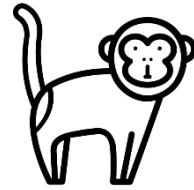


Oefening zoeken naar data: biodiversiteit in de Amazone met behulp van DNA

Links:

Re3data.org	https://www.re3data.org/
Elsevier DataSearch	https://datasearch.elsevier.com/#/
DataCite	https://search.datacite.org/
WorldWideScience	https://worldwidescience.org/
Google Data Set Search	https://toolbox.google.com/datasetsearch
ScholExplorer	http://scholexplorer.openaire.eu/



Icon made by Freepik
from Flaticon

Je doet onderzoek naar biodiversiteit in het Amazone gebied en bent op zoek naar onderzoeksdata waarbij gebruik is gemaakt van DNA technieken om de biodiversiteit te bepalen.

1. Kijk in Re3data.org of je een geschikte repository kunt vinden op dit onderwerp. Als je een geschikte repository hebt gevonden, noteer deze dan.
[Zoeken naar repositories op gebied van "biodiversity" levert zeer veel resultaten \(173\) op.](#)
[Combineren met "DNA" resulteert in 12 resultaten, waar mogelijk wel een relevante repository tussen zit.](#) [Combinatie met Amazonia levert één Braziliaans repository op.](#)
2. Je bent vervolgens eerst aan de slag gegaan met het zoeken naar relevante wetenschappelijke literatuur en hebt een interessant artikel gevonden: Ritter CD, Zizka A, Barnes C, Nilsson RH, Roger F, Antonelli A (2019) Locality or habitat? Exploring predictors of biodiversity in Amazonia. *Ecography* 42(2): 321-333. <https://doi.org/10.1111/ecog.03833>

Zoek in het artikel de verwijzing naar de dataset. In welke repository is de dataset opgeslagen?
[In de materiaal en methoden sectie staat een verwijzing naar de data depositie.](#)
[De dataset is opgeslagen bij Dryad.](#)

3. Open de pagina van de dataset bij Dryad en bekijk uit welke onderdelen de dataset bestaat. Je hoeft de dataset niet te downloaden.
[De dataset bestaat uit 5 bestanden, 4 bestanden met DNA data en 1 met R scripts.](#)

Open de volledige metadata (onderaan de pagina) en bekijk welke metadata-velden beschikbaar zijn.

[In de metadata kun je vinden wat de geolocatie is, wie de auteurs zijn, waar het bijbehorende artikel te vinden is etc.](#)

4. Ga nu na of je deze dataset ook kunt vinden met de beschikbare data-zoekmachines (Elsevier Data Search, DataCite, WorldWideScience, Google Data Search). Probeer eerst met enkele termen (biodiversity Amazonia DNA) de dataset te vinden. Als het niet lukt kun je de exacte titel gebruiken van het artikel.

a. Elsevier Data Search:

[Je vindt de dataset als je zoekt met de drie termen.](#)

- Bekijk bestandssoorten in de dataset zitten
[Bij Elsevier Data Search zie je in het zoekresultaat direct wat voor type data er in de dataset zit \(Software/Code en Tabular Data \(4\)\).](#)
- Bekijk wat je nog meer aan resultaten vindt; bekijk hierbij ook de filters.
[Er worden veel verschillende soorten bestandstypen gevonden. De meeste resultaten worden gevonden in de repository Zenodo, daarbij gaat het vaak om wetenschappelijke artikelen die in Zenodo gedeponeerd zijn.](#)

b. DataCite:

- Zoek eerst met de drie termen
Zoeken met de drie termen resulteert in een gigantisch aantal resultaten (> 1 mln.), die ook niet echt relevant lijken.
- Gebruik daarna de titel van het artikel (en gebruik eventueel quotes om te zoeken naar te exacte titel).
De quotes zijn nodig om de dataset te kunnen vinden, Zoeken met de woorden uit de titel blijven er te veel en niet relevante resultaten over. Er worden verder weinig metadata getoond in het record. Er wordt verwezen naar de dataset bij Dryad.

c. WorldWideScience:

- Naar welke repository wordt je nu doorverwezen?
Zoek met de drie termen en ga naar het tab 'Datasets'. De gezochte set staat op plaats 4. Je wordt nu echter door verwezen naar Narcis. DANS-EASY fungeert als back up voor Dryad en de metadata worden daarom ook opgenomen in DANS-EASY en via Narcis ontsloten (<https://dans.knaw.nl/nl/actueel/nieuws/dryad-en-dans-kondigen-partnerschap-voor-digitale-duurzaamheid-aan>).

d. Google Data Search

Zoeken met de drie termen geeft de juiste dataset als eerste zoekresultaat.

