Controlled Vocabularies as a Sphere of Influence

By

Anita S. Coleman
Asst. Professor
School of Information Resources & Library Science
University of Arizona
Tucson, AZ 85719
Email: asc@u.arizona.edu

&

Paul Bracke
Head of Systems and Networking
Arizona Health Sciences Library
University of Arizona
Tucson, AZ 85719
Email: paul@ahsl.arizona.edu

This is a preprint of a paper published in Raghavan, K.S. and Prasad, K.N. (Editors). Knowledge organization, information systems and other essays: Prof. A. Neelameghan festschrift. New Delhi: Ess Ess Publications for Ranganathan Centre for Information Studies, 2006. (pages 89 – 110)

Abstract:

Objective: The objective of this citation study is to understand the use and influence of the concept of "controlled vocabularies" in Geographic Information Science (GIS) as part of a larger goal to distinguish information science from information technology.

Methods: Articles with pre-selected descriptors that represented the concept of "controlled vocabularies" within GIS were selected from GeoRef and validated in ISI indexes. Bibliographic coupling and content analysis of the article titles were used to draw clusters and understand the influence of the concept of controlled vocabularies in other discipline such as the geosciences.

Results and Conclusion: The results from this analysis provide one perspective of the LIS sub-domain of "controlled vocabularies" as represented in *GeoRef* and used in the context of GIS research and scholarship. Findings are used to suggest future research directions to address issues related to better understanding of the concept of "controlled vocabularies" and the provision of knowledge organization tools that will promote interdisciplinary understanding. The creation of special, more-finely grained in-depth classifications and thesauri for the concept itself, namely, "controlled vocabulary" is recommended.

Controlled Vocabularies as a Sphere of Influence

This paper describes an exploratory study that examined the existence of the Information Science concept of Controlled Vocabularies in Geographic Information Science (GIS). This study is part of a larger domain analysis that is mapping the scholarly output, in Geographic Information Science from 1980 onwards, with the specific goals of distinguishing Information Science (IS) concepts from those of Information Technology (IT).

The Importance of Controlled Vocabularies

The NISO Z39.19 standard for the construction and maintenance of monolingual thesauri defines a thesaurus as a "controlled vocabulary of terms in natural language that are designed for postcoordination. The need to control the formation and use of terms stems mainly from two basic features of natural language, namely synonyms (different terms representing the same concept) and polysemes or homographs (terms with the same spelling representing different concepts). The controlled vocabulary is established by information specialists or lexicographers and is generally employed in indexing." (NISO, 1993) The purpose of a controlled vocabulary is thus to improve information retrieval in systems of organization. Also, studying, creating and maintaining controlled vocabularies can be said to fall within the scope of the Library and Information Science (LIS) discipline and profession. Buckland (1999) supports this notion and in fact, argues for the centrality of "vocabulary" in LIS. "Vocabulary" he notes, "commonly refers to the stylized adaptation of natural language to form indexes and thesauri. Much of bibliographic access, filtering, and information retrieval can be viewed as matching or translating across vocabularies. Multiple vocabularies are simultaneously present. A simple query in an online catalog normally involves at least five distinct vocabularies: those of the authors; the cataloger; the syndetic structure; the searcher; and the formulated query." Buckland is thus extending the concept of "vocabulary" to include "the range (or repertoire) of values in any field of bibliographic description and, in a more extended sense, the range of types in a set at any level (word, field, collection, and library)." Buckland further suggests that vocabularies are central features of digital libraries.

Controlled vocabularies are thus both an IS and IT-related concept. As digital libraries have developed, the concept of controlled vocabularies and enabling true semantic interoperability and facilitating meaning across different disciplines has become increasingly important as we try to build the semantic web: "an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation." (Berners-Lee, et al., 2001) Investigating the notion of controlled vocabularies in other disciplines is thus timely. As Neelameghan (1972) has eloquently suggested, the universe of knowledge is a "turbulently dynamic continuum" with fragmentation and hybridization of disciplines frequently occurring and the study of other disciplines will help enhance our own understanding of Library Science and it's merger with Documentation or Information Science.

Domain analysis studies that attempt to identify the attributes of subjects and disciplines include citation analysis and bibliometrics (Hjorland, 2001; Neelameghan, op. cit.); it has been used to document *specializations* and *modes of disciplinary growth* (Ellis et al, 1999; Hurt, 1983; White & McCain, 1998), and includes:

1) Identification - for example, what is the intellectual structure of a discipline like Accounting (Bricker, 1991);
2) Fragmentation among disciplines such as Management Information Science (Culnan, 1986, 1987);
3) Fusion in Medical Informatics (Morris & McCain, 1998) and;
4) Specialization to identify the emerging directions and existing coherent structures of research specialties.

Citation analysis has also been used to visualize interdisciplinary fields but studies have typically not sought to isolate Information Science from Information Technology related concepts and problems. Such a distinction is of interest because IS and IT are multi-disciplinary and share many research interests and yet there is little cross citation of their literatures. Also, IS-IT is often used as a tool in other disciplines. We are curious in understanding if the role of citation networks in disciplinary growth can provide clues as to when a new IS-IT intensive discipline becomes a science and is no longer just a research method or a tool. We also wished to use citation analysis data in conjunction with content analysis data to improve the nature of our findings and to investigate the potential, problems and strengths of these methods for helping distinguish between types of concepts that can help in mapping disciplinary structures and pinpointing their growth and status.

Research Questions:

1. How is the concept of controlled vocabularies described in GIS literature?
2. What are the relationships between articles in GIS discussing controlled vocabularies?

IS-IT concepts as an assumption for this research

Anecdotal evidence suggests that basic IT concepts include components of the computer, storage, memory, processing, input, output, hardware, software, multimedia, applications, operating systems, security, ethics, and networking. Similarly, IS concepts include, human information processing, organization, storage, retrieval, and use of information, especially recorded information. In the Dewey Decimal Classification system, Table 1, the Standard Subdivisions table outlines the many Library and Information Science and Computer Science concepts that may be combined with any topic (WebDewey, 2002). Our first step in this research was to identify how a thesaurus and a classification scheme in these disciplines may show disciplinary structures. We examined the ASIS Thesaurus for Information Science (Milstead, 1998) as well as the Dewey Decimal Classification scheme for the phrase Controlled Vocabularies and made a list of the related and narrower classes, terms and phrases and came up categories of IS terms. We also briefly

examined the ACM Classification for categories of IT terms for related concepts and terms (ACM Classification, 2002).

Methods Used:

Identifying concept terms: We qualitatively examined both the recommended core curriculum for GIS by the NCGIA (NCGIA, 2000) and *GeoRef thesaurus* (Goodman, 2000) to identify the terms and phrases used to represent Controlled Vocabularies in these documents. *GeoRef* does not use the phrase Geographic Information Science (it uses "geographic information systems") while the *NCGIA Core Curriculum* emphasizes the discipline as Geographic Information Science and use of Geographic Information Systems as methods. Our final list of selected terms were the indexing descriptors used in *GeoRef* and these include terms such as: *nomenclature, classification, digital data*, and *spatial data*. *Nomenclature* and *classification* are clearly and closely related to the IS controlled vocabulary concept. Nomenclature covers naming conventions, and classification deals with placing entities into a normalized naming scheme based on their characteristics. Digital data and spatial data deal with issues of data standardization that allow for data retrieval and use in the GIS context.

Identifying articles dealing with these concepts: The selected descriptors for controlled vocabularies were searched in *GeoRef* both separately and in conjunction with the term "geographic information systems." Retrieval of relevant articles was limited to English language publications and resulted in approximately 500 citations. From the resulting sets, non-journal articles (for example, books) and articles in journal titles that were not indexed in ISI's *Social Science Citation Index* or *Science Citation Index* were omitted (Web of Science, 2002) leaving us with a total of 62 citations. These limits were done for two reasons: one, to make the set size manageable and two, to limit the articles to high quality research. Selecting articles that were representative of high quality GIS research was important so that they would be representative of high quality research in the discipline. We could not pursue an original idea we had of limiting our analysis to a core set of GIS journals from ISI's *Journal Citation Reports* because coverage of these journals in *GeoRef* was highly selective.

Analysis and Results:

Our test set included 62 articles and these are listed in Appendix A. Of the 62 articles, 3 contained no citations. Additionally, 19 articles contained unique citations and were not coupled with other documents in the set. The remaining 40 articles formed a network of bibliographic couplings. Articles that were cited multiple times, and thus were used to establish couplings are listed in Appendix B. A list of couplings is presented in Appendix C.

Bibliographic coupling and co-citation analysis are citation methods that can be used to identify subject relationships in articles. Figures 1 and 2 show the difference between bibliographic couplings and co-citations. In Figure 1, articles A and B are

bibliographically coupled in article C. In Figure 2, we see that article F cites articles D and E.

Bibliographic coupling is a static method that establishes a relationship between articles through the presence of citations to the same articles in their bibliographies. As seen in Figure 1, if articles A and B both cite C, A, and B are bibliographically coupled. This linking is based on intrinsic, immutable (in a print environment, at least) characteristics of an article – the references contained in its bibliography. Although there are exceptions, the coupling often indicates a subject relationship.

In co-citation analysis, articles become linked when they are both cited in another article. As seen in Figure 2, if articles D and E are both cited by article F, articles D and E are co-cited. This clustering is based on extrinsic characteristics, and patterns may change over time as an article gains or loses favor in a discipline, or as disciplinary lines shift. As with bibliographic coupling, co-citation analysis often indicates a subject relationship

The bibliographic couplings matrix was analyzed using MVSP 3.1. (MVSP 2003) to demonstrate statistical similarities amongst articles. The UPGMA method, Percent Similarity was used to generate the dendogram and draw the map of bibliographic couplings. In addition to statistical cluster analysis, content analysis to overcome the limitations of citation analysis was also done. Placing the key phrases from titles into the categories defined, prior to analysis provides an alternative method to the bibliographic coupling analysis; the rules for content analysis are shown in Table 1. We also drew a map comparing the two clusters.

Figure 1: Bibliographic Coupling

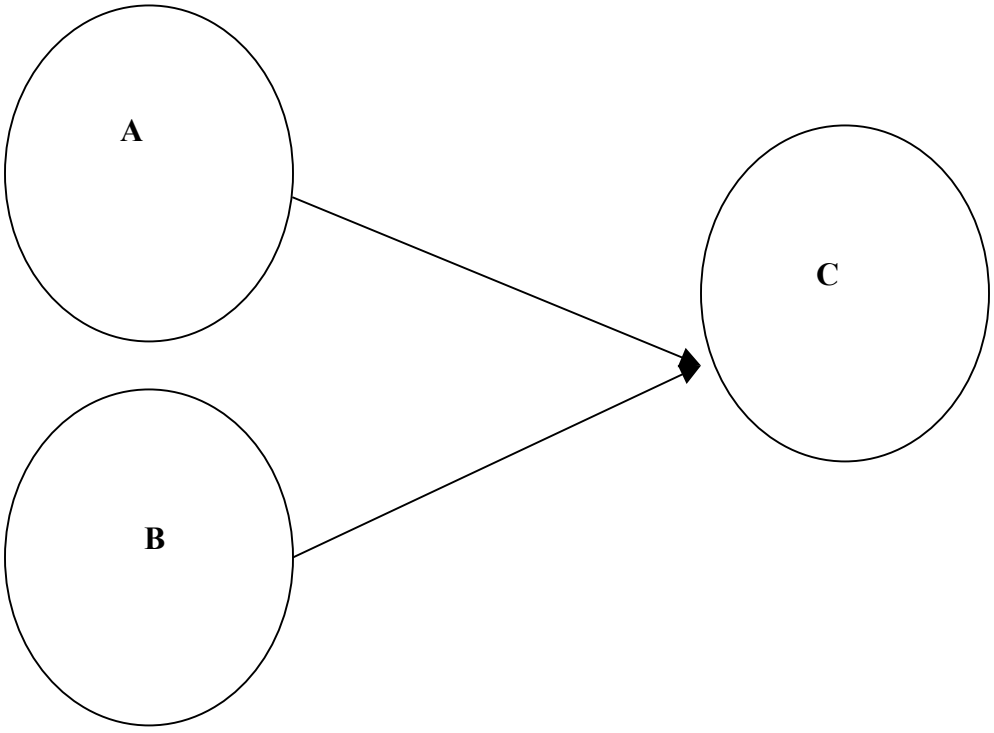


Figure 2: Co-citations

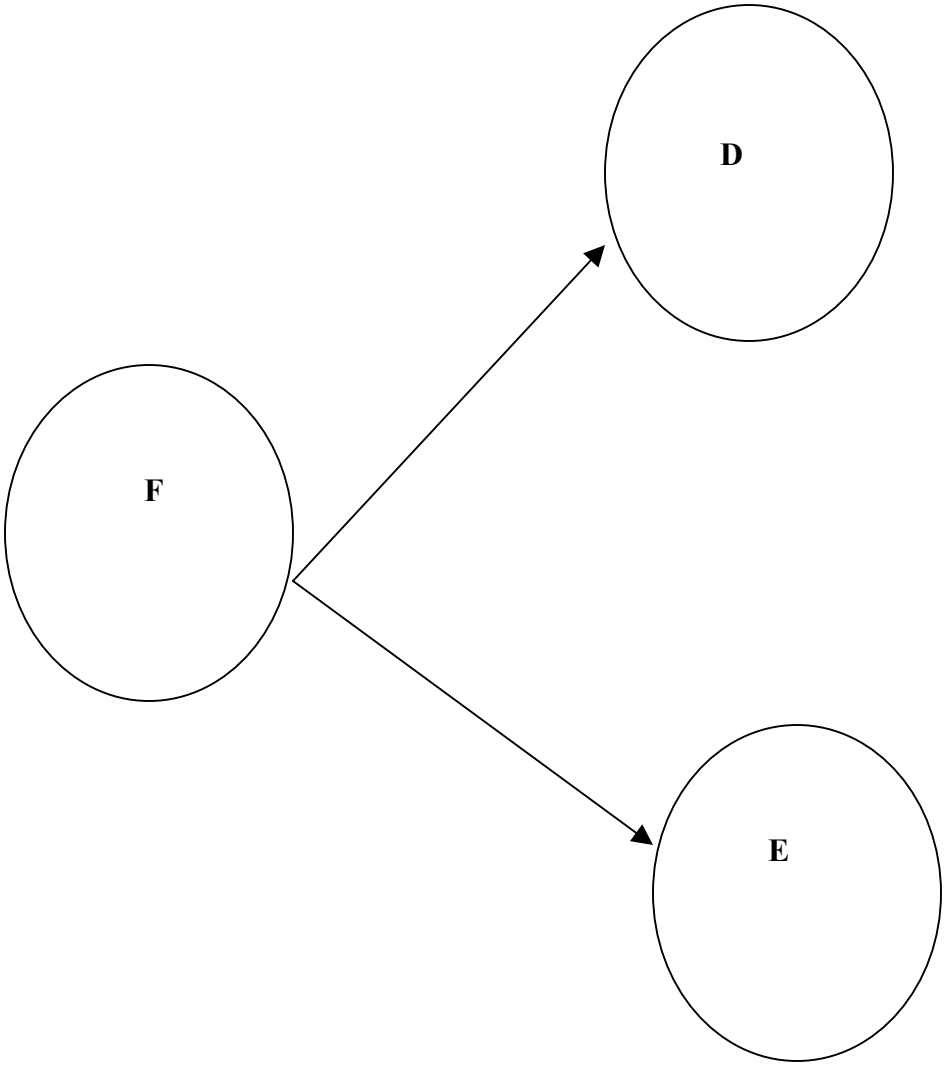


Table 1: Content Analysis Rules

Rule	Terms present	Category	Articles
1.	"map*" or "cartography"	CARTOGRAPHY	J,Y,Z,AS,AZ,BJ
2.	"classification" or "classes" or "codification"	CLASSIFICATION	I,U,AB,AC,AG,BB, BE
3.	("information system" but not "GIS") or "database" of "virtual memory"	INFORMATION SYSTEM	F,L,M
4	"GIS" or "geographic* information system*" or "SDSS"	GIS AS METHOD	N,AA,AD,AE,AI,A L,AO,AT,AU,AW, AY,BC,BD,BO
5.	"analysis" or "stat*"	STATISTICAL ANALYSIS	A,T
6.	"model*"	MODEL	H,R,S,V,X,AF,AH, AJ,AK,AM,AV,AX , BK,BP
7.	Name of Specific Software or "software"	SOFTWARE	C,E,G,AN,BL
8.	Anything else	GEOSCIENCE	B,D,K,O,P,Q,W,AP ,AQ,AR,BI,BM,BN

Table 2: Statistical clusters with strongly correlating content analysis clusters

STATISTICAL CLUSTERS	CORRELATING CONTENT ANALYSIS CLUSTERS
CLUSTER 1 C,F,H,X,AG,AH,AI,AL,BI,BO	GIS AS METHOD N,AA,AD,AE,AI,AL,AO,AT,AU,AW,AY ,BC,BD,BO
CLUSTER 2 J,P,AJ,AK,AU,BP	CARTOGRAPHY J,Y,Z,AS,AZ,BJ
CLUSTER 3 Q,R,V,W,AE,AV,AZ,BJ	MODEL H,R,S,V,X,AF,AH,AJ,AK,AM,AV,AX, BK,BP
CLUSTER 4 K,T,AD,AM,AN,AO,AS,AT,AY,BF,BN	GIS AS METHOD N,AA,AD,AE,AI,AL,AO,AT,AU,AW,AY ,BC,BD,BO
CLUSTER 5 O,U,Y,AA,AF,AW,BD	GIS AS METHOD N,AA,AD,AE,AI,AL,AO,AT,AU,AW,AY ,BC,BD,BO

Five (5) clusters were generated through statistical analysis and the content analysis cluster generated eight (8) clusters. The statistical clusters related exclusively to subject-related clusters of GIS as method, model, and cartography (Table 2). There were no relationships between a statistical cluster and what we had defined as a purely IS-IT cluster (classification, software, system) with overlap of greater than 1. The only instance of overlap with one article turned out to be with software and not classification. Articles that fell into IS-IT clusters based on title words did not correspond to the statistical clusters, indicating little or no commonality with bibliographic couplings. Figure 3 shows the relationships between the statistics and the content clusters. Thus, while the content analysis clusters contain all the articles, the statistical clusters do not. Figure 4 shows map of bibliographic couplings and confirms that the concept of controlled vocabulary is not a research front within GIS. Table 2 shows the statistical and content analysis clusters and we briefly explain the content analysis rules used to derive these clusters.

"GIS as Method" was a category that only required the presence of GIS or geographic(al) information system. The phrase "as method" was added to the category name to indicate that the term almost always appeared in contexts that indicated its use as a research method. In addition, Geoscience "catch-all" category was used for titles that did not seem to form any patterns or natural groupings. Additionally, "Information System" and "Software" could easily be combined into a single category without losing much meaning. Data was not included as a rule in the content analysis because of the varieties of contexts in which it appeared. There were several references, for example, to data sets, but they were usually the object of another process. The term data was also used in the context of data structures, data processing and data handling. Although this limited the analysis of controlled vocabulary concepts in GIS, spatial data and digital data were not as core to the concept of controlled vocabulary as classification, a term that was used as a

term for grouping in content analysis. More work is needed on the exact meaningful use of specific terms such as classes, ontologies, taxonomies, and arrays in the context of GIS data.

Figure 3: Relationships between clusters

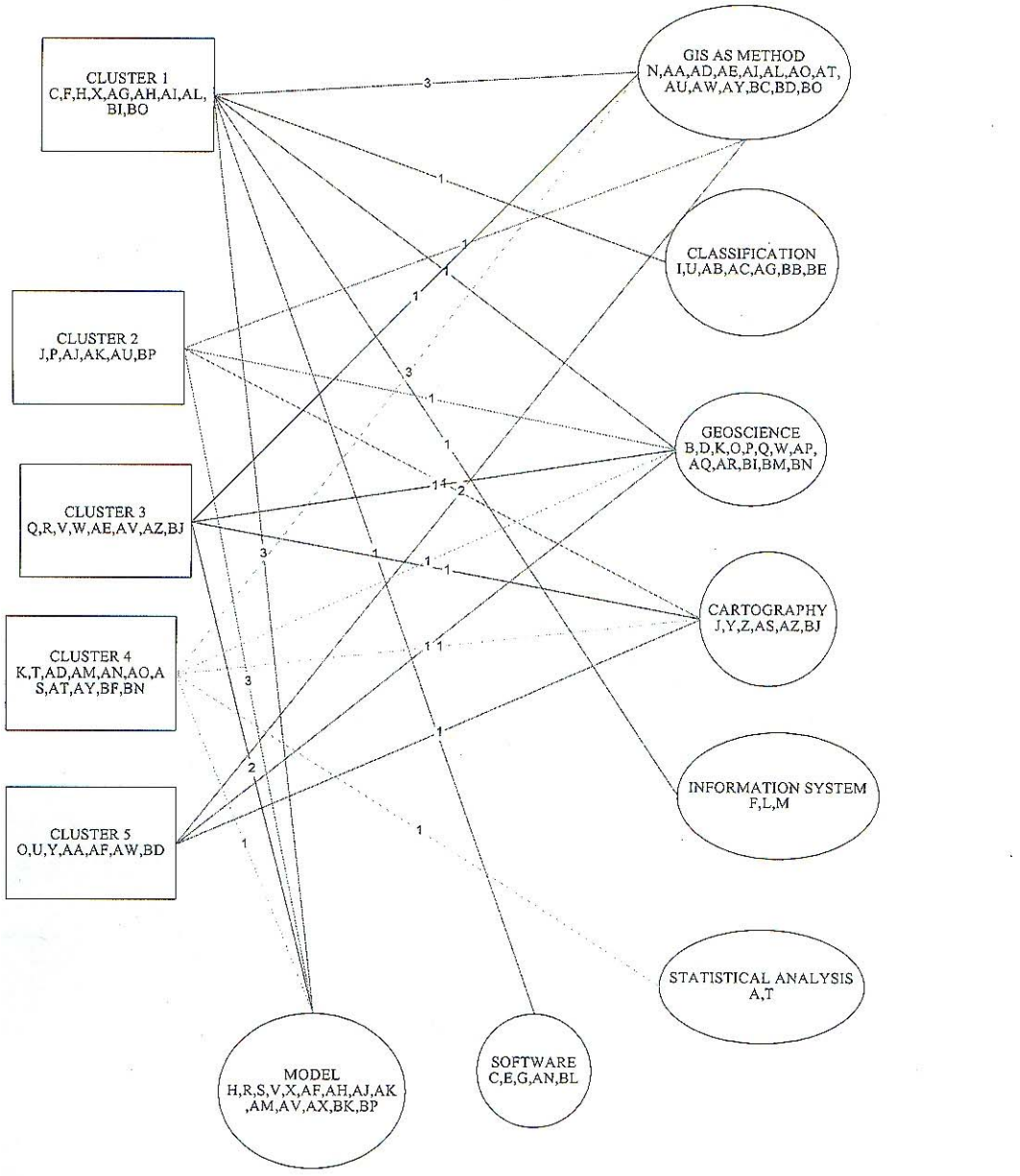
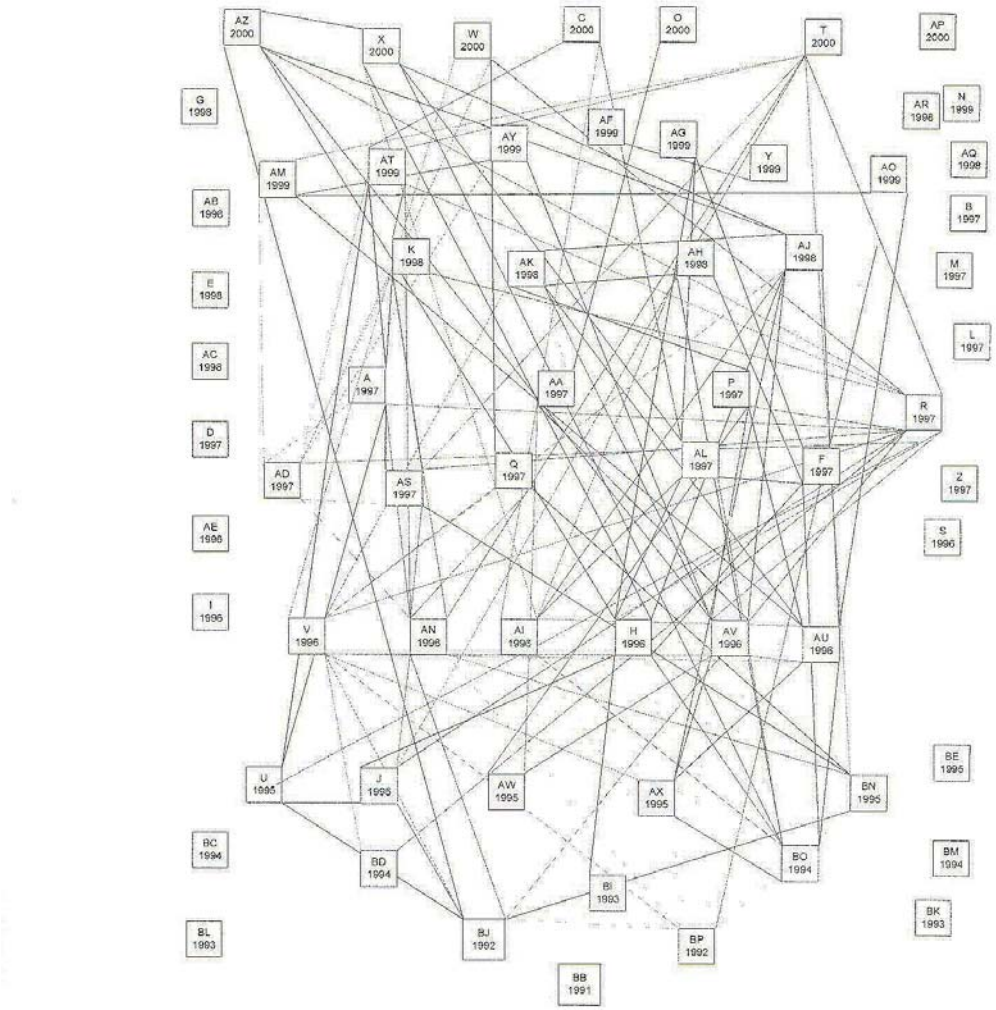


Figure 4: Map of bibliographical couplings



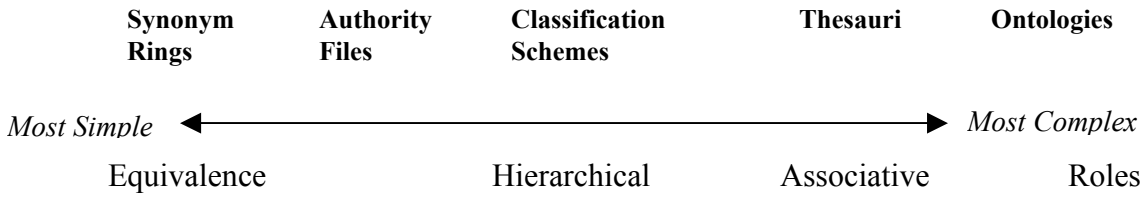
The two analyses of the sets of documents (statistical and content) thus present two different views of the documents. Controlled vocabularies seem to be well integrated with GIS-research when it is necessary in the examination of another topic but their sphere of influence in GIS based on our findings appears to be controversial and limited. Indeed the argument about the "ambiguity" of GIS as science or tool is reflected in these clusters (Wright et al, 1987). Replicating the study with a larger document set for validation is necessary before other conclusions can be made.

We also find that the subject of 'controlled vocabularies" is represented very poorly and at a most superficial level in Library and Information Science thesauri and classification schemes. Interestingly, descriptive and subject cataloging along with authority control of names and titles and vocabulary control of subjects is a widespread and standard

technique used to organize information resources in traditional libraries. It is a technique that is increasingly gaining in popularity in the world of digital libraries and with the use of metadata and the advent of the semantic web. The web world, however, is chaotic and naïve in the use of many of terms and processes related to controlled vocabulary: consider for example, that large scale projects in the United States such as the NSDL or DLESE both of which use flat, short lists of terms and call them a controlled vocabulary use this phrase in the most narrow sense. The small vocabularies (lists of controlled values is a better way to describe them) have few relationships of limited hierarchies; they in no way represent the richness and depth that a true controlled vocabulary that a thesaurus potentially offers. A distinction must be made between simple and complex vocabularies, such as thesauri and ontologies, which have varying types of relationships between terms and can number in the thousands for even a small concept.

A controlled vocabulary, as LIS professionals know, includes a range of knowledge organization tools. Both of these, a list of equivalent terms in the form of a synonym ring or a list of controlled or preferred terms in the form of Authority Files, are types of controlled vocabularies. Additionally, logic and a range of analytical methods can be used to build the controlled vocabulary for any discipline or subject. For example, faceted classification structuring may be used in a thesaurus, which is another form of a controlled vocabulary, making it a faceted thesaurus. Two modern faceted thesauri are the Art & Architecture Thesaurus (A&AT), and Medical Subject Headings (MeSH). MeSH is part of the Unified Medical Language System (UMLS), referred to as an ontology. Figure 5 graphically shows how controlled vocabularies can be placed on a continuum of simple to complex.

Figure 4: Types of controlled vocabularies



Simple vocabularies such as preferred lists of terms have long been proved to be clearly insufficient for efficient information retrieval. The do not represent the range of relationships among terms and phrases. The new NISO Z39.19 standard just released (2003) tries to help increase understanding about the concept of a "controlled vocabulary" and the creation of a true end-user thesaurus (Bates, 1986). If controlled vocabularies are to become a major sphere of influence that LIS exerts on other disciplines, and for LIS professionals to continue to be accepted as the experts in this area by new entrants into the field such as ontologists, taxonomists, knowledge managers, and information architects, we need to do two things: 1) an in-depth and finer-grained controlled vocabulary of the concept of "controlled vocabularies" itself with definitions of terms and their scope and context of all relationships (in particular the equivalence

relationships of synonyms and using qualifies and polyhierarchies), and the many disciplines that are now engaged in this area must be created. This could be represented as a classification scheme or a thesaurus but adapted from currently existing ones. 2) The newly developed controlled vocabulary should be used to replicate the study reported herein in order to examine our sphere of influence – both to see how this important topic has infiltrated other disciplines and how it is being used in others disciplines. Using machine classification methods on the full-text of the articles may also yield a different picture of the influence of this important concept on IT and specific disciplines.

References:

A&AT. Art & Architecture Thesaurus Online. (2005). Retrieved 15 March 2005 from http://www.getty.edu/research/conducting_research/vocabularies/aat/

ACM Classification. (1998). Top Two Levels of the ACM Computing Classification System. Retrieved 15 March 2005 from http://www.acm.org/class/1998/overview.html

Bates, Marcia. (1986). Subject access in online catalogs. Journal of the American Society for Information Science 37 (6): 357-376.

Berners-Lee, Timothy, Hendler, James, Lassila, Ora. (2001). The Semantic Web. *Scientific American*, May.

Bricker, Robert. (1991). Deriving disciplinary structures: Some new methods, models, and an illustration with accounting. *Journal of the American Society for Information Science,* 42 (1): 27-35.

