ICPSR/Retraction Watch, First Look

Ashvin Narayan

26 September 2025

DOIs available

```r
# What proportion of entries in each table have DOIs (ICPSR, Retraction Watch)
c(length(icpsr$DOI[which(!is.na(icpsr$DOI))]) / length(icpsr$DOI),
  length(retractions$OriginalPaperDOI[which(!is.na(retractions$OriginalPaperDOI))]) /
    length(retractions$OriginalPaperDOI))
```

```
## [1] 0.4917015 0.9444636
```

```r
# Duplicate DOIs in retractions?
retractions |>
  drop_na(OriginalPaperDOI) |>
  add_count(OriginalPaperDOI) |>
  filter(n > 1) |>
  distinct() |>
  select("Record.ID", "Title", "OriginalPaperDOI", "n") |>
  nrow()
```

```
## [1] 10821
```

```r
# Retraction Watch database contains 'duplicate' DOIs - if a paper
# was corrected then retracted, there are separate entries for the
# correction and the retraction

# pivot_wider before merging?

# What proportion of entries for each type of retraction reason?
retractions |>
  drop_na(RetractionNature) |>
  group_by(RetractionNature) |>
  summarize(n())
```

```
## # A tibble: 4 x 2
##   RetractionNature `n()`
##   <list>           <int>
## 1 <chr [1]>        61681
## 2 <chr [1]>         1362
## 3 <chr [1]>         3440
## 4 <chr [1]>          158
```

```r
# Duplicate DOIs in ICPSR?

icpsr |>
  drop_na(DOI) |>
```

```
  add_count(DOI) |>
  filter(n > 1) |>
  distinct() |>
  select("Reference.ID", "Title", "DOI", "n") |>
  nrow()
```

[1] 0

```
# Doesn't look like it! Thank you data librarians :)
```

```
# For entries without DOIs in ICPSR, what type of materials are they
```

```
# 'Type.of.Work' unavailable for what proportion of icpsr entries?
icpsr |>
  filter(is.na(Type.of.Work)) |>
  select("Reference.ID", "Title", "DOI", "Type.of.Work") |>
  nrow() / nrow(icpsr)
```

[1] 0.9033972

```
# 'Type.of.Work' unavailable for what proportion of icpsr entries
# that have a DOI?
icpsr |>
  filter(is.na(DOI)) |>
  filter(is.na(Type.of.Work)) |>
  select("Reference.ID", "Title", "DOI", "Type.of.Work") |>
  nrow() / nrow(filter(icpsr, is.na(DOI)))
```

[1] 0.8153003

```
# ... that don't have a DOI?
icpsr |>
  filter(!is.na(DOI)) |>
  filter(is.na(Type.of.Work)) |>
  select("Reference.ID", "Title", "DOI", "Type.of.Work") |>
  nrow() / nrow(filter(icpsr, is.na(DOI)))
```

[1] 0.9619963