

National Coordination Point Research Data Management

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.

more information: www.lcrdm.nl

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 (TUDelft), 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. The original DCN model	7
3.2. Adjusting the model to the Dutch context	8
3.3. Deliverable: matrix with Dutch best curation practices	8
3.4. First analysis	9
4. Survey setup and findings	11
4.1. Survey setup	11
4.2. Main findings	12
4.3. Response	12
5. Conclusions and recommendations	19
5.1. Conclusion	19
5.2. Recommendations	21
Acknowledgements	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
Introductory reading	32
Advanced reading	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 <u>Data Curation Network Project</u> 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 <u>CURATE(D)</u> 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. Check files and read documentation (risk mitigation, file inventory, appraisal/selection)
- Action 2. Understand the data (or try to), if not... (open files, run code/environment, quality assessment/quality control issues, readmes)
- Action 3. Request missing information or changes (tracking provenance of any changes and why)
- Action 4. Augment metadata for findability (DOIs, metadata standards, discoverability)
- Action 5. Transform file formats for reuse (data preservation, conversion tools, data visualisation)
- Action 6. Evaluate for FAIRness (transparent usage licenses, responsibility standards, metrics for tracking use)
- Action 7. Document 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.



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. **T**ransform 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.

- Are you involved in or working for an organisation (also) located in the Netherlands?
 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 reuse 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 eighter	9	10%
4. No answer	9	10%
Total	93	100%

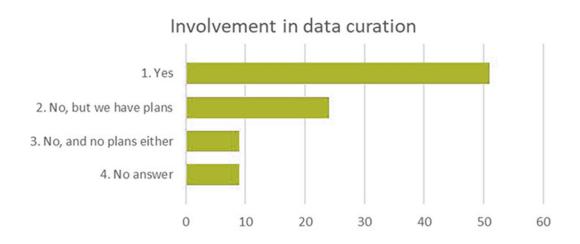


Figure 1 Clustered bar chart question 2

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).

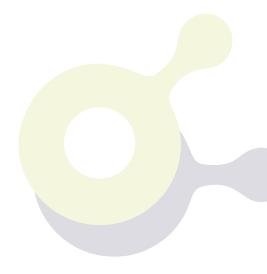
The total response for this question was 54; however, many of the answers mention multiple challenges. In total, the task group identified 94 separate aspects, which could be grouped into nine main challenges. See Appendix C for the complete survey answers to question 3.

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
Answer is out of scope	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

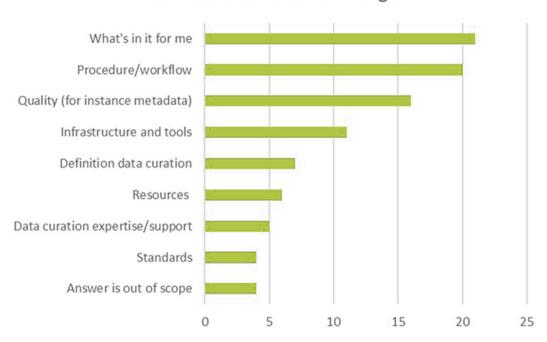


Figure 2 Clustered bar chart question 3

[Question 4]

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

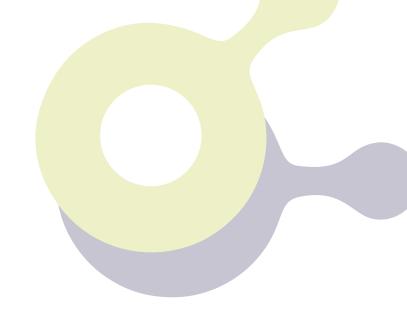
Respondents were asked about their wishes and needs concerning a data curation network in the Netherlands. They had to rank eight activities by dragging and dropping (1 = most useful). They could also fill out optional wishes and needs via "Other ... (free text)". None of the respondents added wishes 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"	4.5
or "checking the data quality"	
To learn what long-term data repositories like 4T∪.ResearchData and	5.7
DANS EASY offer and expect	
Other <free text=""></free>	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 $$
C	Check files and read documentation (risk mitigation, file inventory, appraisal/selection)	All organisations check files and documentation
U	Understand the data (or try to), if not (run files/ environment, quality assurance/quality control issues, readmes)	Organisations differ in workflows
R	Request missing information or changes (tracking provenance of any changes and why)	All organisations communicate directly with the researcher regarding missing information or changes
A	Augment metadata for findability (DOIs, metadata standards, discoverability)	Organisations favour generic metadata schemes
T	Transform file formats for reuse (data preservation, conversion tools, data visualisation)	Organisations differ in workflows
Ė	Evaluate for FAIRness (transparent usage licenses, responsibility standards, metrics for tracking use)	FAIRness is checked by all organisations, with focus on F(indability)
D	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 A CURATE(D) MATRIX

