National Coordination Point
Research Data Management

Dutch Data Curation Network

LCRDM task group

DRAFT version, October 2019
https://doi.org.10.5281/zenodo.3466632

Report on the state of the art of data curation in the Netherlands and the feasibility of creating a dedicated Dutch Data Curation Network

Inge Slouwerhof (Radboud University), Mijke Jetten (Radboud University), Christina Elsenga (University of Groningen), Nynke de Groot (Erasmus University Rotterdam), Marjan Grootveld (DANS), Lena Karvovskaya (Utrecht University), Marcel Ras (Dutch Digital Heritage Network), Madeleine de Smaele (TUDelft), Richard Visscher (Inholland University of Applied Sciences), Boudewijn van den Berg (LCRDM), Ingeborg Verheul (LCRDM)

Content

1. Introduction

Triggered by coordinated data curation activities abroad, an LCRDM (National Coordination Point Research Data Management) task group[1] recently investigated the interest, necessity and feasibility of a Data Curation Network in the Netherlands. The aim was to find out whether data curation in the Netherlands could benefit from sharing expertise and experiences in a dedicated, lightweight professional network. It seemed that data curation processes in research institutions were not very well standardised yet and a certain degree of standardisation might enhance data curation as an important aspect of the research life cycle. To this end a dedicated network might be valuable.

The definition of "data curation" we agreed upon in the task group is as follows:

> *the activity of managing the use of data from its point of creation to ensure it is available for discovery and reuse in the future. Examples of data curation range from adding, verifying and improving metadata to checking if files open as expected and recording who did what with the dataset in a repository. Researchers, research support staff and repository staff carry out this kind of activity, in different phases of the research data life cycle.*

We specifically focused on the needs and practices of research support and repository staff, and start at the moment when the dataset is being prepared for publication and "something should be done with the data".

The task group concentrated on the following activities:

- Describe current data curation practices by means of the CURATE(D) model[2].
- Carry out a survey among research organisations involved in data curation in the Netherlands.

In this report, we present the outcomes of the task group: the overview of the current Dutch data curation practices, the survey outcome, and recommendations for next steps. The full overview of practices (matrix), the survey questions, and a basic reading guide on the topic of data curation can be found in the appendices.

2. Drivers for data curation

Following their ambitions towards Open Science, research institutes, journals and funders more and more require researchers to publish their data. Archives and repositories help not only to archive data but also to make data available for the long term: open when possible and restricted when necessary. The aim of data publication is to serve both reusability and research transparency. However, data without any context or documentation is of little value. Therefore, data publishing requires a clear process of data curation. Generally, curation is done by either by the researchers themselves, and/or by the research support offices of research organisations, or by external archival staff. The process of data curation therefore affects the daily practice of (data) scientists, data supporters (stewards, managers, librarians) and data archive staff.

The FAIR Guiding Principles[3] for scientific data management and stewardship offer basic criteria for data curation, such as the presence of rich metadata and persistent identifiers. So, first steps have surely been taken and the goals are set: Open and FAIR data. However, the FAIR principles are - by definition - principles and don't describe practice. The LCRDM task group aims to provide a picture of the current Dutch data curation practices. Is it uniform or does it show a great variety in the quality, structure, content, and context of data curation at the different data archives, universities (for applied sciences) and research organisations?

Based upon the experience and information already available in the US, initiated by the Data Curation Network Project[4] (DCN), the time seemed right to combine forces in the Netherlands and to investigate the possibility to initiate a similar project: the Dutch Data Curation Network.

3. The CURATE(D) model

3.1. The original DCN model

With the example of the US initiated Data Curation Network Project[5] in mind, and focused on joining forces and exploring the idea of a *Dutch* Data Curation Network, the task group used the CURATE(D)[2] model of the Data Curation Network as reference point.

On the Data Curation Network website, this model is described as follows: "the DCN developed a standardized set of C-U-R-A-T-E steps and checklists to ensure that all datasets submitted to the Network receive consistent treatment. The CURATE checklists[6] were drafted in the planning phase of the project (read the 2018 post[7]) and further enhanced by members of the DCN at the First Annual All Hands Meeting in July, 2018. These checklists are works in progress. The main goal for designing CURATE checklists was creation of training materials for the future curators".

The CURATE(D) acronym consists of seven “actions”. The D of CURATE(D) was added later, with particularly archives in mind, therefore the brackets in the acronym. For detailed information on the original Data Curation Network actions, check their project website.

- Action 1. **C**heck files and read documentation (risk mitigation, file inventory, appraisal/selection)
- Action 2. **U**nderstand the data (or try to), if not… (open files, run code/environment, quality assessment/quality control issues, readmes)
- Action 3. **R**equest missing information or changes (tracking provenance of any changes and why)
- Action 4. **A**ugment metadata for findability (DOIs, metadata standards, discoverability)
- Action 5. **T**ransform file formats for reuse (data preservation, conversion tools, data visualisation)
- Action 6. **E**valuate for FAIRness (transparent usage licenses, responsibility standards, metrics for tracking use)
- Action 7. **D**ocument all curation activities throughout the process

3.2. Adjusting the model to the Dutch context

The original model was slightly adjusted to the curation practices and needs in the Dutch community; however, all seven actions of the CURATE(D) acronym were kept intact, including their main content, structure and order. This process included consultation with DCN representatives, to ensure that the model was well understood.[*]

Originally, as cited above, the CURATE(D) model was designed as a training methodology for data curators. To be able to use it as an assessment model for Dutch data curation practices, the original model was adjusted:

- Adjustment 1. *Questions*: all *actions,* which originally had the form of statements, were reformulated into questions, to be able to actively disclose practices in a community.

[*] Teleconference with Lisa Johnston and Cynthia Vitale, US Data Curation Network.

