An introduction to

Authors: Boudewijn Alink Tom De Schryver
Maarten van Bentum *(contact person)*

Date: 16-6-2014

Table of Contents

1 For the purpose of data DVN for dummies

management for scientific research, B & A forwards a secured software platform DVN. This report is written to understand how groups of researchers can make use of DVN in the best possible way. The report is mainly intended to support information specialists who need to assist researchers in sharing and publishing their data via DVN. Complementary to this report a workshop on June, 12 2014 is scheduled.

First we explain what DVN is and discuss the most important features of DVN. Thereafter we present the decision tree that we developed for information specialists as a tool to help them uploading and organizing research data in DVN. Then we illustrate the use of the decision tree by means of some use cases as possible real world implementations. This report ends with a user manual to explain how to perform the most important actions in DVN.

1.1 What is DVN?

The Dataverse Network (DVN)[1] is an open source application to publish, share, reference, extract and analyze[2] research data. Its goal is to make research data available to others. At a first glance DVN seems to be a complex application, which is not very user-friendly, but there is a legitimate underlying reason. This is driven by the advantage of storing your data in the secure environment of DVN and permit, in a controlled way, multiple persons to access your data. Therefore, access is managed through a set of access permissions. Think of each of these as a lock, which together determine whether or not a user can access a particular data file (Section 1.3.1), study (Section 1.3.2) or dataverse (Section 1.3.3). The application of these access permissions requires some knowledge, which will be provided in this document.

1.2 Features of DVN

To fulfil this goal, DVN makes it possible to:
- Create dataverses and studies (Section 1.3)
- Create a Handle (persistent URL) for each study (Section 1.3)
- Create collections within your dataverse with studies from other dataverses (Section 1.3)
- Publish research data with access restrictions and permissions to specific users (Sections 1.4 and 1.5)
- Provide metadata corresponding to your study (Section 1.6)
- Search for studies and research data of others within the entire network (Section 1.7)
- Guestbook to collect information about users before they can download research data files (Appendix 5.1.2.1)
- Download tracking data – reports with statistics about which users downloaded your research data files (Appendix 5.1.2.2).

In the next sections an explanation of all these features is given.

1.3 Basic structure of DVN

In order to get research data into DVN, one first has to create the place where the data will be put. DVN uses a structure of four different containers to store and organize data: dataverses, studies, collections and data files. These containers are related to each other through the following structure: First, DVN hosts multiple *dataverses*. Second, each dataverse contains *studies* and *collections* of studies. Third, each study contains cataloguing information, the metadata that describes the study,

---

[1] Although not correct, in fact, DDN (Dutch Dataverse Network) is often used interchangeably with DVN. DDN is the installation of the Dataverse Network hosted, operated and used by and for the Dutch Universities. Because the name DDN is not used widespread through the application, DVN is used instead in this document.

[2] This function is not present in the current version of DVN installed by DANS. N.B. In the opensource version of Harvard University it is available.

plus the actual data and complementary files (*data files*). Thus, to store data files, one first needs to create a dataverse and then a study. Finally, each study can be put in one or more collections. Each of these containers will be addressed in short below.
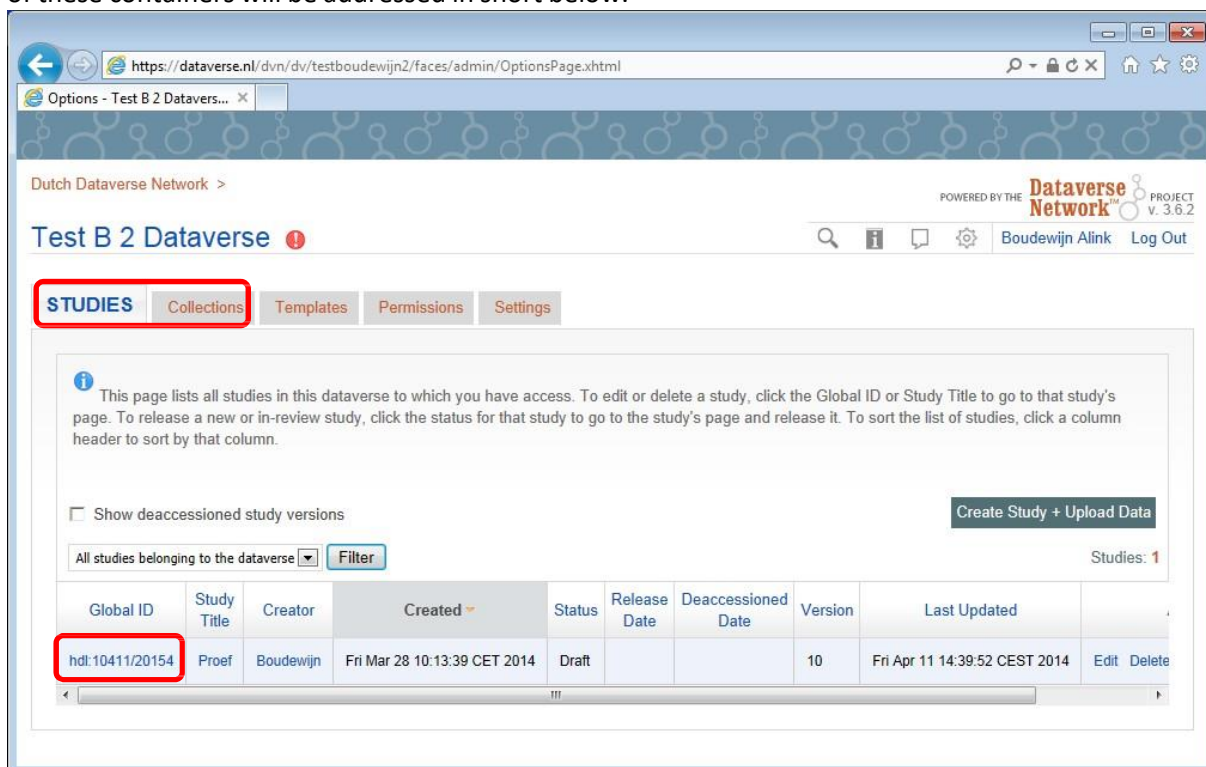

Figure 1: Showing a dataverse with the tabs for the containers within the dataverse: Studies and Collections. Also the unique handle of the study is shown.

1.3.1 Data file
A data file is the data that a researcher supplies to DVN. It can be of any type and format, and technically up to 2 GB[3] each. It can contain anything related to the study, e.g. explaining documentation regarding the study, data used for or gathered during research, used code for analyses, or the text for publishing.

1.3.2 Study
Data files must be stored in a study. A study is a container where you can store research data, i.e. one or more data files. It includes cataloguing information, data and complementary files (data files). Different versions of a study can be created, released and archived (see Appendix 5.4, Version Management). Once released, this version becomes available to the public, if necessary with restricted access permissions.

An important feature of studies created in DVN is the unique Handle (persistent URL) that is assigned to it (see fig. 1). This Handle can be used to refer to the study and its accompanying data files, e.g. in publications.

1.3.3 Dataverse
A dataverse is a container of multiple studies, and/or collections of studies (see 1.3.4). A dataverse can be either "Released" or "Not released". Releasing a dataverse implies that the dataverse is publicly accessible from the network homepage. This is an important setting, because once a dataverse is released, everyone can see – even without an account in DVN – all the content of released versions of studies within this dataverse.

---

[3] Although the standard SLA describes a technical limit of 500 MB per data file, in practice data files can be up to 2 GB each.

1.3.4 Collection

A collection is a compilation of released studies. There are two types of collections: Study collections and Linked collections. Linked collections are carbon copies of a collection in another dataverse, managed and edited by the administrator of that dataverse.

Study collections are managed by yourself. You can add released studies from your own dataverse and from other released dataverses in DVN. There are two ways to add additional studies to a Study collection: (1) in a static way you can browse or search manually in your or other dataverses, (2) in a dynamic way you determine selection criteria to let DVN fill the collection automatically[4].