Buckland, Michael. (1999). Vocabulary as a central concept in library and information science. In Arpanac, T. et al. (Eds.), *Digital libraries: Interdisciplinary concepts, challenges, and opportunities. Proceedings of the Third International Conference on Conceptions of Library and Information Science* [CoLIS3], May 23-26, 1999, Dubrovnik, Croatia, (pp 3-12). Zagreb: Lokve. Retrieved March 15, 2005, from http://www.sims.berkeley.edu/~buckland/colisvoc.htm

Culnan, M. (1986) The intellectual structure of management information systems, 1972-1982: a co-citation analysis, *Management Science*, 32, 2, 156-172.

Culnan, M. (1987) Mapping the intellectual structure of MIS, 1980-1985: a co-citation analysis, *MIS Quarterly*, 11 (3): 341-353.

DLESE. Digital Library for Earth System Education. Retrieved March 15, 2005 from http://ww.dlese.org/

Ellis, D., David, A., & Wilson, T. (1999) Information Science and Information Systems: Conjunct subjects disjunct disciplines. *Journal of the American Society for Information Science* 50 (12): 1095-1107.

Goodman, Barbara, Editor. (2000). GeoRef Thesaurus. 9th edition. AGI.

Hjorland, Birger. (2002). Domain analysis in Information Science: Eleven approaches – traditional as well as innovative. *Journal of Documentation* 58 (4): 422-462.

Hurt, C.D. (1983) A comparison of a bibliometric approach and an historical approach to the identification of important literature. *Information Processing and Management*, 19, 151-157.

MeSH. Medical Subject Headings Home Page. Retrieved 15 March 2005 from http://www.nlm.nih.gov/mesh/

Milstead, Jessica L. , Editor. (1998). ASIS Thesaurus of Information Science and Librarianship. 2nd edition. Medford, N.J.: Information Today.

Morris, Thodore A., and McCain, Katherine W. (1998). The Structure of Medical Informatics. Journal of the American Medical Informatics Association, 5: 448 – 466.

MVSP (2002). Multi-Variate Statistical Package. Retrieved 15 March 2005 from http://www.kovcomp.co.uk/mvsp/.

NCGIA. (2000). NCGIA Core Curriculum in GIScience. Retrieved March 15, 2005 from http://www.ncgia.ucsb.edu/education/curricula/giscc/

Neelameghan, A. 1972. Systems approach in the study of the attributes of the universe of subjects. *Library Science with a Slant to Documentation*. 9 (4): 445-472.

NISO Z39.19. (1998). Guidelines for the constuction and maintenance of monolingual thesauri. Revision of Z39.19-1980 and of 1993. Retrieved March 15, 2005 from http://www.niso.org/standards/resources/Z39-19.html

NISO Z39.19. (2003). Guidelines for the constuction and maintenance of monolingual thesauri. Revision of Z39.19-1980. Retrieved March 15, 2005 from http://www.niso.org/standards/resources/Z39-19.pdf

NSDL. National Science Digital Library. Retrieved March 15, 2005 from http://nsdl.org/

UMLS. (2005). Unified Medical Language System. Retrieved 15 March 2005 from http://www.nlm.nih.gov/research/umls/umlsmain.html

Web of Science. (2003). Retrieved with subscription March 15, 2005 from http://www.isiknowledge.com/

WebDewey. (2003) Dewey Decimal Classification. 21st edition. Dublin, OH: OCLC . Retrieved with subscription March 15, 2005 from http://connexion.oclc.org/

White, H.D. and McCain, K.W.. (1998) Visualizing a discipline: An author co-citation analysis of Information Science, 1972-1995. *Journal of the American Society for Information Science* 49 (4): 327-355.

Wright, D.J., Goodchild, M.F., & Proctor, J.D.. (1997) Demystifying the persistent ambiguity of GIS as "tool" versus "science". *Annals of the Association of American Geographers* 87 (2): 346-362.

Appendix A: Document Set Used in Study

A	Bourgault, G, A G Journel, J D Rhoades, D L Corwin, and S M Lesch. "Geostatistical Analysis of a Soil Salinity Data Set." Advances in Agronomy 58 (1997): 241-92.
B	Steel, Kevin, and Bette L Beswick. "Getting the Big Picture." Canadian Mining Journal 118, no. 1 (1997): 6-7.
C	Bivand, Roger S. "Using the R Statistical Data Analysis Language on Grass 5.0 Gis Database Files." Computers and Geosciences 26, no. 9-10 (2000): 1043-52.
D	Dong, Pinliang. "Implementation of Mathematical Morphological Operations for Spatial Data Processing." Computers and Geosciences 23, no. 1 (1997): 103-07.
E	Knox, Robinson Carl M, and Stephen J Gardoll. "Gis-Stereoplot; an Interactive Stereonet Plotting Module for Arcview 3.0 Geographic Information System." Computers and Geosciences 24, no. 3 (1998): 243-50.
F	McCormack, J E, and J Hogg. "Virtual-Memory Tiling for Spatial Data Handling in Gis." Computers and Geosciences 23, no. 6 (1997): 659-69.
G	Wulder, M A. "The Prediction of Leaf Area Index from Forest Polygons Decomposed through the Integration of Remote Sensing, Gis, Unix, and C." Computers and Geosciences 24, no. 2 (1998): 151-57.
H	Townsend, Philip A, and Stephen J Walsh. "Estimation of Soil Parameters for Assessing Potential Wetness; Comparison of Model Responses through Gis." Earth Surface Processes and Landforms 21, no. 4 (1996): 307-26.
I	Host, George E, Philip L Polzer, David J Mladenoff, Mark A White, and Thomas R Crow. "A Quantitative Approach to Developing Regional Ecosystem Classifications." Ecological Applications 6, no. 2 (1996): 608-18.
J	Lathrop, Richard G, Jr., John D Aber, and John A Bognar. "Spatial Variability of Digital Soil Maps and Its Impact on Regional Ecosystem Modeling." Ecological Modelling 82, no. 1 (1995): 1-10.
K	Nathanail, C P, and M S Rosenbaum. "Spatial Management of Geotechnical Data for Site Selection." Engineering Geology 50, no. 3-4 (1998): 347-56.
L	Cook, Jerome S, Michael P Fahy, and Frank Tian. "Profile in Gis; Construction of Water Rights Database at El Paso Water Utilities." Environmental and Engineering Geoscience 3, no. 4 (1997): 588-94.
M	Bartsch, H U, J Kues, J Sbresny, and J Schneider. "Soil Information System as Part of a Municipal Environmental Information System." Environmental Geology 30, no. 3-4 (1997): 189-97.
N	Doerfliger, N, P Y Jeannin, and F Zwahlen. "Water Vulnerability Assessment in Karst Environments; a New Method of Defining Protection Areas Using a Multi-Attribute Approach and Gis Tools (Epik Method)." Environmental Geology 39, no. 2 (1999): 165-76.

