**Data Stewardship: Efficiënt en Gestructureerd Beheren**

**Waarom is dit een belangrijk onderdeel van onderzoek doen?**

Een goede manier van data beheren is essentieel in onderzoek om **reproduceerbaarheid, transparantie en betrouwbaarheid** te waarborgen. Als gegevens goed zijn georganiseerd en beheerd, kunnen anderen het onderzoek begrijpen, controleren en herhalen. Daarnaast helpt goed beheer om fouten te voorkomen en maakt het het werk efficiënter.

**Wat is een duidelijke mappenstructuur en waarom houd je dat aan?**

Een duidelijke mappenstructuur zorgt ervoor dat bestanden logisch en overzichtelijk zijn opgeslagen. In dit project werk ik bijvoorbeeld met aparte mappen voor:

* **Data\_raw**: onbewerkte data
* **Data\_processed**: verwerkte data
* **Scripts**: gebruikte codes
* **Output**: grafieken, resultaten en tabellen

Door deze structuur te volgen, is het direct duidelijk waar informatie te vinden is, wat de kans op verwarring of verlies van data minimaliseert.

**Wat is een handige manier voor de naamgeving van bestanden en waarom?**

Bestandsnamen moeten **duidelijk, kort en gestructureerd** zijn. Bijvoorbeeld: count\_matrix\_day3.csv of volcano\_plot\_wc3.png.

* Dit voorkomt dubbele bestanden en verwarring.
* Het maakt snel duidelijk wat de inhoud is.
* Door consistente naamgeving (bijvoorbeeld altijd project\_onderdeel\_datum) blijft alles logisch en makkelijk terug te vinden.

**Wat houdt versiebeheer in en waarom doe je dit?**

Versiebeheer betekent dat je bijhoudt welke wijzigingen zijn gemaakt in bestanden en scripts, bijvoorbeeld met behulp van **Git en GitHub**. Dit is belangrijk omdat:

* Je kunt altijd terug naar een vorige versie.
* Je kunt goed samenwerken met anderen zonder bestanden te overschrijven.
* Het maakt de ontwikkeling van je project inzichtelijk voor anderen.

**Hoe zorg je voor goede documentatie in scripts?**

Goede documentatie betekent dat je in je scripts voldoende **commentaarregels** plaatst waarin je uitlegt:

* Wat je doet (per stap)
* Waarom je het doet
* Welke parameters of bestanden je gebruikt

Dit maakt het script begrijpelijk en bruikbaar voor anderen, en helpt jezelf om later je eigen werk terug te begrijpen.

**Wat voor platforms kan je gebruiken om scripts te delen?**

* **GitHub**: meest gebruikt voor het delen en beheren van scripts, met goede versiebeheerfunctionaliteiten.
* **GitLab**: vergelijkbaar met GitHub, soms gebruikt door bedrijven.
* **Zenodo**: kan gekoppeld worden aan GitHub om datasets en scripts officieel te publiceren.
* **OSF (Open Science Framework)**: platform voor het beheren van volledige onderzoeksprojecten.

**Welke tools gebruik je voor data-analyse, zoals bijvoorbeeld voor het veilig opslaan van data?**

* **R en RStudio**: voor data-analyse en statistiek.
* **GitHub**: voor versiebeheer en scriptopslag.
* **OneDrive / SharePoint**: voor het veilig opslaan van data met back-ups.
* **CSV-bestanden**: voor het opslaan van de tellingen en resultaten.
* Eventueel gebruik van **gecodeerde opslag** of **versleutelde drives** voor gevoelige data.

**Hoe zou je omgaan met gevoelige of persoonlijke data?**

* Gegevens anonimiseren waar mogelijk.
* Opslaan op beveiligde locaties (zoals OneDrive van school of een beveiligde server).
* Nooit gevoelige gegevens op openbare platforms zoals GitHub zetten.
* Toegang beperken tot alleen de betrokken onderzoekers.
* Controleren of het onderzoek voldoet aan privacywetgeving (AVG).

**Wat is het belang van open data en het publiceren van datasets?**

* **Open data** maakt onderzoek transparant en controleerbaar.
* Het versnelt wetenschappelijke vooruitgang doordat anderen de data kunnen hergebruiken.
* Het voorkomt dat onderzoek dubbel wordt gedaan.
* Het verhoogt de betrouwbaarheid en de impact van je werk.