Metadata Template

Christian Krestel

2023-06-27

Contents

Version number: 3

Template: Code repository catalogue

1. Create new project: Type script
2. Screenshot of the *complete* first page
3. For a published paper, prepare the link to the repository. For a WP prepare a zip that contains all data and code; 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: Second name only: Smith

2. Abbreviation: PS

3. Name: Smith

4. Affiliation: World Bank

5. Production date: Year-Month-Day

6. Version number: 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
      ```

      ii. A corresponding PRWP does not exit: FORMAT

      ```
      PP_CountryCodeIso3_Year_EXT−EXTno_versionno
      Example :  PP_CRI_2018_EXT−1_v01
      ```

      EXTno is a unique running number assigned by Dime.

   (b) Working Paper: FORMAT

   ```
   RR_CountryCodeIso3_Year_PRWP−PRWPno_versionno
   RR_CRI_2018_PRWP−8525_v01
   ```

2. Other identifiers: DOI and https://doi.org/number

3. Title: Of the following form

   ```
   Reproducibility  package  for  TITEL( paper )
   ```

4. Subtitle:

   - WP: World Bank Policy Research Working Paper (PRWP) number
   - PP: DOI number (no https)

Authors and contributors

Sufficient are

1. Authoring entity: author, affiliation and ORCID number and email when available; For world bank authors use **World Bank**. Example:

   ```
   Christoph Lakner World Bank ORCID 0000, clakner@worldbank.org
   ```

2. Contributors: like RAs (not conference participants). Also coeditors if specifically mentioned. FORMAT

```
Coeditor
```

3. Sponsors: All financial contributions and individual/general grants

4. Acknowledgements: Usually empty; Example: Corporate Taxation: special mentioning ministry employees for data assistance (could go to acknowledgment statement)

5. Acknowledgement Statement:

```
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, bc it is needed for the front page of cataloge

2. Version date: FORMAT: YEAR-MONTH-DAY

   - PP: empty
   - WP: data of publication as research paper in WB Year-Month-Day

Scope and Coverage

Sufficient are

1. Abstract: Abstract of the article; Hyphens are a problem. Check all abstracts. Also carriage returns are a problem, direct copy pasting leads to bad results

2. Geographic areas: Format: Name, Code, Type

```
Lebanon, LBN, ISO−3166 Alpha−3
```

3. Keywords: No keywords if none are given; Yes if given.

4. JEL codes:

```
ID D63
Name Equity, Justice, Inequality
Parent ID (empty)
Vocabulary JEL Classifications
URL https://www.aeaweb.org/econlit/jelCodes.php?view=jel
```

Processes and Output

The following fields exist

1. Process: empty for PP, tbd for WP

2. Institutional review board: empty

3. Approval process: empty

4. Reviews and comments:
   - WP: If available (tbd, 2023-06-22)
   - PP: empty

5. 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, Petty Smith
   - Description: Add kind and No of external working paper if applicable

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)

6. Language: English, EN

Data

Sufficient are

1. Data statement:[1]. 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:[2] Not used for PP, will be used for WP, if they data sets are of proper value and stored in another WB repository (like microdata library). No such case arose so far. TBD
   - (a) Name
   - (b) Data set ID: (empty)

---

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

[2] Description 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)"

(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

Fields that are not listed can be discarded.

1. Methods: Econometric methods used. This field will learn from past entries. Possibly add notes (more details about the context)

2. Software: Stata 17.0, R, etc

3. Scripts:

   (a) File name (WP+PP): Name of the zip file. Naming convention: PrimaryID_prg_v01.zip

   (b) Zip package (WP+PP): Same as file name[3]

   (c) Title:

   ```
   Reproducibility package (data and code) for Title
   ```

   Title may be shortened if adequate.

   (d) Authors: WP+PP: empty[4] 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)

   (f) Software: Stata

   (g) Description: empty

   (h) Methods: This is methods used in the script. Usually empty

   (i) Dependencies: Here enter all packages necessary:[5]

   ```
   coefplot, distinct, lincomeest, lprobust.ado
   ```

   (j) Instructions: Write: "See readme in reproducibility package"

   (k) Repository: ICPSR, Harvard Dataverse, Journal webpage etc. For ICPSR the correct description is

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

   (l) Notes: Here who verified the package.

---

[3]File name and zip package are identical, because we do not list files separately, for which the field file name is there.

[4]According to descr: A list of authors and coauthors of the script

[5]It 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

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

    (m) License:

- PP

    - ICPSR:

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

- WP: Modified BSD3

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: empty

    (b) Technology Requirements: Standard, Cluster...; include runtime if available

    (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.
```

Access and Rights

1. Disclaimer: Perhaps not necessary

```
The  views  expressed  here  are  the  authors,  and  do  not  necessarily  reflect  the
    official  views  of  the  World  Bank,  its  Executive  Directors,  or  the  countries
    they  represent.
```

2. Confidentiality: YES or NO, or PARTIAL. Do not add description

3. Citation requirement: (empty)

4. 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/
```

5. Copyright: (empty)

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

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 shortened[6]

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

   (d) Description:

   - WP: empty
   - 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. All else emtpy.

---

[6]It 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

Open Issues and Remarks

1. ResolvedTags: 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) ResolvedCreating 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

        i. ResolvedKept, bc maybe useful for WP

        ii. Unused fields for PP: Should we use these for WP? An alternative is not to use these (in particular Process) and add working paper and published version separately.

        - Process: Description:

        ```
        This element is used to document the life cycle of the research project
            , from its design and inception to its conclusion. This can include
             phases of fundraising, approval by a review board, concept note
            review, data acquisition, analysis, publishing of a working paper,
            peer review, publishing in journal, presentation to conferences,
            publishing, evaluation, reporting to sponsors, etc. It is
            recommended to provide these steps in a chronological order.
        ```

        - Approval process:

        ```
        The "Approval process" is a group of six elements used to describe the
            formal approval process(es) (if any) that the project had to go
            through. This may for example include an approval by an Ethics
            Board to collect new data, followed by an internal review process
            to endorse the results.
        ```

    (d) Methods, software and scripts:

        i. 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.

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 on 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) In Scripts, License, License is not displayed, although a value is supplied

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

Licenses (Which to use?)

```
Economic Studies
This article is published and distributed under the terms of the Oxford University
    Press, Standard Journals Publication Model (https://academic.oup.com/journals/pages
    /open_access/funder_policies/chorus/standard_publication_model)
```