JLSC

ISSN 2162-3309 | JLSC is published by the Iowa State University Digital Press | http://jlsc-pub.org

Volume 12, 1 (2024)

Developing and Improving the Capability of Scientific Communities in Open Research Data: A Maturity Continuum Model: Concepts, Elements, and Applications

Marielle Guirlet & René Schneider

Guirlet, M. & Schneider, R. (2024). Developing and Improving the Capability of Scientific Communities in Open Research Data: A Maturity Continuum Model: Concepts, Elements, and Applications. *Journal of Librarianship and Scholarly Communication*, 12(1), eP16320. https://doi.org/10.31274/jlsc.16320

This article underwent fully anonymous peer review in accordance with JLSC's peer review policy.

ISSN 2162-3309 10.31274/jlsc.16320

THEORY ARTICLE

Developing and Improving the Capability of Scientific Communities in Open Research Data: A Maturity Continuum Model: Concepts, Elements, and Applications

Marielle Guirlet
UNIRIS – Department of Information Resources and Archives, University of Lausanne, Switzerland
René Schneider
Information Science, Haute Ecole de Gestion, HESSO, Geneva, Switzerland

ABSTRACT

This article presents a new tool developed to assess and increase the maturity level of a research community in terms of Open Research Data (ORD) practices and culture. This model, called the Maturity Continuum Model, is divided into four successive steps, starting from the awareness and initiation of the community to its full maturity in ORD, through the building and strengthening of this community. Each step has its own intermediate objectives, and each benefits from tools and means to support the process of gaining maturity. The steps are separated by evaluation boundaries that are used to assess whether the process can move to the next step. Key stakeholders are diverse and contribute at different stages of the process. The Maturity Continuum Model was inspired by the Data Curation Continuum Model (Treloar et al., 2007; Treloar and Klump, 2019). Considering the aspects in which these two tools differ, they can be seen as complementary tools in the broad field of research data management. Our model may be used at the level of a specific research community, with several possible purposes, such as assessing, developing, and monitoring its maturity. It can also be used to compare the maturity level of several communities at the same time and to identify the success factors of more advanced research communities in order to transfer them to less advanced ones. We have designed and created this model primarily to get an overview of a complicated topic, to find a common language, to provide a common basis for discussion, and to enable implementation over a long period of time. Successful implementation of the model would also allow for consolidation of scientific communities in terms of ORD, provide standards, and allow for evaluation.

Keywords: Open Science, Open Research Data, data sharing, Research Data Management, FAIR principles, (meta)data standards, scientific communities, research communities, data stewards, data librarians, Data Curation Continuum, Maturity Continuum Model

Received: 04/28/2023 Accepted: 11/05/2023

IMPLICATIONS FOR PRACTICE

1. The model presented here is a tool for academic libraries and other stakeholders in scholarly communication to support research communities in achieving Open Research Data (ORD) literacy.
2. Readers may gain a better understanding of the specific needs of a research community in terms of support, training, and resources at each step of the process, from the creation of the community, through the development of its ORD maturity level, until it reaches full autonomy in ORD practices and culture. Readers may also gain an awareness of the specific role of various stakeholders in contributing to the process and the best timing in the process for that contribution.
3. This tool can also be applied to several communities at the same time in order to compare them and apply the success factors of the more advanced communities to the less advanced ones. It then helps to rationalize the efforts and resources spent on developing and improving ORD practices and culture.

INTRODUCTION

In this article, we present and describe a new model, the Maturity Continuum Model (MCM), which conceptualizes the maturity level of a research community with respect to Open Research Data (ORD). This model can be used to increase this maturity level and to strengthen the use of standards and standardization in this area, i.e., ORD, by research communities. It can also be used to find a common language and shared conceptualizations.

We have already briefly presented this model in Guirlet et al. (2023). In the current article, we provide more detailed and precise information about the model, and we discuss the full range of future implementations. We describe the involvement of key stakeholders and their potential contributions. We also discuss the human aspects that need to be considered in the context of MCM. Finally, we compare our model with the Data Curation Continuum (DCC), which provided us with a lot of inspiration for the conceptualization and development of our model (Treloar et al., 2007; Treloar and Klump, 2019).

Our motivation to develop this instrument was based on several observations derived from the literature and from a recent study commissioned by swissuniversities on the ORD practices and needs of Swiss researchers, which is fully described in Bongi et al. (2021).

The results of an international survey by Tenopir et al. (2020) show that scientists are generally willing and even enthusiastic about sharing their data. However, their practices do not

really reflect this positive attitude toward open data. The authors suggest that "stronger organizational involvement in providing training and support of good practices," with some support from data managers and data librarians, would help to close this gap.

It is recognized that a community facilitates and stimulates the ORD practices of its members (Berman et al., 2013; Cooper and Springer, 2019). The ORD practices of researchers who are members of a community are more advanced than those of researchers who are not members of a community (Bongi et al., 2021, p. 44).

However, the same study highlights that the maturity level of research communities in the Swiss academic environment is very diverse. The least advanced communities are barely organized, showing an incipient or even non-existent data sharing culture.

Other more advanced communities are better organized, but with limited data sharing practices. However, they have established policies for data curation, i.e., sharing their data through repositories and preserving data in newly built archives. And the most advanced communities show widespread and systematic adoption of ORD best practices by their members. Some of them have been very advanced for a long time. These communities have established their own standards over time (see, e.g., Berman et al., 2013, on how the structural biology community has shaped, established, and developed the Protein Data Bank over the last 40 years).

