

Process data from reproducibility service

Lars Vilhuber
2021-03-16

- Citation
- Requirements
- Data
 - The workflow
 - Raw process data
 - Anonymized data
 - Publishing data
 - Preserving some minimal data for internal purposes
- Describing the Data
 - Variables
 - Sample records
 - R session info

Note: The PDF version (<https://aeadataeditor.github.io/processing-jira-process-data/README.pdf>) of this document is transformed by manually printing from a browser.

Citation

Vilhuber, Lars. 2021. "Process data for the AEA Pre-publication Verification Service." *American Economic Association [publisher]*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-03-16. <https://doi.org/10.3886/E117876V2> (<https://doi.org/10.3886/E117876V2>)

```
@techreport{10.3886/e117876v2,  
  doi = {10.3886/E117876V2},  
  url = {https://www.openicpsr.org/openicpsr/project/117876/version/V2/view},  
  author = {Vilhuber, Lars},  
  title = {Process data for the AEA Pre-publication Verification Service},  
  institution = {American Economic Association [publisher]},  
  series = {ICPSR - Interuniversity Consortium for Political and Social Research},  
  year = {2021}  
}
```

Requirements

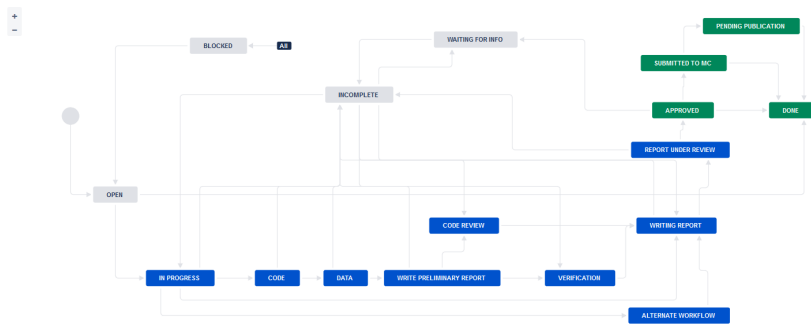
This project requires

- R (last run with R 4.0.4)
 - package here (>=0.1)

Other packages might be installed automatically by the programs, as long as the requirements above are met, see Session Info.

Data

The workflow



Workflow stages

Raw process data

Raw process data is manually extracted from Jira, and saved as

- export_MM-DD-YYYY.csv (for detailed transaction-level data)

The data is not made available outside of the organization, as it contains names of replicators, manuscript numbers, and verbatim email correspondence.

At this time, the latest extract was made 12-22-2020.

Anonymized data

We subset the raw data to variables of interest, and substitute random numbers for sensitive strings. This is done by running 01_jira_anonymize.R:

```
source(file.path(programs,"01_jira_anonymize.R"),echo=TRUE)
```

Publishing data

Some additional cleaning and matching, and then we write out the file

```
source(file.path(programs,"02_jira_anon_publish.R"),echo=TRUE)
```

Preserving some minimal data for internal purposes

For analytics, we preserve the unmodified manuscript number. This file is produced here, but not published.

```
source(file.path(programs,"03_jira_manuscript.R"),echo=TRUE)
```

Describing the Data

The anonymized data has 15 columns.

Variables

```
##
## --- Column specification -----
## cols(
##   name = col_character(),
##   label = col_character()
## )
```

| name | label |
|---------------------|--|
| ticket | The tracking number within the system. Project specific. Sequentially assigned upon receipt. |
| date_created | Date of a receipt |
| date_updated | Date of a transaction |
| mc_number_anon | The (anonymized) number assigned by the editorial workflow system (Manuscript Central/ ScholarOne) to a manuscript. This is purged by a script of any revision suffixes. |
| Journal | Journal associated with an issue and manuscript. Derived from the manuscript number. Possibly updated by hand |
| Status | Status associated with a ticket at any point in time. The schema for these has changed over time. |
| Software.used | A list of software used to replicate the issue. |
| received | An indicator for whether the issue is just created and has not been assigned to a replicator yet. |
| Changed.Fields | A transaction will change various fields. These are listed here. |
| external | An indicator for whether the issue required the external validation. |
| subtask | An indicator for whether the issue is a subtask of another task. |
| Resolution | Resolution associated with a ticket at the end of the replication process. |
| reason.failure | A list of reasons for failure to fully replicate. |
| MCRRecommendation | Decision status when the issue is Revise and Resubmit. |
| MCRRecommendationV2 | Decision status when the issue is conditionally accepted. |

Sample records

| ticket | date_created | date_updated | mc_number_anon | Journal | Status | Software.used | received | Changed.Fields | external | subtask | Resolution | reason.failure | MCRrec |
|-------------|--------------|--------------|----------------|-----------|--------|---------------|----------|-------------------------------|----------|---------|------------|----------------|--------|
| AEAREP-1676 | 2020-12-21 | 2020-12-21 | 481 | AEJ:Micro | Open | | No | Journal | No | NA | | | |
| AEAREP-1676 | 2020-12-21 | 2020-12-21 | 481 | | Open | | No | Manuscript Central identifier | No | NA | | | |
| AEAREP-1676 | 2020-12-21 | 2020-12-21 | 481 | | Open | | Yes | | No | NA | | | |
| AEAREP-1674 | 2020-12-18 | 2020-12-21 | 480 | AER | Open | | No | openICPSR Project Number | No | NA | | | |
| AEAREP-1674 | 2020-12-18 | 2020-12-18 | 480 | AER | Open | | No | Journal | No | NA | | | |
| AEAREP-1674 | 2020-12-18 | 2020-12-18 | 480 | | Open | | No | Manuscript Central identifier | No | NA | | | |

R session info

```
sessionInfo()

## R version 4.0.4 (2021-02-15)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 20.04 LTS
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblas-p0.3.8.so
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=C
##  [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] readr_1.4.0  knitr_1.31   tidyr_1.1.2  stringr_1.4.0 dplyr_1.0.4
##
## loaded via a namespace (and not attached):
## [1] rstudioapi_0.13  magrittr_2.0.1  hms_1.0.0      tidyselect_1.1.0
## [5] here_1.0.1       R6_2.5.0        rlang_0.4.10   highr_0.8
## [9] tools_4.0.4      xfun_0.21       cli_2.3.0      DBI_1.1.1
## [13] htmltools_0.5.1.1 ellipsis_0.3.1  yaml_2.2.1     rprojroot_2.0.2
## [17] digest_0.6.27    assertthat_0.2.1 tibble_3.0.6   lifecycle_1.0.0
## [21] crayon_1.4.1     purrr_0.3.4     vctrs_0.3.6    glue_1.4.2
## [25] evaluate_0.14    rmarkdown_2.6   stringi_1.5.3  compiler_4.0.4
## [29] pillar_1.4.7     generics_0.1.0  pkgconfig_2.0.3
```