMKit) to provide data management across the stage of the RDM Lifecycle. These are typically tools that the ELIXIR Nodes combined to support RDM that can be picked up and used by others. An example is the Norwegian eInfrastructure for Life Sciences (NeLS). In future work assembly patterns (blueprints) will be distilled from real examples from the Nodes to show how tools generally can be combined.
3. **Data Stewardship Wizard**, helping users of the RDMKit with the considerations that are mentioned in the web pages through decision trees. The ELIXIR Data Stewardship Wizard³ is built around a hierarchical knowledge model that helps researchers and data stewards choose the right tools, resources and practices when planning their data management. Our vision is that the Data Stewardship Wizard bi-directionally connects at key places in its knowledge model to the pages of the RDMKit, such that users of the Wizard are referred to the RDMKit for background information and overview, and users of the RDMKit are referred to the Wizard to help make context-dependent decisions between the alternative options.

3.1 Scope of Deliverable

The first phase of the CONVERGE project has **concentrated on the RDMKit**. This provides rapid and maximum benefit to the Node's Life Science communities and effectively coordinates and integrates the work of the CONVERGE WPs around a common goal. It also provides a timely resource for funders such as the EC's Open Science Unit during development of the HorizonEurope programme,

¹<https://rdmtoolkit.jisc.ac.uk/>

²<https://fairtoolkit.pistoiaalliance.org/>

³<https://ds-wizard.org/>

and the UK's UKRI Data Science Training in Health and Bioscience programme:

- **For researchers** - RDMKit is a one stop shop of information, advice and signposting to research data management know-how, tools, examples and best practice for those developing and executing project data management plans.
- **For data managers and data stewards in laboratories, facilities or universities** - RDMKit is a resource for data stewards, a complement to their own resources and a guide to the specific challenges of and solutions for RDM in life sciences.
- **For project consortium coordinators** - RDMKit is a resource for research data management planning, and a signpost to specific examples of data management platforms, practice and potential partners.
- **For funding agencies and policy makers** - RDMKit is a resource that can be used to assess and evaluate Data Management Plans, develop open science policies and highlight open science requirements. By using RDMKit, and better managing data, the outcomes of investments will contribute to strengthening regional, national and European economies.

This deliverable reports on the conception, design, implementation of the RDMKit, and its open contribution, review and editorial processes.

The RDMKit will be beta released internally to ELIXIR for review on 31 Jan 2021 and public beta released 31 March 2021.

3.2 Relationship with other WPs

The RDMKit has created a powerful and unifying focus for WPs 1-5 and for the CONVERGE project as a whole. All these WPs have contributed to the design, content and oversight of the RDMKit.

- WP1 Data Management experts contribute best practice and guidelines;
- WP2 Training experts lead work on defining persona and took charge of the role sections and have connected the RDMKit to TeSS, the ELIXIR Training Portal.
- WP4 Assisted in outreach, the RDMKit communication planning and connections with EC policy makers
- WP5 Use Cases are the focus of the RDMKit domain pages.

4. Description of work accomplished

4.1 RDMKit Design Methodology

The requirements gathering and design approach follows standard best practices from software development. The approach is agile, progressing through storyboarding, prototyping, the development of a Minimal Viable Product, and then the incremental improvements against a list of requirements and features. Five overarching principles have been established to guide the development:

1. **User interviews, stories and observation** to gather and review user requirements and feedback rather than surveys. From stories we can extract, synthesize and distil requirements, challenges, etc. From observations we can see obstacles, missing information



and common navigation routes. Surveys place the burden on the users rather than the surveyors and rarely produce useful results.

2. **Keep the RDMKit design and implementation simple.** The temptation is to over-engineer in an attempt to produce the perfect generic universal solution. For widespread and sustainable contribution and maintenance, and for greatest usability, the design must be as simple as possible and not overreach.
3. **Keep the content tightly scoped at the right level.** The RDMKit aims to provide a roadmap to research data management know-how, practices and resources, to give examples of tool assemblies and domain specific practices and to signpost to other resources. The provided information needs to be focussed and concise as too much detail swamps the users.
4. **Leverage other resources.** The RDMKit is an important consolidating overview to other tools and resources and a link to their deep and rich content and functionality. The toolkit will work together with other registries, resources and tools inside and outside ELIXIR to go deeper. The design will use registries for tools (Bio.tools), standards and data repositories (FAIRSharing), training (TeSS), computational workflows (WorkflowHub); will link to EOSC resources (e.g. openAIRE, EOSC Portal) and national resources and forthcoming ELIXIR resources (e.g. FAIR Cookbook) and will complement the ELIXIR Data Stewardship Wizard (see Milestone M3.5⁴).
5. **Plan for continuous and incremental development.** The internal beta release will have missing and erroneous content and organisational issues. New insights and technologies will emerge in life sciences research. The design, implementation and management of the RDMKit must embed sustainable continuous development by a wide team of contributors.

4.1.1 The RDMKit Team and Management

WP3 leads the RDMKit activity. The activity incorporates and bridges all Tasks and has a tight collaboration with WP1, WP2 and WP5. The RDMKit team operates as a goal-driven, cross WP and cross task single collective. Representatives of WPs attend WP3 meetings and WP3 representatives attend theirs. WP3 runs a dedicated toolkit virtual meeting that meets bi-weekly, with open minutes⁵. Since July 2020 this is complemented by a smaller RDMKit editorial team (see Appendix 10.1). Documents are managed through a shared Google drive and communications are encouraged through an active Slack channel. All members of CONVERGE are welcome.

The RDMKit project management is supported through our GitHub⁶, which is also the mechanism for the RDMKit implementation (see section 4.2) and accompanying contribution processes (see section 4.3). A link to the GitHub is on the toolkit website. Development of the RDMKit is continuous with regular focusing hackathons, some specialised as “contentathons” concentrating on content acquisition: see subsection 4.3.2 and milestones M3.2.1 and M3.2.2.

Once the definition of the ELIXIR Common Data Management Toolkit had been separated into the three components, the conceptualisation and design work was organised into four main sub-tasks: defining the Research Data Lifecycle; Tool and research landscaping; RDMKit requirements gathering through Users Stories and Personas; and RDMKit storyboarding and prototyping.

⁴https://docs.google.com/presentation/d/1s-9s9jmYeFNBMpcQcPN4H2fMIsUDWPYoaSxO3TiHJ7k/edit#slide=id.g75cc253543_3_33

⁵https://docs.google.com/document/d/12bsNo7s38j_08OCaoN433IHh1oE75w26_cNYcpiwq4/edit

⁶<https://github.com/elixir-europe/rdmkit>



4.1.2 Research Data Lifecycle

The Research Data Lifecycle forms the core organisation for the RDMKit (Figure 1). Such a cycle is recognised as a common and powerful way of organising Research Data Management practice. In reality the world is not so neat and no-one actually operates a Data Lifecycle in such a systematic way and in step order. Nonetheless it is a familiar and recognisable framework for the RDMKit itself and the work of WP3. The Lifecycle was first developed for the Tool landscaping subtask drawing upon: the Jisc RDM lifecycle⁷, DCC curation lifecycle⁸, the DAMA-DMBOK2 Data Management Framework wheel⁹, the Jisc reference architecture for RDM¹⁰ and so on, and subsequently revised a number of times throughout the design process.



Figure 1: The finalised RDMKit Research Data Lifecycle

4.1.3 Tool and Resource Landscaping

The Tool and Resource landscaping sub-task was designed to serve both aspects of the WP3 toolkit: the web RDMKit and the Tool Assemblies and assembly patterns. We surveyed (breaking rule 1 of our principles) the Nodes to identify tools and resources for Research Data Management. These were categorised against:

- **RDM Steps:** Plan, Collect, Process, Analyse, Preserve, Publish/Share. Reuse was added later.
- **Functional areas:** Storage, Security, Privacy, Transfer, Metadata Management, Sustainability, compliance and other. These have since been revised and extended, labelled as “Your Problems” in the RDMKit.
- **Scope and maturity:** Status (production-level, under development etc) and Reach (widespread use, sub-domains, within a node etc).

19 Nodes participated in the survey and over 160 distinct tools and resources were identified, many of which were not registered in the bio.tools registry. The results are reported in M3.1 Assessment of tools and services landscape for WP3 integration in the toolkit, linked to the first landscape list¹¹ and

⁷<https://rdmtoolkit.iisc.ac.uk/research-data-lifecycle/>

⁸<https://www.dcc.ac.uk/guidance/curation-lifecycle-model>

⁹<https://tdan.com/the-data-centric-revolution-semantics-and-the-dama-wheel/25576>

¹⁰<https://researchdata.iiscinvolvement.org/wp/2015/07/30/makes-ideal-research-data-management-system/>

¹¹<https://docs.google.com/spreadsheets/d/1VXhSE2le0J5TAUZY9Se3mb2JACUg0vavGoJML0BlakY/edit#gid=>

the initial review¹² presented at All Hands 2020 workshop on Data Management.