Looking closely at specific communities, the adoption and use of standards appropriate to a particular community are good proxies for the maturity of that community with respect to ORD. The level of adoption and use of these standards is also a lever for strengthening good practices in ORD within the community (Bongi et al., 2021, p. 44).

We infer from these results that stimulating and facilitating the formation of research communities, as well as encouraging the definition and adoption of standards within those communities, significantly contributes to the improvement of ORD practices by researchers in those communities. Based on these observations, we have developed a model that can be used to assess the ORD maturity level of a given research community and to develop that maturity level.

DATA COMMUNITIES, RESEARCH COMMUNITIES—CONCEPTS & DEFINITIONS

So far in this article we have used the term "research community" or "scientific community," although it seems more natural to use the term "data community" when focusing on ORD and data sharing. We will now discuss these definitions in more detail and our rationale for using the term "research community."

Data sharing involves first establishing a dialog with researchers and building services at the scale of a community (Rinehart, 2022). According to Cooper and Springer (2019), "successful data sharing happens within data communities" (p. 4). And the policies, practices, and tools (standards and infrastructure) for data "usually refer to the communities associated with those data," as Borgman (2015) notes. However, defining exactly what a data community is remains a challenging task, and consensus on its exact meaning has yet to be reached.

For Borgman (2015), a data-related community can be a community of interest, as also described by Power (2021): "Data communities are specialized communities made up of organizations that have data and organizations that want data tied together based on an intent-driven use of the data." It can also be a community of practice where one learns and shares knowledge and gains expertise (Borgman, 2015).

The discipline does not necessarily match the data community, as a particular discipline may not be interested in a particular dataset (Rinehart, 2022), and some datasets may be of interest to multiple disciplines. For example, the National Snow and Ice Data Center collects and stores data from different research areas (seismology, glaciology, oceanology, hydrology, climatology, anthropology, etc.) and for the potential benefit of scientists from different disciplines (National Snow and Ice Data Center, 2023). Similarly, the Global Earth Observations System of Systems shares environmental data from different observing systems for the benefit of all researchers focused on Earth processes (GEO, n.d.).

Bongi et al. (2021) also highlighted that the organization of communities is heterogeneous (in the Swiss academic environment). They can be organized in the form of an association, a project, an infrastructure, or, less frequently, a social network. However, most of them are informal networks without real boundaries. Some researchers feel that they belong to more than one of these informal communities (Bongi et al., 2021, pp. 19–20).

Nevertheless, Borgman (2015) emphasizes that collaboration is an important aspect as it reveals the communities: "By examining the role that data play in collaborations, the boundaries, scope, agreements, and disagreements of communities come into view" (p. 36). Cooper and Springer (2019, p. 4) define these data communities as "formal or informal groups of scholars who share a certain type of data with each other, regardless of disciplinary boundaries."

However, there is currently little evidence that such communities are a reality, at least in the Swiss academic environment (Bongi et al., 2021, p. 16).

Furthermore, given that sharing one's data in practice implies the existence and use of several other elements, concepts, or tools, we believe that the communities we focus on in the context

of this model should be considered as a broader group that shares more than just data of a specific type.

We therefore propose to use the more general term "research community" instead of "data community" and its definition given in Bongi et al. (2021, p. 16): "networks of researchers sharing the same (type of) data and/or the same data sharing tools and infrastructures, the same ORD practices, and the same open science culture, but not necessarily belonging to the same research discipline."

FURTHER DEFINITIONS

To avoid any misunderstanding or confusion, we define several key concepts that will be addressed and developed throughout the model description.

Standards

The standards mentioned here obviously refer to the standards relevant to data sharing. These standards may relate to metadata, other documentation, file formats, persistent identifiers, data licenses, etc.

Infrastructures

In the following, the term infrastructure refers to data sharing infrastructure, i.e., data storage and sharing solutions. Beyond the minimum, infrastructures very often provide specific features and tools to support the repository process, to make the data FAIR (Wilkinson et al. 2016), and to archive them. By default, the term "infrastructures" will also include these features and tools.

Services

The term "services" or "support services" in our context refers to services provided to researchers to adopt and improve their research data management (RDM) practices. These services may include information, training, coaching, assistance in preparing data before uploading to a repository, etc.

MODEL DESCRIPTION

The development and specification of our MCM was inspired by the DCC (Treloar et al., 2007; Treloar and Klump, 2019). Although the DCC focuses specifically on research data and

metadata and aims to support the sharing of data or other research objects, its components can be used *sui generis* in a broader context and the model can be adapted to another environment.

MCM has thus been developed by transferring the DCC's focus on standards and other concepts and tools related to ORD, and with the aim of supporting the creation and development of research communities to improve ORD practices and broaden their adoption and use.

In the following paragraphs, we describe the key features of the MCM (Figure 1). We also highlight the similarities and differences between the MCM and DCC. We then discuss the potential applications and future prospects for the MCM.