- Klopt de bewering van Google Data Search over het aantal artikelen dat de dataset citeert?
Volgens Google Scholar wordt de dataset door 15 artikelen geciteerd, maar dat blijken citaties naar het artikel te zijn. Google Scholar is niet in staat om goed onderscheid te maken tussen artikel en bijbehorende dataset. Dit komt waarschijnlijk omdat de dataset grotendeels dezelfde titel heeft als het artikel.

5. Ga nu naar de repository Figshare.com. Zoek daar met behulp van de titel van het artikel naar een dataset. Wat vindt je?

De dataset blijkt ook hier in te zitten. Er wordt wel weer verwezen naar de bron bij Dryad.

Wij hebben deze opdrachten zorgvuldig opgesteld, maar door de dynamiek van de zoekmachines en databases kunnen de antwoorden afwijken van de zoekresultaten na afloop van de workshop.

Oefening zoeken naar data: monitoring van intensive care patiënten

Links:

Re3data.org	https://www.re3data.org/
Elsevier DataSearch	https://datasearch.elsevier.com/#/
DataCite	https://search.datacite.org/
WorldWideScience	https://worldwidescience.org/
Google DataSet Search	https://toolbox.google.com/datasetsearch
ScholExplorer	http://scholexplorer.openaire.eu/



Icon made by [FreePik](#)
from Flaticon

Je bent op zoek naar een dataset met data afkomstig van intensive care patiënten zoals vitale kenmerken, medicatie, laboratorium analyses, observaties en notities van medisch personeel.

1. Kijk in Re3data.org of je een geschikte repository kunt vinden op dit onderwerp. Als je een geschikte repository hebt gevonden, noteer deze dan. Tip zoek naar fysiologische data. Noteer een van de gevonden repositories.

Bijv. [PhysioNet](#), [PhysioBank](#), [InCHIANTI](#)

Open een van de repositories en kijk wat je dan tegenkomt.

Als je PhysioNet nog niet had gevonden, zoek hem op via Re3data.org. Ga vervolgens naar dit repository. Zoek in dit repository naar "intensive care". Je vindt nu o.a. de MIMIC III database. Open de link naar de database. Waar kom je nu terecht? En wat kun je hier vinden?

<https://mimic.physionet.org/>

Dit is een website over de MIMIC III database. Hier kun je alle informatie vinden over deze database, je vindt de data hier en ook de scripts waarmee je data kunt analyseren.

2. Op deze site staat een verwijzing naar een data paper in het tijdschrift Scientific Data. In dit data paper wordt de MIMIC III database uitvoerig beschreven: Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. - H., Feng, M., Ghassemi, M., . . . Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3 doi: [10.1038/sdata.2016.35](https://doi.org/10.1038/sdata.2016.35).

Ga naar het data paper. In het paper wordt verwezen waar de data te vinden zijn (Data Citation). Open de link naar de data. Waar kom je nu terecht?

<https://physionet.org/physiobank/database/mimic3cdb/>

De doi verwijst naar een oude versie van de pagina die je eerder hebt gevonden via PhysioNet.

Waar staan de scripts op de data te analyseren?

De code is te vinden via GitHub. (<https://github.com/MIT-LCP/mimic-website>). Dit is een platform waar software code kan worden gedeeld, de versies worden beheerd en anderen kunnen voortbouwen op bestaande code.

3. Om toegang te krijgen tot deze database niet je eerst toestemming vragen aan de eigenaar. We bekijken de dataset dus verder niet voor de opdracht, maar proberen hem wel via verschillende zoekmachines te vinden.

