The Version of Record of this manuscript has been published and is available in the *Journal of Agricultural & Food Information*, 17(1), 2016. Available at http://www.tandfonline.com/ with the DOI: 10.1080/10496505.2015.1120213

Practices, Policies, and Persistence: A Study of Supplementary Materials in Crop Science Journals

Sarah C. Williams
University of Illinois at Urbana-Champaign

Abstract

This study compared practices and policies of 24 crop science journals to selected NISO/NFAIS recommendations for online supplemental journal article materials. The studied recommendations include the display of supplementary materials, DOIs for supplementary materials, and clear preservation statements regarding supplementary materials. This study also investigated missing supplementary materials on 18 of the journal websites. The findings reveal some potential roles for librarians and libraries, especially those with institutional repositories, which could better facilitate long-term access, data citation, and data reuse.

Introduction

Supplementary materials are additional files (e.g., text, figures, data, multimedia, code), which support or enhance a published journal article. Supplementary materials are an important element of scientific scholarly communication and data sharing. Many journals have an increasing percentage of articles with supplementary materials, in some cases over 90% in recent years (Kenyon & Sprague, 2014). Researchers also frequently mention traditional scholarly outputs as a data sharing method in data management plans (Mischo, Schlembach & O'Donnell, 2014), and while Mischo et al. (2014) did not specifically mention supplementary materials, the

author of this study regularly sees supplementary materials mentioned alongside publications in data management plans. Particularly in crop sciences and more broadly in the agricultural sciences, publications and supplementary materials are typical avenues for data sharing (Diekmann, 2012; Williams, 2012).

Supplementary materials are not without complications, however. Standards for data organization and file formats are lacking, even for similar types of information, which can make reuse difficult (Santos, Blake & States, 2005). Broken links to supplementary materials can make files inaccessible (Anderson, Tarczy-Hornoch, & Bumgarner, 2006; Evangelou, Trikalinos, & Ioannidis, 2005). In other cases, articles mention supplementary materials, but the files cannot be found on the journal website (Anderson et al., 2006; Williams, 2012). Long-term accessibility, preservation, and archiving of supplementary materials also are of concern (Laue, 2010; Rosenthal & Reich, 2010; Schwarzman, 2010).

Efforts to improve the untenable situation of supplementary materials are ongoing. One notable example is the work of the NISO/NFAIS Working Group on Supplemental Journal Materials, specifically *Recommended Practices for Online Supplemental Journal Article Materials* (NISO/NFAIS, 2013). The document provides recommendations for business policies and practices and for technical considerations and implementations for supplementary materials, covering issues such as discoverability and findability, persistent identifiers, and preservation. Roles and responsibilities for the primary publishing parties and related parties are also outlined. For example, publishers should be clear about the level of delivery and preservation provided, editors should encourage authors to deposit non-integral supplementary materials in repositories that provide preservation assurances and bidirectional linking, and libraries could function as repositories for the research output of their institutions.

Given the importance of supplementary materials in agricultural research publications, this study focuses on journals in which crop scientists frequently publish, including crop science journals and broader science journals. One purpose of the current study is to compare current journal policies and practices to selected areas of the guidance in the NISO/NFAIS (2013) recommended practices document, including the display of supplementary materials, DOIs for supplementary materials, and clear preservation statements regarding supplementary materials. The study also investigates the issue of missing supplementary materials that was noted in an earlier, smaller-scale study of crop sciences publications (Williams, 2012). Based on these findings, the author recommends roles for librarians and libraries related to supplementary materials.

Literature Review

It is important for librarians to be aware of supplementary material trends and practices and to consider what roles librarians and libraries can play. Kenyon and Sprague (2014) studied 12 years of published material in 60 key environmental science journals, finding a substantial rise in the percentage of articles with supplementary materials - in 2011, a median percentage greater than 40% for biology, plant science, and geology journals. They also reviewed publisher and journal policies related to supplementary materials, including peer review, metadata, file formats, file sizes, access issues, and long-term archiving. As subject librarians, they noted that the findings informed their efforts to advise researchers on scholarly communications issues. In 2011, Lorbeer and Klusendorf considered the indications that some journals would no longer accept supplementary materials and suggested that libraries and their institutional repositories would be good mechanisms for disseminating and preserving supplementary materials.

