

# David S. Dubin

## Business Address:

Graduate School of Library and Information Science  
University of Illinois at Urbana-Champaign  
Champaign, IL 61820  
Phone: 217-244-3275  
WWW: <http://people.lis.illinois.edu/~ddubin/>

## Education

Ph.D. Information Science, 1996, University of Pittsburgh.

M.S. Library and Information Science, 1990, Drexel University, Philadelphia, PA.

B.S. Humanities and Communications, 1988, Drexel University.

## Employment History

**University of Illinois at Urbana-Champaign, Graduate School of Library and Information Science.** Research Associate Professor, Since August 2007.

**University of Illinois at Urbana-Champaign, Graduate School of Library and Information Science, Information Systems Research Laboratory.** Senior Research Scientist, August 2001 – August 2007. Specialist in electronic publishing, information retrieval, data analysis, and visualization. Teaching areas: data analysis and information processing.

**University of Illinois at Urbana-Champaign, Graduate School of Library and Information Science.** Assistant Professor, September 1996 – August 2001. Teaching areas: information organization and access, information storage and retrieval, data structures, library automation, document processing, networked information systems, research methods.

**University of Pittsburgh, Department of Information Science.** Teaching Fellow, September 1995 – August 1996, January 1995 – May 1995, January 1994 – September 1994, May 1991 – August 1993. Teaching areas: information retrieval, online searching, behavioral models, human information processing, systems analysis.

**University of Pittsburgh, Department of Library Science.** Teaching Fellow, May 1995 – September 1995. Teaching area: microcomputer applications.

**University of Pittsburgh, Department of Information Science.** Visiting Lecturer, September 1994 – December 1994. Teaching area: information storage and retrieval.

**dh Molde College, Molde, Norway.** Research Fellow in Computing Science, August 1993 – December 1993. Teaching area: human information processing.

**University of Pittsburgh, Department of Library Science.** Adjunct Instructor, September 1992 – December 1992. Teaching area: microcomputer applications.

**University of Pittsburgh, Department of Information Science.** Graduate Student Assistant, September 1990 – May 1991.

**Drexel University, Philadelphia, PA, College of Information Studies.** Graduate Assistant, September 1988 – June 1990.

**Delaware Valley Transplant Program, Philadelphia, PA.** Cooperative Education Intern (technical communication), November 1984 – June 1987.

#### **Ph.D. Committees as Research Director**

Larry S. Jackson (Library and Information Science, University of Illinois at Urbana-Champaign, 2009) *Website Structure*.

#### **Ph.D. Committees as Reader**

Karen M. Wickett (Library and Information Science, University of Illinois at Urbana-Champaign, 2012) *Collection/Item Metadata Relationships*.

Jin Ha Lee (Library and Information Science, University of Illinois at Urbana-Champaign, 2008) *Analysis of Information Features in Natural Language Queries for Music Information Retrieval: Use Patterns and Accuracy*.

Jun Wang (Library and Information Science, University of Illinois at Urbana-Champaign, 2006) *Computational Approaches to Linguistic Consensus*.

Bei Yu (Library and Information Science, University of Illinois at Urbana-Champaign, 2006) *An Evaluation of Text Classification Methods for Literary Study*.

Qin He (Library and Information Science, University of Illinois at Urbana-Champaign, 2000) *Component Study of Co-Word Analysis*.

Jianhua Dong (Library and Information Science, University of Illinois at Urbana-Champaign, 2000) *Combination of Multiple Web Search Results and its Effect on the Search Performance*.

Damien Guillaume (Astronomy, University of Strasbourg, 2000) *Distributed Information Retrieval, Search and Processing in Astronomy*.

#### **Professional Activities**

Co-Chair (with Bridget Almas and Sayeed Choudhury) Research Data Provenance Interest Group, Research Data Alliance, September 2013 – present.

Co-Chair (with Carolyn Anderson), Annual Meeting of the Classification Society of North America, June 2007.

Board of Directors, Classification Society of North America, 2002–2004.

Editorial Board, Classification Literature Automated Search Service, Classification Society of North America.

Local Organizing Committee, Ninth Biennial Meeting of the International Federation of Classification Societies, July 2004.

Program Committee, Annual Meeting of the Classification Society of North America, June 1998.

Demonstrations Chair, Second ACM International Conference on Digital Libraries, July 1997.

