Data management plan - ACT

June 2022

****

**Names**: Jakko-Jan van Ek, Julia Sipkema, Mark Boeve, Niels Verouden & Raimon Bach Pareja

**Commissioner**: S.M.J. Arts, Ministerie van Defensie

**Coach**: Sytze de Bruin

**Contents**

[**Organisational context**](#_5ks6lrb7gnqm) **3**

[**Project description**](#_qmgcrjpmem1h) **3**

[**Data management roles**](#_na9vtb5q4hd8) **4**

[**Project data**](#_1o8ch66uv69e) **4**

[**Short term storage solutions**](#_hzxhueujegi6) **5**

[**Data and information structuring**](#_bomg8ccn8p1j) **6**

[**Documentation and metadata**](#_h6sxhzo48mia) **6**

[**Sharing and ownership**](#_gojt2nace0wv) **7**

[**Long term storage**](#_e1cwpnstynms) **8**

[**Annex I: Agreement between project group and WU-supervisor.**](#_4rvrsa1p2wjw) **9**

# 

# Organisational context

| **Name** | Group RGIC22-05 |
| --- | --- |
| **Date** | May 2022 |
| **Chair group** | Geo-information Science and Remote Sensing (GRS) |
| **Commissioner/ GRS-supervisor** | S.M.J. Arts, Ministerie van Defensie |
| **Start date of the project** | 09-05-2022 |
| **File name of this DMP** | DataManagementPlanACT.docx |

# Project description

| **Title** | Flood risk assessment using time series analysis |
| --- | --- |
| **Abstract** | In recent years, climate change has become an important factor for conflict. To mitigate these conflicts and to find suitable locations to set up camps, information about natural disasters, such as floods, is valuable to the Ministry of Defence. Therefore, in this project a time series analysis method will be created showing the spatio-temporal characteristics of flood events. This report serves to give the reader an intensive background on the methods used in the project and the choices that were made.  In this report multiple methods of time series are explored based on the use of sentinel-1 SAR data. These methods are thresholding, random forest, k-nearest neighbour, gaussian naive bayes classifier, support vector machines, average backscatter and difference maps. The last two methods were specifically created to detect urban floods, because urban floods could not be detected easily by the methods that were created. This resulted in two parts of the model: one for urban floods and one for non-urban floods.  The machine learning methods turned out to be the best methods to classify floods. From the machine learning methods, k-nearest neighbours is the most accurate, but all of these methods do not differ that much in accuracy. Both machine learning and thresholding methods had difficulties detecting floods in urban areas, whereby two other methods were created: image differencing of SAR imagery and a time series analysis. The time series analysis had the best results and the image differencing method was found to be too inaccurate, therefore its use is not recommended. In conclusion, machine learning methods are recommended to use. However, due to the different backscatter in different countries and areas, polygons need to be created by the user to use as training data. When this is done, the machine learning method could be applied globally. |

# Data management roles

| **Roles** | Project manager: Mark Boeve  Controller: Julia Sipkema  Secretary: Jakko-Jan van Ek  Member: Niels Verouden  Member: Raimon Bach Pareja |
| --- | --- |
| **Who is collecting the data?** | All members from project group 5 |
| **Who is analysing the data?** | All members from project group 5 |
| **Other**  **(Is there a person in the project group with a specific responsibility for data management? Do other people contribute, for example by writing code?)** | No |
| **What is the role of your supervisor?** | Mainly supervise the functioning of the team and provide some feedback on written reports. No part in data collection and analysis. |

