Available online at www.sciencedirect.com
SCIENCE DIRECT®
Information Processing and Management 42 (2006) 805–825
INFORMATION PROCESSING & MANAGEMENT
www.elsevier.com/locate/infoproman

Collaborative bibliography

David G. Hendry [a,*], J.R. Jenkins [b], Joseph F. McCarthy [c]

[a] *Information School, University of Washington, Box 352840, Suite 370 Mary Gates Hall, Seattle, WA 98195-2840, USA*
[b] *Serials Solutions, 444 NE Ravenna Blvd, Suite 211 Seattle, WA 98115, USA*
[c] *Interrelativity, Inc, 16705 226th Avenue NE, Woodinville, WA 98077, USA*

Received 10 November 2004; accepted 16 May 2005
Available online 12 July 2005

Abstract

A bibliography is traditionally characterized by the judgments, bounded by explicit selection criteria, made by a single compiler. Because these criteria concern the attributes ascribed to a work and the needs of readers, bibliographic work is largely conceptual even across technological eras and domains. Yet, the development of networked information services, made possible by WWW infrastructure, has enabled very large numbers of people to discover, organize, and publish information, including bibliographies. Indeed, bibliographies, or at least bibliography-like artifacts, are a common genre of website, often published by people without specialized skills in information organization who follow non-rigorous selection procedures. Nevertheless, even if the items from these lists are poorly selected and described, this publishing activity is fundamentally important because it structures information locally, creating a patchy network of secondary access points. In turn, these access points enable information discovery, the formation and development of communities of interest, the estimation of document relevance by search engines, and so on. In sum, this activity, and the enabling technical infrastructure, invites bibliographies to take on new interactive possibilities. The aim of this article is to extend the traditional view of bibliography to encompass collaborative possibilities for wide, or narrow, participation in the shaping of bibliographies and the selection of items. This is done by examining the nature of bibliography on the Web, by proposing a conceptual model that opens bibliography to participatory practices, and by discussing a case study where a team sought to develop a bibliography of electronic resources. This examination reveals splendid opportunities for expanding the notion of bibliography with participatory policies while remaining true to its ancient roots.

Keywords: Bibliography; Collection development; Computer-supported cooperative work; Electronic resource collections; Information gateways

* Corresponding author. Tel.: +1 206 616 2316; fax: +1 206 616 3152.
E-mail addresses: dhendry@u.washington.edu (D.G. Hendry), jr@serialssolutions.com (J.R. Jenkins), joe@interrelativity.com (J.F. McCarthy).

0306-4573/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ipm.2005.05.007

1. Introduction

In January 2004, over 50 million people in the USA published content to the Web, including photographs, messages, and files of various kinds and approximately 5 million people maintained a website (Lenhart, Horrigan, & Fallows, 2004). While this survey-based research did not examine the motivations, objectives and tasks surrounding this publication activity, we can assume that a significant number of people, perhaps millions of people, invested time in creating outbound links to websites. We can consider the work associated with selecting links, writing titles and descriptions, arranging the items on Web pages, and maintaining the integrity of the links a form of bibliographic work. Given the ancient history of bibliography and its importance in library studies and literary scholarship (Francis, 1973; Stokes, 2003), this position is admittedly idealistic. Nevertheless, we shall argue that making this conceptual move—that is, analyzing the production and organization of information in terms of the standards of bibliography—leads to productive ways for examining popular classes of information artifacts and for conceptualizing improved tools for the collection, and organization of electronic resources.

Websites, of course, serve different purposes and audiences and therefore present selective material and experiential opportunities. Thus, we can expect that the process of selecting, describing, and organizing links to electronic resources will vary depending on the genre of the website. The term genre is used by Web designers to characterize classes of websites that share critical features, such as personal e-commerce, daily news, community conference and so on (Crowston & Williams, 1997; Van Duyne, Landay, & Hong, 2003; Yates & Orlikowski, 1992). With experience, readers are assumed to readily identify the genre of a website and orient their expectations concerning the nature of the material, information-seeking approaches, and options for interaction. When the genre of a website is difficult to detect, it can be disorienting to the reader. At the same time, the genre places constraints on the form and topic of the communication and the nature of the audience. Such constraints simplify the work of the website developer. For developers with experience in bibliography or collection development, these constraints might be carefully articulated in policies and procedures. More typically, however, the constraints of the genre, while being important for readers and readily observable through signals of various kinds, are implicit and not written down in policies. Thus, for example, a teenager's personal Web log of daily happenings, known as a *blog* (Blood, 2000, 2004), is likely to be readily distinguished from an elementary school teacher's list of homework links suitable to a 5th grade class. While the visual design and editorial style are salient signals of a genre so too is the nature and form of the outbound links. In short, the elementary school teacher will employ selection criteria and structuring and formatting standards that support his or her pedagogical objectives. Students might be asked to visit some links in a particular order whereas other links will be optional; at the same time, link descriptions and cited sources will be coordinated with the curriculum. The blogger, on the other hand, has a different audience in mind and different intentions for including outbound links in her entries. Nevertheless, in some important senses these are both bibliographies because pointers to resources are collected, collocated, and arranged. Indeed, the early blogs of circa 1998 were bibliography-like, rather than personal commentaries, with the twist that they focused on fast breaking news and newly available information (Blood, 2000). Still, today among the many journal-like blogs are blogs that predominantly contain lists of links with pithy descriptions that act as instruments for information discovery (Bar-IIan, 2004; Schiano, Nardi, Gumbrecht, & Swartz, 2004).

A common website genre, referred to as *resource page* in this article, draws significant benefit from bibliographic practices. We define a resource page as a website that contains an organized set of outbound links on a given topic. (Of course, a resource page might consist of multiple pages and frequently a resource page is one element of a website with a broader purpose.) Librarians might use the term *directory* to refer to this concept. The etymology of the term is unclear but may have been coined in distinction to *home page*, which typically contains personal information, by the early users of the Web who needed a genre to capture material concerned with projects or topics. A Google exact-match search for *Resource Page* restricted to

page titles turns up approximately 39 000 evidently distinct sites (search performed 1 July 2004). In general, the status assigned to resource pages by information professionals is often low because professional practices for their creation and maintenance are often not followed.

As a genre, resource pages are centrally concerned with the description and organization of links. Labels that may indicate other neighboring and overlapping genres include hotlist, portal, gateway, directory, webliography, WebQuest, and so on. The categories denoted by these labels are certainly fuzzy and in flux, and we know of no systematic attempt to clearly define these terms, although the importance of formal definitions of systems that make up this space has been noted (Goncalves, Fox, Watson, & Kipp, 2004; Smith, 2004). A further definitional difficulty is that many websites that serve largely bibliographic purposes draw upon other collaborative and interactive features such as message boards, blogs, chat rooms, and so on.

Resource pages are often developed by knowledgeable, and sometimes fanatic, individuals who gather and organize material concerned with a common subject. Like a bibliography, a resource page enables people to readily discover items relevant to their information needs by collocating selective materials on a given subject area. A prototypic example is the Human–Computer Interaction Bibliography (Perlman, 2004). This project, lead by Gary Perlman since its inception in 1988, organizes various resources, including printed and online-only materials, of the field into such categories as companies, consultants, education, online columns by usability experts, and so on. In addition, approximately 28 500 entries, mostly citations to academic papers, were added to the bibliography by a grassroots effort outside of the publishing industry. Volunteers have helped sustain the bibliography since 1988 by submitting and verifying items (Perlman, 1999). While the advent of the ACM Digital Library (ACM, 2004) has rendered the bibliography of published papers somewhat obsolete, it nevertheless collocates materials, especially from non-ACM journals and third party resources, in very useful presentations.

As a case study, the HCI Bibliography can be used to illustrate four general claims about the evolution of resource pages. First, resource pages are often started by a single person, the director, who has an interest in a subject area but is typically not trained in bibliographic work. Second, the director of the resource page may recruit help in making the collection grow and thus he or she may organize a community of contributors around the development of the resource page. Third, in general, to maintain a high degree of relevance the resource page must evolve, and this is especially the case for subject areas that are themselves evolving rapidly. Forth, in addition to supporting information finding, a resource page can support the development of communities of interest, enabling groups to socialize and collaborate. At this general level of description, the development of resource pages appears to be relatively straightforward. We shall see, however, that this is not the case.

