

Core Research for the Networked University Digital Library (NUDL)

Edward A. Fox
Department of Computer Science
Virginia Tech

Contact Information

Edward A. Fox
Department of Computer Science, M/C 0106
Virginia Tech, 660 McBryde Hall
Blacksburg, VA 24061
Phone: (540) 231-5113
Fax: (540) 231-6075
Email: fox@vt.edu; <http://fox.cs.vt.edu>

WWW PAGES

<http://www.nudl.org>

List of Supported Students and Staff

Marcos Andre Gonçalves

Ryan Richardson

Collaborators in Germany (working on Dissertations On-Line, with funding from DFG):

- Prof. Diepold, (Education Science, Humboldt University Berlin)
- Prof. Hilf, (Physics, University of Oldenburg)
- Prof. Toerner, (Mathematics, University of Duisburg)
- Prof. Gasteiger (Chemistry, University of Erlangen)
- Dr. Schirnbache (Computer Center of Humboldt University Berlin)

Project Award Information

- **Award Number:** IIS-9986089 (SGER)
- **Duration:** 05/15/2000 – 03/01/2002*
- **Title:** Core Research for the Networked University Digital Library (NUDL)

Keywords

Networked Digital Library of Theses and Dissertations (NDLTD), Computer Science Teaching Center (CSTC), ACM Journal of Educational Resources in Computing (JERIC), NSDL, GMD, CONACyT

Project Summary

The Networked University Digital Library (NUDL) is a worldwide initiative that addresses making the intellectual capital produced in universities more accessible, stimulating international collaboration across all disciplines. In collaboration with the Dissertations Online project in Germany, we began to develop a worldwide, multilingual, federated digital library to facilitate sharing of research and educational materials among universities. The NUDL testbed is focused first on electronic theses and dissertations but is being extended to include courseware and other university digital resources.

This SGER project provides core support for NUDL, emphasizing digital library research of particular relevance to the German-US partnership – especially regarding information retrieval, metadata, markup (XML), character sets, languages, and multimedia. The concrete result will be an effective multilingual federated search system for a variety of languages, starting with English and German, but readily extensible to others such as Chinese, Greek, Italian, Japanese, Korean, Portuguese, Russian, and Spanish. This effort will demonstrate interoperability across a number of protocols, especially with the Harvest system employed by the German group. It will analyze and reconcile the varying approaches in the USA and Germany to handling metadata and markup so as to facilitate data sharing and integrated searching and browsing. Other supporting research will involve German-US collaboration in the human-computer interaction arena, on requirements engineering, and on user interface design, development, and evaluation.

Our long-term, scalable, and sustainable vision is to make submission to and usage of digital libraries a universal practice, as part of students' educational experiences. The goal is for all dissertations, theses, reports, portfolios, and other works to be entered by those students into personal, university, regional, discipline-wide, or other digital libraries. All of the world's future researchers and scholars can become accustomed to contributing to and making use of our testbed, which may become one of the largest federated digital libraries, with hundreds of thousands of multimedia works submitted

each year, used by millions of researchers in every nation. We will approach this vision by targeting an annual production of at least 10,000 electronic dissertations or theses in USA and Germany.

We build upon work begun in 1987 and funded in 1996 by the U.S. Department of Education that has led to over 100 members of the Networked Digital Library of Theses and Dissertations (NDLTD), in more than a dozen countries. We will collaborate with the German team that has been extending work on electronic theses and dissertations since 1997. In the courseware-related part of NUDL, we will link together all interested parts of the NSF Division of Undergraduate Education funded National SMETE (science, mathematics, engineering, and technology education) Digital Library – especially our work on the Computer Science Teaching Center (CSTC) and Curriculum Resources in Interactive Multimedia (CRIM) – with various international efforts such as the distributed physics worldwide retrieval system PhysDoc of the EPS (European Physical Society), which includes the PhysDis dissertation system.

Publications and Products

BOOKS:

- E. Fox, Joseph M. Moxley, and Christian R. Weisser, eds., *The ETD Sourcebook: Theses and Dissertations in the Electronic Age*, in preparation for publication by Marcel Dekker

JOURNAL/MAGAZINE ARTICLES:

- E. Fox and G. Marchionini. *Digital Libraries: Extending Traditional Values*. Guest Editors' Introduction to special section on Digital Libraries. *Commun. of the ACM*, May 2001, in press
- L. Cassel and E. Fox. *Introducing the ACM Journal of Educational Resources in Computing (JERIC)*, editor-in-chiefs' introduction. 1(1), April 2001, in press.
- E. Fox. Short pieces (In Brief) in *D-Lib Magazine* (www.dlib.org):
 - Fox, Edward A., *Open Archives Initiative*, <http://www.dlib.org/dlib/june00/06inbrief.html#FOX>
 - Fox, Edward A., Fernando Adrian Das Neves, Robert France, Marcos Gonçalves, and Hussein Suleman, *Digital Libraries 2000*, Fifth ACM Conference on Digital Libraries, <http://www.dlib.org/dlib/july00/07inbrief.html#FOX>

BOOK CHAPTERS:

- E. Fox and S. Urs. *Digital Libraries*. ARIST 2001, in process.
- E. Fox, M. Goncalves, and N. Kipp. *Digital Libraries in Education: Background, Theory, Case Studies, and Prospects*. In *Handbook on Information Technologies for Education & Training*, in Springer series "International Handbook on Information Systems", ed. Heimo Adelsberger, Betty Collis, Jan Pawlowski, 2000, 18 pages, in press
- E. Fox. *Overview of Digital Library Components and Developments*. Chapter 4 in *Digital Libraries and Virtual Workplaces: Important initiatives for Latin America*

in the Information Age, edited by Johann van Reenen, Washington: Organization of American States, 2001, in press.

- E. Fox. Overview of a Guide for Electronic Theses and Dissertations. Chapter 5 in Digital Libraries and Virtual Workplaces: Important initiatives for Latin America in the Information Age, edited by Johann van Reenen, Washington: Organization of American States, 2001, in press.

REFEREED CONFERENCE/WORKSHOP PAPERS:

- M. Gonçalves, R. France, Edward A. Fox, Eberhard R. Hilf, Michael Hohlfeld, Kerstin Zimmermann, Thomas Severiens. Flexible Interoperability in a Federated Digital Library of Theses and Dissertations. In Proceedings of the 20th World Conference on Open Learning and Distance Education, "The Future of Learning - Learning for the Future: Shaping the Transition", ICDE2000, Duesseldorf, Germany, 01 - 05 April 2001
- MARIAN: Searching and Querying Across Heterogeneous Federated Digital Libraries, by Marcos Andre Gonçalves, Robert K. France, Edward A. Fox, and Tamas E. Doszkocs. Reviewed workshop paper presented by E. Fox, DELOS Workshop, Zurich, Switzerland, Dec. 12, 2000. Talk:
<http://www.ndltd.org/talks/delosfox.ppt>, paper:
http://www.ercim.org/publication/ws-proceedings/DelNoe01/11_Fox.pdf
- Canós, J. H., Fox, E.A., Gonçalves, M. and France, R., NDLTD: una biblioteca digital global de tesis doctorales y de licenciatura. Actas de las I Jornadas de Bibliotecas Digitales. Valladolid, noviembre de 2000. Translation: NDLTD: a global digital library of theses and dissertations. Proceedings of the 1st Spanish workshop on digital libraries (6-7 October, 2000, Valladolid, Spain), ISBN: 84-8448-066-6
- Hussein Suleman, Edward A. Fox, Marc Abrams. "Building Quality into a Digital Library", <http://purl.org/net/repository>, short paper in Proc. ACM Digital Libraries '2000, June 4-7, 2000, San Antonio.
- Fernando Das Neves and Edward A. Fox. "A study of user behavior in an immersive Virtual Environment for digital libraries", in Proc. ACM Digital Libraries '2000, June 4-7, 2000, San Antonio.

REVIEWS:

- E. Fox. Review of "Digital Libraries" by William Y. Arms, MIT Press, 2000, in press for Kluwer Journal "Information Retrieval", 2001.

Project Impact

RELATED CONFERENCES/WORKSHOPS RUN:

- Chair: Electronic Theses and Dissertations Standards Meeting. OCLC, Dublin, Ohio, Jan. 9-10, 2001.
- Chair: Extending Interoperability of Digital Libraries: Building on the Open Archives Initiative. ECDL Workshop, Lisbon, Portugal, Sept. 21, 2000.
<http://purl.org/net/oa Sept00>

- Chair: Extending Interoperability of Digital Libraries: Building on the Open Archives Initiative. Workshop for ACM Hypertext'2000 and ACM Digital Libraries'2000. San Antonio, TX, June 3, 2000, <http://purl.org/net/oaijune00>
- Chair, Steering Committee meetings of NDLTD on 9/15/2000, 3/21/2001
- Member, Coordination Group (led by R. Steinmetz, with W. Effelsberg and N. Georganas), "Multimedia for Multimedia: Learning and Teaching in the Next Decade", Dagstuhl seminar, Castle Dagstuhl, Germany, June 11-16, 2000

RELATED COURSES TAUGHT:

- Digital Libraries (for graduate students, CS6604, Fall 2000)
- Digital Libraries (for undergraduate honors students, UH3004, Fall 2000)
- Digital Library Courseware, WWW site, Blacksburg, VA: Virginia Tech Department of Computer Science, <http://ei.cs.vt.edu/~dlib/>

KEYNOTE/BANQUET/INT'L INVITED/DISTINGUISHED SPEAKER

PRESENTATIONS:

- Terabyte scale multilingual digital library services and harvesting from Open Archives. Invited plenary for NIT'2001, Beijing, China, May 29-31, 2001, to be given
- Status Report on the NDLTD. Invited closing plenary for Fourth International Symposium on Electronic Theses and Dissertations (ETDs), Caltech, Pasadena, CA, March 22-24, 2001, <http://www.ndltd.org/talks/etd01key.ppt> (plus helped run standards meeting, chaired 2 sessions, and hosted a "networking" table on NDLTD future directions)
- Digital Libraries: From Theory to Applications in Education and Business. Invited keynote presentation, 2000 Asian DL conference, ICADL'2000, Seoul, Korea, Dec. 6-8, 2000. <http://www.ndltd.org/talks/koreainv.ppt>
- Digital Libraries: Extending and Applying Library and Information Science and Technology, invited keynote for ACM CIKM 2000, McLean, VA, Nov. 9, 2000. <http://www.ndltd.org/talks/cikm2000.ppt>
- The Co-Evolution of Digital Libraries and the WWW. Invited keynote for WebNet '2000, San Antonio, Nov. 4, 2000. <http://www.ndltd.org/talks/webnet.ppt>
- Digital Libraries: Extending the Impact of Information Science, sole keynote for final DISSAnet Conference, Pretoria, South Africa, Oct. 26-27, 2000. DISSAnet is a project between the Royal School of Library and Information Science in Copenhagen and the Universities of Pretoria and the North in South Africa. It is funded by DANIDA. <http://www.ndltd.org/talks/prolissa.ppt>

TUTORIALS:

- Practical Digital Libraries Overview. Half-day tutorial, scheduled for ACM SIGIR'2001, New Orleans, LA, Sept. 7-12, 2001, to be given
- Practical Digital Libraries Overview. Full-day tutorial, scheduled for JCDL'2001, Roanoke, VA, June 24, 2001, to be given
- Practical Application of Digital Libraries. Two-day course hosted by ASPI, Caracas, Venezuela, June 11-12, 2001, to be given
- Digital Libraries from Theory to Framework to Application: 5S Model, Open Archives Initiative, and Networked Digital Library of Theses and Dissertations,

- Dec. 6, 2000, Seoul, S. Korea, half-day tutorial for 2000 Asian DL conference, ICADL'2000, <http://www.ndltd.org/talks/foxtutkr.ppt>
- Practical Digital Libraries Overview. Half-day tutorial, ACM Multimedia 2000, Los Angeles, CA, Oct. 30 - Nov. 4, 2000, <http://ei.cs.vt.edu/~dlib/tut/MM00.htm>
 - Digital Libraries. One-and-a-half day (12 hour) tutorial for DISSAnet workshop, Pretoria, South Africa, Oct. 23-24, 2000, <http://ei.cs.vt.edu/~dlib/tut/SA.htm>
 - Practical Application of Digital Libraries. Full-day tutorial, RRTC Advanced Technology Seminar, Rosslyn, VA, October 13, 2000, <http://ei.cs.vt.edu/~dlib/tut/RRTC4.htm>
 - Digital Libraries: An Overview. Full-day tutorial, with ECDL'2000, European Conference on Digital Libraries, Lisbon, Portugal, Sept. 18, 2000, <http://ei.cs.vt.edu/~dlib/tut/ECDL2000.htm>
 - Digital Libraries: An Overview. Two half-day tutorials (intro., advanced), ACM Digital Libraries '2000, June 4-7, 2000, <http://www.ndltd.org/talks/DL2000i.ppt>, <http://ei.cs.vt.edu/~dlib/tut/indexi.htm>, <http://ei.cs.vt.edu/~dlib/tut/indexa.htm>

ORAL PRESENTATIONS:

- Parts 1 and 6, statement for panel 1: Developing the UNESCO International Guide for ETDs: A Multilingual Training Resource of Best Practices, for Fourth International Symposium on Electronic Theses and Dissertations (ETDs), Caltech, Pasadena, CA, March 22-24, 2001, <http://www.ndltd.org/talks/etd01p1.ppt>
- Collaboration on Digital Libraries. Invited presentation at NEC RI, Princeton, NJ, Dec. 27, 2000. <http://www.ndltd.org/talks/nec2000.ppt>
- Collaboration on Digital Libraries. Invited presentation at Hofstra U., Hempstead, NY, Dec. 22, 2000. <http://www.ndltd.org/talks/hofstra2.ppt>
- Digital Libraries: Starting with Authors/Creators. Invited presentation for Digital Strategies 2000, National Archives, College Park, MD, Nov. 16-17, 2000. <http://www.ndltd.org/talks/nara.ppt>
- Plenary panel 1 slides for ACM Multimedia '2000: Curricula and Resources for Courses about Multimedia, Oct. 30 - Nov. 4, 2000, Marina del Rey, CA, <http://www.ndltd.org/talks/mm00pan1.ppt>
- Digital Libraries, OAI, ETDs, and NDLTD. Invited presentation for SEALS ETD Meeting, Oct. 25, 2000, Port Elizabeth, S. Africa. <http://www.ndltd.org/talks/seals.ppt>
- Extending Interoperability of Digital Libraries: Building on the Open Archives Initiative, OAI Workshop at ECDL'2000, Lisbon, September 21, 2000, <http://www.ndltd.org/talks/oaiecdl.ppt>
- Networked Digital Library of Theses and Dissertations (www.ndltd.org). Invited presentation for ECDL 2000 Panel 3 "Digital Libraries Programs and International Cooperation", September 20, 2000, Lisbon. <http://www.ndltd.org/talks/ecdlpnl.ppt>
- Open Archives initiative and the Networked Digital Library of Theses and Dissertations. Invited presentation for seminar at UCSD, Geisel Library, Seuss Room, Aug. 9, 2000 <http://www.ndltd.org/talks/UCSD20000809.ppt>
- Virginia Tech's Digital Library Research Laboratory: From Hardware to Software to Projects to Theory, invited seminar talk at IBM T. J. Watson Research Center, July 26, 2000. <http://www.ndltd.org/talks/ibm20000726public.ppt>

- Digital Libraries, invited presentation requested by Virginia Center for Innovative Technology for AOL - University Research Day: Finding Common Ground, Reston, VA, July 14, 2000 (and separate breakout session for those desiring further info.)
- Digital Libraries and the Open Archives Initiative, invited presentation for Louisiana State U., June 30, 2000. <http://www.ndltd.org/talks/lsuoai.ppt>
- Networked Digital Library of Theses and Dissertations (NDLTD), invited presentation for Louisiana State U., June 30, 2000. <http://www.ndltd.org/talks/lsundltd.ppt>
- OAI, Electronic Theses, and Scholarly Communication: The Way to the Virtual University, Sonder-Kolloquium Einladung, Carl von Ossietzky Universitat Oldenburg, June 19, 2000, <http://www.ndltd.org/talks/oldnbrg.ppt>