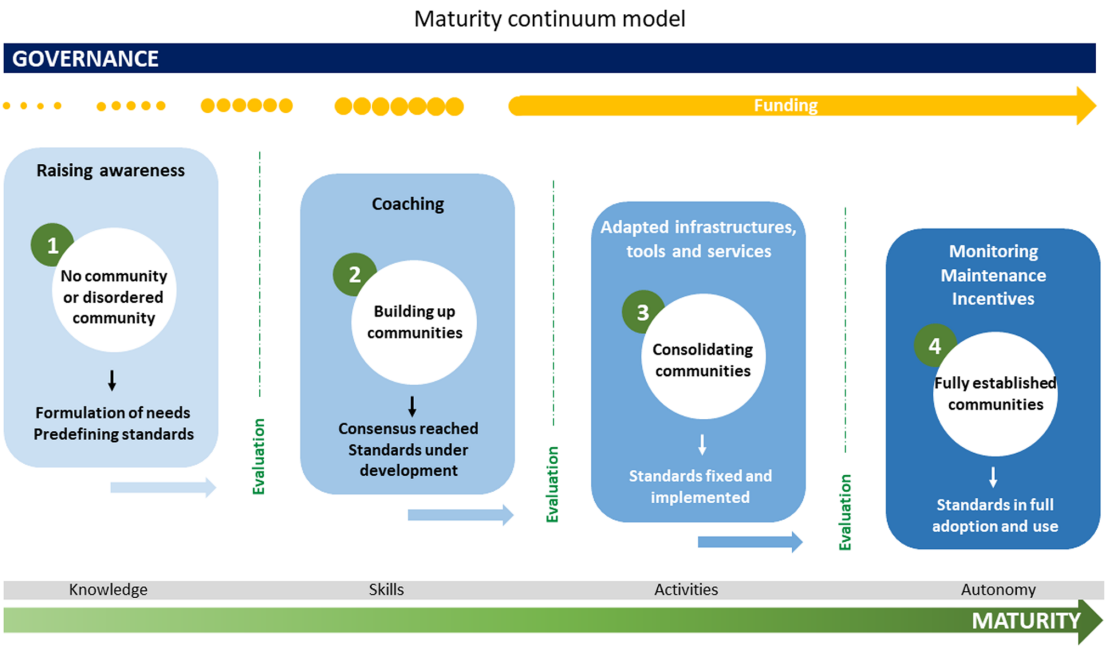

Figure 1. Maturity continuum model for Open Research Data.

A continuum segmented in four successive steps

As mentioned earlier, there is a wide range of ORD maturity levels between research communities (at least in Swiss academic research). The broad spectrum of ORD maturity would therefore be better reflected by a continuum. However, as a continuum is difficult to handle as such, the model is segmented into four domains or steps, which are separated by boundaries. These boundaries also serve as evaluation steps to assess whether the objectives of the current step have been achieved and whether it is appropriate to move on to the next step.

Each step addresses a specific aspect and has its own objectives. However, all steps are underpinned by the standards and standardization of data sharing practices. As discussed earlier, these elements are both maturity indicators and leverage tools for developing and strengthening the community in terms of ORD. From the pre-definition of these standards in the first step to the widespread adoption and use in the last step, efforts are focused on implementing a systematic use of standards and standardization as far as possible within the community.

Step 1: Raising awareness

The first step can be seen as a "nudge" to start the whole process. In this step, it is observed that there is no community yet, or that the community is very disorganized when it comes to managing and sharing its research data. The goal here is to initiate the creation and organization of the community, or to reorganize it in a more structured way.

In order to trigger the development of a common culture and practices, awareness raising will first focus on the potential benefits for research and researchers. Once the benefits are recognized by researchers, efforts can focus on formulating more specific needs for the community and pre-defining standards that are appropriate for the community.

Researchers need to be heavily involved in this first step. This is because, for one, performing these tasks together is a way to rally them around a common goal as early in the process as possible and to initiate networking practices and a common culture. A second reason is that it is critical that future users (e.g., community members) drive and shape the development of standards for their own domain to ensure broad and sustainable adoption in the future. A parallel can be drawn here with the approach of the Research Data Alliance (RDA) and the involvement of its communities of practice in all of its work and the development of tools, guidelines, and policies in specific disciplines or research areas (see, e.g., RDA, 2024).

Following the pre-definition of these standards, the next steps focus on the development of standards appropriate for the community and their adoption and systematic use by community members.

Step 2: Coaching

The community initiated by the nudge of the first step is now being built.

This second step aims to characterize the standards that are appropriate for the community. For this purpose, coaching support is needed. This coaching involves encouraging participation and facilitating dialog among community researchers until a broad consensus on

standards is reached within the community. From this consensus, standards development can begin.

The role of coach could possibly be played by a data steward. The Dutch Techcentre for Life Sciences defines data stewardship as "the responsible planning and executing of all actions on digital data before, during and after a research project, with the aim of optimizing the usability, reusability and reproducibility of research data" (Scholtens et al. 2019, p. 11). This broad mission of data stewards translates into a variety of tasks, contributions, and means of support for researchers. Verheul et al. (2019, p. 7) distinguish between the role of embedded and operational data stewards and the role of generic and advisory data stewards. The latter provide support and advice to researchers on RDM activities, as well as information and training. Given their social and interpersonal skills—communication, building relationships and networks, developing and maintaining trusted relationships (see, e.g., Armstrong et al., 2021; Gruber et al., 2021)—generic data stewards could successfully play this role of coach for the MCM process.