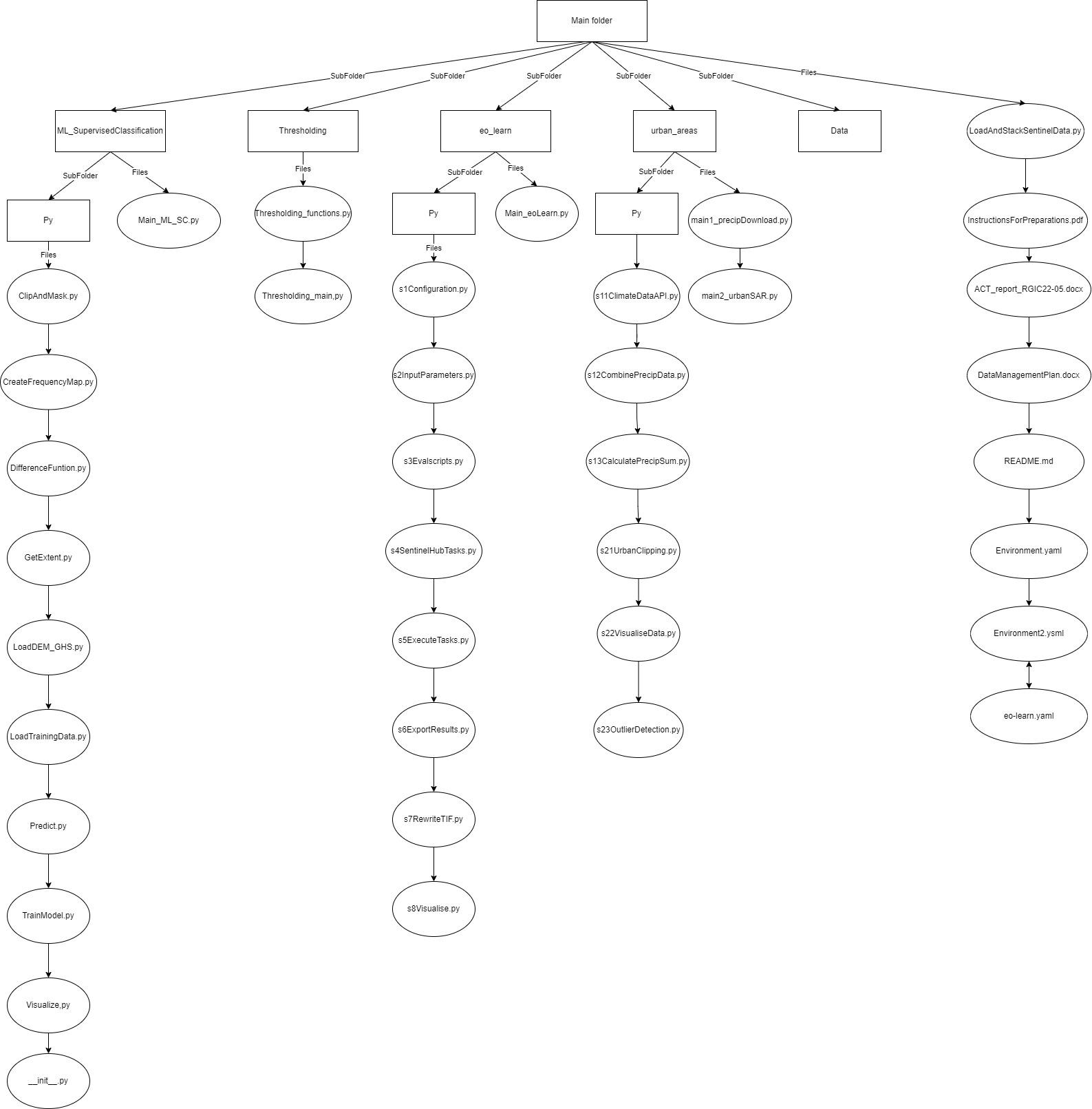
# Project data

| **Data stage** | **Specification of type of project data** | **Software choice** | **Data size/ growth** |
| --- | --- | --- | --- |
| **Source data** | * Sentinel-1 data * DEM * Global Surface Water data * Global Human Settlement data * precipitation data * Creation training polygons | Python, ArcGIS Pro | *Dependent on spatial and temporal extent*. *For three months (17 images) estimated: 614 mb.* |
| **Result data** | Flood maps, frequency maps | Python |  |
| **Models/ code** | Workflow model  flood maps | Diagrams.net & Python |  |
| **Other?** |  |  |  |

# Short term storage solutions

| **Datastage** | **Storage location** | **Backup procedures** |
| --- | --- | --- |
| **Source data** | Personal laptops | Google Drive, personal laptops, WeTransfer |
| **Result data** | Personal laptops | Google Drive, personal laptops |
| **Models/code** | Personal laptops, GitLab | Google Drive, personal laptops, GitLab |
| **Other?** |  |  |

# Data and information structuring



# Documentation and metadata

The SAR data is derived from EO browser and is pre-processed by EO Browser as well. The VV - decibel gamma0 - radiometric terrain and VH - decibel gamma0 - radiometric terrain are used and in the script a speckle filter is executed over these images. Later, the water bodies are masked out and the images are cropped to decrease the processing time of the model. Further, a DEM and Global Human Settlement data is stacked on these SAR images to improve the classification of the area. The DEM is downloaded from the EO browser as well in a grey scale. The Global Human Settlement data is downloaded from the European Commission’s website: GHSL - Global Human Settlement Layer. The precipitation data, which is used to relate rainfall to floods is downloaded via an API built in the script. This data is derived from NASA POWER: The Power Project. More information about the data is in the report in the section *Materials*. An explanation about how to download and pre-process the data is in the file *Instructions for Running the Time Series Application*.

