

---

# **Data Repository Documentation**

***Release 1.0***

**Data Repository support <[helpdesk@surfsara.nl](mailto:helpdesk@surfsara.nl)>**

**May 26, 2020**



---

## Contents

---

<b>1</b>	<b>General</b>	<b>3</b>
1.1	About the Data Repository . . . . .	3
1.2	How to get access . . . . .	4
<b>2</b>	<b>Basic Usage</b>	<b>7</b>
2.1	Basic Usage . . . . .	7
2.2	Deposit . . . . .	12
2.3	Downloads . . . . .	15
<b>3</b>	<b>Advanced</b>	<b>19</b>
3.1	Advanced Usage . . . . .	19
3.2	Account . . . . .	24
3.3	Objects . . . . .	26
<b>4</b>	<b>Best practices</b>	<b>27</b>
4.1	Best Practices . . . . .	27
<b>5</b>	<b>Appendix</b>	<b>29</b>
5.1	Appendix . . . . .	29





Welcome to the user documentation of the Data Repository service provided by [SURFsara](#). The information in this tutorial will help you get started with using the Data Repository, learn best practices and techniques for publishing your data and stay up-to-date with the service developments. We welcome your comments at [helpdesk@surfsara.nl](mailto:helpdesk@surfsara.nl) to help us improve the documentation.



## 1.1 About the Data Repository

In this page you will find general information about the Data Repository, how it works and how to get access:

### Contents

- *About the Data Repository*
  - *Introduction to Data Repository*
  - *For what data*

### 1.1.1 Introduction to Data Repository

Data Repository Service is a trusted digital repository that ensures long-term data preservation. The Data Repository provides an online data publication platform service to share and publish research data. The aim of the service is to provide support for researcher to publish research data and to preserve published datasets for the long-term, thereby satisfying the requirements of publishers, financiers and (Dutch) science organisations regarding data publication. Furthermore, the user is complete freedom to decide on the structure of the published data, attached metadata schemas, community and access levels.

There is a direct connection between the data repository service and other SURFsara services such as the Data Archive. The data that you publish in the data repository will eventually end up on the data archive for long-term storage and preservation. The repository is tape-aware, that means that data can automatically be put offline when it is not required for a certain amount of time.

This document is the user guide for basic and advanced usage of the data repository service and also provide best practice in preparing your data and using the service.

### 1.1.2 For what data

The Data Repository service is suitable for publishing research data especially in the field of Big Data sciences. The repository imposes no limits on the data volume and number of files, but does have some guidelines on how to efficiently store your data for publication.

## 1.2 How to get access

On this page you will find general information about getting access to the Data Repository service.

### Contents

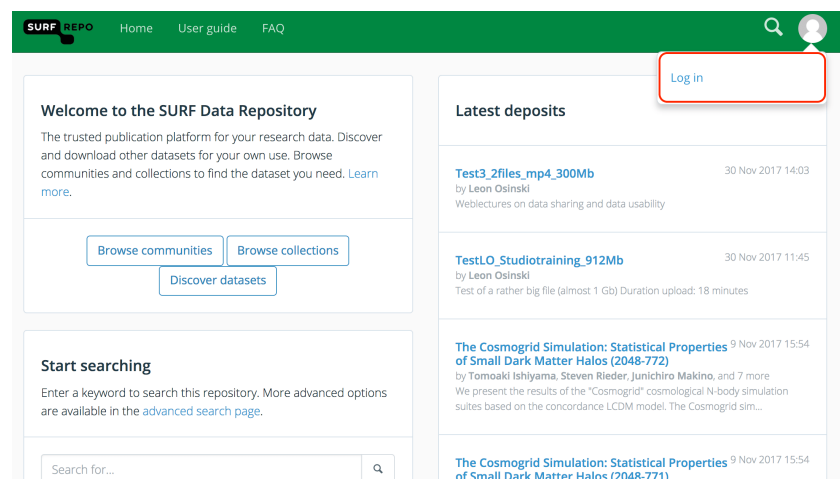
- [How to get access](#)
  - [General access](#)
  - [Obtain an account](#)

### 1.2.1 General access

The repository provides access to mostly open access data which can be found and downloaded by anyone interested without requiring registration.

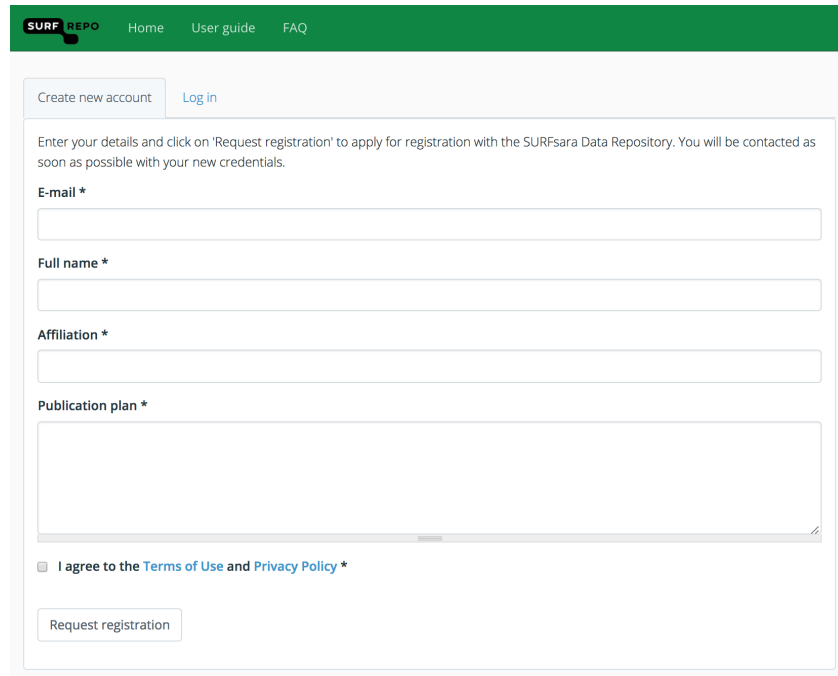
### 1.2.2 Obtain an account

To be able to administer publications and other objects in the repository service, you need to login as a registered user. To obtain user credentials you should first send a registration request via the Login page. Select 'Log in' from the top right menu under the avatar image.



Then click on **Request registration** tab. Enter your required information and agree to the **Terms of Use** and **Privacy policy** of the service. Then click on **Request registration** to apply for registration with the SURFsara Data Repository. You will be contacted as soon as possible by the SURFsara advisors. If your request is approved, your credentials will be sent to the email address you provided during the request.