A community manager, as described in the Center for Scientific Collaboration and Community Engagement's (CSCCE) Community Participation Model (Woodley and Pratt, 2020), could also contribute to this role of coach. This person encourages, facilitates, and supports members' participation in the process. The CSCCE model also includes a "champion mode," which is used by a community member who temporarily complements the work of the community manager (Woodley and Pratt, 2020, pp. 5–6). In the context of MCM, this role corresponds to the "community champion," which we describe in more detail below.

Step 3: Providing adapted infrastructures, tools, and services

The standards selected or defined and developed in the previous step can now be implemented within the community. This should be accompanied by the provision of appropriate infrastructure and tools to put the use of standards into practice. Services to researchers are essential in this step, where researchers are expected to change their behavior and adopt new practices.

Step 4: Monitoring, providing maintenance and incentives

The community is now fully established. It is operational in terms of ORD practices, with a very wide use of the standards. In the same way as the benefits of ORD practices were highlighted in the first step, incentives are provided here to stimulate researchers' motivation to use good ORD practices. They may include or relate to policies, rewards, financial incentives, recognition in academic career evaluation and academic hiring, and support for "community champions."

To ensure that the new community is sustainable over the long term, it is important to monitor its practices. Significant changes may occur, such as the use of new standards, the use of standards other than those agreed upon in Step 2, or a general decline in the use of standards. If such situations are observed, short-term "maintenance" may be provided, such as redefinition of standards or some timely and specific training. It is also possible to re-enter the process at a previous stage (e.g., Step 2 for the development of standards).

Some external and general changes coming from the research environment or the main stakeholders of the research domain may also imply to start the process all over again from the beginning (Step 1).

Steps separated by evaluation boundaries

At several stages in the process, assessing the ORD maturity level of the community is critical.

Before beginning the entire process, assessing this level for a particular community allows us to identify the appropriate step at which that community should first enter the model. Between steps, the assessment provides the basis for deciding whether or not to proceed to the next step. At each step, it is important to identify and assess the appropriate level of contribution required, e.g., from long-term stakeholders in terms of governance and funding. More generally, this can provide useful information for planning and prioritizing the human and financial resources needed to support the development of the community in terms of ORD.

The question of how this evaluation should be carried out, according to what criteria and what metrics, remains to be specified. However, it could be based on elements that aim to characterize the ORD practices of researchers, such as those used and presented in Bongi et al. (2021, p. 39), for example, the level of knowledge of standards, the extent of use of these standards, and the skills and the level of autonomy in performing other activities related to ORD. These other activities may include the following: writing a data management plan, preparing data for uploading to a repository, uploading and sharing data via a repository, searching for and reusing data produced by other researchers, and reproducing research results produced by others. For each task, researchers could then self-assess their level of skill and autonomy by rating themselves: "I do not know how to do this"; "I am able to do it with some support"; "I am able to do it myself without any help."

Key actors

Throughout the process, at specific steps or in a less timely manner, several stakeholders play a critical role in supporting the development and strengthening of the community maturity

level. We present here the main features of their involvement as first outlined in Bongi et al. (2021, Section II. A.).

Research communities. Research communities obviously play a central role in the process. In particular, they should drive the process once it has been initiated and keep it active in the long term. In practice, they can contribute to raising awareness among researchers, reporting on current practices, and gathering needs. These tasks could rely on community members who are particularly advanced in ORD culture and active in ORD practices and who are likely to bring together other researchers. The creation of a formalized network of these community champions (analogous to the Data Champions; EPFL, 2023) would give them visibility and increase their positive impact on other researchers through peer stimulation.

Research institutions. Research institutions, through their institutional policies, can push communities to strengthen their ORD culture and improve their ORD practices. We refer to Bryant et al. (2020) for a detailed description of the role and potential contribution to research support services of typical faculty departments such as research administration, library, IT services, faculty affairs and governance, and communications. Research institutions can recognize the effort, time and skills that researchers invest in ORD practices, e.g., by considering these aspects in the hiring and career development process. They formally commit to using these criteria by signing the Declaration on Research Assessment (DORA, n.d.), endorsing the Hong Kong Principles on Research Integrity (World Conferences on Research Integrity, n.d.; Moher et al., 2020), and joining the Coalition for Advancing Research Assessment (CoARA, 2022).

Funding agencies. Funding agencies provide financial support for data infrastructures and the necessary services associated with these infrastructures. They may also support the training of researchers to improve their RDM skills. By including in their calls for proposals some evaluation criteria based on ORD practices (and similar to some in DORA for research evaluation), they could also encourage the improvement of practices.

Data infrastructures. Data infrastructures can support better ORD practices through the services they provide to researchers throughout the data life cycle. By using specific and adapted standards (and by requiring infrastructure users to do so), they also contribute to the promotion of these standards.

Data librarians. As Bryant et al. (2020) point out, the academic library is "deeply embedded in all phases of the research life cycle." It is a key player in supporting research activities (Bryant et al., 2020, p. 11, p. 18, p. 24), and in some organizations it actively contributes to data management, although it may not be recognized as the primary actor in support services (Tenopir et al., 2020, p. 22; Bongi et al., 2021, p. 40, p. 44, p. 48). Data librarians are

involved in data-related ethical issues and are familiar with data integrity and data security issues. They have expertise in metadata, data licensing, and data discoverability (Rinehart, 2022). As part of the implementation of MCM, they could therefore contribute to the setting of standards (Step 2).