# Sharing and ownership

| **Sharing and ownership** | **(With) who(m), what and how?** |
| --- | --- |
| **Data sharing**  - Do you expect that others may be interested in re-using your data? Do you have plans to share your data with these parties?  - How are you going to make sure your data files will be accessible once you leave the department? Who will take care of your data? | The project is executed at the request of the Ministry of Defence, but it is useful for other stakeholders as well. For example, local governments. NGO’s and the people living in the flood-prone areas. All the data that is used for the project is open source data, so others can freely use the same data as the team does for this project. The model, however, is not directly freely available. The team does not have plans for sharing the data with these third parties, because it is not clear if these people know how to use the model and it is hard to contact those third parties. For the Ministry of Defence, the data will be handed over after the project is finished. One day after the project presentations, the team will go to the commissioner and give a last presentation about the project. Also, the data and the model will be given via an USB-stick. |
| **Data ownership**  - Any funder’s requirements to share your data, or to impose an embargo?  - Are there agreements on how the data will be used and shared within your group or with other parties involved in this project? (outside your group or outside Wageningen UR) | The team does not have any funders, except for the Wageningen University which will compensate for expenses made by the group such as travelling and printing costs. As mentioned before, all the data is open source data so there are no restrictions in sharing the data. |
| **Privacy**  - Are there privacy or security issues, and if there are, how are you dealing with them? | For this project the team has signed a Non-Disclosure Agreement (NDA) which was created by the Wageningen University. On purpose, the Ministry of Defence did not share any classified information or data, therefore the team is not able to share any of this data. |

# Long term storage

An overview of the project data and the argument(s) whether it is valuable to store the data long term or not are given below. The data archives we intend to use are primarily the personal laptops. However, for long term storage, the archives consist of an USB stick and Google Drive as a cloud-based storage and file-sharing service.

| **Yes or no?** | **Argumentation** |
| --- | --- |
| Yes | **What:**  Workflow model of time series analysis  **Why:**  The workflow model of the time-series analysis has value for long term storage since the workflow model shows how the analysis can be reconstructed. This is highly valuable for the commissioner since they are interested in ultimately conducting the analysis and creating a risk assessment of flood-prone areas. Moreover, the workflow model can be useful for the group members as well whenever a time-series analysis has to be conducted in future research. |
| No | **What:** Intermediate data  **Why:** Open-source data that can easily be downloaded from the internet will not be stored after the project is completed since this takes up redundant storage space. |

# 

# Annex I: Agreement between project group and WU-supervisor.

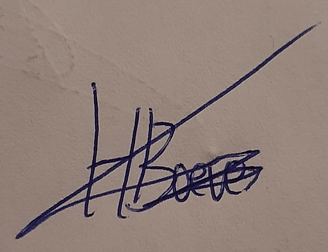
We, the undersigned, agree upon the following points:

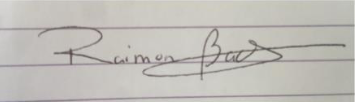
**1)** The project group RGIC22-05 provides a Data Management Plan (DMP) as part of the Project Proposal. If no Project Proposal has to be written (anymore), the project group provides a DMP by 28-06-2022, as agreed with the WU-supervisor.

**2)** The project group provides core data sets to the daily supervisor (or commissioner) at the end of the project or earlier as agreed upon with the WU-supervisor (or commissioner).

**3)** All data, scripts, products and results are owned by Wageningen University and custodianship is with Wageningen University. The project group has the right to use them.

As agreed upon,

Project group: 



Name: Raimon Bach Pareja Name: Mark Boeve

Title: Student Title: Student

Date: 19-05-2022 Date: 19-05-2022

Name: Jakko-Jan van Ek Name: Julia Sipkema

Title: Student Title: Student

Date: 19-05-2022 Date: 19-05-2022



Name: Niels Verouden

Title: Student

Date: 19-05-2022

Signature(s):

Name(s) WU-supervisor/Commissioner:

Title:

Date: