

ScheldeMonitor Handleiding: Gebruik van de RStudio omgeving

Geschreven, onderhouden en geüpdated door het ScheldeMonitor team van VLIZ - Versie 3
(24/08/2021)

Contents

1 Over de RStudio omgeving	1
2 Verbinding maken met de RStudio omgeving van ScheldeMonitor	2
3 Werken met de RStudio omgeving van ScheldeMonitor	5
3.1 Richtlijnen voor de workspace	5
3.2 Richtlijnen voor scripts	13
3.3 Richtlijnen voor de code	18
4 Data van ScheldeMonitor gebruiken in RStudio	24
4.1 Data gebruiken van gedownloade databestanden	28
4.2 Data gebruiken van generic webservices	30
5 Helpdesk	31

1 Over de RStudio omgeving

Dit document werd geschreven als handleiding voor het gebruik van de online RStudio omgeving van het ScheldeMonitor informatie en dataportal. Deze omgeving werd toegankelijk via de ScheldeMonitor website sinds 2021. Het voorziet geaccrediteerde onderzoekers en de partners van ScheldeMonitor van een gecentraliseerd RStudio hub om analyses uit te voeren en scripts te schrijven, rechtstreeks gebaseerd op de data en informatie opgeslagen in het portaal en de onderliggende database.

In deze handleiding worden richtlijnen gegeven voor het gebruik van RStudio voor nieuwe en onervaren gebruikers maar ook enkele algemene tips om gestructureerd en duidelijk te werken in RStudio. In het laatste hoofdstuk van deze handleiding wordt stapsgewijs uitgelegd hoe data van ScheldeMonitor kan ingevoegd worden in de RStudio workspace.

De RStudio omgeving is toegankelijk via de website, via volgende link. Om toegang te krijgen zijn inloggegevens vereist. Deze gegevens kunnen hieraangevraagd worden, of door een mail te zenden naar info@scheldemonitor.org met vermelding waarom het gebruik van de RStudio omgeving nodig is.

Het is ook mogelijk om de RStudio omgeving te linken met een bestaand project van de ScheldeMonitor GitHub organisatie. Hoe u kan werken met GitHub in combinatie met RStudio wordt uitgelegd in een bijkomende GitHub handleiding.

2 Verbinding maken met de RStudio omgeving van ScheldeMonitor

Met uw persoonlijke inloggegevens kunt u toegang krijgen tot de ScheldeMonitor omgeving op (<https://rstudio.scheldemonitor.org/auth-sign-in>)

The image shows a web-based sign-in form titled "Sign in to RStudio". It contains fields for "Username" and "Password", a "Stay signed in" checkbox, and a prominent blue "Sign In" button.

Deze inloggegevens zijn dezelfde als diegene die gebruikt worden voor alle andere tools op het ScheldeMonitor platform zoals de data download toolbox en de E-room.

Nieuwe gebruikers moeten zich eerst registreren om hun inloggegevens te ontvangen. Dit kan gedaan worden door deze link te gebruiken of door de helpdesk van ScheldeMonitor te contacteren met vermelding van de reden voor het gebruik van de RStudio omgeving.

De volgende gegevens zijn vereist om te registreren voor de RStudio omgeving van ScheldeMonitor:

The image shows a registration form titled "RShiny/RStudio registration". It requires users to enter their first name, last name, password, and email. There is a checkbox for "my details are listed in IMIS". A section titled "Request access to" lists several options with checkboxes, including "Access to European ocean Tracking Network (ETN) data (ETN_data)", "Able to use the IMIS fairchecker (Rshiny app) (fairchecker)", "Labeling Images (MOC) (labeling_app)", "Full access to Lifewatch Uva-BiTS birds data (Lifewatch_birds)", "Download raw data from Flemish Banks Monitoring Network (mvb)", and "Access Rstudio server (RStudio)". A "Purpose" field is present with a note about GDPR. At the bottom, there is a "Register" button and links for "[Login]" and "[Lost password]".

Na uw registratie moet uw account worden goed gekeurd door een moderator van de ScheldeMonitor RStudio omgeving. Dit proces kan één of twee werkdagen duren.

RShiny/RStudio edit account

Your account has been updated successfully.

Access to the new group(s) has been requested, **please allow a moderator to approve it.**

Na goedkeuring kunt u met uw persoonlijke gegevens inloggen op de RStudio omgeving van ScheldeMonitor.

Eenmaal ingelogd kunt u uw persoonlijke workspace in RStudio zien. Deze workspace kan leeg zijn (bv. als u een nieuwe gebruiker bent) of de structuur en inhoud laten zien waaraan u gewerkt had tijdens uw vorige sessie, als u de workspace heeft opgeslagen bij het beëindigen van uw laatste sessie.

Een persoonlijke workspace is standaard samengesteld uit vier vensters, die uw scripts of dataframes, uw omgeving, de console en uw project en persoonlijke bestandstructuur weergeven. Het geeft ook aan welke gebruiker is ingelogd en welk project is gelinkt aan uw workspace:

Script & file viewer

Console

The screenshot shows the RStudio interface with two main red-highlighted sections: the "Script & file viewer" (top) and the "Console" (bottom).

Script & file viewer (Top): This pane displays an R Markdown script. The code includes YAML front matter, R code chunks (e.g., `library(knitr)`), and Rmd content. A red box highlights the entire content area.

```
1 ---  
2 title: "ScheldeMonitor Manual: Using the RStudio environment"  
3 output:  
4   html_document:  
5     highlight: haddock  
6     toc: true  
7     toc_float: true  
8     number_section: true  
9     css: "style.css"  
10 date: "Written, maintained and updated by VLIZ - Version 3 (`r format(sys.time(), '%Y-%m-%d %H:%M')`)"  
11 ---  
12  
13 ````{r setup, include=FALSE}  
14 library(knitr)  
15 library(kableExtra)  
16 knitr::opts_chunk$set(echo = TRUE, fig.align = 'center')  
17 ````  
18  
19 # About  
20  
21 This document was written as a manual on the use of the online RStudio environment on the ScheldeMonitor website. It has been accessible through the ScheldeMonitor website since 2021. It provides accredited users with the possibility to do analysis and build up scripts directly based on the data and information available.  
22  
23 Within the manual, guidelines are provided for new users or users that are inexperienced with R. It shows how to keep work in RStudio structurized and comprehensible. The last chapter of this manual shows how to import data from the ScheldeMonitor into the RStudio workspace.  
24  
25 The RStudio environment can be accessed from the website, using the following [link]. No specific credentials are required. These credentials can be requested [here] (https://rshiny.lantern.be) or by email to info@scheldemonitor.org with a statement on the reason why use of the RStudio environment.
```

Console (Bottom): This pane shows the R startup message and the R prompt. A red box highlights the entire content area.

```
R version 4.0.2 (2020-06-22) -- "Taking Off Again"  
Copyright (C) 2020 The R Foundation for Statistical Computing  
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.

> |
```

3 Werken met de RStudio omgeving van ScheldeMonitor

3.1 Richtlijnen voor de workspace

De standaard werking van R voor het omgaan met .RData bestanden en workspaces stimuleert en vergemakkelijkt een model voor het opsluiten van werkcontexten in verschillende working directories. Dit impliceert dat de gebruiker een bepaalde map in zijn lokale directory kan selecteren om te gebruiken als locatie waar bestanden, bewerkt door RStudio, worden opgeslagen. Deze lokale directory, of workspace, kan op elk moment veranderd worden door de gebruiker.

In versie v0.95 van RStudio werd een nieuwe 'Projects' functie ingevoerd om het beheer van meerdere working directories eenvoudiger te maken. Het is aangeraden om deze functie te gebruiken, maar dit hoofdstuk legt ook het gebruik op de standaard manier van uw persoonlijke workspaces uit.

Zoals bij de lokale RStudio installatie gebruikt de online RStudio omgeving van ScheldeMonitor standaard de gebruikers lokale home directory als workspace. Naar deze workspace wordt meestal verwijzen met ~ in R. Wanneer RStudio opstart gebeurt het volgende:

- Voert het .Rprofile (indien aanwezig) uit van de standaard working directory.
- Laad het .RData bestand (indien aanwezig) vanuit de standaard working directory in de workspace.
- Voert de andere acties uit beschreven in R Startup.

Wanneer RStudio afgesloten wordt en er wijzigingen zijn aangebracht aan de workspace, vraagt een dialoogkader of deze wijzigingen moeten opgeslagen worden in het .RData bestand in de huidige working directory. Door op "save" te klikken zullen de wijzigingen opgeslagen worden en verschijnen zoals ze waren de volgende keer u inlogt in de RStudio omgeving.

3.1.1 Workspace opstellen

RStudio toont de huidige working directory in het titelgebied van de console. U kunt het commando getwd() uitvoeren in de RStudio console om uw huidige working directory te controleren:

```
getwd()
```

```
## [1] "/data/home/innovauth/jeller@vliz.be/ScheldeMonitor-Manuals (organization)"
```

Om van working directory te veranderen kunt u het commando setwd() uitvoeren in de RStudio console, met een nieuwe directory als string ingevoerd:

```
setwd("//fs/HOME/jeller")
```

U kunt ook de werk map veranderen door het "Session" menu en "Set Working Directory" te selecteren. Houd rekening met de gevolgen van het veranderen van uw working directory:

- Relatieve bestands verwijzingen in uw code (voor data, bron bestanden, etc.) zullen ongeldig worden wanneer u working directories wijzigt.
- De locatie waar .RData is opgeslagen zal bij het afsluiten veranderen naar de nieuwe directory.

Doordat deze gevolgen voor verwarring en foutmeldingen kunnen zorgen, is het beter om te starten in de working directory geassocieerd met uw project en deze te behouden voor de volledige sessie.