DUTCH DATA CURATION PRACTICES

		ı									
		Radboud University	4TU.ResearchData	TU Delft	Groningen	DANS	Hogeschool InHolland	Utrecht University	SURFsara	Utrecht University DataverseNL	
		We custe only the distances that are and to us for authorizing it the Dooks impossible y- ring the proposition of the state of the this system (types y-leven x allemans) with system (type y-leven x allemans) with system (type y-leven x allemans) with system (type y-leven x allemans) and the system of the support that should the system of the support that should consider the time of the support that should the system of the support that should consider the system of the support that should the system of the support that should the system of the system of the system of the system of the system of the system of the system of the system of the system of the system o	long-term access and curation of receasth datasets, with a focus or data from ocience, engineering and	as a front office for eTU Receashbota. This means that all feedback regarding dataset	or data curation in it data policy (2015). Most research institutes have a protocol for long term	http://tinyuri.com/yzuwftip	Hope-chool introduced the metabolist part on the part long metabolist produced part part part to the part long metabolist control part part part part part part part part	YODA. The recearcher output archived in YODA is checked by a data manager. Utsech University also has an agreement	services for long-term preservation, sharing and publication of recearch data. The Data Archive provides low	All researchers have the option to register for a baseweeks account using their CU credentals. They can add data to a disaset, but there-datasets are checked before publication by RDM Support. The checking is:	Het Meedens Instituut is onderdeel van de KNRW. Het beleid lissles de discountei is vastgeleigd in de Dissonatie (ODIE) Het Instituut sluit daarbij aan bij de diss-principes en het dissibeleid van de KNRW (pie: https://www.knaw.kl/si/Phematics/) (approchass/logendist 240).
		control form to curate the dataons. At least 2 colleges at the RBM support team check the dataons operately from each other. For	in the technology, Every researcher, both in the technology and abroad, can upload data to the data archive or access and download data for use	provided via the data officer in the RDS team. The data officer receives the metadata quality review from	conventions, metadata, codebooks etc. The purpose is serving research integrity and re-use within the		op dataset it er nog net maar dat zie it nu we komen. In het verleden zij er ruim 1000 publicatiec ingevoerd waarbij de metadata nog vank	through Dataverse, the dataset quotiened through Dataverse are checked by the local admin. As far as I know, here is no special curation service if the researcher	dataset, while the total Repository service provides a self-service platform for researchers to share	publication by RDM Support. The checking is more high level and wouldn't qualify as curation.	Major (Ferres Arian in) Information () approximation operation § 80 Ratio (Ferres in extraor line is information () following free earth- data management, Daamaast is Het Meeters Britishut is
	Provide a short overview of how data curation is set up at your institution (see are interested in	researcher, there is a manual for this process: https://www.nunl/research-information- senices/manuals/step-archiving-dataset/	in their research. The publication workflow may differ clightly depending on whether we are in direct contact with the researcher	the moderator of cTU.RecearchData The RDS team works closely with the faculty data stewards who are providing domain-specific support	recearch group. Archaeology curates part of its data in Dataversells, GELPES has its own repository irectricted access. The		onvolledig en comic onjuist is: Sit kwam vooral door onbekendheid en onervarenheid sas ondersoekers (Bij shholland is er oprake van decentrale invoer door ondersoekers es	wants to publish at DANS EASY or make use of any other repository. The RDM team is ready to help with any questions, around data publication, but we don't	and publish dataset of any size with annotations and persistent identifiers. A separate assisted workflow enables base-scale dataset		2 PM The process of
	data curation is set up at your institution (see are interested in quality assessments in place for the data that is being active-of-published within the institution and/or by the employers of the institution). Which		or with a front office.	for ROM.	RDD curates data in distaversets, and supports same other repositoried, list recommended repositoried, the RDD also curates metadata in Pure (with help of the Pure team).		controlle va metadata achteral door de onderspekondersteuners), spa prooss en istues zie ik daarom wel saskvakken met dota norma. De shaet is inspecial vanual de door oor	have any strict procedures for that so many cases, it is important for the research: to use a repository that is well- ertablished in the field for the country.	publications: that a curation is in pilot only on a technical level, i.e. the use is forced to annotate all data and is limited in choice of file forcests. But		ondersoek controllentuar en reproduceestaar ternaken. taankaart daan wij datavet op voor huidig of toekonistig ondersoek. Voor dat laasta hanteven wij een acquistiele model dat 2018 Dr. door wij in processoraal met da
	data is curated and in which circumstances?				metadata in Pure (with help of the Pure-team).		gewents course in.b.t. data cursion	type of data.	request, the researcher is supported in curating new or existing data publications. SURF cara is setting up	1	collections committeement environment environment environment floramment claim or discharged op over charged of selectioning of ordercoler. Voor der basson historiern wij een coquistie- monde (set 2015), orden wijn in camergrank met de ondercoleriert (machtig we ook viragen of de set complete it, of et occumentaties, of et populationises it ext., in bedar granten sijn de ondercoelers sidonal sit het gast on seboud en kvaliteit van de datasent 20 ook bij code 29e ook het collectivijste.