After creating a dataverse, a standard Root collection is created and, by default, it will contain all released studies created in the dataverse. You can also create new collections in your dataverse. New Linked collections are created on the same level as the Root collection. New Study collections will always be a child of the Root collection. By creating a child within a child it is also possible to make a nested organization of collections.

1.4 Layers of access permission

A dataverse in DVN has a structure with three different layers. Think of each of these as a lock, with dataverse being the outer most, study the next and finally the data file level (Figures 2 to 5 show the three layers of access restrictions). This structure facilitates to release research data with restricted access permission to specified groups or persons (see Section 1.5). These access permissions determine whether or not a user can access a particular data file, study or dataverse. For example, in a dataverse a user can only have access to a data file if he has got granted explicit permission to the corresponding dataverse and study.

We describe the situation in case of an <u>unreleased</u> dataverse[5].

On the lowest level, access to a data file can be set to <u>restricted</u> or <u>public</u> (default setting). When public, the data file can be downloaded by everyone who has access to the study to which it belongs. When restricted, the data file can only be downloaded by users who got granted explicit permission by the administrator of the dataverse.

Also a study can be set <u>restricted</u> or <u>public</u> (default setting). When public, the study can be accessed by everyone who has access to the dataverse to which it belongs. When restricted, the study can only be accessed by users who got granted explicit permission by the administrator of the dataverse.

At the highest level, the administrator of the dataverse must grant access to the dataverse to individuals or groups of users.

The above process of making the right security settings and granting permissions requires special attention of the dataverse administrator. When configuring file access, it might be helpful to approach this from the dataverse access level first and so on. To assist in this decision making process we will present in Chapter 2 a decision tree that helps to set the right levels of security.

---

[4] You can create a query that gathers studies into a collection based on matching criteria, and keep the contents current. If a study matches the query selection criteria one week, then is changed and no longer matches the criteria, that study is only a member of the collection as long as it's criteria matches the query.

[5] The situation for a <u>released</u> dataverse is similar; i.e. as soon as a dataverse is released, the standard setting is that everyone has access to the dataverse, study and data files, unless they are restricted explicitly on the corresponding level.
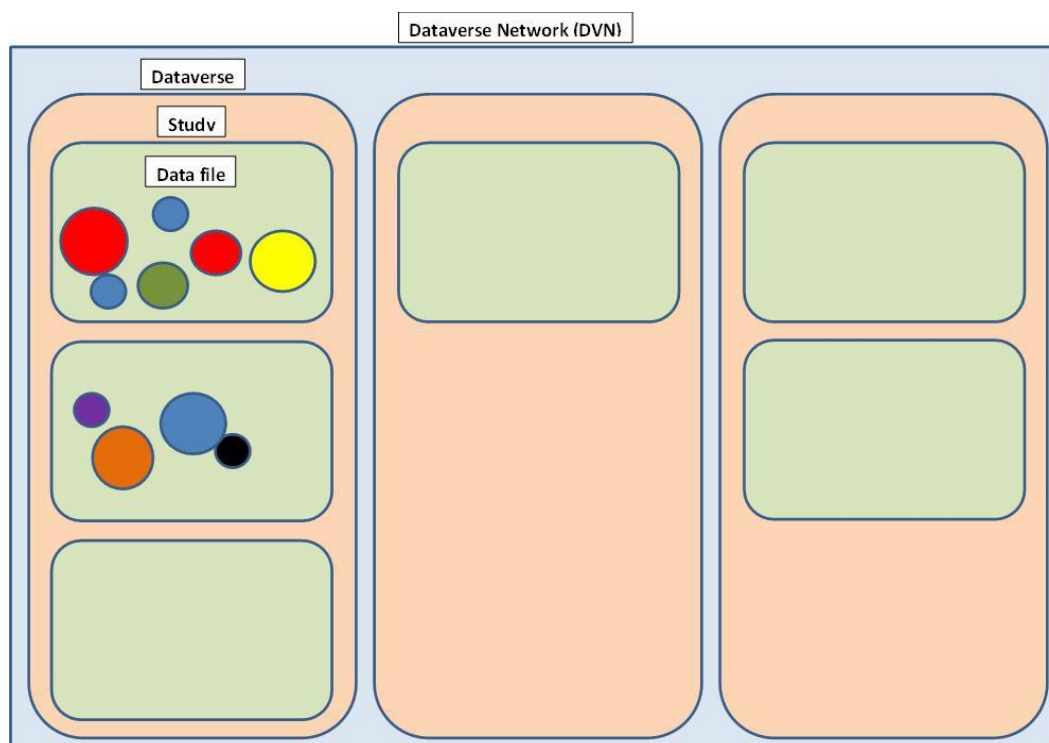
Figure 2: Situation with data files uploaded in a study and being part of a dataverse. This dataverse is one of the many dataverses in DVN.
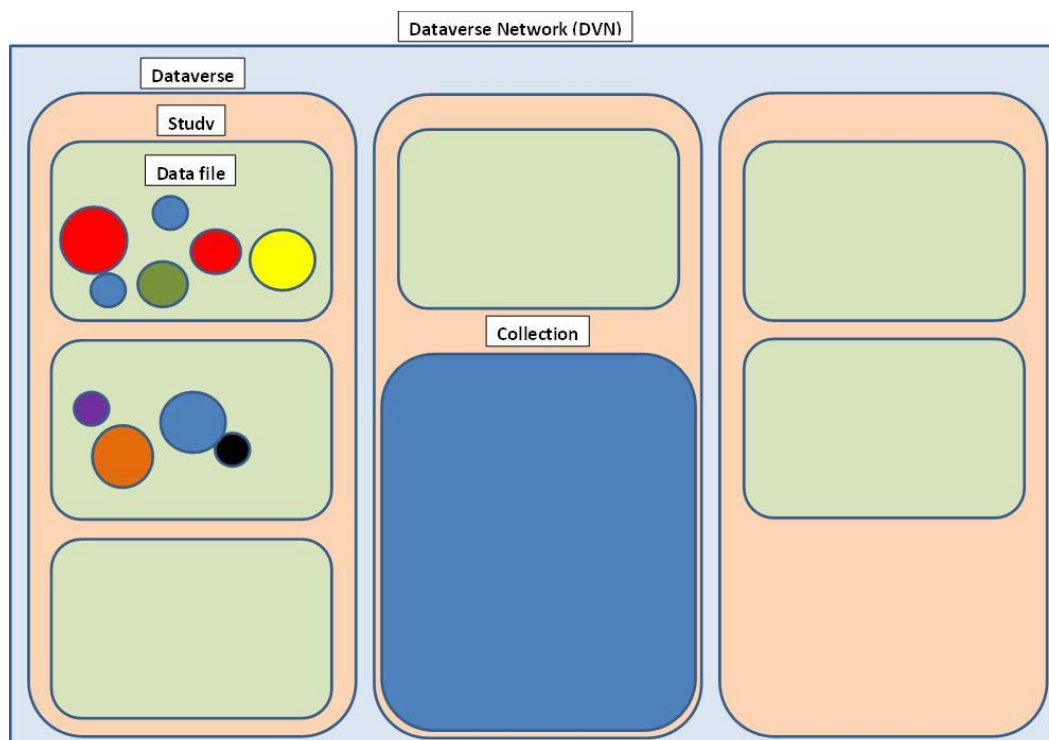

