Publishing Your Research and Data

Chris Diaz
Digital Publishing Librarian
Northwestern University Libraries
chris-diaz@northwestern.edu
Digital Scholarship Services

digitalscholarship@northwestern.edu

Cunera Buys  Data Management  c-buys@northwestern.edu

Liz Hamilton  Copyright  emhamilton@northwestern.edu

Josh Honn  Digital Humanities  josh.honn@northwestern.edu

John Dorr  Department Head  john.dorr@northwestern.edu

Chris Diaz  Digital Publishing  chris-diaz@northwestern.edu
Agenda

• Scholarly Communication
  Digital Scholarship Services

• Data Management
  DMPTool

• Managing Copyrights
  Journal Publishing Agreements

• Preservation & Access
  Arch | Digital Repository
the process by which academics share their research findings with the public
the process by which academics share their research findings with the public
Scholarly Journals

• Over 80,000 active English-language scholarly peer-reviewed journals

• Over 12,000 in Journal Citation Reports, representing the “core literature” in the sciences and social sciences

• Between 2.5 and 3 million journal articles are published per year

Sources
• Ulrich’s Periodical Database
• InCites Journal Citation Reports
Scholarly Journal Market

2014: $11.5 Billion in STM Journal Subscription Sales

Five companies control 40% of the global market
1. Reed Elsevier: 17%
2. Thomson Reuters: 7%
3. Springer: 6%
4. John Wiley & Sons: 5%
5. IHS: 5%

Roughly 75% of revenue comes from academic library subscriptions

Scholarly Journal Prices

<table>
<thead>
<tr>
<th>DISCIPLINE</th>
<th>AVERAGE PRICE PER TITLE</th>
<th>DISCIPLINE</th>
<th>AVERAGE PRICE PER TITLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry</td>
<td>$5,105</td>
<td>Technology</td>
<td>$2,239</td>
</tr>
<tr>
<td>Physics</td>
<td>4,508</td>
<td>Zoology</td>
<td>2,221</td>
</tr>
<tr>
<td>Engineering</td>
<td>3,244</td>
<td>Math &amp; Computer Science</td>
<td>1,895</td>
</tr>
<tr>
<td>Biology</td>
<td>3,104</td>
<td>Health Sciences</td>
<td>1,801</td>
</tr>
<tr>
<td>Food Science</td>
<td>2,729</td>
<td>General Science</td>
<td>1,717</td>
</tr>
<tr>
<td>Astronomy</td>
<td>2,718</td>
<td>Geography</td>
<td>1,713</td>
</tr>
<tr>
<td>Botany</td>
<td>2,418</td>
<td>Agriculture</td>
<td>1,687</td>
</tr>
<tr>
<td>Geology</td>
<td>2,400</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SOURCE: LJ PERIODICALS PRICE SURVEY 2016

Data Loss in the Reference List

Paper-based reference covers assets existing in print

• journal article → journal article

Web-based referencing covers a range of assets used during the research process

• journal article → software, data sets, blogs, video, slide decks
The case of the missing data

• Survey of authors of 516 ecology papers

• Two years from publication, the chance of data accessibility falls by 17% per year

• "In theory, the data still exist, but the time and effort required by the researcher to get them to you is prohibitive."

Source:
Reference Rot

Reference rot = link rot and/or content drift

- Link rot: the contents of a referenced web page no longer exist
- Content drift: the contents of a reference web page have changed from the time at which they were first referenced
Reference Rot: One in Five Articles

3.5 million articles (1997-2012) from arxiv, Elsevier, and PubMed Central

Articles suffering reference rot:

- Articles from 2005-2012: 70-80%
- Articles from 2009-2012: 20%
Data Publishing

Driven by funder and publication policies...
Federal Funding in the U.S.