Local Arrangements Committee, Annual Meeting of the Classification Society of North America, June 1993.

Registration Chair, Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, June 1993.

### **Refereed Publications**

Berti, M., Almas, B., Dubin, D., Franzini, G., Stoyanova, S., and Crane, G. R. (2015). The linked fragment: TEI and the encoding of text reuses of lost authors. *Journal of the Text Encoding Initiative*, (8).

Dubin, D. and Jett, J. (2015). An ontological framework for describing games. In *Proceedings of the 2015 Joint Conference on Digital Libraries*. New York.

Dubin, D., Senseney, M., and Jett, J. (2013). What it is vs. how we shall: complementary agendas for data models and architectures. In *Proceedings of Balisage: The Markup Conference 2013*, volume 10 of *Balisage Series on Markup Technologies*. Montréal, Canada.

Thomer, A. K., Baker, K. S., Sacchi, S., and Dubin, D. (2012). Completeness, coverage & equivalence in scientific data records. *Proceedings of the American Society for Information Science and Technology*, 49.

Wickett, K. M., Sacchi, S., Dubin, D., and Renear, A. H. (2012). Identifying content and levels of representation in scientific data. *Proceedings of the American Society for Information Science and Technology*, 49.

Dubin, D., Wickett, K. M., and Sacchi, S. (2011). Content, format, and interpretation. In *Proceedings of Balisage: The Markup Conference 2011*, volume 7 of *Balisage Series on Markup Technologies*. Montréal, Canada.

Sacchi, S., Wickett, K., Renear, A., and Dubin, D. (2011). A framework for applying the concept of significant properties to datasets. In *Proceedings of the American Society for Information Science and Technology*, volume 48 of *ASIS&T Annual Meeting Proceedings*. American Society for Information Science and Technology.

Dubin, D. (2010). Encoded descriptions at face value. In A. Grove, editor, *Proceedings of the 73rd Annual Meeting of the American Society for Information Science and Technology*, volume 47 of *ASIS&T Annual Meeting Proceedings*. Information Today, Inc., Pittsburgh, PA.

Renear, A. H., Wickett, K. M., Urban, R. J., Dubin, D., and Shreeves, S. L. (2008). Collection/item metadata relationships. In J. Greenberg and W. Klas, editors, *Proceedings of the International Conference on Dublin Core and Metadata Applications, Berlin*, pages 80–89. Dublin Core Metadata Initiative, Goettingen University Press, Goettingen.

Renear, A. H. and Dubin, D. (2007). Three of the four FRBR group 1 entity types are roles, not types. In A. Grove, editor, *Proceedings of the 70th Annual Meeting of the American Society for Information Science and Technology*. Information Today, Inc., Medford, NJ.

Dubin, D., Futrelle, J., and Plutchak, J. (2006). Metadata enrichment for digital preservation. In B. T. Usdin, editor, *Proceedings of Extreme Markup Languages 2006*. Montreal, Quebec.

Dubin, D. and Birnbaum, D. (2004). Interpretation beyond markup. In B. T. Usdin, editor, *Proceedings of Extreme Markup Languages 2004*. Montreal, Quebec.

Dubin, D. (2003). Object mapping for markup semantics. In B. T. Usdin, editor, *Proceedings of Extreme Markup Languages 2003*. Montreal, Quebec.

Dubin, D., Sperberg-McQueen, C. M., Renear, A., and Huitfeldt, C. (2003). A logic programming environment for document semantics and inference. *Literary and Linguistic Computing*, 18(2):225–233. (This is a corrected version of an article that appeared in 18:1 pp. 39-47).

Lee, J. and Dubin, D. (2003). Vocabulary mapping in the NASA ADS: Prospects for practical subject access. In B. Corbin, E. Bryson, and M. Wolf, editors, *Library and Information Services in Astronomy IV*, pages 249–256. U.S. Naval Observatory, Washington, DC.

Renear, A. and Dubin, D. (2003). Towards identity conditions for digital documents. In S. Sutton, editor, *Proceedings of the 2003 Dublin Core Conference*. University of Washington, Seattle, WA.

Renear, A., Dubin, D., Sperberg-McQueen, C. M., and Huitfeldt, C. (2003). XML semantics and digital libraries. In C. C. Marshall, G. Henry, and L. Delcambre, editors, *Proceedings of the third ACM/IEEE-CS joint conference on Digital libraries*, pages 303 – 305. IEEE, Los Alamitos, CA.

