IJDC | *Brief Report*

Proposal for a Maturity Continuum Model for Open Research Data

Marielle Guirlet
HEG-GE // HESSO

Gaia Bongi
HEG-GE // HESSO

Elise Point
HEG-GE // HESSO

Grégoire Urvoy
HEG-GE // HESSO

René Schneider
HEG-GE // HESSO

Abstract

As a contribution to the general effort in research to generalize and improve the practices of Open Research Data (ORD), we developed a model conceptualizing the degrees of maturity of a research community in terms of ORD. This model may be used to assess the ORD capacity or maturity level of a specific research community, to strengthen the use of standards with respect to ORD within this community, and to increase its ORD maturity level.

We present the background and our motivations for developing such an instrument as well as the reasoning leading to its design. We present its elements in detail and discuss possible applications.

Submitted 12 June 2022 ~ *Accepted* 13 June 2022

Correspondence should be addressed to Marielle Guirlet, 17, rue de la Tambourine, CH-1227 Carouge. Email: marielle.guirlet@hesge.ch

This article is based on a lightning talk at International Digital Curation Conference IDCC22, Edinburgh, 13-16th June 2022

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Context and Motivations

We were commissioned in 2021 by swissuniversities (the umbrella organization of the Swiss universities) to investigate the ORD capability (or maturity level) of academic scientific communities. This maturity level is to be understood as an indicator of advancement in terms of ORD culture and practices. The results of this study are expected to contribute to the next ORD strategy by swissuniversities, with the underlying aim to maximize the efficiency of support brought to the research communities in order to improve their ORD practices and culture.

To address this, we also focused on some more specific objectives, such as to:

- identify and characterize the needs of communities
- specify possible key actors and means to address those needs
- make some recommendations to rationalize the efforts in bringing support to communities

The data collected during the course of this study as well as some outputs from their analysis led us to conceptualize and develop an ORD maturity model. We hereby present the elements and the possible applications of this model.

Approach

According to Cooper and Springer (2019, reporting for STEM research), successful data sharing happens within data communities. They define those data communities as "formal or informal groups of scholars who share a certain type of data with each other, regardlesss of disciplinary boundaries". The same authors, as well as Berman et al. (2013), recognize that the environment of a research community stimulates the ORD practices of its members.

Consistently, based on a massive on-line survey addressed to all Swiss academic researchers, we stressed with our study that ORD practices of data community members are more advanced than the ones of non-members (Bongi et al., 2021).

Our results also highlighted that standards or norms in ORD (such as file formats, metadata schemas, etc.) are more widely used (about three times more) within data communities. We then recommend to facilitate the creation of research communities and to empower them, and to support the definition and the adoption of standards within the communities.

Going one step further, we conceptualized these elements and this reasoning by designing a Maturity Continuum Model. The development and the specifications of this model were partly inspired by the Data Curation Continuum (Treloar et al. 2007, Treloar and Klump 2019).

Model description

The results of our study show that the ORD maturity level of the communities investigated is very diverse. Some of them are very much advanced and have been so for a long time. Some of

them are well behind, with limited data sharing practices. And others are barely organized, showing an even poorer data sharing culture.
This landscape would have been better reflected by a continuum between the most advanced communities and the less advanced ones. However, a continuum being difficult to handle as such, we segment it in four separate domains, each of them addressing a specific maturity level (Figure 1):

- 1: No community or disordered community
- 2: Building a community
- 3: Consolidating a community
- 4: Fully established community

Transition between the domains happens through impulses, services or funding, allowing to step from one domain to the other.

