

Selecting and cataloging Internet resources in Danish public libraries

Marianne Lykke Nielsen & Erik Thorlund Jepsen
The Royal Danish School of Library & Information Science

Abstract

The Danish Netkat project involved four public libraries. The libraries developed selection and registration policies for Internet resources and cataloged 270 resources. The selection policy included the development of a combined evaluation and preregistration scheme, which was filled out and used by subject specialists. The core elements for selection of resources, were reliability, uniqueness and ability to fill gaps in the library collection. The level of registration was developed in consideration of the dynamicality of most Internet resources. The utility and usability of OPAC's facilitating direct access to Internet resources was tested by 58 patrons. Test persons were interviewed to reveal the adequacy of selection criteria and to obtain further knowledge about usability aspects.

Keywords

Internet resources ; cataloging ; selection policy ; evaluation ; level of registration ; Denmark ; public libraries ; user tests

Introduction

This paper presents the background and findings of the Netkat project a Danish study concerned with the identification, selection and cataloging of Internet resources for use in public libraries. The Netkat project involved four Danish public libraries (Aalborg, Allerød, Hillerød and Hjørring), The Royal Danish School of Library and Information Science and The Danish Library Centre. The project was carried out from February 1998 to January 1999. The aim of the project was to examine the feasibility of incorporating records of Internet resources in public library catalogs, and providing direct access to these resources from the catalog. The project focused on obtaining indications for both practical acceptability as well as utility and usability of OPAC's facilitating access to all types of materials.

The specific objectives was to identify important aspects for - and develop procedures for - identifying and selecting Internet resources; to establish an adequate level of registration; and to test patron's and library staff's comprehension of, use of and satisfaction with the searchability of Internet resources in library catalogs.

Mediation of Internet resources

Public libraries are natural centres for mediation off all kinds of information, it is therefore necessary to include Internet resouces in the libraries selection, registration and mediation policies.

Libraries can choose among different mediation strategies for Internet resources.

One strategy is to let search engines and web guides do the job! This solution is relatively cost free for the technical departments but lays the burden upon the reference staff. Search engines are large databases, which often index several millions of documents. The indexes in these databases are built automatically. Software devices "robots" traverse the Web looking for resources to index as well as updates to previously found resources. The search engines are valuable tools due to their size. The automatic indexing however results in some major disadvantages. First off all data are only to a minor degree qualified. Keywords are statistically - not intellectually - identified, and can only be regarded as indicators for subjects, that may or may not be dealt with in the retrieved documents. Moreover users are not able to separate different types of data. For example searches for "Microsoft" will retrieve both documents created by "Microsoft" as well as documents about "Microsoft". Finally the automatic indexing produces descriptions that are both inconsistent and insufficient for identification and selection purposes. A recent survey at Seton Hall University [1] revealed, that out of 786 students and faculty members 38.5 % found, that too many hits were retrieved when searching the Internet, 49.2 % did not find the information requested, and 43.7 % found, that none of the full-text information found could be cited for academic study and/or research.

Opposed to search engines, most web guides or "directory search tools" are based on intellectual selection and indexing. There exists a great number of both larger (universal) and smaller (topical) hierarchically structured lists of documents, which are often useful due to the expert-based selection of documents. These lists are easy to use, since users are not forced to create any search statements, but simply browse through

the categories in the list. Yet the web guides often suffer from the same lack of appropriate descriptions as the search engines.

Instead of referring patrons to search engines and/or web guides, libraries can choose to create and/or reuse bibliographic records of Internet resources, and provide access to these records through separate databases or through the library catalog. Whether a separate database for Internet resources, or incorporation of records in the actual catalog is chosen, patrons will benefit the most, if unified access to all kinds of materials is facilitated.

The Netkat-project was based on the following assumptions:

Cataloging Internet resources can benefit both libraries and patrons:

- Intellectual selection among resources will benefit both patrons, who are less distracted by noise (documents of none or only inferior value), in comparison to searching search engines, and reference staff, who will have immediate access to intellectually selected materials from the Internet.
- Patron's retrieval of Internet resources will be eased through the use of known search techniques and searching among qualified data. Furthermore, patrons will be able to retrieve all types of publications using only one search tool.
- Relevance judgement is eased through intellectually compiled descriptions, as opposed to the automatically extracted descriptions, offered by most search engines and Web guides.
- Incorporating Internet resources in the catalog can even minimize library costs, if selected Internet resources replaces publications in more expensive formats.
- Facilitating access to intellectually selected Internet resources will benefit patrons, who are able to use the same documents at the same time, in contrast to other materials where obtainability is a frequent problem.