O	Nogues, J, J Herrero, Ochoa R Rodriguez, and J Boixadera. "Land Evaluation in a Salt-Affected Irrigated District Using an Index of Productive Potential." Environmental Management (New York) 25, no. 2 (2000): 143-52.
P	Turner, Graham W, Rod M C Ruffio, and Mark W Roberts. "Comparing Environmental Conditions Using Indicators of Pollution Hazard." Environmental Management (New York) 21, no. 4 (1997): 623-34.
Q	Zhang, Minghua, Shu Geng, Susan L Ustin, and Kenneth K Tanji. "Pesticide Occurrence in Groundwater in Tulare County, California." Environmental Monitoring and Assessment 45, no. 2 (1997): 101-27.
R	Corwin, Dennis L, Peter J Vaughan, and Keith Loague. "Modeling Nonpoint Source Pollutants in the Vadose Zone with Gis." Environmental Science and Technology, ES and T 31, no. 8 (1997): 2157-75.
S	Tsanis, I K, C Valeo, J Wu, and S Boyle. "Managing Contaminated Sediments Using a Hydrodynamic Model and a Gis." Environmental Technology 17, no. 8 (1996): 877-83.
T	Dolan, David M, Shaarawi Abdel H El, and Trefor B Reynoldson. "Predicting Benthic Counts in Lake Huron Using Spatial Statistics and Quasi-Likelihood." Environmetrics 11, no. 3 (2000): 287-304.
U	Lark., R M, and P H T Beckett. "A Regular Pattern in the Relative Areas of Soil Profile Classes and Possible Applications in Reconnaissance Soil Survey." Geoderma 68, no. 1-2 (1995): 27-37.
V	Tim, U S, D Jain, and H H Liao. "Interactive Modeling of Ground-Water Vulnerability within a Geographic Information System." Ground Water 34, no. 4 (1996): 618-27.
W	Shukla, Sanjay, Saied Mostaghimi, Vernon O Shanholt, Michael C Collins, and Burton B Ross. "A County-Level Assessment of Ground Water Contamination by Pesticides." Ground Water Monitoring and Remediation 20, no. 1 (2000): 104-19.
X	Endreny, T A, E F Wood, and D P Lettenmaier. "Satellite-Derived Digital Elevation Model Accuracy; Hydrological Modelling Requirements." Hydrological Processes 14, no. 2 (2000): 177-94.
Y	Zakir, F A, M H T Qari, and M E Mostafa. "A New Optimizing Technique for Preparing Lineament Density Maps." International Journal of Remote Sensing 20, no. 6 (1999): 1073-85.
Z	Wood, William B. "A Jeffersonian Vision for Mapping the World." Issues in Science and Technology 14, no. 1 (1997): 81-86.
AA	Navas, A, and J Machin. "Assessing Erosion Risks in the Gypsiferous Steppe of Litigio (Ne Spain); an Approach Using Gis." Journal of Arid Environments 37, no. 3 (1997): 433-41.
AB	Barnhardt, Walter A, Joseph T Kelley, Stephen M Dickson, and Daniel F Belknap. "Mapping the Gulf of Maine with Side-Scan Sonar; a New Bottom-Type Classification for Complex Seafloors." Journal of Coastal Research 14, no. 2 (1998): 646-59.

AC	Cooper, J A G, and S McLaughlin. "Contemporary Multidisciplinary Approaches to Coastal Classification and Environmental Risk Analysis." Journal of Coastal Research 14, no. 2 (1998): 512-24.
AD	Chiueh, Pei Te, Shang Lien Lo, and Chin Dee Lee. "Prototype Sdss for Using Probability Analysis in Soil Contamination." Journal of Environmental Engineering 123, no. 5 (1997): 514-19.
AE	Siddiqui, Muhammad Z, Jess W Everett, and Baxter E Vieux. "Landfill Siting Using Geographic Information Systems; a Demonstration." Journal of Environmental Engineering 122, no. 6 (1996): 515-23.
AF	Cheng, Qiuming. "Spatial and Scaling Modelling for Geochemical Anomaly Separation." Journal of Geochemical Exploration 65, no. 3 (1999): 175-94.
AG	Verdin, K L, and J P Verdinb. "A Topological System for Delineation and Codification of the Earth's River Basins." Journal of Hydrology 218, no. 1-2 (1999): 1-12.
AH	Wang, Xinhao, and Zhi Yong Yin. "A Comparison of Drainage Networks Derived from Digital Elevation Models at Two Scales." Journal of Hydrology 210, no. 1-4 (1998): 221-41.
AI	Desmet, P J J, and G Govers. "A Gis Procedure for Automatically Calculating the Usle Ls Factor on Topographicaly Complex Landscape Units." Journal of Soil and Water Conservation 51, no. 5 (1996): 427-33.
AJ	Arnold, J G, R Srinivasan, R S Muttiah, and J R Williams. "Large Area Hydrologic Modeling and Assessment; Part I, Model Development." Journal of the American Water Resources Association 34, no. 1 (1998): 73-89.
AK	Srinivasan, R, T S Ramanarayananan, J G Arnold, and S T Bednarz. "Large Area Hydrologic Modeling and Assessment; Part Ii, Model Application." Journal of the American Water Resources Association 34, no. 1 (1998): 91-101.
AL	Blaszczynski, Jacek S. "Landform Characterization with Geographic Information Systems." Photogrammetric Engineering and Remote Sensing 63, no. 2 (1997): 183-91.
AM	Chung, Chang Jo F, and Andrea G Fabbri. "Probabilistic Prediction Models for Landslide Hazard Mapping." Photogrammetric Engineering and Remote Sensing 65, no. 12 (1999): 1389-99.
AN	Varekamp, C, A K Skidmore, and P A B Burrough. "Using Public Domain Geostatistical and Gis Software for Spatial Interpolation." Photogrammetric Engineering and Remote Sensing 62, no. 7 (1996): 845-54.
AO	Whitman, Dean, Timothy Gubbels, and Linda Powell. "Spatial Interrelationships between Lake Elevations, Water Tables, and Sinkhole Occurrence in Central Florida; a Gis Approach." Photogrammetric Engineering and Remote Sensing 65, no. 10 (1999): 1169-78.
AP	Newson, M D, and C L Newson. "Geomorphology, Ecology and River Channel Habitat; Mesoscale Approaches to Basin-Scale Challenges." Progress in Physical Geography 24, no. 2 (2000): 195-217.

AQ	Duncan, Christopher C, Andrew J Klein, Jeffrey G Masek, and Bryan L Isacks. "Comparison of Late Pleistocene and Modern Glacier Extents in Central Nepal Based on Digital Elevation Data and Satellite Imagery." Quaternary Research (New York) 49, no. 3 (1998): 241-54.
AR.	Li, Zhen, Wenxin Sun, and Qunzhu Zeng. "Measurements of Glacier Variation in the Tibetan Plateau Using Landsat Data." Remote Sensing of Environment 63, no. 3 (1998): 258-64.
AS	Frangi, J P, and D Richard. "Heavy Metal Soil Pollution Cartography in Northern France." The Science of the Total Environment 205, no. 1 (1997): 71-79.
AT	Korre, A. "Statistical and Spatial Assessment of Soil Heavy Metal Contamination in Areas of Poorly Recorded, Complex Sources of Pollution; Part 2, Canonical Correlation Analysis and Gis for the Assessment of Contamination Sources." Stochastic Environmental Research and Risk Assessment 13, no. 4 (1999): 288-316.
AU	Sharma, K D, M Menenti, J Huygen, and A Vich. "Modeling Spatial Sediment Delivery in an Arid Region Using Thematic Mapper Data and Gis." Transactions of the ASAE 39, no. 2 (1996): 551-57.
AV	Tim, U Sunday. "Emerging Technologies for Hydrologic and Water Quality Modeling Research." Transactions of the ASAE 39, no. 2 (1996): 465-76.
AW	Adamus, Christine L, and Martinus J Bergman. "Estimating Nonpoint Source Pollution Loads with a Gis Screening Model." Water Resources Bulletin 31, no. 4 (1995): 647-55.
AX	Julien, Pierre Y, Bahram Saghafian, and Fred L Ogden. "Raster-Based Hydrologic Modeling of Spatially-Varied Surface Runoff." Water Resources Bulletin 31, no. 3 (1995): 523-36.
AY	Correia, Francisco Nunes, Maria da Graca Saraiva, da Silva Fernando Nunes, and Isabel Ramos. "Flood Plain Management in Urban Developing Areas; Part Ii, Gis-Based Flood Analysis and Urban Growth Modelling." Water Resources Management 13, no. 1 (1999): 23-37.
AZ	Zhu, A Xing. "Mapping Soil Landscape as Spatial Continua; the Neural Network Approach." Water Resources Research 36, no. 3 (2000): 663-77.
BA	Adamus, Christine L, and Martinus J Bergman. "Estimating Nonpoint Source Pollution Loads with a Gis Screening Model." Water Resources Bulletin 31, no. 4 (1995): 647-55.
BB	Delong, Michael D, and Merlyn A Brusven. "Classification and Spatial Mapping of Riparian Habitat with Applications toward Management of Streams Impacted by Nonpoint Source Pollution." Environmental Management (New York) 15, no. 4 (1991): 565-71.
BC	Eash, David A. "A Geographic Information System Procedure to Quantify Drainage-Basin Characteristics." Water Resources Bulletin 30, no. 1 (1994): 1-8.
BD	Hammen, John L, and Philip J Gerla. "A Geographic Information Systems Approach to Wellhead Protection." Water Resources Bulletin 30, no. 5 (1994): 833-39.

