Journals' policies of storage and reuse of raw research data

Rafael Aleixandre (www.rafaelaleixandre.es; aleixand@uv.es)

In collaboration with Antonio Vidal, Antonia Ferrer, Fernanda Peset, Andrea Sixto, Rut Lucas, Adolfo Alonso, Enrique Sánchez-Pérez, Lourdes Castelló & Máxima Bolaños

UYESIS (CSIC-UV). Ingenio (CSIC-UPV)
Purpose

1. To analyse open-data policies concerning the availability of raw data in journals
2. 5 subject categories of the JCR
3. Relation of these policies with the impact factor and the quartile
Revision of the websites of **451 journals** included in the following JCR subject categories:

- Food Science and Technology (FS&T) (124 journals)
- Pediatrics (115 journals)
- Information Science & Library Science (IS&LS) (85 journals)
- Substance Abuse (39 journals)
- Dentistry (88 journals)

**Methods**

Journals' policies of storage and reuse of raw research data
Methods

For each journal we documented rules related to public availability of sharing research data as stated in the instructions to authors.

Journals' policies of storage and reuse of raw research data.
Methods

Journals' policies of storage and reuse of raw research data

Springer accepts electronic multimedia files (animations, movies, audio, etc.) and other supplementary files to be published online along with an article or a book chapter. This feature can add dimension to the author's article, as certain information cannot be printed or is more convenient in electronic form.

Before submitting research datasets as electronic supplementary material, authors should read the journal's Research data policy. We encourage research data to be archived in data repositories wherever possible.
Methods

Variables analysed:

a) **Statement on deposit of supplementary material attached with the manuscript**
b) **Acceptance of reuse of data included as supplementary material**
c) **Acceptance of storage in thematic or institutional repositories**
Methods

Variables analysed:

- Impact factor
- Quartile in JCR
Table 1. Acceptance of the variable “Statement of supplementary material”

<table>
<thead>
<tr>
<th>Quartile in JCR</th>
<th>Dentistry</th>
<th>FS&amp;T</th>
<th>IS&amp;LS</th>
<th>Pediatrics</th>
<th>Substance abuse</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10 (11.4%)</td>
<td>30 (24.2%)</td>
<td>16 (18.8%)</td>
<td>24 (20.9%)</td>
<td>4 (10.3%)</td>
</tr>
<tr>
<td>2</td>
<td>4 (4.5%)</td>
<td>24 (19.5%)</td>
<td>13 (15.3%)</td>
<td>25 (21.7%)</td>
<td>4 (10.3%)</td>
</tr>
<tr>
<td>3</td>
<td>3 (3.4%)</td>
<td>16 (12.9%)</td>
<td>7 (8.2%)</td>
<td>26 (23.6%)</td>
<td>4 (10.3%)</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3 (2.4%)</td>
<td>7 (8.2%)</td>
<td>19 (16.5%)</td>
<td>4 (10.3%)</td>
</tr>
<tr>
<td>Total*</td>
<td>17 (19.3%)</td>
<td>73 (59%)</td>
<td>43 (50.6%)</td>
<td>94 (81.7%)</td>
<td>16 (41.2%)</td>
</tr>
</tbody>
</table>

**Journals' policies of storage and reuse of raw research data**
Figure 1. Acceptance of the variable “Statement of supplementary material”
Results

“Statement of supplementary material”

• From 82% in Paediatrics to 19% in Dentistry

• Statistical significant differences have been found in this variable through the four quartiles, with a higher acceptance in the upper quartiles (Q1 and Q2) and lower in Q3 and Q4.

• The highest values for journals ranked in the Q1 correspond to FS&T (24.2%) and Pediatrics (20.9%).
Table 2. Acceptance of the variable “Reuse”

<table>
<thead>
<tr>
<th>Quartile in JCR</th>
<th>Dentistry</th>
<th>FS&amp;T</th>
<th>IS&amp;LS</th>
<th>Pediatrics</th>
<th>Substance abuse</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9 (10.2%)</td>
<td>31 (25%)</td>
<td>18 (21.2%)</td>
<td>6 (5.2%)</td>
<td>8 (20.5%)</td>
</tr>
<tr>
<td>2</td>
<td>4 (4.5%)</td>
<td>27 (21.7%)</td>
<td>19 (22.3%)</td>
<td>9 (7.8%)</td>
<td>6 (15.4%)</td>
</tr>
<tr>
<td>3</td>
<td>1 (1.1%)</td>
<td>25 (20.2%)</td>
<td>12 (14.1%)</td>
<td>13 (11.3%)</td>
<td>5 (12.8%)</td>
</tr>
<tr>
<td>4</td>
<td>3 (3.4%)</td>
<td>8 (6.5%)</td>
<td>6 (7.1%)</td>
<td>7 (6.1%)</td>
<td>3 (7.7%)</td>
</tr>
<tr>
<td>Total*</td>
<td>17 (19.3%)</td>
<td>91 (73.4%)</td>
<td>55 (64.7%)</td>
<td>35 (30.4%)</td>
<td>22 (56.4%)</td>
</tr>
</tbody>
</table>
Figure 2. Acceptance of the variable “Reuse”
Results