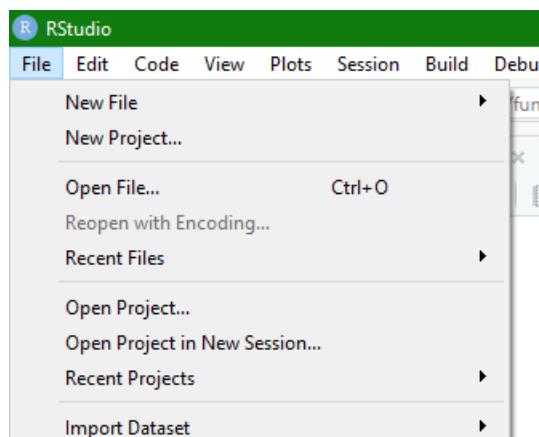
De beste werkwijze is echter om de RStudio omgeving te koppelen met een bepaald 'project'. Dit zorgt voor een beter overzicht van de working directories en de verschillende zaken waaraan u werkt in dezelfde RStudio omgeving. De volgende segmenten beschrijven hoe dergelijke projecten tot stand komen.

3.1.2 Start een lokaal project

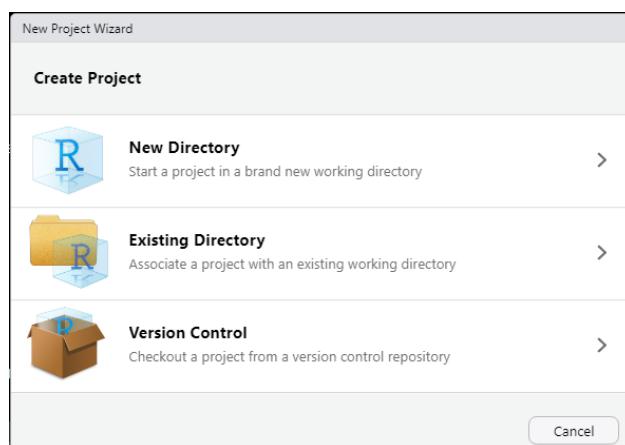
Elke erkende gebruiker kan de RStudio omgeving gebruiken om te beginnen of verder te werken met zijn of haar persoonlijk project met de data van ScheldeMonitor. Hierdoor kunnen gebruikers starten aan een nieuw lokaal project, een bestaand project downloaden van hun eigen GitHub, of hun eigen werk koppelen aan de GitHub organisatie van ScheldeMonitor.

Met een lokaal project starten is de makkelijkste manier om met uw werk te beginnen. Dit project is opgeslagen op een lokale drive van de gebruikers hardware, en kan enkel heropgestart worden door de drive te openen. Om een lokaal project te beginnen moeten gebruikers de volgende stappen volgen:

1. Open het 'File' menu en selecteer 'New Project'.

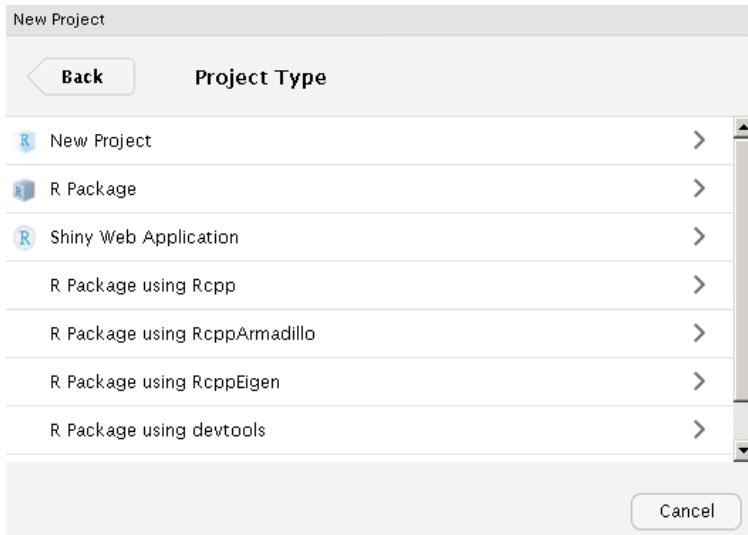


2. Selecteer 'New Directory' in de 'New Project Wizard'. Een nieuw project zal gestart worden en zal opgeslagen worden op de lokale drive van de gebruikers hardware.

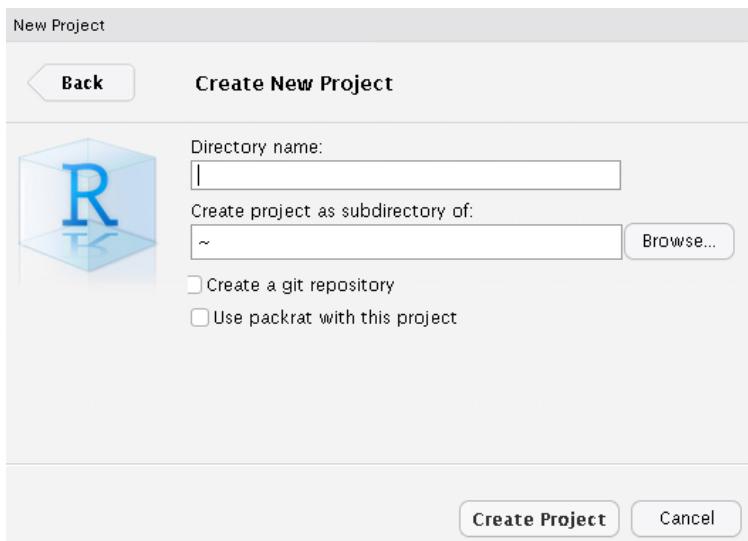


3. De gebruiker kan kiezen welk type project gestart moet worden

- New Project: Basis R project waar verschillende scripts kunnen opgesteld worden.
- R Package: Project waar gebruikers R Packages kunnen maken en publiceren voor andere gebruikers.
- Shiny Web Application: Een R project waar alle scripts voorgeschreven zijn voor gebruikers om Shiny web applicaties te creëren en runnen.



4. Tenslotte kan de gebruiker het project een naam geven en een directory selecteren waarin het project een subdirectory zal maken. De optie “create a git repository” selecteren, zal het voor de gebruiker mogelijk maken om een lokaal geïnstalleerde versie te gebruiken. Deze optie is niet nodig wanneer er met een online GitHub space gewerkt wordt.



3.1.3 RStudio omgeving verbinden met GitHub met behulp van SSH-verificatie

3.1.3.1 Disclaimer

Deze handleiding werd geschreven voor de gebruikers van de online RStudio omgeving van ScheldeMonitor. De procedure en de bijbehorende scripts werden geïmplementeerd op de server van deze omgeving zodat gebruikers niets moeten installeren om de procedure uit te voeren. Dit betekent dat de procedure niet zal werken op andere (lokale) RStudio omgevingen of servers. Gebruikers die de procedure willen uitvoeren in een andere RStudio omgeving, niet verbonden met de ScheldeMonitor server, kunnen een aanvraag indienen voor een Git bundel van de scripts bij info@scheldemonitor.org.

3.1.3.2 Redenen voor SSH-verificatie

In 2020 kondigde GitHub aan dat het niet langer account paswoorden zou accepteren voor verificatie met de

REST API en dat het gebruik van token-gebaseerde authenticatie vereist zal zijn. Dit impliceert dat wanneer het gebruik van ScheldeMonitor RStudio workspace gecombineerd wordt met de GitHub repository, alle acties (bv. pull, push, commit) een andere soort authenticatie vereisen dan standaard wordt gebruikt.

VLIZ heeft vastgesteld dat het instellen van een SSH-sleutel de beste manier is voor gebruikers om verder te gaan. Deze handleiding legt in detail de eenmalige procedure uit dat nodig is voor een correcte installatie. Deze procedure omvat vier scripts die moeten gerund worden door een gebruiker in de terminal van de RStudio omgeving van ScheldeMonitor.

3.1.3.3 Toegang tot de RStudio terminal

De ScheldeMonitor omgeving kan worden geopend door de stappen te volgen beschreven in het hoofdstuk Verbinding maken met de RStudio omgeving van ScheldeMonitor

Zodra u in de workspace bent, schakelt u de console over naar het terminalvenster door op de tab 'Terminal' bovenaan het venster te klikken. In dit venster zal je de nodige commando's uitvoeren:



3.1.3.4 SSH-sleutel instellen

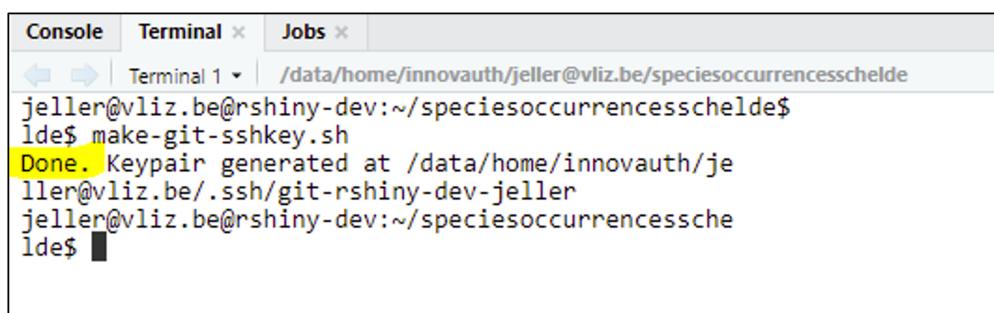
1. Sleutelpaar genereren

Indien gebruikers nog nooit een SSH-sleutel aangemaakt hebben in deze RStudio omgeving, moeten ze het 'make-git-sshkey.sh' script uitvoeren. Dit script zal een standaard sleutelpaar genereren om te gebruiken om verbinding te maken met git services.

Om het script te runnen, voer volgend commando uit in de terminal:

```
make-git-sshkey.sh
```

Een bericht zal verschijnen dat het sleutelpaar werd gegenereerd:



2. Sleutelpaar registreren

Vervolgens zal de gebruiker het gegenereerde sleutelpaar moeten registreren op GitHub online. Om dit te doen kan het script 'connect-sshkey.sh' gebruikt worden om lokaal correct te configureren en de gebruiker te adviseren over hoe het openbaar deel van de sleutel openbaar te registreren ten dienste van de gebruikers keuze (standaard in GitHub).

Voer het volgende commando in de terminal uit om te runnen:

```
connect-sshkey.sh
```

Een bericht zal verschijnen en gedetailleerd weergeven hoe de sleutel moet worden geregistreerd op GitHub:

```
Config Ready. (for key in /data/home/innovauth/jeller@vliz.be/.ssh/git-rshiny-dev-jeller)
Instructions to connect your ssh-key to the git service at github.com

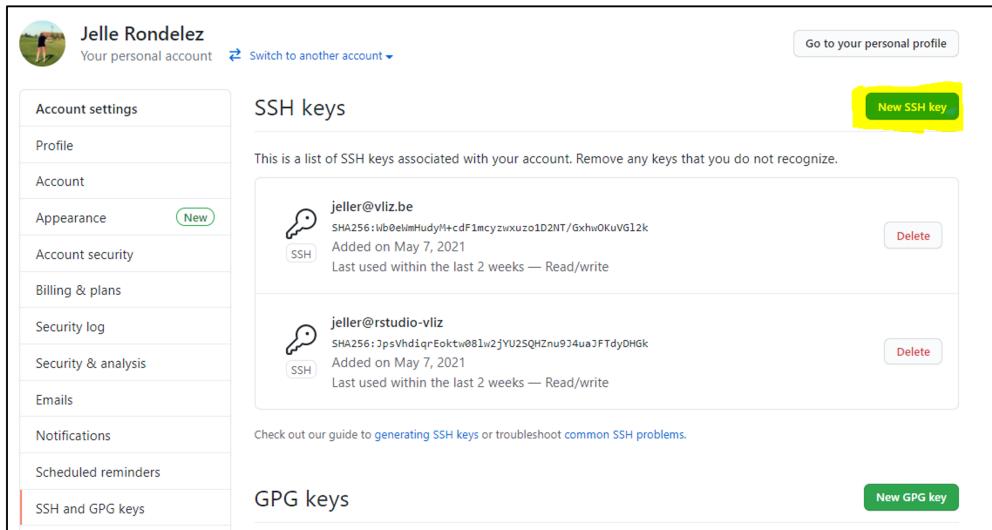
Surf to https://github.com/settings/keys and verify if the ssh key for jeller@rshiny-dev is available.
If not register it on that page, by pasting the section between (not including) the '----' lines below.
----
ssh-rsa AAAAB3NzaC1yc2EAAAQABAAQC8ehYjbnPnA2nbH0eRJ+wMwnp3YKK6XsLSD+dPlrtUEgR4zo2W3nEZqct
e6t3Xc8QdXNu5Gqxe5B1KGqiFuY7qrJJd+2MQ4FJ2DcfMgAbh4j9sLuAS3z60PflVhnusN/rU3B236UE7u+AIHBSc+Vv2w
Oh7deMPtT7agfxDbSUmK/6wv2/xGljrttKaIw2YY80r059yKwKmzsri1xKB4einLibsDuelDcoSmVeRJtze01vAiEDtrh
qm2xqNHe5T/a63eBs+CnPktJ2t5Yj0V4RKGnK3H1V670KrS7LVTF3wJOSSsC5pkL9erCBLXKdmPiJh1TbTmMHL9EZcKRCBM7
jeller@rshiny-dev
---
After this, please verify the setup with check-gitssh.sh
jeller@vliz.be@rshiny-dev:~/speciesoccurrenceschelde$
```

3. Sleutel kopiëren

Zoals beschreven in het bericht, kopieer de tekst tussen de twee '----'.

4. sleutel toevoegen aan GitHub

Surf naar <https://github.com/settings/keys>, en selecteer de 'New SSH key' knop in de rechterbovenhoek.



The screenshot shows the GitHub user profile of Jelle Rondelez. On the left, there's a sidebar with account settings like Profile, Account, Appearance (which has a 'New' badge), Account security, Billing & plans, Security log, Security & analysis, Emails, Notifications, and SSH and GPG keys. The main area is titled 'SSH keys' and contains two entries. Each entry has a key icon, the key name (jeller@vliz.be or jeller@rstudio-vliz), the SHA256 fingerprint, the date it was added (May 7, 2021), and a 'Delete' button. Below the keys, there's a note about generating SSH keys and troubleshooting common SSH problems. At the bottom right, there's a 'New GPG key' button.

Een nieuw venster met twee tekstvakken zal openen. Plak de tekst dat u heeft gekopieerd van de RStudio terminal in stap 3 in het 'Key' veld.

Optioneel kunt u een naam geven aan de sleutel in het 'Title' veld, die van toepassing is op de RStudio omgeving waarin u de sleutel heeft gemaakt (het beste is dat elke omgeving zijn eigen sleutel heeft).

Druk op de 'Add SSH key' knop wanneer u klaar bent. Dit leidt u terug naar de het vorige venster, waar u kan zien dat de nieuwe sleutel is toegevoegd.

The screenshot shows the 'SSH keys / Add new' dialog box. It has two main sections: 'Title' and 'Key'. The 'Title' section contains an empty input field. The 'Key' section contains a large text area with a long ssh-rsa key. Below the text area is a green 'Add SSH key' button.

5. SSH-sleutel controleren

Het script 'check-gitssh.sh' kan gebruikt worden om na te gaan of de vorige stappen correct werden uitgevoerd. Dit script controleert of de verbinding correct ingesteld is. Deze stap kan zo dikwijls als nodig of gewild worden uitgevoerd.

Voer het volgende commando uit in de terminal van RStudio om te runnen:

```
check-gitssh.sh
```

Indien uitgevoerd voor de eerste keer kan de terminal vragen om de actie te controleren. Om dit te doen, schrijf 'yes' na het vraagteken en voer uit:

```
jeller@vliz.be@rshiny-dev:~/speciesoccurrencesschelde$ check-gitssh.sh
The authenticity of host 'github.com (140.82.121.3)' can't be established.
RSA key fingerprint is 16:27:ac:a5:76:28:2d:36:63:1b:56:4d:eb:df:a6:48.
Are you sure you want to continue connecting (yes/no)? yes
```

Indien alles correct is verlopen keert de terminal terug naar de juiste gebruikersnaam:

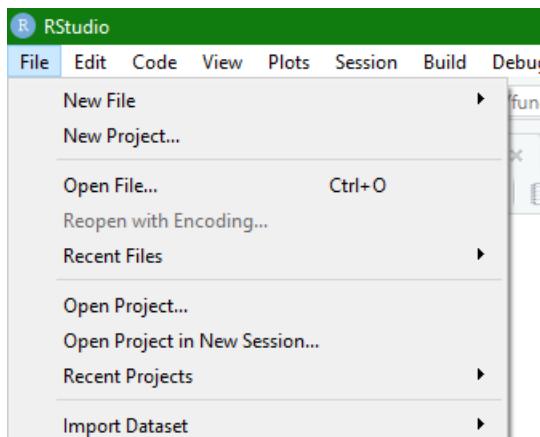
```
jeller@vliz.be@rshiny-dev:~/speciesoccurrencesschelde$ check-gitssh.sh
Looking Good. Well Done!
Apparently the service @github.com connects you to this account: Djro33
jeller@vliz.be@rshiny-dev:~/speciesoccurrencesschelde$ []
```

De workspace is nu verbonden via SSH. Nu kan de gebruiker via SSH aan de slag in nieuwe projecten of al aangesloten projecten omzetten naar SSH. Beide methoden worden uitgelegd in de volgende hoofdstukken.

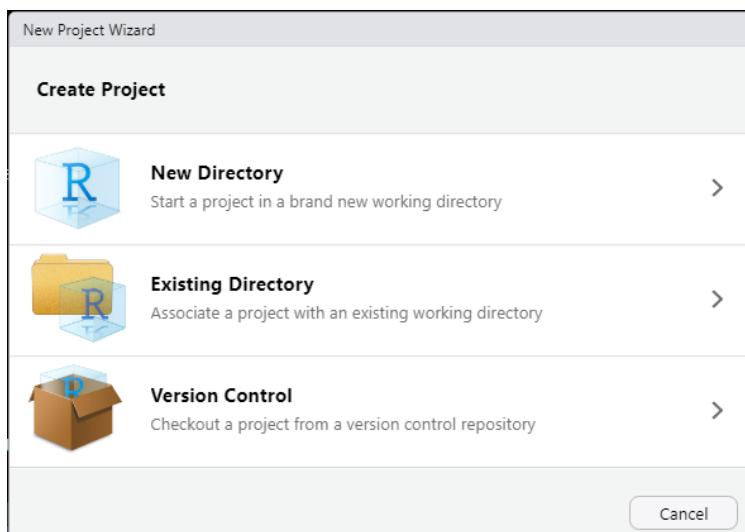
3.1.3.5 SSH authenticatie toepassen op nieuwe projecten

Vanaf dit moment zouden nieuwe projecten onmiddellijk moeten starten door de juiste remote origin-url te gebruiken. De onderstaande stappen leggen uit hoe het moet.

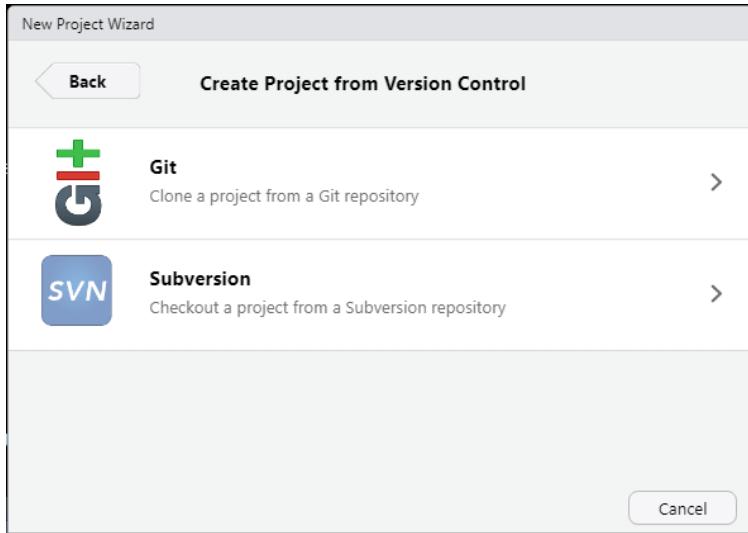
1. Open het 'File' menu en selecteer 'New Project'.



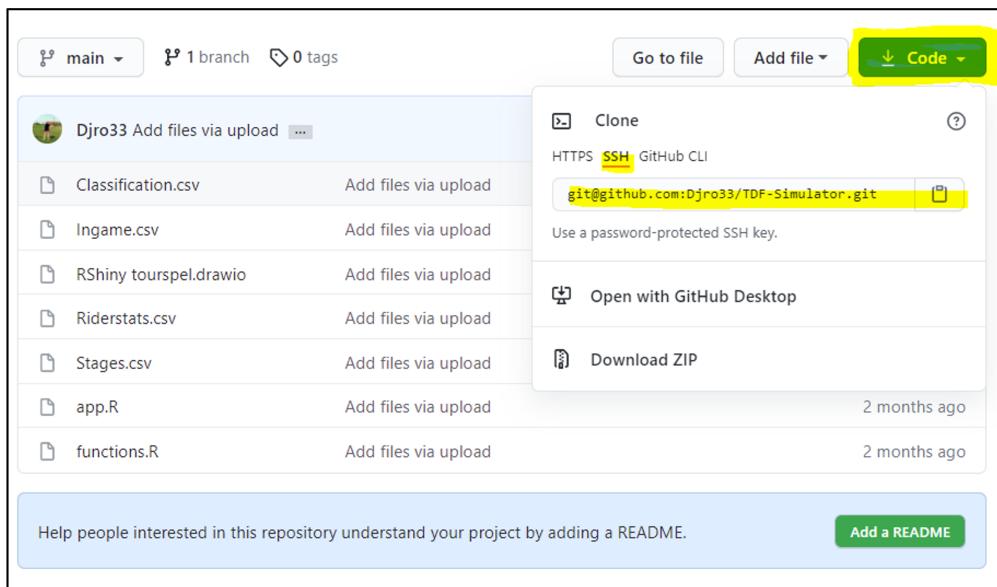
2. Selecteer de optie "Version Control" in de 'New project Wizard'.



3. Selecteer vervolgens "Git", aangezien dit de versie gecontroleerde methode is die gebruikt wordt in de context van ScheldeMonitor.



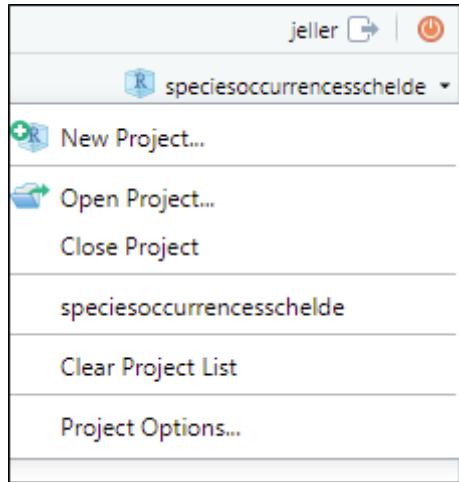
4. Plak de URL van de repository in het 'Repository URL' veld. In vele gevallen plakken gebruikers de HTTPS URL van de repository. Wanneer echter SSH authenticatie gebruikt wordt moet de SSH URL ingevoerd worden. Deze SSH URL kan terug gevonden worden op de home pagina van de online repository door de groene 'Code' knop en de SSH optie te selecteren.



Het project kan aangemaakt worden nadat een naam en een locatie voor uw lokaal project is gekozen.

3.1.3.6 Bestaande projecten naar SSH-authenticatie omzetten

Projecten die alreeds verbonden waren met de RStudio workspace voor de omzetting naar SSH, zullen nog steeds werken via HTTPS. Om dit te wijzigen, open een bestaand project dat u wenst om te zetten in de RStudio omgeving. Dit kan gedaan worden in de rechterbovenhoek van het scherm:



Eenmaal het project is geopend, kan het script 'fix-gitssh.sh' gebruikt worden, dat eenvoudig de verbinding naar uw project zal veranderen naar gitt-ssh, en tijdens het proces uitzoeken met welke dienst het verbonden is.

`fix-gitssh.sh`

De terminal zou naar de originele verbinding moeten terugkeren en een bericht weergeven dat het de verbinding met git-shh herstelt:

```
jeller@vliz.be@rshiny-dev:~/speciesoccurrencesschelde$ fix-gitssh.sh
Current git repo is linked to remote origin at https://gitlab.vliz.be/datac/rshiny/scheldemonitor/speciesoccurrencesschelde
Fixing the link for remote origin to git@gitlab.vliz.be:datac/rshiny/scheldemonitor/speciesoccurrencesschelde
jeller@vliz.be@rshiny-dev:~/speciesoccurrencesschelde$
```

3.2 Richtlijnen voor scripts

Een working directory of project in RStudio kan een groot aantal scripts en bestanden bevatten om mee te werken. Het is belangrijk om deze scripts gestructureerd te houden intern en in de directory om het werk georganiseerd en productief te houden. De segmenten hieronder stellen richtlijnen voor die onderzoekers kunnen helpen om hun werk transparant te houden voor zichzelf en andere gebruikers.

3.2.1 Directory structuur

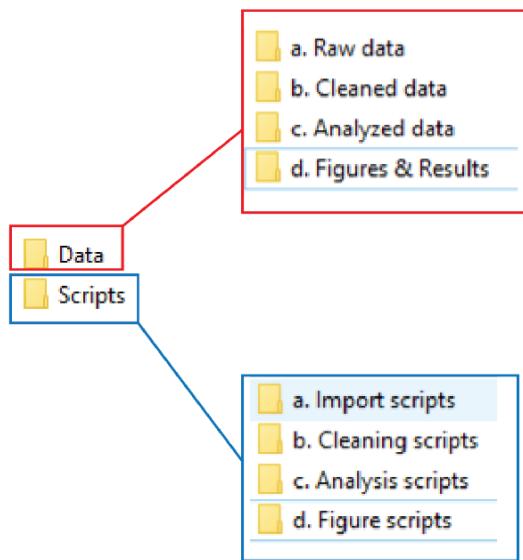
Een working directory of project is vergelijkbaar met andere mappen op de lokale drive van uw hardware. Dit betekent dat zo'n directory kan bestaan uit mappen en submappen. Het is echter noodzakelijk dat mappen worden aangemaakt volgens een bepaalde structuur of idee om de scripts en bijbehorende data gemakkelijk vindbaar te maken voor uzelf en andere gebruikers. Er zijn meerdere niveaus waarop een directory gestructureerd kan worden.

Ten eerste, als je werk in RStudio verbonden is aan een bepaalde publicatie of rapport zou de structuur ongeveer hetzelfde moeten zijn als de koppen van het rapport. Hier is een voorbeeld van het T2015 rapport van de Schelde waarvoor de project directory gestructureerd was conform de titels en de ondertitels in het gepubliceerd rapport:

- _Algemeen
- 3_Dynamiek waterbeweging
- 4_Bevaarbaarheid
- 5_Plaat- en geulsysteem
- 6_Waterkwaliteit
- 7_Leefomgeving
- 8_Flora en Fauna
- 9_Ecologisch functioneren

Toch is het nog belangrijker om een uniforme structuur te hebben op het laagste niveau van de working directory, waar alle bestanden worden opgeslagen. Vooral voor projecten dat niet gelinkt zijn aan een vast rapport, en waarvoor de hierboven vermelde structuur niet van toepassing is.

Gewoonlijk worden databestanden en scripts in verschillende mappen opgeslagen. Hoewel het misschien handiger lijkt om deze bestanden bij elkaar te houden, heeft het algemene overzicht meer voordeel bij de twee-mappen structuur. Scripts en databestanden hebben vaak geen 1 op 1 verhouding, aangezien één script meerdere databestanden kan gebruiken terwijl deze databestanden door meerdere scripts kunnen gerund worden. De structuur van elke map zou echter hetzelfde moeten zijn met één map voor elke fase van het project:



Door gebruik te maken van deze structuur kan een uniforme werkwijze in de project directory gerealiseerd worden. De werkwijze bestaat uit vier stappen die worden uitgelegd in volgende tabel:

	Using data from:	Using scripts or functions from:	Saving new data or results in:
Step 1 - Import data (if necessary)	n/a	a. Import scripts	a. Raw data
Step 2 - Clean data	a. Raw data'	b. Cleaning scripts	b. Cleaned data
Step 3 - Analyze data	'b. Cleaned data	c. Analysis scripts	c. Analyzed data
Step 4 - Create figures or results	c. Analyzed data	d. Figure scripts	d. Figures & Results

Het is mogelijk dat gebruikers liever een single script runnen om al deze stappen te doorlopen, vooral in kleinere projecten. In dit geval kan een 'Main.R' script opgeslagen worden naast de mappen 'Data' en 'Scripts'. Dit main script kan vervolgens al deze stappen zelfstandig doorlopen, terwijl het verschillende databestanden en functies uit onderliggende mappenstructuur zoekt. Het laatste is vooral belangrijk in grotere projecten om ervoor te zorgen

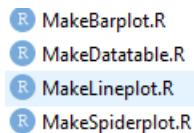
dat de lengte en leesbaarheid van het main script optimaal is. Daarbij is het van groot belang dat het main script goed gestructureerd is en van aantekeningen is voorzien, zoals verder zal worden uitgelegd in Script structuur en Script aantekeningen hieronder.