The purpose of this article is to examine the conceptual and technical issues associated with developing sustainable bibliographies of electronic resources by non-specialists. It will be argued that this is an important area of inquiry because many non-professionals, working alone and in small groups, are creating bibliographies yet we know little about nature of these artifacts and the practices used to develop and sustain them. Further, this appears to be a long-term phenomenon which the Web enables. In the next section, we establish the background for this inquiry, beginning with a discussion of professional bibliographic practice, examining the case that bibliography-like publishing is extremely common on the Web, and reviewing the literature on electronic collection development, which, in general, is critical of the efforts of non-professionals. In Section 3, we propose a framework for organizing the full range of online bibliographic efforts and sample several key positions in this space. This analysis is then abstracted into a conceptual model that extends the traditional conceptualization of bibliography to include the notion of participatory development. In Section 4, we then use this model to reflect upon our own efforts to develop a resource page. This case study allows us to factor out the technical issues associated with the creation, maintenance, and publishing of bibliographies from lasting conceptual issues which, in turn, are more readily identified and fully understood. The article concludes with a discussion of the design challenges associated with the development of tools that support the participatory development of bibliographies.

2. Background

2.1. Professional bibliographic practices

Francis (1973) introduces the practice of bibliography thus: "The tasks of the compiler of a bibliography are (1) to find what books on a particular subject exist; (2) to describe them item by item; and (3) to assemble the resulting entries into useful arrangements for reference and study" (p. 978). This deceptively simple procedure reveals a remarkable degree of conceptual universality. By replacing the word *book* with *scroll* or *Web page* we can see that bibliography applies equally well in the age of Alexandria and in the age of the dot.com, and by clarifying the notion of *reference and study* we can, in principle, address any information need and context. Further, it is a procedure that is accessible to anyone with a basic level of literacy and some familiarity with writing and publishing tools. The only essential prerequisite is the motivation to collect and list. Yet, clearly, this procedure alone does not guarantee the quality, or more specifically the fitness of purpose, usefulness, or versatility of the resulting bibliography on a particular subject. To meet these quality criteria, a great deal of judgment is required.

Drawing on the theoretic work of Wilson (1968), Bates (1976) clarifies this judgment for the professional librarian. She defines a bibliography as "A list or sequence of descriptions of graphic materials on a given subject or area." (p. 9) and proceeds to present, in splendid detail, a decision-making process that leads to *rigorous systematic bibliography*. The central claim of this process is that the specifications for creating a bibliography, and similar information artifacts such as directories, must be made explicit and rigorously followed during its creation. As represented in Fig. 1, this process is decomposed into six types of specification. According to Bates, the task of the bibliographer is to make explicit decisions concerned with each type of specification and to state these decisions and their rationale to the readers of the bibliography. When this is done, readers are able to understand the purpose and coverage of the bibliography and this contextual information informs readers for how they should employ it for information discovery.

Like documents in general, bibliographies have a logical and physical structure (Andre, Furuta, & Quint, 1989). The logical structure consists of the explicitly modeled entities, attributes, and relationships that formally specify the bibliography for computational representation and manipulation whereas the physical structure concerns the presentation of the bibliography, that is, its appearance. Bates (1976) does not clearly identify this distinction and, indeed, for her audience of information professionals in the mid 70s, developing bibliographies for print, it is irrelevant. Perhaps reflecting this period, Bates uses the terms "list or

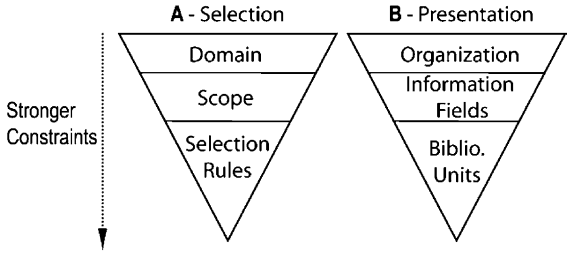

Fig. 1. Bates (1976) specification types for bibliography (after Wilson, 1968). This figure structures Bates' six specifications types. The top half of the figure shows the specifications that enable conceptual narrowing where *domain* concerns the space of information searched, *scope* concerns the topical boundaries for the bibliography, and *selection rules* concern the rules for selecting specific items to appear in the bibliography. Thus, the selection process is decomposed into three levels with each lower level applying stronger selection constraints than above. The left part of the figure shows the specification types that enable effective presentations of bibliographies, where decisions about the *organization* of the bibliography place constraints on the *information fields* to be used for each item, which, in turn, place constraints on the *bibliographic unit* (i.e., granularity) of the item to be described. The two half's of the figure, (A) and (B), meet at the individual bibliographic unit; later in Fig. 3, we shall add a third triangle, (C), to this figure.

sequence'' to characterize the arrangement of bibliographic items but, of course, other arrangements might be useful; for example, a node-link graph of items indicating intellectual borrowing, such as the beautifully rendered chart of influence relationships amongst popular styles of music and album sales between 1955 and 1974 (Tufte, 1997, p. 90). This observation, however, is not new. Bradshaw, a seminal figure in descriptive bibliography, wrote in 1870: ''The books are arranged strictly according to their respective countries, towns and presses; and, the year of the earliest dated production of each country, town or press, being affixed, it is hoped that the arrangement will at once afford its own explanation'' (Stokes, 2003, p. 283). Thus, different arrangements reveal different information because they collocate items in different ways, the classic arrangements being alphabetical by author or chronological by publication date.

In general, people can happily ignore the distinction while nevertheless benefiting from it. On the other hand, computationally, it is crucial because the logical structure of a bibliography, or bibliography-like artifact, constrains the range of physical arrangements that can be presented. Indeed, this is fundamental in the field of information visualization (Card, Mackinlay, & Shneiderman, 1999), which focuses on techniques for creating dynamic displays that support specific tasks. To illustrate the potential payoff of examining the logical structure of familiar information artifacts for bibliography-like forms, consider a listserv where a moderator sends out weekly messages concerning some well defined topic. Often, these messages will contain links to relevant resources. Within an archive of such messages lies a kind of bibliography that is not easily recognized as such because the links are intermixed within other written material. Yet, by collecting the links together in one place and by describing and formatting the items in a consistent fashion this otherwise latent structure surfaces. In turn, through a single list of links, the archive of messages is potentially made more valuable.

Returning to the model, Bates (1976) paints a picture of a lone compiler working diligently, even painstakingly, to select items to be presented in the relatively fixed form of print. She does not address the possibility of sustaining bibliographies over time by, for example, outlining specification types that cover such matters as collecting and responding to implicit and explicit audience feedback, controlling the evolution of the bibliography's scope, highlighting recent selections and de-selections, and so on. Yet, such matters can be very important. Certainly, this is the case when specifications of scope are provisional or contentious, when the number of potential items is large and the relative value of an item can change quickly as a function of time, or when items are subject to change and authoritative sources for their discovery do not exist. Settings where such constraints apply are common. Examples include bibliographies intended to support the formation of a new field (e.g., Ubiquitous Computing), to support decision-making in design and development contexts, or to support reflection or discussion on a provisional body of materials. Characteristic of these settings is fluidity.

Broadening the picture to an institutional level, approaches for sustaining collections can be readily identified in collection development policies (e.g., Evans, 2000; Mack, 2003; Orr, 2003). Fundamentally, these policies set out management plans for how an institution will engage its patron community. These plans cover such matters as how the institution will budget for and prioritize the collection of materials, how items are suggested, selected and de-selected and by whom, how disagreements about selections will be resolved, and so on. It must also be acknowledged, however, that creating and maintaining a collection development policy is exceedingly difficult, even for elite institutions (e.g., see Spohrer, 2003).

When a resource page is published and attracts an audience, it becomes a dynamic social artifact. People visit it and might want to discuss it and contribute to it. Interestingly, resource pages simultaneously draw on the properties of bibliographies and collection development policies. We shall see that this is the key to understanding a wide range of bibliography-like artifacts that are found online.

2.2. Descriptions and critiques of online collections

For librarians, the Internet has thrown up pressing challenges for the collection and organization of documents, especially third party electronic resources (Levy & Marshall, 1995; Pitschmann, 2001; Torok, 2003;

Wells, Calcari, & Koplow, 1999). Janes et al. (1999, chap. 2) outline the issues and offer practice-oriented solutions for implementing electronic collections. For electronic resources, they give specific guidance on such matters as documenting selection policies, cataloging items, developing access points, implementing a working system with Web infrastructure, and sustaining the collection over time. Taking a conceptual stance, Lee (2000) discusses how technological advances, especially the infrastructure that enables digital libraries, have disrupted traditional notions of collections, including tangibility of resources, clarity of ownership, and uniformity of user community. In short, when so much is available through a commercial search engine, why collect? In Lee's view, disintermediation, that is, people using technology to discover resources rather than relying on humans or human-crafted artifacts presents a major challenge to online collections. Lee frames the controversy with two competing points of view, where the technological-optimistic view is held by one side. Under this view, it is predicted that in the future all information will be reachable digitally and that, with tools for search and organizing information, people will be able to directly access information that they need. Simply stated, information technology will become more powerful and usable, enabling end-users to work directly with primary sources. On the other side is the professional-practice view. Here, the position is that significant value, that is, more effective access to information for time expended, can be added by creating selective, highly focused collections. Simply stated, human expertise in information seeking, selection, and description add sufficient value to make intermediation viable.