Out of the 160 tools identified in our survey 96 are listed in the RDMKit¹³ website. The editorial team ensures that tools identified in the survey are introduced to the website with accompanying best practice guidance, this way each tool is placed in context: be it a particular life science domain, a particular problem, or a particular stage of data management. Tags are used to automatically organise their appearance in pages. A link to their entry in the ELIXIR registries bio.tools, TeSS or FAIRsharing is given where available (Figure 2).

All tools and resources [Edit me](#)

This is the main tool and resource list of our website. This is a curated list which means that not all tools or resources that exist for a certain topic are listed here. This is mainly because we do not intend to be a registry. In most cases you will only find back the tools or resources that are mentioned in the different pages. Most pages will show a filtered list of this main table at the end of the page. If you see a missing link with one of the registries or a mistake, please open an [issue](#) or check our [how to add a tool or resource guide](#).

Show entries Search:






Tool or resource	Description	Tags	Registry
BBMRI-ERIC's ELSI Knowledge Base	The ELSI Knowledge Base is an open-access resource platform that aims at providing practical know-how for responsible research.	data protection data classification policy officer data manager human data	
Beacon	The Beacon protocol defines an open standard for genomics data discovery.	researcher data manager IT support human data	
Bioconda	Bioconda is a bioinformatics channel for the Conda package manager	IT support data analysis	
Bitbucket	Git based code hosting and collaboration tool, built for teams.	data organisation data manager IT support	
BMRB	Biological Magnetic Resonance Data Bank	IDP researcher	
BrAPI	Specification for a standard API for plant data: plant material, plant phenotyping data	IT support plants	
BRENDA	Database of enzyme and enzyme-ligand information, across all taxonomic groups, manually extracted from primary literature and extended by text mining procedures		 
Bulk Rename Utility	File renaming software for Windows	data organisation data manager researcher	
CATH	A hierarchical domain classification of protein structures in the Protein Data Bank.		 

Figure 2: RDMKit Full Tools and Resources listing, with links to registry entries (e.g. to course materials in TeSS) where available and tagged with persona, domain, lifecycle step, functional area to appear automatically in other pages.

¹²<https://docs.google.com/presentation/d/1SqeMEtP5Srab-mnfs0VWHDzrxwNsom72ucf6PRAmJXE/edit?usp=sharing>

¹³https://rdm.elixir-europe.org/all_tools_and_resources.html



4.1.4 RDMKit Users Stories and Persona

The User Requirements process for the RDMKit is based on the Personas who take part in the Research Data Lifecycle and their User Stories. The aim is to: elicit functional and non-functional user requirements; extract feature sets and prioritize them; and use the requirements to check the usefulness and usability of the toolkit.

User Stories are high level requirements that are translated into the technical requirements for the toolkit. Persona are the characters who appear in the RDMKit user stories. They represent the researchers, data steward experts, trainers, developers and so on.

The **User stories** were structured using a commonly used template along the lines of who/what/why.

- As a `<who?persona/role>` I can/want to `<what?capability>` so that `<why?receive benefit>`

For example

- As a `<researcher>` I want to `<find marine reference datasets>` so that `<I can compare my data against it>`
- As a `<data steward>` I want to `<know where I can find community standards and ontologies>` so that `<I have guidelines when curating metadata>`

Users Stories were sourced from

- CONVERGE WP5: Marine Metagenomics¹⁴, Human Data¹⁵ and Plant Sciences¹⁶
- Externally: GO-FAIR Discovery Implementation Network¹⁷

A **Persona** is a representation of the goals and behavior of a hypothesized group of users, synthesized from interviews and from the experience of the partners. They are captured in short descriptions that include behavioral patterns, goals, skills, attitudes, and sometimes with fictional personal details to make the persona a realistic character. These persona vary in their skillset, agendas and their place in the Data Lifecycle. Persona can take on many roles. For example, a researcher can have the role of data creator, data user, data analyst or data curator. A researcher can also have the role of supervisor.

Persona were developed from a number of sources

- the ELIXIR Communications team¹⁸ for the purpose of designing the ELIXIR website
- for RDMKit¹⁹, led by WP2.

The RDMKit beta release currently focuses on three persona:

1. Researchers,
2. Data Stewards with three sub-types: policy focused, research focused, and IT focused,
3. Funders.

All Persona are used to steer the design of the toolkit and the content. The first two explicitly appear in the RDMKit as “Your Role”. Between the internal and external beta releases the toolkit will be

¹⁴https://docs.google.com/spreadsheets/d/1rjv2BZ_b1aqpS_tr774BiEt6aL0rnn9/edit#gid=1925866563

¹⁵<https://docs.google.com/spreadsheets/d/1nWBVpieoLCdxCiGILwgfVBiZWE9nL0pzShQgZz52d5E/edit?usp=sharing>

¹⁶<https://drive.google.com/file/d/1vQHIEd-KKAsFxN5y1rGgAJAIBxDHfUu8/view?usp=sharing>

¹⁷<https://docs.google.com/spreadsheets/d/1gK8sNP-zL9wVOWMrsX8dv2Tv5oia4sKk4pjLKBM9KMA/edit#gid=0>

¹⁸<https://drive.google.com/drive/u/0/folders/OBziWaXXbS5X2azRISIFyRVpZd1U>

¹⁹https://drive.google.com/drive/folders/1cDn-eHZiTPVZUgtm63rIK_piRgUToFbo



revised for the Project Consortium Coordinators persona in preparation for Horizon Europe guidance. The RDMKit will be reviewed at regular intervals against existing and new Persona.

4.1.5 RDMKit Storyboarding and Prototyping

Storyboarding for software design, and in particular interactive web based systems, is a tool borrowed from the visual storytelling media of film making and comics. Using storyboards and mock-ups we are able to experiment with layout, flow and user interaction without the cost of implementation. Once things become implemented there is a reluctance to make changes. Design-feedback loops are rapid and low cost.

Existing RDM Toolkits were analysed²⁰ for features, and wireframe storyboards²¹ were sketched for the initial RDMKit designs. Storyboarding is currently being used to develop the interactions between the RDMKit and ELIXIR's Data Stewardship Wizard (see Milestone M3.5).

The RDMKit has been through several iterations of prototyping and reviewing. However carefully plans and designs are made, it is only when the site is implemented that issues are identified with the structure, usability, maintainability and ease of content contribution.

4.2 RDMKit Implementation

Our selection of the underlying technical infrastructure was driven by requirements to be sustainable, and to quickly build a community and attract outside collaborators:

- to accept and coordinate contributions under an open authoring and review process;
- to be able to support and track versioning and contributor credit;
- to be familiar to as wide a contributor community as possible;
- to be sustainable beyond the lifetime of the project, and to be lightweight enough to be sustainable with limited resources;
- to be freely accessible and open source;
- to align with similar sites being developed in association with ELIXIR as far as possible.

We surveyed similar sites and were inspired by the Turing Way²², a handbook for reproducible data science developed by a grassroots community, which has also inspired the FAIRplus FAIR Cookbook²³. Both these sites use GitHub in combination with GitHub-pages. The pages are written in Markdown and translated to HTML using Jekyll. This allows contributors to focus on content and site developers to impose a consistent presentation of the content across the site. GitHub features such as version control, issue tracking and discussion threads support collaborative and trackable contribution.

GitHub-pages have many benefits over Content Management Systems, but do have limitations. Notably, pages are static - which means we have to be smart about how we automate navigation paths, organise pages, and incorporate lists of tools and resources. We rely on YAML headers, tagging and adherence to our contribution practices²⁴. We provide empty page templates in Markdown to ensure contributions follow a common structure.

²⁰<https://docs.google.com/spreadsheets/d/1iXtZG8ZgL3b9Bn9vP7tZF-8v8J9mOIBvCvavBoIWkg8/edit#gid=1463104415>

²¹<https://docs.google.com/spreadsheets/d/1iXtZG8ZgL3b9Bn9vP7tZF-8v8J9mOIBvCvavBoIWkg8/edit#gid=1463104415>

²²<https://www.turing.ac.uk/research/research-projects/turing-way-handbook-reproducible-data-science>

²³<https://fairplus.github.io/cookbook-dev/intro.html>

²⁴https://rdm.elixir-europe.org/how_to_contribute.html



4.2.1 Content Structure and Navigation

The RDMKit is implemented through two main components:

- A table of RDM tools and resources that are marked with tags.
- A set of RDM best practice guidelines organised as web pages that are reachable via a navigation structure of topics. Each guideline page contains a filtered view of the tools table depending on the tags appropriate for the page topic.