Federal Research Funding Obligations for FY2017: $145.4 Billion

Down from $160 Billion in 2010

Top Departments (in Billions):
- Department of Defense: $69.1
- Health and Human Services: $32.7
  - NIH: $31.4
- Energy: $14.3
- NASA: $12.9
- NSF: $6.5

Source
- NSF. Survey of Federal Funds for Research and Development, FY 2015-17
What is/are Data?

• NSF: “Research data means the recorded factual material commonly accepted in the scientific community as necessary to validate research findings” [2 CFR § 200.315(3)]

• NIH: “Recorded factual material commonly accepted in the scientific community as necessary to document and support research findings” [Data Sharing Guidance | NIH]
Culture of Sharing Data

• Before 2010: 9% of papers from the top 50 journals published data sets in support of research articles

• Between 2011 and 2015: 400% increase in number of data sets published online in data repositories

• “the static PDF as a primary format for scholarly communication will be superseded by formats that can do better justice to the needs of the researcher” [eLIFE]
“Innovation points out paths that are possible; replication points out paths that are likely; progress relies on both.”

– Open Science Collaboration (2015)
Peer Review

Journal peer review:
• Is the research within the journal’s aims & scope?
• Is the research methodologically sound?
• Is the research important?
• Is the research complete?

Data peer review [SpringerNature]:
• Are the data in a public repository?
• Are the data presented in a useful file format?
• Are the data described with useful metadata?
• Are the data methodologically sound?
• Are the data free from personally identifiable, sensitive, or inappropriate information?

Further reading on peer-review practices for publishing research data: https://arxiv.org/abs/1704.02236
Data Management
Data Sharing Policy Examples

Funders

- National Institutes of Health
- National Science Foundation
- Gates Foundation
- Arnold Foundation
- Wellcome Trust

Browse Article and Data Sharing Requirements by Federal Agency [SPARC]

Publishers

- American Economic Association
- Public Library of Science (PLOS)
- SpringerNature / BioMedCentral
- Science
- eLIFE

Browse Journal open data policies by Title [OAD | Simmons]
Data Management Planning

Benefits

• Good practice: transparency, accountability, reproducibility
• Competitive edge: data citations, larger contribution to the field
• Prevents data loss

It’s probably required.
Data Management Planning

Data management plans are a standard piece in research grant proposals. They document what researchers will do with data they collect during and after the project, such as:

- the **types** of data that will be created
- the **standards and metadata** that will be used with the data
- how these data will be **accessed and shared**
- policies and provisions for the **reuse** of the data
- plans for the **long term archiving** of the data
Metadata

Helps make your data set more discoverable online and understandable by other researchers

Technical documentation, codebook

Provides context
  • Means of creation of the data
  • Purpose of the data
  • Time and date of creation
  • Creator or author of the data
  • Standards used
The project will leverage existing metadata standards currently stored in Ecological Metadata Language (EML) format for the NutNet project. We will add additional metadata entries for the arthropod community composition and arthropod stoichiometry; field notes taken during the time of collection will be recorded. Morpho software will be used to generate the metadata file in EML. We chose EML format for our metadata since it allows integration with existing NutNet data housed in the Knowledge Network for Biocomplexity (KNB) data repository.

Source: [Arthropod Responses to Grassland Nutrient Limitation](http://example.com)
Metadata: Resources

• Social Science: ICPSR Guide to Social Science Data Preparation and Archiving

• Earth and Environmental Sciences: DataONE Best Practices

• Clinical Data: Preparing raw clinical data for publication

• Human Subjects: DRYAD FAQ

Check Journal Publisher or research funder policies for guidance
A free online DMP generator that uses funder and institution-specific templates for private foundations and government agencies.

Each template draws on funder requirements for DMPs.

Online library of real DMPs shared by researchers.

Website: [http://dmptool.org](http://dmptool.org)

Library Support: [http://libguides.northwestern.edu/datamanagement](http://libguides.northwestern.edu/datamanagement)
A free online DMP generator that uses funder and institution-specific templates for private foundations and government agencies.

