

Metadata Template

Christian Krestel

2023-07-06

Contents

Meetings	1
Protocol Meeting 2023-06-28	1
Changes	2
Implemented changes on qa and adpted protocol following the the meeting on 2023-06-28 . .	2
Open changes	2
Template: Code repository catalogue	3
Create a package	3
Information on metadata	3
Project description	3
Title statement	3
Authors and contributors	4
Version statement	4
Scope and Coverage	4
Processes and Output	5
Data	5
Methods, software and scripts	6
Access and Rights	8
Contacts	8
Tags (To be discarded)	8
External Resources	8
Open Issues and Remarks	9
Errors and Issues	10

Version number: 6

Meetings

Protocol Meeting 2023-06-28

1. Confidentiality: Cannot only be Yes, No or partial. Has to involve more information or empty
2. Technology requirement: Do not put Standard. Put only if something more is needed: Cluster etc
3. No DOI in the subtitle, no Working paper number in the subtitle.
4. DOI is not for the paper, it is for the package.

5. Geographic region in scope: to be used with the codes provided by the editor (WB codes)
6. Keep separating donors from the
7. Delete the field contributors. Go into Acknowledgement.
8. Keep English EN (and fill it it), but it will not be displayed
9. Disclaimer is for the package (and not from the paper) to be drafted by legal
10. Get rid of the tags
11. Get rid of the project website
12. There is a field for repository, there one puts potential Github repositories
13. Corresponding author is not a role. The corresponding author is the author that shows in Contacts
OK
14. Type in country to be used only if it is not a country, ocean, city, province etc. New codes are provided
OK (use world bank ones)

Changes

Implemented changes on qa and adpted protocol following the the meeting on 2023-06-28

1. Remove confidentiality if public: OK
2. Leave Technology environment empty (instead of Standard): OK
3. DOI of paper deleted: OK
4. All subtitles deleted
5. Disclaimer: deleted, waiting for new one OK
6. Project website OK (aarts github is now in repository)
7. Geographic region: changed the type to cty; use WB codes now
8. Coeditors in field contributor deleted
9. Added to template that the field Language will not be displayed
10. Tags have been changed, but are still there. Naming is now more meaningful: Fully Reproducible, Partially Reproducible, Not Reproducible
11. Excel file for document number and methods has been created and linked to the metadata scheme

Open changes

1. Contributors: RAs are still in place; must adjust ACK if to be deleted
2. Sponsors: still in place; must adjust ACK if to be deleted
3. Tags are still there; labeling is now more meaningful; perhaps we can add them to the meta data but hide them

Template: Code repository catalogue

Create a package

1. Create new project: Type script
2. Screenshot of the *complete* title page
3. For a published paper, prepare the link to the repository. For a WP prepare a zip that contains: readme, code files, all public data files, data statement (if not included in readme) and license. For **DIME verified** packages also the reproducibility verification statement

Information on metadata

This is information about the creator of the metadata. Sufficient fields

1. Name: Surname only: Smith
2. Abbreviation: First and Last initials: PS
3. Affiliation: Department, World Bank
4. Production date: Year-Month-Day
5. Version number of the metadata: Entered as 1,2,3

Project description

Title statement

Sufficient fields are

1. Primary ID: Ideally a DOI or an organization consistent scheme;

(a) Published Paper:

- i. A corresponding PRWP exists: FORMAT

PP_CountryCodeIso3_Year_PRWP-PRWPno_versionno Example: PP_CRI_2018_PRWP-8524_v01

with

- For ISO codes see [Iso Codes](#)
- versionno refers to the version number of the PRWP/PP
- Year refers to the year of publication of the published paper.

- ii. A corresponding PRWP does not exist: FORMAT

PP_CountryCodeIso3_Year_EXT-EXTno_versionno Example: PP_CRI_2018_EXT-1_v01

EXTno is a unique running number assigned by the metadata producer. Please see [Qa Excel](#) for the current number.

- (b) Policy Research Working Paper: FORMAT

RR_CountryCodeIso3_Year_PRWP-PRWPno_versionno RR_CRI_2018_PRWP-8525_v01
--

2. Other identifiers: DOI of the package and <https://doi.org/number>¹
3. Title: Of the following form

Reproducibility package for TITLE (full title of the paper)

Authors and contributors

Sufficient are

1. Authoring entity: author, affiliation and ORCID number and email when available. For World Bank authors use **World Bank** and always provide email address (query in staff directory as needed).
2. Sponsors: All financial contributions and individual/general grants
3. Acknowledgement Statement: All acknowledgments except financial support

We are grateful to Benu Bidani, Francisco Ferreira, Luis-Felipe..

.. and seminar participants at the World Bank's Poverty ...