Selecting Internet Resources

Internet resources are characterized by differing quality and stability and often by a high degree of complexity.

Publishing Internet resources is not limited by the filters, that influence publishing in other medias, primarily because resources are normally not evaluated by publishers. Furthermore Internet resources are seldom reviewed in journals or newspapers. Exhaustive reviews of all monographic textual material, published in Denmark, is compiled and published by a central institution "Indbindingscentralen". These reviews are primary tools for selection in Danish public libraries. Caused by the absence of such selection aids, libraries must establish their own selection criteria and develop their own procedures and guidelines for selection of Internet resources.

An often stated argument is that selection criteria must equal the criteria used for other media, "*If an Internet resource were published in any other media, say paper, and it met your library's selection and collection development criteria, would you catalog it? If the answer is yes, then select the internet resource and catalog it*" [2].

According to this statement, selection should be based upon summary evaluation of the resources. To minimize costs, such evaluations can be standardized by evaluation schemes. The criteria, accounted for in such schemes, must equal libraries traditional selection criteria along with specific criteria, which are crucial for patrons relevance judgements of Internet resources.

There is an increasing amount of literature about evaluating Internet resources. Most literature concerned with selection of Internet resources and most selection schemes account for *content* (audience, coverage, currency, quality, uniqueness), *design* (communication mode, skills needed, software required) and *access* (access mode, restrictions) [3]. Peggy Johnson [4] identified the following criteria of relevance for selection: relevance to program needs, scope, ability to fill gaps, uniqueness, quality of scholarship, quality of physical product, currency of information and frequency of updating, accessibility of information, language, cost reputation of author/reliability of publisher and special features.

Most of these criteria are general criteria, which could be applied to any media, but especially "accessibility" is rather unique for electronic resources. Resources that satisfy all other criteria can be poorly organized, and only usable for computer experts or likewise. Such resources should probably not be selected. However there are good reasons for selecting resources, which are difficult to use, if the content is unique, reliable, updated and of high importance for patrons. The evaluation scheme used In the Netkat-project included comments on

interface, navigation, search and output facilities and problems (e.g. if the naming of hyperlinks were misleading). Such information was however not used to exclude resources satisfying a majority of the content related criteria. Instead essential information for navigating or searching selected resources was stated in notes in the bibliographic record.

The criterion "reputation of author/reliability of publisher" implies some complications when dealing with Internet resources. Statements of responsibility are often absent or superficial in Internet resources. Although such information about responsibility is an essential aspect in determining the validity of a given resource, the lack of such information was not used alone to exclude resources from registration in the project. Some resources, lacking any information of responsibility, was included because of their uniqueness.

A specific selection criterion for Internet resources is the stability of the resource under consideration. Some resources, such as monographs (e.g. reports) are very static, while other resources tend to change on a day to day basis. The stability of a resource can influence the reliability of the catalog incorporating Internet resources, because records, pointing to dead links or to resources very unlike the ones described in the record, will reduce patron's confidence in the catalog.

The dynamics of Internet resources can be expressed in 1,2 or all 3 of the following ways:

"1) *Overwriting the content (existing content removed and replaced)*

2) *Supplementing the content by integrating new content (accumulation)*

3) *Supplementing the content by adding new content as independent sections/supplements"* [5]

It was chosen to include both static and dynamic resources in the Netkat project. The NETMIND-server, that monitor resources and informs about changes in extent and location of resources, was used to identify important changes. Problems with very dynamic documents were solved by minimizing the level of registration, and for some resources by registering the dynamic resource as part of a larger collection.

The complexity of Internet resources is another aspect, which influences selection, since the determination of which entities to select and register depends on how hypertext documents are organised. A website can contain some valuable documents along with parts of minor value. Important single documents can be selected and registered alone, or the entire collection can be chosen as the bibliographic entity to register. Diane I. Hillman points out, that when attempting to gain bibliographic control of the World Wide Web it is necessary to soften the view on bibliographic entities: "*The archival community has traditionally been more comfortable with making collection level decisions for their materials than general library selectors and catalogers have been with theirs*" [6, p.100]. In the Netkat project some single documents – e.g. country studies published by The Federal Research Division - were selected, primarily because of their stability and adequacy for grammar and business school students. However the general strategy was to select and register entire websites [see figure 1]. This selection of larger entities furthermore limits the chances, of selected resources changing radically in proportions to the description given in the bibliographic record.