A middle ground, not discussed, is that of non-professionals creating useful, broadly used collections, perhaps with the assistance of professionals. Borgman (2001) argues, and we agree, that current trends run contrary to the general predictions of disintermediation—librarians and information professionals are assuming critical roles in the development of digital libraries, in organizing resources, and assisting people in finding information. Yet, as discussed below, non-professionals seem to be playing interesting intermediation roles as well.

Reviews of online collections typically begin by asserting quality criteria for collections. Robinson and Bawden (1999) and Bawden and Robinson (2002), for example, drawing on common collection development standards (e.g., Evans, 2000), propose that Internet subject gateways should: (1) clearly describe the subject scope; (2) explicitly document the criteria for selecting resources; (3) describe or annotate all resources; (4) categorize or index all resources; and (5) clearly define the roles and responsibilities of people involved in the creation and maintenance of the collection. Then, the collections are reviewed against these criteria. In their review of academically oriented gateways, these writers found considerable variation in the degree that these criteria are met. Franco (2003) and Wells et al. (1999) give similar reviews.

When such criteria are used to evaluate resource pages prepared by non-professionals they, perhaps not surprisingly, are found to be inadequate. Kirriemuir (1999), for example, in a similar survey of gateways says "It is somewhat worrying that a ridiculously large number of people, organizations and other entities are still creating lists of links to resources, often with little or no quality criteria, standards for describing resources, or thought that they might be duplicating better work done elsewhere" (section, Excluded resource discovery systems). Yet, at the same time, Bawden and Robinson (2002) regrettably state: "[quality-controlled subject gateways] remain an under-used, under-valued, and, in large part, unknown resource, for Internet users... It remains an important task for information specialists to promote the value of these resources" (p. 161).

2.3. Bibliographic work by non-professionals

Little appears to be known about how non-professionals create bibliographies. In fact, we know of no studies that have investigated discretionary practices of "amateur bibliographers", people, without professional training, who compile lists of resources. A great number of resource pages have been published; yet, we seem to know little about who creates them, why are they created, what practices are followed, who uses them and for what purpose? While much literature relates to this activity, including personal information

management (e.g., Jones, 2004), information architecture (e.g., Garrett, 2002; Rosenfeld & Morville, 2002), and website design (e.g., Van Duyne et al., 2003), research has not addressed resource pages as a specific genre. This omission in the literature is puzzling. Perhaps, a major reason for it is simply due to the fact that lists, resource pages, and other bibliographies created by amateurs do not meet the normative standards of professionals.

Nevertheless, bibliography-like work appears to be a very common activity. Several sources of evidence back this claim. First, anecdotally when one visits personal home pages, pages created by community organizations, and so on, one invariably encounters lists of links under headings such as "related links", "my favorites", "see also", and so on. At commercial sites, such as amazon.com, users are prompted to submit lists of favorite books or favorite sites and descriptions in order to create interesting material (i.e., a bibliography) that attracts other customers. Finally, inspecting the archives of a newsgroup over an appropriate period of time will turn up scores of submitted links to potentially relevant resources. At a first level of approximation, these various lists of links are similar to the abstraction of trails, proposed by Vannevar Bush. Indeed, perhaps he anticipated much of this bibliographic activity when he said: "There [will be] a new profession of trail blazers, those who find delight in the task of establishing useful trails through the enormous mass of the common record" (Bush, 1945, Section 8, para. 3). The key observation is that many users create lists of descriptive pointers that adhere to some scope whether described or not.

Second, structural analyses of the Web, which have sought to identify cohesive citation structures with graph-algorithmic techniques, have found a fair degree of structure (Kleinberg & Lawrence, 2001). In fact, using data from 1998, Kumar, Raghavan, Rajagopalan, and Tomkins (1999), discovered approximately a hundred thousand cohesive communities as defined by a high degree of inter-site citation compared to intra-site citation. Random inspection of a small sample of these "communities" found that the cohesiveness of the linking structure was due to citation patterns around very narrow topics, such as resources on "oil spills off the coast of Japan" (Kumar et al., 1999, p. 403). Of course, popular search engines such as Google exploit this "latent" structure to improve the precision of search results (e.g., Google's PageRank algorithm, Brin & Page, 1998). And, in the future, the dynamic analysis of the structure of the scientific literature, especially in connection with message groups, blogs, download requests, and other online activ-

Table 1
Popularity estimate of resource page genre

Term	No. hits (1000s)
Directory	218 000
Collection	75 100
Portal	48 900
Gateway	19 600
Search Engine	10 600
Bibliography	9700
Pathfinder	3940
Hotlist	1050
Information Portal	260
Resource Directory	205
WebQuest	161
Information Directory	46
Resource Page	39
Information Gateway	19
Webliography	2

Shown are rough estimates of the number of websites that are in the resource page genre. These estimates were made by searching for the number of websites that contain an exact match for the target term in the page title. The search was performed on 1 July 2004 at Google using the query template *allintitle*: "*X*", where *X* is the term or phrase.

ities, will become more important for identifying important trends and assessing the impact of research (Henzinger & Lawrence, 2004; Kleinberg, 2004). An obvious point is that this macro structure emerges as people create localized micro-structures through the selection, description, and arrangement of links—bibliographic work is clearly important here!

A final sign is the frequency of use of the term *bibliography* and related terms (see Table 1). Informal samples of the resulting hits shows a wide range of bibliography-like information artifacts, which often appear to be created by non-professionals. Irrespective of the general quality of these lists of resources, they are important because they locally structure the Web and, in turn, enable the discovery of resources both by machines and by individuals.

In sum, the Web has enabled a class of amateur bibliographers to arise; yet, their products have not received systematic investigation. This observation is striking in light of the research being directed towards blogging (e.g., Bar-IIan, 2004; Blood, 2004; Lindahl & Blount, 2003; Schiano et al., 2004). In the next section, we propose a framework for organizing the space of collection development efforts and review a sample of projects.

3. Online bibliographic projects

Fig. 2 presents a framework for comparing collections where projects are plotted by the number of people who select resources versus the number of resources available. Because of its generality, we can chart a diverse range of projects and systematically sample projects. The figure shows four major regions: (1) the area where a relatively large number of people create large collections, (2) the area where individual people create collections of sizes which span five orders of magnitude, (3) the area where a large number of people create collections of small to moderate size, and (4) the area where a moderate number of people create collections of moderate size. Using this framework, we can examine, at least to a course grain, how labor is organized and tools employed to create collections of various sizes.

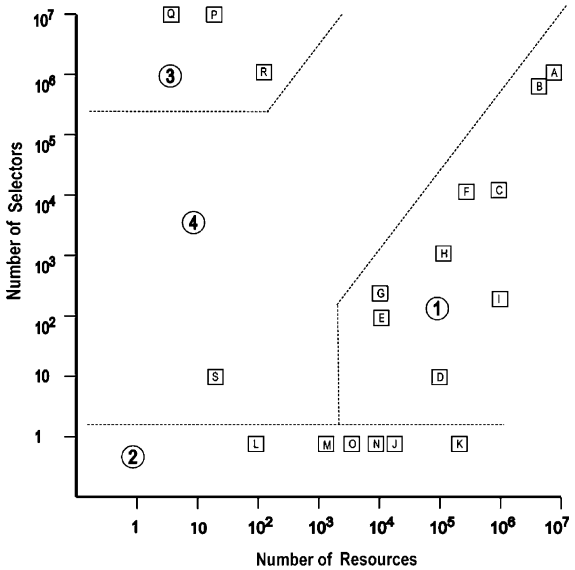

Fig. 2. Framework of bibliographic projects. This plot shows the relationships between the number of selectors participating in a bibliographic project versus the number of resources collected in the project.

This section consists of two parts. In the first part we sample selected projects from this framework in a purposive fashion; that is, we identify examples that illustrate different points in the framework. This sampling is neither exhaustive nor systematic except to the extent that we aim to identify a reasonable spread of systems. Then, for these examples we discuss basic issues concerning the organization of intellectual effort, the types of tools and structures used, and general conformance to traditional bibliographic practices. Next, in the second section, we propose a conceptual model that accounts for the participatory nature of many of the projects. Further, we show how this new model expands the traditional view of bibliography in an interesting way.

The estimates of the number of selectors and the number of resources are taken from statements contained in the site, for example, on the front page, About section or Help section. To improve the readability of this presentation, in general, only the front page for the site project is referenced in Table 2 and not specific pages. These estimates were first made January 2004 and verified in July 2004. Finally, capital letters, beginning with A, are used to reference particular sites below, in Fig. 2, and in Table 2.

3.1. Examples

Yahoo Groups (A) enable people to create Web-based environments for information management and for collaborating in small groups. Once registered, anyone can create a group that can be made open to the public or restricted to a list of members. Groups contain a variety of information management tools, including a bookmark tool for organizing links to third party resources, a photo-album for updating and organizing pictures, a directory for uploading and organizing files, and a database application for the definition and management of times that can be represented with records and fields (e.g., list of recipes, class assignments, etc.). In addition, chat, message boards, and a group calendar are available for enabling socialization and collaborative work. Approximately, two million groups, all sharing the Yahoo look-and-

Table 2
List of bibliography-like projects

Region #1	
A	Yahoo Groups ~ www.groups.yahoo.com
B	LiveJournal.com ~ www.livejournal.com
C	Open Directory Project ~ www.dmoz.org
D	Infomine ~ infomine.ucr.edu
E	Librarian's Index to the Internet ~ www.lii.org
F	Wikipedia ~ www.wikipedia.org
G	Virtual Library ~ www.vlib.org
H	U.S. GenWeb ~ www.usgenWeb.com
I	About.com ~ www.about.com
Region #2	
J	Human–Computer Interaction ~ www.hcibib.org
K	Cyndi's List ~ www.cyndislist.com
L	Champion Fishing Links Directory ~ www.championbass.com/links/
M	Encyclopedia of Psychology ~ www.psychology.org
N	Esoterism.com Search Engine ~ esoterism.com/links/
O	Ubiquitous Computing Resource Page ~ www.ucrp.org
Regions #3 and #4	
P	Lycos Fifty ~ 50.lycos.com
Q	Google Zeitgeist ~ www.google.com/press/zeitgeist.html
R	Technorati ~ www.technorati.com
S	SlashDot ~ www.slashdot.com

The identifiers, A–S, correspond to the points in Fig. 2.

feel, have been created and are organized into a directory, evidently by Yahoo's editors. Assuming that 10–100 resources are collected within each club gives an impressive estimate of the total number of resources collected. Yahoo Groups, in short, illustrates what can happen when localized selectors and centralized integrators collaborate. Of course, this does not mean that items are easily found. LiveJournal (B), a community of blogs, is of similar scale albeit blogs have few provisions for collecting resources, and more generally, information management.

The Open Directory Project (C), at dmoz.org, aims to be the definitive catalog of Web resources by recruiting volunteers to organize electronic resources and to freely distribute the resulting catalog. Following the open source model, people apply to be an editor for a topic and volunteer their time to select, organize, and describe resources. Typically, editors begin by collecting links for small categories and as they develop seniority they can gain responsibility for larger and more general categories. In turn, senior editors review the work of more junior editors; thus, editors are organized into a hierarchy much as the categories themselves. Users of the directory are encouraged to submit links which are reviewed for inclusion by editors. Approximately, 63 000 editors manage four million links in over a half million categories; thus, this project can be seen as an effort to organize a large number of bibliographers, with all the associated problems of maintaining a rigorous selection process. The Yahoo Directory, of course, is of a similar scale but it does not publish the size of the directory nor promote the editors that sustain and grow it.

Infomine (D), a digital library of electronic resources suitable to academic users, is led and run by librarians. The aim of this project is to create a very high-quality collection of indexed resources by following traditional practices in librarianship, complemented by automatic and semi-automatic tools for discovering, describing, and classifying items. Of the approximately 100 000 resources, 25 000 resources were created by librarians and the remainder by automatic means.

The Librarians' Index to the Internet (E), which began in the early 1990s, aims to provide well-organized access to high-quality Internet resources. This collection of materials is created by librarians, following a rigorous process of selection, description, and cataloging. Detailed selection criteria and style guides are published at the site. The site accepts recommendations from the public but only a select group of approximately 100 contributors can submit records which are then reviewed and verified by an even smaller group of editors. Unlike Infomine, the use of computational tools to augment manual practices is minimal. Approximately, 14 000 resources are available. The Internet Public Library (Janes et al., 1999) follows a similar approach. Franco (2003) and Wells et al. (1999) give reviews of additional portals created by librarians.

The Wikipedia (F) aims to develop a comprehensive, multi-lingual encyclopedia of human knowledge where each article meets very high standards of scholarship. This project follows the Open Source model where anyone is able to contribute or revise an article, defined as a piece of text of 200 characters or longer. A version control system tracks the revision history and enables revisions to be rolled back. This mechanism and other collaborative tools have enabled the full multi-lingual encyclopedia to grow rapidly in three years to approximately 659 000 articles written by 17 000 contributors. (The English version has approximately 261 000 articles written by approximately 7000 contributors.) It is claimed that many of the articles are of poor quality and the pressing goal in its development is to improve the quality of the articles.

The Virtual Library (G), claimed to be the oldest catalog of the Web, was developed through a loose federation of volunteers where a single editor manages the top-level categories and a second tier of editors manage the topic pages. It aims to create a catalog of academic-oriented resources, particularly in science and engineering. To contribute, an editor develops a list of links on a particular topic and asks to be included in the top-level categories. In contrast to the Librarians Index to the Internet and Open Directory Project, the Virtual Library does not publish selection criteria, record formatting guidelines or style guidelines and no de facto standard has emerged; thus, each topic page has different record formats for items and visual styles. To indicate that a page is part of the virtual library, editors use a standard logo and URL that points to the top-level of the virtual library. This up-link is the only available mechanism for structuring the

catalog. The Virtual Library is structured with approximately 15 main categories which organize approximately 500 topics and an estimated 50 000 links (assuming 100 links per topic). While difficult to confirm, it appears that this collection is no longer growing and may be a fossil.

The USGenWeb Project (H), which aims to create a catalog of genealogy resources, follows a similar model where a single national coordinator is responsible for the national page, which lists pages for each of the states. In turn, state coordinators are responsible for managing the list of links to county pages. Finally, at the bottom of the directory, county coordinators are responsible for collecting and presenting the genealogy information resources for each county. The project has published bylaws and a mediation process for resolving conflicts amongst coordinators but does not publish detailed guidelines on selection criteria, record format, or visual style. Like the Virtual Library, lower levels of the catalog employ up-links to structure the information resources at the county level.

About.com (I) recruits and carefully selects *guides*, people who collect and organize resources and write original material for specific topics. A consistent information structure and visual presentation is enforced by About.com, which claims that it has approximately one million resources, 50 000 topics, and 500 guides. Topics contain a significant amount of advertising and guides are remunerated based, in part, on the number of page views that their topics generate. It attracts approximately 40 million unique visitors per month which puts it among the top 10 most popular websites.

So far, the examples come from region #1 of Fig. 2 where institutional and loosely collaborative organizations marshal labor for the purpose of building relatively large collections. Yet, most bibliographies that are published on the Web appear to be largely single-person efforts. The Human–Computer Interaction Bibliography (J), introduced earlier, is one example. Cyndi's List (K), a resource page for genealogy research resources, is another and is noteworthy because of its combination of scale and simplicity of technical infrastructure. This site, which began in 1996 as a bookmark file of 1000 links, has grown to over 250 000 categorized links and is maintained by a single person, using very simple writing and publishing tools. The compiler writes: "I have created each and every page on my site by hand. I write the HTML code myself and use Word for Windows to do so". (Howells, 2004). It is a remarkable example of what a highly motivated person can accomplish with the most rudimentary of tools but it is also exceptional—we know of no similar examples.

For a single person, order 10^5 number of items is very large. More typically, bibliographies that are created, or at least controlled, by an individual editor seem to range in size from order 10 to 10^3. Determining the distribution of the sizes of bibliographies is important problem, outside the scope of this analysis. This is not an easy problem because of the sheer size and diffuseness of the Web but also because even the definition of a Web page is controversial (e.g., see Henzinger & Lawrence, 2004). In any case, the nature of these resources pages varies to a very large degree. If in doubt of this, perform a Web search on *Bibliography* or *Directory*, restricted to the title of a page and then sample the results. To dodge this problem, one can investigate the nature of bibliographies that are all generated from a common application. One such application Links Engine (Gossamer Threads, 2004), a server-side code library that enables editors to create, manage, and organize items in a collection environment and to then publish these items at a website. Examples of resource pages that range in size from order 10^1 to 10^4 are easily found with a Web search such as "Links Engine". At increasing size the following sites are typical: Champion Fishing Links Directory (L), Encyclopedia of Psychology (M), and Esoterism.com Search Engine (N). In each case, the resource page consists of a relatively small number of categories and subcategories with resources listed. Readers can submit new resource descriptions which, in turn, are reviewed by an editor for acceptance and readers can submit general comments to the editor. In addition, readers can rate the quality of items and the click-through traffic each item receives is tallied and reported to the editor. Thus, the editor is given feedback from the reading audience on the quality of the resources. The application also provides a dynamic list of newly added items, an alerting service, and the ability to sample randomly selected items, an idea that was suggested by Bates (1976). Many other resource pages, whether handcrafted or created

from proprietary tools, have similar features. The Ubiquitous Computing Resource Page (O) is one example that is discussed below.