Schaffer and Jackson's study (2004) of 94 science and engineering journals involved reviewing author instructions regarding supplementary materials, checking recent issues for supplementary materials, and contacting select publishers about whether they permitted supplementary materials. While 69% of the journals permitted supplementary materials, less than half appeared to have supplementary materials in recent issues. The article also made recommendations to improve the accessibility of supplementary materials. Notably, some of their recommendations from 2004 are similar to the NISO/NFAIS (2013) recommendations, such as the indication of supplementary material in tables of contents, linking between articles and supplementary materials, and clear policies about archiving supplementary materials.

One challenge of particular relevance to this study is the availability of supplementary materials. Anderson et al. (2006) used two methods to explore the persistence of all supplementary materials links in biomedical journals (i.e., links internal and external to the journal website). With the first method, they manually checked 655 supplementary material links from PubMed abstracts between 1998 and 2005 and found about 26% of the manuscripts had inaccessible links, with 7% of the inaccessible links internal to the journal website. With the second method, they manually checked links in 162 full-text manuscripts published in 2004 in three high-impact journals. They found that 17% of the links were inaccessible approximately one year after publication and 45% of those inaccessible links were internal to the journal website. One example of this was the journal *Genetics*, in which some articles said that supplementary data was available at genetics.org, but the data could not be found. The article concluded with five recommendations to improve the persistence of supplementary materials, including stronger, more consistent journal policies and support for systems designed for long-term preservation. A study by Evangelou et al. (2005) provides additional evidence that broken

links to supplementary materials are a problem. They examined supplementary material links in 244 articles from six top-cited science journals in 2000 and 2003. Using link-checking software, they found 15 articles with 21 broken links; 14 of the broken links pointed to external websites.

Methodology

This study was designed to focus on journals in which crop scientists frequently publish and a variety of publishers. The author considered using a combination of the top 10 journals in the *Journal Citation Reports* categories: agronomy, horticulture, plant sciences, and soil science, but many core titles (e.g. *Crop Science*, *HortScience*, *Journal of Economic Entomology*, *Soil Science Society of America Journal*, *Weed Science*) would be excluded. Another consideration was to use a combination of top-ranked core journals from *The Literature of Crop Science* (Olsen, 1995) and *The Literature of Soil Science* (McDonald, 1994), but some newer or interdisciplinary titles (e.g., *GCB Bioenergy*, *PLOS ONE*) would not be included. To accommodate the interdisciplinary nature of crop sciences research, the author used journals reported in an earlier study of data practices in crop sciences faculty publications (Williams, 2012). Journals in this newer study included only those wherein University of Illinois Crop Sciences faculty members were published at least twice - a study set of 24 journals from a variety of publishers (Table 1).

Each journal was reviewed, in particular to compare current journal practices and policies to selected recommendations in the NISO/NFAIS (2013) report. Issues and articles were browsed to study these recommended practices:

- provide consistent presentation in tables of contents, display in the article, and display in the supplementary materials
- use consistent naming conventions for supplementary materials

- provide context for supplementary materials
- provide DOIs for supplementary materials separate from that of the article

Journal and publisher policies and webpages were searched for any mention of archiving or preservation of supplementary materials, and author instructions were reviewed for recommendations as to file format, file size, and data deposit for supplementary materials. Journal-specific term(s) for supplementary materials (e.g., supplemental material, supporting information, e-Xtra) were recorded to facilitate the search for missing supplementary materials.

Missing supplementary materials were defined as those mentioned in an article as available on the journal website but not actually found there. This definition did not include broken links to supplementary materials (on the journal website or an external website). Link rot has been studied previously (Anderson et al., 2006; Evangelou et al., 2005), so the author focused specifically on supplementary materials missing on journal websites, in part because the issue surfaced in an earlier, smaller-scale study (Williams, 2012).

