Quantifying and understanding the human impacts on catchment sediment yield at the global scale

A Data Management Plan created using DMPonline.be

Creator: Florence Tan

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Florence Tan

Grant number / URL: https://www.fwo.be/en/fellowships-funding/phd-fellowships/phd-fellowship-fundamental-research/

ID: 200952

Start date: 01-01-2022

End date: 31-10-2026

Project abstract:

Sediment fluxes are a key component of the Earth system but are also associated with a wide range of environmental and geomorphic challenges. Human activities have a rapidly growing impact on these sediment fluxes (e.g. through land use/cover changes, reservoirs and mining). These impacts are complex, scale-dependent and often counteracting. They thus remain poorly understood, especially at the global scale. Here, we aim to tackle this research gap. For this, we will first construct the hitherto largest database of contemporary catchment yield (SY) observations and develop a framework to assess the uncertainties on these measurements. Next, we will develop a model that can simulate the 'natural' SY of a catchment, i.e. the SY that could be expected under the current climatic, geomorphic and tectonic conditions but without human intervention (e.g. natural land use, no reservoirs). For this, we will draw from recent advancements in landscape evolution and SY modelling. We will calibrate and validate the model using well-selected SY observations from (nearly) natural catchments and measured long-term erosion rates. In a final phase, we will apply the model to thousands of catchments with contemporary SY data available. This will allow us to obtain unprecedented insights into the degree of human impacts on current SY values at the global scale; to identify the most sensitive and impacted river systems in the world; and to better understand the factors controlling these impacts.

Last modified: 06-07-2023

Quantifying and understanding the human impacts on catchment sediment yield at the global scale
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, ... • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Global contemporary sediment yield (SY) database	Compiled SY observations and their relevant metadata	Reuse existing data	Digital	Compiled/aggregated data	.csv	<100MB	
GIS files of SY-associated catchments	Vector files of delineated catchment boundaries and relevant catchment properties	Generate new data	Digital	Simulation data	.shp	<100GB	
GIS files of relevant environmental variables	Raster files of global hydro-environmental variables	Reuse existing data	Digital	Simulation data	.geotiff	<5TB	
GIS files of model output	Raster files of final model runs on natural SY and the estimated degrees of human impacts worldwide	Generate new data	Digital	Simulation data	.geotiff	<1TB	
Code and programs for analysis	Python scripts to run and validate our model, assess uncertainties on SY data, etc.	Generate new data	Digital	Software	.text	<100MB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Our database of SY observations will likely include a few dozens existing individual datasets. Below is a list of some the existing datasets we have compiled so far.
- Vanmaercke et al. 2014 (DOI 10.1016/j.earscirev.2014.06.004)
- Vanmaercke et al. 2011 (DOI 10.1016/j.geomorph.2011.03.010)
- Li et al. 2021 (DOI 10.1126/science.abi9649)
- Yang et al. 2021 (DOI 10.1002/rra.3896)
- Milliman and Farnsworth 2011 (DOI 10.1017/CBO9780511781247)
- Liu et al. 2022 (DOI 10.1029/2022WR032365)
- National Water Data Archive (https://www.canada.ca/en/environment-climate-change/services/water-overview/quantity/monitoring/survey/data-products-services/national-archive-hydat.html)
- United States Geological Survey On-Line Instantaneous Fluvial Sediment and Ancillary Data (https://water.usgs.gov/osw/sediment/index.html)

Source for the GIS files of global hydro-environmental variables:
- Linke et al. 2019 (DOI 10.1038/s41597-019-0300-6)

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

Metadata documents in the form of README.txt files and/or spreadsheets will be created for each dataset. Two levels of metadata will be included: 1) general information about the overall study (e.g. study title, study description, study design and methodology, keywords, time period, date of creation, spatial extent and resolution, projection, contact information, etc.) and 2) descriptions of individual variables and observations (including units of measurements).

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

The choice of a metadata standard or guidelines will ultimately be determined based on the data repository used. We expect, however, that the metadata will closely follow field-specific standards such as the Federal Geographic Data Committee’s Content Standard for Digital Geospatial Metadata, NASA’s Earth Science Data Systems’ ASCII File Format Guidelines for Earth Science Data, or the International Organization for Standardization’s (ISO) 19115-1:2014 Geographic Information – Metadata – Part 1: Fundamentals.

3. Data storage & back-up during the research project

Where will the data be stored?

The data will be stored on OneDrive (2TB capacity), SharePoint (5TB capacity) and on external hard drives kept at different locations (home of the PhD candidate + office of the promotor).

How will the data be backed up?

The data will be backed up on a bi-weekly basis using scripts developed for that purpose.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

Yes, there is currently sufficient storage and backup capacity during the project (2TB capacity on OneDrive + 5TB capacity on SharePoint + 2TB capacity on hard drive), as we expect to store around 3TB of relevant data in total.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Access to the original data and the backups will be strictly limited to the principal investigator (i.e., the PhD candidate) and the promotor. If, under specific circumstances, portions of the data need to be shared (e.g., to Bachelor's and Master's students), a new version of the data will be created for that purpose and will be password-protected.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The expected costs for data storage and backup during the research project is ~200 euros (price of external hard drives). These costs will be covered by the PhD candidate's FWO bench fee and/or the promotor's dedicated funding.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

All final output (and important mid-output) data will be retained for at least ten years.

Where will these data be archived (stored and curated for the long-term)?

The data will be preserved on the large volume storage services of KU Leuven (https://icts.kuleuven.be/sc/english/storage/large-volume-storage) and on external hard drives, kept at different physical locations.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

The expected costs for data preservation during the expected retention period is ~300 euros/year. The promotor of the research has dedicated funding to cover these costs.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository

Most of the data will be made available for reuse after (or during the later stages of the project):
- SY database: comma-delimited file of compiled SY observations and their relevant metadata
- GIS files: vector files of delineated catchment boundaries and relevant catchment properties; raster files of relevant environmental variables and final model output on natural SY and the estimated degrees of human impacts worldwide
- Code and programs: Python scripts developed to run our model, assess uncertainties on SY data, etc.

If access is restricted, please specify who will be able to access the data and under what conditions.

NA

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- Yes, Other

Some of the SY observations we will collect will have been shared with us through personal communication. Without explicit, written consent from the original authors of the data to publish their data as part of our open-access SY database (with proper credit), these observations will be removed from our published output.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

Our key results (e.g. SY database, models, final maps) will be made freely available via Open Access platforms (e.g. GitHub, Zenodo, CSDMS platform, KU Leuven RDR platform) and as appendices to our publications.

When will the data be made available?

The data will be made available during and after the final stages of the research project.

Which data usage licenses are you going to provide? If none, please explain why.

As we abide to an open-access philosophy, we plan to provide an Open Database License (ODbL) for all final output.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

NA

What are the expected costs for data sharing? How will these costs be covered?

The expected costs for data sharing is ~200 euros. These costs will be covered by the PhD candidate's FWO bench fee and/or the promotor's dedicated funding.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

During the research project, the principal investigator (i.e., the PhD candidate) will manage data documentation and metadata.

Who will manage data storage and backup during the research project?

During the research project, the principal investigator (i.e., the PhD candidate) will manage data storage and backup.

Who will manage data preservation and sharing?

After the research project, the promoter, in consultation with the principal investigator (i.e., the PhD candidate), will manage data preservation and sharing.

Who will update and implement this DMP?

The principal investigator (i.e., the PhD candidate), in consultation with the promoter, will update and implement this DMP.