The navigation structure to reach the pages was based on the landscaping survey (subsection 4.1.3) and is comprised of the following dimensions (see section 5):

- **Data Lifecycle: RDM Lifecycle stages.** The RDM lifecycle introduces the reader to the phases passed through by data before, during and after the end of a research project. Each phase then points the reader to the most common data management problems encountered during that phase by listing related “your problem” pages.
- **Your Problem: Problems encountered** when dealing with research data. These pages describe in detail the most frequent data management functional areas and challenges and give concrete solutions by explaining the use of existing tools and resources. The problems described concern different Persona, such as life sciences researcher, data manager, IT specialist and policy officer.
- **Role: Persona** associated with the researcher role, as well as the roles and responsibilities of data stewards. These pages explicitly identify the researchers as a tool user, and identify the different hats that data stewards wear and roles they play when providing RDM support in their respective institutions. These pages identify the tasks of each role and how the toolkit can be used by stewards in different roles.
- **Your Domain: RDM Challenges specific to particular life science domains.** Challenges specific to domains such as data types, species or areas are highlighted in these pages e.g. Human-Subject research, Marine Metagenomics. These pages list core considerations and solutions for each domain.
- **Tool Assembly: Examples of tool combinations** to target RDM problems, domains and national requirements. Tool assemblies and their underlying common patterns will be the focus of work after the public release of the RDMKit. Currently we have a pilot example tool combination that has been deployed as a national solution in Norway.

Tags are a controlled list of terms and are used as a simple way to automatically support the navigation, cross-linking and population of pages given that the underlying GitHub platform is inherently static. Groups of tags have been defined for each aspect of site navigation. The overview of this tagging system can be found [here](#)²⁵.

By using this tagging system we can automatically build lists and navigation paths. For example:

- List “Your Problem” pages under “Problems to be addressed at this stage” in each RDM lifecycle stage page.
- List “Your Problem” pages under “Related topics” in each “Your Role” page.
- List appropriate tools under each “Your Problem”, “Your Domain” and “Your Role” page.

Keywords (mostly synonyms of page titles) have been added to the pages when needed, to increase the findability of the pages through the website search box.

4.2.2 Web Development Best Practices

The RDMKit website follows best practices in design and development, including:

²⁵https://docs.google.com/spreadsheets/d/16PkQsXPfWbORwtHPu5jifM_VHyyBUVIQBiGkECiWpT6A/edit#gid=277251697



- **Accessibility:** We aim to make the site accessible to all. We follow the EU Web Accessibility Directive²⁶ (2016) and aim to meet the WCAG 2.1²⁷ AA level of accessibility.
- **Privacy/data protection:** any personal data is processed in compliance with GDPR²⁸.
- **Security:** we have kept the site technically simple, with static HTML pages, to make the site as secure and robust as possible. This ensures it remains available to users.
- **Search engine optimisation (SEO):** we want people to find and use this website so we make sure people can find the site in search engines. For example, we use terms the target audience would actually use in a web search.
- **Consistent branding:** to make the site look professional and credible we keep the colour palette, fonts, page layouts, illustrations and writing style consistent. The Editorial Board has defined a style guide²⁹ for contributors and page templates to make sure the text is consistent.

4.2.3 Design Challenges

During our piloting and contentathons we have addressed a number of design challenges, and will continue to address them and new ones we expect to emerge as the site matures. Developing recommendations and guidelines and operating oversight is the responsibility of the Editorial Board, in collaboration with the WP3 team. Challenges fall into several areas:

- **Navigation complexity.** The use of tags for automating navigation and lists is that they must be used and maintained with discipline. There is a balance between the richness of navigation and over complicating the tagging system.
- **Content scope.** The scope of the “Your Domain” content varies and is contextually sensitive. Some domains are wide and complex (Human Data) and other narrow and niche (IDP). The edges of the Domains are fuzzy and overlap. These pages are likely to be the most heterogeneous in the RDMKit. The scope of Your Problem and Lifecycle stage pages have to be focused on the persona and purpose of the site and avoid a tendency to too much technical detail. The RDMKit is not intended to replace the Data Stewardship Wizard or e.g. provide detailed information on AAI: it is intended to complement tools and point people to further resources.
- **Regional and national specialisations.** A little localised content on a page is acceptable but too much and the pages become too long with too much irrelevant material. Examples include pointers to nationally-limited resources such as data stores, sensitive data policies or DMP planning resources. Solutions to this are still being developed.
- **List overload.** A side effect of comprehensive tagging is list overload. If all tools and resources are labelled with all tags then the discriminatory nature of the tags becomes compromised and every page has a long list. We foresee this as a challenge during the beta-testing of the site, as providers will argue for their resources.

4.3 Content processes

4.3.1 Editorial Processes

Contribution to the RDMKit website needs to be as open as possible in order to gain and maintain

²⁶<https://eur-lex.europa.eu/eli/dir/2016/2102/oj>

²⁷<https://www.w3.org/TR/WCAG21/>

²⁸https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

²⁹https://martin-nc.github.io/martinc-rdm-toolkit/style_guide.html



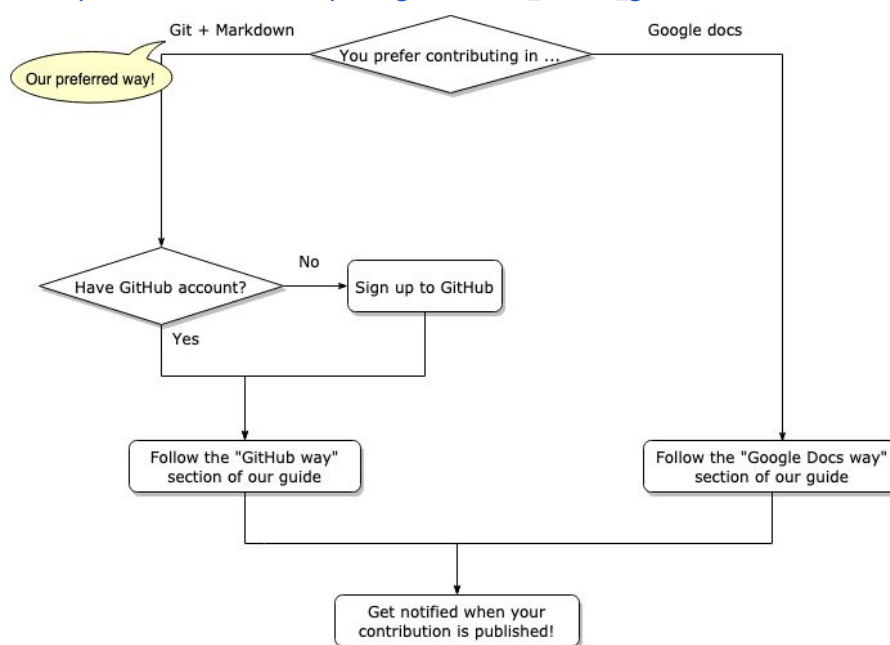
content, and to support the activities of the DM experts (see Deliverable D1.1) and trainers in WP2. On the other hand quality control and coherence needs to be maintained. To accept and coordinate contributions from a wide group of RDM experts we have developed an open authoring and review process.

An Editorial Board including representatives of WP1, WP2, WP3 and WP5 was established in July 2020 to design the contribution and editorial processes, oversee and review content, handle contribution requests, ensure due credit for contributors and manage the governance of the site including issues of plagiarism and copyright infringements. Current board membership is given in the Appendix (Section 10). The processes were reviewed, revised and endorsed at the bi-weekly WP3 meetings and the contentathons. They will be continually refined.

Figure 3 illustrates the contribution workflow. There are three ways to contribute content³⁰:

- The preferred way is to join the co-development on GitHub³¹ with step-by-step instructions³² for the less confident and advanced instructions³³ for those familiar with git.
- For contributors that are not comfortable with git a second path³⁴ uses GoogleDocs and templates. This approach requires e-mail communication with the editors, who are responsible for transferring content to the git repo.
- A third route enables comments and suggestions using a form on the website.

The editorial board meets bi-weekly. Minutes of the board are [here](#)³⁵. New members can apply to join by contacting the board at rdm-editors@elixir-europe.org. Our Editorial Board Guide is available on the RDMKit at https://rdm.elixir-europe.org/editorial_board_guide.html.



³⁰https://rdm.elixir-europe.org/how_to_contribute.html

³¹<https://github.com/elixir-europe/rdm-toolkit>

³²https://rdm.elixir-europe.org/github_way

³³https://rdm.elixir-europe.org/working_with_git

³⁴https://rdm.elixir-europe.org/google_doc_way

³⁵https://docs.google.com/document/d/1eGRE26W2dgBpNGXsfZYvrOzCC_e1_o7oJdTJK7gW5uM/edit#heading=h.2rz9r53i6150

Figure 3. The RDMKit’s alternative contribution ways

For both ways of contributing the toolkit editors ensure that contributions are announced on the toolkit website and a review process is followed. This process involves one or more contributors, at least one reviewer and the editor coordinating between the two. The process, when contributing via Google docs, is illustrated in Figure 4. The detailed instructions on respective processes are on the toolkit website.