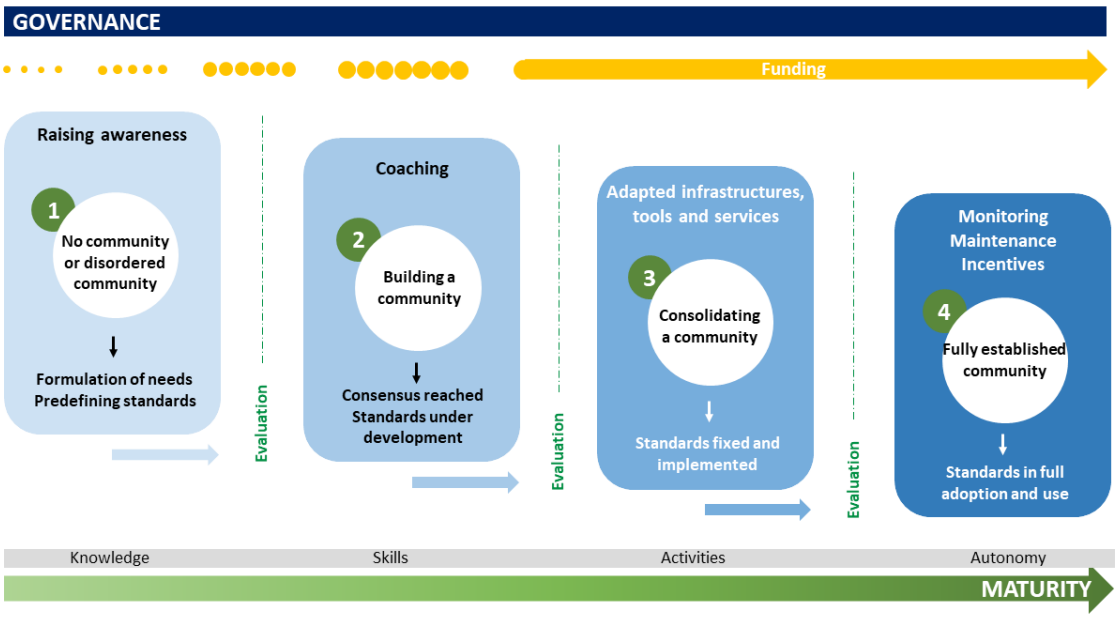

Figure 1: Maturity Continuum Model for Open Research Data

Domains

1: No community or disordered community: Raising awareness
This first domain may be seen as a 'nudge' one to trigger the whole process. At this stage, it is observed that there is no community or that the community is very much disorganized when it comes to manage and share its research data. The objective here is thus to support the creation and the organization of the community, or to re-organize it in a more structured way and to initiate a common culture of data sharing. It involves raising awareness and convincing of benefits brought by a data sharing culture (with (vocational) training). It may also involve identifying, observing, understanding and supporting nascent data sharing practices, even small-scale efforts if already existing.

Resources or tools needed for sharing more the research data and following better practices are formally identified, possibly from already expressed desires of researchers. This leads to the predefinition of standards and norms fit to the community. This is also the appropriate time for setting up a dialogue between the community researchers. By federating them and initiating a

common culture of data sharing among them, this dialogue should contribute to their long-term involvement and ensure a wider adoption and a more sustainable process.

2: Building a community: Coaching
The community is now building up, with support brought by means of coaching. This coaching focuses on the characterization of standards, with the aim of reaching a wide consensus within the community. This implies a continuous dialogue between the community researchers, a work done to lead to the development of standards.

3: Consolidating a community: Providing adapted infrastructures, tools and services
At this stage, a new community enters the process of consolidation. The standards defined and developed at the previous stage can now be fixed and implemented. This should be made along with providing appropriate infrastructures and tools, allowing to put into practice the use of standards. Services to researchers (information, training etc.) are essential here as researchers are expected to change their behaviour and to adopt new practices.

4: Fully established community: Monitoring, providing maintenance and incentives
The new community is now fully established and operational in terms of ORD practices, with a very wide use of the standards. In the same way as benefits of ORD practices were made visible to researchers in the first stage, relevant incentives are introduced or highlighted in order to stimulate them in using best practices. To make sure that the new community lasts on the long-term, monitoring its practices is needed. In case of significant changes, such as the decrease of use of already existing standards or the use of new ones, specific support may be provided (for instance the redefinition of standards, or some timely and specific training). A re-entry in the process in a previous domain is possible (the second one for instance, to develop more adequate standards). Some external and overall changes coming from the research environment or from the main stakeholders may also imply to consider reapplying the process.