Probeer eerst met enkele termen ("critical care" patients "vital sign") de dataset te vinden. Als het niet lukt kun je de exacte titel gebruiken van de dataset (MIMIC-III Clinical Database). Beschrijf de opmerkelijk zaken per zoekmachine.

a. Elsevier Data Search

De dataset is niet vindbaar via Elsevier Data Search. Dit komt omdat DataCite niet als bron gebruikt wordt door Elsevier Data Search.

b. DataCite

De dataset is te vinden door quotes te gebruiken bij de exacte titel. Anders blijven er teveel records over, ook bij het gebruik van filters. Er is nagenoeg geen metadata aanwezig in het record, wat het vinden van de MIMIC III database moeilijk maakt.

c. WorldWideScience

In WWS is de dataset alleen te vinden via de titel. Er is verder geen metadata aanwezig, waardoor de dataset dus niet op andere termen gevonden kan worden.

d. Google Data Search

In Google Data Search moet je ook zoeken met de titel van de dataset.

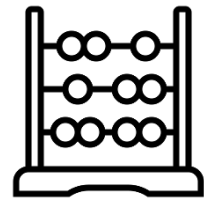
- Hoe vaak wordt deze dataset geciteerd volgens Google Scholar?
19 citaties
- Bekijk de citaties in Google Scholar. Op welke manier wordt in het artikel "Machine learning landscapes and predictions for patient outcomes" naar de dataset wordt verwezen.?
In dit artikel wordt verwezen naar het data paper in Scientific Data.

Wij hebben deze opdrachten zorgvuldig opgesteld, maar door de dynamiek van de zoekmachines en databases kunnen de antwoorden afwijken van de zoekresultaten na afloop van de workshop.

Oefening zoeken naar data: wiskundeonderwijs in de 21ste eeuw

Links:

Re3data.org	https://www.re3data.org/
Elsevier DataSearch	https://datasearch.elsevier.com/#/
DataCite	https://search.datacite.org/
WorldWideScience	https://worldwidescience.org/
Google DataSet Search	https://toolbox.google.com/datasetsearch
ScholExplorer	http://scholexplorer.openaire.eu/



Icons made by Eucalyp
from Flaticon

Je hebt gehoord van een internationaal onderzoek naar de opleiding van wiskundeleraren in de 21^{ste} eeuw, door onderzoeker William Schmidt. Schmidt vroeg toekomstige wiskundeleraren in een questionnaire naar hun ideeën over pedagogie. Je wilt Schmidts data onderliggend aan dit onderzoek vinden, en eventuele artikelen die op deze data hebben voortgebouwd.

1. Probeer de dataset te vinden via Elsevier DataSearch, DataCite, WorldWideScience en Google DataSet Search, door in alle vier de zoekmachines te zoeken naar: "mathematics teaching" "21st century". (Tip: de dataset staat in de repository ICPSR.)

Noteer voor elke zoekmachine of je de data kunt vinden. Schrijf per zoekmachine ook op wat je verder opvalt aan de zoekopties en/of de resultaten.

<u>Elsevier DataSearch</u>	De dataset staat bovenaan, wel alleen maar metadata terwijl Elsevier DS meestal previews geeft.
<u>DataCite</u>	De dataset staat nergens bovenaan.
<u>WorldWideScience</u>	Geeft twee resultaten met dezelfde titel, V1 en Archival version, maar die gaan naar dezelfde dataset van Schmidt.
<u>Google Dataset Search</u>	De dataset als enige resultaat.

2. Je wilt weten of er ook artikelen zijn gepubliceerd die deze dataset citeren. Probeer dit te achterhalen via ICPSR en vanaf ScholExplorer. In ICPSR doe je dit vanaf de pagina die de dataset omschrijft (<https://doi.org/10.3886/ICPSR34430.v1>), en in Scholexplorer door eerst naar de dataset te zoeken. Wat vind je in de twee systemen?

In ICPSR vind je onder het tabje Data-related publications' 37 bronnen. Maar de meeste zijn van voordat de dataset en zelfs het eerste rapport gepubliceerd waren, en de dataset citeert deze niet.

ScholExplorer geeft twee keer links naar hetzelfde artikel maar andere ingang (Crossref DOI en OpenAire identifier). Artikel is van voor de dataset (2011); citeert niet de dataset maar wel hun eerste rapport (Schmidt et al 2007).

Dus: beide geven gerelateerde literatuur maar niet per se directe links.