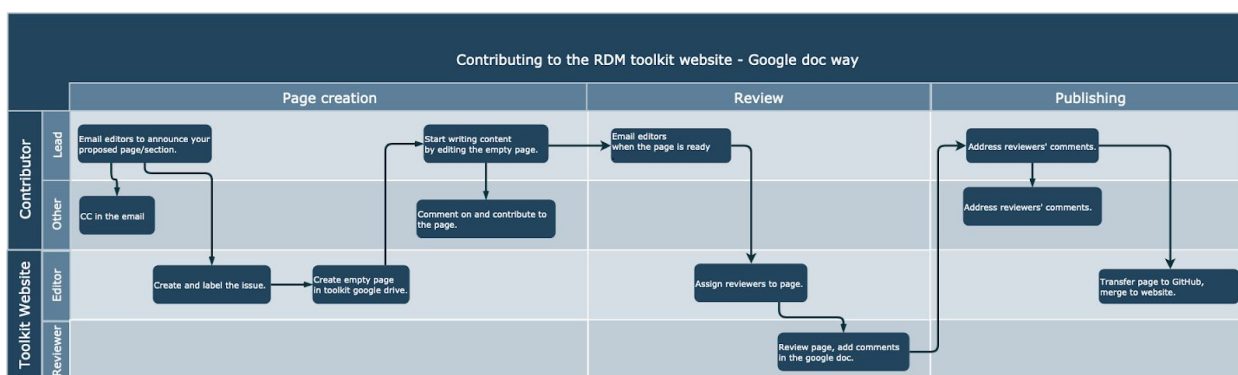


Figure 4. Contribution pathway using Google docs

4.3.2 Hackathons and Contentathons

WP3 organised the following events to build the starter toolkit :

- Hackathon July 2020³⁶: Bootstrapping of the toolkit website, sketching of toolkit design ideas, drafting of working procedures, identification of content areas.
- Contentathon October 2020³⁷: 1st CONVERGE-wide training/orientation to the toolkit, initial content writing in the areas of RDM lifecycle stages, problems, domains and roles.
- Contentathon January 2021³⁸: 2nd CONVERGE + ELIXIR focus groups focused content contribution; further content writing, initial toolkit gap analysis.

Table 1: Hackathon and Contentathon statistics

Event	Attendance	WP Coverage	Node Coverage
Hackathon July 2020	Day 1: 29 Day 2: 29 Day 3: 28 Day 4: 23	WP1, WP2, WP3, WP5	BE, CH, CY, DE, EE, ELIXIR Hub, FI, FR, IL, IT, LU, NL, NO, PT, SE, SI, UK
Contentathon October 2020	Day 1 : 47 Day 2 : 44	WP1, WP2, WP3, WP4, WP5	BE, CH, CZ, DE, ELIXIR Hub, ES, FI, FR, GR, IL, IT, LU, NL, NO, PT, SE, SI, UK
Contentathon January	Day 1: 31	WP1, WP2, WP3, WP5	BE, CH, DE, EE, ELIXIR

³⁶<https://docs.google.com/document/d/102fbQiaZ63Gpl9ZHYLcg1HQmW9CJgTiBi5hJQmzslos/edit?usp=sharing>

³⁷<https://docs.google.com/document/d/1x3J2XeDQUxCcKxxu7rllgQsMy3vf-O7NIUQDh-y-usk/edit?usp=sharing>

³⁸<https://docs.google.com/document/d/1p1cLTwnv6HZJzYdzTGIB4LhHuE3cGZcnGITEJnTXCLQ/edit?usp=sharing>

2021	Day 2: 27		Hub, ES, FI, FR, IT, LU, NL, NO, PT, SE, UK
------	-----------	--	---

Hackathons are reported in Milestones M3.2.1³⁹ and M3.2.2⁴⁰. Training materials for the starter toolkit are reported in M3.4⁴¹.

4.3.3 Community Development

We focused on building a community around the toolkit through contentathons. We identified the roles of participation in the toolkit's development as content contributor, reviewer and editor. Contribution pathways and expected interaction between roles (discussed in Section 4.3.1) have been described on the website. To ensure smooth communication within the community we outlined a code of conduct⁴² on the website.

We started planning community outreach activities jointly with WP4. As a first activity we asked the community's input for choosing a name for the toolkit. At the beginning of the January 2021 Contentathon we asked participants to propose names for the toolkit and at the end of the event the proposals were put to a vote. The name "RDMKit" was selected by majority vote.

4.4 Workshops, Dissemination and External Engagement

Dissemination will be undertaken in earnest once the RDMKit is released Spring 2021. However, we have started dissemination, notably an All Hands 2020 Workshop in June

- Within CONVERGE
 - Initial meeting with WP5 and WP7 to discuss the representation of the CONVERGE Human Data use case within the toolkit. [slides⁴³]
- ELIXIR-wide
 - ELIXIR ALL HANDS 2020 WORKSHOP 13, Data management⁴⁴, 10 June 2020 (101 attendees)
 - ELIXIR Human Data Communities, presentation and call for contributions [slides⁴⁵]
- External
 - Presentation to European Commission Open Science Unit, organised by WP4 (14 January 2021). Discussions regarding RDMKit's role in HorizonEurope RDM recommendations and policy.
 - Dissemination to EU IMI FAIRplus AGM (14-15 January 2021). Discussions regarding integration with the FAIRplus FAIR Cookbook.

³⁹https://drive.google.com/file/d/1ZcFPSnszU0I8Y4FIj5G9eihYk_yAD3YZ/view?usp=sharing

⁴⁰https://docs.google.com/presentation/d/1t7tqvMkeEUZNSLWpTEo5rSZSbTEyMKZQkjRmHvsvdwbY/edit#slide=id.g75cc253543_3_33

⁴¹https://docs.google.com/presentation/d/1R_onNTRvWaPoEE7kRuy_dWqG3z11EggHiTRZUKPAi_w/edit?usp=sharing

⁴²https://rdm.elixir-europe.org/CODE_OF_CONDUCT.html

⁴³https://docs.google.com/presentation/d/1F2QqUdkLxA2S2TdyOchBPlx3dfo_8yWeD2G-gjhBczY/edit#slide=id.ga192857a77_0_56

⁴⁴https://docs.google.com/document/d/1gNiZkw_OZuSwLO1hyLLxqKBuwulRJQZQqKAKwZJZDHI/edit

⁴⁵https://docs.google.com/presentation/d/1-a_8zC3SYgD2GJkUX6JBgkO3AvgGW5TeN4q1CPhkJ8/edit#slide=id.ga192857a77_0_56



- ☺ Awareness raising in meetings with EU projects FAIRsFAIR and EOSC-Enhance.

5. Results

The RDMKit technical infrastructure has been established and the editorial processes put in place. The technical infrastructure and content has been assembled through the hackathons/contentathons (section 4.3.1). The name was chosen by ballot from a crowd-sourced set of nominations. The logo is being designed with the ELIXIR-Hub communication team and WP4, and will be finalised by the public beta-release.

The RDMKit link: <https://rdm.elixir-europe.org/> and <https://rdmkit.elixir-europe.org/>

The RDMKit Github link: <https://github.com/elixir-europe/rdmkit>

5.1 Landing Page

Figure 5 gives an annotated image of the RDMKit landing page.