To investigate missing supplementary materials, journal issues from 2013, 2011, 2009, 2007 and 2005 were reviewed on the journal websites by the author and a research assistant. For each issue, the information recorded included:

- volume
- issue (when available)
- year
- the total number of articles (excluding brief items, e.g., letters to the editor, introductions)
- the number of articles with supplementary materials
- the number of articles with missing supplementary materials

To determine the number of articles with supplementary materials, the full-text of each article was searched for the supplementary material term(s) used by that journal. If an article indicated that supplementary material was available on the journal website, it was counted as such, and the supplementary materials mentioned in the article were compared against the supplementary materials available on the journal website. The searchers were careful about false positives, because many crop science articles include variations of the word supplement to refer to supplemental irrigation, fertilizer, etc. When an article was missing supplementary material, the article title, first author, DOI, volume and issue ID number, and any notes were recorded.

After reviewing 18 of the 24 journals (75%), it was clear that missing supplementary materials were not a common problem with the journals in this sample. Continuing the search and adding hundreds of articles to the dataset seemed to follow the law of diminishing returns, so the author decided to stop the search for missing supplementary materials. Table 1 indicates the 18 journals that were examined.

All data for both study components (i.e., comparisons to recommended practices and missing supplementary materials) were recorded in a Microsoft Access database designed by the author. The data were collected from June 2014 through June 2015.

Results

Supplementary Material Practices and Policies

The NISO/NFAIS (2013) report recommended consistent presentation in three areas: tables of contents, display in the article, and display in the supplementary materials. Many (18) of the 24 journals followed the recommendation to indicate the presence of supplementary materials on the

table of contents page. The two observed methods were links to individual supplementary material files or links to supplementary material landing pages.

The recommendations did not make a distinction between display in HTML and PDF articles. Yet in practice, journals have different displays for these formats, so this study recorded the display in the HTML article separately from the display in the PDF article. For HTML articles, most journals met the NISO/NFAIS recommendation of placing supplementary material links near the top of the page, although some of these links were hidden because article navigation options were collapsed by default. Supplementary material links were most commonly in a left or right navigation column. Other implementations included links at the top in the center, supplementary material tabs, and supplementary material sections before acknowledgement sections. Several journals used a combination of these implementations.

The findings for PDF articles were mixed. Only nine journals indicated supplementary materials on the first page, which included a logo in the top corner, a note in the author information section, a note with the article citation and DOI, and a note above the abstract. More commonly, PDF articles had a supplementary materials section near the end (e.g., near the acknowledgements and references). As with HTML articles, some journals used a combination of these indicators. With three journals (*HortScience*, *Weed Science* and *Weed Technology*), supplementary materials were typically only mentioned within the text of the PDF article.

Regarding display within the supplementary materials, the NISO/NFAIS (2013) report recommended navigational elements to match the elements available at the article level. Differences in supplementary material access methods made this complicated to assess, but most journals met this recommendation. Links to supplementary materials were usually on the HTML article page or a supplementary materials landing page or tab, and these typically shared

navigational elements with the article. A different, but related, recommendation was that supplementary material files should include contextual information (e.g., article citation and DOI). Of the limited downloads checked, very few referenced the article. Two exceptions were the *Journal of Environmental Quality* and the *Journal of Experimental Botany*. The author instructions for both of these journals stated that supplementary material files should include the article title and authors, and in most cases they did.

The NISO/NFAIS (2013) report also recommended that publishers use consistent naming conventions for supplementary materials in the tables of contents and in the articles. While different publishers used a variety of terms, they were usually consistent within a journal and sometimes across journals. The two Elsevier journals in this study (*Biomass & Bioenergy* and *Crop Protection*) were exceptions. For example, *Biomass & Bioenergy* used "Supplementary Content" in the tables of contents, "Supplementary Material" in the appendix subtitle, and "Supplementary Data" in the appendix description.

Journal and publisher policies and webpages infrequently addressed archiving or preservation of supplementary materials. Only 5 of the 24 journals, from two publishers/platforms – Wiley-Blackwell and American Chemical Society (ACS)--specifically mentioned supplementary materials. The Wiley-Blackwell long-term preservation policy noted that the publisher currently archives some of the data that supports articles and is participating in discussions concerning a longer-term plan. An ACS Legacy Archives page acknowledged the value of supporting information and stated that supporting information from 1970-1995 would be added to the ACS Legacy Archives. Of the remaining 19 journals, 15 indicated participation in CLOCKSS, LOCKSS, Portico, or independent archives but did not specifically mention supplementary materials. Four journals did not appear to address archiving or preservation at all.