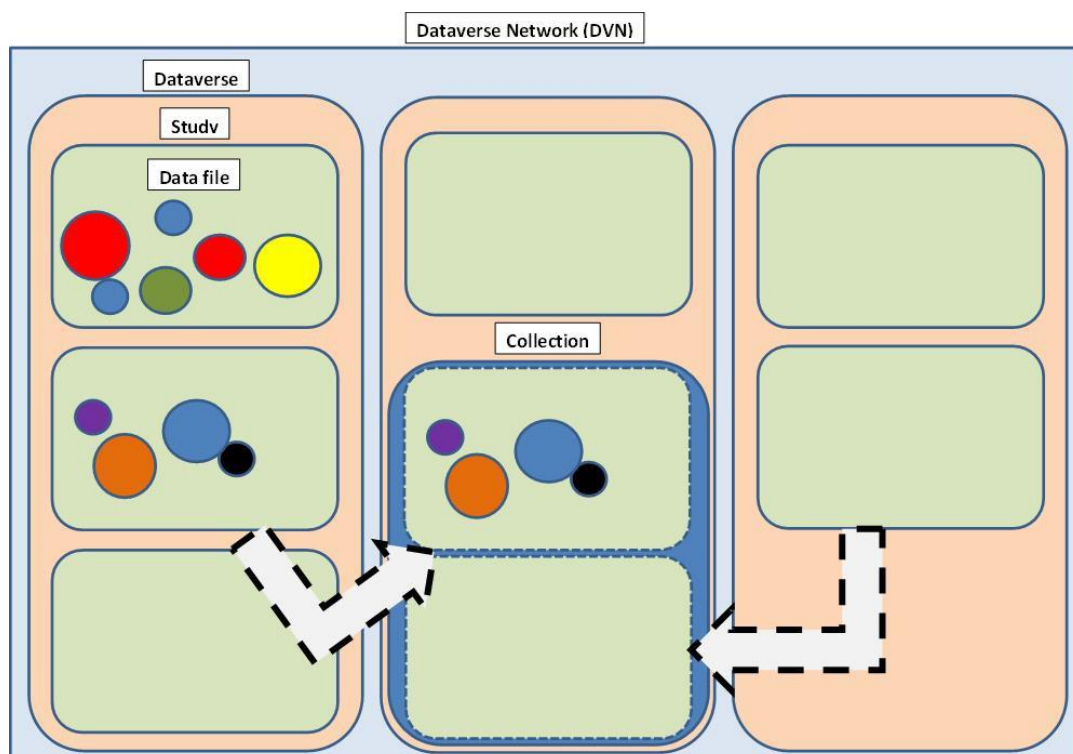**Figure 3: In another dataverse there is a collection created.**

Figure 4: This collection can contain studies from its own dataverse and from other dataverses to create a place where e.g. similar studies or studies of personal interest are "collected".
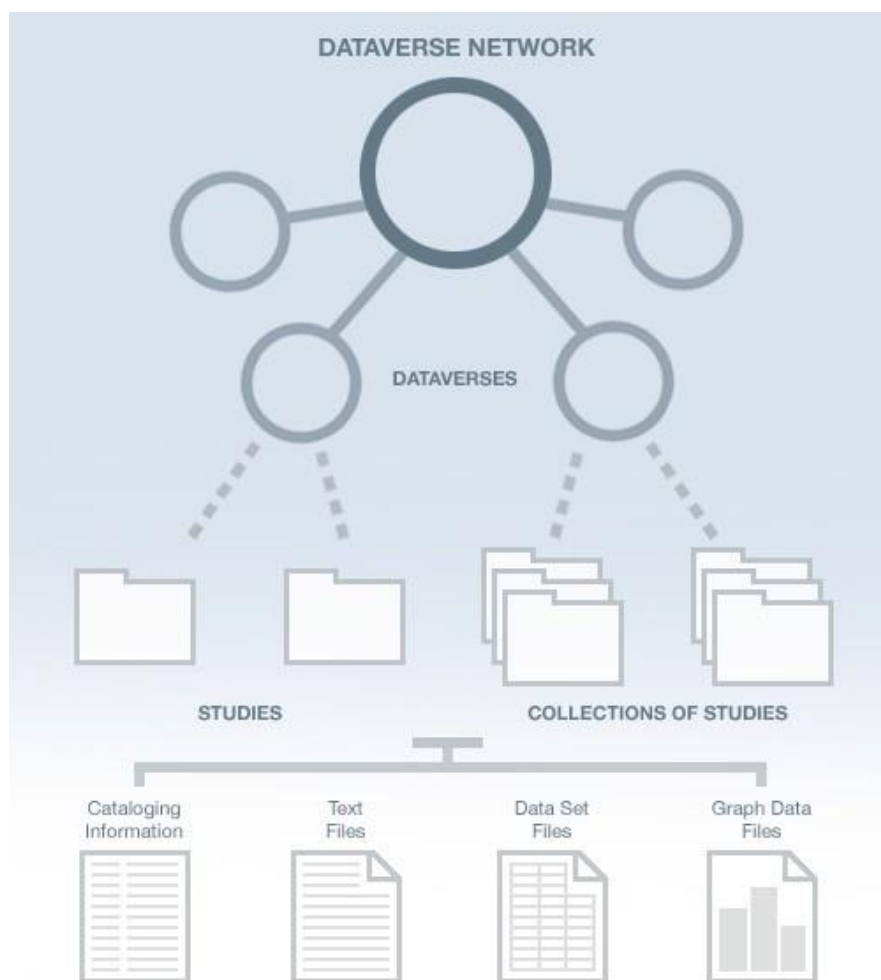

Figure 5: Hierachic structure of DVN

1.5 User Roles and Dataverse Access Permissions

To distinguish between permissions, there are different user roles in DVN: Admin, Curator, Contributor, Access Restricted Site. In the next sections the main standard permissions per user type are explained. Some permissions (*italic*) can be restricted by the (network) administrator.

1.5.1 Admin[6]

DVN: Set up and manage contributions to your dataverse, manage the appearance of your dataverse, organize your dataverse collections.

A user with Admin permission to the dataverse has all permissions at dataverse level and below. So, the Admin can:

- manage all permissions for users of the dataverse regarding the studies and data files it contains;
- create new users, studies and collections;
- manage settings, metadata and the appearance of the dataverse (release a dataverse) and studies (release or de-accession a study);
- add, delete and download data files;
- organize collections of the dataverse;

1.5.2 Curator

DVN: Summarize related data, organize data, or manage multiple sets of data.

A user with Curator permission to the dataverse has less permissions than the Admin at the dataverse level. He has the same permissions as the Admin at the study level and below. So, the Curator can:

- create new studies and collections;
- within studies manage all permissions regarding the data files it contains;
- manage the appearance of studies (release or de-accession a study);
- add, delete and download data files;
- organize collections of the dataverse;
- manage metadata of the studies;

1.5.3 Contributor

DVN: Distribute data and receive recognition and citations to it.

A user with Contributor permission to the dataverse has less permissions than the Admin and Curator on dataverse and study level. So, the Contributor can:

- create new studies;
- contribute to self-created studies (permission to contribute to all studies in dataverse is optional);
- submit self-created studies for review to Curator/Admin for release (permission to submit all studies in dataverse is optional);
- add, delete and download data files;
- manage metadata of the studies;

1.5.4 Access Restricted Site

DVN: Download and analyze all types of data.

An Access Restricted Site user can, after getting granted explicit permission of the Admin, only access a study to see and download research data files. This user cannot make any adjustments to the

---

[6] Not to be confused with the Network Administrator who gives access to DVN.

study, or release or submit the study for review. This is the appropriate role for a user who has no part in the research, but only wants to use the research data.

1.6 Providing metadata

DVN provides a wide range of different metadata fields[7] for the study, of which only the Title field is always required. For publishing, sharing, referencing, extracting and analyzing research data, it is important to provide sufficient metadata with each study. Moreover, providing these is a necessary condition for users to find the study by using the search function (see Section 1.7). The more and better these metadata are provided, the better the study is findable for other users.

1.6.1 Templates

A template is a tool to help to reduce the work needed to add a study, and to apply consistency to studies within a dataverse. For example, you can create a template to prepopulate any of the Cataloguing Information fields of a new study with default values, so that every study has the same values for that metadata, or enforce filling in specific required fields in a certain format. Every dataverse has its own default study template, but a user who adds a new study can select any template. Hence, although you can define a default template for a dataverse, it is not possible to enforce the use of it.[8]

1.7 Searching studies in DVN

To find studies in DVN, it has an extensive search function, which is also available without the need to login. Only released studies in released dataverses are included in the search results.