Version statement

Sufficient are

1. Project completion date: FORMAT: YEAR-MONTH
 - PP: Date at which the article was published: Year-Month
 - WP: Must be supplied for the catalogue: use cover page of PRWP if published, o.w. month of review

Scope and Coverage

Sufficient are

1. Abstract: Abstract of the article; Delete any hyphens from the abstract.²
2. Geographic areas: Format: Name, type (cty...country, reg...region (WLD))

Lebanon , LBN, cty

World Bank codes. The editor provides a list for these.

3. Keywords: No keywords if the paper/journal homepage does not provide keywords; Yes if keywords are provided.
4. JEL codes:

ID D63
 Name Equity, Justice, Inequality
 Parent ID (empty)
 Vocabulary JEL Classifications

¹Requires a later update.

²If the abstract has hyphens, copy and pasting the abstract will always copy the hyphen (no matter if Word is used in between). The window for the abstract resizes with the browser window, therefore the hyphen will almost surely be in the wrong place. Delete.

URL <https://www.aeaweb.org/econlit/jelCodes.php?view=jel>

Note: Some journals provide JEL codes on the homepage for the paper, but not in the paper itself (The Economic Journal is an example)

Processes and Output

Fields Process, Institutional review board, Approval process, Reviews and comments seem more apt at describing the outcome of a process other than working papers and published papers. These fields are currently not used and therefore not included in the template, except for this remark.

Sufficient fields are

1. Output:

- Type
 - WP: Working Paper
 - PP: Published Paper (not paper, not article, not Article)
- Title: Title of the paper
- Authors: Authors: all comma separated; first name, last name. Patty Smith, Patty Smith, Patty Smith
- Description: PRWP + PRWP number, external Working paper series etc

Toulouse School of Economics Working Papers No 1391
- DOI: YES (if it exists)
- URL: YES (Editor displays an error at https, but it is accepted)

2. Language: English, EN; enter, it will not be displayed

Data

Sufficient are

1. Data statement:³ For external packages, the following rule:

- If all data is public:
 - Fill in: "All data is public and contained in the (external) reproducibility package."
 - Add data availability statement if available from journal package; often separate from the zip, has to be copied over manually.
- If some data is confidential: Paste from the readme notes about confidentiality.

2. Datasets:⁴ Used both for PP and WP, if they data sets are of proper value and stored in another WB repository (like microdata library or MDL/DDH).⁵

(a) Name

³ Description: An overall statement on the data used in the project. A separate field is provided to list and document the origin and key characteristics of the datasets

⁴Description of this field: "This field is used to provide an itemized list of datasets used in the project. The data are not documented here (specific metadata are available for documenting data of different types, like the DDO for microdata, the ISO 19139 for geographic datasets etc)"

⁵Process may differ between PP and WP (involvement in cataloging data) No cases so far, 2023-06-27.

- (b) Data set ID: (empty)
- (c) Note: (empty)
- (d) Access policy:
 - Published elsewhere: Note access/license
 - Published with package: Authors must confirm there is no confidentiality.
- (e) License (empty)
- (f) License URL (empty)
- (g) Data URL

Methods, software and scripts

Sufficient are:

1. Methods: Econometric methods used. This field will learn from past entries. Possibly add notes (more details about the context) in the field Note. The following link collects past Method entries, [Qa Excel](#)
2. Software:
 - Name: Stata, R, Matlab, Dynare
 - Version: 17
3. Scripts:
 - (a) File name: Name of the zip file. Naming convention: PrimaryID_prg_v01.zip
 - (b) Zip package: Same as file name⁶
 - (c) Title:

Reproducibility package (data and code) for Title

Title may be shortened if adequate.
 - (d) Authors: empty⁷ If authors of the paper and the code differ and authors want to indicate this, add here list of authors for the code.
 - (e) Date: Year-Month (usually the completion/version date of the paper)
 - (f) Software: Stata
 - (g) Description: Simple description like a tree of what is in the zip file
 - (h) Methods: This are specific methods on the script level (and not econometric methods)
 - (i) Dependencies: Here enter all packages necessary:⁸

coefplot , distinct , lincomeest , lprobust .ado
 - (j) Instructions: Write: "See readme in reproducibility package"
 - (k) Repository: ICPSR, Harvard Dataverse, Journal webpage etc. For convenicnes, for ICPSR; Also github repositories for the project go here use

⁶File name and zip package are identical, because we do not list files separately, for which the field file name is meant.

⁷According to descr: A list of authors and coauthors of the script

⁸It may be preferable to enter this in Software, libraries: **Error: Must be an array.** Does not accept more than one entry, and also not a string. Behavior of the editor his random here

AEA Data and Code Repository (ICPSR)
Reproducible Research Repository (World Bank)

For Harvard Dataverse use

TBC

For github

TBC

(l) Notes: Here who verified the package:

Checked for reproducibility externally at the journal.
Checked for reproducibility at Development Impact Evaluation (DIME), World Bank

(m) License:

- PP

– ICPSR:

Name: Creative Commons Attribution 4.0 International (CC BY 4.0)
License
URL: <https://creativecommons.org/licenses/by/4.0/>

- WP:

Name: Modified BSD3
URL: <https://opensource.org/licenses/bsd-3-clause/>

4. Repository: Harvard, ICPSR- Type empty - URL, Journal webpage

AEA Data and Code Repository
<https://www.openicpsr.org/openicpsr/aea>

5. Reproducibility:

(a) Technology environment:

- WP + verified by Dime: Environment in which the Rep check was conducted
- PP: usually empty (bc not available)

(b) Technology Requirements: empty; except for sth unusual: Cluster, large run time etc

(c) Reproduction instructions: No details:

A readme file with detailed instructions is part of the (external) reproducibility package.