The screenshot shows the SURF REPO registration page. At the top is a green navigation bar with the SURF REPO logo and links for Home, User guide, and FAQ. Below the navigation bar, there are two buttons: 'Create new account' and 'Log in'. The main content area contains a registration form with the following fields: 'E-mail \*', 'Full name \*', 'Affiliation \*', and 'Publication plan \*'. Each field has a corresponding text input box. Below the 'Publication plan \*' field is a checkbox labeled 'I agree to the Terms of Use and Privacy Policy \*'. At the bottom of the form is a 'Request registration' button.

An alternative is to send an email to [helpdesk@surfsara.nl](mailto:helpdesk@surfsara.nl) with a description of your request.



### 2.1 Basic Usage

This page provides information on the basic usage of the data repository, such as depositing, searching or downloading data. It also describes the online deposit workflow which are the steps that you take to deposit data into the Data Repository.

#### Contents

- *Basic Usage*
  - *Logging in*
  - *Homepage*
  - *Deposit landing page*
    - \* *Export Metadata*
  - *Collection landing page*
  - *Search and discover datasets*

#### 2.1.1 Logging in

Login with your user credentials in the [Login](#) page of the Data Repository Service.

If you do not have access, go to [How to get access](#) page for more information.

#### 2.1.2 Homepage

After you login to your account, you will be redirected to your homepage where you can quickly deposit data, find data to download or create collections of data.

### Note:

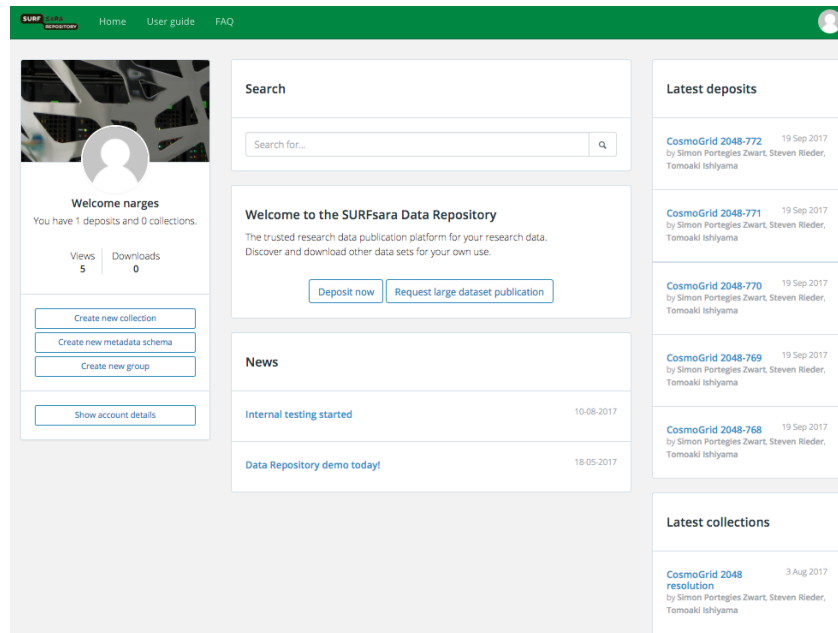
**Deposit** is an object containing one or more files and corresponding metadata

**Collection** is a bunch of deposits

On the right column you can see the latest deposits and collections that are uploaded on the data repository.

On the top left you see some statistics of your account such as the number of deposits and collections you have. “Show account detail” redirects you to your profile page.

On the bottom left column, there are shortcuts for defining new collections, metadata schemas and groups, explained more in the *Advanced usage*.



### 2.1.3 Deposit landing page

The landing page of a deposit is created after completion of the online deposit workflow. In the deposit landing page you can see the deposit's basic details, additional metadata and information about the status of the files contained in the deposit, for example if the data is currently online or offline and therefore needs to be staged.

[HOME](#) | [USER GUIDE](#) | [FAQ](#) | [CONTACT](#)

Hello

Download as

Add to basket

Favourite

### The Cosmogrid Simulation: Statistical Properties of Small Dark Matter Halos (512-001)

Tomoaki Ishiyama, Steven Rieder, Junichiro Makino, Simon Portegies Zwart, Derek Groen, Keigo Nitadori, Cees de Laat, Stephen McMillan, Kei Hiraki, Stefan Harfst

11 January 2011

We present the results of the "Cosmogrid" cosmological N-body simulation suites based on the concordance LCDM model. The Cosmogrid simulation was performed in a 30 Mpc box with 2048 particles. The mass of each particle is  $1.28 \times 10^5$  Msun, which is sufficient to resolve ultra-faint dwarfs. We found that the halo mass function shows good agreement with the Sheth & Tormen fitting function down to  $\sim 10^7$  Msun. We have analyzed the spherically averaged density profiles of the three most massive halos which are of galaxy group size and contain at least 170 million particles. The slopes of these density profiles become shallower than -1 at the inner most radius. We also find a clear correlation of halo concentration with mass. The mass dependence of the concentration parameter cannot be expressed by a single power law, however a simple model based on the Press-Schechter theory proposed by Navarro et al. gives reasonable agreement with this dependence. The spin parameter does not show a correlation with the halo mass. The probability distribution functions for both concentration and spin are well fitted by the log-normal distribution for halos with the masses larger than  $\sim 10^8$  Msun. The subhalo abundance depends on the halo mass. Galaxy-sized halos have 50% more subhalos than  $\sim 10^{11}$  Msun halos have.

[astrophysics](#)
[cosmology](#)
[simulation](#)
[nongalactic astrophysics](#)

FILESRELATIONSHIPSSTATISTICS

File name	Size	Status	Action	
Snapshot.001.hdfs.tar	2.04 GB		Download	<input type="checkbox"/>
Snapshot.001	264 bytes		Download	<input type="checkbox"/>
Snapshot.001.boundary	20.39 KB		Download	<input type="checkbox"/>

Showing 1 to 3 of 3 entries

Object properties

Number of files

3

Data size

2.04 GB

Access level

public

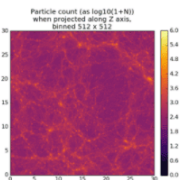
Conditions of use

License

MIT

## Export Metadata

To export metadata of an object go to the object's landing page and click on the export link on the top-right corner of the metadata panel. You can choose to export all the metadata or make a specific selection such as the default (Dublin Core) or community-specific metadata.