- Adjustment 2. *From closed to open actions*: as we searched for information on *how* curation is incorporated in organisations, we preferred open questions to the closed, checkbox questions that were included in the original model. However, the content of the questions remained unchanged.
- *Simplifying the presentation*: all actions had a general description ("CURATE action") and a detailed checklist ("curator checklist"). To keep it simple, when drafting the model, the general description was left out. The detailed checklist seemed to be elaborate enough.
- *Deleting items*: as some of the items were unfamiliar to the task group or seemed not relevant to the Dutch context from the perspective of the task group, these items were left out. This concerned among others visualisation of data, preservation packages and repository collection metadata.

It needs to be emphasised that these changes were used for the purpose of the work of the current task group. For further use, however, it may be advisable to return to the original CURATE(D) model again.

3.3. Deliverable: matrix with Dutch best curation practices

After the CURATE(D) model was adjusted to the goals of the task group, it was used to create an overview of Dutch best curation practices, starting with the task group institutes: each task group member or another representative described his/her organisation's curation practice in terms of the model. This resulted in a matrix of CURATE(D) questions answered by ten organisations.

For a number of reasons, the matrix is rich and diverse:

- Some representatives answered the questionnaire from the perspective of their specific function, while others provided an overview of curation activities performed by their organisation in general.
- Not all representatives were familiar with the CURATE(D) model. By using it for assessment, (for which it was not originally designed), the adjusted model turned out to be multi-interpretable and opened possibilities for various types of answers.
- Not all representatives were data curators, which made it harder to interpret and answer questions about data curation. Related to this, some but not all task group members described their institutional practice with help from a local data curator.
- There was also a lot of diversity in the informativeness of the answers: some answers were very detailed and contained explanations, while other questions were only answered with a yes or no.

However, regardless of its multi-interpretable character, the matrix offers a rich overview of current data curation practices in Dutch organisations. The full matrix is included in Appendix A.

3.4. First analysis

Based on the matrix, the following analysis of Dutch data curation practices can be made:

- The matrix includes the practices of ten Dutch organisations. Some of those are research organisations, such as Radboud University[8], TU Delft[9], University of Groningen[10], Utrecht University[11], Inholland University of Applied Sciences[12] and the Meertens Institute[13]. Others are actual archives, such as 4TU.ResearchData[14], DANS[15], SURFsara[16] and YODA/Dataverse Utrecht[17]. Curation practices vary a lot among those Dutch organisations.
- This is explained by the level of maturity of data curation services, and the priority the process of data curation has within an organisation. Additionally, it also depends on the extent to which an organisation can rely on services offered by in-house or by external data

archives that do the job for them. DANS and SURFsara, for instance, host a data archive themselves, and the 4TU.ResearchData archive is an in-house service of among others TU Delft, while Radboud University closely cooperates with the DANS archive, and Utrecht University has its own archive YODA/Dataverse Utrecht.

- Another explanation is the difference between data curation as a central service or a decentralised effort made by local research communities. In the former situation the library, for instance, provides the curation; in the latter case there typically is a central data cataloguing service.

Regardless of how data curation is positioned within the organisational structure, the CURATE(D) model helps to show similarities in data curation processes across Dutch organisations:

- *Action 1. **C**heck files and documentation*: almost all organisations check the data files and the corresponding documentation in the data package. An exception is TU Delft, which delegates data curation to the 4TU.ResearchData archive.
- *Action 2. **U**nderstand the data*: in all organisations the main responsibility for the content of the dataset and the quality of the documentation remains with the researcher. Some organisations, like Radboud University, 4TU.ResearchData and DANS, check in detail the usability of the dataset and the quality of documentation. Others, like Groningen, Utrecht University and SURFSara, try to check the documentation, but also point out that domain-specific knowledge is not always available and the checks might have a somewhat sporadic nature. For some institutions, like Inholland University of Applied Sciences, these kinds of checks go beyond the scope of data support at the moment.
- *Action 3. **R**equest missing information*: communication with the researcher who deposits the data in the repository is seen as an essential part of the process by all organisations. The exact procedures differ though. For example, 4TU.ResearchData uses the front office team to communicate with the researcher. In some institutions, the researchers are only contacted by the curators if specific changes in the dataset need to be performed. The researchers may receive replies per e-mail, and at some institutions, the communication about the dataset might take place in person or per telephone. The institutions emphasise the importance of explaining why changes are necessary.
- *Action 4. **A**ugment metadata*: in most organisations, generic metadata schemes like Dublin Core and/or Datacite are used in data curation, while structuring and presenting metadata in a domain-specific format is often not part of the curation processes. The University of Groningen, Utrecht University and Dataverse Utrecht use domain-specific metadata in some cases.
- *Action 5. **T**ransform file formats*: advice on transferring data files into formats better suitable for reuse is not always part of data curation. Some organisations follow a list of preferred formats, while other institutions advise on using the preferred formats but don't insist on transformation.
- *Action 6. **E**valuate for FAIRness*: almost all organisations evaluate the dataset for compliance with the FAIR principles. Findability is seen as an essential part of the data curation. A lot of attention is paid to open access to the data.
- *Action 7: **D**ocument processes*: five organisations have an internal service workflow for the curation process (Radboud University, 4TU.ResearchData, University of Groningen, DANS and the Meertens Instituut), while others are working on developing these workflows.

4. Survey setup and findings

4.1. Survey setup

To investigate the idea of a Dutch Data Curation network, the task group set up a short survey. The survey ran between June 21 and July 17, 2019. It was promoted via the LCRDM website[18] and the Dutch RDM mailing list[19]. Members of the task group and subscribers to the mailing list have distributed the survey in their own networks.