"Reuse"

• From 65% in ISLS to 19% in Dentistry

• The acceptance is higher in the upper quartiles

• The highest values for journals ranked in Q1 correspond to FS&T (25%), IS&LS (21.2%) and Substance abuse (20.5%).
Table 3. Acceptance of “Storage in thematic or institutional repositories”

<table>
<thead>
<tr>
<th>Quartile in JCR</th>
<th>Dentistry</th>
<th>FS&amp;T</th>
<th>IS&amp;LS</th>
<th>Pediatrics</th>
<th>Substance abuse</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14 (15.9%)</td>
<td>30 (24.2%)</td>
<td>20 (23.5%)</td>
<td>20 (17.4%)</td>
<td>8 (20.5%)</td>
</tr>
<tr>
<td>2</td>
<td>9 (10.2%)</td>
<td>25 (20.2%)</td>
<td>16 (18.8%)</td>
<td>21 (18.38%)</td>
<td>6 (15.4%)</td>
</tr>
<tr>
<td>3</td>
<td>5 (5.7%)</td>
<td>20 (16.1%)</td>
<td>12 (14.1%)</td>
<td>22 (19.1%)</td>
<td>6 (15.4%)</td>
</tr>
<tr>
<td>4</td>
<td>5 (5.7%)</td>
<td>6 (4.8%)</td>
<td>9 (10.6%)</td>
<td>14 (12.1%)</td>
<td>3 (7.7%)</td>
</tr>
<tr>
<td>Total*</td>
<td>33 (37.5%)</td>
<td>81 (65.4%)</td>
<td>57 (67%)</td>
<td>77 (67%)</td>
<td>23 (59%)</td>
</tr>
</tbody>
</table>

Journals' policies of storage and reuse of raw research data
Figure 3. Acceptance of “Storage in thematic or institutional repositories”

The diagram shows the acceptance rates across different fields and categories. The categories include Dentistry, Food Sciences & Technology, Information Sciences & Library, Pediatrics, and Substance Abuse. The acceptance rates range from 0% to 30%, with specific percentages indicated for each category and subcategory.
“Acceptance of storage in thematic or institutional repositories”

- From > 65% in IS&LS, Pediatrics and FS&T to only 37% in Dentistry
- The acceptance is higher in the upper quartiles.
CONCLUSIONS

• Different behaviours according to the thematic areas, including within the health sciences areas (from 82% in Paediatrics to 19% in Dentistry)

• Journals with a better position in the impact factor ranking have an open policy about scientific data.

• This behaviour has been also observed in previous works analysing policies of open research data in high impact factor journals (Alsheikh-Ali et al, 2011)
CONCLUSIONS

- If journals are the appropriate location for data sharing, it would be helpful to know not only what journals' policies are but also what their actual practices are.
Limitations and questions

1. Only in five subject categories

2. We do not know the real frequency of data sharing among scientists

3. If journals permit but do not require data to be shared, what percentage of authors choose to share their data?
Future research

• To analyse policies in other subject categories of JCR.

• To investigate the rate of papers in each subject category that really provide data for reuse

• To identify the nature of these data
Advancement of future research

- 697 (4.7%) of 14,911 original articles published in 39 Substance Abuse journals included in Pubmed Central repository contained supplementary material.

Journals' policies of storage and reuse of raw research data.
Advancement of future research

- The percentage of papers with supplementary material is higher in upper quartiles (7% in Q1; 4% in Q2; 3% in Q3)
Advancement of future research

- Text and .pdf files are the most frequent especially tables and figures.
- Spreadsheets and raw data (i.e., .xls or .xlsx files) represent 7.7%
Advancement of future research

41% of journals: Statement of supplementary material

5% of papers contain s.m.

8% of papers with s.m. contain raw data

Journals' policies of storage and reuse of raw research data
ACKNOWLEDGEMENTS

Supported by the National R+D+I of the Ministry of Economy and Competitiveness of the Spanish Government (projects: CSO2012-39632-C02-01 and CSO2015-65594-C2-2-R) and 2015-Networks of Excellence call (project CSO2015-71867-REDT).

UISYS (CSIC-UV)

Thanks to Antonio Vidal, Antonia Ferrer, Fernanda Peset, Andrea Sixto, Rut Lucas, Lourdes Castelló, Adolfo Alonso, Máxima Bolaños & Enrique Sánchez-Pérez