

Dutch Data Curation Network

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

LCRDM

The National Coordination Point Research Data Management (LCRDM) is a national network of experts on research data management (RDM) in the Netherlands. The LCRDM connects policy and daily practice. Within the LCRDM experts work together to put RDM topics on the agenda that ask for mutual national cooperation.



Colophon

Dutch Data Curation Network

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

PUBLICATION DATE | December 2019

DOI | 10.5281/zenodo.3557237

LCRDM Task Group 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 (TU Delft), Richard Visscher (Inholland University of Applied Sciences), Boudewijn van den Berg (LCRDM), Ingeborg Verheul (LCRDM)

DESIGN | Nina Noordzij, Collage, Grou

TRANSLATION | Gosse van der Leij

COPYRIGHT

all content published can be shared, giving appropriate credit



creativecommons.org/licenses/by/4.0

LCRDM



LCRDM supported by



Content

1. Introduction	5
2. Drivers for data curation	6
3. The CURATE(D) model	7
3.1. <i>The original DCN model</i>	7
3.2. <i>Adjusting the model to the Dutch context</i>	8
3.3. <i>Deliverable: matrix with Dutch best curation practices</i>	8
3.4. <i>First analysis</i>	9
4. Survey setup and findings	11
4.1. <i>Survey setup</i>	11
4.2. <i>Main findings</i>	12
4.3. <i>Response</i>	12
5. Conclusions and recommendations	19
5.1. <i>Conclusion</i>	19
5.2. <i>Recommendations</i>	21
<i>Acknowledgements</i>	22
Appendix A. CURATE(D) matrix: Dutch data curation practices	23
Appendix B. Survey	24
Appendix C. Survey answers to question 3	26
Appendix D. Survey answers to question 5	30
Appendix E. Literature reading guide	32
<i>Introductory reading</i>	32
<i>Advanced reading</i>	32





1 INTRODUCTION

Triggered by coordinated data curation activities abroad, an LCRDM (National Coordination Point Research Data Management) task group 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 appeared 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 could be valuable.

The task group agreed on the following definition of “data curation”:

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, starting from the moment 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](#).
- 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: an overview of the current Dutch data curation practices, the survey outcome, and recommendations for next steps. A 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

In their ambition to facilitate Open Science, research institutes, journals and research funding organisations increasingly 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 undertaken by the researchers themselves, and/or by the research support staff of research organisations, or by external archival staff. The process of data curation therefore affects the daily practice of (data) scientists, data support staff (stewards, managers, librarians) and data archive staff.

The [FAIR Guiding Principles](#) for scientific data management and stewardship offer basic criteria for data curation, such as the presence of rich metadata and persistent identifiers. So, the first steps have certainly been taken and goals have been 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](#) (DCN), the time seemed right to combine forces in the Netherlands for investigating the possibility of initiating 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](#) in mind, and focusing on joining forces with other institutes at a national level to explore the idea of a Dutch Data Curation Network, the task group used the [CURATE\(D\)](#) model of the Data Curation Network as reference point. (<https://datacurationnetwork.org>).

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](#) were drafted in the planning phase of the project ([read the 2018 post](#)) 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 to create training materials for 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 meet the curation practices and needs of the Dutch research 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.¹

Originally, as cited above, the 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, in order to actively disclose practices in a community.
- *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 remains 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:* because some of the items were unfamiliar to the task group or in their view seemed irrelevant to the Dutch context, they were omitted. 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 current task group’s work. 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 suit the goals of the task group, it was used to create an overview of Dutch best curation practices, starting with the institutes affiliated with the task group : each task group member or other 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.

¹ Teleconference with Lisa Johnston and Cynthia Vitale, US Data Curation Network.