									DAMS.		Who are the state of the state
									All answers below are for the Data Repository service.		
				No	Data curated by the RDO: yes,	yes, we do this for all files. Some details: A)	Tec, if possible, I don't suppose that we have all	yes	Because of the self-service nature of	We advise to ctick to the preferred formatcas:	Op dit moment controleren wij of de files zijn aangeleverd
	Which checks do you perform to check if files in familiar formats can be opened? And in the case of unfamiliar formats (when it is not immediately clear which offware is required to open them??	tion, we do this for all the files, even if unknown and/or unfamiliar software is needed to open	Yes		unless for short term re-use specific formats are helpful that cannot be read by saftware affered on our workstations.	yes, we do this for all files, some details: a) when the original software is too expensive for us (e.g. State), the work-around it to convert the (ii) this case to SPIS) and check that version. B) we contact depositor in case of	te took to oper all sorts of files.		the senice, there are currently no checks other than ministype determination to see if a file is as indicated by the file extension. The accepted file formats are limited	fiched by DANS. Or use an other widely used formati for the field. Also common formatic- MIS Office - are accepted.	Op dit moment controleren wij of de files zijn zangelewerd in de perferved formats (http://www.meerses.Linaw.ni/cms_(enzgen/storien/stata/P- nelersed/constatth pdf), dat it nu nog manuel. Wij zijn ook besig met een traject om een geautomatiseerd cycteen met checksuum oo a visten.
	immediately dear which software is required to open them(?	the files.				damaged file.			to knows and well-established formats (see below).		
				No	10	yes for executables and yes when we have the required coftware (if any). Otherwise we only try to open it. This issue also relates to our	Yes, if possible.	Aec	There is no specific test for this during the creation of the digital objects on the platform. A user can	We advise to focus on reproducibility and to include the needed code plus a readme.txt to ensure data can be reused.	HER MANISON INCIDENT IS CONSISTED AND THE MANISON HER DESIGNATION IN VALUE OF THE MANISON HER CONTROL OF THE MAN
						demand to use so-called "preferred formatc".			ectablish a link to an external source for the code. For large-scale-datasets this is done prior to publication		possity mer measure and stated pain of an easily precision of the first databased was do KNRM (sie: https://www.knaw.nij.el/thenaturch/apenside-ooe/apendix.a-en.
											https://www.kinaw.irl/eli/hematists/j.lepensis-only/pensist 3-20/ https://www.meetens.kinaw.ir/coss/jal/collecties/leseasch- https://www.meetens.kinaw.ir/coss/jal/collecties/leseasch- date-management/, Dawnaads is net Meetenses is include is gover/ficesed meet de Conferious/Seal de crowde emasur de collecties digitate de sepa accesso and to bleeden set Meetens incitates is een suitable Contes. Wij slaan data op your trees meetens: sood de anderzoelessis on het
	What do you do if there is code provided within the data set?	tes, all code provided within the dataset is checked and numed to see if no errors occur.									Meethor (Indition) is en CLARINI (Certin, W) caal data op voor twee redesen; voor de andervoelenz can het ondercoek controllentaar en reproducerstaar te maken. Daarnaart daan w) ditazent op voor hulling of toekonstig ondercoek; voor dat bastot hanteren w) een augustite model (uit 2019). Dit doen wij in zamengraak met de
C Check files and read documentation											ondecoek. Voor dat laatste hanteren wij een acquisitie model (uit 2019). Dit doen wij in samenigraak niet de ondecoekers (waarbij we ook vragen of de oet campieet is, of er documentatie is, of er gepubliceerd is etc.). In beide
											model (set 2013), tirt doen wij in zameegynak met de ondersoawer (zamely see ou wrzygen for set tiensjiere it, of er documentzier is, of er gepublichend is etc.), in teiles gestlich is je de ondersoaken telendari je bet gest zem gestlich is je de ondersoaken telendari je de beginnen bet het collectrajate. The proposition of the proposition of the proposition of entrejatere per sentence kan en (from fingerge) publicate enjoyen entrejatere upd. De metadata van de stateser wordt gegenemende na persoaken de stateser wordt gegenemende na persoaken de stateser wordt metadata is een eigen standarid waar Dublin Core en CMD1 metadata is een eigen standarid waar Dublin Core en CMD1 metadata is een eigen standarid waar Dublin Core en CMD1 metadata is een eigen standarid waar Dublin Core en CMD1 metadata is een eigen standarid waar Dublin Core en CMD1 metadata is een eigen standarid waar Dublin Core en CMD1 metadata is een eigen standarid waar Dublin Core en CMD1 metadata is een eigen standarid waar Dublin Core en CMD1 metadata is een eigen standarid waar Dublin Core en CMD1 metadata is een eigen standarid waar Dublin Core en CMD1 metadata is een eigen standarid waar Dublin Core en CMD1 metadata is een eigen standarid waar Dublin Core en CMD1 metadata is een eigen standarid waar Dublin Core en CMD1 metadata waar een een word word to persoaken en een een een een een een een een e
											ectioplanew.pdf. De metadata van de dataset wordt gegenemend en gecantzisiend door de afseing collecties. De metadata is een eigen standaard waar Dublin Corv es CMDI meddata van sraener end wordt.