Turning to region #3, resource pages that contain a few links but are created by a large number of people, we see only a few examples, which particularly stretch the meaning of resource page. Nevertheless, this region presents interesting opportunities for identifying the most popular or highest quality material because it plots systems where the behaviors or opinions of many people are aggregated. This region is particularly suited to identifying new trends. One approach in this vein, followed by the Lycos Fifty (P) and Google Zeitgeist (Q), is to analyze the search queries of millions of users and identify the queries that are growing most rapidly. The top queries are listed to allow users also to search on these topics. Another approach is to analyze the linking structures (i.e., who is citing who) of dynamic, online conversations and uncover the topics being discussed. Technorati (R), for example, computes lists of the top blogs, books, and news articles by analyzing, in near real-time, the text and linking structures of over three million blogs. A final example is Slashdot (S), where a select group of editors accepts news items that are submitted by a broad, technically oriented audience and where the quality of those news items can be rated by the audience. Systems for peer review and social book-marking (Hammond, Hannay, Lund, & Scott, 2005) are other examples that generally fall within region #4 where a moderate number of people create moderate sized collections.

In sum, this framework charts the terrain of collection development efforts in terms of the number of selectors and size of the collection and shows three significant regions. While far from exhaustive, this survey reveals a tremendous diversity of tools and practices that shape the development of collections. Next, we ask: Is it possible to identify general dimensions that underlie these systems?

3.2. Conceptual model

Focusing on region #1 and #2, we extend Bates' (1976) model by identifying and labeling the collaborative aspects employed by these projects when selecting items. In this way, we begin to identify the general dimensions and better see the regularities across the full spectrum of information systems. In turn, this conceptual model provides an effective ground for exploring system requirements and analyzing existing tools.

First, we need to introduce the notion of participatory practices in relation to the selection of bibliographic items and suggest the types of collaborative policies that are available to the compiler. To begin, consider this scenario:

Top-Ten Problem Scenario. *A researcher in Ubiquitous Computing wishes to create a bibliography of the 10 central papers on this field. Together, the papers should map the field and enable researchers in neighboring fields to recognize the applicability of their specialties.*

To develop a solution to this problem, several different kinds of approaches are available for establishing a participatory selection policy. First, the researcher could review her own information sources, select 10 items, and publish them. Then, based on feedback elicited from the audience, she could revise the list. Second, the researcher could select a sample of 30 sources and ask a small group of trusted colleagues to rate the items. Based on the ratings and, perhaps additional unsolicited items, she could make a final selection. Third, she could ask her colleagues to submit their top-ten picks and make a final selection from this material. As a final example, the researcher could post a general call to a large list of researchers in the field, asking for the best papers and make the final selection from the items submitted. Each of these approaches represents a different kind of participatory policy for selecting items but, even for this trivial scenario, it is far from obvious which policy will result in the most effective, high-quality top-ten list for effort expended. In all likelihood, the researcher will select the most expedient approach and this will likely be largely determined by the available technical resources (e.g., access to e-mail, listserv, etc.).

Now, we can seek a more abstract view. Following Bates (1976), the aim of the conceptual model is to explicate the ideal process for selecting items in a technologically neutral fashion. Here, however, the focus

is on how the bibliographer can augment the specification types associated with individual items with a participatory process. To do this, the original model, shown in Fig. 1, is extended to model the life-cycle of a bibliographic item in Fig. 3. This move associates a life-cycle status with each item, thereby explicating how items pass through a series of stages, from item-proposed to item-de-selected. By observing changes in item status, stakeholders can take the appropriate actions to move an item through its life-cycle. This approach falls under the class of event-based workflow systems, where one designs a workflow by specifying a series of allowable events, actions and responsibilities (Anderson, Anderson, Wadhwaini, & Bartolo, 2003).

In a strict manifestation of the workflow, this life-cycle is a waterfall where the process forces items to pass through a series of states with no backtracking. In practice, of course, backtracking is often needed to respond to oversights and changes in judgment; furthermore, an item's state might be conditioned on other criteria (e.g., the availability of the item has been shown to be reliable). A more important point concerns the subset of states that an item can pass through and the ordering of states. Should it be possible, for example, to accept an item before it is described? What happens if any proposed item is accepted? How are the judgments of multiple selectors coordinated for a final decision? By reasoning from the abstractions in Fig. 3, we can examine these questions and explore the consequences of difference design choices at some level of detail. We can, furthermore, discuss the similarities and differences of the above projects in terms of the choices made in how items pass through each stage of a life-cycle. Thus, by examining the workflow associated with items, the regularities of the above projects can become apparent.

For example, the Librarians' Index to the Internet ensures quality by imposing a rigorous workflow for how items pass through the phases of proposed, described, verified and selected. Further, the workflow is relatively centralized. With the Open Directory Project, the workflow is simplified and the responsibilities for selection are distributed. In both cases, automatic tools might be used to periodically check the availability of the resource, compute an access-reliability metric, and highlight items that fall below a threshold. In addition, if someone disagrees with an editor's selection, a note can be sent to the responsible editor, asking that the item be removed. For the Open Directory Project, this is crucial because, with audience feedback, the relatively lax selection criteria can be balanced by the ability to send corrective feedback to the editor. This allows the number of items to scale better than, for example, the Librarians' Index to the Internet at the cost of maintaining consistency in selection criteria. With the Wikipedia, a completely open policy is followed. For this to work, a motivated pool of knowledgeable writers and a version control

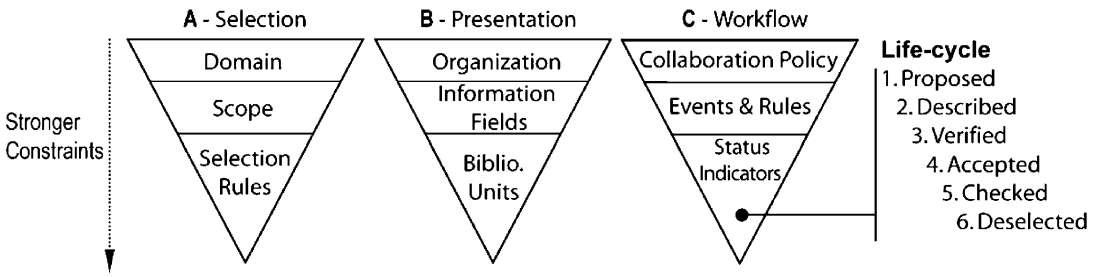

Fig. 3. Conceptual model for collaborative bibliography. This model extends Batess' (1976) model, shown in Fig. 1 by adding workflow constraints. The *collaboration policy* concerns intention and the general manner in which people are to collaborate in connection with the bibliography. The *events and rules* level concern the specific protocols for supporting the intentions that are outlined in the collaboration policy. The lowest level, *status indictors*, concern the properties that are associated with items and which are used to model the states of items and trigger new events. As an example, the allowable states assigned to an item might be: Proposed, Described, Verified, Accepted or Declined, Checked, and De-selected. The *events and rules* level specifies the rules for moving an item through this workflow. Finally, the *collaboration policy* level describes the intention of the workflow and the expected quality and character of the bibliography. Thus, we see three types of constraint, from Bates' original model, selection (A) and presentation (B), and from this work, workflow (C).

system is required so that changes can be tracked and rolled back. In sum, we can identify a major trade-off for ensuring quality: A centralized, multi-stage process for selecting items followed by a relatively high level of stability versus a distributed, relatively simple initial process for selecting items followed by the ability to take corrective actions. In the first case, the workflow leads to an incrementally fixed collection; in the second, it leads to a forever provisional collection.

Additional options are possible with this conceptual model. Some workflows, for example, will proceed in parallel with multiple editors making judgments that are hidden from each other. After the rating phase, these judgments and rationale might be made visible for a period of debate or kept confidential. In other cases it might be desirable to make all deliberations visible in real time. Another option concerns the provisionality of some state transitions. In some workflow scenarios, state transitions will occur subject to some conditions where, for example, an item might be accepted subject to a revised description or accepted items might remain accepted subject to the reliability of the resource and a yearly check of the resource's content. Another kind of option concerns the visibility of items as they move through the phases. Should, for example, proposed or rejected items be made publicly available? Making proposed items available, allows the audience to give the selectors some initial feedback. Showing rejected items can clarify the scope, or other selection criteria, of the bibliography but, in some cases, might cause unwanted consequences for the authors of the item. This analysis of the space of collaborative practices in bibliography shows that at least the following decisions need to be made to create a collaborative policy for creating and sustaining bibliographies: What are the (1) the allowable item states; (2) the allowable transitions between item states; (3) the conditions necessary for transitioning between states; and (4) the authorities and responsibilities entailed by the conditions.

