1. ABOUT THE DATASET

------------

Title: Dataset supporting: “A systematic map of cassava farming practices and their agricultural and environmental impacts using new ontologies: Agri-ontologies 1.0”

Creator(s): Amelia S. C. Hood1,2, Alec P. Christie1, William J Sutherland1, Gorm E. Shackelford1

ORCID IDs

Amelia S. C. Hood: <https://orcid.org/0000-0003-3803-0603>

Alec P Christie: <https://orcid.org/0000-0002-8465-8410>

William J Sutherland: <https://orcid.org/0000-0002-6498-0437>

Gorm E Shackelford: <https://orcid.org/0000-0003-0949-0934>

Organisation(s):

1 Department of Zoology, University of Cambridge, Cambridge, United Kingdom

2 Centre for Agri-environmental Research, School of Agriculture, Policy and Development, University of Reading, Reading, United Kingdom

Rights-holder(s): University of Reading, University of Cambridge

Publication Year: 2023

Description: These data support a systematic map of scientific studies about cassava farming practices, which was made with the aim of identifying knowledge gaps and clusters. A secondary aim of the study was to develop a hierarchical classification system for [1] farming interventions, and [2] agricultural, economic, and environmental outcomes. This standardized classification system for agricultural metadata can facilitate dataset reuse and promote research efficiency across syntheses. Following our published protocol [2], we searched eight publication databases/repositories using the search string “cassava OR mandioca OR manihot OR manioc OR yuca” in December 2017. We screened 36,580 records at title and abstract and then at full text stage, and included publications that measured the impact of cassava farming practices on agricultural or environmental outcomes, including: yield, soil, water, wildlife, pests, pollutants, profits, and labour. We classified the resultant 1,599 publications by interventions, outcomes, study location, study years, and study design. We assessed coding consistency using Kappa scores. This map is available online via an interactive database: https://www.metadataset.com/ (registration required). The Kappa scores indicated that we successfully developed a consistent intervention and outcome ontology that can be applied to other systems.

Cite as: HoodA.S.C., Christie A.P., Sutherland W. J. & Shackelford G.E. (2023) Dataset supporting: A systematic map of cassava farming practices and their agricultural and environmental impacts using new ontologies: Agri-ontologies 1.0. University of Reading. Dataset. [DOI: [10.17864/1947.000429](https://doi.org/10.17864/1947.000429)]

Related publications:

1. HoodA.S.C., Shackelford G. E., Christie A. P., Usieta H. O., Martin P. A. & Sutherland W. J (2023) A systematic map of cassava farming practices and their agricultural and environmental impacts using new ontologies: Agri-ontologies 1.0. *Ecological Solutions and Evidence.* 4(2).
2. Shackelford G. E., Haddaway N. R., Usieta H. O., Pypers P., Petrovan S. O & Sutherland W. J. (2018) Cassava farming practices and their agricultural and environmental impacts: a systematic map protocol. *Environmental Evidence.* 7(30).

Contact: Amelia Hood [a.s.hood@reading.ac.uk](mailto:a.s.hood@reading.ac.uk)

Acknowledgements: We thank our co-authors Hope O Usieta and Philip A Martin. We are grateful to Dr Neal Haddaway, Dr Silviu Petrovan, and Dr Pieter Pypers for their help writing the protocol that this systematic map was based on.

2. TERMS OF USE

------------

Copyright 2023 University of Reading and University of Cambridge. This dataset is licensed under a Creative Commons Attribution 4.0 International Licence: https://creativecommons.org/licenses/by/4.0/.]

3. PROJECT AND FUNDING INFORMATION

------------

We are grateful to the A. G. Leventis Foundation, Arcadia, and the David and Claudia Harding Foundation for funding. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

4. CONTENTS

------------

File listing

1. “FullTexts.csv”: a spreadsheet of all records (included and excluded) showing the article Authors, Year (published), Title (article), Journal, Volume, Issue, DOI, URL (on the metadataset website, where the systematic map is hosted), Abstract, Status (whether the article was included in the map or not), and ExclusionReason (if the article was excluded, why it was excluded).
2. “CountByCountry.csv”: a spreadsheet showing the number of records according to study location (country). This is the country where the experiment or field study was located rather than the country the article was published in. iso\_a3 (Country shorthand) and Country are shown. count-by-country is the number of records.
3. “InterventionsList.docx”: a hierarchical classification of farming practice interventions.
4. “Interventions.csv”: a spreadsheet showing the number of records according to study intervention. Each row is a publication that we coded an intervention for, and publications that we coded multiple intervention for (at the lowest level of the intervention hierarchy – see “InterventionsList.docx”) are coded over multiple rows, with one row for each intervention. The columns are:

* “Publication”: the unique publication ID we prescribed to each publication on www.metadataset.com
* “Authors”: the publication authors
* “Year”: the publication year
* “Title”: the publication title
* “Journal”: the publication journal
* “Volume”, “Issue”, “Pages”, and “DOI”: the publication volume, issue, pages, and DOI
* “Country”: the country the experimental study was conducted (i.e. not the country where it was published)
* “Start year” and “End Year”: the start and end year of the experimental study
* “Intervention level 1”: the intervention we coded at the highest (i.e. least specific) level of the intervention hierarchy
* “Intervention level 2”: the intervention we coded at the upper-middle level of the intervention hierarchy (this is a subset of “Intervention level 1”).
* “Intervention level 3”: the intervention we coded at the lower-middle level of the intervention hierarchy (this is a subset of “Intervention level 2”).
* “Specific intervention”: the intervention we coded at the lowest, or most specific, level of the intervention hierarchy (this is a subset of “Intervention level 3”).
* “Abbreviated intervention for plotting”: the same as “Specific intervention”, but with the terms abbreviated to facilitate plotting in the manuscript.
* “Blocked”, “Replicated”, “Randomised”, “Controlled”, “Correlated”, “Before-and-After”,

and “Paired”: each column is a feature of the study design used to test the specific intervention coded in each row. If “TRUE” this study design was used, if “FALSE” it wasn’t used, and if blank the study design was not tracked. “Controlled”, “Correlated”, and “Before-and-after” are study comparators. “Controlled” includes comparisons of plots with treatment applied vs plots without treatment applied, “Correlated” includes comparisons of plots with a treatment (pre-existing application; e.g., before the study began) with plots without a treatment (pre-existing control)), and “Before-and-after” includes comparisons of plots through time before and after treatment is applied. “Blocked” studies are controlled studies with treatments next to each other spatially, whereas “Paired” studies are correlated studies with sites that are selected to have similar characteristics. “Replicated” refers to studies that have more than one replicate, and “Randomised” to controlled studies where treatments are allocated randomly.

1. “OutcomesList.docx”: a hierarchical classification of agricultural, environmental, and economic outcomes.
2. “Outcomes.csv”: a spreadsheet showing the number of records according to study outcome. Each row is a publication that we coded an outcome for, and publications that we coded multiple outcomes for (at the lowest level of the outcome hierarchy – see “OutcomesList.docx”) are coded over multiple rows, with one row for each outcome. The columns are:

* “Publication”: the unique publication ID we prescribed to each publication on www.metadataset.com
* “Authors”: the publication authors
* “Year”: the publication year
* “Title”: the publication title
* “Journal”: the publication journal
* “Volume”, “Issue”, “Pages”, and “DOI”: the publication volume, issue, pages, and DOI
* “Country”: the country the experimental study was conducted (i.e. not the country where it was published)
* “Start year” and “End Year”: the start and end year of the experimental study
* “Outcome level 1”: the outcome we coded at the highest level of the outcome hierarchy
* “Outcome level 2”: the outcome we coded at the middle level of the outcome hierarchy (this is a subset of “Outcome level 1”).
* “Specific outcome”: the outcome we coded at the lowest, or most specific, level of the outcome hierarchy (this is a subset of “Outcome level 2”).
* “Abbreviated outcome for plotting”: the same as “Specific outcome”, but with the terms abbreviated to facilitate plotting in the manuscript.

5. METHODS

-----------

Full details are provided in the related pre-registered protocol [2] and systematic map [1] which are published open access.

Following our published protocol [2], we searched eight publication databases/repositories using the search string “cassava OR mandioca OR manihot OR manioc OR yuca” in December 2017. These repositories were: *Scopus; Web of Science Core Collection; AGRICOLA; AGRIS; Conservation Evidence; Google Scholar* (first 500 results sorted by “relevance”)*; the Document Repository of the Food and Agriculture Organisation (FAO)* (first 500 results); and *the repository of the Consultative Group on International Agricultural Research (CGIAR) Centres, in CGSpace: A Repository of Agricultural Research Outputs* (first 500 results). We screened 36,580 records at title and abstract and then at full text stage, and included publications that measured the impact of cassava farming practices on agricultural or environmental outcomes, including: yield, soil, water, wildlife, pests, pollutants, profits, and labour. We excluded duplicates (16,872 records), articles that were not accessible according to University of Cambridge journal subscriptions (301 records), articles that were not found (1,204 records), articles that were not in English (58 records), articles that were secondary literature (231 records) and articles that were not relevant to our aims. We classified the resultant 1,599 publications by interventions, outcomes, study location (country), study years, and study design. We included four aspects of study design: reporting standards (whether the article reported the experimental start date and study location), experimental design, study duration, and publication delay.

We did not code outcomes, years, or experimental design for the intervention “10.10.10.TT.20. Planting a different variety/cultivar (e.g., a disease-resistant variety)”. This was due to time constraints. This intervention was the most common and therefore would have been the most time-consuming (939 intervention occurrences, 24% of intervention occurrences). Furthermore, it would be a difficult to apply a synthesis to this intervention as there are numerous cultivars, which are largely regional. For publications that tested this intervention and others, we coded the other interventions.

Many publications tested multiple different interventions and outcomes, or the same intervention multiple times (n = 111), which meant that the number of interventions (3,890) and outcomes (4,236) tested was greater than the number of mapped publications (1,599). We mapped the country-of-study of 998 publications; 76 were excluded as they did not report the country and 525 were excluded as they only tested the intervention “*Planting a different variety/cultivar (e.g., a disease-resistant variety)*”. Country occurrences (*n* = 1,113) is greater than mapped publications (*n* = 998) as 47 publications were conducted in multiple countries.

We assessed coding consistency using Kappa scores, which show a measure of agreement between two coders. The Kappa scores indicated that we successfully developed a consistent intervention and outcome ontology that can be applied to other systems (Title and abstract: Kappa 0.69 and 89% agreement, Full text: Kappa 0.76 and 91% agreement).

This map is available online via an interactive database: https://www.metadataset.com/ (registration required).