	How do you evaluate the richness.			No	We consiste recearchers to describe data as they would have liked it, if they would have found it. We consistence add metadata	yes. When metadata is week, documentation is other week too, and we contact depositor.	Yes, this is necessary, next ours if we do agree on what is a nequired minimal set of metadatal	make sure the fields in the metadata editor are filled	The only requirement is that all required fields of the metadata schemas are filled in	We check if there are fields left open that could be filled, e.g. software can be specified very after.	
	How do you evaluate the richness, accuracy and completeness of the metadata?	Yes	Tec		(important variables, methods, fields, scientific names of species as key-words). Check the right affiliation in Pure. We-do not aim						
	What do you expect to be present in the documentation? (reading, codebook, data dictionary, other?)	196	Yes	No		yer. When documentation is weak, metadata is often weak too, and we contact depositor.	avt	yes	We expect the researcher to provide at least a concise description of the district in the metallics. Any added	Moor data is related to a publication, the descriptions should clearly state what data and/or code is included and what not.	
	What kind of charts do you pool- to			No	Yes. We shook for direct and indirect identifyable data, albeit the	yes. We also check if the data is anonymised and if needed involve our legal expert. The	It, als onderdeel van ontwikkelen van datamanagemension	yes.			Personsgegevens worden door oes behandeld conform de Allemene Verordening Gegevensbeschenning Motils. De
	know if there are human collects involved? In case there are human subjects involved, how do you check for direct and indirect identifiable data?	tex, if it is a distanct with a real patention of containing personal data, it separate people perform a privacy check to prevent data leakage.	Tec		lact is by some rules of thumb rather than analytical tools. We work on a checkler. We offer a DPA service for a final check and	yer. We also check if the data is anonymised and if needed invalve our legal expert. The depositor is responsible and in principle bands doesn't do anonymisation or previouslymisation.				recearcher is responsible, and we after advise on what how to anyonimise or exclude personal information.	Personingsgevens worden door on behanded conform de Algemene Virrordening degevendeuchenning (MVG), De privagywicklasing van bet Meestens Instituut is de koese op die welcoler van GE MVM (Telss //www.kraw.el/el/de- kszan/grivagywicklasing-kraw).
				No	We do check for headings and explanation of headings, but not for mixing data in the datafiles. We do	yet. Quality soluments not conference, but Conference, but Conference, but Conference, but Conference and Confe	avt	make sure that files mentioned in the documentation are submitted and there are no files that are not mentioned.	Check data precentation manually. Depending on the ingest workflow we can make sure all files are	We-don't check actively for mixing data or failing code, but we make clear that all data in the dataset will be published and potentially reused.	
	What usability criteria de you consider? (missing data, ambiguous headings, code execution failures, etc)	tec. All of the items described here are gast of our data cursion process: we check if there is missing data and ask the invanisher to document why this data is missing. We also execute the code to see if it must without continues.	Tes		check missing metadata. If a lot is missing we try to add surveives (KED)/Pure-team) or ask the occasion to miss more	interdependent. Ambiguous headings: we propose for demand) changes to the depositor. Code execution failures: back to depositor.		Check data presentation	present.	record.	
	Which metadata do you extract from	of content, but on the level of the data quality.	Yes, if not already added in the	No	information. No checks on code execution yes, but we do not search "forever". that wild before propert in these	yet. Eg. we search for publication(s) and	avt	The data manager does not necessarily has domain specific knowledge to evaluate what is important for neuce	Currently no metadata is extracted from cubasimal files	Only the built-in options on size and mdS checksum.	Zie 3. Op collectioniveau DC en CMDI. Op een lager niveau
U Landerstand the data (or try ta), if eat				No.	allows for validation, improvements and re-salidation.	dataset, in cace of "known" deposition we search for related datasets in our repository.		evaluate what is important for reuse			
	What ways do you have to determine if the documentation of the data is cufficient for a user with similar qualifizations to the author's to understand and rouse the data?	ter, the quality of the documentation is one of the most important aspects we check on. We make recommendations towards the researcher, but we do not create any documentation ourselve.	Tec	_	more if the policy and the researcher indicate the gives info ic good enough. We are no expert in	station in missele stations on improving yet, this is a major goal of our data custom. We provide nch information on the velocities ("yes larger") stray ("larsa stans villey/deposit, infor- mation-about-appointing". Attalnet, languagenes: We dan't create additional documentation ourselves yet, and we report a code-book equiaming headers, variables etc.		The data manager does not necessarily has domain specific knowledge: the qualifications are not the same as qualifications of the researcher.	does not have dontain-specific knowledge for file formats and contents.		
		documentation ourselves.		No	Headings are checked. Structure is	data/set_language-en. We don't create additional documentation ourselved yet, and we expect a codebook explaining	avt	yes, form and completeness are checked	There are no checks in place for	We only do some spot checks at the moment.	
	Which parameters of the tabular data do you check? (structure, definitions of headers, codebook) How do you communicate to the	196.	Tec	The list of questions is provided by		yes, Esseded. See http://deyuri.com/yzuwitip					