In the Netkat project an evaluation scheme, which also was used as a preregistration form - specifying the most frequently used MARC-tags, was used to evaluate and select the resources.

The most important selection criteria were uniqueness and the resource's ability to fill gaps in the collection. The primary target group were grammar and business school students. A secondary target group were adults - "The common patron". Subjects covered were widespread: Danish authors; prostitution, country profiles, incest, films, drugs..a.s.o. Both danish (19%) and english (81%) resources were selected.

Even though quality aspects like reliability and currency were given weight, some resources, which are very unique within their topic but less reliable and up to date, were selected.

Subject specialists at each of the four libraries used the evaluation and preregistration scheme. 270 resources were selected for registration. The largest problem identified, was that most resources were either too superficial or too academic to suit the target group. The subject specialists used approximately one hour evaluating and filling out the evaluation and preregistration scheme for each resource in question. This lack of effectiveness was primarily caused by problems with the preregistration elements, since subject specialist are normally users of data, not creators of such.

Interviews with 58 patrons, used as test persons, revealed, that factors concerning content such as reliability and currency were very important for patrons. Only 36% of the respondents relied fully on the resources, while 52% were unable to judge the reliability of the retrieved resources. Several of the respondents stated, that reliability problems would be diminished, if they were told, that the resources was evaluated and selected

by library staff. Comprehension of the retrieved documents was not identified as a problem, which can be interpreted as a successful selection procedure, rather than a general attribute of Internet resources.

Currency

Yes	No	Dont know
69%	7%	24%

Table 1. Do you think that you found the most recent information

Reliability

Fully reliable	Dont know	Not reliable	Unanswered
36%	52%	9%	3%

Table 2. Would you rely on the information on the Internet

Subject level

Exhaustive	Appropriate	Superficial	Dont know
54%	30%	4%	12%

Table 3. The subject level in retrieved resources

Readability

High	Appropriate	Low	Language problems	Unanswered
63%	24%	5%	5%	3%

Table 4. Level of readability

Cataloging Internet Resources

The Netkat project was carried out under some constraints. First of all cataloging Internet resources in Denmark is eased, due to the 1998 edition of the Danish Cataloging Rules [7], which take into account some of the problems encountered, registering internet resources. Furthermore Danish public libraries in general assign descriptors using a universal controlled vocabulary, developed and maintained by The Danish Library Centre [8]. A study "Indoreg" performed in 1996 has established the level for national bibliographic registration of Internet resources [5]. The Netkat project has profited from some of the findings in the Indoreg project, but the project has produced its own level of registration.

A final, but major, constraint for the project, was the library system and OPAC's employed in the four libraries. Problems with the library systems primarily concerned the possibilities for displaying records in comprehensive and usable formats.

A major issue, when cataloging Internet resources, is to establish an adequate level of registration. The registration must facilitate the tasks of finding, identifying, selecting and obtaining the resources cataloged [9]. Yet the dynamic character of Internet resources calls for a minor level of registration, to ensure, that data do not become obsolete or misleading.

In establishing the registration level for the Netkat project, some problems occurred: Determination of entities to register - level of analytic registration, authorship/responsibility, titles, publication data, obtainability, description of content and indexing.

What makes an entity? Regarding the choice of entities, some resources were registered on a collection level. Some of the individual documents in these collections were also registered in single records. This is primarily a selection problem, since the choice of analytical registration depends on the individual documents uniqueness and ability to fill gaps in the library collection.

Authority control. Only persons and corporations responsible for the intellectual content in resources were registered. Persons, that carry out functions of technical or administrative character (e.g. web masters and programmers), were not mentioned in the records. Efforts were made to trace and identify authorships, due to the importance of relevance judgements. Authority control was performed, when names were traceable in the Danish National Bibliography "Danbib". Otherwise authority control was only performed to ensure access by different parts of compound names.

Titles. A majority of Internet resources have an entrance page or a regular title page. Titles identified in title pages often differ from the HTML-title, that are identified and displayed by Internet browsers. To ensure equivalence between the record and the resource, title page information was preferred. HTML-titles diverging the title page were cited and made searchable as variant titles

Publication. Since most Internet resources are dynamic, simply stating the publication year is not sufficient to inform about the currency of the resource. Information about updates were cited along with the indication of origin. All records were added a note stating, when the resource was registered, because information about