[Add to basket](#)
[Favourite](#)
[Edit](#)

### CosmoGrid 512-010

Simon Portegies Zwart, Steven Rieder, Tomoaki Ishiyama

11 January 2011

We present the results of the "Cosmogrid" cosmological N-body simulation suites based on the concordance LCDM model. The Cosmogrid simulation was performed in a 30Mpc box with 2048<sup>3</sup> particles. The mass of each particle is 1.28x10<sup>9</sup> Msun, which is sufficient to resolve ultra-faint dwarfs. We found that the halo mass function shows good agreement with the Sheth & Tormen fitting function down to ~10<sup>-7</sup> Msun. We have analyzed the spherically averaged density profiles of the three most massive halos which are of galaxy group size and contain at least 170 million particles. The slopes of these density profiles become shallower than -1 at the inner most radius. We also find a clear correlation of halo concentration with mass. The mass dependence of the concentration parameter cannot be expressed by a single power law, however a simple model based on the Press-Schechter theory proposed by Navarro et al. gives reasonable agreement with this dependence. The spin parameter does not show a correlation with the halo mass. The probability distribution functions for both concentration and spin are well fitted by the log-normal distribution for halos with the masses larger than ~10<sup>9</sup> Msun. The subhalo abundance depends on the Galaxy-sized halos have 50% more subhalos than ~10<sup>11</sup> Msun halos have.

**Keywords:** astrophysics, cosmology, simulation, nongalactic astrophysics

FILES | RELATIONSHIPS | STATISTICS

Filename	Size	Status	Action
Snapshot.010.hdf5.tar	2.08 GB	offline	Request
Snapshot.010	265 bytes		Download
Snapshot.010.boundary	20.39 KB	offline	Request

[Add to basket](#)

#### Metadata

Dublin Core | Community | Collection

EPIC PID	21.12996/4F05ED19-36F6-11E7-95DB-14109FDCCF91
Title	CosmoGrid 512-010
Creator	Simon Portegies Zwart, Steven Rieder, Tomoaki Ishiyama
Description	We present the results of the "Cosmogrid" cosmological N-body simulation suites based on the concordance LCDM model. The Cosmogrid simulation was performed in a 30Mpc box with 2048 <sup>3</sup> particles. The mass of each particle is 1.28x10 <sup>9</sup> Msun, which is sufficient to resolve ultra-faint dwarfs. We found that the halo mass function shows good agreement with the Sheth & Tormen fitting function down to ~10 <sup>-7</sup> Msun. We have analyzed the spherically averaged density profiles of the three most massive halos which are of galaxy group size and contain at least 170 million particles. The slopes of these density profiles become shallower than -1 at the inner most radius. We also find a clear correlation of halo concentration with mass. The mass dependence of the concentration parameter cannot be expressed by a single power law, however a simple model based on the Press-Schechter theory proposed by Navarro et al. gives reasonable agreement with this dependence. The spin parameter does not show a correlation with the halo mass. The probability distribution functions for both concentration and spin are well fitted by the log-normal distribution for halos with the masses larger than ~10 <sup>9</sup> Msun. The subhalo abundance depends on the halo mass. Galaxy-sized halos have 50% more subhalos than ~10 <sup>11</sup> Msun halos have.
Keyword	astrophysics, cosmology, simulation, nongalactic astrophysics
Publisher	Leiden University
Publication date	11 January 2011
Type	Dataset
File format	HDF5

### 2.1.4 Collection landing page

A landing page of a collection is similar to those of a deposit. In the collection landing page the basic details of the collection are listed, as well as the member objects of that collection, which can be other collections or deposits.

The screenshot shows the SURF Repository interface. At the top, there's a navigation bar with 'HOME', 'USER GUIDE', 'FAQ', and 'CONTACT'. A search bar and user profile are on the right. The main content area features the 'CosmoGrid' logo and the title 'The Cosmogrid Simulation: Statistical Properties of Small Dark Matter Halos 512<sup>3</sup> resolution'. Below the title, it lists authors: Tomoaki Ishiyama, Steven Rieder, Junichiro Makino, Simon Portegies Zwart, Derek Groen, Keigo Nitadori, Cees de Laat, Stephen McMillan, Kei Hiraki, Stefan Harfst, and the date '11 January 2011'. A detailed abstract follows, describing the simulation results. At the bottom, it provides the EPIC PID and DOI. On the left, there are tabs for 'COLLECTION' and 'RELATIONSHIPS', and a 'Subcollections (0)' section. On the right, there's a 'Collection search...' bar and an 'Object properties' section showing 'Number of files: 147'.

## 2.1.5 Search and discover datasets

To search for data use the search functionality on the home page. The text entered can be part of a title, keyword, abstract or any other metadata. Both registered and unregistered users can search for data. Unregistered users can only search for data sets that are publicly accessible. You can also make search within specific communities, which means the search results will be limited to that community.

The screenshot shows the SURF Repository home page. A red rectangle highlights the search bar in the top right corner. The page includes a 'Welcome narges' message with 'Views: 5' and 'Downloads: 0'. There are buttons for 'Create new collection', 'Create new metadata schema', 'Create new group', and 'Show account details'. A 'Welcome to the SURFara Data Repository' section provides a brief description and buttons for 'Deposit now' and 'Request large dataset publication'. A 'News' section lists 'Internal testing started' and 'Data Repository demo today!'. A 'Latest deposits' section lists several CosmoGrid datasets with their dates and authors.

Advanced searches can be performed by clicking the Search button, then entering the additional search criteria on the page that is shown. You can set advanced search criteria to search for deposits, collections, schemas, communities or combinations of any of them. You can also order the search results by Title, Creator, Identifier, or Creation date.

**Note:** If you have the PID (Persistent Identifier) of the data you can directly search in the [Handle Server](#) and get the url to the location of the data.

## 2.2 Deposit

This page explains how to get your data into the data repository using the online deposit workflow.

### Contents

- *Deposit*
  - *Deposit data*
    - \* *Prepare data*
    - \* *Deposit workflow*

### 2.2.1 Deposit data

Deposit is the act of uploading data to the Data Repository. Only registered users are permitted to deposit new records.

#### Prepare data

To prepare data for a deposit there are some best practices for the file format, file size, metadata, data documentation and organisation. For more information, visit the [Best Practices](#) page.

#### Deposit workflow

Clicking the deposit link in the main interface opens the first of a three stage process required to upload data. The steps are as follows:

1. In the Create a new deposit page, select files to deposit and then click on Upload to upload the file.


---

**Note:** Make sure that you are uploading an acceptable file format. Valid file formats can be found on the [File formats](#) documentation page. If the data format you are trying to upload is not supported, you can always contact us for support.

---

Provide the basic metadata for your deposit such as the title, creator(s), keywords (for searching purposes) and description. The fields in the upper section are mandatory fields, while the optional fields are shown in the lower section.






[Home](#)
[User guide](#)
[FAQ](#)

## Create a new deposit

Select and upload files and provide basic information about your deposit.

### Files for publication

Select one or more files to deposit and click the upload button to upload them.

Name	Type	Size	
deposit1.png	image/png	70.59 kB	 

Showing 1 to 1 of 1 entries

[+ Select files](#)
[Upload](#)

Valid file extensions: .jpg, .jpeg, .gif, .png, .txt, .text, .md, .pdf, .odt, .ods, .odp, .mp4, .avi, .hdf5, .h5, .hdf4, .nc, .cdl

### Basic metadata

Provide the necessary details by filling out the form, fields marked with an \* are mandatory.

Title \*

Creator \*

+

Keyword \*

+

Description \*

Publisher

Publication date \*

2017-09-25

Type \*

Dataset

▼

Language \*

No linguistic content

▼

License \*

Select

☒ Open access
 

Indicate whether your data publication should be openly accessible

Next >>

In this step you should also select the type of data and language of the data and a license for publication. The license can be selected through a built-in wizard. If you are not sure which license to choose, answer the questions on the top of the built-in wizard form to find the appropriate license.

**Choose a License**

Answer the questions or use the search to find the license you want

[Start again](#) [←](#) [→](#)

What do you want to deposit?

[Software](#) [Data](#)

Search for a license...

**Public Domain Mark (PD)**  
The work identified as being free of known restrictions under copyright law, including all related and neighboring rights.  
[Publicly Available](#)

**Public Domain Dedication (CC Zero)**  
CC Zero enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.  
[Publicly Available](#) [OPEN](#) [DATA](#)

**Creative Commons Attribution (CC-BY)**  
This is the standard creative commons license that gives others maximum freedom to do what they want with your work.  
[Publicly Available](#) [OPEN](#) [DATA](#)

**Creative Commons Attribution-ShareAlike (CC-BY-SA)**  
This creative commons license is very similar to the regular Attribution license, but requires you to release all derivative works under this same license.  
[Publicly Available](#) [OPEN](#) [DATA](#)

**Creative Commons Attribution-NoDerivs (CC-BY-ND)**  
The no derivatives creative commons license is straightforward; you can take a work released under this license and re-distribute it but you cannot change it.  
[Publicly Available](#)

Click 'Next >>' to go to the second step of the online deposit workflow.

- Next you can optionally select a community, collection and/or metadata schema. If you are a member of a community and you want to deposit data in that community, select the community name. The collection and metadata schema will be prepopulated based on which community you choose. If you are not member of any communities, you will see the collections and schemas defined by yourself. This step is optional and can be skipped. In the right you can the basic metadata you defined for the deposit.

**CREATE** [Home](#) [User guide](#) [FAQ](#)

**Create a new deposit (2)**

Optionally select a community, collection under which you want to publish your dataset. This might require annotation of extra metadata according to corresponding metadata schemas.

Select an additional metadata schema to provide more detailed information on your publication.

Select community: (None)

Select collection: (None)

Select metadata schema: (None)

[<< Back](#) [Next >>](#)

**Metadata overview**

<b>Title</b>	CosmoGrid 512-001
<b>Creator</b>	Simon Portegies Zwart, Steven Rieder, Tomoaki Ishiyama
<b>Keyword</b>	astrophysics, cosmology, simulation, nongalactic astrophysics
<b>Description</b>	We present the results of the "CosmoGrid" cosmological N-body simulation suites based on the concordance $\Lambda$ CDM model. The CosmoGrid simulation was performed in a 30Mpc box with 2048 <sup>3</sup> particles. The mass of each particle is $1.28 \times 10^{15} M_{\odot}$ , which is sufficient to resolve ultra-faint dwarfs. We found that the halo mass function shows good agreement with the Sheth & Tormen fitting function down to $\sim 10^{12} M_{\odot}$ . We have analyzed the spherically averaged density profiles of the three most massive halos which are of galaxy group size and contain at least 170 million particles. The slopes of these density

Click 'Next >>' to go to the third and final step of the online deposit workflow.

- In this final step you can fill in optional information such as the embargo data of your data. You can also specify links that are related to this publication. If you have any metadata that does not fit the basic or community metadata fields, you can add them here. For every field there needs to be a unique field name and a value.

Please carefully read the [Terms of use](#) and [Data Producer Agreement](#) before you agree by checking the checkbox before depositing data. In the end, click on the **Complete** button to finalize your deposit.

**Create a new deposit (3)**

Provide information on privacy, relations and other identifiers.

**Privacy settings**

Provide an embargo date or end date for this publication

Embargo

Publish until

**Linked information**

Provide any links that are relevant for this publication

ORCID

Github

Article

**Additional metadata fields**

Provide additional metadata field values for any field not defined in a metadata schema.

☐ I accept the [Terms of Use and Data Producer Agreement](#)

**Metadata overview**

Title	CosmoGrid 512-001
Creator	Simon Portegies Zwart, Steven Reider, Tomoaki Ishiyama
Keyword	astrophysics, cosmology, simulation, nongalactic astrophysics
Description	We present the results of the "CosmoGrid" cosmological N-body simulation suites based on the concordance $\Lambda$ CDM model. The CosmoGrid simulation was performed in a 30Mpc box with 2048 <sup>3</sup> particles. The mass of each particle is $1.28 \times 10^{15} M_{\odot}$ , which is sufficient to resolve ultra-faint dwarfs. We found that the halo mass function shows good agreement with the Sheth & Tormen fitting function down to $\sim 10^{17} M_{\odot}$ . We have analyzed the spherically averaged density profiles of the three most massive halos which are of galaxy group size and contain at least 170 million particles. The slopes of these density profiles become shallower than -1 at the inner most radius. We also find a clear correlation of halo concentration with mass. The mass dependence of the concentration parameter cannot be expressed by a single power law, however a simple model based on the Press-Schechter theory proposed by Navarro et al. gives reasonable agreement with this dependence. The spin parameter does not show a correlation with the halo mass. The probability distribution functions for both concentration and spin are well fitted by the log normal distribution for halos with the masses larger than $\sim 10^{18} M_{\odot}$ . The subhalo abundance depends on the halo mass. Galaxy-sized halos have 50% more subhalos than $\sim 10^{11} M_{\odot}$ halos have.
Publisher	
Publication date	2017-09-25
Type	Dataset
Language	English
License	Public Domain Mark (PDM)

**Files (2)**

Name	Type	Size
deposit1.png	image/png	70.59 kB
deposit2.png	image/png	73.89 kB

## 2.3 Downloads

This page provides information on downloading files from the Data Repository. There are two methods to get data out of the repository: download files directly from an object's landing page, or use the download script that allows downloads of multiple files or even complete deposits and collections selected in your basket or favourites.