Due to access restrictions it might be that:
* Studies appear in the search results with only the study title and a red lock icon next to it. Without explicit access permission these studies are not accessible. If you have access permission for this study, you need to be logged in.
* After accessing a study, the research data files are visible (file names only), but not available for download. If the owner has the option enabled, it is possible to send an email for access permission to these files.
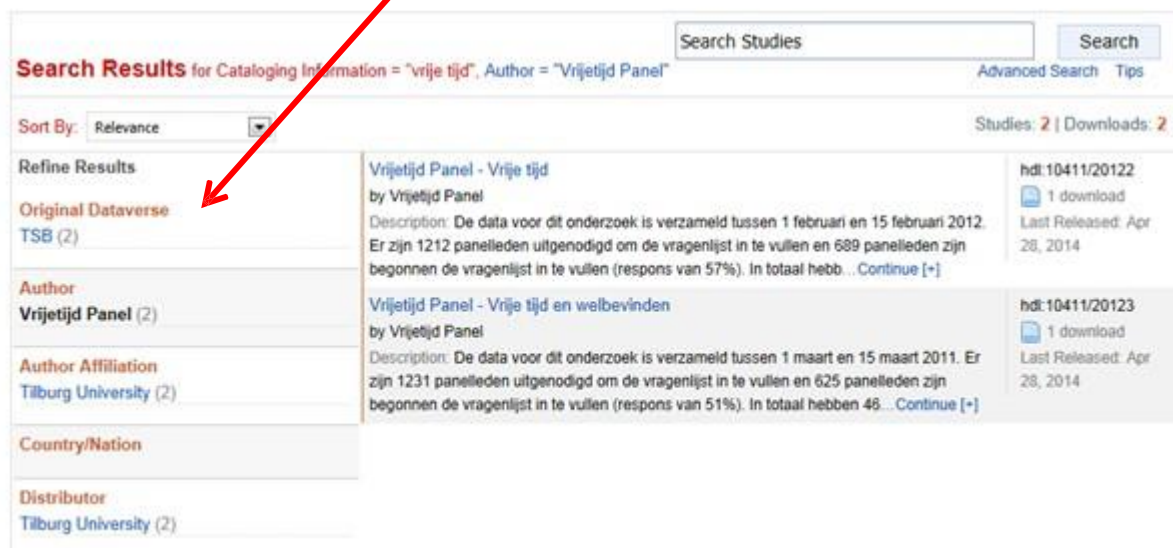
There are two methods for searching: the "Basic Search" and the "Advanced Search". Both methods can be found on the top-right of the network homepage. With the "Advanced Search" you can be more specific on which metadata fields to search in. A very useful option in both methods is to refine your search by using facets (see Figure 6).

Search terms are not case sensitive, i.e. it is ignored whether the search term is in uppercase or lowercase. Though, terms need to be exact, i.e. parts of a word, wildcards (*) or approximations do not work. In addition, the default search operator is "AND". Therefore, when searching with more search terms, it can be an effective strategy to first search on one search term and then use the facets to refine.

---

[7] For an overview of all fields see appendix 5.2.
[8] As mentioned, the Dataverse administrator can create a template with required and recommended fields, hide some of the available fields, or already fill some fields with pre-defined values. So, it might be that not all fields are available when creating a new study based on the template.

Figure 6: Refining search results by using facets. The original search term was "vrije tijd". Subsequently, the search results were refined by using the facet for author.

1.7.1 Basic search

It is possible to perform a Basic search from the network homepage. Simply type a search string into the search field on the top-right of the network homepage and click "Go". This method supports searching in any metadata field of the studies. As mentioned, it is possible to refine the results by using the facets.

1.7.2 Advanced search

In an advanced search, you can refine your criteria by choosing in which metadata fields to search. You can also apply logic to the field search, e.g. for text fields you can specify that the field either "contains" or "does not contain" the text you enter. With the "Advanced Search" method you can search the study metadata fields mentioned below. Other fields cannot be searched separately.

- **Title** – Title of study.
- **Author** – Information in any of the Author fields (Name, Affiliation) of study.
- **(Study) Global ID** - ID assigned to study.
- **Other ID** - A different ID previously given to the study by another archive.
- **Description** - Any words in the description of the study.
- **Keyword** - A term that defines the nature or scope of a study. For example, elections.
- **Topic Classification** - One or more words that help to categorize the study.
- **Producer** - Institution, group, or person who produced the study.
- **Distributor** - Institution that is responsible for distributing the study.
- **Production Date** - Date on which the study was created or completed.

- **Distribution Date** - Date on which the study was distributed to the public.
- **Time Period Cover Start** - The beginning of the period covered by the study.
- **Time Period Cover End** - The end of the period covered by the study.
- **Country/Nation** - The country or countries where the study took place.
- **Geographic Coverage** - The geographical area covered by the study. For example, North America.
- **Universe** - Universe of interest, population of interest, or target population.
- **Kind of Data** - The type of data included in the file, such as survey data, census/enumeration data, or aggregate data.
- **Publication, Replication for**
- **Related Publications**
- **Variable Information** - The variable name and description in the studies' data files, given that the data file is subsettable and contains tabular data. It returns the studies that contain the file and the variable name where the search term was found.[9]

1.7.3 Searching within own dataverse or collections

The "Basic" and "Advanced" search methods are, with the same rules and restrictions as mentioned above, also available to search for studies only within collections in your own dataverse. Simply open your dataverse or a lower level collection and find the search options on the top-right of the page. Click the checkbox "within this collection" with the "Basic Search" or "Search Only the Selected Collections" with the "Advanced Search".

---

[9] This search field is of no value in the current version of DVN (3.6.2), because it does not support the analysis of tabular data files.

2 Decision tree

In order to decide on the different settings to be made regarding accessibility of data you upload to DVN, we developed a decision tree as a tool to assist. By simply answering the questions on your way through the tree, you receive the configuration that should be made with respect to the dataverse, study and data files, eventually accompanied with side-implications of this configuration. A preview of the decision tree is presented below in Figure 6; a full-size version will be provided separate from this report and is available via this link: Decision tree.
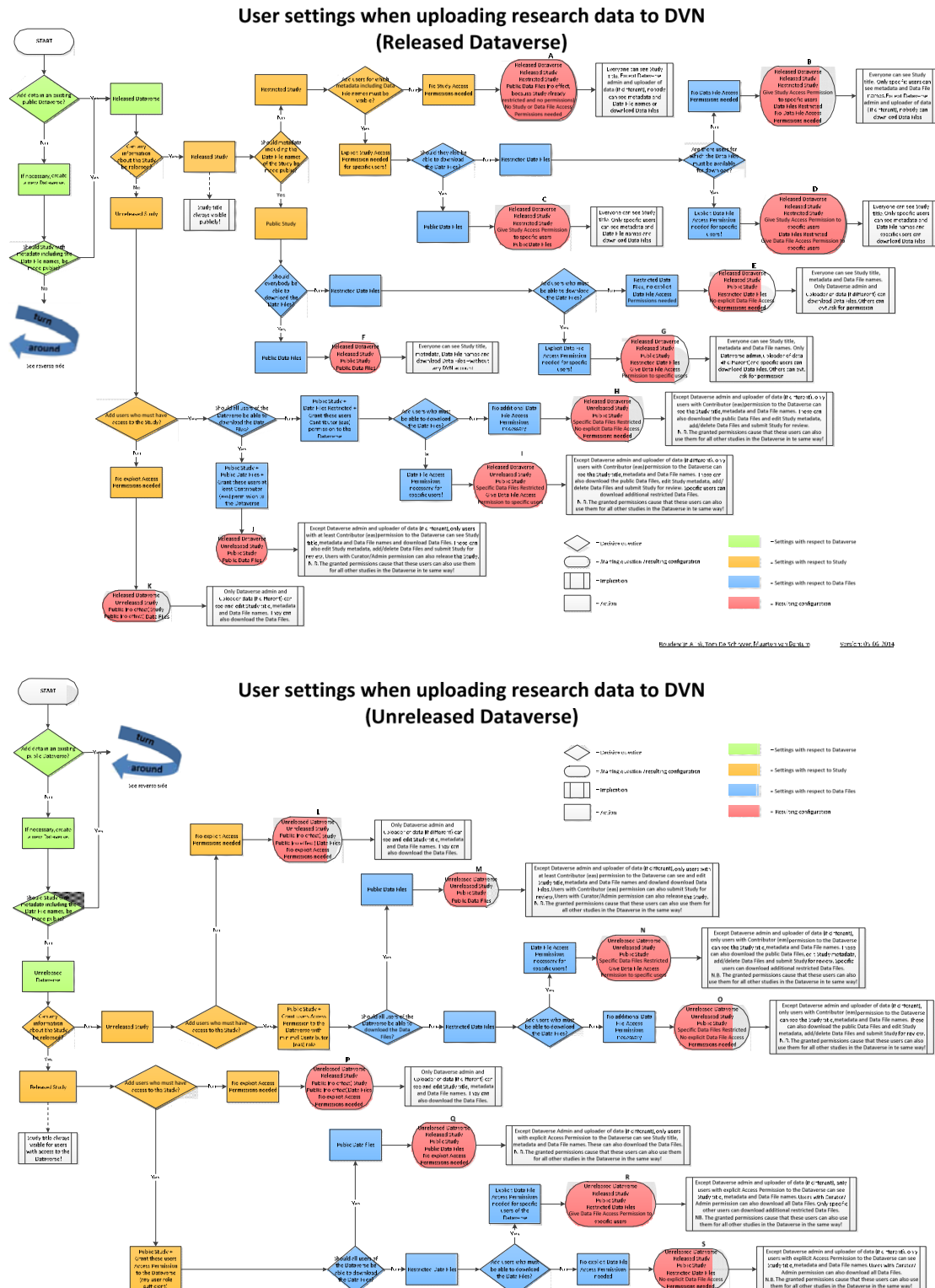

Figure 7: Decision tree to determine access permission settings for dataverse, study and data files

3   Use Cases

Now we know what DVN is and what it can be used for, we will have a look at how to use it. By elaborating on several practical situations and playing with some variables, the different configurations in DVN will become clear. The following use cases elaborate on settings to be made regarding the dataverse, studies and data files and its implications.

If we assume that, as in the use cases below, not the PhD-students themselves will upload the data, but the professor or someone else who is the admin of the dataverse on behalf of the department, the admin can determine the settings that need to be made in DVN to accomplish the intended situation by using the decision tree of Chapter 2.

3.1   Two PhD-students in same department must store and share research data

3.1.1   Description
Think of a situation with a professor (P1) and two PhD-students (S1 and S2). The PhD-students both performed their own study. When they have finished their study, they are asked by their professor to upload the research data to the department's dataverse in DVN.

3.1.2   Requirements:
- The professor, as well as S1 and S2, should be able to view and download the research data files of both studies;
- Research data of S1 should be publicly available;
- Research data of S2 should only be available for the P1, S1, S2 within the department.

3.1.3   Dataverse settings
While all studies of the department will be placed in the same dataverse, and the study of S1 is intended to be publicly available, this leads us to a Released Dataverse.

3.1.4   Study settings
Since the studies of S1 and S2 are finished, after uploading the data files, the studies can be released.

Hence, the decision must be made whether, besides the title, the other metadata and file names may be public (even visible without a DVN account). Because for the study of S1 the metadata and file names must be public, the study should be set as a public study. Now, the study settings for the study of S1 are complete.

For the study of S2 it was determined that metadata and file names should only be available for a restricted group of users. This means that, as we can see in the decision tree, the study of S2 should be set as a restricted study. Because the research data of S2 must though be available for P1, S1 and S2, access permissions to the study are required for them. The admin of the dataverse should give these users explicit access permission to the study in the study settings. Now, the study settings for the study of S2 are complete.

3.1.5   Data File settings
After making the right study settings, it must be set whether the data files are available for download and for which users.

The situation created above for the study of S1 is that all data files are publicly available for download (even without a DVN account). This is in accordance with the requirements, so no additional access restrictions regarding the availability of data files are required.

For the study of S2, in the situation created above all data files are available for download for P1, S1 and S2. This is also in accordance with the requirements, so no additional access restrictions regarding the availability of data files are required.

3.1.6   Resulting configuration
Following the branches of the decision tree, this leads to configuration **F** for the study of S1 and configuration **C** for the study of S2.

3.1.7   Implications
The implication of the selected configuration for the study of S1 is that: Everyone can see the study title, metadata, Data File names and download Data Files – even without any DVN account.

Since this is what was intended, there seems to be no problem. It is only important to realise!

The implication of the selected configuration for the study of S2 is that: Everyone can see study title, but only specific users (P1, S1 and S2) can see other metadata and data file names, and also download the data files.

Since this is what was intended, there seems to be no problem. It is only important to realise!

3.2   Two PhD-students can only download own data files, and need to ask for permission before downloading other's data files

3.2.1   Description
Think of the situation of use case 3.1, with a professor (P1) and two PhD-students (S1 and S2). When making a slight adjustment in the requirements, a different configuration is needed. Again, the PhD-students both performed their own study. When they have finished their study, they are asked by their professor to disclose their research data by uploading it to the department's dataverse in DVN. However, in this use case they can only download research data files from their own study. If they want to download a research data file from the other study, they first have to ask for permission.

3.2.2   Requirements
-   The professor should be able to view and download the research data files of both studies;
-   S1 and S2 should be able to view and download the research data files of their own studies;
-   S1 and S2 should be able to access each other's study, but not download the research data files;
-   When S1 wants to download a research data file of S2, or vice versa, he needs to ask for permission.

3.2.3   Dataverse settings
There is no change in this setting, because still all studies of the department will be placed in the same dataverse and we assume that at least one other study in the department's dataverse is intended to be public, so this leads us to a Released Dataverse.

3.2.4   Study settings
Again, since the studies of S1 and S2 are finished, after uploading the data files, the studies can be released.

Hence, the decision must be made whether, besides the title, the other metadata and file names may be public (even visible without a DVN account). This is the point where the configuration first becomes different from use case 3.1!

Both studies should not be public, but restricted, because only a restricted group of users should be able to access the studies. Because P1, S1 and S2 must be able to access both studies, access permissions to the study are required for them. The admin of the dataverse should give these users explicit access permission to the study in the study settings of both studies. After doing this, the study settings for the studies of S1 and S2 are complete.

3.2.5 Data File settings
After making the right study stings, it must be set whether the data files are available for download and for which users. The requirements regarding downloading the data files are also different with respect to use case 3.1!

In the situation created above for the studies of S1 and S2 all users who have access to a study, which are for both studies P1, S1 and S2 in this case, can also download the data files of that study. However, it is required that besides the professor (P1), S1 and S2 can only download the data files from their own study! To accomplish this, the admin of the dataverse must set the data files to restricted. Thereafter, the admin should in the study of S1 grant permission to P1 and S1 for each data file, and in the study of S2 grant this permission to P1 and S2.

If S1 or S2 wants to download a file from a study that is not his own study, he first has to ask for permission. To enable a request for permission to a restricted file, the admin of the dataverse must check the option to "Allow users to request access to restricted files for this study" in the data file settings. Now the implementation is in accordance with the requirements of this use case.

3.2.6 Resulting configuration
Following the branches of the decision tree, this leads to configuration **D** for both studies.

3.2.7 Implications
The implication of the selected configuration for the studies of S1 and S2 is that: Everyone can see study title, but only specific users (P1, S1 and S2) can see other metadata and data file names. In addition, only specific users (P1 and S1 for study of S1; P1 and S2 for study of S2) can download the data files.

Since this is what was intended, there seems to be no problem. It is only important to realise!

3.3 Two PhD-students in same department work together on research and must both upload their research data

3.3.1 Description
Again think of the situation of use case 3.1, with a professor (P1) and two PhD-students (S1 and S2). Now, S1 and S2 work together on a study, which ultimately must be made publicly available, including its research data. Meanwhile they have to upload their research data themselves, as soon as they have it, in the department's dataverse in DVN. During the research, the professor and both students must be able to access the study and download the data files.

3.3.2 Requirements
- The professor, S1 and S2 should be able to view and download the research data files of the study;
- S1 and S2 should be able to upload their research data files to the study;
- When the study is finished, the study, including the research data files, must be publicly available.

3.3.3 Dataverse settings
While the study must ultimately be made publicly available, this leads us to a Released Dataverse.

In addition, because P1, S1 and S2 must be able to upload the research data into an unreleased study (see 3.4.4 Study settings), they should be granted permission on the dataverse with at least the contributor role, together with the setting that contributors can create and edit all studies in the dataverse.

3.3.4 Study settings
Since the study of S1 and S2 is not finished yet, the study must stay unreleased.

At this point the decision must be made whether there should be users who have access to the study. This should be the case for P1, S1 and S2. Therefore, additional permission settings should be made in the dataverse settings (see 3.4.3 Dataverse settings). Because the study should ultimately be made publicly available, it should be set as a public study. With this the study settings are complete.

3.3.5 Data File settings
After making the right study settings, it must be set whether the data files are available for download, and for which users.

Because P1, S1 and S2 are in this use case the only users of the dataverse, there are no restrictions necessary regarding data file access. The uploaded data files can be set to public, also taking into account that it is defined that the study ultimately must be public, and the data files must be publicly available for download. With this setting the data file settings are complete.

3.3.6 Resulting configuration
Following the branches of the decision tree, this leads to configuration **J** for this study.

3.3.7 Implications
The implication of the selected configuration is that besides the admin of the dataverse, all users with at least the contributor role on the dataverse (and the edit all studies setting), can see study title, metadata and data file names and download data files. They can also edit study metadata, add/delete data files and submit the study for review. As long as P1, S1 and S2 are the only users within the dataverse, this seems to correspond with the intended situation, so there is no problem. However, it should be noticed that S1 is able to edit the study metadata and delete the data files of S2 from the study, and vice versa.

As soon as there will become other users within the dataverse with these roles, e.g. for other studies, they will also have the same access permissions on this study! A solution can be to set the study or data files to restricted and grant the applicable users explicit permission to the study for the time that the study is not finished. As soon as the study is finished, this setting must be undone, because after the release, the study and data files must be made publicly available!

4 User Manual

When you would like to store your research data within DVN, this user manual guides you through using DVN, step by step. From the log in procedure to creating your dataverse, studies, collections, and uploading data files.

4.1 How to Log in?

After opening the website www.dataverse.nl, there are two ways to log in to DVN.
- Federated log in - Most common way to log in (for staff and students).
- DVN account - For users with a guest account for DVN *and for the Network Administrator*.

4.1.1 Federated log in

Log in via the "Log in"-button on the top-right of the Homepage of DVN. Log in with your m- or s-account details of the University of Twente. On first log in share your data with DVN (does not mean your data becomes public).

4.1.2 Guest account

Log in via the "Log in with DVN account"-button on the top-right of the Homepage of DVN. Log in with received/created log in details. Change your password on first login.

4.2 Dataverses

4.2.1 Edit a dataverse

After creating a dataverse, permissions and settings can be changed by users granted Admin permissions to the dataverse. Also studies, collections and templates can be created.

- When logged in, click on your account name on the top-right of the page.
- On the tab "Dataverses" search the applicable dataverse to edit.
- Click on the Edit Settings-button [⚙] in the Actions column.

The 'control panel' of your dataverse appears. From this 'control panel' you are able to manage your studies, collections, templates, permissions and settings of the dataverse via the applicable tabs.

4.2.2 (Un)Release a dataverse

A dataverse is created as an "Unreleased dataverse", so it is still not publicly accessible. To make it public, the dataverse must be released via the dataverse settings tab.

- When logged in, click on your account name on the top-right of the page.
- On the tab "Dataverses" search the applicable dataverse to edit.
- Click on the Edit Settings-button [⚙] in the Actions column.
- Select the Settings tab, change the Dataverse Release Settings from "Not Released" to "Released", or vice versa, and click "Save".

4.2.3 Change access permissions for a dataverse

To give users a specific user role with corresponding permission on the dataverse, it is necessary to define them explicitly via the dataverse permissions tab.

- When logged in, click on your account name on the top-right of the page.
- On the tab "Dataverses" search the applicable dataverse to edit.
- Click on the Edit Settings-button [⚙] in the Actions column.
- Select the Permissions tab.

- In the User Permissions section enter the applicable Username and Permission Setting and click Add.
  - Don't forget to click on Add! Otherwise the settings will not be saved!
  - It is necessary to enter the Username in the correct spelling (case sensitive!)
- Click "Save Changes" to apply the new permissions.

4.3 Studies

4.3.1 Create a study
When a dataverse already exists, a new study can be created.

- When logged in, click on your account name on the top-right of the page.

- On the tab "Dataverses" click on the Create Study-button to create a new study.
  - If the screen with "Terms of Use" appears, read the terms, select "I agree and accept these terms of use" and continue.
- Fill out at least the required fields (with red *) and click "Save"
- Your study is ready for use and has an own unique Handle (=persistent URL) for use in publications.

N.B. The study is created as an "Unreleased study", so it is still not publicly accessible. To make it public, the study must be released via the "Release" button in the Study editing screen, see "Release a Study".

4.3.2 Edit study settings
You can change the study settings of a draft version of a study.

- When logged in, click on your account name on the top-right of the page.
- On the tab "Studies" search the applicable study and click on it to edit.
- See the different editing options on the top-right of the screen.

4.3.2.1 Cataloguing Information
- Click on the Edit Cataloguing Information-button to adjust the metadata of the study.
  - If the screen with "Terms of Use" appears, read the terms, select "I agree and accept these terms of use" and continue.
- The screen with the Cataloguing Information appears.
- Make the necessary changes and click on "Save".

4.3.2.2 Permissions
To define whether a study should be publicly accessible or restricted, and to grant permissions on restricted studies, select the options for permissions.

- Click on the Permissions-button to adjust the permissions for the study.
  - If the screen with "Terms of Use" appears, read the terms, select "I agree and accept these terms of use" and continue.
- The screen with the Permission settings appears.
- Select whether the study should be "Public" or "Restricted".
- When "Restricted" is selected type a specific username for who must have access to the study in the applicable field and click on "Add".
- Click on "Save".

4.3.3 Release a study
When a study is ready to publish the study can be made permanent by releasing the study.

- When logged in, click on your account name on the top-right of the page.
- Now there are two ways to release a study:
    1. On the tab "Studies" search the applicable study and click on the Release-button in the Action column.
    2. Click on the study name and subsequently click on the Release-button on the top-right of the screen.
- Fill in your study version notes, if applicable, and click on "Save".

4.3.4 Deaccession a study

In case it is necessary to withdraw an already released study, it is not possible to delete, but only to deaccession the study. The study will remain in Archive.

- When logged in, click on your account name on the top-right of the page.
- Now there are two ways to deaccession a study:
    o On the tab "Studies" search the applicable study and click on the Deaccesion-button in the Action column.
    o Click on the study name and subsequently click on the Deaccesion-button on the top-right of the screen.
- Fill in your comments and a link a to refer to another study, if applicable, and click on "Save".

4.4 Data Files

4.4.1 Upload a Data File

After creating a study, data files can be uploaded in this study.

- When logged in, click on your account name on the top-right of the page.
- On the tab "Studies" select the applicable study to edit
- On the top right of the page click on "Add File(s)"
    - If the screen with "Terms of Use" appears, read the terms, select "I agree and accept these terms of use" and continue.
- In the drop-down menu select the Data Type "Other"
- Click on "Browse…" to select the file to upload
- Click "Save" to confirm the upload.

4.4.2 Edit file information

- To edit the information about already uploaded files to the study click on the Edit/Delete File+Information-button.
- Change the existing information about the files and click on "Save".

4.4.3 Delete uploaded files from study

- To edit the information about already uploaded files to the study click on the Edit/Delete File+Information-button.
- Check the checkboxes of the files you want to delete from the study
- Click on "Save" to delete the files from the study and save your changes.

4.4.4 Permissions

To define whether data files should be publicly accessible or restricted, and to grant permissions on restricted data files, select the options for permissions.

- When logged in, click on your account name on the top-right of the page.
- On the tab "Studies" search the applicable study and click on it to edit.
- Click on the Permissions-button on the top-right of the screen to adjust the permissions for the data files.
    - If the screen with "Terms of Use" appears, read the terms, select "I agree and accept these terms of use" and continue.
- The screen with the Permission settings appears.
- Click on the "Files" tab for the permission settings on the data files.
- Select the checkboxes in front of the files for which you want to change the permissions.
- Select the new permission setting from the File Permission-dropdown menu and click "Update Permissions".
- Click on "Save".

4.4.4.1   *Grant access to restricted data files*

- Select the checkboxes in front of the files for which you want to change the permissions.
- Enter the username of a specific user who must have access to the data files in the applicable field and click on "Grant Access".
- Click "Save" to save your changes.

4.5   Collections

4.5.1   Create a static Study Collection
- When logged in, click on your account name on the top-right of the page.
- On the tab "Dataverses" find the applicable dataverse to edit.
- Click on the Edit Settings-button [⚙] in the Actions column.
- Select the tab "Collections".
- Click on "Create Study Collection".
- Select the type "Static".
- Give the collection a name.
- Select the collection that will be the parent of this new Study Collection.
- Start selecting studies in the pane "Studies to choose from".
    - You can choose from any Released dataverse in DVN.
- After your selection is complete click on "Save" to create your new collection.

4.5.2   Create a dynamic Study Collection
- When logged in, click on your account name on the top-right of the page.
- On the tab "Dataverses" find the applicable dataverse to edit.
- Click on the Edit Settings-button [⚙] in the Actions column.
- Select the tab "Collections".
- Click on "Create Study Collection".
- Select the type "Dynamic".
- Give the collection a name.
- Enter a query to define the selection criteria for your collection.
- Select the scope for the query.
- Click on "Save" to create your new collection.

4.5.3   Create a Linked Collection
- When logged in, click on your account name on the top-right of the page.
- On the tab "Dataverses" find the applicable dataverse to edit.

- Click on the Edit Settings-button ⚙ in the Actions column.
- Select the tab "Collections".
- Click on "Create Linked Collection".
- Select the dataverse to which the existing collection belongs.
- Select the collection you want to link.
- Click on "Save" to create your new collection.

4.6 Study templates

4.6.1 Create a study template
- When logged in, click on your account name on the top-right of the page.
- On the tab "Dataverses" find the applicable dataverse to edit.
- Click on the Edit Settings-button ⚙ in the Actions column.
- Select the tab "Templates", the DVN default template and other already available templates show up.
- Click on "Clone" behind the template you want to use a basis for your new template.
- Give the template a "Template Name" and eventually a "Template Description"
- Now you can select on the right of each field if it is "Required" (with red *), "Recommended" (with green *), "Optional" or "Hidden" when a new study is created.
- When desired, it is also possible to fill out a standard value in the fields of the user form, which is already pre-entered when a new study is created.
- Click on "Save" to store your new template.
- To make the template available for use, click on "Enable".

4.6.2 Set default study template
- When logged in, click on your account name on the top-right of the page.
- On the tab "Dataverses" find the applicable dataverse to edit
- Click on the Edit Settings-button ⚙ in the Actions column.
- Select the tab "Templates", the DVN default template and other already available templates show up.
- Click on "Make Default" behind the template you want to use a the default template.
- Click on your account name on the top-right of the page or the Edit Settings-button ⚙. Otherwise your changes are not saved!

4.6.3 Enable/Disable a study template
Note: a template can only be disabled when it is not the default template!

- When logged in, click on your account name on the top-right of the page.
- On the tab "Dataverses" find the applicable dataverse to edit.
- Click on the Edit Settings-button ⚙ in the Actions column.
- Select the tab "Templates", the DVN default template and other already available templates show up.
- Click on "Enable" ("Disable") behind the template you want to enable (disable).

4.6.4 Edit a study template
- When logged in, click on your account name on the top-right of the page.
- On the tab "Dataverses" find the applicable dataverse to edit.
- Click on the Edit Settings-button ⚙ in the Actions column.

- Select the tab "Templates", the DVN default template and other already available templates show up.
- Click on "Edit" behind the template you want to edit.

4.6.5 Delete a study template

Note: a template can only be deleted when it is not the default template!

- When logged in, click on your account name on the top-right of the page.
- On the tab "Dataverses" select the applicable dataverse to edit
- Click on the Edit Settings-button in the Actions column.
- Select the tab "Templates", the DVN default template and other already available templates show up.
- Click on "Delete" behind the template you want to delete.

5 Appendices

5.1 Appendix – Additional features of DVN

5.1.1 Analyze data within DVN

DVN claims to enable users to analyze and subset data within the application, even without requiring the user to have a license on the original software of the data file. Actually, although this was the case in previous versions, because of a bug this feature of DVN is not present in the current version. It might be that it will return in future versions.

The intention of this feature is:

A study might contain documentation, data, or other files. Tabular and Network Data files can be subset and analyzed using the Dataverse Network analysis tools. When the study contributor uploads data files of the supported file types to the Network, those files are converted to *.tab tab-delimited files. These *.tab files are subsettable, and can be subsetted and analyzed online by using the Dataverse Network application.

Data files of the type .xml also are considered to be subsettable, and can be subsetted and analyzed to a minimal degree online. An *.xml type file indicates social network data that complies with the GraphML file format.

5.1.1.1 Supported tabular/subsettable file types

The supported file types are *.dta (STATA), *.sav, or *.por (SPSS), and *.xml (GraphML). These files are processed as subsettable data files, which can be analyzed online by using the Dataverse Network analysis tools.

5.1.2 Guestbook, Download Tracking Data and other statistics

Interesting features of DVN for the dataverse administrator are the Guestbook and the Download Tracking Data, which gives him information about the use of his research data files by others. The network administrator of DVN can implement another feature to draw more statistics about the use of DVN, and the dataverses it contains, by means of using Google Analytics.

5.1.2.1 Guestbook

The administrator of a dataverse can enable the guestbook feature for the dataverse to collect information on all users before they can download or subset contents from the dataverse. Once it has been enabled it will be shown to any user for the first file a user downloads from a given study within a single session. If the user downloads additional files from the study in the same session, a record will be created in the guestbook response table, using data previously entered.

Which information is requested from the user can be modified by adding custom questions. Information entered in the guestbook can be viewed and downloaded by the dataverse administrator in the "Download Tracking Data" section.

5.1.2.2 Download Tracking Data

Besides the guestbook, data will be collected silently based on the logged-in user or anonymously. The data displayed includes user account data or the session ID of an anonymous user, the global ID, study title and file name of the file downloaded, the time of the download, the type of download and any custom questions that have been answered. A comma separated values file of all download tracking data may be downloaded by clicking the Export Results button.

5.1.2.3 Google Analytics

On the level of the network administrator, DVN provides the option to compile and analyze site usage through Google Analytics. Any page access along with associated browser and user

information can be recorded by Google, after enabling the option and embed a small amount of code, provided by Google. Later analysis of this compiled access data can be performed using the Google Analytics utility.