	(1994): 833-39.
BE	Heinzmann, Uta, and Gaby Zollinger. "Validation of Representativeness with Relief Parameters Based on the Comparison of Two Landuse Classifications." Catena (Giessen) 24, no. 1 (1995): 69-87.
BF	Julien, Pierre Y, Bahram Saghafian, and Fred L Ogden. "Raster-Based Hydrologic Modeling of Spatially-Varied Surface Runoff." Water Resources Bulletin 31, no. 3 (1995): 523-36.
BG	Lark, R M, and P H T Beckett. "A Regular Pattern in the Relative Areas of Soil Profile Classes and Possible Applications in Reconnaissance Soil Survey." Geoderma 68, no. 1-2 (1995): 27-37.
BH	Lathrop, Richard G, Jr., John D Aber, and John A Bognar. "Spatial Variability of Digital Soil Maps and Its Impact on Regional Ecosystem Modeling." Ecological Modelling 82, no. 1 (1995): 1-10.
BI	Mitchell, James E. "The Influence of (X,Y) Uncertainty on Prediction Error and Contour Lines from a Three-Dimensional Surface." Water Resources Bulletin 29, no. 5 (1993): 863-70.
BJ	Odeh, Inakwu O A, Alex B McBratney, and David J Chittleborough. "Fuzzy-C-Means and Kriging for Mapping Soil as a Continuous System." Soil Science Society of America Journal 56, no. 6 (1992): 1848-54.
BK	Orzol, Leonard L, and Timothy S McGrath. "Summary of Modifications of the U. S. Geological Survey Modular, Finite-Difference, Ground-Water Flow Model to Read and Write Geographic Information System Files." Water Resources Bulletin 29, no. 5 (1993): 843-46.
BL	Richards, Christopher J, Honesto Roaza, and Roaza Ruth Montgomery. "Integrating Geographic Information Systems and Modflow for Ground Water Resource Assessments." Water Resources Bulletin 29, no. 5 (1993): 847-53.
BM	Spaling, Harry. "Analyzing Cumulative Environmental Effects of Agricultural Land Drainage in Southern Ontario, Canada." Agriculture, Ecosystems and Environment 53, no. 3 (1994): 279-92.
BN	Stein, A, C Varekamp, Egmond C van, and Zoest R van. "Zinc Concentrations in Groundwater at Different Scales." Journal of Environmental Quality 24, no. 6 (1995): 1205-14.
BO	Tachikawa, Y, M Shiiba, and T Takasao. "Development of a Basin Geomorphic Information System Using a Tin-Dem Data Structure." Water Resources Bulletin 30, no. 1 (1994): 9-17.
BP	Tim, U S, S Mostaghimi, and V O Shanholtz. "Identification of Critical Nonpoint Pollution Source Areas Using Geographic Information Systems and Water Quality Modeling." Water Resources Bulletin 28, no. 5 (1992): 877-87.

Appendix B: Articles establishing couplings (cited by at least 2 articles)

BOGGY CREEK WATER QU 1987
CELL BASED MODELING 1992 DISTRICT WATER MANAG 1994
FL DOT550010001A DEP 1985
NATIONAL ENG HDB 1972
PHASE I REPORT MASTE 1992
STATISTICAL ANAL SYS 1985 URBAN STORMWATER ANA 1990
*FAO FAO SOILS B 32 1976
*NAT RES COUNC GROUND WAT VULN ASS 1993
*SAS I INC SAS STAT US GUID VER 1686 1990
*USDA NAT SOIL HDB 1983
ABBOTT MB J HYDROL 87 61 1986
ABER JD ECOL MODEL 67 37 1993
ABER JD OECOLOGIA 92 463 1992
ADAMUS CL 1 ST JOHNS RIV WAT M AKIMA H ACM T MATH SOFTWARE 22 362 1996
ARNOLD JG J HYDROL 142 47 1993
ARNOLD RW SSSA SPEC PUBL 28 1 1991
ARONOFF S GEOGRAPHIC INFORMATI 1989
AYDIN K IN PRESS J APPL METE 34 1995
BAILEY T INTERACTIVE SPATIAL 1995
BAND LE ECOL MODEL 56 171 1991
BECKETT PHT J SOIL SCI 22 466 1971
BECKETT PHT SOIL FERTILIZERS 34 1 1971
BEVEN KJ HYDROL SCI B 24 43 1979
BLISS NB J SOIL WATER CONSERV 44 30 1989
BONHAMCARTER GF GEOGRAPHIC INFORMATI 1994
BRESLER E WATER RESOUR RES 24 381 1988
BURGESS TM J SOIL SCI 31 315 1980
BURKE IC BIOSCIENCE 41 685 1991
BURKE IC LANDSCAPE ECOL 4 45 1990
BURROUGH PA J SOIL SCI 22 368 1971
BURROUGH PA THESIS U OXFORD 1969
BURROUGH PA PRINCIPLES GEOGRAPHI 1986
BURROUGH PA J SOIL SCI 43 193 1992
BURROUGH PA J SOIL SCI 40 477 1989
BURROUGH PA SSSA SPECIAL PUBLICA 48 19 1996
BUTLER BE SOIL CLASSIFICATION 1980
CHANDRASEKAR V J APPL METEOROL 32 1288 1993
CHANDRASEKAR V J ATMOS OCEAN TECH 7 621 1990
CHOI GW WATER RESOUR RES 29 321 1993
CHOW VT APPL HYDROLOGY 1988
CHOW VT HDB APPLIED HYDROLOG 1964
CHRISMAN NR CARTOGRAPHICA 21 79 1984
CLINE TJ THESIS COLORADO STAT 1988

CLINE TJ WATER RESOUR BULL 25 641 1989
CRESSIE N ENV MODELING GIS 414 1993
CRESSIE N MATH GEOL 17 563 1985
CRESSIE NAC STAT SPATIAL DATA 1991
DELREGNO KJ LAKE RESERVOIR MANAG 4 17 1988
DEUTSCH CV 18 GSLIB GEOSTATISTICAL 340 1992
DEVANTIER BA J WATER RES PL-ASCE 119 246 1993
DICKINSON WT 25 COL STATE U HYDR 1967
DOE WW 7TH P ANN GRASS US C 91 1992
DOE WW THESIS COLORADO STAT 1992
DRAPER NR APPL REGRESSION ANAL 1981
EVANS BM J SOIL WATER CONSERV 45 242 1990
FEDRA K IAHS PUB 211 1993
FISHER PF INT J GEOGR INF SYST 5 193 1991
GERSMEHL PJ CARTOGRAPHICA 18 79 1981
GOODCHILD MF ENV MODELING GIS 1993
GOODRICH DC WATER RESOUR RES 27 995 1991
GORGUCCI E J ATMOS OCEAN TECH 11 586 1994
GUSTAFSON DI ENVIRON TOXICOL CHEM 8 339 1989
HAINING R SPATIAL DATA ANAL SO 1990
HALL FG LANDSCAPE ECOLOGY 2 3 1988
HALL GF SSSA SPEC PUBL 28 9 1991
HARLIN JM GEOGRAPHIC INFORMATI 1993
HARPER HH ESTIMATION STORMWATE 1992
HE CS WATER RESOUR BULL 29 891 1993
HESSION WC J SOIL WATER CONSERV 43 264 1988
ISAAKS EH INTRO APPL GEOSTATIS 561 1989
IZUNO FT J ENVIRON QUAL 20 608 1991
JAMES WP WATER RESOUR BULL 26 587 1990
JARVIS MG SOILS OXFORDSHIRE 1 1982
JENSON SK PHOTOGRAMM ENG REMOT 54 1593 1988
JETT SC P HYDR TRANSP S ST J 127 1979
JOHNSON BE USGS934018 WAT RES I 9 1993
JOURNEL AG MATH GEOL 18 119 1986
JOURNEL AG MINING GEOSTATISTICS 600 1978
JULIEN PY CER9091PYJBS12 COL S 1991
JURY WA VADOSE ZONE MODELING 245 1986
JURY WA J ENVIRON QUAL 12 558 1983
JURY WA J ENVIRON QUAL 16 422 1987
KELLOGG RL J SOIL WATER CONSERV 49 294 1994
KHAN MA ENVIRON MANAGE 13 233 1989
KLAZURA GE B AM METEOROL SOC 74 1293 1993
KNISEL WG 26 USDA CONS RES 1980
KOVAR K IAHS PUBLICATION 211 1993
LARK RM IN PRESS INT J REMOT 1995
LASLETT GM J SOIL SCI 38 325 1987