Renear, A., Phillippe, C., Lawton, P., and Dubin, D. (2003). An XML document corresponds to which FRBR Group 1 entity? In B. T. Usdin, editor, *Proceedings of Extreme Markup Languages 2003*. Montreal, Quebec.

- Renear, A., Dubin, D., Sperberg-McQueen, C. M., and Huitfeldt, C. (2002). Towards a semantics for XML markup. In R. Furuta, J. I. Maletic, and E. Munson, editors, *Proceedings of the 2002 ACM Symposium on Document Engineering*, pages 119–126. Association for Computing Machinery, McLean, VA.
- Sperberg-McQueen, C. M., Dubin, D., Huitfeldt, C., and Renear, A. (2002). Drawing inferences on the basis of markup. In B. T. Usdin and S. R. Newcomb, editors, *Proceedings of Extreme Markup Languages 2002*. Montreal, Quebec.
- Lee, J. and Dubin, D. (1999). Context-sensitive vocabulary mapping with a spreading activation network. In M. Hearst, F. Gey, and R. Tong, editors, *Proceedings of the 1999 ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 198–205. Association for Computing Machinery, ACM, New York.
- Lee, J., Dubin, D. S., and Kurtz, M. J. (1999). Co-occurrence evidence for subject vocabulary reconciliation in ADS databases. In D. M. Mehringer, R. L. Plante, and D. A. Roberts, editors, *Astronomical Data Analysis Software and Systems VIII*, volume 172 of *A.S.P. Conference Series*, pages 287–290. Astronomical Society of the Pacific, San Francisco.
- Dubin, D. (1998). The search for structure and the search for meaning. In R. Schwartz, editor, *Advances in Classification Research Volume 6*, pages 13–20. Information Today, Inc., Medford, NJ.
- Dubin, D. S. (1998). Addressing the heterogeneity of subject indexing in the ADS databases. In U. Grothkopf, H. Andernach, S. Stevens-Rayburn, and M. Gomez, editors, *Library and Information Services in Astronomy III*, volume 153 of *A.S.P. Conference Series*, pages 77–83. Astronomical Society of the Pacific, San Francisco.
- Korfhage, R. R., Dubin, D., and Housman, E. M. (1998). Computer-aided interactive classification: applications of VIBE. In P. Solomon, editor, *Advances in Classification Research Volume 7*, pages 83–101. Information Today, Inc., Medford, NJ.
- Korfhage, R. R., Dubin, D., and Housman, E. M. (1998). What good is visualization: three experiments. In R. F. Erbacher and E. Pang, editors, *Visual Data Exploration and Analysis V*, pages 196–207. SPIE, Bellingham, WA.
- Dubin, D., Kwaśnik, B. H., and Tangmanee, C. (1996). Elicitation techniques for classification research. In R. Fidel, C. Beghtol, B. H. Kwaśnik, and P. J. Smith, editors, *Advances in Classification Research Volume 5*, pages 33–68. Information Today, Inc., Medford, NJ.
- Dubin, D. S. (1996). Attribute selection for visualizing multidimensional document spaces: a progress report. In D. S. Ebert, editor, *Workshop on New Paradigms in Information Visualization and Manipulation (NPIV '96)*, pages 8–11.
- Dubin, D. (1995). Document analysis for visualization. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 199–204. ACM SIGIR, Association for Computing Machinery, New York.

Dubin, D. (1994). Applying similarity measures to texts. *TEXT Technology*, 4(4):283–291.

Olsen, K. A. and Dubin, D. (1994). Maintaining a personal reference library with a word processor and a scanner. *TEXT Technology*, 4(2):149–154.

Song, I. Y. and Dubin, D. (1991). An intensional query processor in prolog. In M. H. Hamza, editor, *Computer Applications in Design, Simulation, and Analysis*, pages 204–207. ACTA Press, Anaheim, CA.

### Unrefereed Publications

Dubin, D., Futrelle, J., Plutchak, J., and Eke, J. (2009). Preserving meaning, not just objects: semantics and digital preservation. *Library Trends*, 57(3):595–610.

Renear, A. H. and Dubin, D. (2008). FRBR as an interdisciplinary high-middle range theory. In A. De Cenzo, editor, *Proceedings of iConference 2008*. Los Angeles.