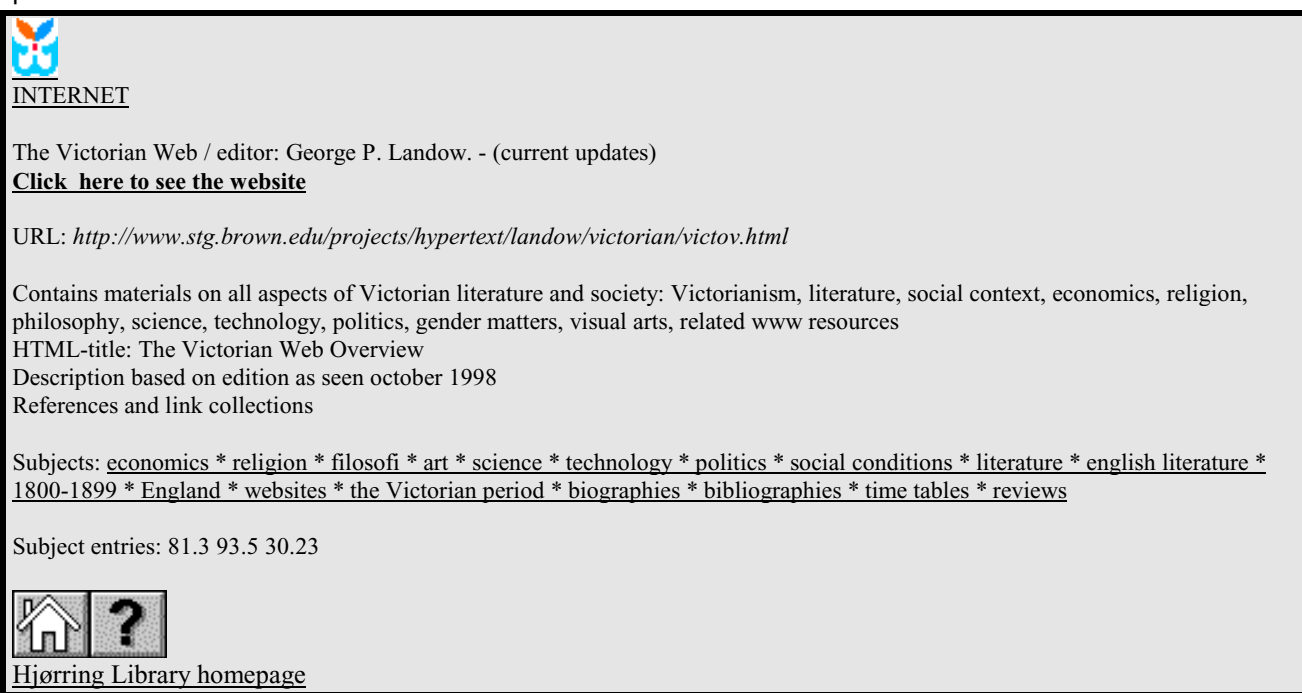
updates is bound to become obsolete. Furthermore information about registration dates facilitates updating strategies for the bibliographic records.


Obtainability. The MARC field “856” was used to state a clickable address to the cataloged resource. A standard formulation “Click here to see the website” was given. Tests revealed, that Internet novices had difficulties manoeuvring between the catalog and the Internet resources.

Notes. Notes was primarily used for content information. Short descriptive content notes were stated in the primary language of the resource. Notes also included information about collections of references and/or links.

Indexing. Descriptors were generated following the traditional Danish public library policy. Indexing was performed on document - in contrast with analytical - level. The specificity was determined by the vocabulary developed by The Danish Library Centre [8], which accounts for subjects, form and target groups. Especially the assignment of descriptors troubled the subject specialist, who preregistered the Internet resources.

Assumably, the project evaluation will give special attention to problems concerning indexing policy and procedures.




INTERNET

The Victorian Web / editor: George P. Landow. - (current updates)
Click here to see the website

URL: <http://www.stg.brown.edu/projects/hypertext/landow/victov.html>

Contains materials on all aspects of Victorian literature and society: Victorianism, literature, social context, economics, religion, philosophy, science, technology, politics, gender matters, visual arts, related www resources
 HTML-title: The Victorian Web Overview
 Description based on edition as seen october 1998
 References and link collections

Subjects: economics * religion * filosofi * art * science * technology * politics * social conditions * literature * english literature * 1800-1899 * England * websites * the Victorian period * biographies * bibliographies * time tables * reviews

Subject entries: 81.3 93.5 30.23



 
 Hjørring Library homepage

Fig.1. ISBD display of record, as shown in the Catalog of Hjørring Public Library (translated)

Rebecca Green emphasises the need for “*an increased role for bibliographic description in the electronic world*” [10, p.242]. She points out the problems regarding selection, which are inherent in most Internet search tools. Interviews with patrons in the Netkat project revealed, that the descriptions, found in the catalogs, met the requirements of even very experienced Internet users to a high degree.