In elk geval mag er maar één ‘Main.R’ bestand aanwezig zijn om verwarring te voorkomen.

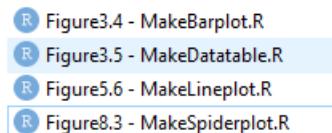
3.2.2 Script benaming

Scripts moeten zo worden genoemd zodat gebruikers makkelijk het doel ervan kunnen afleiden, om niet alle scripts in de RStudio omgeving te hoeven openen om te weten waarvoor ze gebruikt worden. Dit is voornamelijk belangrijk wanneer er gewerkt wordt met een main script dat tijdens verschillende fasen functies uit andere scripts zoekt.

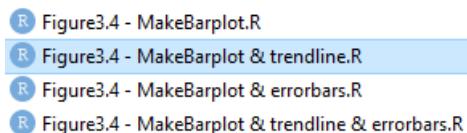
Bijvoorbeeld, wanneer er verschillende scripts gebruikt worden voor verschillende grafieken, moet de naam duidelijk aangeven welke plot gecreëerd wordt bij gebruik van het script:



Daarnaast, als het werk in de RStudio verbonden is aan een bepaald rapport of publicatie, kan het figuur nummer van de publicatie in de bestandsnaam worden opgenomen:



Het is ook mogelijk dat meerdere scripts voor dezelfde figuur worden gebruikt, bijvoorbeeld als gebruikers zowel de originele als de nieuwe plot op een later tijdstip willen kunnen tonen. Toch moet de nomenclatuur duidelijk het verschil in de verschillende scripts aangeven:



Welke nomenclatuur ook wordt gekozen, het moet bestaan uit een vaste en uniforme naamgevingsconventie. Er zijn verschillende opties mogelijk, gelijkaardig aan degene beschikbaar voor de code nomenclatuur zoals uitgelegd in Naamgevingsconventies:

- alllowercase: e.g. makebarplot
- period.separated: e.g. make.barplot
- underscore_separated: e.g. make_barplot
- lowerCamelCase: e.g. makeBarPlot
- UpperCamelCase: e.g. MakeBarPlot

3.2.3 Script structuur

Gelijkwaardig aan de directory kan een individueel script veel baat hebben bij een vaste en uniforme structuur. Deze structuur moet duidelijk de verschillende secties in het script afbakenen, wat de lezer een snel overzicht van de inhoud geeft, maar er ook de gebruiker verzekert dat alle acties en functies in een vaste volgorde worden uitgevoerd. De script structuur kan bijna onmiddellijk worden bereikt door koppen in de code te gebruiken. Deze worden op dezelfde manier ingevoerd als de aantekeningen. Idealiter zouden alle scripts dezelfde koppen moeten hebben om mee te beginnen:

- Wie, wanneer, wat en hoe: Dit is een grote kop die in het begin van elk script moet staan, en aangeven wie het script schreef, wanneer het werd geschreven, hoe de schrijver gecontacteerd kan worden en wat het doel is.
- 0 – Load libraries: In deze sectie worden alle ‘libraries’ opgeliist die moeten geladen worden voordat het volledige script wordt gerund. Deze sectie kan ook wat meer uitleg geven over het gebruik van de ‘libraries’.
- 1 – Static part: In dit deel worden alle statische zaken uitgevoerd zoals databestanden laden, deze data voorbereiden voor analyse, zoeken naar andere scripts functies of argumenten benoemen die later in het script zullen worden gebruikt.
- 2 – Script: Deze sectie bevat de eigenlijke code die ervoor zorgt dat het script zijn doel vervult.

```
#####
## This is an example for the manual
##
## written by Jelle Rondelez of VLIZ
## info@scheldemonitor.org - Oct 2020
#####

#####
# 0 - Load libraries
#####
library(dplyr) # package to clean datatable
library(lubridate) # package to change date formats

#####
# 1 - Static part
#####

#Assign variable
newvar <- ""

#Source script from within directory
source("Script/a. Import scripts/ImportWFS")

#Open datafile
datafile <- read.csv(file = "Data/b. Cleaned data/dataRWS.csv")

#####
# 2- Scripts
#####
code...
```

Merk op dat de gezochte bestanden in het voorbeeld hierboven de directory structuur gebruiken zoals beschreven in Directory structuur.

Deze koppen geven niet enkel een vaste structuur en orde aan alle scripts in het project, maar het heeft ook het extra voordeel dat secties indien nodig kunnen samengevouwen of uitgebreid worden. Vooral voor langere

scripts, waarin bepaalde codes niet van belang zijn voor de gebruiker, kan dit de leesbaarheid van het script enorm verhogen:

```
1 #####
2 #####
7 #####
8 #####
9 #####
10 # 0 - Load libraries
11 #####
15 #####
16 # 1 - Static part
17 #####
29 #####
30 # 2 - Script
31 #####
32 code...
~~
```

Grottere scripts kunnen meer baat hebben bij een uitgebreide structuur met extra koppen. Dit geldt voornamelijk voor 'Main.R' scripts die alle fasen van het project in één script doorlopen, zoals besproken in Directory structuur. Dit soort scripts zoeken en gebruiken meestal een groot aantal verschillende functies en bestanden. Een uitgebreide structuur kan de scripts beter leesbaar maken en het zoeken naar specifieke functies of acties makkelijker maken:

```
#####
## This is an example for the manual
##
## written by Jelle Rondelez of VLIZ
## info@scheldemonitor.org - Oct 2020
#####

#####
# 0 - Load libraries
#####
library(dplyr) # package to clean datatable
library(lubridate) # package to change date formats

#####
# 1 - Static part
#####

#Assign variable
newvar <- ""

#Source script from within directory
source("Script/a. Import scripts/ImportWFS")

#Open datafile
datafile <- read.csv(file = "Data/b. Cleaned data/dataRWS.csv")

#####
# 2- Scripts
#####
code...

#####
# 3 - Analysis part
#####
```

```

code...

#####
# 4 - Make plots & Figures
#####
code...

```

3.2.4 Script aantekeningen

De code van aantekeningen voorzien is om een aantal redenen belangrijk. De belangrijkste reden is voor de gebruiker persoonlijk bij het terugkijken op wat was gecodeerd. Het helpt in detail uit te leggen wat een regel, chunk of zelf een sectie van de code probeert te bereiken. Dit kan ook helpen voor andere mensen die de code lezen. De uitleg van wat een code regel doet kan handig zijn voor anderen die de code willen aanpassen voor hun eigen werk, of wanneer iemand een code chunk controleert of evalueert. Aantekeningen maken in de code wordt gedaan met `#` (hashtag). Aantekeningen kunnen gewoonlijk boven een volledige chunk gemaakt worden, zoals bij het uitleggen van het doel van een bepaalde functie.

```

#Reactive values for uses locations
data_of_click <- reactiveValues (clicked = NULL)
longitude_click <- reactiveValues (lng = NULL)
latitude_click <- reactiveValues (lat = NULL)

#if user clicks on map, new coordinates are saved and maps is adjusted
observeEvent(input$Map_click, {
  data_of_click$clicked <- input$Map_click
  longitude_click <- input$Map_click$lng
  latitude_click <- input$Map_click$lat
  leafletProxy('Map') %>%
    clearMarkers() %<%
    addMarkers(lng = input$Map_click$lng,
               lat = input$Map_click$lat,
               popup = paste("Longitude=", round(input$Map_click$lng, 2),
                             "and",
                             "Latitude=", round(input$Map_click$lat, 2)))
})

```

3.3 Richtlijnen voor de code

Helaas heeft R, in tegenstelling tot andere programmeertalen, geen algemeen aanvaarde coderingswerkwijze. In plaats daarvan zijn er een aantal pogingen geweest om enkele regels samen te stellen. Dit hoofdstuk probeert deze tekortkomingen aan te vullen door samen te vatten wat belangrijk werd gevonden in die verschillende pogingen.

3.3.1 Hardcoding

Een bestand of map oproepen vanuit het script wordt meestal gedaan door ‘hardcoding’, e.g. de locatie van het bestand geven als string. De gebruikers worden echter sterk aangeraden om hoeveelheid hardcoding te minimaliseren, aangezien het minder moeite kost om een script te veranderen wanneer een directory locatie veranderd met minder hardcoding. Als uw code data zal inlezen van een bestand, definieer vroeg in de code een variabele dat het pad naar dat bestand opslaat. Hierbij het volgende voorbeeld:

```

input_file <- "data/data.csv"
output_file <- "data/result.csv"

```

```

#read input
input_data <- read.csv(input_file)

#get number of samples in data
sample_number <- nrow(input_data)

#generate results
results <- some_other_function(input_file, sample_number)

#write results
write.table(results, output_file)

```

heeft voorkeur boven:

```

#read input
input_data <- read.csv("data/data.csv")

#get number of samples in data
sample_number <- nrow(input_data)

#generate results
results <- some_other_function("data/data.csv", sample_number)

#write results
write.table(results, "data/result.csv")

```

3.3.2 Naamgevingsconventies

R heeft geen naamgevingsconventies voor algemene variabelen en functies. Als nieuwkomer in R is het handig om te beslissen welke naamgevingsconventie aan te nemen. In het algemeen zijn er vijf naamgevingsconventies om uit te kiezen. Het is belangrijk om één conventie te kiezen en deze te gebruiken voor het volledige project:

- alllowercase: e.g. adjustcolor
- period.separated: e.g. plot.new
- underscore_separated: e.g. numeric_version
- lowerCamelCase: e.g. addTaskCallback
- UpperCamelCase: e.g. SignatureMethod

Bovenal, en naast de naamgevingsconventie, is het belangrijk om een korte en betekenisvolle naam te kiezen voor de variabele en de functie.