### Contents

- [Downloads](#)
  - [Download files directly](#)
  - [Adding objects or files to your basket](#)
  - [Adding objects to your favourites](#)
  - [Download script](#)

### 2.3.1 Download files directly

To download files directly from the Data Repository you need to navigate to the landing page of the deposit you want to download the files from. Unregistered users can only download files of objects that are publically accessible. Files

of objects that are privately shared can only be downloaded by users that have explicit access to these objects. If an object's access level is set to registered, only registered and logged in users can access the files.

You can download single files by clicking on the **download** link next to the file name. If the status of the file is **offline**, it means it cannot be downloaded directly, but first needs to be staged. A request to stage the file can be made by clicking the **request** link.

FILES

RELATIONSHIPS

STATISTICS

Show

20

entries

Search:

+ File name	Size	Status	Action	
➤ Snapshot.770.0-63.hdf5.tar	3.82 GB	offline	Request	<input type="checkbox"/>
➤ Snapshot.770.1024-1087.hdf5.tar	4.03 GB	offline	Request	<input type="checkbox"/>
➤ Snapshot.770.1088-1151.hdf5.tar	4.03 GB	offline	Request	<input type="checkbox"/>
➤ Snapshot.770.1152-1215.hdf5.tar	4.01 GB	offline	Request	<input type="checkbox"/>
➤ Snapshot.770.1216-1279.hdf5.tar	3.83 GB	offline	Request	<input type="checkbox"/>
➤ Snapshot.770.128-191.hdf5.tar	4 GB	offline	Request	<input type="checkbox"/>

## 2.3.2 Adding objects or files to your basket

To download several files at the same time, you can add objects or files to your basket. This can be done by clicking on the 'Add to basket' button in the top-right corner, or by selecting individual files in the files table and then clicking on the basket icon button directly above the table.

Once you have selected all the objects and/or files you want, you can download them all at once using the download script provided by the repository. Go to your account page and select the basket tab to download the script by clicking on the 'Download script' button.

SURF

REPOSITORY

HOME

|

USER GUIDE

|

FAQ

|

CONTACT

5

1

Hello

Profile

Administrator

Review

Deposits

Collections

Groups

Communities

Metadata schemas

Basket

Favourites

API tokens

Messages

Basket

Download script

Download JSON

Remove all

An overview of all the items you currently have in your basket. Go to the landing page of an object to add it to your basket.

Show

10

▾

entries

Search:

Title	Type	Total size	Status	Container	Actions
<a href="#">Snapshot.772.boundary</a>	File	164.92 KB	online	The Cosmogrid Simulation: Statistical Properties of Small Dark Matter Halos (2048-772)	<a href="#">🔗</a>
<a href="#">Snapshot.772.512-575.hdf5.tar</a>	File	4.12 GB	online	The Cosmogrid Simulation: Statistical Properties of Small Dark Matter Halos (2048-772)	<a href="#">🔗</a>
<a href="#">Snapshot.772.320-383.hdf5.tar</a>	File	4.11 GB	online	The Cosmogrid Simulation: Statistical Properties of Small Dark Matter Halos (2048-772)	<a href="#">🔗</a>
<a href="#">Snapshot.772</a>	File	267 bytes	online	The Cosmogrid Simulation: Statistical Properties of Small Dark Matter Halos (2048-772)	<a href="#">🔗</a>
<a href="#">Snapshot.001.0-63.hdf5.tar</a>	File	1.96 GB	online	The Cosmogrid Simulation: Statistical Properties of Small Dark Matter Halos (1024-001)	<a href="#">🔗</a>

Showing 1 to 5 of 5 entries

Previous

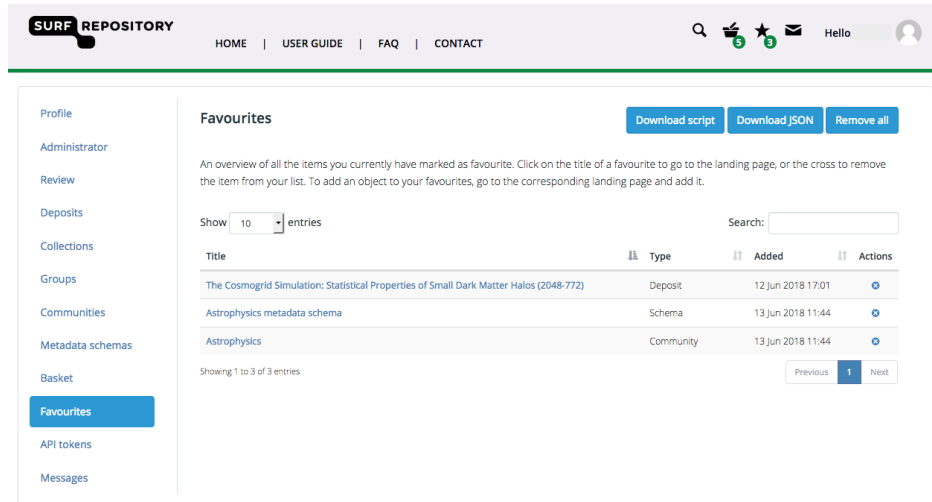
1

Next

### 2.3.3 Adding objects to your favourites

You can add any object to your favourites by clicking on the ‘Favourite’ button in the top-right corner of the landing page of an object. Go to the favourites tab in your account overview to see the list of favourited items.

You can download all downloadable objects directly using the download script. Click on the ‘Download script’ button on the same tab to download the script.

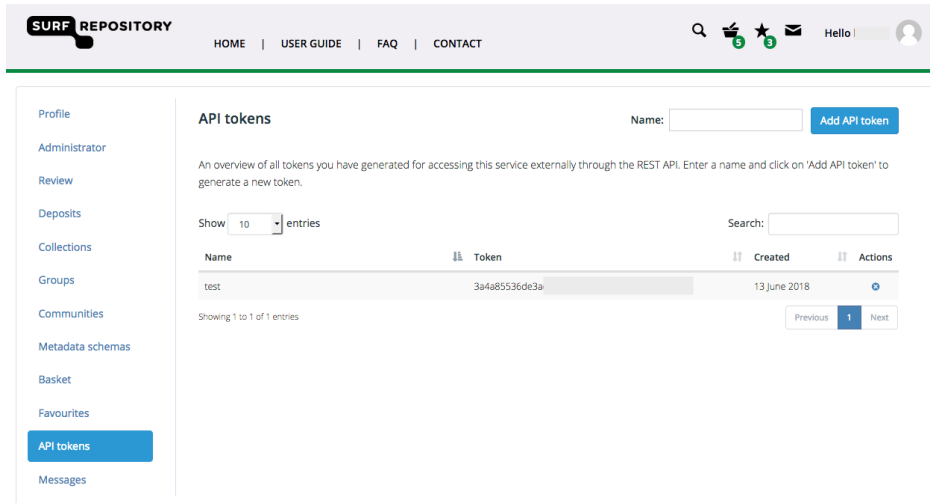


### 2.3.4 Download script

The download script as provided by the Data Repository service is a tool that automatically downloads all objects and files in your basket or favourites. It is written in Python and therefore needs to have the Python runtime installed. You can download Python from its [website](https://www.python.org/).

