PUBLISHER ENGAGEMENT/DEVELOPMENT REPORT: 2005—April 2006

The 14-month period from January 2005 to the present has seen the coming-of-age of Columbia CI's initial work encouraging publishers to contribute metadata to NSDL. This has come about in 2005 through significant improvement of our local, phase-one technical capability to process, convey, and/or create publisher metadata; expansion of our organizational partners beyond academic and research publishers to include educational software firms, establishment of important contractual protocols and their attending language in conjunction with our CI colleagues at Cornell and UCAR, launch of local, enterprise-scale technical capacity to handle large collections, and the beginnings of original editorial development of publisher content through the Library.

My present sense of Columbia CI's position in terms of continued work with publishers is quite optimistic, particularly in reference to what I think of as the baseline publisher-engagement activity: convincing publishers, broadly defined, to willingly participate in making their metadata available to NSDL. Nearly all the organizations I approach respond to my solicitation and a significant number go on to enter into discussions with NSDL (vide infra). These discussions concern technical and legal/logistical issues connected with moving raw bibliographic or product-description data in various formats from the content creators' or aggregators' servers to the NDR--going through the Columbia technical facility for processing and conveyance to Ithaca, if necessary.

A highly significant challenge we have overcome in the engagement/accession process is our current technical capability to work with data sets in almost any consistent, orderly format and the new server's capacity to handle very large collections of metadata (e.g. from journals publishers, many with hundreds of serial titles that each publish dozens to hundreds of articles per year.)

An intermittent challenge that remains is publishers' occasional articulation of unusual, unexpected, or logistically overwhelming conditions over the use of their data. In the area of these lingering challenges is where some patterns, which we must anticipate, for future relationship lie. Important examples here are publishers demanding that we restrict access of their contributed metadata in certain ways and the possible need to obtain permission from multiple agents when approaching aggregators.

> * Phase-one publisher metadata work: This capability came at the very end of 2004 with the creation of the first EPIC/NSDL server at http://grackle.cc.columbia.edu/cwis/. "Grackle" employs the NSDL service project software, CWIS. During 2005, this facility has been fine-tuned and now serves publisher metadata via OAI-PMH v. 2 to the main NSDL metadata repository, now the NDR, for numerous partner organizations based on the Dublin Core element set-Spec. "Publisher." Currently metadata for five partner organizations are processed and transmitted to NSDL using grackle: Cambridge University Press (books), John Wiley (books), Prentice Hall (text companion websites), Houghton-Mifflin (web-based educational software), Tool Factory, (educational software). An additional improvement in phase-one metadata work is a streamlined

system to convert the publishing standard ONIX to Dublin Core.

*Organizational partners: Along with working with academic and research publishers to create records in NSDL that describe and locate books and articles in journals and other serials, in 2005 EPIC began work with educational software firms. Presently, we have processed metadata for two new media firms, Tool Factory Educational software, http://www.toolfactory.com/ and Tom Snyder Productions, http://www.tomsnyder.com/, a software division of Scholastic, and are in discussion with a third company, Riverdeep Interactive.

*Contract protocols: In 2005, two publishers, one under contract and the other prospective, requested provisions and contractual language for possible scenarios not addressed in the boiler plate letter of agreement or expressed previously by any other publishers. These included restriction of further OAI harvest from NSDL by third parties and differentiation between [the] publishers' metadata made available to the Library under the agreement at hand and other metadata records pointing to that publishers resources derived from other sources (i.e. NSDL harvest of other aggregated collections). EPIC further modified the template agreement to include very specific articulation of how the Library uses metadata in connection with authorized service providers.

*New server: In March of 2005, EPIC hired a systems analyst to focus on technical issues related to work with publishers and NSDL. Paramount among this individual's accomplishments over the past year has been the alpha launch of an enterprise-scale server to process and convey large collections of frequently cycling publisher metadata (e.g. capable of working with article-level metadata updated monthly from numerous research publishers publishing thousands of journals). The first metadata processing/conveyance activity slated for the new system is an archive of articles from the Scientific American ranging from 1993 to 2005 (> 5000 articles) followed by Blackwell Publishing and Cambridge University Press (together, nearly 1000 journals). Apropos of this enhanced capability, EPIC is presently in discussions with other large STEM journal publishers including Oxford University Press, Taylor and Francis, SAGE Publications, and Springer.

*Educational overview: In connection with NSDL partnerships with research publishers and Core Integration content redistribution and contextualization efforts made possible by the new Fedora-based NDR, CI Columbia has embarked on the early stages of a STEM primary research exhibit that seeks to include pedagogic overviews of original research articles across the STEM disciplines. Herein NSDL partner publishers will make titles, abstracts, citation information, and links to the full text (open access) of select paper available for this feature. Initial candidate publishers include the life-science aggregator BioOne, Blackwell Publishing, and Cambridge University Press.

Information as of Feb. 2006. Updates (in parentheses) as of April 2006

 1. Negotiation phase: Bedford, Freeman & Worth
 2. Negotiation phase: Oxford University Press Books
 3. Negotiation phase: Oxford University Press Journals
 4. Negotiation Phase: Taylor & Francis
 5. Negotiation Phase: SAGE Publication
 6. Data processing phase: Scientific American
 7. Data processing phase: John Wiley
 8. Data processing phase: Blackwell
 9. Data processing phase: Tool Factory
10. Data processing phase: Cambridge University Press Books
11. Data processing phase: Cambridge University Press Journals
12. Data processing phase: Houghton-Mifflin
13. Data processing phase: Nature Publishing Group
14. Data processing phase: Tom Snyder
15. Data processing phase: O'Reilly Media/Safari
16. Data processing phase: BioOne (agreed to participate educational overview exhibit)
17. Data processing phase: National Academy Press
18. Pre-negotiation phase: American Psychological Association
19. Pre-negotiations phase: Elsevier (books) (recently moved to data processing)
20. Pre-negotiations phase: Riverdeep Interactive Learning
21. Pre-negotiation Phase: Springer (recently moved to data processing)

Michael Luby
Publisher Relations