5.2 Appendix – List of all study metadata fields

Production Place			
Software	Software	Software Version	𝐎
Funding Agency			
Grant Number	Grant Number		𝐎
	Grant Number Agency		
Distributor *	Distributor	Abbreviation	𝐎
	Affiliation		
	URL		
	(http://example-domain.edu/)		
	Logo URL		
	(http://example-domain.edu/image)		
Contact *	Contact	Affiliation	
	E-mail		
Distribution Date *			
	'OfYY or YYYY-MM or YYYY-MM-DD; AD or BC optional)		
Depositor			
Deposit Date *	1201+03-28		
	'OfYY or YYYY-MM or YYYY-MM-DD; AD or BC optional)		
Series	Series		
	Series Information		
Version	**Version**		
	Version Date		
	'OfYY or YYYY-MM or YYYY-MM-DD; AD or BC optional)		

Description and Scope

Description *	𝐎 Copying and pasting from a Word document can create errors when you save this page_	𝐎

Description

Description Date

'OfYY or YYYY-MM or YYYY-MM-DD; AD or BC optional)

Keyword *	Keyword Vocabulary		𝐎
	URL		
	(http://example-domain.edu/)		

Topic Classification*	Topic Classification	Vocabulary	**O**
	URL		
	(http://example-domain.edu/)		

Related Material*		**O**

Related Studies*		**O**

Other References		**O**

Time Period Covered - Start*

\fYYYY or YYYY-MM or YYYY-MM-DD; AD or BC optional)

Time Period Covered - End*

\fYYYY or YYYY-MM or YYYY-MM-DD; AD or BC optional)

Date of Collection - Start*

\fYYYY or YYYY-MM or YYYY-MM-DD; AD or BC optional)

Date of Collection - End*

\fYYYY or YYYY-MM or YYYY-MM-DD; AD or BC optional)

Country/Nation*

Geographic Coverage*

Geographic Unit*

Geographic Bounding Box

@ Use the CGA Geo-location Finder to find the coordinates_

West Longitude	East Longitude

North Latitude	South Latitude

Unit of Analysis

Universe *

Kind of Data *

Data Collection Methodology

Time Method

Data Collector

Frequency

Sampling Procedure

Major Deviations for Sample Design

Collection Mode

Type of Research Instrument

Data Sources

Origin of Sources

Characteristic of Sources Noted

Documentation and Access to **Sources**

Characteristics of Data Collection Situation

Actions to Minimize Losses

Control Operations

Weighting

Form fields shown:
- Cleaning Operations
- Study Level Error Notes
- Response Rate
- Estimates of Sampling Error
- Other Forms of Data Appraisal

Data Set Availability
- Data Access Place
- Original Archive
- Availability Status
- Size of Collection
- Study Completion

Terms of Use
- Confidentiality Declaration
- Special Permissions
- Restrictions
- Contact
- Citation Requirements
- Depositor Requirements
- Conditions
- Disclaimer

Other Information
- Notes
  - Note Type
  - Subject
  - Text

Save Cancel

5.3   Appendix – Open/Wiki Dataverse

5.3.1.1   Open Dataverse

In an Open dataverse anyone who has a DVN account can, without explicit access permission of the dataverse administrator, create and edit their own studies and submit them for review (to release). However, only the dataverse administrator and curator users of the dataverse can release these studies.

5.3.1.2   Wiki Dataverse

In a Wiki dataverse anyone who has a DVN account can, without explicit access permission of the dataverse administrator, not only create and edit their own studies and submit them for review (to release), but can also do this with all other studies in the dataverse. However, only the dataverse administrator and curator users of the dataverse can release these studies.

5.4   Appendix – Version management

A feature of DVN is automatic version management of studies, with preservation of data sets belonging to each version. This means that as soon as a study is released, the corresponding data set is made definitive for that version, i.e. the data set for this version no longer editable.

By editing the data set, e.g. adding, replacing or deleting a data file, a new draft version of the study and corresponding data set is created and its version number is automatically increased by 1. This version of the study and its data set is again editable as long as it is draft, i.e. not released again.

By releasing this draft study again, the version number is again increased by 1, the corresponding data set is made definitive, and the previous released version is stored as archived, but still accessible for users who got granted permission.

Also the unique handle corresponding with a study remains the same when new versions of a study are created. So, when already referred to the study with the unique handle, the new version of the study is still traceable in this way. In case of de-accessioning a study, it is possible to include a comment with a reason for de-accessioning and eventually a link to another applicable or replacing study.

Important to note is that versions and automatic version numbers are used for studies (data sets) and not for individual data files. When a new version of a data file must be uploaded, several options are available:
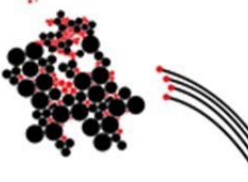1. Use a different file name for the newer version and add the new data file to the study
2. Delete the old version of the data file from the study and upload the new version with the same name
3. Release the study to fix its data set and keep the old version of the file in this version of the study. De-accession the study to prevent undesirable public access. Delete the original and edited data file from the study, and upload the new version.

In the situation of the first and third option the old version of the data file stays available within DVN. With the second option there stays no copy of the data file in archive for future use or reference.

Another important remark is that once a study is released, the set of data files belonging to it at that moment is made fixed, including all settings regarding permissions for downloading these files. This cannot be changed in the archived version with retroactive effect. So, deleting a data file in a later version of a study has no effect on the availability of this file in a previous version, and settings regarding permission to this file cannot be changed anymore after deletion, which can have positive as well as negative implications. After all, the data file shall always be downloadable, even if this is not the intention, because there might be good reasons for deleting the data file. At least, it might be desirable to be able to set the data file access permission to "Restricted", but this is not possible anymore.

5.5   Appendix – Slides workshop 12/6/2014

INLEIDING

- Dael:
    - Workshop geven
    - Informatiespecialisten beter beeld geven van mogelijkheden en beperkingen DVN
    - Rapport met bevindingen en handvatten
        - Beslisboom

WAT IS DVN?

- Open source application to ... research data:
    - Publish
    - Share
    - Reference
    - Extract
    - *Analyze*

WAT IN DVN?

- Data File
- Study
- Dataverse
- Collection

UNIVERSITY OF TW'ElflE.   SO<.dcWO'IM"lt   l:.'6.-:01'   S

WAT IN DVN?

- Gebruikersrollen
  - Admin
  - Curator
  - Contributor
    - Edit all/own studies
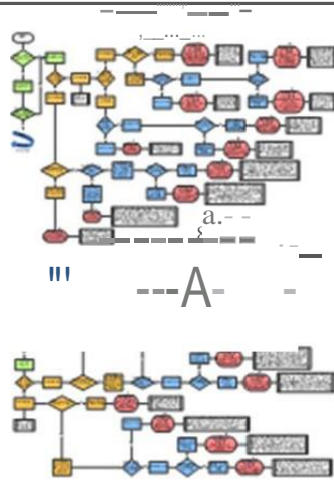 - Access Restricted Site

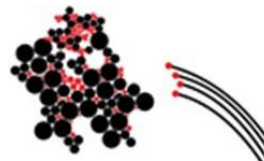UNIVERSITY OF TW'ElfTE.   "I'III"*   l:.oe.-:01'   6

BESLISBOOM

UNIVERSITY OF TWENTE .

WORKSHOP DVN

KORTE DEMONSTRATIE

- Homepage: www.dataverse.nl
- Inloggen
- Dataverses, studies, data files, collections
- Settings and permissions (gebruikersrollen)

ACCOUNTS AANMAKEN, INLOGGEN EN PERMISSIES

- Account aanmaken
- Voor het eerst inloggen
- Permissies geven op bestaande dataverse

HOE GEBRUIK JIJ DVN? (1)

- 2 PhD-students (S1 en S2), Professor (P1)

- S1 and S2 performed both their own study

- Professor requires that research data becomes available in department's dataverse in DVN

- Research data files of study of S1 must become publicly available

- Research data files of study of S2 must only be available for P1, S1 and S2

HOE GEBRUIK JIJ DVN? (2)

- 2 PhD-students (S1 en S2), Professor (P1)

- S1 and S2 performed both their own study

- Professor requires that research data becomes available in department's dataverse in DVN

- Research data files of study of S1 must become publicly available

- Research data files of study of S2 must only be available for P1, S1 and S2

- Give researcher from Japan permission to S2?

HOE GEBRUIK JIJ DVN? (3)

- 2 PhD-students (S1 en S2), Professor (P1)

- S1 and S2 performed both their own study

- Professor requires that research data becomes available in department's dataverse in DVN

- Research data files of study of S1 must only be available for P1 and S1 (student s2 can ask for permission!!!)

- Research data files of study of S2 must only be available for P1 and S2 (student s1 can ask for permission!!!}

UNIVERSITY OF TW'ElflE.   Boudewijn Alink   12/6/2014   13

HOE GEBRUIK JIJ DVN? (4)

- Voorstel?

UNIVERSITY OF TW'ElfTE.   Boudewijn Alink   12/6/2014   14

Reflectie / Vragen?