R Request missing information or changes	How do you communicate to the researcher what changes need to be made and what know, errors need to be fixed? (Cruffy, by enail, crasting a lot of necessary changes, implementing the changes younce! and discussing the results?)	tos, after we complete our standardized controlle form, we callect a list of questions and suggestions and mail the depositor of the dataset.	Tes	data officer of the MS team (fort office).				neceascher on the same dataset before it is published. Questions and issues can be discussed. I am not sure if questions are secoded in floor of a "list".	and structure of the files as necessary. This will be communicated in person or via annual for coff-contra insert there is	We usually right wis e-mail and offer further 6 support by mail righton. We clearly come the changes than end to made, change that are optional and mention the CC Science.	
	implementing the changes yourself and discussing the results?]			No.	yes, especially keywords. Topics	yes, e.g. we add extra Subject terms (+	Size	The metadata forms are provided by YCCA environment or by DataVerse	no such process, but their might be communication after the publication the metadata fields are structured	We look for additional identifiers and ask the	Zeker. Dist does wij is overleg. Zowel mondeling bij de
	Describe how you enchance metadata to facilitate findability	We only enhance metadata when we don't change the content. For example, if laywords are not separated by semicolons, we add them for the researcher. However, if we think that there should me more keywords, we make this	Tec, every deposited dataset is undergoing a metadata quality major funnation for		often need to be included. Also general description or method may need more info. In Pure we try to nebte to other research output.	yet, e.g., we add extra Subject terms (+ keywordd) and Location (if we're cure). Also, we add Relations to e.g., external websites, seland datasets, publication etc. in addition to what we do surrollves (if seeded), we may request to rename unclear file sames and/or		YOOA environment or by busiverse	according to a metadata schema which enforces certain formats and allows linking to other data sources. Some-fields are tied to (controlled)	researcher to add this.	Zeker. Dar does wij is overlag. Zowel mondeling bij de istake zis uchstrelijk.
	metadata to facilitate findability (correcting evens, adding keywords, linkages to related datasets, etc.)	for the researcher. However, if we think that there should me more keywords, we make this suggestion to the researcher.	improvement of the metadata are seturned to the depositor.		This requires checking several times since workflows and timing my differ. We still need NARCIS to read our Pure metadata on datasets.	request to rename unclear file names and/or recommend to sig-files into the decired, clear folder structure.			vocabularies.		
A Augment metadata for findability	to which cases do you structure and precent metadata in domain-specific schemas to facilitate interoperability with other systems?	No, out for specifice datasets.	No	No	Yes	not on the level of datasets. In specific cases we adopt metadata for community harvesters, so that they can aggregate the metadata of all datasets relevant for their community. This	evt	Various YODA environments follow their own standards, which correspond to what is agreed upon in a particular research community.	Communities and domains can define their own metadata schemas to allow them to make their data interoperable within their domain.	We use the domain-specific fields in dataverse	
	interoperability with other systems? How do you evaluate that linkages are sufficient? (link to report/paper, to related data sets, to source data,	tec. We check if the researcher has made a link		No	Yes, especially in Pure. We do not have sufficient capacity to do the	dataset newart for their community. This costs makes >1 yet, this is a major goal of our data curation. See other fields above.	avt	We try to ask the researboers if there are			dat controlerer wij bij de intake.
	to related data cets, to source data, etc)	to the corresponding research paper. If existing data is used, we also check whether there is a proper inference.	publications.	No	came for Estaversets. This does not happen in local research group archives yes, sometimes we add generally	yet, this is why we have so-called preferred	avt			We advise to use the DANS preferred formats,	
		ter. We check if we jobta curation team) can			readable formats.	yes, this is why we have co-called perferred formans. Documentation should contain sufficient possible, or which a good description of what's reeded to access and use the table of the contained on the contained of the contained on the contained		the researchers are free to archive formats they prefer, focus on safeware should be in the documentations	between accepted and preferred formats. Any other file formats are not accepted upon ingest in the self- senice portal. For the massive inges	We advise to use the DMAS girden of Sonnace, or another well known florings in the Seld.	
	Which criteria do you have on operiodized file formats and their restrictor? (e.g., to the software fixedy available? Link to it or archive it alongside the data)?	Yes. We check if we jobts oursion team) can open all the files. This includes checking if needed coftware is available and can be easily sistalled. Needed callware should be mentioned in the documentation, accompanied with information on where to download the software and how to open the files with the software and how to open the files with the	rec, we check if the dataset is provided in a preferred file format. See our list of preferred formatic https://researchdata.etu.			data https://dans.knaw.ni/en/deposit/informa tion-about-depositing-data/before- depositing/file-formats is the current version			workflow any file format can be concidend after careful consideration with the data producer. Generally, file formats that are open, free to	•	