3.3.3 Spacing

Als het neerkomt op code schrijven in R is er zoals de naamgevingsconventie geen syntax conventie. Grote scripts hebben echter baat bij het gebruik van een duidelijke en consistente syntax, aangezien het de code meer open en leesbaar maakt. De juiste spacing gebruiken in uw code maakt een waardevol verschil in de syntax. Het kan uitgevoerd worden door deze regels te volgen:

- Plaats altijd een spatie na de komma, nooit ervoor, net als in normaal Nederlands.

```
# Good
x[, 1]
```

```
# Bad
x[,1]
x[ ,1]
x[ , 1]
```

- Plaats geen spatie binnen of buiten de haakjes voor normale functie calls.

```
# Good
mean(x, na.rm = TRUE)
```

```
# Bad
mean (x, na.rm = TRUE)
mean( x, na.rm = TRUE )
```

- Plaats een spatie voor en na () wanneer gebruikt met 'if', 'for' en 'while'.

```
# Good
if (debug) {
  show(x)
}
```

```
# Bad
if(debug){
  show(x)
}
```

- Plaats een spatie na () gebruikt voor functie argumenten.

```
# Good
function(x) {}
```

```
# Bad
function (x) {}
function(x){}
```

- Meeste infix operatoren (=,+,-,<-, etc.) moeten altijd omgeven zijn door spaties.

```
# Good
height <- (feet * 12) + inches
mean(x, na.rm = 10)
```

```
# Bad
height<-feet*12+inches
mean(x, na.rm=10)
```

Het is echter belangrijk om niet te veel spacing te gebruiken. Extra spacing toevoegen kan helpen, maar alleen als het de uitlijning van = of <- verbeterd. Plaats geen extra spaties op plaatsen waar het niet behulpzaam is.

3.3.4 Code blokken

Net als wanneer we het hebben over de algemene structuur van een script, is de hiërarchie binnen de code zelf even belangrijk. Om de belangrijkste hiërarchie te definiëren worden accolades gebruikt. Om de hiërarchie echter transparant te houden voor uzelf en andere gebruikers is een consistente syntax nodig bij het gebruik van accolades. De syntax is op drie regels gebaseerd:

- '{' moet het laatste karakter op de regel zijn. Verwante code (e.g. een if-clausule, een functieverklaring, een komma achteraan, ...) moet op dezelfde regel staan als de openingsaccolade.
- De inhoud moet worden ingesprongen met twee spaties.
- '}' moet het eerste karakter op de regel zijn.

```
# Good
if (y < 0 && debug) {
    message("y is negative")
}

if (y == 0) {
    if (x > 0) {
        log(x)
    } else {
        message("x is negative or zero")
    }
} else {
    y^x
}
```

3.3.5 Lange code regels

Gebruikers worden aangeraden om altijd de code te beperken tot 80 karakters per regel. Om dit te doen, kan het gebruik van een beknopte en efficiënte naamgevingsconventie al een belangrijke stap zijn. Als een functie call te lang is om op een regel te passen, gebruik één regel voor elke functienaam, elk argument, en het sluitingshaakje. Dit maakt het makkelijker om de code te lezen en aan te passen:

```
# Good
do_something_very_complicated(
    something = "that",
    requires = many,
    arguments = "some of which may be long"
)

# Bad
do_something_very_complicated("that", requires, many, arguments, "some of which may be long")
```

3.3.6 Pipes

Zelf met het gebruik van de juiste spacing en adequate structurering van de code blokken, kan het moeilijk zijn om het script te begrijpen. Dit is voornamelijk het geval bij scripts waar veel verschillende operaties en functies worden gebruikt. Wanneer de code voornamelijk gevormd wordt uit functionele taal, heeft het een groot aantal accolades en argumenten per functie. Dit kan de code extreem complex maken en moeilijk om te begrijpen.

Om dit probleem op te lossen worden gebruikers aangeraden om ‘piping’ te gebruiken voor meerdere acties op hetzelfde argument. Piping gebruikt de ‘%>%’ operator en kan toegepast worden door het installeren van de ‘magrittr’ of ‘dplyr’ library. Het is het beste uit te leggen aan de hand van deze drie simpele regels:

- $f(x)$ kan geschreven worden als $x \%>% f$
- $f(x, y)$ kan geschreven worden als $x \%>% f(y)$
- $h(g(f(x)))$ kan geschreven worden als $x \%>% f \%>% g \%>% h$

```
#Import 'dyplr' library
library(dyplr)

#Load the data
data(babynames)

#Count how many young boys with the name "Taylor" are born
sum(select(filter(babynames, sex=="M", name=="Taylor"), n))

#Do the same but now with '%>%
babynames %>% filter(sex=="M", name=="Taylor") %>%
  select(n) %>%
  sum
```

3.3.7 Tidyverse stijlgids & add-ons

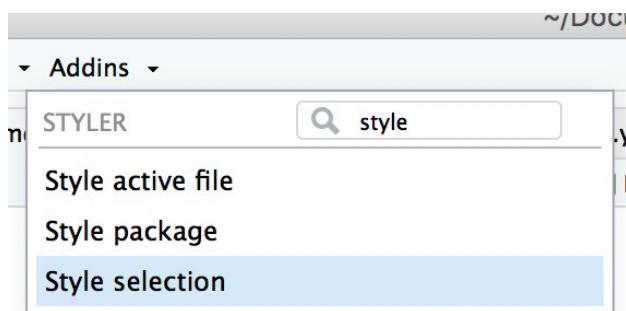
De R-gemeenschap heeft meerdere gidsen over hoe u uw code moet opmaken en beheren om het leesbaar en schoon te maken. Al deze stijlgidsen zijn fundamenteel eigenwijs. Sommige beslissingen maken de code echt makkelijker om te gebruiken, maar vele beslissingen zijn willekeurig. Het belangrijkste van een stijlgids is dat het consistentie biedt, waardoor de code makkelijker te schrijven is omdat u minder beslissing hoeft te nemen.

Gebruikers van de RStudio omgeving van ScheldeMonitor worden aanbevolen om de tidyverse stijl gids te gebruiken, aangezien het een van de meest gebruikte gidsen is. De regels die hierboven in deze gids werden aangehaald maken ook deel uit van de tidyverse stijlgids.

Er zijn twee tools die kunnen geïnstalleerd worden door gebruikers die het gemakkelijker maken om deze stijlgids te implementeren, de ‘styler’ en ‘lintr’ packages. De installatie van het ‘tidyverse’ package is niet nodig voor deze applicaties. The ‘styler’ en ‘lintr’ packages kunnen geïnstalleerd worden met de volgende R code:

```
#Install packages 'styler' and 'lintr'
install.packages("styler")
install.packages("lintr")
```

- Het ‘styler’ package maakt het mogelijk om de geselecteerde tekst, bestanden of het volledige project interactief te restylen. Deze bevat een RStudio add-in, de makkelijkste manier om de bestaande code te restylen.



- Het 'lintr' package kan automatische checks uitvoeren om te bevestigen dat de code conform de stijlgids is. Deze check wordt automatisch vertoont in het RStudio 'Markers pane'. Om dit deelvenster te zien, ga naar het 'Tools' Menu en selecteer 'Global Options ...'. Een venster met de titel 'Options' zal verschijnen. In dat venster: selecteer aan de linkerkant 'Code'; selecteer het 'diagnostics' tabblad; Klik 'show diagnostics for R' aan.

Het volgende venster is nu zichtbaar:

The screenshot shows the RStudio interface. The top panel displays the 'bad.R' script with some syntax errors highlighted. The bottom panel shows the 'Markers' pane listing 20 specific diagnostic errors related to the code in 'bad.R'.

```

bad.R
1 fun <- function(one) {
2   one.plus.one <- oen
3   four <- newVar <- matrix(1:10, nrow = 2)
4   four[1, ] <- txt
5   three <- two + 1
6   if(txt == 'hi') 4
7   5
8
9
10
11

6:14 fun(one) ▾
```

Markers

~/Dropbox/projects/lintr/bad.R

- Line 1 Use <- , not =, for assignment.
- Line 2 Opening curly braces should never go on their own line and should always be followed by a colon.
- Line 3 Words within variable and function names should be separated by '_' rather than '.'.
- Line 3 local variable 'one.plus.one' assigned but may not be used
- Line 3 no visible binding for global variable 'oen', Did you mean 'one'?
- Line 4 Variable and function names should be all lowercase.
- Line 4 local variable 'newVar' assigned but may not be used
- Line 4 Commas should always have a space after.
- Line 5 Do not place spaces around code in parentheses or square brackets.
- Line 6 Only use double-quotes.
- Line 7 local variable 'three' assigned but may not be used
- Line 7 no visible binding for global variable 'two', Did you mean 'txt'?
- Line 7 Put spaces around all infix operators.
- Line 8 Place a space before left parenthesis, except in a function call.
- Line 8 Only use double-quotes.
- Line 9 Closing curly-braces should always be on their own line, unless it's followed by a closing brace.
- Line 9 Trailing whitespace is superfluous.
- * Line 10 unexpected end of input

Het is aangeraden om het 'styler' package eerst te gebruiken, gevuld door het 'lintr' package. Omdat het 'styler'

package automatisch de stijl fouten verbeterd zoals het onjuist gebruik van spaties en komma's. Daardoor is de lijst met fouten, gegenereerd door het 'lintr' package, dat handmatig verbeterd moet worden, korter.