No personal data were collected in the survey. The name and e-mail address of the task group chair were provided in case of questions; she received no questions or feedback.

The online survey was drafted in Qualtrics in both English and Dutch and contained five questions. See Appendix B for the complete survey text.

1. Are you involved in or working for an organisation (also) located in the Netherlands? [1. Yes; 2. No]
2. Is your organisation involved in data curation? [1. Yes; 2. No, but we have plans; 3. No, and no plans either]
3. What, in your experience, is the main data curation challenge? [free text]
4. A Dutch Data Curation Network would be useful to (…) [rank 8 options, including 8. Other ... (free text)]
5. The members of the LCRDM task group Dutch Data Curation Network started to describe their curation practices by means of a US data curation spreadsheet <link added>. [1. I will add my organisation to the spreadsheet; 2. The spreadsheet is not useful because … (free text)]

The survey included the definition of data curation, introduced in section 1 of this report:

> *the activity of managing the use of data from its point of creation to ensure it is available for discovery and reuse in the future. Examples of data curation range from adding, verifying and improving metadata to checking if files open as expected and recording who did what with the dataset in a repository. Researchers, research support staff and repository staff carry out this kind of activity, in different phases of the research data life cycle.*
>
> We specifically focus on the needs and practices of research support and repository staff, and start at the moment when the dataset is being prepared for publication and "something should be done with the data".

To those respondents who answered question 1 with "2. No", or who answered question 2 with "3. No, and no plans either", the subsequent questions weren't presented and the survey ended there. The task group assumed these respondents would not be interested in contributing to a potential Dutch Data Curation Network.

4.2. Main findings

The respondents expect that a data curation network primarily would be useful to reuse guidance that other organisations have made (e.g. how to's or instructions), to create such guidance together, and to define minimal good practices for data curation in the Netherlands. These activities emerged at the top of the ranking question (see question 4 below).

Respondents identified three main challenges in curation. First of all, building awareness and having a rewarding system in place, summarised into "what is it for me?". Designing proper and workable procedures, and setting quality standards came second and third.

4.3. Response

During the 27 days that the survey was open via Qualtrics, 98 respondents started the survey. 37 respondents participated in the English version, and 61 respondents in the Dutch version. Since the content of both versions was equal, and the language difference was only to facilitate respondents, we have combined the Dutch and the English responses in the analyses. As participants were (depending on their answers) introduced to subsequent questions, the response per question becomes smaller.

[Question 1] *Are you involved in or working for an organisation (also) located in the Netherlands?* [1. Yes; 2. No]
n = 98

Table 1. Working for an organisation in the Netherlands

	n	**%**
1. Yes	93	95%
2. No	5	5%
Total	**98**	**100%**

The task group assumed that respondents involved in a Dutch organisation would be more likely to be interested in getting involved in a Dutch Data Curation Network. To those respondents who selected "2. No", namely 5 respondents, the subsequent questions weren't presented and the survey was ended. This means that, out of a total of 98 respondents, for the purpose of a *Dutch* network 93 responses were relevant.

[Question 2] *Is your organisation involved in data curation?* [1. Yes; 2. No, but we have plans; 3. No, and no plans either]
n = 84

The task group wanted to find out which part of the respondents was already involved in data curation or was about to. For reasons unknown, 9 respondents didn't complete this question.

Table 2. Involvement in data curation

	n	**%**
1. Yes	51	54%
2. No, but we have plans	24	26%
3. No, and no plans either	9	10%
4. No answer	9	10%
Total	**93**	**100%**

The task group assumed that only those already involved in or those who have plans to get involved in data curation might be interested in contributing to a potential Dutch Data Curation Network (n = 75). To those respondents who selected "3. No, and no plans either", namely 9 respondents, the subsequent questions were not presented and the survey was ended.

[Question 3] *What, in your experience, is the main data curation challenge?* [free text]
n = 54

This was a free-text question on data curation challenges, intentionally presented before the question 4 ranking of benefits of a network, in order to collect as much information from the respondents as possible. The drawback of a free-text question typically is that clustering and analysing the answers is difficult, which in this case is increased by using multiple languages (English and Dutch).

The total response for this question is 54; however, the answers often include mentioning multiple challenges. In total, the task group identified 94 individual items, which could be grouped into nine main challenges. See Appendix C for the complete survey answers to question 3.

Table 3. Main data curation challenges

Challenge	**n**
What's in it for me	21
Procedure/workflow	20
Quality (for instance metadata)	16
Infrastructure and tools	11
Definition data curation	7
Resources	6
Data curation expertise/support	5
Standards	4
Answer is out of scope	*4*
Total	**94**

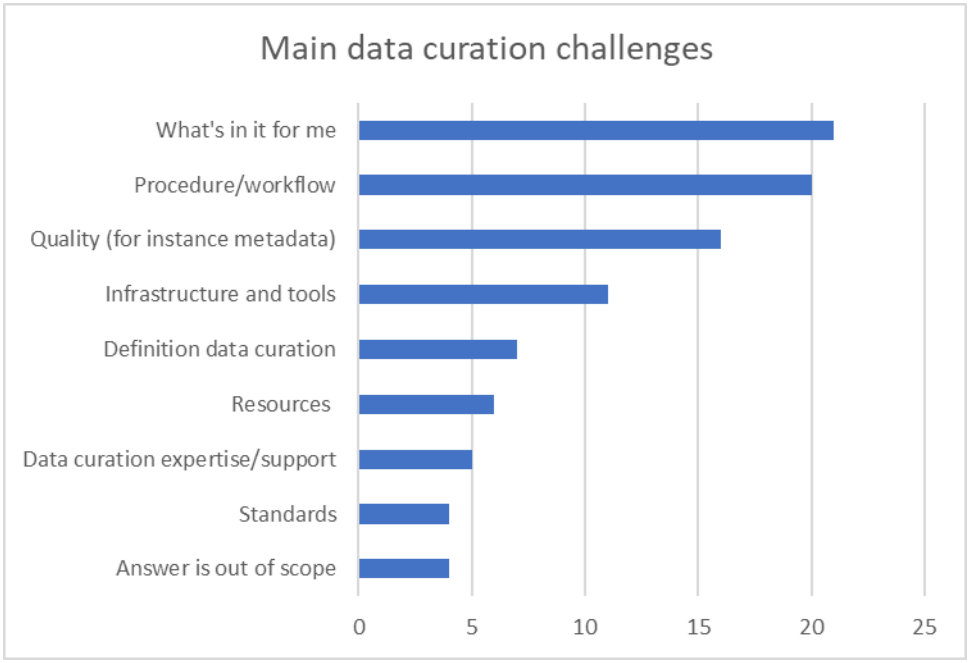

Figure 1 Clustered bar chart question 3

Challenges such as awareness of the organisation or researchers, incentives or rewards for researchers were all grouped under "what's in it for me", referring to the need for a "business case" for getting involved in data curation. Challenges concerning FAIR data and metadata were grouped under "quality". Four answers were considered related to research data management, but not typically to data curation, and are therefore considered out of scope of this project.

[Question 4] *A Dutch Data Curation Network would be useful to (…)* [rank 8 options, including 8. Other ... (free text)]
n = 52

Respondents were asked about their wishes and needs concerning a data curation network in the Netherlands. They had to rank eight activities by dragging and dropping (1 = most useful). They could also fill out optional wishes and needs via "Other … (free text)". None of the respondents added wishes or needs, so the task group assumes that the presented list of benefits is fairly complete.

Table 4. Ranking of benefits of a Dutch Data Curation Network

Benefits as ranked by the respondents	**Mean**
To reuse guidance (e.g. how to's or instructions) that other organisations have made	3.1
To create guidance (e.g. how to's or instructions) together	3.3
To define minimal good practices for data curation in The Netherlands	3.5
To compare our curation practice with others	4.4
To make data training for researchers more effective	4.4
To compare and discuss examples, e.g. of so-called "rich metadata" or "checking the data quality"	4.5
To learn what long-term data repositories like 4TU.ResearchData and DANS EASY offer and expect	5.7
Other … <free text>	7.1

Ranking: a lower mean implies a higher perceived benefit

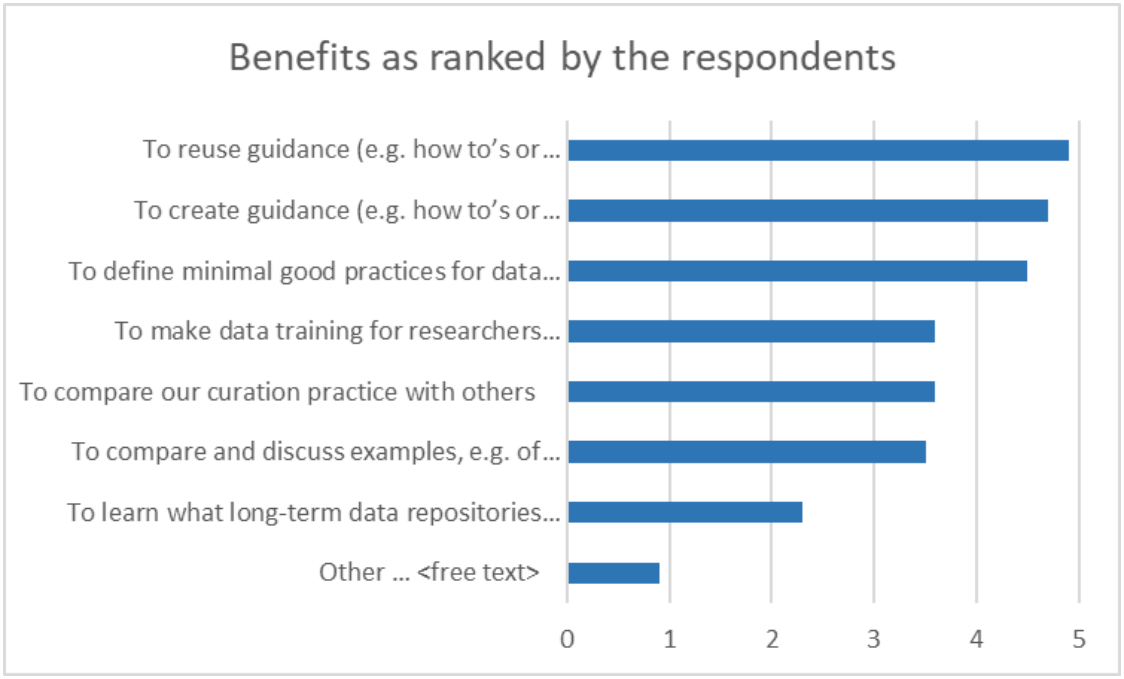

Figure 1 Clustered bar chart question 4
The ranking has been recoded. The highest perceived benefit has been given the highest value (mean scores)

[Question 5] *The members of the LCRDM task group Dutch Data Curation Network started to describe their curation practices by means of a US data curation spreadsheet <link added>*. [1. I will add my organisation to the spreadsheet; 2. The spreadsheet is not useful because … (free text)]
n = 48

See Appendix A for the spreadsheet that the question refers to. Despite answering that they would add their organisation to the spreadsheet, none of the respondents has actually done this. The task group is not sure why respondents didn't, but it may have to do with the external link to the spreadsheet, or that the spreadsheet required information that a respondent would have to sort out in their organisation, before being able to complete it.