A variety of operations can be applied against a particular workflow. In terms of human-oriented operations, selectors and readers might want to *inspect*, *rate*, or *comment on* a resource, and *check* various properties, including citations, history of editorial changes made to the item, click-through traffic received by the item, and so on. The audience might be prompted to *submit new items* or *subscribe to syndicated updates* of newly added resources. Readers, furthermore, might benefit from a *scoped search* in which just the sites in the bibliography are searched and where search results are presented in the context of the bibliography. More complex activities include the participatory development of schemes for structuring the bibliography, revising structures, and highlighting alternative or contentions structures and enabling discussion and negotiation. For machine-oriented operations, selectors might expect automatic checking for duplicates, link reliability, readability, accessibility, and other such properties. Or, the structure and content of the bibliography might be used to drive automatic or semi-automatic resource discovery (Hendry, in press). The true power of these otherwise ordinary operations comes from how they function in an interrelated way within a particular workflow.

Overall, this analysis reveals a profound challenge. To enable various forms of participation in the creation of a bibliography, the bibliographer must first design the policy, just as with the selection criteria, but selecting among the available collaboration policies and implementing a policy is exceedingly difficult. Indeed, in general to implement a policy requires a Web development project which, in turn, transforms the bibliographic project into a complex systems building effort. In short, the tools available for implementing collaborative bibliographic policies lag our abilities to invent policies. In the next section we discuss some of these implementation challenges.

4. Case study: Ubiquitous computing resource page

In this section, a case study is presented. The aim of this case is to describe the collaborative challenges of creating bibliographies in a particular technology milieu and thereby specifically illustrate the conceptual issues of the previous section. This case comes from our own development efforts and summarizes a 10

month project to design and implement a resource page on Ubiquitous Computing (Ubicomp). Ubicomp is a quickly evolving, multi-disciplinary field of research that explores computing technology as it moves beyond the desktop environment and becomes increasingly interwoven into the fabrics of our lives. The increasing power and decreasing size and cost of a variety of technologies enables computers to be carried, handheld, worn, or embedded in things, places and even people. The field of ubiquitous computing is concerned with the design, implementation and deployment of these technologies, as well as the impact these technologies have on people and society. The central idea for the Ubicomp Resource Page was to support researchers by collecting and organizing selected items that focused on the *people*, *projects*, and *organizations* of the field (McCarthy, Jenkins, & Hendry, 2003).

The project consisted of two key members: A developer with a background in technical project management who was just completing his master's degree in Library and Information Science (the second author) and an expert in ubiquitous computing (the third author). The project was originally scheduled to run for six months but was extended and ran for nine months at approximately 20 h/week for the developer. Four goals drove the project: (a) gather together the key resources on Ubicomp for the purpose of supporting researchers in the field and neighboring fields; (b) present these items in a consistent, integrated fashion to promote directed and serendipitous exploration of the field; (c) position the resource page in such a way that it complements regularly convened conferences on Ubicomp; and, most challenging of all; (d) codify selection policies such that the growth of the collection would be sustainable.

The project was divided into two phases. For phase I, collection development, the aim was to assemble a collection of resources of sufficient size to attract members of the Ubiquitous Computing community and to describe these resources in a consistent scheme. To structure this work, it was decided that the collection would focus on *people*, *projects*, and *organizations*. Using search engines and a variety of secondary access points, provisional items were selected. The access points included a list of home pages for researchers in the field, bibliographies found at project sites, and the proceedings from the Ubicomp conferences, workshops, and journals items were selected. To organize this process, an application for managing bookmarks, PowerMarks (PowerMarks, 2004), was used. This application provided convenient methods for describing resources, tagging them with keywords, and saving entries. As sites were visited and determined to be relevant, the URL of the site was captured along with title, description, and institution metadata. Images, if found, were also saved. Then, the site's other outbound links were inspected and additional pages were added if appropriate. In this way, PowerMarks was used to organize the process of collecting resources and developing a general awareness of the size, scope, and relationships within the Ubicomp community.

As this list grew, regular meetings were held to review the appropriateness and quality of the items and to clarify the subject scope of the collection. These review meetings were crucial for establishing a shared vision for the subject scope and quality criteria for the items to be included in the collection. This was particularly important because the boundaries of the field have yet to be established and appear to be controversial (e.g., ubiquitous computing is not yet in the ACM Computing Classification System). Incidentally, while statements on subject scope and quality criteria were not written, they were reviewed and rehearsed frequently in project meetings.

As items were collected, the team explored how items falling under the categories of *people*, *projects* and *organizations* could be interrelated. Initially, one goal was to explicitly model the many-to-many relationships amongst these items, thus allowing readers, for example, to access a Ubicomp researcher and then discover the projects that the researcher has participated in and the organizations that the researcher belongs to. It was felt that explicitly modeling these connections would be a valuable for enabling readers to readily explore the Ubicomp research community. Furthermore, interesting possibilities for visualization could be pursued. Two problems, however, soon became apparent. First, clearly defining the meaning of the categories *project* and *organization* proved to be problematic. For example, projects span from small, single person efforts to multi-institutional, government-funded initiatives. Even more difficult, however, was deciding what conditions needed to be met for a *person* to be considered a participant on a *project*

or for a *project* to be within an *organization*. The team discovered various categories of relationship including people who were affiliated with but not members of an organization, people with joint appointments, people who had been members of different organizations at different times, and people who played major roles on Ubicomp projects but where not themselves researchers in Ubicomp. In short, deciding when to associate a *person* with a *project* or *organization* became a political act and initial feedback from our colleagues made clear that our judgments were not always seen to be objective, despite our best efforts. Second, a database schema that would represent these many-to-many relationships was a complex design and develop task, outside the scope of the project. Thus, even if the conceptual problems associated with defining these classes and making reliable judgments could be solved, maintaining these relationships manually over time would be very time-consuming. Thus, for conceptual and technical reasons, the team decided to abandon the idea of explicitly modeling the many-to-many relationships between *people*, *projects*, and *organizations*.

At this point, it was decided that the site would emphasis these three categories. The *people* items consist of the fields *name*, *affiliation*, *home page*, *photo*, and *publications*. A list of publications is generated by a query to the ACM Digital Library (ACM, 2004) and the DBLP Bibliography Server (DBLP Bibliography Server, 2004). We felt that photographs of the researchers would make the collection more interesting and a list of photographs of many Ubicomp researchers was available.

Turning to *project* items, they consist of the fields *name*, *project homepage*, *photo*, and brief *description*, which is typically copied from text found on the home page or publication of the project. An important observation was the remarkable diversity of description and presentation of projects, organizations, and people at particular sites, and a major contribution of the project was to standardize the description of these items. As discussed above, the team decided to not associate *people* and *organizations* with *projects*. The photo field was added in response to the observation that many of the projects included photographs illustrating the system and, like the images of the researchers, these photographs made a long list of entries more interesting to scan.

Finally, turning to *organization* items, they consist of the fields *name* and *home page*. Organizations represent another category that has varying levels of granularity (as with projects and initiatives). For example, academic researchers are associated with their university, their school, their department, and any number of centers, groups or labs. Again, for conceptual and technical reasons, distinctions and connections among these different kinds of organizations were not modeled. In addition to these major components, the resource page also presents *event* and *news* items lists. (These lists are described further.)

The next consideration was how to arrange the items in each of these lists. The team estimated that for the first version, there could be as many as 200 items in each of the lists and that this number could grow to 500 items and perhaps more, especially for the project list. Thus, the team felt it was important to introduce some method of categorization. The team, however, decided to defer this issue and decided on an alphabetic ordering by name of person, project, and organization, clearly a suboptimal arrangement for people who are uncertain of what they are searching for. The reasons were twofold. First, informed by the early discussions on the subject scope for the field, the team knew that creating thematic categories would be very difficult and sufficient time would not be available to gather feedback on the quality of the categorization scheme. Second, the task of developing a categorization scheme was in competition for time to create tools that would enable members of the Ubicomp community to submit and discuss items. This second task was given priority and the team decided, if time permitted, to revisit the categorization problem later.

After approximately three months, the team transitioned into phase II, community development. The original project brief called for a static collection of resources, accepting that sustaining the collection would be the job of a single person working alone. As the project evolved, however, a shift in thinking occurred. It began with an observation and a simple future scenario.

The observation was that it was very challenging to decide what items were within the subject scope of the collection and what were clearly outside. In weekly discussions, we attempted to find prototypic

examples of items that were clearly in and out of scope and we attempted to find boundary cases that revealed potential controversies. The team soon found itself wanting to elicit outside opinions. In turn, this led to the future scenario: Members of the Ubicomp community might want to discuss the resources and help shape the collection. If this were so, could the resource page explicitly support discussion and, secondarily, could the ability to register, submit items, and review items enable the collection to be sustained? The project was extended and the team began to explore options for creating a relatively open-ended, collaborative environment for collection development.