Very good	Good	Suitable	Unsuitable	Very unsuitable
33%	28%	21%	16%	2%

Table 5. Satisfaction with display/description

The 10 patrons unsatisfied with the descriptions, were all Internet novices, while the majority of respondents satisfied with the descriptions were highly experienced Internet users.

Testing the project

One of the major purposes of the project was to gather indications of the utility and usability of incorporating Internet resources in the public library catalogs. To test the usability of OPAC’s facilitating direct access to Internet resources, 58 patrons were observed performing searches for Internet resources. The methodology used was “Simulated information need” [11], where users carry out tasks, equvaling what they could experience in real life situations. Test persons were primarily students, who were asked to retrieve Internet

resources, which could be used for school assignments such as writing a report about “employment of children”, “death penalty”, “Spain” a.s.o.

The test revealed how well users were able to comprehend the functionality of the catalog incorporating Internet resources. The user tests were followed up by interviews trying to delineate the users subjective experiences with the system.

The majority of the test persons experienced few or no problems searching the catalog for Internet resources covering “their” subject. The interviews however revealed, that 12% of the respondents found the catalog difficult to use. Problems primarily arose due to the web interface design, which led test persons to activate links to subjects and all Internet records, instead of the actual resource described in the record.

Very easy	Easy	Neither...nor	Difficult	Very difficult
50%	29%	9%	10%	2%

Table 6. Did the system ease your retrieval of information requested?

Conclusions

Selecting and cataloging Internet resources in public libraries, requires carefully developed selection and registration policies. In the Netkat project selection policies primarily accounted for the Internet resources ability to fill gaps in the collection.

The actual registration of resources was impeded by some characteristics of the Internet and the resources situated here. The chosen level of registration and the catalog maintenance policy accounted for potential changes in both location, content and design of the entities registred.

Although incorporating Internet resources in the catalog seems expensive, at least two of the four libraries involved in the project have chosen to continue the selection and cataloging of Internet resources. The two libraries are confident, that selection and registration cost will be minimized gradually, as subject specialists and cataloging staff are becoming familiar with the medium. Furthermore the utility for patrons, who are enabled to use the same resource at the same time, will release reference staff, since they are no longer troubled with a lot of students seeking materials for the same assignment.

References

- [1] **Xue-Ming Bao**. Challenges and Opportunities: A Report of the 1998 Library Survey of Internet Users at Seton Hall University. *College & Research Libraries*, 59(6), 1998, 535-543.
- [2] **Jul, Erik**. Now that we know the Answer, What are the Questions? *Journal of Internet Cataloging*, 1(3), 1997
- [3] <http://www.haworthpressinc.com/jjic/jic1nr3-42.html>
- [4] **Caywood, Carolyn**. Library Selection Criteria for WWW Resources. Last revised 2/99
- [5] <http://www6.pilot.infi.net/~carolyn/criteria.html>
- [6] **Johnson, Peggy**. Selecting Electronic Resources: Developing a Local Decision-Making Matrix. *Cataloging & Classification Quarterly*, 22(3/4), 1996, 9-24
- [7] **INDOREG : Internet Document REGistration : Project report**. The Danish Library Centre, 1997
- [8] <http://purl.dk/rapport/html.uk/>
- [9] **Hillman, Diane I.** "Parallel Universes" or Meaningful Relationships: Envisioning a Future for the OPAC and the Net. *Cataloging & Classification Quarterly*, 22(3/4), 1996, 97-103
- [10] **Katalogiseringsregler og bibliografisk standard for danske biblioteker** (Cataloging Rules and Bibliographic Standard for Danish Libraries). Biblioteksstyrelsen, 1998
- [11] <http://www.kat-format.dk>
- [12] **DBC's emneordliste** (Danish Library Centre subject vocabulary). Danish Library Centre, 1997
- [13] **Functional Requirements for Bibliographic Records**: Draft Report for World-Wide Review. Recommended by the IFLA Study Group on the Functional Requirements for Bibliographic Records. Frankfurt am Main: Deutsche Bibliothek, 1996
- [14] **Green, Rebecca**. Description in the Electronic Environment. *Advances in Knowledge Organization*, 5, 1996, 242-248
- [15] **11.Borlund, Pia and Ingwersen, Peter**. The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 53 (3), 1997, 225-250