18 respondents selected option 2 ("The spreadsheet is not useful because …") and gave the following explanation. See Appendix D for the complete survey answers to question 5.

Table 5. Explanation for not completing the spreadsheet/matrix

Category	**n**
I'm not the right person to fill this out	5
This spreadsheet comes too early for me/us	3
The spreadsheet is not relevant (enough)	3
The spreadsheet is too complex	2
I don't have time	2
Answer is out of scope	*3*
Total	**18**

5. Conclusion and recommendations

5.1. Conclusion

Triggered by coordinated data curation activities abroad, an LCRDM task group, focused at finding out whether data curation in the Netherlands could benefit from sharing expertise and experiences in a dedicated, lightweight professional network. The task group concentrated on the following activities:

- Describe current data curation practices by means of the CURATE(D) model.
- Carry out a survey among research organisations involved in data curation in the Netherlands.

The conclusions are fourfold:

1. *With minor adjustments, the CURATE(D) model proves to be useful as an assessment model for Dutch data curation practices.*

   The original CURATE(D) model is designed as a training methodology for data curators. To be able to use it as an assessment model for Dutch data curation practices, the model was slightly adjusted to the curation practices and needs in the Dutch community. However, all seven actions of the CURATE(D) acronym were kept intact, including their main content, structure and order. It needs to be emphasised that these changes were used for the purpose of the work of the current task group. For further use, however, it may be advisable to return to the original CURATE(D) model again.

2. *Having organisations describe their curation practices in terms of the (adjusted) CURATE(D) model results in a rich and diverse overview with Dutch curation practices, that can well serve as a 'good practice' or 'use case study'. However, it's a bit too early for standardisation of data curation practices in the Netherlands.*

   The ten Dutch organisations that together shaped the matrix give a *diverse picture* for many reasons: different perspectives (specific function versus the organisation in general), multi-interpretability of the CURATE(D) model (as a fairly new model), different backgrounds (not only data curators completed the matrix) and diversity in the informativeness of the answers (short versus detailed answers).

   At the same time, the picture of these ten organisations is *rich*: it shows that curation practices vary a lot, based on the level of maturity, the priority given to data curation, whether an organisation can rely on services offered by in-house or external data archives and whether data curation is a central or decentralised effort.

   For standardisation of data curation practices in the Netherlands, it seems to be too early, as the CURATE(D) model shows that many of the organisations are just starting to formalise their workflows and procedures for data curation.

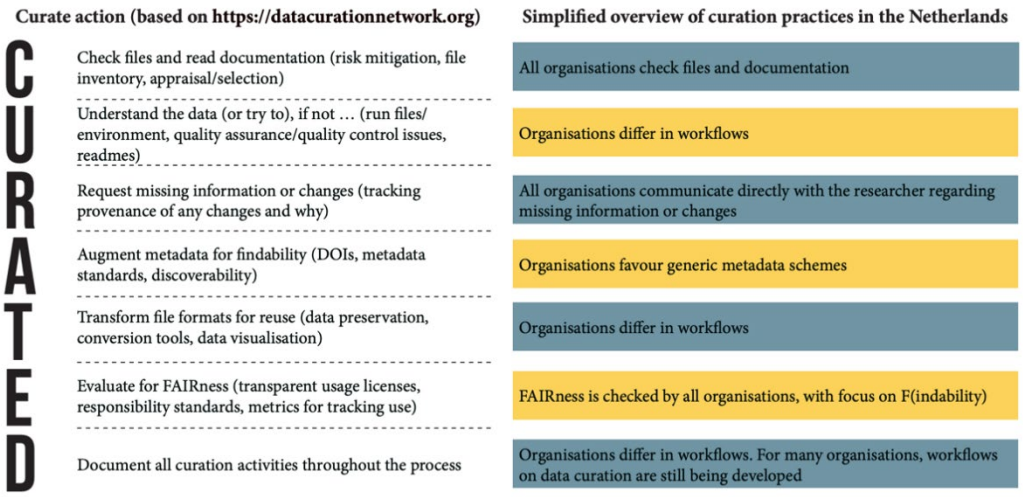

Figure 2 Simplified overview of data curation practices in the Netherlands based on the CURATE(D) model. "all" = 10

3. *Based on the survey, three main challenges in data curation in the Netherlands were identified: what's in it for me, workflows/procedures and quality of for instance metadata.*

   Based on the size of the nationally coordinated and broadly used mailing list in which the survey was distributed and the number of respondents of other surveys distributed in the same mailing list, with 98 respondents, the response rate is high.

   According to the Dutch community, the main challenges in data curation are (1) building awareness and having a rewarding system in place ("what is it for me?") (2) designing proper and workable procedures, and (3) setting quality standards.

4. *Based on the survey, creating a Dutch Data Curation Network would be beneficial for at least three reasons, namely reuse guidance, create guidance and define good practices.*

   According to the Dutch community, the main benefits of creating a data curation network are reusing guidance that other organisations have made (e.g. how to's or instructions), creating such guidance together, and defining minimal good practices for data curation in the Netherlands. Clear guidelines may be considered a prerequisite for benchmarking and training researchers, as these latter two benefits were considered less important.

5.2. Recommendations

With the work of the task group, the initial steps towards investigating the feasibility of a Dutch Data Curation Network have been taken. The recommendations of the task group can be grouped in two categories: on the one hand, recommendations for the national coordination of data curation practices in the Netherlands, and on the other hand, recommendations for individual Dutch organisations.

1. *Recommendations for the national coordination of data curation practices in the Netherlands:*

   - In the context of the National Coordination Point Research Data Management (LCRDM) that facilitates the current task group and is one of the main coordinating initiatives on

research data management in the Netherlands, the task group recommends to *set up a next LCRDM task group*.

- This task group should – like the current task group – encompass a broad spectrum of task group members, including repository curators and data stewards from various disciplines. They should represent research organisations, archives and possibly digital heritage organisations.
- The main task of this new task group on data curation practices in the Netherlands would be to set up an initial Dutch Data Curation Network. The current task group has explored the feasibility, challenges and the usefulness of a Dutch Data Curation Network. A next task group could capture what such a network in the Netherlands should do (building on the outcomes of the survey), which stakeholders and organisations should ideally be involved and what challenges should be addressed to start with. It should also be decided where to embed the Network, for instance under Research Data Netherlands, the Network Digital Heritage, or LCRDM.
- Another important task of this next task group would be to explore the application and use of the CURATE(D) model. Could it be used as a framework for training? Or for creating shared guidance? Or for standardisation of data curation practices in the Netherlands? From those perspectives, the CURATE(D) model seems to be really promising. A selection of other literature can be found in Appendix E.
- A final recommendation in the context of the national coordination of data curation practices in the Netherlands is to create an overview/page, based on the CURATE(D) matrix created in the current task group, of current curation practices in Dutch organisations, and make that overview/page available for the broad Dutch RDM community via the LCRDM website[20].

2. *Recommendations for individual organisations in the Netherlands:*

- The matrix delivered by this task group (see Appendix A) could serve as good practice for Dutch organisation to professionalise data curation practices, and to explain what data curation is about.
- The matrix can also be used as a benchmark to compare the data curation practices of the own organisation to these of other organisations in the Netherlands.
- We recommend individual organisations and their data supporters to exchange experiences, initiatives and efforts taken with regard to data curation.

Acknowledgements

The task group Dutch Data Curation Network thanks the following people for their support:

- Cynthia Hudson-Vitale, Pennsylvania State University, Data Curation Network
- Lisa Johnston, University of Minnesota Libraries, Data Curation Network
- Everyone who completed the matrix
- Everyone who participated in the survey

Questions regarding this task group and its deliverables can be asked via the task group's chair, Inge Slouwerhof, at i.slouwerhof@ubn.ru.nl.

References

1. https://www.edugroepen.nl/sites/RDM_platform/LCRDM/Wikipages/Taakgroep%20Data%20Curatie%20Netwerk.aspx
2. https://datacurationnetwork.org/resources/resources-2/
3. https://www.nature.com/articles/sdata201618
4. https://datacurationnetwork.org/
5. https://datacurationnetwork.org/about/
6. https://docs.google.com/document/d/1RWt2obXOOeJRRFmVo9VAkl4h41cL33Zm5YYny3hbPZ8/edit
7. https://datacurationnetwork.wordpress.com/2018/01/25/dcn-curation-checklist-and-fairness-scorecard/
8. https://www.ru.nl/english/
9. https://www.tudelft.nl/en/
10. https://www.rug.nl/?lang=en
11. https://www.uu.nl/en
12. https://www.inholland.nl/inhollandcom/
13. http://www.meertens.knaw.nl/cms/en/
14. https://researchdata.4tu.nl/en/
15. https://dans.knaw.nl/en/front-page?set_language=en
16. https://repository.surfsara.nl/
17. https://www.uu.nl/en/university-library/services/dataversenl
18. https://www.lcrdm.nl/en/questionnaire-dutch-data-curation-network
19. https://www.edugroepen.nl/sites/RDM_platform/netwerken/SitePages/On%20using%20the%20RDM%20mailing%20list.aspx
20. https://www.edugroepen.nl/sites/RDM_platform/Datamanagement/Wikipages/Home.aspx

Appendix A. CURATE(D) matrix: Dutch data curation practices

Click on the icon below to open the CURATE(D) matrix:

Dutch current practices.pdf

Appendix B. Survey

Question 1

Dear colleague,

The LCRDM task group Dutch Data Curation Network is eager to find out whether data curation in the Netherlands could benefit from sharing expertise and experiences in a dedicated, lightweight professional network. This survey was created to get input from relevant organisations involved in data curation in the Netherlands. We thank you in advance for participation in the survey.

Please pay attention: there are some links in this survey. You are advised to open them in a new tab!

Are you involved in or working for an organisation (also) located in the Netherlands?

- ◯ Yes
- ◯ No

Question 2

Is your organisation involved in data curation?

By data curation we mean the activity of managing the use of data from its point of creation to ensure it is available for discovery and reuse in the future. Examples of data curation range from adding, verifying and improving metadata to checking if files open as expected and recording who did what with the dataset in a repository. Researchers, research support staff and repository staff carry out this kind of activity, in different phases of the research data life cycle. In this task group, however, we specifically:

a) focus on the needs and practice of research support and repository staff and
b) start at the moment when the dataset is being prepared for publication and "something should be done with the data".

- ◯ Yes, my organisation is already involved in data curation
- ◯ No, but my organisation has plans to get involved in data curation
- ◯ No, and my organisation also has no plans to get involved in data curation

Question 3

What, in your experience, is the main data curation challenge?

Question 4

A Dutch Data Curation Network would be useful ... [use drag and drop to rank from most relevant to least relevant]

- To compare our curation practice with others
- To re-use guidance (e.g. how to's or instructions) that other organisations have made
- To create guidance (e.g. how to's or instructions) together
- To compare and discuss examples, e.g. of so-called "rich metadata" or "checking the data quality"
- To define minimal good practices for data curation in The Netherlands
- To learn what long-term data repositories like 4TU.ResearchData and DANS EASY offer and expect
- To make data training for researchers more effective
- Other

Question 5

The members of the LCRDM task group Dutch Data Curation Network started to describe their curation practices by means of a US data curation spreadsheet. Select one option

- ◯ I think the spreadsheet with Dutch curation practices is useful and I will use the link to add this kind of information for my organisation as soon as possible. If I have additional questions, I can contact Inge Slouwerhof via i.slouwerhof@ubn.ru.nl
- ◯ I don't think the spreadsheet with Dutch curation practices is useful, because

Appendix C. Survey answers to question 3

	Answer	**Challenge**
1	Define what kind of data curation activities fall within the field of data curation and development of tools for specific data curation activities	Definition data curation Infrastructure and tools
2	Sustainability of formats	Standards
3	There is no established procedure of data curation for researchers at my institution. If they want to share their data, they wouldn't even know that they could consult us on that. The only exception is data sharing at the university repository, this activity involves data curation. However even there there is no standardized procedure for quality control	Definition data curation Procedure/workflow Data curation expertise/ support
4	Incentives for researchers, disciplinary specific data sharing infrastructure	Infrastructure and tools
5	A clear definition to start with	Definition data curation
6	Lack of sustained funding for long term data curation. Lack of crediting system for scientists spending time on data curation	Resources What's in it for me Infrastructure and tools
7	1. Metadata: getting it clear and good enough. 2. Linking relevant material persistently	Quality (for instance metadata) Infrastructure and tools
8	How to avoid data curation to a large extent by making data FAIR at the source	Quality (for instance metadata) Procedure/workflow
9	There is no direct *reward* for data curation and there are no penalties involved when data curation does not happen. Researchers need to do the necessary steps during their research time; they procrastinate [stellen uit] those activities	What's in it for me Procedure/workflow
10	For experimental data, it is the richness of data (to allow reuse of data for different questions). For knowledge structures, it is convincing the right experts that it is OK to contribute	Quality (for instance metadata)
11	Making data F.A.I.R	Quality (for instance metadata)
12	Awareness. A lot of researchers and staff involved do not prioritize data curation because: - They do not know what data curation is - They do not see the potential of reusable data - They feel it is not worth the effort	Definition data curation What's in it for me
13	Lack of disciplinary expertise to review the data + researchers who don't want to be troubled with long discussions / going back and forth to improve their datasets. Researchers are advised to use institutional /national data repositories, instead of discipline-specific repositories which might be more suitable homes for their data	Data curation expertise/support Infrastructure and tools Procedure/workflow
14	Long term interoperability and disciplinary metadata standards lacking in many fields	Quality (for instance metadata) Standards
15	Having researchers practice good data management. We have resources to archive most research data output but getting researchers on board with best practices for them to do that is the main challenge	What's in it for me

16	Getting (senior) researchers educated in the FAIR-datacycle	Data curation expertise/support What's in it for me Quality (for instance metadata)
17	Raising awareness for services	Data curation expertise/support
18	Adding sufficient metadata for reusability, and persistent storage	Quality (for instance metadata)
19	2 main challenges: - finding resources to do it - figuring out what should be curated and what shouldn't (we can't curate everything!)	Resources Definition data curation Procedure/workflow
20	Het proces goed inrichten	Procedure/workflow
21	Open Access publiceren	*Answer is out of scope*
22	Geld besparen	Resources
23	Heb geen ervaring	*Answer is out of scope*
24	De bereidheid van onderzoekers om hun data te willen/kunnen delen	What's in it for me
25	Hoe zet je een werkbare en stabiele workflow op voor medewerkers en onderzoekers	Procedure/workflow
26	In onze organisatie wordt aan data curatie gedaan maar niet op grote schaal. Die taken liggen op dit moment op decentraal niveau, bij data- of lapmanagers die data invoeren in bijv. dataverse. We gaan op korte termijn wel meer data lokaal opslaan voor de langere termijn en zullen dan zeker meer met datacuratie te maken krijgen. Een van de uitdagingen zal zijn om voldoende informatie (metadata) over de datasets te krijgen en voldoende capaciteit om de datasets te beschrijven	Quality (for instance metadata) Resources
27	Inzicht krijgen in aanpak en bewustzijn creëren	Procedure/workflow What's in it for me
28	Bewustwording, organisatiebrede inrichting en professionalisering	Procedure/workflow
29	Data voor langdurig behoud opslaan	Procedure/workflow
30	Moeilijk te zeggen, uitdagingen liggen op vele verschillende vlakken (ook organisatorisch, bewustwording in omgang met data etc. kennis op peil houden)	Procedure/workflow Data curation expertise/support What's in it for me
31	De FAIR-principes concreet maken en naleven, met name de R van Reusable	Quality (for instance metadata)
32	Structuur van data opslag, beperktheid van opslagquotum	Infrastructure and tools
33	- Grootste uitdaging: Het belang van datacuratie en de daarbijbehorende verplichtingen (AVG, FAIR, DMP, etc) goed onder de aandacht brengen van het onderzoeksdomein vanuit het vertrekpunt 'verleiden in plaats van dwingen'. - In het verlengde daarvan is de grote uitdaging om dit zowel technisch (tooling, infrastructuur) als organisatorisch te regelen (invulling van research support in de brede zin van het woord; i.e. voor alle fases van het onderzoek (idea, preparation, conduct, closure) en specifiek in relatie tot het DataCuration Continuum model van Treloar	What's in it for me Infrastructure and tools Procedure/workflow
34	Om te zorgen dat alleen relevante data worden geselecteerd en gepresenteerd.	Quality (for instance metadata)