The script needs to be run using a terminal or command prompt (depending on the operating system you are using). If you are not familiar with the Linux or MacOS terminal, please refer to for example this [introductory tutorial](https://computers.tutsplus.com/tutorials/navigating-the-terminal-a-gentle-introduction-mac-3855) or for Windows, this [introduction to the Windows command prompt](https://www.bleepingcomputer.com/tutorials/windows-command-prompt-introduction/).

Before you can start downloading using the script, you need to create an API access token that allows authentication through the service API using the tool. Go to the API tokens tab in your account page and click ‘Add API token’ after filling in a unique name for the token.



The screenshot shows the 'API tokens' page in the SURF Repository interface. The page has a sidebar on the left with a list of navigation items: Profile, Administrator, Review, Deposits, Collections, Groups, Communities, Metadata schemas, Basket, Favourites, API tokens (highlighted in blue), and Messages. The main content area is titled 'API tokens' and includes a 'Name:' input field and an 'Add API token' button. Below this, there is a table with columns: Name, Token, Created, and Actions. The table contains one entry with the name 'test', a token '3a4a85536de3a...', and a creation date of '13 June 2018'. The table is paginated, showing 'Showing 1 to 1 of 1 entries' and navigation buttons for 'Previous', '1', and 'Next'.

**API tokens** Name:  [Add API token](#)

An overview of all tokens you have generated for accessing this service externally through the REST API. Enter a name and click on 'Add API token' to generate a new token.

Show  entries Search:

Name	Token	Created	Actions
test	3a4a85536de3a...	13 June 2018	<a href="#">ⓘ</a>

Showing 1 to 1 of 1 entries [Previous](#) [1](#) [Next](#)

Copy the randomly generated token to your clipboard and add it after the script name's statement in your terminal or command prompt. When executing the tool will automatically download all files contained in the selected objects.

### 3.1 Advanced Usage

This page provides information on the advanced usage of the data repository, such as Metadata schema definition, and community/group memberships. It also describes the online deposit workflow which is the steps that you take to deposit data on the Data Repository.

#### Contents

- *Advanced Usage*
  - *Metadata schema*
  - *Collections*
  - *Groups*
  - *Communities*
  - *API tokens*

#### 3.1.1 Metadata schema

A metadata schema is a logical plan showing the relationships between metadata fields, normally through establishing rules for the use and management of metadata. Metadata schema makes sure that your deposit or collection has complete and understandable information attached to it. The first question to ask before making a metadata schema is: Is it necessary to create a new metadata schema? or are there already existing metadata schemas which can be adapted for use? We advice to reuse existing schemas. You can find here the **'best practices for defining your metadata schema'**.

To define a metadata schema for your data, go to the profile tab of account overview page click on **Create new metadata schema** to go to the metadata schema creation form.

**Create new metadata schema**

Provide the title, description and metadata field descriptions of this schema object.

**Title \***  
Schema name

**Creator \***  
narges

**Description**  
Description

**Metadata fields**

Field	Label	Description	Default value	Type	Optional	Multivalued	Actions
				string	no	no	

**Administrators**

**Other options**

☒ Public metadata schema

Cancel Create

You need to provide a title, creator, and description for the metadata schema. Then add metadata fields to the schema. Metadata field contains information about the field such as description, default value, type and if the field is optional or not. To save the field click on the **Save** icon under **Actions**. In **Other options**, you can also specify if it is a public metadata schema. In the end hit **Create** button to finalise the creation of the metadata schema.

### 3.1.2 Collections

Collection is a useful way to bundle a large number of deposits. Collections itself can also be bundle in another collection. Any structure is possible.

To create a new collection, go to the profile tab of the account overview page and click on **Create new collection** to go to the collection creation form. Provide the title, creator, description and optionally the parent collection of this collection object. Then hit **Create** to save your collection. The new collection will now be available during the online deposit workflow.



**CREATE NEW COLLECTION**

Provide the title, description and optionally the parent collection of this collection object.

**Title \*** ⓘ  
Collection name

**Creator \*** ⓘ  
narges

**Description** ⓘ  
Collection description

**Parent collection** ⓘ  
(None)

**Part of community** ⓘ  
(None)

**Metadata schema** ⓘ  
(None)

**Administrators** ⓘ  
Start typing a user name

Cancel Create

To see the list of your current collections, go to your account page and then choose the **Collection** tab on the left.

### 3.1.3 Groups

A group is a simple bundling of one or more users. They can be used to provide access or administrative privileges to a bunch of people using only a single relation.

To create a new group or to see the list of current groups are member of, go to your account page and then choose the **Groups** tab on the left.

**Groups**

An overview of all groups you are the owner of. Click on 'Create new group' to add a new group, or search for a group to become a member of an existing one.

Show 10 entries

Search:

Title	Members	Created	Owner	Actions
You currently don't have any groups				

Showing 0 to 0 of 0 entries

Create new group

A new page will be open where you need to provide name and description for the group, invite other members to join the group, and assign admins to the group. After you enter the information, click on the **Create** button.

### 3.1.4 Communities

Communities bundle collections and deposits under a single entity. With a community, you can add policies to deposits that make sure publications are up to the standards of your community.

To create a new group or to see the list of current groups are member of, go to your account page and then choose the **Communities** tab on the left. You can also create the community from the profile tab of your account overview page click on **Create new community** to go to the community creation form.

A new page will be open where you need to provide more information about the new community you make such as the title, creator, and description. You should assign at least one member to the community. You can also choose who can be the administrators of the community. If you are a community administrator, you have special privileges regarding the reviewing of objects, and editing their metadata. More information such as description, default collection and default metadata schema can be defined for the community. In the end you can choose the policies you want to apply to the community. After you enter the information, click **Create** to make the community.

### Create new community

Provide the title, description of this community object.