4 Data van ScheldeMonitor gebruiken in RStudio

De meeste data van ScheldeMonitor kan vrij gebruikt worden, en gebruikers worden aangemoedigd om de RStudio omgeving van ScheldeMonitor te gebruiken om onze data collectie verder te analyseren en valideren. Om dit te kunnen doen, moet de data eerst in de RStudio omgeving geladen worden. Dit kan gedaan worden door gedownloade data bestanden zoals CSV of TXT op te laden, of door de generic webservices van ScheldeMonitor te gebruiken. Voor beide methoden is toegang tot het Data Download Toolbox van ScheldeMonitor nodig, welke te verkrijgen is door de volgende stappen te volgen:

1. Ga naar het startscherm van de toolbox en kies tussen biotische of abiotische data.

A screenshot of the ScheldeMonitor Data Download Toolbox. At the top, there is a navigation bar with the ScheldeMonitor logo, Home, Zoeken op thema, Tools, and Project. Below the navigation bar, the URL 'Toolbox Home > Selecteer een Data type' is visible. There are two main sections: 'Biotische Data (periodieke metingen)' which shows a close-up image of a shrimp-like crustacean, and 'Abiotische Data (periodieke metingen)' which shows a hand holding a white cylindrical device connected to a tube.

2. De toolbox biedt verschillende criteria aan om de databank van ScheldeMonitor te filteren. Deze criteria verschillen voor biotische en abiotische data. In het explore tabblad kan je data selecteren gebaseerd op de databron, geografisch gebied, of tijdsperiode. Meer details over de inhoud van de dataset is te verkrijgen via de 'more info' knop (informatie teken) rechts van de dataset naam.

The screenshot shows a search interface for datasets. On the left, there's a search bar with fields for 'free text', 'datasource' (set to 'Any'), and 'data origin' (set to 'Any'). Below the search bar is a table titled 'Dataset Name' listing various monitoring datasets. On the right, a 'Selection Criteria' panel shows filters applied: 'Datasources' (Climatological monitoring network o...), 'Spatial' (Scheldemonding), and 'Temporal' (2016-01 to 2021-01). It also displays 'Approx. records: 10 000' and a green 'Next >' button.

Dataset Name	
AZ monitoring water quality of the Scheldt ⓘ	+/-
Benthos Westerschelde (MOVE) ⓘ	+/-
Chemical quality of dredged material in Zeeschelde (and Westerschelde) ⓘ	+/-
Climatological monitoring network of the Royal Netherlands Meteorological Institute (KNMI) ⓘ	+/-
Continuous measurements in Deurganckdok ⓘ	+/-
De Pauw: The environment and plankton of the WesterScheldt estuary ⓘ	+/-
ENDIS-RISKS fysico-chemical measurements and hyperbenthos in the Scheldt river, 2002-2006. ⓘ	+/-
Gillard: The micro-organisms of the Leie and the Scheldt in Ghent ⓘ	+/-
Maintenance of ports and channels: Western Scheldt ⓘ	+/-
Marine Information and Data Acquisition System: Underway & Cruise data ⓘ	+/-

1 to 10 of 32 records Previous 1 2 3 4 Next

Geographical area Time period

Het is niet verplicht om een databron, geografisch gebied of tijdsperiode te selecteren in het explore tabblad. In het volgende tabblad (beschikbaar via de groene 'next' knop) kan een specifiek taxon (biotische data) of parameter (abiotische parameter) geselecteerd worden. Datasets, parameters, of taxa kunnen toegevoegd worden aan uw selectie met het plus teken rechts van de dataset, parameter, of taxon.

Wanneer criteria geselecteerd zijn, toont de teller aan de rechterkant van het scherm het resterende aantal records dat voldoet aan de gekozen criteria.

The screenshot shows the 'Parameters Search' section of the toolbox. It includes fields for 'category' (set to 'Any') and 'parameter' (empty), a 'Search' button, and a table of parameters. The table has columns for 'Parameter Name' and actions. One row is highlighted with a blue border:

Parameter Name	
Elmaalgemiddelde windsnelheid (in 0.1 m/s)	+
Hoogste uurgemiddelde windsnelheid (in 0.1 m/s)	+
Laagste uurgemiddelde windsnelheid (in 0.1 m/s)	+
Vectorgemiddelde windrichting in graden	+
Vectorgemiddelde windsnelheid (in 0.1 m/s)	+

To the right, the 'Selection Criteria' sidebar shows 'Datasources' (Climatological monitoring netw...), 'Spatial' (Scheldemonding), 'Temporal' (2016-01 to 2021-01), and 'Parameters' (approx. 10 000 records). A green 'Next' button is visible at the bottom.

- Eenmaal alle gewenste criteria geselecteerd zijn, klik op de groene “next” knop om een samenvatting van uw data te zien in de toolbox.

The screenshot shows the 'Zoek op parameter' (Search by parameter) section. It includes fields for 'categorie' (set to 'Alles') and 'parameter' (empty), and a 'Zoek' (Search) button. Below is a table of parameters:

Parameter	
1,1,1,2-Tetrachloorethaan in mg/kg drooggewicht in sediment	+
1,1,1,2-Tetrachloorethaan in ug/l in oppervlaktewater	+
1,1,1-Trichloorethaan (Methylchloroform) in mg/kg drooggewicht in sediment	+

To the right, the 'Uw Selectie' (Your Selection) sidebar lists datasets: Databron, Geografisch, Temporeel, and Parameters (1 item: 1,1,1,2-Tetrachloorethaan in mg/kg...). A green 'Volgende' (Next) button is visible at the bottom.

- De toolbox toont een samenvatting van de gekozen datasets, samen met verschillende opties om de data te downloaden of te visualiseren. De volgende acties kunnen ondernomen worden in de toolbox:

- *Download Data*: een data bestand in csv-format zal gedownload worden. Meer gedetailleerde informatie is beschikbaar in het segment 'A: Data gebruiken van gedownloade databestanden'.
- *View on Map*: dit visualiseert de data in een dynamische kaart viewer.
- *Upload to MDA*: slaat uw specifieke daselectie op in de Marine Data Archive, zodat deze daselectie later hergebruikt kan worden.
- *Save selection*: slaat een JSON bestand op die uw specifieke selectie beschrijft.
- *Share*: maakt een URL link van uw selectie.
- *Webservice URL*: genereert een WFS URL (web feature service) dat kan gebruikt worden om automatisch data op te laden in het script of medium. Meer gedetailleerde informatie is beschikbaar in het segment 'B: Data gebruiken van generic webservices'.
- *Load selection*: laad in een eerder opgeslagen data selectie met behulp van een JSON bestand.

Data Download

Data Summary

1 to 8 of 10 records

Station Name	Parameter Name	Dataset Title	Record count
Vlaakte van de Raan	Hoogste uurgemiddelde windsnelheid (in 0.1 m/s)	Detail	1404
Cadzand	Laagste uurgemiddelde windsnelheid (in 0.1 m/s)	Detail	1461
Vlaakte van de Raan	Laagste uurgemiddelde windsnelheid (in 0.1 m/s)	Detail	1404
Cadzand	Hoogste uurgemiddelde windsnelheid (in 0.1 m/s)	Detail	1461
Cadzand	Vectorgemiddelde windrichting in graden	Detail	1461
Vlaakte van de Raan	Vectorgemiddelde windsnelheid (in 0.1 m/s)	Detail	1404
Vlaakte van de Raan	Vectorgemiddelde windrichting in graden	Detail	1404
Cadzand	Vectorgemiddelde windsnelheid (in 0.1 m/s)	Detail	1461

Actions

[Download Data](#) [Upload to MDA](#)
[View on map](#) [Save Selection](#)
[Share](#) [Get webservice url](#)
[New Selection](#)

Selection Criteria

- 1 Datasources
Climatological monitoring network
- 1 Spatial
Scheldemonitoring
- 1 Temporal
2016-01 to 2021-01
- Taxa
- Traits
- Data Precision
- Parameter Filters

< Previous

- De meeste, maar niet alle data in ScheldeMonitor is openbaar. Sommige data kan alleen gezien worden door gebruikers met geschikte inloggegevens. Voor deze datasets worden geen waarden weergegeven bij het downloaden van de data via zowel de databestanden en webservices. Daarom is de Toolbox voorzien van een 'Login' knop in de rechterbovenhoek. Deze knop zal gebruikers naar een login scherm leiden, waar inloggegevens kunnen ingevoerd of aangevraagd worden. Ga terug naar de toolbox na een succesvolle login. Nu zullen alle waarden zichtbaar zijn wanneer de dataset gedownload wordt.

E-mail

Password

Login

If you have any problems logging in, please contact us at account@vliz.be

[\[Register\]](#) [\[Login\]](#) [\[Lost password\]](#)

4.1 Data gebruiken van gedownloade databestanden

De gebruikers kunnen nu kiezen om de data te downloaden van de ScheldeMonitor toolbox als databestanden in een CSV bestand format. Om dit te doen en om deze te data te gebruiken in de RStudio omgeving kan de gebruiker volgende stappen uitvoeren:

1. De gebruiker selecteert de “Download Data” knop en vult alle nodige gegevens in om te beginnen met zijn/haar download:
 - Organization: selecteer het type organisatie waarvoor u werkt. Dit is niet verplicht.
 - Email: Vul uw email adres in waarnaar een melding kan gestuurd worden over de gereedheid van uw download.
 - Country: Selecteer het land waarin de download wordt uitgevoerd.
 - Data purpose: Selecteer voor welk doel de download is.

organization:

email: *

Get a notification by email when the download is ready.

country:

data purpose:

I agree to ScheldeMonitor keeping my personal data to inform me about the ScheldeMonitor project and its products. (optional)

Annuleren **Download voorbereiden**

2. Nadat de nodige informatie werd ingediend, zal uw data voorbereid worden voor de download. De status van de voorbereiding kan gevolgd worden in de rechterbovenhoek van het scherm. Nadat de data voorbereid is, zal een knop verschijnen waardoor de download kan beginnen. Voor grotere databestanden kan een mail verzonden worden naar het opgegeven adres om de gebruiker te melden dat de download volledig voorbereid is.