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 included 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](#), [TU Delft](#), [University of Groningen](#), [Utrecht University](#), [Inholland University of Applied Sciences](#) and the [Meertens Institute](#). Others are actual archives, such as [4TU.ResearchData](#), [DANS](#), [SURFSara](#) and [YODA/Dataverse Utrecht](#). Curation practices vary widely among these 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. 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 their own data archive. The 4TU.ResearchData archive is an in-house service for among others TU Delft, while Radboud University closely cooperates with the DANS archive. Utrecht University has its own archive YODA/Dataverse Utrecht.
- Another explanation is the difference between data curation as a central service as opposed to a decentralised initiative set up by local research communities. In the former situation the library, for instance, is responsible for curation; in the latter case there is usually 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 among Dutch organisations:

- **Action 1. Check 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. Understand 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, make a detailed check of the usability of the dataset and the quality of documentation. Others, like the University of Groningen, Utrecht University and SURFSara, try to verify the documentation, but also point out that domain-specific knowledge is not always available and that 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 this current time.
- **Action 3. Request omitted information:** communication with the researcher who deposits the data in the repository is seen as an essential part of the process by all organisations although exact procedures differ. 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 made. The researchers may receive replies per e-mail while at some institutions, communication about a dataset might take place person to person or by telephone. Nonetheless all institutions emphasise the importance of explaining why changes are necessary.
- **Action 4. Augment 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 process. The University of Groningen, Utrecht University and Dataverse Utrecht use domain-specific metadata in some cases.
- **Action 5. Transform file formats:** advice on transferring data files into formats better suited to reuse is not always part of data curation. Some organisations stipulate a list of preferred formats, while other institutions advise on using certain preferred formats but don't insist on transformation.
- **Action 6. Evaluate for FAIRness:** almost all organisations evaluate a dataset for compliance with the FAIR principles. Findability is seen as an essential part of data curation. Open access to data is given considerable attention.
- **Action 7. Document 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 such 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 site](#) and the [Dutch RDM mailing list](#). Members of the task group and subscribers to the mailing list distributed the survey via 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 described their curation practices with the help 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 focus specifically on the needs and practices of research support- and repository staff, beginning at the moment when the dataset is being prepared for publication and “something should be done with the data”.

Those respondents who answered question 1 with “2. No”, or who answered question 2 with “3. No, and no plans either”, were not given any further questions to answer, thereby concluding the survey. 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 would be useful primarily to re-use guide lines established by other organisations (e.g. how to’s or instructions), to draw up such guide lines together, and to define basic good practices for data curation in the Netherlands. These actions were ranked most important (see question 4 below).

Respondents identified three main challenges facing data curation. First of all, building awareness and establishing a reward system that can be characterized as “what’s in it for me?”. Designing proper and workable procedures, and setting up quality standards came second and third.

4.3. Response

During the 27 days that the survey was open via Qualtrics, 98 respondents took part in the survey. 37 respondents completed the English version, and 61 filled-out the Dutch version. The content of both versions was identical; the language difference was only to facilitate respondents. We have therefore combined the Dutch and the English replies in the analyses. As participants progressed in filling out the survey and depending on their answers, were presented with subsequent questions, replies per question decrease.

[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 associated with a Dutch organisation would be more likely to be interested in participating in a Dutch Data Curation Network. Those 5 respondents who replied “2. No”, were not asked any further questions thereby concluding the survey. Therefore, out of a total of 98 participants, 93 responses were relevant for establishing a Dutch network.

[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 how many respondents were already engaged in data curation or had intentions to that end. For reasons unknown, 9 respondents failed to 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%

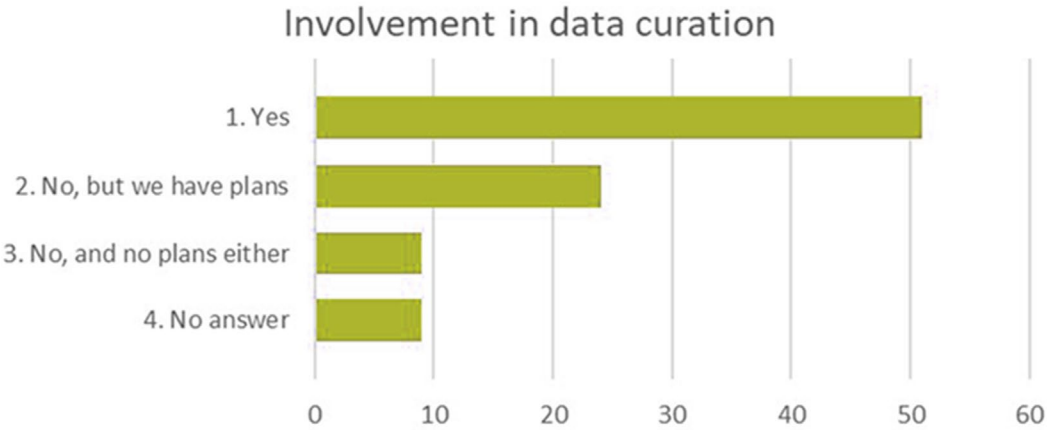


Figure 1 Clustered bar chart question 2

The task group assumed that only those already engaged in data curation or those who had plans to that end might be interested in contributing to a potential Dutch Data Curation Network (n = 75). Those respondents who selected “3. No, and no plans either”, namely 9 respondents, were asked no further questions, thereby concluding the survey.



[Question 3]

What, in your experience, is the main challenge facing data curation? [free text]

n = 54

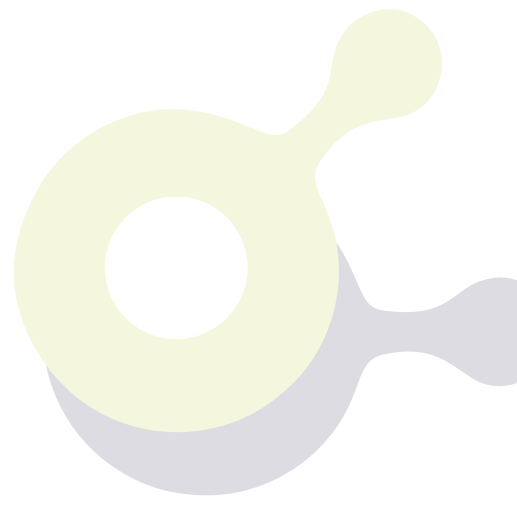
This was a free-text question about the challenges of data curation, intentionally inserted before question 4 which seeks to rank the benefits of a data curation network, in order to collect as much information from respondents as possible. The drawback of a free text question is typically that grouping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

The total response for this question was 54; however, many of the answers mention multiple challenges. In total, the task group identified 94 separate aspects, 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
<i>Answer is out of scope</i>	4
Total	94

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 to relate to research data management, but not necessarily to data curation, and are therefore considered beyond the scope of this project.