		Procedure/workflow
35	De juiste balans vinden tussen begrijpelijkheid van de data en de tijdsinvestering van de onderzoeker. (Het begrijpelijk en herbruikbaar maken van een dataset voor een ander vergt erg veel documentatie en dus tijd van de onderzoeker)	Quality (for instance metadata) Resources What's in it for me
36	Goede begrip van de achtergronden en keuzes bij dataverzamelingen om het nut van hergebruik te beoordelen. - het technisch gezien 'live' houden van data services	Definition data curation Infrastructure and tools What's in it for me
37	Datamanagement beleid concreet maken met de juiste service en faciliteiten	Infrastructure and tools Procedure/workflow
38	Dat er weinig passende repositories zijn voor medisch onderzoek. Eigen repository of aansluiten bij ene grotere? Lastig AVG: wanneer anoniem en mag wel gedeeld worden en wanneer niet	Infrastructure and tools Procedure/workflow
39	Veilige archivering van data waarbij de onderzoeker ook het vertrouwen heeft en de bereidheid om zijn data beschikbaar te stellen	Infrastructure and tools
40	Het erkennen dat datacuratie een taak is, die in de toekomst nodig is, is de eerste stap die onze organisatie moet nemen.	Definition data curation What's in it for me
41	Zorgvuldigheid in het proces. Alle lectoren/onderzoekers het belang van openheid hierin voorleggen Het op een goede manier opslaan Het (laten)invullen van metadata	Procedure/workflow
42	Data verzameld dusdanig opslaan dat deze voldoen aan FAIR. Maw FAIR data in FAIR repository voor die duur die verplicht is en voorzien van geode metadatering	Quality (for instance metadata)
43	"Rich" metadata genereren voor zoveel mogelijk datasets. D.w.z. Duidelijke beschrijvingen, exacte info over tijd en plaats, info over data gebruik, linken naar andere bronnen, workflow informatie, keywords met bijbehorende vocabulaires, etc. etc. Alle metadata moeten machine readable zijn	Quality (for instance metadata) Procedure/workflow What's in it for me
44	Iedereen binnen de organisatie op 1 lijn krijgen	*Answer is out of scope*
45	Een belangrijke uitdaging is om instellingsbreed processen/workflows voor datacuratie in te richten en in kaart te brengen. Daarnaast zou het goed zijn om te standaardiseren. Minimale eisen te stellen aan een dataset waarmee een kwaliteitsstandaard ontstaat die acceptabel is en voldoet aan de FAIR principles. Hierbij geldt de i van FAIR als grootste uitdaging	Procedure/workflow Quality (for instance metadata) Standards
46	Financien	Resources
47	Het staat hier nog in de kinderschoenen en het begint te komen	*Answer is out of scope*
48	Metadata in orde krijgen. Dit doen we door vooraf zo veel, correct en duidelijk mogelijk de meta-data te verzamelen	Quality (for instance metadata)
49	Bewustwording bij onderzoekers mbt belang van datacuratie wie gaat datacuratie uitvoeren /formatie	What's in it for me
50	Betrokkenen overtuigen van het belang hiervan, zodat zij tijd en energie hierin willen steken	What's in it for me
51	Inzicht krijgen van alle aanwezige data binnen de organisatie en het aan boord krijgen van "data-naieve" medewerkers	What's in it for me
52	Formaten en versiebeheer	Procedure/workflow
53	Response van onderzoekers krijgen	What's in it for me
54	Onderzoekers kwalitatief correcte meta-data laten opstellen	Quality (for instance metadata) Procedure/workflow

Appendix D. Survey answers to question 5

	Anwer	**Category**
1	Het biedt geen praktische handvatten waar ik iets mee kan	Not relevant (enough)
2	De matrix is waardevol, maar wij zitten nog in een pilotfase	Too early
3	Ik vind het wel nuttig, maar verricht op dit moment deze taken nog niet	Too early
4	Wel nuttig, maar geen tijd!	No time
5	Is niet aan mij om in te vullen	Not the right person
6	Het is onoverzichtelijk en ik begrijp het nut er niet helemaal van	Too complex
7	Ik vind het wel nuttig maar moet invullen coördineren met anderen	Not the right person
8	XXX	*Answer is out of scope*
9	Wij zijn zover nog niet	Too early
10	Ik weet dit niet	Not the right person
11	Ik niet de tijd heb om even snel te bekijken wat het is.	No time
12	I have not seen it	*Answer is out of scope*
13	See answer to previous question	*Answer is out of scope*
14	I don't know	Not the right person
15	I think that disciplinary practices (which are international) are much more relevant	Not relevant (enough)
16	It is barely navigable it needs to be presented in a different format so that it is readable	Too complex
17	It is not widely disseminated	Not relevant (enough)
18	I am struggling to see how I can make use it myself. I see the value for your organization however. Good stuff! But hard to exploit on my side	Not the right person

Appendix E. Literature reading guide

Introductory reading

- **Data curation definition**
  https://dictionary.casrai.org/data_curation

- **Digital humanities data curation**
  https://guide.dhcuration.org/
  Introductions to key topics, including annotated links to important standards, articles, projects

- **Leren preserveren**
  https://lerenpreserveren.nl/
  Dutch introduction course: the first steps towards sustainable storage, management and accessibility of digital heritage

Advanced reading

- **SPEC Kit 354: Data Curation**
  https://doi.org/10.29242/spec.354
  The SPEC kit from the Association of Research Libraries (USA) explores the infrastructure different institutions use for data curation, which data curation services are offered, who may use them, which disciplines demand curation services most, library staffing levels, policies and workflows, and the challenges of supporting these activities

- **Case study: the University of Glasgow's digital preservation journey 2017-2019**
  http://doi.org/10.1629/uksg.461

- **Research Data Curation Bibliography**
  http://digital-scholarship.org/rdcb/rdcb.htm
  The *Research Data Curation Bibliography* includes over 750 selected English-language articles, books, and technical reports that are useful in understanding the curation of digital research data in academic and other research institutions. It covers topics such as research data creation, acquisition, metadata, provenance, repositories, management, policies, support services, funding agency requirements, open access, peer review, publication, citation, sharing, reuse, and preservation