Each template draws on funder requirements for DMPs.

Online library of real DMPs shared by researchers.

Website: [http://dmptool.org](http://dmptool.org)

Library Support: [http://libguides.northwestern.edu/datamanagement](http://libguides.northwestern.edu/datamanagement)
Copyright

Understanding your rights as an author
What does copyright protect?

“original works of authorship fixed in any tangible medium of expression” -- 17 USC § 102 (a).

Copyright starts when work is fixed!

Notice (Copyright © 2016 by Chris Diaz) and registration not required, but it is a clearer statement of ownership.

Most foreign works (published outside the U.S.) receive the same protection in the U.S. as works published in the U.S.
What does copyright protect?

• to reproduce the copyrighted work in copies or phonorecords;
• to prepare derivative works based upon the copyrighted work;
• to distribute copies or phonorecords of the copyrighted work to the public by sale or other transfer of ownership, or by rental, lease, or lending;

... 

-- 17 USC § 106.
Copyright Policy

Northwestern University upholds the principle that our faculty retain copyright ownership for traditional works created by the faculty, such as books, films, musical scores, and other works of art. A group of faculty from throughout the University community created the copyright policy with that principle in mind. In addition to traditional works, the creators of the policy provided for the University to share in the benefits of copyrightable software when the University provides extraordinary resource support for the development of the software and when research sponsors assign the copyright to the University.

Source: https://invo.northwestern.edu/policies/copyright-policy
Publisher-Author Agreements

- Transfer of all rights in perpetuity
- Exclusive license for distribution
- License of certain rights on a nonexclusive basis
- Self-archiving restrictions
- Divvying of rights by manuscript version
- Embargo, i.e., you have to wait X months before you can use the publisher PDF
- You can participate in our open access program if you pay an additional author fee

NOTE: If you transfer all rights to a publisher, you may have to ask permission to use your own work!
SHERPA/RoMEO: Policy Lookup Tool

Database that tracks self-archiving policies for over 2,300 scholarly publishers

http://www.sherpa.ac.uk/romeo/index.php

<table>
<thead>
<tr>
<th>ROMEO colour</th>
<th>Archiving policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>green</td>
<td>can archive pre-print and post-print or publisher’s version/PDF</td>
</tr>
<tr>
<td>blue</td>
<td>can archive post-print (ie final draft post-refereeing) or publisher’s version/PDF</td>
</tr>
<tr>
<td>yellow</td>
<td>can archive pre-print (ie pre-refereeing)</td>
</tr>
<tr>
<td>white</td>
<td>archiving not formally supported</td>
</tr>
</tbody>
</table>

80% of publishers allow some form of self-archiving research papers
Elsevier: Copyright Policy

Rights granted to Elsevier

For both subscription and open access articles, published in proprietary titles, Elsevier is granted the following rights:

- The exclusive right to publish and distribute an article, and to grant rights to others, including for commercial purposes.
- For open access articles, Elsevier will apply the relevant third party user license where Elsevier publishes the article on its online platforms.
- The right to provide the article in all forms and media so the article can be used on the latest technology even after publication.
- The authority to enforce the rights in the article, on behalf of an author, against third parties, for example in the case of plagiarism or copyright infringement.

Source: https://www.elsevier.com/about/our-business/policies/copyright
Manuscript Versions

Pre-print/ Draft
• The manuscript as it is submitted to the journal

Post-print/ Author’s Accepted Manuscript
• The manuscript after revisions have been incorporated as the result of peer-review

Publisher’s version/PDF
• The version of record, typeset, found in databases and publisher’s website
Elsevier: Article Sharing Policy

Preprint

- Authors can share their preprint anywhere at any time.
- If accepted for publication, we encourage authors to link from the preprint to their formal publication via its Digital Object Identifier (DOI). Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
- Authors can update their preprints on arXiv or RePEc with their accepted manuscript.