Main data curation challenges

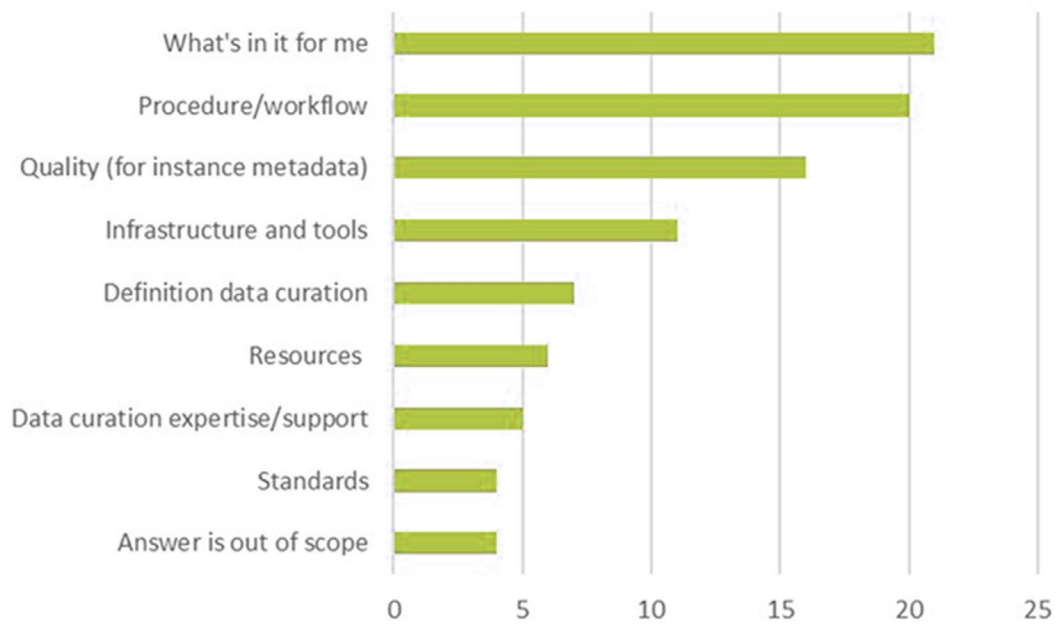


Figure 2 Clustered bar chart question 3

[Question 4]

A Dutch Data Curation Network would be useful to (...)

[rank 8 options, including 8. Other ... (free text)]

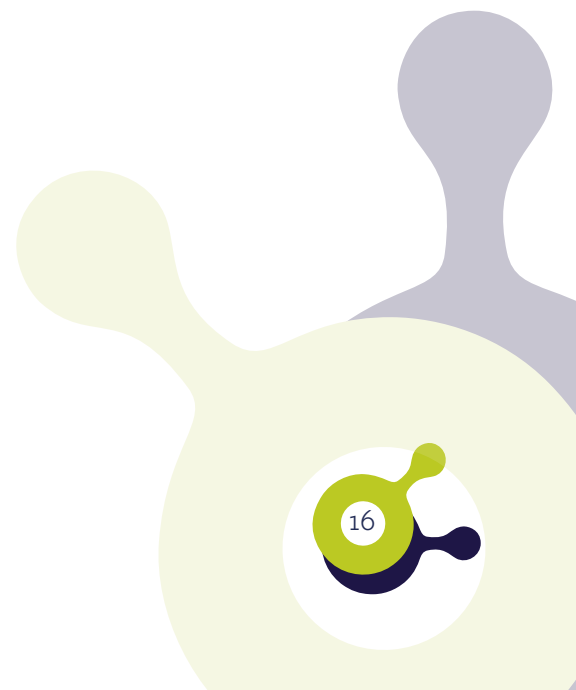
n = 52

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 list of benefits given is fairly complete.

TABLE 4. RANKING OF BENEFITS OF A DUTCH DATA CURATION NETWORK

Benefits as ranked by the respondents	Mean
To reuse guide lines (e.g. how-to's or instructions) that other organisations have made	3.1
To create guide lines (e.g. how- to's or instructions) together	3.3
To define basic 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





Benefits as ranked by the respondents



Figure 3 Clustered bar chart question 4

The ranking has been recoded. The highest perceived benefit has been given the highest value (mean scores)

[Question 5]

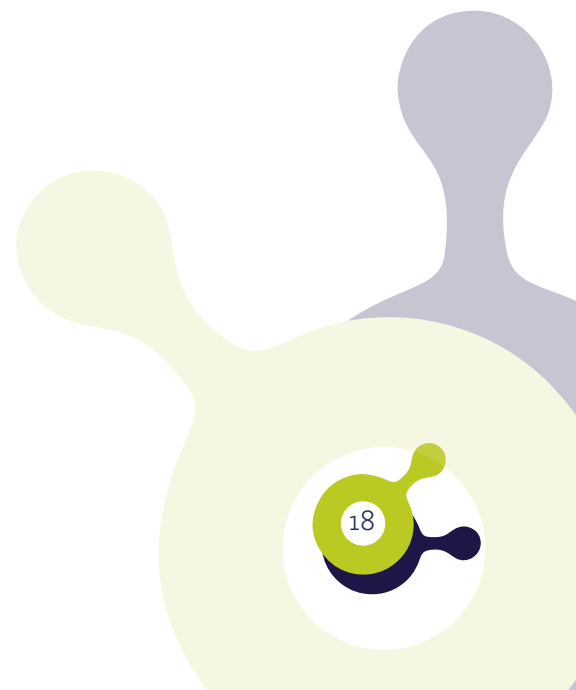
The members of the LCRDM task group Dutch Data Curation Network started to describe their curation practices with the help of by means of a US data curation spreadsheet <link added>. [1. I will add my organisation to the spreadsheet; 2. The spreadsheet is not useful because ... (free text)]

n = 48

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 on 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 engaged 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. For its use as an assessment model for Dutch data curation practices, the model was slightly adjusted to suit the curation practices and needs of the Dutch research 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 the adjustments made were specifically for the purpose of the current task group. For further use, however, it may be advisable to return to the original CURATE(D) model.

2. Having organisations describe their curation practices in terms of the (adjusted) CURATE(D) model, results in a rich and diverse overview of Dutch curation practices, that can well serve as 'good practice' or 'useful case study'. However, it's still too early to standardize 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 ten organisations make a *rich* and prolific impression: it shows that curation practices vary widely, due to differing levels 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 undertaking.



It appears too early to attempt standardisation of data curation practices in the Netherlands, as the CURATE(D) model shows that many of the organisations have just starting to formalise their workflows and procedures for data curation.

Curate action (based on https://datacurationnetwork.org)	Simplified overview of curation practices in the Netherlands
Check files and read documentation (risk mitigation, file inventory, appraisal/selection)	All organisations check files and documentation
Understand the data (or try to), if not ... (run files/ environment, quality assurance/quality control issues, readmes)	Organisations differ in workflows
Request missing information or changes (tracking provenance of any changes and why)	All organisations communicate directly with the researcher regarding missing information or changes
Augment metadata for findability (DOIs, metadata standards, discoverability)	Organisations favour generic metadata schemes
Transform file formats for reuse (data preservation, conversion tools, data visualisation)	Organisations differ in workflows
Evaluate for FAIRness (transparent usage licenses, responsibility standards, metrics for tracking use)	FAIRness is checked by all organisations, with focus on F(indability)
Document all curation activities throughout the process	Organisations differ in workflows. For many organisations, workflows on data curation are still being developed

Figure 4 Simplified overview of data curation practices in the Netherlands based on the CURATE(D) model

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.

The high response rate of 98 participants compares favourably with other surveys conducted using the same nationally coordinated and broadly used mailing list.

According to the Dutch research community, the main challenges in data curation are (1) building awareness and establishing a reward system ("what's in it for me?") (2) designing suitable 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 to draw up guidelines for re-use and creation of data and good practices.

The Dutch research community considers the main benefits of creating a data curation network to be guidelines for reuse that other organisations have drawn up (e.g. how to's or instructions), drafting such guidelines together, and defining basic good practices for data curation in the Netherlands. Clear guidelines may be considered a prerequisite for benchmarking and training researchers, as the latter two benefits of a data curation network were considered less important.



5.2. Recommendations

Now the task group has completed its work, the initial steps towards investigating the feasibility of a Dutch Data Curation Network have been taken. The recommendations of the task group can be divided into two categories: recommendations for the national coordination of data curation practices in the Netherlands and 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 acts as one of the main coordinating initiatives for RDM in the Netherlands, the task group recommends that *a new LCRDM task group be set up*.
- This task group should include a diverse group of members, including repository curators and data stewards from various disciplines.
- 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. In the previous months, the current task group has explored the feasibility, the usefulness and the challenges facing a Dutch Data Curation Network. A subsequent task group could outline what such a network in the Netherlands should do, which stakeholders and organisations should be involved and what challenges should be addressed.
- Another important task of the following task group could 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 guidelines? Or for standardisation of data curation practices in the Netherlands? Viewed from these perspectives, the CURATE(D) model seems to be very promising.
- A final recommendation regarding the national coordination of data curation practices in the Netherlands is to create an overview/page, based on the CURATE(D) matrix adapted by the current task group, of contemporary curation practices in Dutch organisations. Via the LCRDM website the overview/page could be accessed by the broad Dutch RDM community.

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

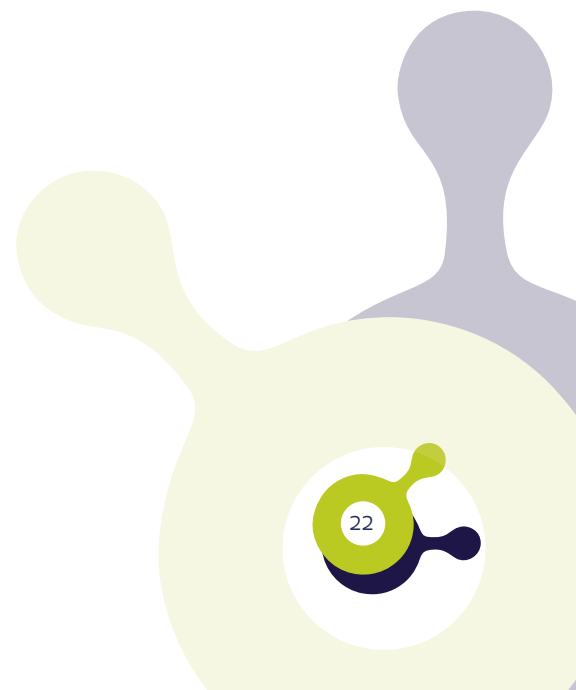
- The matrix supplied by this task group (see appendix A) could serve as good practice or use case for Dutch organisation in order to professionalise data curation practices, and to explain what data curation is about.
- The matrix can also be used as a benchmark to compare the data curation practices of the own organisation with that of other organisations in the Netherlands.
- We recommend that individual organisations and their data support staff exchange experiences, initiatives and actions taken with regard to data curation.
- We recommend that individual organisations and their data support staff become (stay) involved in national initiatives on data curation in the Netherlands.



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
- Alastair Dunning, TU Delft, liaison LCRDM advisory group
- Everyone who completed the matrix
- Everyone who participated in the survey
- And the members of this task group.



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]

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

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 ANSWER 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 a 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 (e.g. metadata) Infrastructure and tools
8 How to avoid data curation to a large extent by making data FAIR at the source	Quality (e.g. 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 (e.g. metadata)

11	Making data F.A.I.R.	Quality (e.g. 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-data cycle	Data curation expertise/support What's in it for me Quality (e.g. metadata)
17	Raising awareness for services	Data curation expertise/support
18	Adding sufficient metadata for reusability, and persistent storage	Quality (e.g. 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	<i>Answer is out of scope</i>
22	Geld besparen	Resources

23	Heb geen ervaring	<i>Answer is out of scope</i>
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 (e.g. 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 (e.g. metadata)
32	Structuur van data opslag, beperktheid van opslagquotum	Infrastructure and tools
33	Grootste uitdaging: Het belang van datacuratie en de daarbij behorende 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 (e.g. 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 (e.g. metadata) Resources What's in it for me
36	Goede begrip van de achtergronden en keuzes bij data-verzamelingen 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 goede metadatering	Quality (e.g. 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 (e.g. metadata) Procedure/workflow What's in it for me
44	Iedereen binnen de organisatie op 1 lijn krijgen	<i>Answer is out of scope</i>

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 principes. Hierbij geldt de i van FAIR als grootste uitdaging	Procedure/workflow Quality (e.g. metadata) Standards
46	Financien	Resources
47	Het staat hier nog in de kinderschoenen en het begint te komen	<i>Answer is out of scope</i>
48	Metadata in orde krijgen. Dit doen we door vooraf zo veel, correct en duidelijk mogelijk de meta-data te verzamelen	Quality (e.g. 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/work flow
53	Response van onderzoekers krijgen	What's in it for me

APPENDIX D

SURVEY ANSWER TO QUESTION 5

ANSWER	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	<i>Answer is out of scope</i>
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	<i>Answer is out of scope</i>
13	See answer to previous question	<i>Answer is out of scope</i>
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

INTRODUCTORY READING• **Data curation definition**

https://dictionary.casrai.org/data_curation

• **Digital humanities data curation**

<https://guide.dhcurator.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