DataGURU

LUND UNIVERSITY
Faculty of Science

Adding data to DataGURU: Step by step

DataGURU is a service to disseminate spatio-temporal data to colleagues in your field and in other fields. Its aim is to ease the access to data that you decided to be freely available, by allowing the user to define the extend, resolution and format in which the user will get the data.

You are eligible to add your data if:

- your data is a geo-located raster (so far we haven't added shapefile functionality, but if required we might do so)
- it is of potential use to others
- you own the rights to your data
- you have the required meta-data to your data (location, time, resolution, format, etc...)

The decision whether your data will be made available is ultimately made by the DataGURU team. The data stays your intellectual property and you will remain responsible for providing correct data.

Basically, DataGURU is a server which allows the data owner to enter the meta-data of the data to be disseminated and at a request of the user, the server downloads the data and performs the required conversions to deliver the data in the user required form.

To be able to do this DataGURU stores the meta-data of the dataset in its own data base.

There are two options for the storage of the actual data.

1. If you have already a web space, either on a web page or an ftp server, you just provide the link to this server and the data will be downloaded once requested by the user.
   This allows you to continuously update your data, but it also requires that you make sure that the data will be available over a long term.
2. If you do not have such an opportunity or would like us to handle the data storage, we can store the data for you. In this case you need to inform us and we will either store it on our server, or -in the case of dataset larger than a few terrabytes- we will try to find a suitable storage location together with you.

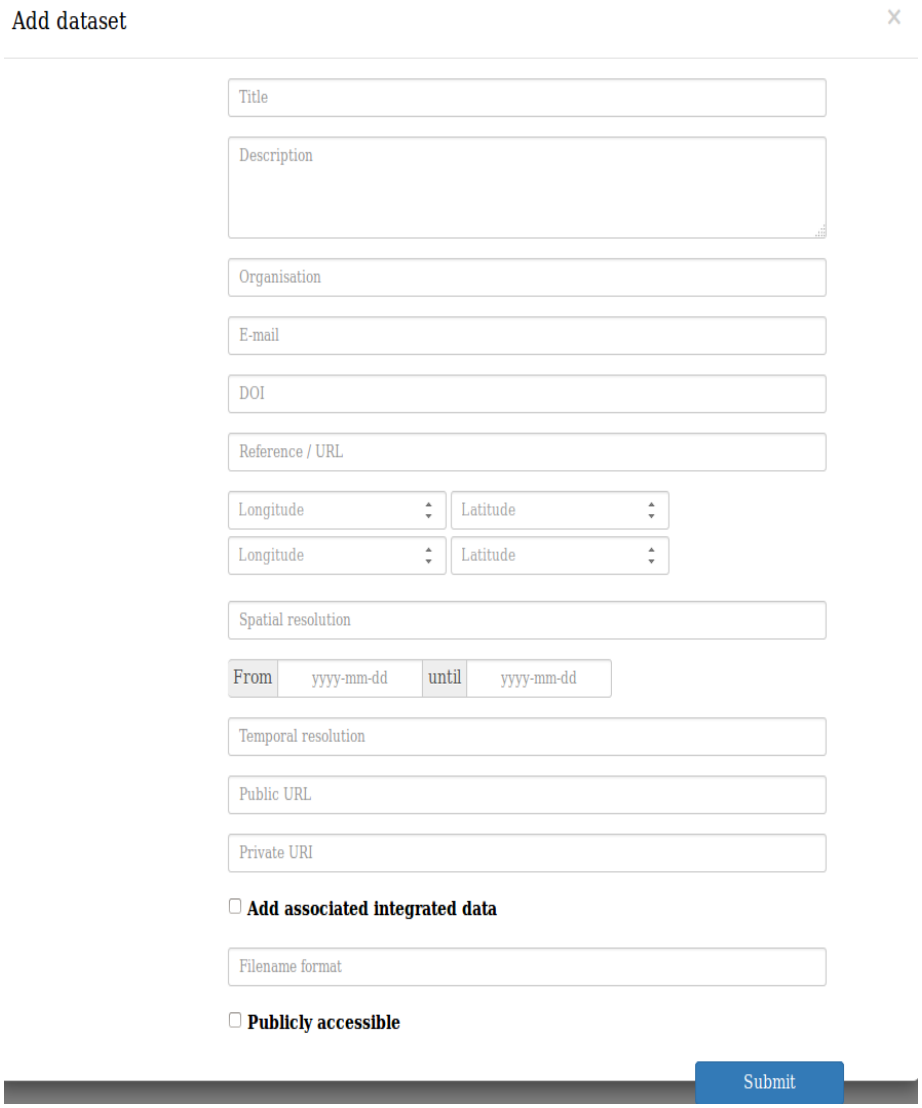
To be able to submit your dataset (as well as to download any of the datasets already stored in DataGURU) you need to have an account. We currently use 'Persona' as an identification service. Once you navigated to dataguru.nateko.lu.se please click on the login symbol:

You will have to provide a valid email to create an account at DataGURU and a password. Check you email for a verification link and this is all that is required to get a DataGURU account.
Once you are logged in the sign will disappear and it will be replaced by one sign indicating your account management (left) and one for logout (right):

To add (or in a later stage edit) your meta-data, click on the 'manage' icon (the toothed wheel). Once you are on the 'Manage your account' page:

you can change your 'Username', check the 'Recent jobs' (data requests) and submit new 'Datasets'. Pushing 'Datasets' and afterwards 'Add a new dataset' open the meta-data editor:

Adding a new dataset can not be fully automated giving the large variety in dataset formats and specifications. Depending on the structure of the dataset this list might not be sufficient to give us all details of your data (especially if it is a projected data set). However it gives us the most important meta-data and will allow us to start the process of integrating your data.
When you move the mouse over the fields, a help text will appear.



First provide a title of your data set.

The description that is asked for in the next line should be a short description of the data. This text will be fully search-able. If you want a certain project to be associated with the data, add the project name and any user searching for the project will find the data set. This is also the place where links to the data to be used in publications can be placed. For example for the habitat projections published in GEB I placed LehstenGEB2014 in the text and in the publication I wrote:

"The simulated data presented in this study, as well as the projections of land use by the Dyna-CLUE model, can be accessed from the DataGURU server (http://dataguru.nateko.lu.se/app#LehstenGEB2014) and can be used to project species occurrences based on broad habitat.

Hence this is the place to put a tag that you can list in your publication when submitting it to the reviewer, to show the reviewer that you will be sharing the data.

In the next lines please provide your Organization, email, DOI of the dataset if you have one and the Reference URL of the publication describing the dataset.

The western, northern, southern and eastern limits of your dataset are to be provided in latitude-longitude in the fields below.

The next fields asks for the temporal extend of your data. Please use the format given : year-month-day.
The temporal resolution accepts the entries 3hour, 6hour, 1day, 1month, 1year, 10year.
If your data has a different resolution, leave this free and write the resolution in the comments field that will pop up later.

The Public URL is the root path containing all variables. So if you have your data in a folder on an ftp-server, this would be the place where this folder needs to be typed in. If it is a web page, the address of the file would be required here.

If you want DataGURU to store your data and you have it in a number of different folders organized, e.g. one folder for each year, then the name of the folder should be listed in the 'Private URL' field.

The naming convention organizing your files need to be specified in the list text field. The convention is {rel_path}_{year?}{month?}{day}.nc?
For example the climate data from WFDEI would have a file name like Rainf_daily_WFD_CRU_190104.nc indicating that the variable name is Rainf (for rainfall) followed by some descriptors of the dataset (daily_WFD_CRU_) and a year-month. Finally the file format is indicated. So for this example one would have to write :
{Rainf}_daily_WFD_CRU_{year}{month}.nc .
If in doubt please try to explain our naming convention in the comment field and we will check this.

Once you filled all those fields, and press the submit button, a small window will pop up and allow you to add all information that you think would be required that is not provided here.
These could include meta-data about geographical projections, or any properties that you could not enter in the form. If you want DataGURU to store the data please also indicate this in the comment box.

Pressing the submit button will generate an email and we will contact you to see how we can add

your data in DataGURU and make it available to everyone.

The DataGURU team.