	it alongside the data)?	software and how to open the files with the software.	sten/prefered_file_formats.pdf						the transact that are appea, twests were, considered standard floreats in relevant communities and commonly used are perferred. Accepted formatic can be stared, but are not custed other than bitwise preservation.		
Transform file formats for reuse				No	Nut as a standard procedure. If we	yet, if needed we convert e.g. Word, Sucel, DBF	evt	So far, no such transformations are	are not curated other than bitwise preservation. Transformations can be performed	We advise on using the preferred, but also	Nog geer specifele criteria. Behalve dat het kan zijs dat een
	Which offsels do you have an preferred file format; and transformation into open, ean- proprietary file format; that braden the potential audience for reuse.	tes, we use the list of the DANS archive. Often, we also store the original files so that no data is lost.	We don't do this ourselves.		we usually upload both versions as some information may get lost in the transformation process.	yet, if needed we convert e.g. Word, Sucel, DBF and audiovicual data to preferred formats. We setain the originals.		deriving copies in preferred formats when possible	a (number of) datacet(x). The transformation should always be to a preferred format.	Estavenent is not very long term storage.	Nog geer specifiele-critaria. Behalve dat het kan zijs dat een onderoorker een specifiels format of colheurs wil gebruiken die niet duutzaam it. Ook in despleijke gevollen moeten slaan wij die data op minners, het is per definite zu dat nieuw onderzoek ook gedoon kan worden door nieuwe
				No	defenitely within the Uti-domain and preferably general.	yet; ceraiss "check if code runs", in case of unclear coffware version or unclear coffware at all, we ask for convention to a better documented and specific version of a specific application.	eve	For some datasets (inc. dec-labs) only specific software can deal with the data.	There is currently no such specific criterium; for the preferred formats	No criteria at the moment.	datasets en tooks die finog) niet duurzaam zijn of als duurzaam zijn dekwalifioeerd.
	Which criteria do you have on the availability of software needed to open the dataset?	tec. If not, we advise to close it with the dataset.						For some datasets (iss. dec-labe) only specific software can deal with the data. So far, it have't been problemosticad and it is assumed that it should be possible to archive those datasets as well.	accepted farmatic, it chould be broadly used and accepted in the community		***
	Name same of the metadata fields you expect to find next to author/title/date.	ter. We check if most of the metadata fields are filled in by the recentive. We also encourage researchers to extend the metadata when we find it too brief.	Tec.	_	Even in the lacal research incitiuses archives more is required (n as stated in their protocols/policies).	OVERALL COMMENT about 6-FAIR: DANS "provides" overall FAIR qualities from this list, so we don't "check" for their in the datasets. That we receive. That we receive.		fields	botacite fields can be filled in. They are mandatory or recommended according to the batacite guidelines.	author / author identifier / institue / title / related publication	
				No	We check the references.	is we don't "Check" for them is the district. That we receive. The, we check that all mandatory fields are filled in and recommend that all bubble Core fields are filled we provide a DOI.	evt	yes, add PRD for authors and contributors as much or countries	DOX and SPIC PIDs are automatically promped to the district chief and	# All datasets in CoroverseAS get a Handle.	Als CLARR B centre voegen wij PD's zan de datasets die handrijkhor worden som vijf voor (* ARM (* budse)
	How do you make ours the dataset is findable with a PID?	All datasets curated by us are archived within the DANS EASY archive and will have a DOI.	tec, every distance is provided a DO upon publication.	No.	All first sections of security free.	was accounted all controllers of all districts up the			accigned to the digital object and files upon completion and publication of the object.	bscaversets is indexed by Google buts Search	
	Was diversity and	Alderen constitu	We support the One PMH prynocol		should be described in Pure (which is not the case yet). The metadata may be readable for the public, within the Uti-domain or bad-	we provide all metadata of all datasets via the CAN-PMM protocol to search regimes and aggregators		data services	An CRA-PMM endpoint is provided for harvesting and the site can be indexed by any web crawler.		dese zaken ook mee genomen worden.
	How do you make ours that the dataset will be discoverable via web search engines?	All datasets curated by us are archived within the DANS EASY archive and the discoverability is therefor guaranteed.	to abow the harvesting of our metadata for integration in search engines. We have also embedded schema.org metadata in the		office only. General access levels are specified too (open, restricted, closed etc.) Public metadata in Pune is indexed by e.g. Google. We want						
	Was diversity and		schema.org metadita is the dataset beding pages, so that it can be indexed in Google Distaset Search	No	very much to be harvested by numbers. Yes.	we provide all metadata of all datasets via the	evt	YCCs and DataWare comply with this	The website is fully compliant. Only	is hanked via DANS-Custaverse	Het Meertens Instituut heeft datasets die webbased
	How do you make cure that the distanct is retrievable via a standard protocol (e.g., HTTP)?		Yes, https: protocol is used.	No	We check all possible access levels	we provide all metadics of all dataons via the CAN-PMM protocol to search engines and aggregators the DANS mosts "Open when possible,	II, m.b.r. publicative	YCOs and basiswase comply with this requirement The crisical lading page is checked for mistakes.	PITTP's connections are accepted. A REST API allows for automated interaction All digital objects have a dedicated	It handed via DANG-Cistaveroe By default all datasets are licensed willt CCD, we encourage the use of CCD of CC-by	Het Meertens Instituut heeft distaats die webbond beschiktaar zijn: http://www.meertens.knaw.nlj/cns/inljozilecties/dissibanke n