With an important role in education (Bryant et al., 2020, p. 18), data librarians can also contribute to raising awareness at the beginning of the process (Step 1) and to training researchers (Steps 2 and 4) to lay the foundation for data literacy. They can facilitate and coordinate community champions networks within communities.

Data librarians typically interact with different types of audiences. "Sitting at the intersection between groups" (Rinehart, 2022), they could then play an important role in the overall coordination of key actors in the MCM process.

Other elements and concepts of the model

Overall and long-term support by means of governance and funding. In addition to the support means and tools mentioned earlier at specific steps of the model (awareness raising, coaching, monitoring, maintenance; infrastructure, services, tools, incentives, training), governance and financing contribute throughout the process, but with varying intensity. The appropriate level of governance and funding at each step should be derived from the evaluation results. Funding is likely to be needed initially on an ad hoc basis, and then in a more continuous and sustainable way.

Maturity level and its four components. All steps contribute to the improvement of the community's ORD maturity, but each step focuses on a more specific aspect: increasing knowledge, improving skills and practices, and achieving autonomy. However, these components are not strengthened or improved in a completely separate and sequential way but rather in an interrelated and overlapping way.

Twofold implementation strategy

Based on the various aspects discussed earlier, such as the involvement of the community to initiate and drive the process and the need for governance, we recommend a twofold strategy for implementing the process:

- Top-down: since governance and organizational and financial support will create the framework for developing the community and defining the standards.

- Bottom-up: because the involvement and engagement of community members as early as possible in the process will greatly contribute to the adoption of the appropriate standards and to their use in the long term.

HUMAN ASPECTS AND CULTURE CHANGE

So far, we have assumed that the implementation of MCM would follow the same path for the whole community, following fixed sequences (although it is always possible to go back or re-enter a previous step, as discussed in the description of Step 4). However, the "human aspects" inherent in the community (researchers) and other key actors can add unpredictable elements to the process, making it more dynamic and fluid.

This issue is well addressed by the CSCCE model (Woodley and Pratt, 2020). This model conceptualizes member engagement and information flow in STEM communities. Similar to MCM, the CSCCE model is organized into four sequential modes of engagement: convey/consume, contribute, collaborate, and co-create. It also includes a "champion" mode. This mode is used by a community member who contributes to all four other modes at a more leadership level, both within and outside the community (Woodley and Pratt, 2020, pp. 5–6). According to the authors of the CSCCE model, multiple modes can exist in the same community, with members using more than one mode and subsets of members using different modes of participation than other subsets (Woodley and Pratt, 2020, pp. 9–10).

The communities to which MCM will be applied are heterogeneous. At any given time, they are likely to include individuals with different levels of ORD maturity, different skills and resources, and perhaps different interests and goals. Therefore, at any given time, some subsets of the community will be using different steps. They may then move from one step to another at a different time than the rest of the community. The composition of the community may also change over time, as members join and others leave.

In addition to the heterogeneity of the community, characteristics of the academic environment can influence the dynamics of the process. Universities have been described by William B. Rouse as "complex adaptive systems" (Rouse, 2016, as cited in Bryant et al., 2020, p. 4). The behavior of individuals in this ecosystem can then be "random and chaotic" and independent. Individuals can organize themselves into self-organizing groups with different interests and needs (Bryant et al., 2020, p. 5). Although the communities relevant to MCM do not strictly correspond to groups or structures within a particular academic environment, their members share the same professional cultural background and are therefore likely to exhibit the same types of behaviors.

At this stage, it is very difficult to predict and assess the extent to which these human aspects may affect the ability of MCM to support the community in achieving full ORD maturity. This will certainly be an important objective of the validation task.

The culture change that underpins the entire process is also related to the human aspect at the collective level. As the community is initiated and organized, its (new) members are coached and trained to change their practices or adopt new ones. In this article, we mention some elements that facilitate this culture change in the context of MCM (education, coaching, training). In order to put this model into practice and apply it more successfully, the aspect of culture change must be considered in depth and in its entirety. Interesting inputs could be derived, for example, from the work of Nosek (2018). The latter discusses and suggests what can be done to stimulate the change in research toward open science and at which stage of the process researchers are most likely to join. In their report on social interoperability in research support for building cross-campus relationships, the authors recommend consulting "early and often with other stakeholders," sharing "ideas and drafts early in the process," and collecting "preliminary feedback from stakeholders" (Bryant et al., 2020, p. 32). This is the approach we have taken for MCM (see description of Step 1 and Bottom-up implementation earlier).

COMPARISON WITH THE DATA CONTINUUM MODEL (TRELOAR ET AL., 2007; TRELOAR AND KLUMP, 2019)

Although some of the features of MCM were inspired by DCC, the two models differ in a number of ways. In order to emphasize the complementarity of the two models and to clarify their specific areas of application in the broad field of RDM, we now discuss their similarities and their differences.

1. Model elements

Both DCC and MCM conceptualize a process or an approach aimed at sharing and opening research data (at least in the first version of DCC). The conceptual elements, the means, the methods, and the tools at play in the two models are those of RDM and ORD: storage and sharing of research data, description of data with metadata, standards to be respected, infrastructures, and services.

2. Scale of the process

In its 2019 version, the scope of DCC extends to all research objects. Its goal is to make research objects as FAIR as possible at the scale of a research project (Treloar and Klump,

2019, p. 11). The goal of MCM is to prepare communities and support them in opening their research data.

