FWO DMP Eva Boonaert

Project Name My plan (FWO DMP) - FWO DMP Eva Boonaert
Project Identifier 0000-0002-6526-5304
Grant Title 89916
Principal Investigator / Researcher Eva Boonaert
Project Data Contact eva.boonaert@kuleuven.be
Description My PhD aims to analyze the economic, social and environmental impact and trade-offs of voluntary sustainability standards. I will do this for multiple agricultural commodities (bananas, avocados, mangos, coffee, cocoa, quinoa and cassava) and multiple VSS over time. I will focus on the case of Peru due to the importance of agriculture for the countryâ€™s GDP and the livelihoods of their population. I will use secondary panel data for the economic and social impact from the national household survey (ENAHO) collected by the Instituto Nacional de EstadÃstica e InformÃitica and GIS data for the environmental impact from various online sources. This will be combined with certification data from the 32 standard-setting agencies that are active in Peru, which I will contact.
Institution KU Leuven

1. General Information

Name applicant
Eva Boonaert

FWO Project Number & Title
The sustainability impact of voluntary sustainability standards in the food sector: evidence from Peru (nr. 89916)

Affiliation

- KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Reuse existing data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Workpackage	Data origin	Data type	Format	Volume and storage	How created?	Source
WP2, WP3 and WP5	Secondary	Derived quantitative data from an original household survey collected by the Instituto Nacional de Estadística e Informática, Peru	Numerical: .dbf and .dta	10 GB, stored at the KU Leuven I- and L-drive (50 GB and 2 TB)	Computer analysis via STATA and R	http://iinei.inei.gob.pe/microdatos/ (free online available)
WP4 and WP5	Secondary	Derived quantitative data from original sattelite images	Numerical: .shp and .tif	200 GB, stored at the KU Leuven I- and L-drive (50 GB and 2 TB)	Computer analysis via QGIS	*Global Land Analysis & Discovery. Dataset GLAD. https://glad.umd.edu/dataset (2020). *IUCN. Resources International Union for Conservation of Nature. https://www.iucn.org/ (2020). *BirdLife. BirdLife Data . http://datazone.birdlife.org/home (2020).
WP2, WP3, WP4 and WP5	Secondary	Derived quantitative data from an original database of farmers taking part in certification	Numerical: .csv or .rtf.	5 GB, stored at the KU Leuven J-drive (50 GB) encrypted with Bitlocker	Computer analysis via MS EXCEL	I will contact the different certification organizations to request for the necessary certification data via email.

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

I will perform the analyses at district-level. Hence, no personal data is needed. I will do this by linking i) aggregated certification data obtained from the certification agencies at district-level with ii) districts in the anominized data from the national household survey between 2004-2019 from the Instituto Nacional de Estadística e Informática, Peru.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or

animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

- No

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- No

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- Yes

Yes, I will contact the certification agencies to acquire data on certified farmers in Peru at district-level. In particular, I will request for an anonymous list of certified producers in Peru for the period 2004-2019 for one of the following seven crops: banana, avocado, mango, cacao, coffee, quinoa and cassava. Of the producers, I would like to know the crop(s) that are certified, the certified area per certified crop, the year since when they have been certified, and the location of the producers (in the form of the district of the farm). The data will probably be obtained using a data sharing agreement set up with my host university.
In addition, the Instituto Nacional de Estadística e Informática allows users of the online available data to publish the data.

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

1. For the data processing of the raw online available national household survey from the Instituto Nacional de Estadística e Informática (INEI), Peru, a document with the definition of the variables is already provided by the INEI. I will process the household survey data in the scripted language software R and analyze it in the scripted langugage STATA to be able to track all changes. In addition, I will continuously update a ReadMe file to describe the contents of my data files. This will allow a secondary researcher to use the cleaned data accurately and effectively.
2. For the data processing of the raw sattelite images, I will store the original raw data files in the cloud and at an external storage place. In addition, I will continuously update a ReadMe file to describe all my processing files.
3. For the data tha I will acquire from the certification agencies, I will save the original files in the cloud and at an external storage place. Moreover, I will make a data dictionary and save the information on the data collection methodology in a ReadMe file.

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- Yes

Yes, I will use the Data Documentation Initiative standard for the ReadMe files.

5. Data storage and backup during the FWO project

Where will the data be stored?

During the research, the data will be stored at the university's central network drive (i.e. I-, J- and L-drive) (50 GB, 50 GB and 2 TB). The I- drive is my personal drive. The J-drive is a secured and encrypted drive in accordance with KU Leuven's IT Standards Framework which will be used for the confidential data. The L-drive is a personal drive for large datasets. This will be enough for the estimated 10 GB of household data, 200 GB of satellite data and 5 GB of certification data. After the research, the data will be stored on the university's central network drive (i.e. K-drive) in which I will add a secured map with strict access.

How is backup of the data provided?

The data will be stored on the university's central servers with automatic daily back-up procedures.

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, the available storage capacity for the university's central network drive is 50GB for the personal I- and J-drive and 2 TB for the personal L-drive (extra storage for large datasets)

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

For most data repositories there are no extra costs. Whenever needed (e.g for additional external hard drives), the FWO bench fee will be used to cover the costs incurred.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The data will be stored in the university's secure environment for private data with only controlled access possible and are secured by the ICTS service of KU Leuven with a mirror copy.

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

Yes, the data will be retained for a minimum period of 10 years according to te KU Leuven Research Data Management Policy.

Where will the data be archived (= stored for the longer term)?

The data will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven Research Data Management Policy.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

There are no expected costs related to the storage of the foreseen size of the data.

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- Yes. Specify:

Yes, I will contact the certification agencies to acquire data on certified farmers in Peru. However, to protect the privacy of the farmers, we will ask for aggregated data at district-level, probably using a data sharing agreement with my host university.

Which data will be made available after the end of the project?

I will share all my data files and metadocumentation (i.e. ReadMe files), possibly at a higher spatial scale, depending on the arrangements that will be made with the certification bodies.

Where/how will the data be made available for reuse?

- In an Open Access repository

I aim to publish in open access journals to enhance the visibility of the research.

When will the data be made available?

- Immediately after the end of the project

The data will be made available in the form of supplementary material of the Open Access articles.

Who will be able to access the data and under what conditions?

The full data of unpublished articles will be available upon request after the end of the project, provided that they give appropriate credit to the creators.

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

For most data repositories there are no extra costs. Whenever needed (e.g for additional external hard drives), the FWO bench fee will be used to cover the costs incurred.

8. Responsibilities

Who will be responsible for data documentation & metadata?

I myself will be responsible for the data collection, documentation and metadata.

Who will be responsible for data storage & back up during the project?

I myself will be responsible to store the data on the appropriate accommodation provided by KU Leuven. The ICTS service of KU Leuven is responsible for the back-up of the network drives at KU Leuven.

Who will be responsible for ensuring data preservation and reuse ?

I myself will be responsible for ensuring data preservation and reuse.

Who bears the end responsibility for updating & implementing this DMP?

I myself bear the end responsibility of updating & implementing this DMP.