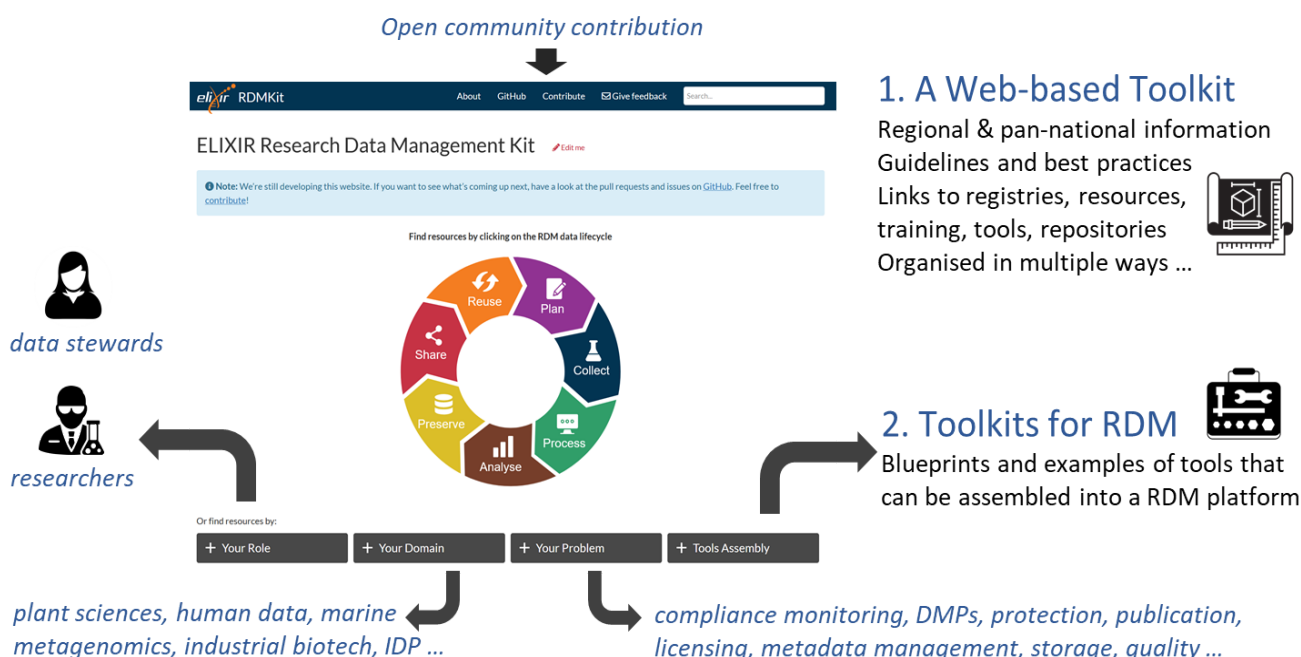
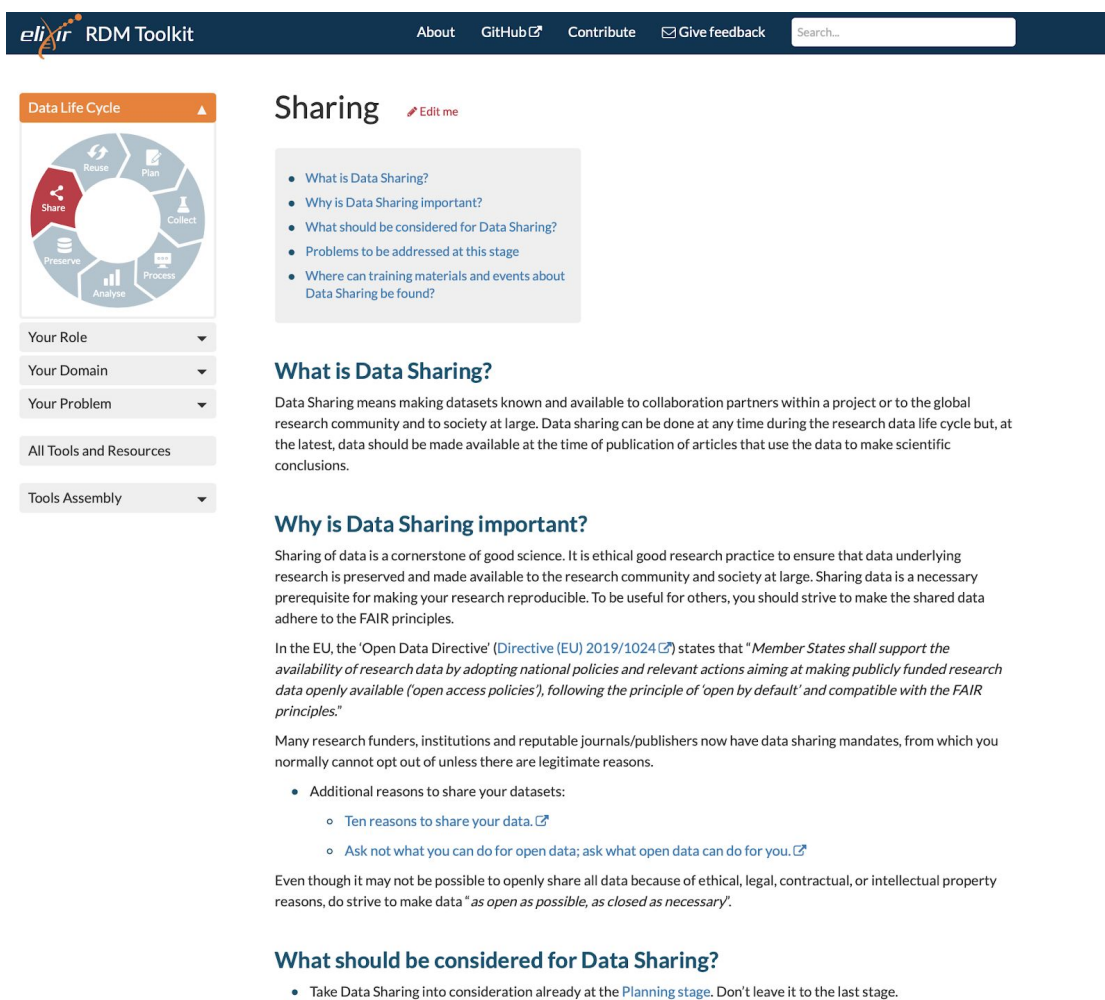


Figure 5. RDMKit Landing page

5.2 RDM Lifecycle

Figure 6 gives the top of a page for a stage of the RDM Cycle, selected from the landing page. Note the menu on the left provides content orientation and site navigation. The page has a menu at the top for rapid in page summary and navigation. All pages have this feature.



Sharing [Edit me](#)

- [What is Data Sharing?](#)
- [Why is Data Sharing important?](#)
- [What should be considered for Data Sharing?](#)
- [Problems to be addressed at this stage](#)
- [Where can training materials and events about Data Sharing be found?](#)

What is Data Sharing?

Data Sharing means making datasets known and available to collaboration partners within a project or to the global research community and to society at large. Data sharing can be done at any time during the research data life cycle but, at the latest, data should be made available at the time of publication of articles that use the data to make scientific conclusions.

Why is Data Sharing important?

Sharing of data is a cornerstone of good science. It is ethical good research practice to ensure that data underlying research is preserved and made available to the research community and society at large. Sharing data is a necessary prerequisite for making your research reproducible. To be useful for others, you should strive to make the shared data adhere to the FAIR principles.

In the EU, the 'Open Data Directive' ([Directive \(EU\) 2019/1024](#)) states that "Member States shall support the availability of research data by adopting national policies and relevant actions aiming at making publicly funded research data openly available ('open access policies'), following the principle of 'open by default' and compatible with the FAIR principles."

Many research funders, institutions and reputable journals/publishers now have data sharing mandates, from which you normally cannot opt out of unless there are legitimate reasons.

- Additional reasons to share your datasets:
 - [Ten reasons to share your data.](#)
 - [Ask not what you can do for open data; ask what open data can do for you.](#)

Even though it may not be possible to openly share all data because of ethical, legal, contractual, or intellectual property reasons, do strive to make data "as open as possible, as closed as necessary".

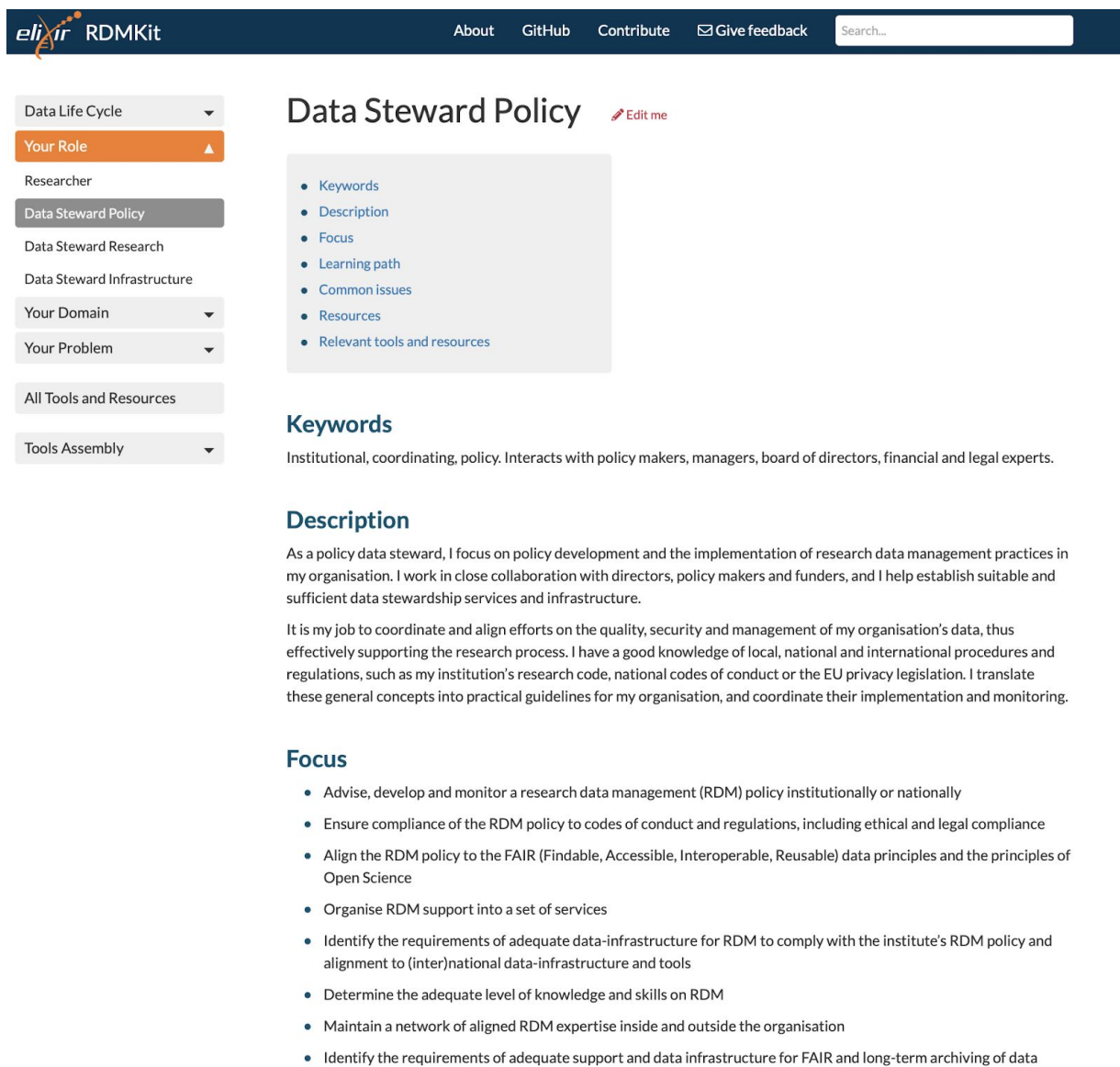
What should be considered for Data Sharing?