This difference has implications for the timeframe of the process. In the case of DCC, the time scale is that of the research projects, which is known in advance. In the case of MCM, the evolution of the process is paced by the evolution of the research community. This is less tangible and more difficult to anticipate and therefore requires regular monitoring.

3. Process flow

Both models show a similarity in the process flow itself. They include successive domains or steps, separated by boundaries. The transition to the next step is triggered when certain conditions are met or when certain tasks are performed. However, there are some differences.

In the DCC model, the boundaries correspond to the separation between domains, the collaboration domain and the publication domain. The move to the next domain is triggered by the progress of the research (e.g., the maturity of the research data in terms of readiness for sharing). This transition is the responsibility of the research team leader, provided that certain tasks have been completed: selecting objects, migrating or publishing them, adding sharing or public context, migrating and extending or publishing provenance, and documenting selection decisions. These tasks can be performed by the research team, a data specialist, a data curator, or the IT support team (Treloar and Klump, 2019, figure 1).

In the MCM model, steps are separated by evaluation boundaries. Moving to the next step is triggered by the output of this evaluation and is the responsibility of the governance representative. It implies a mix of discrete and specific contributions to each step (defining standards, providing infrastructure, etc.) and more continuous contributions (governance, funding, etc.).

It is possible to enter the process at any stage, depending on the initial maturity of the community under consideration. It is also possible to go back if the monitoring reveals some important changes or needs. On the other hand, the DCC process is linear, from the private domain to the publication domain, via the collaboration domain.

4. Domains and steps

The domains and steps of each model timely coincide and complement each other in the general process of adopting and generalizing research data practices (Figure 2 in this article; figure 2 of Treloar et al., 2007; and figure 1 of Treloar and Klump, 2019):

1. Only raw data is available, and it is kept in a small circle (private research domain, DCC). There is no community, or if there is, it is disorganized (Step 1, MCM).
2. Data is shared outside the core team (collaboration domain, DCC). A community builds, develops, and begins to use standards (Steps 2 and 3, MCM). Data is more structured, standards are applied, and collaboration takes place on a larger scale.
3. Data is publicly shared (publication domain, DCC), communities are fully established, and they have reached a high level of maturity in their ORD practices (Step 4, MCM).

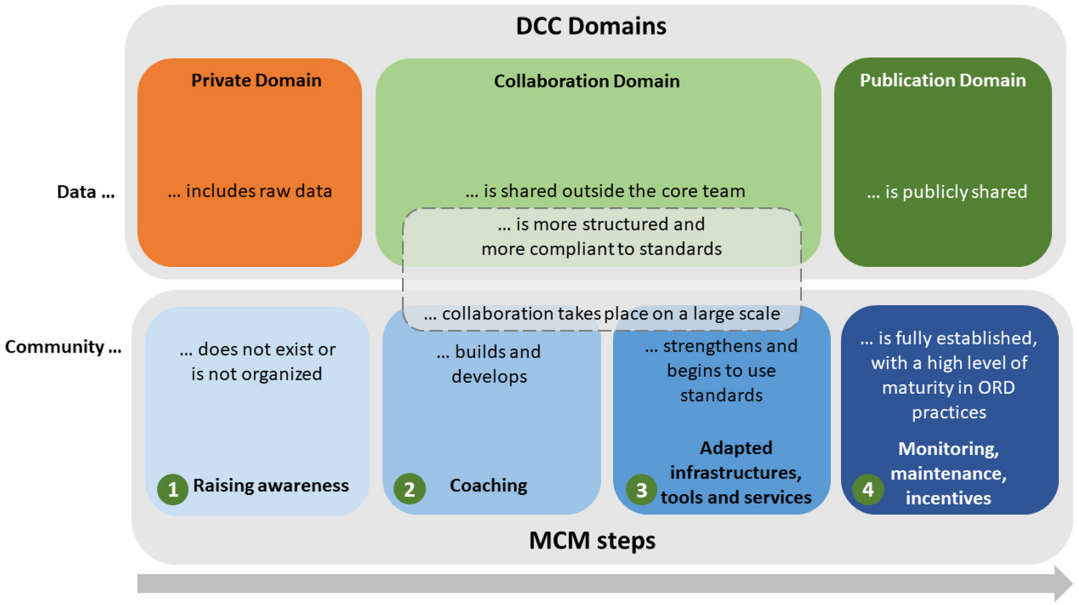

Figure 2. DCC domains (Treloar and Klump, 2019) (top row) and MCM steps (bottom row): how they align and complement each other, from a data perspective (DCC) and from a community perspective (MCM).

5. Stakeholders and key actors

The stakeholders of DCC are mainly data managers, data stewards and data curators or, more generally, specialists who contribute to the curation of data at any stage of the continuum: researchers, records managers, archivists, IT staff, data librarians. The stakeholders of the MCM model, external or internal to the community, their potential contribution, and the timing of their contribution to the process have been precisely specified and presented in an earlier part of this article (see the "Key Actors" section). These stakeholders are diverse, and so are their contributions. These contributions may be long term and continuous in time (governance), long term and discrete in time (funding in the first part of the process), or limited in time.