Assigning DOIs for supplementary materials is another aspect of supporting long-term access and is recommended in the NISO/NFAIS report. Only three journals in this study (*PLOS ONE*, *Weed Science* and *Weed Technology*) assigned separate DOIs for each supplementary material file.

This study also reviewed author instructions for guidance on supplementary materials and data that might affect the library's role with these materials. Figure 1 shows the most commonly recommended or accepted file formats for supplementary material. Of the 24 journals, 9 provided no guidance on file formats. Two journals (*BMC Plant Biology* and *Plant Physiology*) recommended CSV and XLS files for data, which facilitates data reuse, but two other journals (*Agronomy Journal* and *Journal of Experimental Botany*) recommended only PDF and image/audio/video file formats. Another journal (*Soil Science*) recommended converting all non-audio/video files to PDF but did accept other formats. Guidance also varied on file size limits for supplementary materials. Eleven journals provided no guidance. Seven journals limited individual files to 5 MB, 10 MB or 20 MB. Two journals (*Biomass & Bioenergy* and *Crop Protection*) gave different limits for multimedia files (50 MB) and figures (7-10 MB). One journal (*Molecular Biology and Evolution*) limited supplementary materials to 2 MB/file and limited articles to five supplementary material files.

The author instructions for 50% of the journals included expectations or recommendations for sharing data external to the journal website. These journals required or encouraged data to be deposited in public repositories. Specific repositories mentioned include: GenBank (http://www.ncbi.nlm.nih.gov/genbank/), PANGAEA (http://www.pangaea.de/), TAIR (https://www.arabidopsis.org/), and TreeBASE (http://treebase.org/treebase-web/). Institutional repositories were not specifically mentioned, but the two Wiley-Blackwell journals (*GCB*

Bioenergy and *Plant, Cell and Environment*) mentioned that supplementary materials could be displayed on an author's or institutional website, in addition to the journal website. Four journals prohibited supplementary material links to personal, departmental or non-repository websites.

Supplementary Material Persistence

In this study, missing supplementary materials were defined as those mentioned in an article as available on the journal website but not actually found on the journal website, not including broken links to supplementary materials. The major finding was that this type of missing supplementary material was not a common problem on the journal websites in this sample. The sample included 18 journals (Table 1) with 820 issues (publication years: 2013, 2011, 2009, 2007 and 2005) and 17,099 total articles. Of the total articles, 4,776 articles had supplementary materials, and only 56 of those articles (1.2%) were missing supplementary materials on the journal website. Thirteen of the 18 journals did not have any missing supplementary materials in the issues searched, and 1 journal (*Nematropica*) did not appear to have supplementary materials, so this study identified 4 journals with missing supplementary materials.

In analyzing the prevalence of missing supplementary materials, the four journals could be grouped into two categories. In one category, two journals (*Agronomy Journal* and *HortScience*) had only a few articles with missing supplementary materials, but they had few articles overall that included supplementary materials, so their percentages of missing supplementary materials were higher. In the second category, two other journals (*Molecular Biology and Evolution* and *Plant Physiology*) had several articles with missing supplementary materials, but they had hundreds of articles with supplementary materials, so their percentages of missing supplementary materials were lower. In some cases, only one or two supplementary

files were missing for an article with multiple supplementary files; in other cases, none of the supplementary materials were available (often the case with *Plant Physiology*, where the sole link in the supplementary materials section went to the journal homepage).

Evaluating the data over time, most of the missing supplementary materials were from 2005 and 2007, even though those years had the fewest number of articles with supplementary materials. There were 13 articles missing supplementary materials in 2005 (3.5%), 34 in 2007 (5.1%), 3 in 2009 (0.3%), 6 in 2011 (0.4%), and 0 in 2013.