LEENHARDT D EUR J SOIL SCI 45 293 1994
LEONARD RA T ASAE 30 1403 1987
LOAGUE K FIELD SCALE SOLUTE W 175 1990
LOAGUE K J CONTAM HYDROL 8 157 1991
LOAGUE K J CONTAM HYDROL 5 171 1990
LOAGUE K J CONTAM HYDROL 5 405 1990
LOAGUE K J ENVIRON QUAL 25 475 1996
MACLEAN AL PHOTOGRAMM ENG REM S 59 223 1993
MAIDMENT DR ENV MODELING GIS 147 1993
MANDELBROT BB FRACTAL GEOMETRY NAT 1982
MARCUS KB 2 DIMENSIONAL FINITE 1991
MARSMAN B SOILS SHEET SP30 W 2 1973
MCBRATNEY AB J SOIL SCI 43 159 1992
MCBRATNEY AB COMPUT GEOSCI 7 331 1981
MCCORMACK JE INT J GEOGR INF SYST 7 263 1993
MEEKS YJ J WATER RES PL-ASCE 116 693 1990
MEIN RG MODELING INFILTRATIO 43 1971
MITCHELL CW TERRAIN EVALUATION 1973
MITCHELL JK WATER RESOUR BULL 29 833 1993
MOORE ID HYDROL PROCESS 5 3 1991
MULLA DJ J ENVIRON QUAL 25 419 1996
NASH JE J HYDROL 10 282 1970
NIELSEN DR HILGARDIA 42 215 1973
OCALLAGHAN JF COMPUT VISION GRAPH 28 323 1984
ODEH IOA GEODERMA 47 93 1990
ODEH IOA SOIL SCI SOC AM J 56 505 1992
OGDEN FL J HYDROL 158 1 1994
OGDEN FL THESIS COLORADO STAT 1992
OGDEN FL WATER RESOUR RES 29 2584 1993
OLOUGHLIN EM WATER RESOUR RES 22 794 1986
PECK AJ WATER RESOUR RES 13 348 1977
PELLETIER RE J SOIL WATER CONSERV 40 332 1985
PETIT AN APPL STAT 42 185 1993
PICKUS J GEO INFO SYSTEMS NOV 50 1992
PRESTON FW ECOLOGY 43 185 1962
QUINN P HYDROL PROCESS 5 59 1991
RAO DV SJ891 ST JOHNS RIV W 1989
RAO PSC ENV IMPACT NONPOINT 23 1980
RAO PSC P SOIL CROP SCI SOC 44 1 1985
RAWLS WJ T ASAE 25 1316 1982
RAWLS WJ J HYDRAUL ENG-ASCE 109 62 1983
REYBOLD WU J SOIL WATER CONSERV 44 28 1989
RIFAI HS CIVIL ENG 63 44 1993
ROGOWSKI AS WATER RESOUR RES 8 1015 1972
ROWLANDS NJ MATH GEOL 9 383 1977
RUNDQUIST DC GEOCARTO INT 2 51 1991

RUNNING SW ECOLOGY 70 1090 1989
SACHIDANANDA M J ATMOS OCEAN TECH 4 588 1987
SAGHAFIAN B THESIS COLORADO STAT 1992
SCHAAKE JC THESIS J HOPKINS U B 1965
SCOTT RM LAND SYSTEMS ATLAS W 1971
SHARMA ML J HYDROL 45 101 1980
SHARMA ML WATER RESOUR RES 15 1567 1979
SKIDMORE AK INT J GEOGR INF SYST 5 431 1991
SRINIVASAN R WATER RESOUR BULL 30 441 1994
SRINIVASAN R WATER RESOUR BULL 30 453 1994
TAYFUR G J HYDRAUL ENG-ASCE 119 51 1993
TESO RR J ENVIRON QUAL 25 425 1996
TIM US J ENVIRON QUAL 23 25 1994
TIM US WATER RESOUR BULL 28 877 1992
VANDEURSEN WPA PC RASTER PACKAGE 1992
VENTURA SJ WATER RESOUR BULL 29 189 1993
VIEUX BE HYDROL PROCESS 5 101 1991
VILORIA J THESIS U OXFORD 1989
VOLTZ M J SOIL SCI 41 473 1990
WARRICK AW WATER RESOUR RES 13 355 1977
WEBSTER R NATURE 219 680 1968
WEBSTER R STAT METHODS SOIL LA 1990
WEBSTER R SOIL SURVEY TECHNICA 12 1979
WISCHMEIER WH USDA AGR HDB 537 1978
WOOD EF J HYDROL 102 29 1988
WOOLHISER DA T ASAE 12 460 1969
WOOLHISER DA UNSTEADY FLOW OPEN C 2 1975
YOUNG RA 35 USDA CONS RES 1987
ZAWADZKI I ATMOS OCEAN 20 158 1982
ZEVENBERGEN LW EARTH SURF PROCESSES 12 47 1987
ZEVENBERGEN AW INT J REMOTE SENS 9 495 1988
ZHANG WH WATER RESOUR RES 25 2019 1989

Appendix C: Bibliographic Couplings

Please note that BH (J), BG (U), BA (AW), and BF (AX) were duplicates and have been removed; no citations: B, L, Z; no couplings: D, E, G, I, M, N, S, AB, AC, AE, AP, AQ, AR, BB, BC, BE, BK, BL, BM.

A,K
A,R
A,U
A,AD
A,AN
A,AS
A,AT
C, AI
C,AL
C,AT
F,H
F,AG
F,AH
F,AL
F,BO
H,J
H,R
H,X
H,AG
H,AI
H,AJ
H,AL
H,AZ
H,BI
H,BO
J,R
J,T
J,U
J,AN
J,BJ
K,R
K,AD
K,AN
K,AS
K,AT
O,AA
P,R
P,AK
P.AU
P,AV
P,AW
P,BP
Q,V
Q,W
Q,AN
Q,BN,
Q,BP
R,T
R,U
R,V
R,W
R,AD
R,AN
R,AS
R,AT
R,AV
R,AZ
R,BD
R,BJ
T,AM
T,AN
T,AS
T,AT
T,BN
U,AT
U,BJ
V,W
V,AJ
V,AK
V,AN
V,AU
V,AV
V,AX
V,BD
V,BJ
V,BP
W,AD
X,AI
X,AJ
X,AL
X,AZ
X,BO
Y,AF
AA,AV
AA,AW
AD,AM
AD,AN
AD,AS
AD,AT
AD,AY
AG,AL
AH,AI
AH,AK
AH,BO
AI,AJ
AI,AL
AI,AU
AI,BO
AI,BP
AJ,AK
AJ,AU
AJ,AV,
AJ,AX
AJ,AZ
AJ,BP
AK,AU
AK,AV
AK,BP
AL,BO
AM,AO
AM,AU
AM,AY
AN,AS
AN,AT
AN,BJ
AN,BN
AN,BP
AO,AU
AS,BN
AU,AV
AU,AX
AU,BO
AU,BP
AV,AW
AV,AX
AV,AY
AV,AZ
AV,BO
AV,BP
AX,BO
AX,BP
AZ,BJ
BJ,BN
BJ,BP
BN,BP