Dubin, D. (2004). The most influential paper Gerard Salton never wrote. *Library Trends*, 52(4):748–764.

Dubin, D. S., Renear, A., and Sperberg-McQueen, C. M. (2003). Addressing obstacles to the retrieval of structured documents. Technical Report UIUCLIS- -2003/1+EPRG, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, Champaign, IL.

Dubin, D. (2002). Standards and information. In J. R. Schement, editor, *Encyclopedia of Communication and Information*, volume 3, pages 965–967. Macmillan, New York.

Dubin, D. (1999). Toward more robust discrimination-based indexing models. Technical Report UIUCLIS- -1999/7+IRG, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, Champaign, IL.

Dubin, D. (1998). Dimensions and discriminability: The role of controlled vocabulary in visualizing document associations. In P. A. Cochrane and E. H. Johnson, editors, *Visualizing Subject Access for 21st Century Information Resources*, pages 39–44. University of Illinois Graduate School of Library and Information Science, Champaign, IL.

Dubin, D. (1998). Further cautions for the calculation of discrimination values. Technical Report UIUCLIS- -1999/3+IRG, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, Champaign, IL.

Dubin, D. (1997). Measurement in information science (book review). *Journal of Classification*, 14(2):327–330.

Dubin, D. (1996). *Structure in Document Browsing Spaces*. Ph.D. thesis, University of Pittsburgh.

Dubin, D. (1996). Multimedia and imaging databases (book review). *Information Processing and Management*, 32(6):769–770.

Dubin, D. (1995). Search strategies for Internet resources. *School Library Media Quarterly*, 24(1):53–54.

Spring, M. B. and Dubin, D. (1992). *Hands-on PostScript*. Hayden Books, Carmel, IN. (Published in Polish translation by Intersoftland of Warsaw as *PostScript od A do Z*).

Dubin, D. (1989). Online databases put a new universe of resources at our fingertips. *TIES Magazine*, pages 38–43.

### **Presentations**

Almas, B., Berti, M., Choudhury, S., Dubin, D., Senseney, M., and Wickett, K. M. (2013). Representing humanities research data using complementary provenance models. Presented at Building Global Partnerships: Research Data Alliance Second Plenary Meeting, Washington, DC.

Birnbaum, D. J., Dubin, D., and Vakareliyska, C. (2013). Clustering calendars of saints: discriminatory power and variable selection. Presented at the annual meeting of the Classification Society, Milwaukee, WI.

Wickett, K., Dubin, D., Almas, B., and Senseney, M. (2013). Extending the systematic assertion model for humanities research. Presented at the 2013 ASIS&T Research Data Access and Preservation Summit.

Dubin, D. (2012). On data identity and scientific equivalence. Presented at the annual meeting of the Classification Society, Pittsburgh, PA.

Dubin, D. (2012). Internal cohesion and external separation. Presented at the 23rd Annual ASIS&T SIG/CR Classification Research Workshop.

Palmer, C. L., Chao, T. C., Weber, N. M., Sacchi, S., Wickett, K. M., Renear, A. H., Baker, K., Thomer, A., and Dubin, D. (2012). Integrating conceptual and empirical studies of data to guide curatorial processes. Presented at the 2012 ASIS&T Research Data Access and Preservation Summit.

Wickett, K. M., Sacchi, S., Dubin, D., and Renear, A. H. (2012). Representing identity and equivalence for scientific data. Presented at the American Geophysical Union (AGU) Fall Meeting.

Wickett, K. M., Thomer, A., Sacchi, S., Baker, K. S., and Dubin, D. (2012). What dataset descriptions actually describe: Using the systematic assertion model to connect theory and practice. Presented at the 2012 ASIS&T Research Data Access and Preservation Summit.

Dubin, D. (2011). Data theory and scientific data management. Presented at the annual meeting of the Classification Society, Pittsburgh, PA.