In the course of this research, this study also determined the percentage of articles with supplementary materials for the 18 journals (see Appendix), which tracks closely to the data shared by Kenyon and Sprague (2014). Kenyon and Sprague (2014) found that biology, plant science, and geology journals in 2011 had a median percentage greater than 40%, while the journals in this study had a median percentage greater than 10% in 2011 and greater than 20% in 2013. Two journals overlapped between the two studies – *Plant Physiology* and *Journal of Experimental Botany*--and the percentages, and certainly the trends, are similar in the two studies. For *Plant Physiology*, Kenyon and Sprague recorded 34.25% of articles with supplementary materials in 2005 (compared to 34.8% in this study) and 86.78% in 2011 (compared to 91.7% in this study), which is a 153% increase (compared to 164%). For the *Journal of Experimental Botany*, Kenyon and Sprague recorded 8.20% of articles with supplementary materials in 2005 (compared to 8.5% in this study) and 57.89% in 2011 (compared to 65.9% in this study), which is a 606% increase (compared to 675%). The differences in percentages are due primarily to the different total number of articles used. This study counted the number of articles in each issue (excluding brief items), while Kenyon and

Sprague (2014) used the *Journal Citation Reports*' number of citable items. These numbers varied more widely in some years.

Discussion

The comparison of journal practices and policies to the NISO/NFAIS (2013) supplementary material recommendations found a mix of results. Regarding the presentation of supplementary materials, most journals met the recommendations in tables of contents and HTML articles. While supplementary materials were usually consistently indicated in PDF articles, at least within a journal, the indication was rarely on the first page, which did not follow the recommendation. The navigational elements within the supplementary materials typically matched the elements available in the HTML article, but regarding contextual information, the supplementary material files rarely referenced the article, in the limited downloads checked. These opportunities for improvement fall mainly within the responsibility of journals and publishers, although librarians could certainly encourage researchers to reference their articles within supplementary material files, even if the author instructions do not require that.

Other findings in this study, especially those related to long-term access and preservation, could present greater opportunities for libraries and librarians. Archiving or preservation of supplementary materials was rarely directly addressed in journal or publisher policies, echoing the findings of Kenyon and Sprague (2014). Institutional repositories or institutional data repositories, which typically have clear access and preservation policies, could be an alternative to journal websites for disseminating and preserving supplementary materials. Lorbeer and Klusendorf (2011) suggested, during a time of uncertainty about publishers' willingness to accept supplementary materials, that supplementary materials could be deposited in institutional

repositories. Although most publishers currently accept supplementary materials (and often at an increasing rate), institutional repositories could still be promoted for long-term access and preservation. This idea is also supported by the NISO/NFAIS (2013) report, which outlined one role for libraries as a repository for the research output of the institution.

Institutional repositories can provide other long-term access solutions as well. Few journals provided separate DOIs for supplementary materials, so if researchers need persistent identifiers for their documents or data, they could deposit them in institutional repositories, which typically assign a persistent identifier. This could become particularly important, as data citation becomes a more common practice. Persistent identifiers from institutional repositories could also be promoted as a solution for journals that do not allow supplementary material links to personal, departmental, or non-repository websites.

This study also revealed data reuse limitations for some supplementary materials on journal websites. The PDF file format was one of the most frequently recommended or accepted file formats for supplementary materials. Two journals recommended only PDF and image/audio/video file formats, and another journal recommended converting all non-audio/video files to PDF but did accept other formats. Since data shared in PDF files is a barrier to reuse, an alternative could be to deposit a CSV file in an institutional repository. Data disseminated via an institutional repository would also address the point by Herold (2015) that publishers largely control access to data when it is shared as supplementary materials on journal websites. Some publishers only allow journal subscribers to access supplementary materials, while other publishers make them openly available, either immediately or after an embargo period (Herold, 2015).

The search for missing supplementary materials grew out of an earlier, smaller-scale study of crop sciences publications (Williams, 2012), in which there were 29 articles that mentioned supplementary materials on the journal website but, in two cases (7%), they could not be found. To match the earlier study, this study defined missing supplementary materials as those mentioned in an article as available on the journal website but not actually found on the journal website. In this study, 4,776 articles had supplementary materials, and of those, 56 articles (1.2%) were missing some or all of their supplementary material files. This study did not explore broken links to supplementary materials on journal websites, but this has been explored in other studies (Anderson et al., 2006; Evangelou et al., 2005), which found fewer broken links to supplementary materials on journal websites than on external websites.

While the overall percentage of missing supplementary materials was quite low for this study, missing supplementary materials could be an issue with some publishers or journals, but no generalizable conclusions can be drawn about the types of publishers or journals.