Source: https://www.elsevier.com/about/our-business/policies/sharing
Accepted Manuscript
Authors can share their accepted manuscript:

Immediately

- via their non-commercial personal homepage or blog
- by updating a preprint in arXiv or RePEc with the accepted manuscript
- via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
- directly by providing copies to their students or to research collaborators for their personal use
- for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement

Source: https://www.elsevier.com/about/our-business/policies/sharing
Elsevier: Article Sharing Policy

Accepted Manuscript
Authors can share their accepted manuscript:

After the embargo period

- via non-commercial hosting platforms such as their institutional repository
- via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license – this is easy to do, click here to find out how
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy
- not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article

Source: https://www.elsevier.com/about/our-business/policies/sharing
Published Journal Article

Policies for sharing published journal articles differ for subscription and gold open access articles:

Subscription articles

- If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
- Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
- If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others’ research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.
- Otherwise sharing is by agreement only.

Source: https://www.elsevier.com/about/our-business/policies/sharing
Retain Control of Your Copyright

You should retain your copyright, but license the use of your work to publishers.

You have options before you sign that agreement, not afterwards.

Author has non-exclusive rights to his/her work for academic purposes.

After 6 months, can make full use of publisher's copy.

Author has right to grant employing institution rights of reproduction, distribution, display, etc.
Author Addenda

Model language you can modify and attach to publishing agreements

- Right to use, reproduce, distribute the work
- Right to post the PDF online in a digital repository after 6 months
- Right to grant university non-exclusive rights to use, reproduce, distribute, and archive for academic purposes

B1G Academic Alliance Statement on Publishing Agreements

Scholarly Publishing and Academic Resources Coalition
Open Access Fund

The Northwestern Open Access Fund supports Northwestern scholars who wish to make their journal articles openly available immediately upon publication and to support gold open access publishers around the world. The gold open access journal model publishes all of a journal's articles as open access and allows for freely available access immediately upon publication.

• Up to $3,000 toward an Article-processing Change (APC) for an Open Access journal

• Faculty, postdocs, graduate students, staff

http://www.library.northwestern.edu/research/scholarly/open-access-fund.html
Research and Data Archiving

ARCH | Northwestern University’s Institutional Repository
Content

- Pre-prints, scientific and technical reports, journal articles, white papers
- Conference presentations, posters, audio/video recordings
- Data sets supporting publications
- Images, figures, code
The numerology of gender: gendered perceptions of even and odd numbers

Do numbers have gender? Wilkie and Bodenhausen (2012) examined this issue in a series of experiments on perceived gender. They examined the perceived gender of baby faces and foreign names. Arbitrary numbers presented with these faces and names influenced their perceived gender. Specifically, odd numbers connoted masculinity, while even numbers connoted femininity. In two new studies (total N = 316), we further examined the gendering of numbers. The first study examined explicit ratings of 1-digit numbers. We confirmed that odd numbers seemed masculine while even numbers seemed feminine. Although both men and women showed this pattern, it was more pronounced among women. We also examined whether this pattern holds for automatic as well as deliberate reactions. Results of an Implicit Association Test showed that it did, but only among the women. The implicit and explicit patterns of numerical gender ascription were moderately correlated. The second study examined explicit perceptions of 2-digit numbers. Again, women viewed odd numbers as more masculine and less feminine than even numbers. However, men viewed 2-digit numbers as relatively masculine, regardless of whether they were even or odd. These results indicate that women and men impute gender to numbers in different ways and to different extents. We discuss possible implications for understanding how people relate to and are influenced by numbers in a variety of real-life contexts.

Features

- Compliant: Meets funding agency preservation and public access requirements
- Discovery: Shares metadata with academic search engines and indexes
Benefits

• **Free**: Arch is free of charge for Northwestern faculty and research staff
• **Flexible**: Control visibility by setting future release dates for online access
• **Open Access**: Share your research with the world