6. Standards

Standards are obviously key elements in this area of ORD. In the DCC model, standards are used as a tool to make research data (or research objects) FAIR. In the MCM model, the way and extent to which standards are used is an indicator of the maturity of the community with respect to ORD. They are also used as leverage to develop this maturity level.

7. Model validation and evolution

MCM was initially conceived primarily as a useful tool for organizing the elements and concepts collected during the survey on ORD practices and needs of Swiss researchers (Bongi et al., 2021). In the future, it could also be a useful tool for swissuniversities to fulfill its mission with regard to open science and ORD (swissuniversities, 2023). In particular, it would provide a common framework and language for the development of ORD in the Swiss academic environment. We can also assume that, as with DCC, it may prove useful to initiate discussions between stakeholders to clarify their roles and contributions to the process (Treloar and Klump, 2019, p. 95).

MCM has only recently been developed and has not yet been validated or applied. We recognize that, similar to DCC, MCM must first demonstrate its relevance and applicability. In particular, a critical aspect of validation relates to MCM's ability to handle the human aspect of the process, as discussed earlier. Its full value will be confirmed if and when it is widely adopted and successfully applied by key stakeholders.

After its first version in 2007 (Treloar et al., 2007), an updated version of DCC was published in 2019. This new version reflected, among other things, the "increased automation in the research process" (Treloar and Klump, 2019, p. 94), which, according to its authors, made it necessary to modify the model. On the contrary, this aspect, should it continue in the future, has little impact on the way MCM should be implemented and handled. In fact, the five layers of the DCC updated version (namely object, storage, context, provenance, and archival layers; figure 4 from Treloar and Klump, 2019), while very helpful for the curatorial care of research data, play a minor role in our context (with the exception of the context and provenance layers, the latter being a different kind of context). It might be interesting to further investigate the importance of context capture, especially before and during the boundary transition.

Similarly, the Digital Curation Lifecycle (Oliver and Harvey, 2016), with its strong orientation toward curation and archiving, and "The Wheel of Fortune: A 'Cosmic' View of the Repositories Space" (Blinco and McLean, 2004, as cited in Treloar et al., 2007), with its faceted view of repositories, were not considered stereotypes or clichés for our approach.

In any case, MCM should be kept flexible to allow for possible future changes in the environment and possibly refined or improved depending on the feedback from the validation activities.

APPLICATIONS AND PERSPECTIVES

In addition to serving the main purpose for which it was originally designed (e.g., strengthening the maturity level of research communities with respect to ORD), this tool can be used for several more specific and isolated applications. These applications also serve the same overall goal of generalizing and improving ORD practices.

The MCM can be applied at the level of a specific community to

- evaluate its level of maturity by assessing its members' use of standards
- increase its level of maturity
- monitor its maturity level over time, and
- manage and facilitate ORD culture change

until full ORD literacy is achieved.

It can be used by any community, whatever its level of maturity, since, as mentioned earlier, the model can be entered at any stage.

As well as contributing to the development and improvement of ORD practices on a community-by-community basis, this model can also be applied to several communities at once. It is recognized that there is a great diversity in the level of maturity of research communities (disciplines) with respect to ORD. Some are very advanced (Berman et al., 2013), and others are very behind, with the entire spectrum between these extremes (Bongi et al., 2021). By providing an accurate picture of the maturity of specific communities, this tool allows comparison of communities (disciplines), their sharing culture, and their practices.

When applied to more advanced communities, the model can potentially highlight the success factors and characteristics for ORD practices. These, in turn, could be transposed and applied to less advanced communities to support their progress toward improved sharing culture and ORD practices. Applying the tool in this way will help to rationalize efforts and resources by transferring successful practices to a whole set of communities, rather than applying them to one community at a time.

CONCLUSIONS

In this article, we presented our model, called the MCM, which conceptualizes the maturity level of a research community in terms of ORD. This continuum is segmented into four successive stages separated by assessment boundaries, which we described in detail. We identified the main stakeholders and key actors to be involved in the process, and we specified their contribution. We recommended the best implementation strategy for efficient use of the model. We highlighted the similarities and differences between our model and the Data Curation Continuum Model that inspired it (Treloar et al., 2007; Treloar and Klump, 2019), and how these two models can be considered as complementary tools when it comes to RDM in a broad sense.

We reiterate that, from our perspective, this model can be applied at different levels and scales. It can be applied in only one or a few of its steps, possibly in a non-linear way. It can also be used at the scale of one community or across multiple communities to compare and rationalize resources.

We have identified several aspects of the model where it could benefit from further work. The criteria and metrics for conducting the community assessment before entering the process or between steps need to be further specified. Also, a better understanding of how best to support the cultural change of the community that underlies the whole process would certainly increase its effectiveness.

As with any new tool, this model needs to be validated for its various applications. Initial tests need to be carried out to confirm that at least some of its aspects or components are usable, and to highlight the need for adjustments or improvements for others. However, the non-linear and rather general design of the model should allow it to accommodate these adaptations and to take into account possible future external changes in the ORD landscape.

Overall, the model could also be helpful in finding a common language, sharing conceptualizations on a difficult subject, and providing a reference for discussion and implementation of other initiatives in the broad spectrum of ORD management and open science.

ACKNOWLEDGEMENTS

The authors wish to thank Gaia Bongi and Elise Point for their contribution to the conception and the graphic design of the model.