3. Ga weer terug naar de dataset in ICPSR. Probeer nu meer datasets in ICPSR te vinden die over lerarenonderwijs gaan. Lukt dat? Zo ja, hoe heb je gezocht?

Naar subject terms; klikken op 'teacher education'. Of vanaf de homepage 'browse by keyword' of vrij zoeken. Je kunt ook zoeken binnen de variables op bijv. 'teacher education', dan vind je surveys die dit dekken.

Wij hebben deze opdrachten zorgvuldig opgesteld, maar door de dynamiek van de zoekmachines en databases kunnen de antwoorden afwijken van de zoekresultaten na afloop van de workshop.

Oefening zoeken naar data: woordenschat meertaligen in Indonesië

Links:

Re3data.org	https://www.re3data.org/
Elsevier DataSearch	https://datasearch.elsevier.com/#/
DataCite	https://search.datacite.org/
WorldWideScience	https://worldwidescience.org/
Google DataSet Search	https://toolbox.google.com/datasetsearch
ScholExplorer	http://scholexplorer.openaire.eu/



Icon made by Freepik
from Flaticon

Je hebt online een interessant artikel gevonden over de woordenschat van meertaligen: 'Contextualized bilingualism among adolescents from four different ethnic groups in Indonesia', door Sari et al., 2018 (<https://doi.org/10.1177/1367006918803678>). Het onderzoek vergelijkt de woordenschat van vier etnische groepen in Indonesië. Je wilt dit onderzoek graag repliceren, dus wilt zoveel mogelijk van de onderliggende data vinden.

1. Bekijk de pagina waarop het artikel staat. Welke verwijzing(en) naar data kan je hier vinden?

- in menu links 'Supplemental material' --> naar Figshare --> Appendix met grote tabellen
- in menu links 'Figures and Tables' --> toont de tabellen uit het artikel
- in artikel: diezelfde tabellen, maar ook datacollectie/analyse methoden

2. Open nu de PDF van het artikel. Welke verwijzing(en) naar data vind je hier?

In de tekst wordt gelinkt naar 'online appendix' (gaat naar Suppl. material sectie op publisher website)

Verder staat er achterin de tekst alleen dit: Supplemental material for this article is available online.

Geeft aan dat artikel niet altijd een duidelijke referentie geeft.

3. De data die je vanaf het artikel vindt staan in de repository Figshare. Maar er zijn meer data uit dit onderzoek beschikbaar. Die staan in de repository DataverseNL: <https://hdl.handle.net/10411/S6KQQV>.

Probeer de Figshare dataset én de Dataverse dataset te vinden in Elsevier DataSearch, DataCite, WorldWideScience en Google DataSet Search, door in alle vier de zoekmachines te zoeken op de artikel-titel: "Contextualized bilingualism among adolescents from four different ethnic groups in Indonesia".

Noteer voor elke zoekmachine of je de data kunt vinden. Schrijf per zoekmachine ook op wat je verder opvalt aan de zoekopties en/of de resultaten.

<u>Elsevier DataSearch</u>	Nee, 0 resultaten. Zonder "" meer, maar niet de DataverseNL set.
<u>DataCite</u>	Vier resultaten; eigenlijk 2 maar beide dubbel (DOIs zonder en met .v1, maar als je doorklikt gaan beide naar dezelfde) - de Figshare suppl.materials - de Figshare Collection van dit artikel (met daarin de suppl.materials) Dus geen link naar de DataverseNL set.
<u>WorldWideScience</u>	Hier moet je van de 'Papers' naar de 'Data/Software' tab wisselen, anders vind je het artikel. Onder 'Data/Software' vindt je alleen links naar de Collectie in Figshare voor dit artikel (waaronder alleen de suppl.material

	valt), en naar diezelfde suppl.material in Figshare. Niet naar de DataverseNL set.
<u>Google Dataset Search</u>	Nee, 0 resultaten, ook zonder "" (Maar soms krijg je 'm wel naar boven! Google resultaten wisselen steeds)

Wij hebben deze opdrachten zorgvuldig opgesteld, maar door de dynamiek van de zoekmachines en databases kunnen de antwoorden afwijken van de zoekresultaten na afloop van de workshop.