**Title \*** ⓘ

**Creator \*** ⓘ

 +
 

**Description** ⓘ

**Default collection** ⓘ

(None) ▾

**Default metadata schema** ⓘ

(None) ▾

**Members \*** ⓘ

**Administrators** ⓘ

**Other options**

☐ Only members can deposit in this community ⓘ

☐ Review deposit first before publication upon submit ⓘ

☐ Allow external collections ⓘ

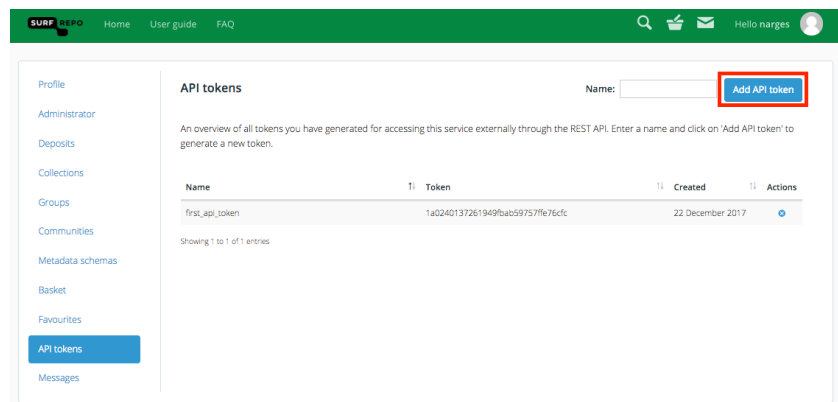
Cancel
Create

Your community is now created and will be available during the workflow of new deposits. It is also visible in the community tab of your account overview page

### 3.1.5 API tokens

The HTTP REST API can be used for interacting with Data Repository via external services or applications, for example for integrating with other web-sites (research community portals) or for uploading or downloading large data sets that are not easily handled via a web browser.

Go to the **API tokens** on the left side of your account page to see an overview of all tokens you have generated for accessing this service externally through the REST API. To create a new API token, enter a name and click on **Add API token** to generate a new token.

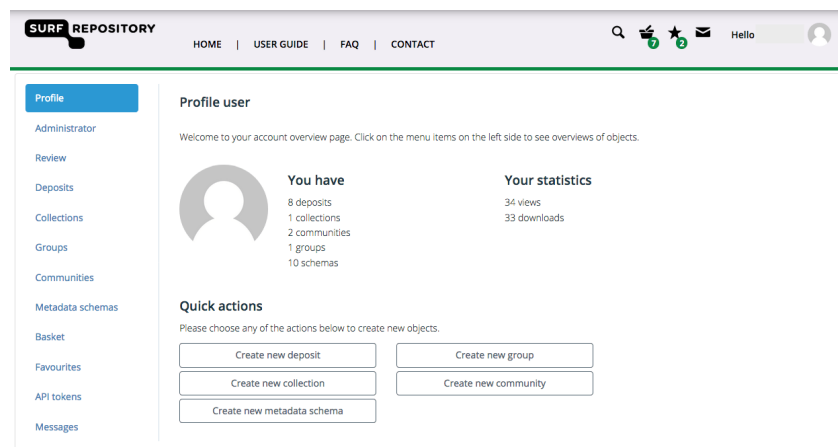


## 3.2 Account

The user account page provides an overview of all the objects you own, administer or are a member of. In this overview, tabular listings of these objects are shown in dedicated tabs.

### 3.2.1 Profile

When you navigate to the account page, you will end up on the profile tab with on the left side a menu to show the other tabs.



The profile tab shows an overview of the number of objects you have per type, and some statistics about the views and downloads of these objects. The quick action buttons allow rapid creation of new objects.

### 3.2.2 Administrator

The administrator tab shows a listing of all the objects you have administrative rights to, but not objects you actually own. The listing includes the type of the object, the access level and current publication status. Click on the edit icon under actions to edit the object.

**SURF REPOSITORY** HOME | USER GUIDE | FAQ | CONTACT

Profile  
**Administrator**  
Review  
Deposits  
Collections  
Groups  
Communities  
Metadata schemas  
Basket  
Favourites  
API tokens  
Messages

### Administrator

An overview of all the objects you currently have administrative rights to.

Show 10 entries Search:

Title	Type	Access	Status	Created	Actions
The Cosmogrind Simulation: Statistical Properties of Small Dark Matter Halos (1024-007)	deposit	public	Published	6 Jun 2018 13:46	
The Cosmogrind Simulation: Statistical Properties of Small Dark Matter Halos (1024-008)	deposit	public	Published	6 Jun 2018 13:46	
The Cosmogrind Simulation: Statistical Properties of Small Dark Matter Halos (1024-009)	deposit	public	Published	6 Jun 2018 13:46	
The Cosmogrind Simulation: Statistical Properties of Small Dark Matter Halos (1024-010)	deposit	public	Published	6 Jun 2018 13:46	
The Cosmogrind Simulation: Statistical Properties of Small Dark Matter Halos (1024-011)	deposit	public	Published	6 Jun 2018 13:46	
The Cosmogrind Simulation: Statistical Properties of Small Dark Matter Halos (1024-012)	deposit	public	Published	6 Jun 2018 13:46	
The Cosmogrind Simulation: Statistical Properties of Small Dark Matter Halos (1024-013)	deposit	public	Published	6 Jun 2018 13:46	
The Cosmogrind Simulation: Statistical Properties of Small Dark Matter Halos (1024-014)	deposit	public	Published	6 Jun 2018 13:46	
The Cosmogrind Simulation: Statistical Properties of Small Dark Matter Halos (1024-015)	deposit	public	Published	6 Jun 2018 13:46	
The Cosmogrind Simulation: Statistical Properties of Small Dark Matter Halos (1024-016)	deposit	public	Published	6 Jun 2018 13:46	

Showing 11 to 20 of 703 entries

Previous 1 2 3 4 5 ... 71 Next

### 3.2.3 Review

The review tab shows a listing of all the objects that need review before they are actually published. The listing is similar to the listing of the administrator tab.

**SURF REPOSITORY** HOME | USER GUIDE | FAQ | CONTACT

Profile  
Administrator  
**Review**  
Deposits  
Collections  
Groups  
Communities  
Metadata schemas  
Basket  
Favourites  
API tokens  
Messages

### Review

An overview of all the objects that are currently under review and need approval.

Show 10 entries Search:

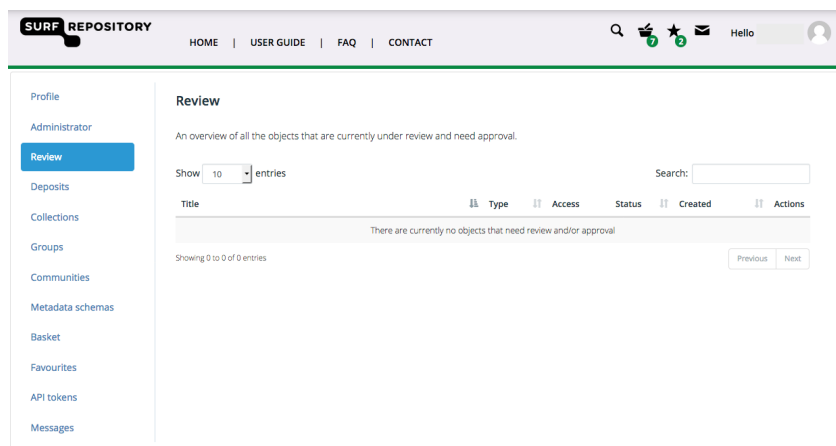
Title	Type	Access	Status	Created	Actions
There are currently no objects that need review and/or approval					

Showing 0 to 0 of 0 entries

Previous Next

### 3.2.4 Deposits

The deposits tab shows a listing of all deposits that you own. Click on the title of a deposit to go to the landing page of that deposit.