Short details:

A readme file and a single do file , AP_final_AEJ. do are provided.

No readme file exists:

Please run the do-file to replicate the tables in the data. The package does not include a readme file.

Access and Rights

The following are sufficient:

1. Disclaimer: To be formulated by legal
2. Confidentiality: Empty if not confidential. Otherwise
 - If data is confidential: Yes, please refer to the data statement and the readme file.
 - If code is confidential: Yes, part of the code is confidential and cannot be published. Please refer to the readme file.
3. License: as in scripts
 - PP
 - ICPSR:

Creative Commons Attribution 4.0 International (CC BY 4.0) License
<https://creativecommons.org/licenses/by/4.0/>

- WP:

Modified BSD3
<https://opensource.org/license/bsd-3-clause/>

Contacts

1. WB author or corresponding author: Fill fields Name, email, affiliation, role Author

Patty Smith
Author
psmith@worldbank.org
World Bank

2. Reproducibility group contact

Reproducibility WBG
reproducibility@worldbank.org
reproducibility.worldbank.org

Tags (To be discarded)

1. Tag: Partially Reproducible, Fully Reproducible, Not Reproducible, Tag group Reproducible
2. Type: PRWP, Published Paper

External Resources

1. Zip file/link for PP and WP
 - (a) File type: prg
 - (b) Title

Reproducibility package (data and code) for TITLE (of paper), PID PrimaryID

Title of paper may be shortened⁹

(c) Date: equal to version date/completion date: FORMAT Year-Month

(d) Description:

- PP:

Externally available on ICPSR

(e) Resource attachment:

- WP: Filename in the format PID_prg_v01.zip (as the zipname in scripts)
- PP: Link to the repository/article

2. For packages checked by DIME, attach the reproducibility verification statement. Title: Reproducibility verification certificate.

3. Readme file. For all packages include the readme file (changing the name to readme if needed)

(a) : Title: README file for the reproducibility package Health Care Rationing in Public Insurance Programs.

(b) Date: equal to version date/completion date: FORMAT Year-Month

(c) Filename: README.pdf

Open Issues and Remarks

1. **Resolved** Tags: We use tag group Reproducible, YES, NO, PARTIAL, shown is then only the value and not the tag group. Do we have to tag like: reproducible, partially reproducible, not reproducible?

2. **Partially Resolved: UPI?** Information on Metadata:

(a) Abbreviation: Fine. But also use an identifier. Perhaps UPI is problematic if this is public. Use a new identifier?

3. Project description

(a) **Resolved** Creating a primary ID: Use country ISO 3 code or not? Use RR for working papers, PP for published papers (see naming convention in Proj_descr, Title statement above)

(b) Authors and contributors

i. Contributors: Should RAs only be named if they contributed to the code?

ii. Acknowledgment: To include or not? (Little work, little value?)

(c) Processes and Output: **Resolved** Fields Process, Institutional Review board, approval process, reviews and comments will not be used

(d) Methods, software and scripts:

i. **Resolved** Remark: Field Methods: There are two method field: one stand-alone, one with Scripts. Current usage: Stand-alone for econometric/statistical techniques (e.g. OLS, random forests...). This field will learn and suggest keywords.

⁹It is not possible to put the PID into the subtitle as we do with the title at the beginning, because the subtitle is not displayed. Putting the PID into the subtitle means dropping the PID

Errors and Issues

1. `proj_descr_script.[0].title`

is missing: this is caused by: in methods, software, scripts a file name was given, but a title is missing.

Title is obligatory, conditional on file name entered

2. **Resolved: Works now.** In Processes and output, in URL, the protocol https is not accepted. The editor sometimes shows an error, but this has no consequence

3. in

External Resource, File type

there is no option package (only prg,...). Perhaps one would want to have this option, because we refer to whole packages, if the source is external

4. There is no real field

Journal

. For published articles, one cannot indicate in which journal it was published.

5. In Scope, Topic it would be nice to have all JEL descriptions available once one enters the code

6. In Software, libraries: sometimes the following error is raised: Must be an array. One cannot enter more than one package

7. (On the external instance of the editor) In Scripts, License, License is not displayed, although a value is supplied

8. One has to supply completion date for the catalogue page to show the year (WLD xxxx). Working papers have no completion date if seen as project. Just a version date.