The initial technical platform for development was a set of XML files and associated style sheets. This was thought to be a straightforward and sufficient approach, consistent with the technical skills of the team and the original project brief. However, with the expansion of the brief to include collaborative features, a new technological platform was needed. The team decided to experiment with Plone (Plone, 2004), an open source development kit for building complex, collaboration-based Web applications. Among the many claimed benefits of Plone were a robust set of community features, a membership model, and a content management framework. After several weeks of effort, the team ran into significant technical problems. First, we discovered that to customize Plone, programming skills in Python and ZClasses were required. These skills, however, fell outside the core competencies of both the project team and the greater community involved with the project. Second, while the community features had much potential, they required users to register and log-in even to read discussions. Loosening this authentication model, however, appeared to be quite difficult. Finally, the Ubicomp Resource Page was being developed for a Microsoft Windows environment. Configuration and extensions to Plone, however, seemed to privilege open source tools and this added a significant level of complexity.

For these technical reasons, the team switched to the ASP.NET Community/Portal Starter Kit (Microsoft, 2004). The starter kit provided many of the same features of the Plone framework, including content management workflow, message boards, RSS feeds, and so on with a much more usable administration console and simpler methods for customizing the application. Because it would be impossible to monitor and manage abuse (e.g., unprofessional or unrelated comments on professional work), the team decided to not implement discussion boards. In short, for largely technical reasons, unrelated to the conceptual problems of bibliography, the Community/Portal Starter Kit was selected.

Reflecting on the project, the following key lessons have been learned. First, with the publication of a collection of electronic resources comes the opportunity, perhaps even expectation, to enable community participation in the collection. Making this move adds a significant degree of technical complexity to the project. But, it also throws up tricky social-cum-political questions such as how should a newly developed collection fit into existing sites that support the Ubicomp community. Second, the conceptual issues associated with defining the scope of the resource page, the structure of the bibliographic items, and the methods of categorization were truly daunting. Yet, so too were the technical issues associated with simply creating and publishing the collection. Regrettably, the technical aspects of this project dominated because the team was unable to find tools that were suitable for developing resource pages and that were commensurate the technical competency of team and scope of the project. Third, the conditions that enable a resource page to be sustained are poorly understood. Early in the project the team acknowledged the tendency of resource pages to quickly become fossilized and wanted to avoid this situation. Indeed, the Intelligent Environments Resource Page, http://research.microsoft.com/ierp/, was the best resource available in the field in 2003 but it reflected a snapshot of research circa 1998. This project sought to avoid this fate. Yet, it soon became apparent that the effort required to do this was beyond the scope of this small project. These lessons prompted us to reflect more broadly on electronic collections and led us to this analysis of collaborative bibliography.

How common are the experiences of this case study? While we do not have data that allows us to generalize, it is important to address this question. From an institutional standpoint, it is clear that managing electronic resources is a difficult problem and that many professional organizations have experienced

significant challenges. Pitschmann (2001) reviews the issues confronted by institutional libraries but he does not consider the participatory forces of networked information that lead to audience expectations for engagement and collaboration. Reasoning that, in general, professionals are better equipped to manage the development of collections, we can expect to see that non-professionals also experience significant challenges. Anectodally, this appears to be the case.

In one example is from our own consulting experience concerns an organization that maintains a collection of electronic resources on a complex, interdisciplinary topic. The aim of the collection is to support a diverse audience, interested in this topic in terms of education, research, and social activism. The collection has been subdivided into four major themes and a different person is responsible for periodically weeding and adding resources to each theme. Technically, the collection consists of four HTML pages that are hand edited to simple standards and the size of collection is approximately 200 links. When an informant, one of the four editors, was asked about what it was like to develop and maintain the site, she responded "[to collect resources and create the pages was] very exciting at first... After you keep it in place for awhile, you need to know that there is value... Doing it in random and with one pair of eyes is not good". Here, we see a person, on the one hand, raising concerns about being connected with the audience, trying to answer the question "Is there value in the collection"? Where, without feedback on use, does the incentive to sustain the collection come from? On the other hand, we also hear the informant indicate the importance of coordination amongst her peers, pointing to the importance of selection criteria and the ability to review and comment upon each other's selection decisions. For this project, one can see both the importance of traditional bibliographic practices but also the need to address collaborative policies that enable the resource collection to be sustained. Collaborative policies concerning the relationship between the selectors and collection's items (i.e., how the selectors work together) and between the audience and the collection (i.e., how the audience uses the collection) are clearly important in this project. Yet, like the Ubicomp project, this team lacks the tools and resources to develop policies that address these two aspects of sustaining a collection over time.

5. Conclusion

This article has examined bibliography from a new perspective, introducing the possibility of collaboration, extending the traditional conceptual view of bibliography to include collaborative practices that enable people to participate in creating, shaping, and sustaining collections. We believe that collaborative policies for bibliography are fundamentally conceptual just as are the traditional criteria of subject scope, selection rules, and bibliographic units. Facilitating participatory decision-making, shaping workflow and making it visible, and enabling audience discussion of bibliographies are enduring, long-term goals. The difficulty, however, is that such conceptual challenges are not easily separated from particular implementation issues. We do not have enough experience with such systems and much work remains to be done to document and to analyze the design space of bibliography-like applications. In the present state of the art, implementing a collaborative policy requires a significant investment in systems development. In turn, this barrier restricts the exploration and adaptation of different collaborative practices to groups with significant technical resources.

A variety of research topics emerge from this work:

Amateur bibliographers. A large number of bibliographies and bibliography-like information artifacts are created by non-professionals. On Web logs, on home pages, and on resource pages millions of people are creating lists of resources. While these lists often fall short of the standards of professional bibliography, they are nevertheless important organizing instruments which, in turn, lead to a great benefit. The various forms of amateur bibliography are, in general, underappreciated and understudied. This is surprising given the evident popularity of list-making in all its various forms and its importance in search-engine algorithms such as Page Rank (Brin & Page, 1998).

Bibliography construction tools. Tools that are designed for single, non-professional users and ad hoc groups of people are lacking. For these users, the common practice is to develop bibliographies with general writing and publishing tools. Even for users with access to technical resources, a great deal of enabling infrastructure must be built. And, as discussed in Section 4, implementing this infrastructure has the potential to deflect attention away from important conceptual issues such as the selection and arrangement of items. In principle, publishing a resource page should be no more difficult than publishing a personal Web log, which in its most basic form is a 5 min process on such sites as www.blogger.com.

Professional intermediation. With many amateurs employing tools to create bibliographies, interesting opportunities exists for professionals to help people in the creation of bibliographies, complementing the intermediation of information-seeking problems. We envision professionals asking questions about collection development policies, selection criteria, and so on. In short, online bibliography presents librarians with a new area for service-oriented, professional practice.

Collection development and bibliography. Collection develop policies primarily concern how an institution will manage its holdings through the delegation of resources and responsibilities. Bibliography is primarily concerned with developing instruments that allow effective information seeking. In general, the bibliography-like artifacts reviewed in this article show that practices are borrowed, deliberately or not, from both of these areas. Because of the dynamic nature of the Web, resource pages can tend to been seen as, and sometime are, institutions, run by communities of interest. It is important to better understand the factors that motivate volunteers to collect resources and how these factors can be supported in policies. Further analysis of the connection between collection development and bibliographic practices may clarify how resource pages can be designed for sustainability.

Finally, we believe that bibliographies and bibliography-like artifacts will play major roles in the future. Fundamentally, bibliographies locally structure the Web into a patchy network of organized areas. These microstructures are invaluable for prompting communication around common topics, for enabling automatic and semi-automatic resource discovery, and for providing creative outlets. They signal the importance of intermediation. As McArthur (1986) as argued with historic examples, the taxonomic urge runs deep. It is an urge, however, that we have yet to fully understand and support in the current technological milieu. Much remains to be learned about the collaborative possibilities of bibliographies.

Acknowledgements

We gratefully acknowledge Allyson Carlyle for suggesting the idea that *resources page*s are *merely* bibliographies and, critically, for directing us to Bates' paper. Marcia J. Bates' comments on a draft of this article improved its clarity. Any remaining faults are, of course, our own. The reviewers of this paper provided extremely helpful criticism. John LaMont, Brian Green, and Jenine Lillian helped with the preparation of this manuscript.

References

ACM (2004). ACM Digital Library. <http://portal.acm.org/dl.cfm> Retrieved 1.6.2004.

Anderson, K. M., Anderson, A., Wadhwaini, N., & Bartolo, L. M. (2003). *Metis: Lightweight, flexible, and Web-based workflow services for digital libraries*. Paper presented at the Proceedings of the 3rd Joint Conference on Digital Libraries, Houston, TX.