HOSTING VISITS (selected):

- Professor Shalini R. Urs, Dept. Library and Information Science, U. Mysore (India), Fulbright Fellowship, 5 months: August 12, 2000 - Feb. 12, 2001
- Professor Mann-Ho Lee, Dept. of Computer Science, CNU (Chungnam National University) in Korea, visiting for 1 year starting Aug. 14, 2000
- Professor Pablo de la Fuente, Departamento de Informatica, University of Valladolid (Spain), visiting 4/27/2000 - 6/24/2000
- Dr. Ahira Maeda, Japan, postdoctoral fellow for Oct. 22, 2000 - March 21, 2001, after graduation from Nara Institute of Science and Technology (NAIST), Japan
- Professor Byongsun (Sunny) Kim, Korean Studies Information Center, The Academy of Korean Studies, Seoul, visiting for 1 year starting Dec. 14, 2000
- RAI-Channel Two News Group from Italy, hosted visit to Digital Library Research Laboratory and other parts of Torgersen Hall, Oct. 10, 2000

AWARDS SERVING AS PRINCIPAL INVESTIGATOR (follow on to this project):

- NSF IIS-0080748: High Performance Interoperable Digital Libraries in the Open Archives Initiative: \$99,999 for two years starting 4/1/2001, Project director: E. Fox (with co-PIs J.Alfredo Sánchez, Universidad de las Américas-Puebla --- UDLA, and David Garza-Salazar, Monterrey Technology Institute --- ITESM, both funded by CONACyT in Mexico)
- NSF IIS-0086227: Open Archives: Distributed services for physicists and graduate students (OAD): \$315,259 for 9/1/2000-8/31/2003. Project director: E. Fox (w. Royce Zia, Physics, VT, and E. Hilf, U. Oldenburg, PI on matching German DFG project)
- AOL: AOL Fellowship in Information Retrieval: \$35,000 for 2001, possibly renewable for 2 additional years. Project director: E. Fox
- UNESCO: International Guide for the Creation of Electronic Theses and Dissertations: \$13,000 for Dec. 2000 to May 2001. Project director: E. Fox.
- SOLINET: Networked Digital Library of Theses and Dissertations: \$1000 to establish Virginia Tech Foundation account 872037 for NDLTD. Project director: E. Fox.
- UNC Wilmington: A Digital Library of Reusable Science and Math Resources for Undergraduate Education: \$78,374 for 5/1/2000 - 4/30/2002. Project director: E.

Fox. (subcontract for what originally was CCRI proposal to NSF 98-63 for \$1,143,282, DLI-2, Undergraduate Emphasis)

- NSF IIS-9986089 (SGER): Core Research for the Networked University Digital Library (NUDL): \$79,997 for 5/15/2000 - 3/1/2002. Project director: E. Fox.

Goals, Objectives, and Targeted Activities

GOALS:

- To achieve interoperability across systems used in NUDL
- To effectively support the addition of new NUDL resources, each with suitable metadata
- To provide efficient, effective, and reliable worldwide access to all NUDL resources

OBJECTIVES:

- To make the currently locally available work of scholars writing in diverse languages more widely used
- To link with other efforts, like:
 - NCSTRL and CoRR, for preprints
 - NSDL (formerly called SMETE-lib), for Science, Mathematics, Engineering, and Technology Education
- To target the next generation of scholars, so they will learn about digital libraries (DLs), through personally submitting their own works
- To integrate prior work on interoperability
- To adapt related metadata efforts
- To extend the use of markup (XML)
- To effectively handle character sets and languages
- To evaluate and enhance NUDL services

TARGETED ACTIVITIES FOR NEXT YEAR:

- We will extend the MARIAN system to be a portable service-provider system for open archives, with multilingual capabilities, tailorable to differing collections and audiences, and efficient operation.
- We will continue work on NDLTD, CSTC, JERIC, and related projects.
- We will continue testing, and shift to production status, regarding our efforts to have an open archive that aggregates metadata from all cooperating NDLTD member sites.