- Take Data Sharing into consideration already at the [Planning stage](#). Don't leave it to the last stage.

Figure 6. Top of the page for the RDM stage for Data Sharing

5.3 Your Role

Figure 7a shows the Role page for the Policy Data steward. Each role page begins with a description of the specific role to help site visitors to identify themselves with the role. Under the focus section these pages list responsibilities of data stewards in each role. These lists are jointly written by data stewards during role-focused break-out sessions.



The screenshot shows the top part of the RDMKit website. The header includes the elixir RDMKit logo, navigation links for 'About', 'GitHub', 'Contribute', and 'Give feedback', and a search bar. On the left sidebar, 'Your Role' is selected, with 'Researcher' and 'Data Steward Policy' as sub-options. The main content area is titled 'Data Steward Policy' with an 'Edit me' link. A list of links includes 'Keywords', 'Description', 'Focus', 'Learning path', 'Common issues', 'Resources', and 'Relevant tools and resources'. The 'Keywords' section defines the role as institutional, coordinating, and policy-oriented. The 'Description' section explains the role's focus on policy development and implementation. The 'Focus' section lists eight key responsibilities, such as developing RDM policies, ensuring compliance, and organizing support services.

Figure 7a. Top part of the role page for Policy Data Steward

Figure 7b shows the bottom part of the Role page for the Policy Data steward. Under Learning Path the role page lists competency areas for the role and lists the resources that can be used to obtain those competencies, such as training resources expert networks. Similar to other pages the Role pages also provide a filtered view of the tools table. The tools listed are those that are mostly used by the role in focus.

Learning path

Institutes across Europe have started hiring professional data stewards. A policy oriented data steward is expected to be competent in the following areas:

- Develop, implement, monitor and evaluate policies regarding research data that are endorsed by researchers and aligned with internal and external stakeholders and effectuate change management
- Communicate about the FAIR data and Open Science principles to researchers, support staff and relevant stakeholders, thus creating awareness within institutes and organisations
- Give advice on RDM and formats for data management plans (DMP) to a broad audience
- Communicate about the RDM policy, explain implications and create awareness
- Translate RDM policy and legislation and codes of conduct concerning research data into practical implications and guidelines for that researchers can understand
- Formulate requirements for RDM support (staff and services)
- Assess and analyse needs from researchers regarding support on RDM for researchers and define new requirements
- Assess RDM knowledge and skills and identify gaps among researchers and relevant stakeholders

If you want to become competent in these areas or build capacity in your institution then the following training resources might be useful:

- [TeSS](#): ELIXIR's training portal
- [RDNL Essentials for Data Support](#)
- [Mantra RDM training](#)
- [GO FAIR resources](#)
- [Data Carpentry lessons](#)
- [RDNL & DCC Delivering RDM Services](#)

Common issues

- [Compliance Monitoring & Measurement](#)
- [Data Classification](#)
- [Data Management Plan](#)
- [Licensing](#)

Resources

- [NPOS/ELIXIR data steward competency framework](#)
- [FAIR guiding principles](#)
- [Science Europe Practical Guide to the International Alignment of RDM](#)
- [ELIXIR Data Managers Network](#)

Relevant tools and resources

Tool or resource	Description	Tags	Registry
BBMRI-ERIC's ELSI Knowledge Base	The ELSI Knowledge Base is an open-access resource platform that aims at providing practical know-how for responsible research.	<ul style="list-style-type: none"> data protection data classification policy officer data manager human data 	

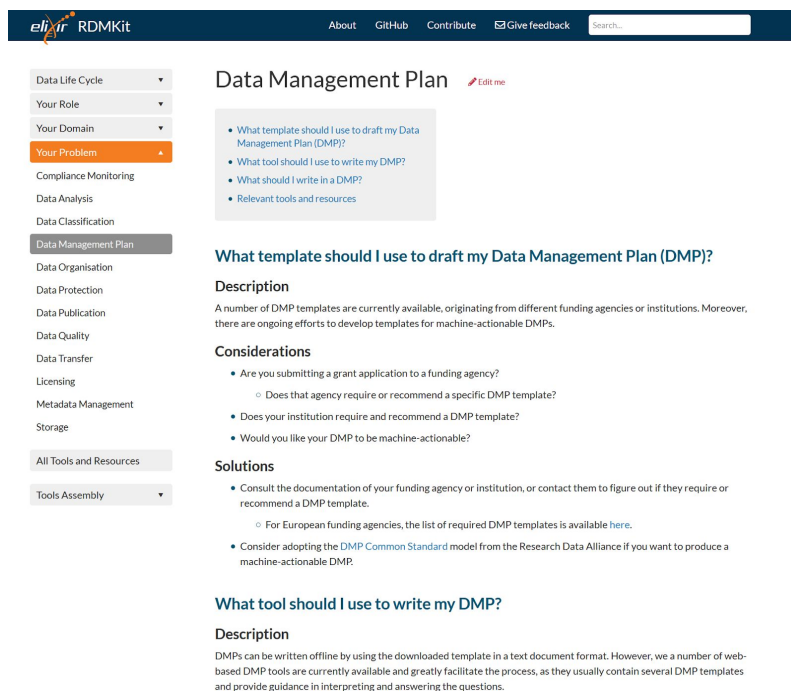
Figure 7b. Bottom part of the role page for Policy Data Steward.



Building on the researcher need’s and competencies for data steward (and potential other RDMKit support roles), WP1, WP2 and WP3 together will continue to work on exploring and integrating different training resources for researchers and data stewards in the toolkit, for instance by improving the links to TeSS.

5.4 Your Problem

Figure 8a gives the top of a Problem page highlighting challenges and approaches. Note the narrative linking for tools and templates. This is hand-written by the contributors and editors.




The screenshot shows the top of a 'Your Problem' page on the RDMKit website. The page title is 'Data Management Plan' with an 'Edit me' link. A navigation menu on the left lists various topics, with 'Your Problem' highlighted. The main content area contains several sections: a list of bullet points for 'What template should I use to draft my Data Management Plan (DMP)?', 'What tool should I use to write my DMP?', and 'What should I write in a DMP?'. Below this is a section titled 'What template should I use to draft my Data Management Plan (DMP)?' with a 'Description' and 'Considerations' subsection. Another section titled 'What tool should I use to write my DMP?' includes a 'Description' subsection. The page is designed to provide guidance and resources for creating a Data Management Plan.

Figure 8a. Top of the Your Problem page for Data Management Planning.

Figure 8b shows the list of relevant tools and resources at the bottom of the page, incorporated through the tagging system. Note that two tools have regional access limitations and one has an entry in the bio.tools registry. The tags are displayed and are active as another mechanism for navigation. Also note the two contributors who wrote this page are credited at the bottom. Credit is recorded and displayed automatically through the GitHub system. Tracking credit for contributors who have not registered in GitHub has to be done by hand. Automated credit appears on all the pages. These pages will be a particular on-going focus for the ELIXIR Platforms.