Andre, J., Furuta, R., & Quint, V. (1989). *Structured documents*. New York: Cambridge University Press.

Bar-IIan, J. (2004). An outsiders view on "Topic-oriented" Blogging. In *Proceedings of the Thirteenth International World Wide Web Conference* (pp. 28–34). New York: ACM Press.

Bates, M. J. (1976). Rigorous systematic bibliography. *RQ, 16*(1), 7–26.

Bawden, D., & Robinson, L. (2002). Internet subject gateways revisited. *International Journal of Information Management, 22*, 157–162.

Blood, R. (2000). Weblogs: A history and perspective. Rebecca's Pocket, 7 September 2000. <www.rebeccablood.net/essays/Weblog_history.html> Downloaded 1.7.2004.
Blood, R. (2004). How blogging software reshapes the online community? *Communications of the ACM, 47*(12), 53–55.
Borgman, C. L. (2001). Where is the librarian in the digital library? *Communications of the ACM, 44*(5), 66–67.
Brin, S., & Page, L. (1998). *The anatomy of a large-scale hypertextual search engine*. Paper presented at the seventh international World Wide Web conference. <http://wwwdb.stanford.edu/~backrub/google.html> Retrieved 1.7.2004.
Bush, V. (1945). As we may think. *The Atlantic Monthly*. <http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm> Retrieved 1.7.2004.
Card, S. K., Mackinlay, J. D., & Shneiderman, B. (Eds.). (1999). *Readings in information visualization*. New York: Morgan Kaufman Publishers.
Crowston, K., & Williams, M. (1997). Reproduced and emergent genres of communication on World-Wide Web. In Proceedings of the 30th Hawaii international conference on system sciences: Digital documents, Maui, HI, USA, January.
DBLP Bibliography Server (2004). <http://www.informatik.uni-trier.de/~ley/db/> Retrieved 1.6.2004.
Evans, G. E. (2000). *Developing library and information center collections* (4th ed.). Englewood, CO: Libraries Unlimited.
Francis, F. C. (1973). Bibliography (15th ed.). *Encyclopedia Britannica* (vol. 2, pp. 978–981). Chicago: Encyclopedia Britannica.
Franco, A. (2003). Gateways to the Internet: Finding quality information on the Internet. *Library Trends, 52*(2), 228–246.
Garrett, J. J. (2002). *The elements of user experience: User-centered design for the Web*. Berkeley, CA: New Riders.
Goncalves, M. A., Fox, E. A., Watson, L. T., & Kipp, N. A. (2004). Streams, structures, spaces, scenarios, societies (5S): A formal for digital libraries. *ACM Transactions on Information Systems, 22*(2), 270–312.
Gossamer Threads (2004). Links Engine 2.0 [Computer Software]. <http://www.gossamer-threads.com/> Retrieved 1.7.2004.
Hammond, T., Hannay, T., Lund, B., & Scott, J. (2005). Social bookmarking tools (I). *D-Lib Magazine, 11*(4). <http://www.dlib.org/dlib/april05/hammond/04hammond.html> Retrieved 1.5.2005.
Hendry, D. (in press). Workspaces for search. *Journal of American Society for Information Science and Technology*.
Henzinger, M., & Lawrence, S. (2004). Extracting knowledge from the World Wide Web. *Proceedings of the National Academy of Sciences, 101*. <www.pnas.org/cgi/doi/10.1073/pnas.0307528100> Retrieved 1.7.2004.
Howells, C. (2004). Cyndi's List. <http://www.cyndislist.com/faq/howdoyou.htm> Retrieved 1.7.2004.
Janes, J., Carter, D. S., Lagace, A., McLennen, M., Ryan, S., & Simcox, S. (1999). *Building online collections*. New York: Neal-Schuman Publishers.
Jones, W. (2004). Finders, keepers? The present and future perfect in support for personal information management. *First Monday, 9*(3). <http://firstmonday.org/issues/> Retrieved 1.7.2004.
Kirriemuir, J. (1999). A brief survey of quality resource discovery systems: Resource discovery network center. <http://www.rdn.ac.uk/publications/studies/survey> Retrieved 1.2.2004.
Kleinberg, J. (2004). Analysing the science in its online context. Nature Web focus on access to the literature. <http://www.nature.com/nature/focus/accessdebate/18.html> Retrieved 1.7.2004.
Kleinberg, J., & Lawrence, S. (2001). The structure of the Web. *Science, 294*, 1849–1850.
Kumar, R., Raghavan, P., Rajagopalan, S., & Tomkins, A. (1999). *Trawling the Web for emerging cyber-communities*. Paper presented at the eighth internal World Wide Web conference. <http://www8.org/> Retrieved 1.7.2004.
Lee, H.-L. (2000). What is a collection? *Journal of American Society for Information Science, 51*(12), 1106–1113.
Lenhart, A., Horrigan, J., & Fallows, D. (2004). *Content creation online*. Pew Internet & American Life Project. <http://www.pewinternet.org/pdfs/PIP_Content_Creation_Report.pdf> Retrieved 1.6.2004.
Levy, D. M., & Marshall, C. C. (1995). Going digital: A look at assumptions underlying digital libraries. *Communications of the ACM, 38*(4), 77–84.
Lindahl, C., & Blount, E. (2003). Weblogs: Simplifying Web publishing. *IEEE Computer, 36*(11), 114–116.
Mack, D. C. (Ed.). (2003). *Collection development policies: New directions for changing collections*. New York: Haworth Information Press.
McArthur, T. (1986). *Worlds of reference*. New York: Cambridge University Press.
McCarthy, J. F., Jenkins, J. R., & Hendry, D. G. (2003). *The ubiquitous computing resource page (ucrp.org)*. Paper presented at the UBICOMP 2003, the fifth annual conference on ubiquitous computing, Seattle, WA.
Microsoft (2004). ASP.NET Community/Portal Server Kit. <http://asp.net/Default.aspx?tabindex=9&tabid=47> Retrieved 1.6.2004.
Orr, C. (2003). Collection development in public libraries. In A. Kent & H. Lancour (Eds.), *Encyclopedia of Library and Information Science*. New York: M. Dekker.
Perlman, G. (1999). The HCI Bibliography: Ten years old, but what's it done for me lately. *Interactions, 1*, 32–35.
Perlman, G. (2004). The HCI bibliography. <http://www.hcibib.org> Downloaded 1.6.2004.
Pitschmann, L. A. (2001). *Building sustainable collections of free third-party Web resources*. Washington, DC: Digital Library Federation, Council on Library and Information Resources. <www.clir.org/pubs/reports/pub98/contents.html> Downloaded 1.7.2004.
Plone (2004). <http://www.plone.org> Retrieved 1.6.2004.

PowerMarks (2004). <http://www.kaylon.com/power.html> Retrieved 1.6.2004.

Robinson, L., & Bawden, D. (1999). Internet subject gateways. *International Journal of Information Management, 19*(6), 511–522.

Rosenfeld, L., & Morville, P. (2002). *Information architecture for the World Wide Web: Designing large-scale Web sites*. Sebastopol, CA: O'Reilly & Associates.

Schiano, D. J., Nardi, B. A., Gumbrecht, M., & Swartz, L. (2004). Blogging by the rest of us. In *Extended abstracts of the 2004 conference on human factors and computing systems* (pp. 1143–1146). ACM Press: New York.

Smith, M. A. (2004). Portals: Toward an application framework for interoperability. *Communications of the ACM, 47*(10), 93–97.

Spohrer, J. H. (2003). The end of an American (Library) dream: The rise and decline of the collection development policy statement at berkeley. In D. C. Mack (Ed.), *Collection development policies: New directions for changing collections* (pp. 33–48). New York: Haworth Information Press.

Stokes, R. B. (2003). Bibliography. In A. Kent & H. Lanjjcour (Eds.), *Encyclopedia of Library and Information Science*. New York: M. Dekker.

Torok, A. G. (Ed.). (2003). *Organizing the Internet* (vol. 52). Library Trends.

Tufte, E. R. (1997). *Visual explanations*. Chesire, CT: Graphics Press.

Van Duyne, D. K., Landay, J. A., & Hong, J. I. (2003). *The design of sites: Patterns, principles, and processes for crafting a customer-centered Web experience*. New York: Addison-Wesley.

Wells, A. T., Calcari, S., & Koplow, T. (Eds.). (1999). *The amazing Internet challenge: How leading projects use library skills to organize the Web*. Chicago: American Library Association.

Wilson, P. (1968). *Two kinds of power: An essay on bibliographical control*. Berkeley, CA: University of California Press.

Yates, J., & Orlikowski, W. J. (1992). Genres of organizational communication: A structurational approach to studying communication and media. *The Academy of Management Review, 17*(2), 299–326.