Conclusion

Supplementary materials and publications are core components of scholarly communication and data sharing in agricultural sciences, and researchers frequently mention traditional scholarly outputs as a data sharing method in data management plans. This study compared crop science journal supplementary material practices and policies to selected NISO/NFAIS (2013) recommendations and investigated missing supplementary materials on journal websites. The findings reveal some potential opportunities for librarians and libraries to engage with crop scientists on scholarly communication and data sharing issues.

The lower percentage of missing supplementary materials in this study, as compared to an earlier study (Williams, 2012), reduces the impetus for these materials to be deposited into institutional repositories. Nevertheless, librarians and researchers should still consider this option, especially if long-term access, data citation, or data reuse are high priorities. Journal or publisher policies rarely directly addressed archiving or preservation of supplementary materials, and in this study the highest percentages of missing supplementary materials were in 2005 and 2007. These findings highlight potential concerns about long-term access to supplementary materials, which institutional repositories with clear preservation policies may be able to address. Since few journals provide separate DOIs for supplementary materials, institutional repositories might better support data citation by providing persistent identifiers for supplementary materials. In the case of journals that restrict supplementary materials to PDF files, institutional repositories might also better facilitate data reuse by accepting data in a variety of file formats. These are important points for librarians to consider when advising researchers on scholarly communications issues and data management plans.

References

Anderson, N., R., Tarczy-Hornoch, P., & Bumgarner, R. E. (2006). On the persistence of supplementary resources in biomedical publications. *BMC Bioinformatics, 7*, Article 260. Retrieved from http://www.biomedcentral.com/1471-2105/7/260

Diekmann, F. (2012). Data practices of agricultural scientists: Results from an exploratory study. *Journal of Agricultural & Food Information, 13*(1), 14-34.

Evangelou, E., Trikalinos, T. A., & Ioannidis, J. P. A. (2005). Unavailability of online supplementary scientific information from articles published in major journals. *The FASEB Journal, 19,* 1943-1944.

Herold, P. (2015). Data sharing among ecology, evolution, and natural resources scientists: An analysis of selected publications. *Journal of Librarianship and Scholarly Communication, 3*(2), eP1244. Retrieved from http://doi.org/10.7710/2162-3309.1244

Kenyon, J., & Sprague, N. R. (2014). Trends in the use of supplementary materials in environmental science journals. *Issues in Science and Technology Librarianship, 75*. doi:10.5062/F40Z717Z

Laue, A. (2010). Hosting supplementary material: Technical challenges and suggested best practices. *Information Standards Quarterly, 22*(3), 10-15.

Lorbeer, E. R., & Klusendorf, H. (2011). Open and accessible supplemental data: How librarians can solve the supplemental data arms race. *Against the Grain, 22*(6), 42-43.

McDonald, P. (Ed). (1994). *The literature of soil science*. Ithaca, NY: Cornell University Press.

Mischo, W. H., Schlembach, M. C., & O'Donnell, M. N. (2014). An analysis of data management plans in University of Illinois National Science Foundation grant proposals.

Journal of eScience Librarianship, *3*(1), Article 3. http://dx.doi.org/10.7191/jeslib.2014.1060

NISO/NFAIS. (2013). *Recommended practices for online supplemental journal article materials* (NISO RP-15-2013). Retrieved August 28, 2015 from http://www.niso.org/publications/rp/rp-15-2013.

Olsen, W. C. (Ed.). (1995). *The literature of crop science*. Ithaca, NY: Cornell University Press.

Rosenthal, D. S. H., & Reich, V. A. (2010). Archiving supplemental materials. *Information Standards Quarterly, 22*(3), 16-21.

Santos, C., Blake, J., & States, D. J. (2005). Supplementary data need to be kept in public repositories. *Nature, 438*(8), 738.

Schaffer, T., & Jackson, K. M. (2004). The use of online supplementary material in high-impact scientific journals. *Science & Technology Libraries, 25*(1-2), 73-85.

Schwarzman, S. (2010). Supplemental materials survey. *Information Standards Quarterly*, *22*(3), 23-26.

Williams, S. C. (2012). Data practices in the crop sciences: A review of selected faculty publications. *Journal of Agricultural & Food Information, 13*(4), 308-325.