Relevant tools and resources

Tool or resource	Description	Tags	Registry
DMP Canvas Generator	Questionnaire, which generates a pre-filled a DMP	DMP researcher data manager	
DMP OPIDoR	Online questionnaire for the development of data management plans; repository of DMPs	DMP researcher data manager	
DMPlanner	Semi-automatically generated, searchable catalogue of resources that are relevant to data management plans.	DMP researcher data manager	
DMPonline	A free tool to write, share and export a data management plan. Built-in data management plan templates for many major funders.	DMP researcher data manager	
DMPonline Belgium	A free tool to write, share and export a data management plan. Instance aimed at Belgian researchers with built-in data management plan templates for the major funders.	DMP researcher data manager	
DMPTuuli Finland	Data management planning tool	DMP researcher data manager	
DS-Wizard	Data Stewardship Wizard	DMP researcher data manager IT support nels	
EasyDMP	DMP creation, versioning and sharing	DMP researcher data manager	
maDMP - Research Bridge	Machine-Actionable Data Management Plan Webinar (2016) on making a good data management plan.	DMP IT support	
Research Management Plan	Machine actionable DMPs.	DMP researcher data manager	

Tags: [data manager](#) [plan](#) [policy officer](#) [researcher](#)

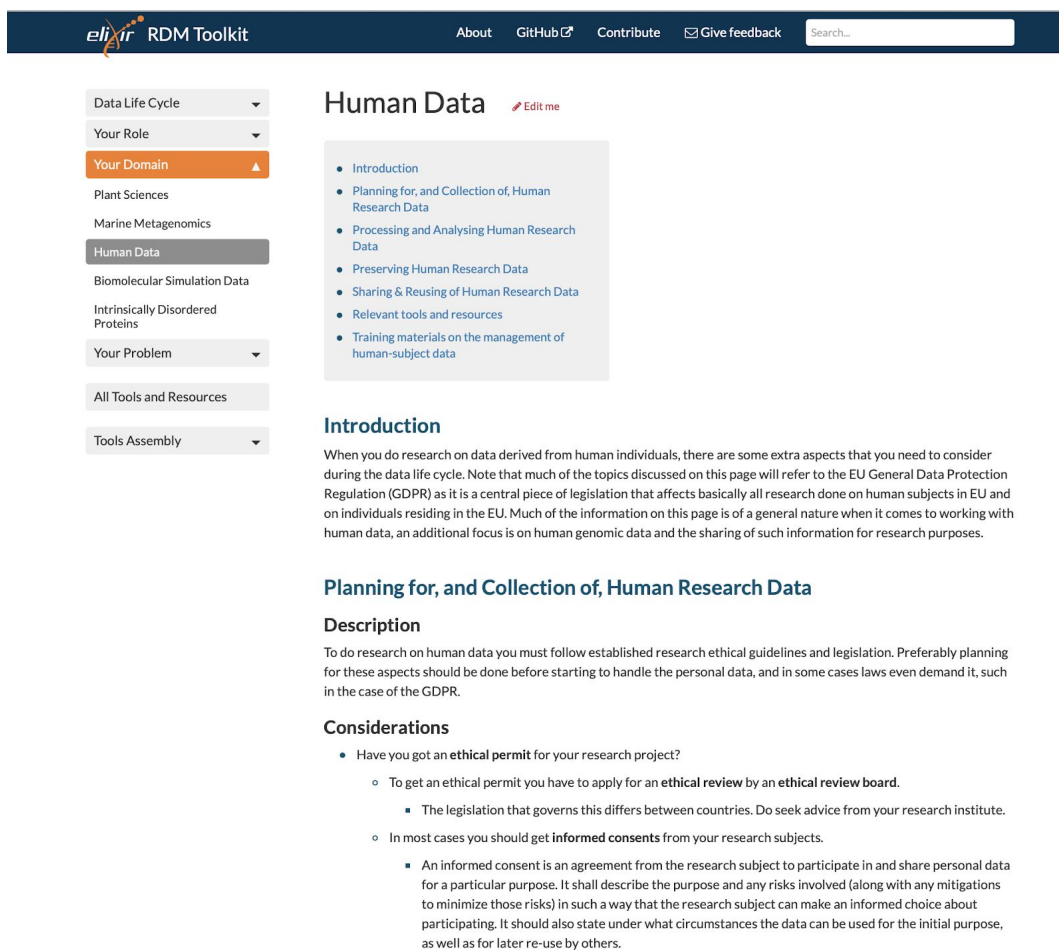
Contributors:  [Flora D'Anna](#),  [Daniel Faria](#)

Figure 8b. Tools tagged for Data Management Planning at the bottom of the page.

5.6 Your Domain

Figure 9 shows the top of the Human Data domain page. The beta release has pages for Plant Sciences, Marine Metagenomics, Human Data, Biomolecular Simulation Data and Intrinsically Disordered Proteins. These domains are demonstrators within WP5 or have stepped up through the ELIXIR Communities and Focus Groups (e.g. ELIXIR Microbial Biotechnology are preparing pages).





Human Data [Edit me](#)

- [Introduction](#)
- [Planning for, and Collection of, Human Research Data](#)
- [Processing and Analysing Human Research Data](#)
- [Preserving Human Research Data](#)
- [Sharing & Reusing of Human Research Data](#)
- [Relevant tools and resources](#)
- [Training materials on the management of human-subject data](#)

Introduction

When you do research on data derived from human individuals, there are some extra aspects that you need to consider during the data life cycle. Note that much of the topics discussed on this page will refer to the EU General Data Protection Regulation (GDPR) as it is a central piece of legislation that affects basically all research done on human subjects in EU and on individuals residing in the EU. Much of the information on this page is of a general nature when it comes to working with human data, an additional focus is on human genomic data and the sharing of such information for research purposes.

Planning for, and Collection of, Human Research Data

Description

To do research on human data you must follow established research ethical guidelines and legislation. Preferably planning for these aspects should be done before starting to handle the personal data, and in some cases laws even demand it, such in the case of the GDPR.

Considerations

- Have you got an **ethical permit** for your research project?
 - To get an ethical permit you have to apply for an **ethical review** by an **ethical review board**.
 - The legislation that governs this differs between countries. Do seek advice from your research institute.
 - In most cases you should get **informed consents** from your research subjects.
 - An informed consent is an agreement from the research subject to participate in and share personal data for a particular purpose. It shall describe the purpose and any risks involved (along with any mitigations to minimize those risks) in such a way that the research subject can make an informed choice about participating. It should also state under what circumstances the data can be used for the initial purpose, as well as for later re-use by others.

Figure 9. The top of the Human Data domain.

5.7 Tool Assemblies

Tool Assemblies are the second stream of the work of WP3. Tool Assemblies are examples of combining tools to cover the “Your Problems” and Lifecycle Stages. They may be domain specialized. The example shown in Figure 10 is our pilot - the Norwegian eInfrastructure for Life Sciences (NeLS). In future work assembly patterns (blueprints) will be distilled from real examples from the Nodes to show how tools generally can be combined. Tool Assemblies are the focus of our future work post the beta public release and we recognise that the pages need further work.

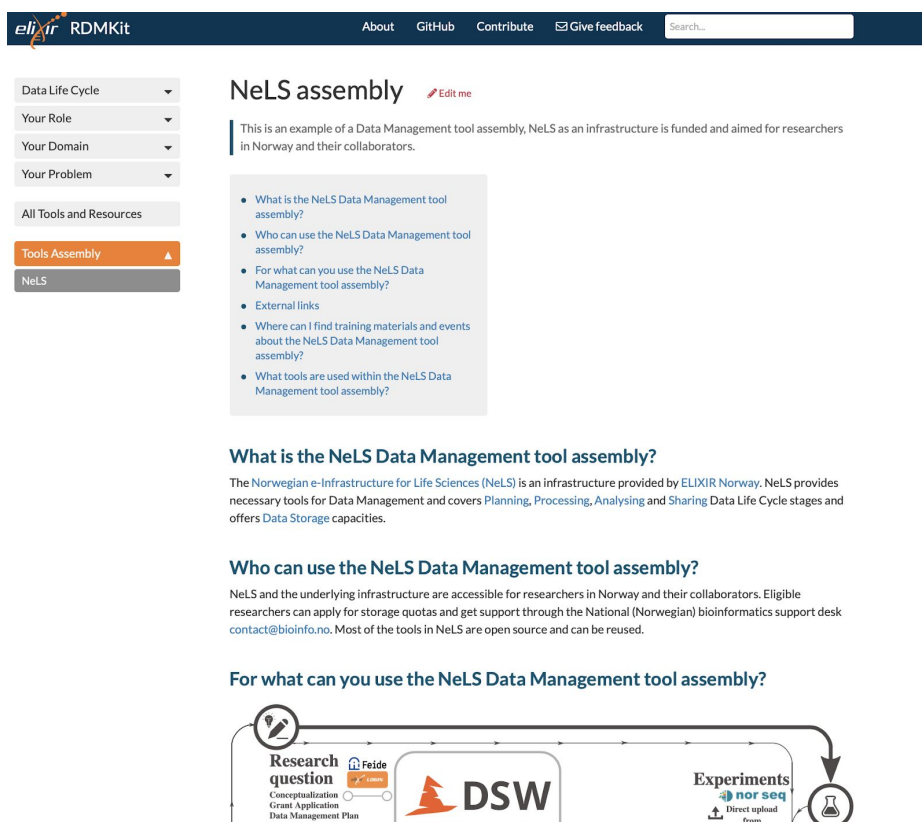


Figure 10. The top of our current example of a Tool Assembly.