- Sacchi, S., Renear, A., and Dubin, D. (2011). One thing missing or two things are confused: An analysis of OAIS representation information. Presented at the International Digital Curation Conference (IDCC).
- Dubin, D. (2009). On the expressive content of games. Presented at the 2009 Joint Conference of the National Popular Culture and American Culture Associations, New Orleans.
- Dubin, D. (2008). Challenges for board game classification. Presented at the 2008 ALISE Annual Conference, Philadelphia, PA.
- Dubin, D. and Birnbaum, D. J. (2008). Reconsidering conventional markup for knowledge representation. Presented at Balisage: the Markup Conference, Montreal.
- Renear, A., Dubin, D., and Wickett, K. (2008). When digital objects change — exactly what changes? Presented at the 2008 Annual Meeting of the American Society for Information Science and Technology, Columbus, OH.
- Renear, A. H., Urban, R. J., Wickett, K. M., Palmer, C. L., and Dubin, D. (2008). Sustaining collection value: Managing collection/item metadata relationships. Presented at Digital Humanities 2008, Oulu, Finland.
- Renear, A. H., Wickett, K. M., Urban, R. J., and Dubin, D. (2008). The return of the trivial: problems formalizing collection/item metadata relationships. In *JCDL '08: Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, pages 464–464. Association for Computing Machinery, ACM, New York, NY, USA.
- Dubin, D. (2007). Instance or expression? another look at reification. Presented at Extreme Markup Languages 2007, Montreal.
- Dubin, D. (2006). Reframing author cocitation analysis. Presented at the tenth biennial meeting of the International Federation of Classification Societies, Ljubljana, Slovenia. (also presented at the 2006 meeting of the Classification Society of North America, Piscataway, NJ.
- Huitfeldt, C., Sperberg-McQueen, M., Dubin, D., and Johnsen, L. G. (2006). Markup languages for complex documents: an interim project report. Presented at Digital Humanities, Paris.
- Downie, J. S., Renear, A., Mathes, A., Medina, K., Dubin, D., and Lee, J. H. (2005). Modelling complex multimedia relationships in the humanities computing context: Are Dublin Core and FRBR up to the task? Presented at ALLC/ACH, Victoria, British Columbia.
- Dubin, D. (2005). Unpacking the interpretation of METS markup. Presented at the Digital Library Federation Fall Forum, Charlottesville, VA.
- Dubin, D. and Birnbaum, D. J. (2005). A declarative framework for modeling pronunciation and rhyme. Presented at ALLC/ACH, Victoria, British Columbia.



- Birnbaum, D. and Dubin, D. (2004). Measuring similarity in the contents of medieval miscellany manuscripts. Presented at the ninth biennial meeting of the International Federation of Classification Societies, Chicago, IL.
- Dubin, D. (2004). Semantic markup or markup semantics? Presented at the University of Pittsburgh School of Information Sciences Colloquium Series.
- Dubin, D. and Lee, J. (2003). Beyond three dichotomies. Presented at the annual meeting of the Classification Society of North America, Tallahassee, FL.
- Dubin, D., Ripoche, G., and Gasser, L. (2003). Organizational dynamics of software development. Presented at the DIMACS Workshop on Algorithms for Multidimensional Scaling, Tallahassee, FL.
- Dubin, D. (2002). Do conceptual spaces have metric structure? Presented at the annual meeting of the Classification Society of North America, Madison, WI.
- Sperberg-McQueen, C. M., Renear, A., Huitfeldt, C., and Dubin, D. (2002). Skeletons in the closet: Saying what markup means. Presented at ALLC/ACH, Tübingen, Germany.
- Dubin, D. (2001). Model, data, and system centered perspectives in information retrieval research. Presented at the annual meeting of the Classification Society of North America, St. Louis, MO.
- Dubin, D. and Rorvig, M. (2001). Classical two-way metric MDS adapted to handle very large datasets. Presented at the DIMACS Workshop on Algorithms for Multidimensional Scaling, Rutgers University, Piscataway, NJ.
- Dubin, D. and Pape, D. X. (2000). Clustering applications and validation: A case study with the Kohonen SOM. Presented at the annual meeting of the Classification Society of North America, Montreal, Canada.
- Dubin, D. and Pape, D. X. (1999). Validation strategies for large-scale clustering applications. Presented at the annual meeting of the Classification Society of North America, Pittsburgh, PA.
- Dubin, D. (1997). Challenges for the future of document clustering. Presented at the annual meeting of the Classification Society of North America, Washington, DC. (Also presented at the 1998 meeting of the International Federation of Classification Societies, Rome, Italy).
- Dubin, D. (1996). Clustering tendency and the cluster hypothesis in information retrieval. Presented at the annual meeting of the Classification Society of North America, Amherst, MA.
- Dubin, D. (1994). POI discovery and the clarity of VIBE displays. Presented at the annual meeting of the Classification Society of North America, Houston, TX.