REFERENCES

Armstrong, M., Davis, M., Dworak, E., Paul, Y., and Shook, E. (2021). *Research data management competencies self-assessment* [unpublished paper]. Boise State University Data Management Services. Paper 8. https://doi.org/10.18122/dataservices.8.boisestate

Berman, H. M., Kleywegt, G. J., Nakamura, H., and Merkley J. L. (2013). How community has shaped the Protein Data Bank. *Structure*, *21*(9), 1485–1491. https://doi.org/10.1016/j.str.2013.07.010

Bongi, G., Guirlet, M., Pelletier, E., Urvoy, G., Schneider, R. (2021). ORD capability of scientific communities - Mandate 2 commissioned by swissuniversities - carried out by HEG, March-October 2021, version V2. https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Hochschulpolitik/ORD/ORD_Mandate2_FinalReport_Master_VF_EN.pdf

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT Press. ISBN: 9780262327862

Bryant, R., Dortmund, A., and Lavoie, B. (2020). Social interoperability in research support: Cross-campus partnerships and the university research enterprise. *OCLC Research.* https://doi.org/10.25333/wyrd-n586

CoARA. (2022). Coalition for Advancing Research Assessment. https://coara.eu/

Cooper, D., and Springer, R. (2019). Data communities: A new model for supporting STEM data sharing. *Ithaka S+R.* https://doi.org/10.18665/sr.311396

DORA. (n.d.). San Francisco Declaration on Research Assessment. https://sfdora.org/read/

EPFL. (2023). EPFL Data Champions. https://www.epfl.ch/campus/library/services-researchers/data-services-expertise-tools-training/epfl-data-champions/#more

GEO (Group on Earth Observations). (n.d.). GEOSS Platform (former GCI). https://www.earthobservations.org/geoss.php

Gruber, A., Schranzhofer, H., Knopper, S., Stryeck, S., and Hasani-Mavriqi, I. (2021). «Kompetenzen von data stewards an österreichischen universitäten» [Competencies of data stewards at Austrian universities]. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare*, *74*(1), 12–32. https://doi.org/10.31263/voebm.v74i1.6255

Guirlet, M., Bongi, G., Point, E., Urvoy, G., Schneider, R. (2023). Proposal for a Maturity Continuum Model for Open Research Data. *International Journal of Digital Curation 2022*. Vol. *17*, no. 1, http://dx.doi.org/10.2218/ijdc.v17i1.821 http://www.ijdc.net/article/view/821

Moher, D., Bouter, L., Kleinert, S., Glasziou, P., Sham, M. H., Barbour, V., Coriat, A.-M., Foeger, N., and Dirnagl, U. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLOS Biology*, *18*, e300737. https://doi.org/10.1371/journal.pbio.3000737

National Snow and Ice Data Center. (2023). Welcome to the National Snow and Ice Data Center. https://nsidc.org/home

Nosek, B. (2018, May 2). *Culture change toward openness and reproducibility* [YouTube video of conference session]. The British Psychological Society Annual Conference, Nottingham. https://www.youtube.com/watch?v=q1JoKB7luPU

Oliver, G., and Harvey, R. (2016). *Digital curation*. 2nd ed. American Library Association. Item Number 978-0-8389-1385-7.

Power, J. (2021 October 4). Data communities. Linkedin.com. https://www.linkedin.com/pulse/data-communities-jerry-power

RDA. (2024). *How the RDA works*. Research Data Alliance. https://www.rd-alliance.org/how-the-rda-works/

Rinehart, A. (2022 December 6). Data communities: Room for everyone, roles for librarians [recorded lecture]. Joseph Leiter NLM/MLA Lectureship. *National Library of Medicine Technical Bulletin.* https://www.nlm.nih.gov/pubs/techbull/nd22/nd22_leiter_mla.html

Scholtens, S., Jetten, M., Böhmer, J., Staiger, C., Slouwerhof, I., van der Geest, M. and van Gelder, C. W. G. (2019). Towards FAIR data steward as profession for the lifesciences. Report of a ZonMw funded collaborative approach built on existing expertise. Zenodo. https://doi.org/10.5281/zenodo.3471708

swissuniversities. (2023). Open research data. https://www.swissuniversities.ch/en/topics/digitalisation/open-research-data

Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R., and Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS ONE*, *15*(3), e0229003. https://doi.org/10.1371/journal.pone.0229003

Treloar, A., and Klump, J. (2019). Updating the data curation continuum. *International Journal of Digital Curation*, *14*(1), 87–101. https://doi.org/10.2218/ijdc.v14i1.643

Treloar, A., Groenewegen, D., and Harboe-Ree, C. (2007 September/October). The data curation continuum: Managing data objects in institutional repositories. *D-Lib Magazine.* Vol. *13*, no. 9/10. https://doi.org/10.1045/september2007-treloar

Verheul, I., Imming, M., Ringerma, J., Mordant, A., van der Ploeg, J.-L., and Pronk, M. (2019, May 6). Data stewardship on the map: A study of tasks and roles in Dutch research institutes. *Zenodo.* https://doi.org/10.5281/zenodo.2669150

Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone,

M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Woodley, L., and Pratt, K. (2020). The CSCCE Community Participation Model—A framework to describe member engagement and information flow in STEM communities. *Zenodo.* https://doi.org/10.5281/zenodo.3997802

World Conferences on Research Integrity. (n.d.). Hong Kong Principles. https://www.wcrif.org/guidance/hong-kong-principles