6. Conclusions

In our first 12 months of work WP3 have matured the Toolkit vision into three streams: **The RDMKit**, a website-based toolkit of guidelines, best practices and resources; **Tool Assembly** examples of combining tools to data stewardship across the stages of the RDM Lifecycle; and integration with the **Data Stewardship Wizard** to further assist data stewardship through decision trees. Each is interlinked. Work on assembling the starter toolkit has chiefly concentrated on the RDMKit website.

WP3 has successfully built a cohesive, productive and committed team. The work on RDMKit has unified and mobilised the WP3 tasks and cross WP work. It gives the CONVERGE project a common goal and a place to gather and disseminate the results of the Data Management experts (WP1) and training (WP2). Work has covered the design and implementation of the site, and the processes for community contribution of content and editorial oversight.

The RDMKit is ready for internal ELIXIR beta-release at the M12 point of CONVERGE. The RDMKit, even in its beta form, already provides a valuable one-stop shop of research stewardship intelligence tailored for the Life Sciences.

We are now well-placed to beta release and continue work on the RDMKit, incorporating other tasks in CONVERGE (e.g. T1.2, data brokerage pipelines) and WPs. At the public beta release we will be in a



position to widen community engagement and launch a dissemination campaign with WP4. We are now in a position to expand on the other workstreams with a clear idea of what the tasks need to do and how all three streams work together.

7. Impact

As outlined in the conclusions the WP3 Toolkit has galvanised the CONVERGE WPs 1-5 to work together around a common goal. Every Node has contributed to some degree and some Nodes have really stepped up.

The RDMKit will support the ELIXIR Nodes in their RDM efforts to serve their Nodes and national communities, build and exchange capacity and capability and showcase data stewardship resources and tools provided by and available to the Nodes.

The RDMKit is an opportunity to provide a single point of knowledge exchange and showcasing for the RDM related activities by and for the ELIXIR Platforms⁴⁶, Communities⁴⁷ and Focus Groups⁴⁸ and EU projects⁴⁹ such as FAIRplus, EOSC-Life, EOSC-Enhance, B1MG, CINECA, and The European Joint Programme on Rare Diseases.

An early example of the impact of the RDMKit on National nodes is a recent award to the ELIXIR-UK node in answer to the call UKRI Innovation Scholars: Data Science Training in Health and Bioscience⁵⁰. The call was highly competitive. The 24 month ELIXIR-UK project (announced January 2021) focuses on FAIR Data Stewardship training for the UK bioscience community, and places the RDMKit as a key pillar of its proposed activities - both as a resource and as a knowledge hub for results, pooling nationally developed materials for the wider good.

Preliminary discussions with European Commission organised by WP4 are steps towards RDMKit's incorporation in EC policy and recommendations for data management planning in Horizon Europe proposals. To quote Kostas Repanas from the Open Science Unit of the Commission "this toolkit does our job... we can include it in our guidelines".

8. Next Steps

Next steps follow several intertwined lines of work.

8.1 RDMKit website release, promotion, extension and continual development

RDMKit internal beta release is 31 January 2021. The release will widen awareness across ELIXIR, in particular ELIXIR Communities, Focus Groups and Platforms, as well as widen awareness across ELIXIR Nodes. This should raise interest in recruiting new contributors and editors, and provide feedback on content, organization and the contribution and editorial processes we have put into place. An internal webinar and internal communications campaign will be coordinated with WP4.

WP3 will run a gap analysis on the current content and actively target experts in CONVERGE and across ELIXIR to fill those gaps. A gap that will be addressed is the inclusion of the Role/Persona of

⁴⁶<https://elixir-europe.org/platforms>

⁴⁷<https://elixir-europe.org/communities>

⁴⁸<https://elixir-europe.org/focus-groups>

⁴⁹<https://elixir-europe.org/about-us/how-funded/eu-projects>

⁵⁰<https://www.ukri.org/opportunity/innovation-scholars-data-science-training-in-health-bioscience/>



project consortium coordinators in preparation for incorporation in Horizon Europe guidance. At least one contentathon will be organised before the public beta release and we will encourage ELIXIR Communities to run their own contentathons. Processes for handling inquiries and feedback will be developed. Processes for extending the RDMKit (see M3.3) will be refined.

RDMKit Public release is 31 March 2021. We will launch a communications campaign with WP4, including an external webinar and promotion in every Node through national channels and events, European channels and through our networks, ELIXIR channels and projects. Processes for content, contribution and enquiries will be monitored and reviewed. Processes for extending the toolkit (see Milestone M3.3) will be reviewed and refined, and applied to extend the toolkit to address additional communities beyond the first wave of WP5 demonstrators, to be reported in Deliverable D3.2. All demonstrators in WP5 should be represented in the RDMKit by Month 18. A review of the RDMKit will be reported in Milestone M3.6 (toolkit review and gap analysis based demonstrator support).

The Editorial Board will continue to meet bi-weekly. The RDMKit will continually evolve as a living site throughout the CONVERGE project duration and beyond. This is a long-term commitment. Content must be current and fresh.

The RDMKit extensions/development in the next reporting period include:

- Defining relationship to ELIXIR Service Bundles⁵¹;
- Planning incorporation of results of the Data Submissions Brokering explorations (Task 1.2);
- Defining relationship to the FAIRPlus FAIR Cookbook⁵²;
- Defining relationship to EOSC resources such as the EOSC Portal Catalogue & Marketplace⁵³ and openAIRE⁵⁴;
- Horizon scanning for resources within ELIXIR, within EOSC-Life and other projects (e.g. the forthcoming FAIRSharing Resource Wizard) and outside ELIXIR;
- Further calls to capture more tools and get tools registered in bio.tools.

8.2 Tool Assemblies from WP5 and pattern development

Tool Assemblies are examples of combining tools to provide data management across the stages of the RDM Lifecycle or a subset of stages, and the distillation of assembly patterns (blueprints) to guide how tools can be combined. We currently have one example for the Norwegian eInfrastructure for Life Sciences. In the next phase of WP3 work on the toolkit is to expand on the examples to include examples from all the demonstrators in WP5, improve the ability of others to implement their own combination tailored to their local environments, and to write guidelines on general patterns for combining tools. The expansion of this part of the toolkit builds on milestone M3.3 and will be reported in Deliverable D3.2 (toolkit extension based on first wave demonstrators).

The explorations in Task 1.2 (Data Submissions Brokering) will significantly contribute to this task. In addition a tools landscape analysis across projects in the EOSC-Life WP3 open call⁵⁵ will provide more examples.

⁵¹<https://elixir-europe.org/services/service-bundles>

⁵²<https://fairplus.github.io/cookbook-dev/intro.html>

⁵³<https://eosc-portal.eu/>

⁵⁴<https://www.openaire.eu/>

⁵⁵<https://www.eosc-life.eu/services/open-call/>



8.3 Data Stewardship Wizard

Task 3.3 matches tools with users. The ELIXIR Data Stewardship Wizard⁵⁶ (DSW) is built around a hierarchical knowledge model that is used to lead data stewards through a decision tree to help them choose the right tools, resources and practices when making data management plans. The DSW and RDMKit are bi-directionally connected. At key places in its knowledge model Wizard users are referred to the RDMKit for background information and overviews. Users of the RDMKit are referred to the Wizard to help make context-dependent decisions between the alternative options. Milestone M3.5 sets out the interaction. The combination of the DSW and RDMKit combines rich and deep decision making with up to date and comprehensive information from the Nodes.

8.4 Training and capacity building

Training on the toolkit will be organised around the toolkit's three streams. Target groups for the training will be the Data Managers Network, CONVERGE demonstrator teams and ELIXIR communities. The primary goal of the training would be to allow the wider RDM community to understand, use and extend the toolkit. Results will be reported in deliverable D3.4 (Dissemination of the toolkit across the ERA. Training and capacity building of all stakeholders).

9. Deviation from Description of Action

This deliverable was originally due in M6. It was delayed due to a number of factors. The most overwhelming impact was from the COVID-19 crisis, which not only disrupted the lives of colleagues but also diverted them to urgent work focused on COVID data infrastructure, data collection, data FAIRification and data analysis. A delay was applied for and accepted.

The development of the understanding of the three different facets of the toolkit was finalised in June 2020. The proposal envisaged the Tool Assemblies and a portal to the assemblies; the RDMKit was realised as the toolkit that was actually needed to link up the WPs and tasks, build community and exchange knowledge.

10. Appendix

10.1 Editorial Board

- Carole Goble (UK)
- Munazah Andrabi (UK)
- Daniel Faria (PT)
- Frederik Coppens (BE)
- Flora D'Anna (BE)
- Bert Droesbeke (BE)
- Rob Hooft (NL)
- Mijke Jetten (NL)
- Niclas Jareborg (SE)
- Pinar Alper (LU)

⁵⁶<https://ds-wizard.org/>



- Ulrike Wittig (DE)
- Laura Portell (ES)
- Martin Cook (ELIXIR Hub)

(<https://github.com/elixir-europe/rdm-toolkit/blob/master/CODEOWNERS>)