Project References

DEMONSTRATIONS:

Building a Unified Library for Theses and Dissertations, by Fox, E. A., France, R. K., Gonçalves, M. A., Suleman, H., and Lee, M. H., demonstration described in Proceedings of the 3rd International Conference of Asian Digital Library, Seoul, December 6-8, 2000

XXDL and CSTC and Virginia Tech, demonstration for NSF-sponsored NSDL workshop, September 22-24, 2000, NSF, Arlington, VA,
<http://www.ndltd.org/talks/nsdlcstc.ppt>

SOFTWARE AND DATABASE PRODUCTS:

MARIAN: The campus production service version of this runs on a collection of PC and NeXTstep machines in the Computing Center and is designed to support scores of simultaneous users. It builds upon data and code from the CED lexicon and CODER (both developed at VT in the 1980s). With support from NLM, and new support from NSF, it is being revamped in Java to become a key part of our research on digital libraries. Integrated with MARIAN is the Envision interface, to graphically manage search results, that is being recoded in JAVA. URL:
<http://www.dlib.vt.edu/products/marian.html>

ETD: Our electronic thesis and dissertation related software and systems assist with building the Networked Digital Library of Theses and Dissertations (NDLTD). This includes extensive online information at <http://etd.vt.edu>, with documentation, automated submission software, multimedia tutorials, SGML/XML document type definitions, conversion tools and services, and adaptation of Open Archives Initiative. Collaboration on these efforts includes staff on multiple grants, personnel in Graduate School, Library, Computing Center, and other universities. The VT Library now maintains the <http://etd.vt.edu> site, while we manage domains ndltd.org and theses.org (also known as dissertations.org) and related listservs. As director of NDLTD, PI Fox coordinates efforts of its over 100 members. According to Don Fischer, Program Officer at FIPSE (US Dept. of Education), in email dated 2/14/2000 "Your work is outstanding. Your outreach efforts are incredibly extensive."

CSTC: The Computer Science Teaching Center (www.cstc.org) is hosted at Virginia Tech where PI Fox serves as editor and is responsible for the software, server, and repository. This supports teachers and learners interested in computing. It has been extended to also support the ACM Journal of Educational Resources in Computing, JERIC.

Open Archives Initiative: This international digital library initiative involves specifications, protocols, and other standards (www.openarchives.org). PI Fox is one of the founding members and serves on the Steering Committee. Students here have developed a range of widely used software to support development, testing, and

operation. We ran 2 workshops for OAI in 2000, attended the public launch meeting 1/23/2001, have several grants related that will continue to yield new tools, and will run a workshop in connection with ACM SIGIR'2001 in New Orleans in September.

Area Background

Digital libraries are being built upon a firm foundation of prior work as the high-end information systems of the future. A component approach is becoming popular, with well-established support for key components like the repository, especially through the Open Archives Initiative. We can separately consider digital objects, metadata, harvesting, indexing, searching, browsing, rights management, linking, and powerful interfaces. Flexible integration will be possible through a variety of architectures, using buses, agents, and other technologies. The field as a whole is undergoing rapid growth, supported by advances in storage, processing, networking, algorithms, and interaction. There are many initiatives and developments, including those supporting education and research.

Area References

Michael Lesk, *Practical Digital Libraries: Books, Bytes and Bucks*, San Francisco: Morgan Kaufmann, 1997

William Y. Arms, *Digital Libraries*, Cambridge, MA: MIT Press, 2000

William Y. Arms, ed. *D-Lib Magazine*, Reston, VA: CNRI, <http://www.dlib.org>

Potential Related Projects

First, every university is invited to join NUDL. A first action typically is to participate in ND LTD, see www.ndltd.org/join.

Second, it is encouraged that all IDM project identify each particular result that might be of pedagogical value, such as to those taking courses on database management or information retrieval. These can be demonstrations, datasets, visualizations, small tutorials, exercises, syllabi, or a broad range of other knowledge modules. These can then be entered into CSTC (www.cstc.org, which also supports submission to ACM Journal of Educational Resources in Computing).

Third, all projects working on interoperability, training/education about information management, preservation, multilingual information retrieval, or other problems related to NUDL, are encouraged to consider NUDL and its emerging collections as a testbed.