The boundaries between the domains are actually evaluation steps. They are used to assess if the objectives of the current domain have been reached and if moving to the following one is appropriate. It is possible to enter or to re-enter the model at any stage.

All domains contribute to the improvement of ORD maturity level in general but each one has its own objective and focuses on a more specific aspect: knowledge, skills, activities until reaching complete autonomy with the last domain.

Standards and standardization

The whole process is underpinned by the standards and the standardization of practices related to data sharing. From their pre-definition with the first domain to the development, their fine-tuning, their implementation within the community, leading finally to the full adoption and the systematic use by the community members at the end of the process. The final effort focuses on the implementation of a systematic use of the standards implying an almost complete and unanimously accepted standardization inside the community.

The standards, or their level of adoption or use, are in the same time indicators of the maturity level of the community, and leverage tools for strengthening and improving the ORD practices and culture, helping as such to increase the community maturity level and allowing to step further to a higher level domain.

Other components

The other essential components of the model include governance and organizational support brought all along the process. Long-term funding is provided, first on an ad-hoc basis, then in a more continuous and sustainable way.

Other kinds of support include (some of them already mentioned) awareness-raising, coaching, monitoring, providing infrastructures, services, tools, incentives, and at several stages, (vocational) training. Details concerning the involvement of key actors and of their roles in specific domains of the process are given in Bongi et al. (2021).

Implementation strategy

Based on the results of our study, we recommend a twofold strategy:

- Top-down, with governance and organizational support to create the frame for the community to develop and for the standards to be defined.
- Bottom-up, since it is a community-driven process, involving the community members from as early as possible for a better adoption and use of the standards and a longer engagement in the process.

Applications and conclusions

The maturity level model may be applied at the scale of one specific community to:

- evaluate its maturity level by assessing the use of standards by its members
- develop, consolidate and monitor its maturity level on the long-term

It is adaptable to any community whatever its maturity level, since, as mentioned above, the model allows entering the process at any time.

Some elements may also be used separately and applied simultaneously to several communities to:

- compare the maturity level of those communities
- transpose success factors of more advanced communities to less advanced ones (standard practices)

It can also be seen as an instrument to manage and facilitate change for an improved ORD culture, leading to the accomplishment 'ORD literacy', and this way as a contribution to the global effort to generalize and improve ORD practices.

References

Berman, H. M., Kleywegt, G. J., Nakamura, H., & Markley, J. (2013). How Community Has Shaped the Protein Data Bank. *Structure, 21, 9, 1485–1491*. doi: 10.1016/j.str.2013.07.010

Bongi, G., Guirlet, M., Pelletier, E., Urvoy, G., & Schneider, R. (2021). *ORD capability of scientific communities.* Report of a mandate commissioned by swissuniversities carried out by HEG, March-October 2021. Retrieved from https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Hochschulpolitik/ORD/ORD_Mandate2_FinalReport_Master_VF_EN.pdf

Cooper, D., & Springer, R. (2019). Data Communities: A New Model for Supporting STEM Data Sharing. *Ithaka S+R*. Retrieved from http://sr.ithaka.org/?p=311396

Treloar, A., Groenewegen, & Harboe-Ree, C. (2007). The Data Curation Continuum: Managing Data Objects in Institutional Repositories. *D-Lib Magazine, 13, 9/10*. doi: 10.1045/september2007-treloar. Retrieved from http://www.dlib.org/dlib/september07/treloar/09treloar.html

Treloar, A. & Klump, J. (2019). Updating the Data Curation Continuum. *International Journal of Digital Curation*, 14, 1, p. 87-101. doi: 10.2218/ijdc.v14i1.643.