## 3.3 Objects

This page provides general information about objects stored in the Data Repository.

### Contents

- *Objects*
  - *Objects*
  - *Object-level access*

### 3.3.1 Objects

Digital objects are stored and maintained by the repository service using a dedicated object storage and definition system. The objects contain metadata defined in XML and in case of deposits also the references to files.

### 3.3.2 Object-level access

By default, an object is publicly accessible and therefore can be seen and downloaded by anyone interested. This includes all files and the metadata, except access statistics and configuration. To restrict access to an object, the owner or an administrator of the object can set the access levels. In the table below these restrictions are listed.

Access level	Per user	Description
Public	No	Object, including files and metadata, is open access and accessible to everyone, even anonymous users
Registered	No	Object is accessible only to registered users. Anonymous users can still see metadata, but cannot download files
Private	Yes	Object is only accessible to users specifically given access to it. Metadata is still open though.

If an object is set to private access, the owner or an administrator can select users that are allowed to access the object. All other users will receive an 'object not found' message.

### 4.1 Best Practices

This page provides information on best practices while using the repository.

#### Contents

- *Best Practices*
  - *Preparing data*

#### 4.1.1 Preparing data

To prepare data for a deposit you need to consider the following points:

- **File formats** By default, the repository does not accept all file formats. This is to ensure future compatibility, understanding and support of published data by the repository itself, but also by other users. The repository distinguishes two kinds of formats: preferred and accepted formats.
  - **Preferred formats** are file formats of which SURFsara is confident that they will offer the best long-term guarantees in terms of usability, accessibility and sustainability. Depositing research data in preferred formats will always be accepted by SURFsara. Here is a list with [SURFsara preferred formats](#).
  - **Acceptable formats** are file formats that are widely used in addition to the preferred formats, and which will be moderately to reasonably usable, accessible and robust in the long term. SURFsara favours the use of preferred formats, but acceptable formats will in most cases also be allowed. Here is a list with [SURFsara acceptable formats](#).

If you think your file format should be accepted by the repository, but is currently not on one of the lists, please contact us. A reason could be that a community uses a file format that is widely used within that community but not outside of it.

- **File size**

- **Maximum file size:** 4 GB using the online deposit form

The maximum deposit data volume is 10 GB and the maximum number of files that can be uploaded via the online deposit is 10 files. If you want to deposit larger volumes of data or larger number of files in a single deposit, please contact us.

- **Data documentation** Data documentation ensures that research data are understood and therefore used by current and future users (including the researcher). It is vital to store the data in a structured and consistent way with appropriate data documentation. The documentation can be a description explaining what the data is, what you can do with it and how it can be used. Even more technical documentation about the exact file format could be of use. Documentation can be provided in one of the accepted or preferred file formats.

- **Metadata** Metadata provides structured information about the data. Depending on the type of data, there are different types of metadata:

- **Descriptive metadata** describes and identifies information resources (minimal metadata required to find a digital object). It includes elements such as title, abstract, author, and keywords
    - \* **Structural metadata** provides information about the internal structure of resources including page, section, chapter numbering, indexes, and relations to other digital objects.
  - **Technical metadata** provides information on the technical aspects of the datasets such as data formats, hardware/software used, calibration, version, authentication, encryption.
  - **Administrative metadata** provides information on user rights and management of digital objects such as license, rights management, and access control.

The repository will only ask for descriptive metadata during the deposit workflow, but it is highly recommended to add links to additional documentation or files containing metadata of any kind. Any other kinds of metadata are automatically added, though not always visible for the end-user.

- **Data organisation**

If you want your research data to be easily traced and interpreted, the collection and deposit structure and the file names used for the data files should be logical and straightforward. Make sure to use the correct file extensions to avoid files being rejected or misinterpreted by other users. It's also a good practice to note the file naming and its meaning in a separate readme file.

You can create collections and collections of other collections to structure your data. There is no limit on the number of deposits in a single collection.

- **Data anonymization** Before you upload the files you should check whether they contain privacy-sensitive information within the meaning of the regulations and guidelines in the [General Data Protection Regulation](#) (GDPR). If you give access to the data publicly or any other user, the data must be completely anonymized.



### 5.1 Appendix

This page provides several terms with their descriptions used in data publication.

#### 5.1.1 Open data

Simply making data available – providing access – is often not enough to be allowed to reuse data. The data has to be available in a format and under a license that allows reuse.

A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike. (see <http://okfn.org/opendata/>).

All publications made in the Data Repository include a license that specifies what a data consumer can or can't do with the data. By downloading data from the service, the user agrees to comply with these terms.

#### 5.1.2 Data Provenance

Data provenance is providing a historical record of the data and its origins. It refers to the process of tracing and recording the origins of data and its movement between databases. (see <http://db.cis.upenn.edu/DL/fsttcs.pdf>).

#### 5.1.3 Persistent identifiers

A unique and persistent code that is coupled to a digital object. With this code, the object can be identified even when the object is moved to a different location.

### 5.1.4 DOI

The digital object identifier (DOI) is a unique and stable (or persistent) identifier that ensures that a digital object can be permanently found by any researcher, regardless of changes in the URL where the object is found. A central registry ensures that the user of a DOI will be referred to its current location (<http://www.datacite.org/>).

All objects stored in the Data Repository will automatically receive a DOI upon publishing. See also ‘Persistent identifiers’.

### 5.1.5 Certification

An archive holding the [CoreTrustSeal certification](#) complies with requirements ensuring that in the future, research data can still be processed in a high-quality and reliable manner (see ).

The Data Repository will be certified with this certification to guarantee a researcher that its data publication is accessible now and in the future.

### 5.1.6 Linked data

A term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information and knowledge on the Semantic Web using RDF. Linked data refers to data published on the web in such a way that it is machine-readable, that its meaning is explicitly defined, that it is linked to other external data sets, and that in turn it can be linked to from external data sets. (see <http://linkeddata.org>).

#### Need help?

Do you need help with the use of the Data Repository or this user documentation? We are more than willing to assist! Just contact us at [helpdesk@surfsara.nl](mailto:helpdesk@surfsara.nl).