The screenshot shows the 'Data Download' interface. On the left, the 'Data Summary' section displays a table of 8 records from 1 to 10. The table columns are 'Station Name', 'Parameter Name', 'Dataset Title', and 'Record count'. The right side features a 'Download Queue' section with a red circle around it, showing an ID and timestamp. Below this is the 'Selection Criteria' section with various filter options like Datasources, Spatial, Temporal, Taxa, Traits, Data Precision, and Parameter Filters. At the bottom are action buttons: 'Download Data' (highlighted in blue), 'Upload to MDA', 'Save Selection', 'Share', 'Get webservice url', and 'New Selection'.

Station Name	Parameter Name	Dataset Title	Record count
Cadzand	Elmaalgemiddelde windsnelheid (in 0.1 m/s)	Detail	30
Cadzand	Hoogste uurgemiddelde windsnelheid (in 0.1 m/s)	Detail	30
Cadzand	Laagste uurgemiddelde windsnelheid (in 0.1 m/s)	Detail	30
Cadzand	Vectorgemiddelde windrichting in graden	Detail	30
Cadzand	Vectorgemiddelde windsnelheid (in 0.1 m/s)	Detail	30
Vlaakte van de Raan	Elmaalgemiddelde windsnelheid (in 0.1 m/s)	Detail	29
Vlaakte van de Raan	Hoogste uurgemiddelde windsnelheid (in 0.1 m/s)	Detail	29
Vlaakte van de Raan	Laagste uurgemiddelde windsnelheid (in 0.1 m/s)	Detail	29

3. Eenmaal het databestand opgeslagen is op de lokale drive kan de gebruiker het bestand laden in de RStudio omgeving en beginnen werken met deze data. Dit kan uitgevoerd worden door gebruik te maken van het basis package van R, door de volgende functies te runnen:

```
data = read.csv("path/file.csv", stringsAsFactors = FALSE)
```

Bijvoorbeeld:

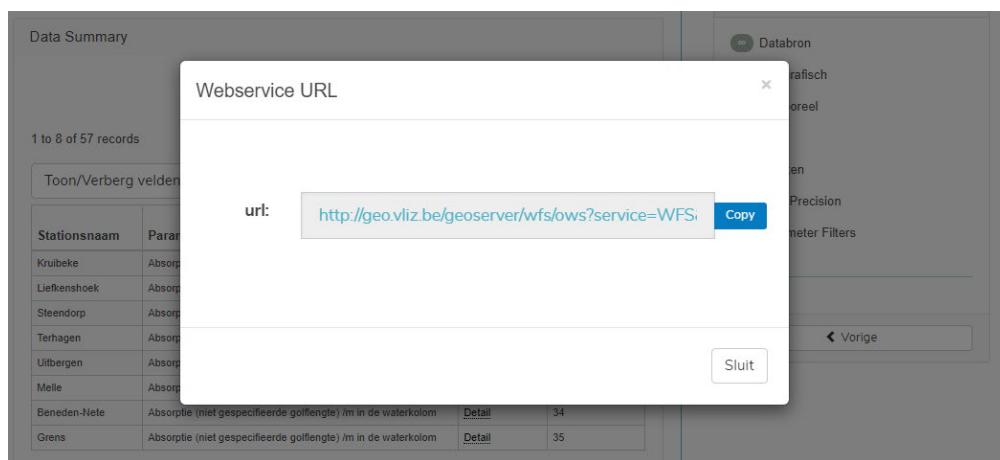
```
data_waterstand = read.csv("Data/hoogwater_combinded.csv", stringsAsFactors = FALSE)
```

4. De databestanden kunnen enkel gedownload worden als CSV bestanden. Dit format heeft echter een limiet van 1.000.000 records. Grottere bestanden zullen records verliezen wanneer de gebruiker het bestand wil openen in MS Excel voordat ze geladen zijn in R. Gebruikers worden daarom aanbevolen om deze grottere bestanden te openen als een TXT bestand in programma's zoals Notepad++.

4.2 Data gebruiken van generic webservices

Gebruikers van de RStudio omgeving van ScheldeMonitor worden echter aangespoord om gebruik te maken van de generic webservices die beschikbaar zijn in de data download toolbox van ScheldeMonitor. Deze webservices zijn een URL format dat de ScheldeMonitor database automatisch doorzoekt zonder menselijke tussenkomst. De samenstelling van deze URL is automatisch gegenereerd, gebaseerd op de selectie gemaakt door de gebruiker in het criteria van de data download toolbox. Het gebruik van webservices heeft het extra voordeel dat er geen databestanden nodig zijn om de dataset te laden in R, en dat de meest recente versie van de database opgezocht wordt. Het laatste betekent dat wanneer nieuwe data wordt toegevoegd in de database van een alreeds gedownloade dataset, dezelfde webservice URL zal in staat zijn om de nieuw toegevoegd data automatisch in te laden. Om de webservices in RStudio te gebruiken:

1. Selecteer de “Webservice URL” optie in de data download toolbox, die u de URL zal geven die nodig is om de geselecteerde dataset te verkrijgen.



2. Eenmaal u de volledige URL heeft gekopieerd, kunt u de data inladen in de RStudio omgeving. U kunt hiervoor een functie uit de 'sf' R-library en de volgende coderegels gebruiken:

```
install.packages("sf")
library(sf)
WebService <- "URL"
data <- data.frame(st_read(WebService))
```

3. Afhankelijk van de grote van de aangevraagde dataset kan het laden van de data in R tot 1 minuut duren. Desalniettemin zal de dataset beschikbaar zijn in de RStudio omgeving. De limiet van de webservice is beperkt tot ongeveer 1.000.000 records per aanvraag. Het is daarom aangeraden dat gebruikers meerdere afzonderlijke URL's genereren in de toolbox als ze meer dan een miljoen records willen analyseren en de data in R zelf samenvoegen.
4. VLIZ heeft een script gemaakt om data voor een bepaalde tijdsperiode (een jaar) voor een bepaalde parameter te lezen. Dit script kan hieronder gevonden worden of op de ScheldeMonitor GitHub pagina. Hoe u toegang kunt krijgen tot de ScheldeMonitor GitHub organisatie wordt uitgelegd in een speciale handleiding over het gebruik van GitHub, beschikbaar op de website.

```
#Created by Jelle Rondelez (VLIZ) on 8/3/21
```

```
#These packages are needed
```

```

install.packages(sf)
install.packages(stringr)
install.packages(dplyr)
library(dplyr)
library(sf)
library(stringr)

dataset <- data.frame()

#Here, the years for which you want to download data should be listed.
#Using the for loop, data can be downloaded per year by default
years <- c("2016", "2017", "2018", "2019", "2020", "2021")

#This is a test string. Replace it with your own string
#The timespan of the original wfs string should run from 1 Jan tot 31 Dec
#no matter which years are selected
wfsstring <- "http://geo.vliz.be/geoserver/wfs/ows?service=WFS&version=1.1.0&request=GetFeature&typeName"

#This for loop results in the download of yearly datasets.
#At the end of the loop, all datasets are both saved separately and appended together.
#If the user wants to download in larger timespans, change the second 'years[i]'
#example: 'years[i+1]' downloads data for two year spans

for (i in 1:length(years)) {
  wfsstring <- str_replace(wfsstring, "(?<=%27).*(?=-12-31)",
                            paste(years[i], "-01-01%27+AND+%27", years[i], sep=""))
  name <- paste(years[i])
  data <- data.frame(st_read(wfsstring))
  dataset <- rbind(data, dataset)
  assign(name, data)
}

```

5 Helpdesk

VLIZ is verantwoordelijk om de RStudio van ScheldeMonitor up-to-date en operationeel te houden. Naast het voorzien van de nodige server en geheugen capaciteit, zal VLIZ dus ook zorgen dat al de nodige R libraries en packages geïnstalleerd zijn op de RStudio server. Indien nieuwe libraries en packages moeten geïnstalleerd worden, kunnen gebruikers VLIZ contacteren om dit te doen.

Om aan deze en andere noden van gebruikers en medewerkers te voldoen, heeft VLIZ een permanente helpdesk. Deze helpdesk kan gecontacteerd worden via het algemene adres van de ScheldeMonitor:

Helpdesk ScheldeMonitor

Data Centre - Local Services & Projects



Vlaams Instituut voor de Zee vzw T +32(0)59340172
Flanders Marine Institute info@scheldemonitor.org
 InnovOcean site, Wandelaarkaai 7 www.vliz.be
 8400 Oostende, Belgium

Voor dringende zaken of vragen, of indien gebruikers en medewerkers het gebruik van de RStudio omgeving willen bespreken voor bepaalde projecten kan de project manager van ScheldeMonitor gecontacteerd worden:

Jelle Rondelez

Project Manager

Data Centre - Local Services & Projects



Vlaams Instituut voor de Zee vzw M +32(0)473510828
Flanders Marine Institute jelle.rondelez@vliz.be
InnovOcean site, Wandelaarkaai 7
8400 Oostende, Belgium www.vliz.be