	How do you assess that the dataset is free, open? How do you make over that it can be downloaded?	orfision accession. Wecasthers can choose detween open access, and restricted access. As the policy of the Madbood University states that distance should be published open access, when possible, we encourage meanthers to publish open access. When we see the researchers choose restricted.	Currently, only Open Access is provided and an embargo date can be set on re		and information on it and copy this is Pure. We may get in touch with the researcher to ask for further information and discuss options.	the DANS motto "Open when possible, instituted when needed" has been widely adopted However, so far the choice is left to the depositor. It is also possible to combine Open and Restricted Access files within one dataset.		metabec	tanding page which displays all open metadata. Depending on the share level the files can be downloaded.	we encourage the use of CCD of CC-by	
E Evaluate for FARkness (**)(%)		over extent when there are nood recessor to					IA, IN.D.T. publication	YCCs and DataWise have their own standards	Dublin Core and DataCite are expected and enforced.	default by Docaversets.	
	How do you evaluate the choice metadata format? How do you check if it follows a standard otherna?	The datasets curated by us are registered and deposited using the MS cyclem. Here, we have a fixed necasitis scheme (suntinistion of stractice and bublinCore). Therefor, we don't sorticularly check the medata scheme.	of standard ontologies and vocabularies, e.g. Dublin Core (dcterno), Soal, owl, wegits								
	How do you make ours that metadata is provided in machine- readable format (OM feed)?	No.	Yes	No	not sure	we provide all metadata of all dataons via the CAN-PMH protocul to search engines and aggregators. Moreover, the metadata can be downloaded as covand will file.		YOOs and discrews comply with this requirement	As CRI-Phile endpoint is provided for harvesting. A MSST API allows for automated interaction using ISON format	default by Escaversons.	
	Which contact info do you expect to be displaced if the di-	No. The author name and affiliation are per default mentioned. We do not provide an entallatives or other contact details in most cases the remothers add it higher contact obtails incide the disport documentation. However we not have a check on whether contact information is provided.	No, but we support CRCID ID for	No	yect. General contact point is the Research Carta Office, so we must be able to find specialists. Responsibilities are described in the	we provide all metadata of all datasets via the CNN-PMN process to search regions and garging passive. Moreover, the menadata can be no, attorupt a deposition of all provide this and is no, attorupt a deposition of an provide this and is also encouraged on their DNI jained in the near future also CRCDI).	avt	different labs have different agreements upon that	the author is registered as a user at subtrains and contact details are to be filled in during metaduta annotation.	Distaveness has the mail address, contact goes via the 'contact' button on the website start.	De adringegevers, sail het Meesters Instituut.
	be displayed (if the direct assistance of the author needed)?	details inside the dataset documentation. However we not have a check on whether contact information is provided.	auchor and contributor names.								
		contact information is provided. Which the medical of the discusser, the right sholder of the discusser is sharper mentioned. In most case, this is the Stational taniestip, Asia, auchier Land on building are a propriet fail. The Asia, contact Land the auchiers of the authors of the discusser Land the authors of the discusses discusser. The disposal of the discusser discusser. The disposal of the discusser, and the discusser is the disposal of the discusser, for managing according expects. However, this is not made clear from the next-state.			mere is still doller work to be done: different roles of researchers, affiliation (many repositories make pigs ears of that). Ownership is	mainly yet: Cristor is a mandatory metadata field, Right cholder is an optional metadata field. A depositor can optionally add a data steward as a Contributors.	w, = 0.1. publicativis maar geen informatie over stewards	Creator is a mandatory field. Right sholder is optional. As a default, affiliation is interpreted as right sholder	wester is a mandatory field. Rightsholder is optional. As a default affiliation is interpreted as right sholder.	the account used to create a distanct is, concidered the main contact and creation.	De sullecties zijn op dit moment op een klein gedeelte za het eigendom van het Meennes instituur. Die solicities die dat niet zijn ander koerte gedeponeend. Va het Meennes motituut kan contact worden gesocht met de eigenoor.
	Which indicators of who created, owns, and stewards the data do you expect in the metadata?	required tiest, the slauge check whether the authors of the datasets and the authors of the corresponding article match. If not, we ack the dataset depositor. The depositor of the dataset,			pige aas of that, Ownership is difficult in Dark laws a researcher has rights of use, as has the university. What is an issue in oursion, are data that belong to third parsies. Sowands are the RDD and parsies. Sowands are the RDD and of the according to the policies of this and meanch inclusive. This is comething to work on, researchis. If the policy of the policy.						
		n we one who columnist the data and handles for example access requests receivery, this is not enade clear from the metadata	Depositors can choose a licence	No	and then according to the policies of CKS and research institute. This is comething to work on,	we provide a urage licence.	II, m.b.r. publicative	ticense is a mandatory field	Licence ic a mandatory field and can	The default is CC-0, we inform and offer advise	
	How do you approach evaluation of usage terms (e.g., a CC Liceose)?	Exta published open across in the banks factor archive is accompanied with a candidad losese. Within our curation process we do not check if additional license-files are used futher the data or published restricted accress.	from a predefined siz. The full range of Creative Commons, sources for distauts, and specifically for software and code,		This is comething to work on, especially if CC-oero, CC-by, CC-NC is not usable.				Se chases via a license selector soci	if a researcher wants to change this or add a data availability statement.	
		ic published restricted access).	three popular opes source licences are supported.	No	This is cometing we try to	yes, from the moment the dataset is submitted	avt	reformation about the approval of	Any changes to metadata are	From creation until publication the dataset	
D	Which provenance information do you record (who did what to the dataset and when)?	Yes. We keep a standardized control form in which we capture all of our findings and actions.			This is cometing we try to implement in our new monorth workpace so this type of data is money or less automatically generated during the research process.			information about the approval of submissions is saved system internally by YDDA	www.matcany logged in the system. File changes are not allowed.	From cristion until publication the dataset has a status "DMP?". The Handle is assigned, but cothing is finished yet. Then when published, it gets a version 1.0 and all changes after that get a new version including a log and a new curation/checking round.	
Document your curation activities (*900)	What is included in acceptioning & depocit records frames, dates, contact information, pubmicsion assessments, etcl?	Not, we have an overview of the deposited distance within our CRS.			process. partly, incomplete.	yes, for keeping provenance information	II, onderdeel van het distansangementplas	Same of this information is caved system intentity by YDDA, to my knowledge there is no separate database	The publication is linked to a data owner which is known by SORF cars	and a new curstion/checking round. All is in the dataset description.	Minimaal DC en verder 30 veel we kunnen (incl. taal, geografie, periode, eigenaar)
	contact information, submission assessment, etcl? Which provenance logs do you keep?	datasets within our CRS. Text. We keep track of different versions of the datasets.	Tec.		If our new research workspace, in Dataversets, (background) and in	yes	Nee		Automated logging	ts kept online in thataverseld	25, voor zover het voorkontt does we zan versiebeheer.
	Do you have a service workflow to follow the curation process?	tion, we follow a standardized control form is which the service workflow is described.	Yes, there is an internal workflow in place.		Yes for Pure and Dataversets.	yes, see http://tinyurl.com/yzuwitsip	Nog niet beschreven		to development.	Not afficially. There is a RDM mailbox, all request end up there. This is handled by three RDM. staff within Dosawerants. UU users can get an own folder and manage and curate their own datasers. By default all is checked by RDM Consultants, mainly this is informing the	Za er is ees intenne workflow
	Describe any other relevant requirements for data curation process at your institution	No		Data policy: https://dirkab?tiqylifs.cloudfront.e et/lubrary/themaportalen/RDM/rec earchdata-framework-policy.pdf	and received institute policies first, then think of how to share data more openly and register in Pure.			The recearcher neds to be employed by the involution to make use of the data curation services.	before and after publication and is offered as a separate package.	Solder and manage and curate their own datasets. By default all is checked by RDM Consultants, mainly this is informing the	
	process at your institution									recearcher on the default license, whether or not to inicide personal information and to be clear when describing the dataset.	



Question 1



Dear colleague,

The LCRDM task group <u>Dutch Data Curation Network</u> 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".

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

Q	ue	25	ti	OI	n	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 <u>Dutch Data Curation Network</u> started to describe their curation practices by means of a US data curation <u>spreadsheet</u> . Select one option
I think the <u>spreadsheel</u> 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



contribute

APPENDIX C 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	 Metadata: getting it clear and good enough. 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	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	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 (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 professio- nalisering	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 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 (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 onderzoe- ker)	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 toe- komst 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	ledereen 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 (e.g. 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 (e.g. metadata)
49	Bewustwording bij onderzoekers mbt belang van data- curatie 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



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	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 GUIDE LITERATURE READING

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/uksq.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