Opening the closed-access medical journals: Internet-based sharing of institutions' access codes on a medical web-site

K Masters

Citation

K Masters. *Opening the closed-access medical journals: Internet-based sharing of institutions' access codes on a medical web-site*. The Internet Journal of Medical Informatics. 2009 Volume 5 Number 2.

Abstract

Introduction: Previous research examined the sharing of non-open access journal articles on a medical web site. This paper examines the sharing of institutions' access codes to non-open access journal databases on the same web site.Method: A 6-month snapshot of the forum postings was analysed to determine the number of access codes shared, the number and geographical locations of the institutions affected.Results: Total access codes shared: 491; total institutions affected: 248 in 40 countries.Discussion and Conclusions: Similar to the previous research, the ethics of such sharing must be considered, especially as ethics is so prominent in medical practice. A major issue, however, concerns security: a large amount of the access was obtained because of extremely poor security at the institutions. This also raises questions regarding the security of other online systems at these institutions. For journal publishers, new models of access to medical and other journals are urgently required.

INTRODUCTION

BACKGROUND

A previous research article published in this journal [1] referred to a medical web site that allowed participants to share journal articles. That article began by discussing some of the issues around Open Access (OA) journals versus Non-Open Access (NOA) journals. These issues included free access to journals, publishing times, readership, number of citings, prestige of journals, legal issues, peer-review, business models, publishing costs, technology infrastructure, business models, indexing services, standards, rewards for researchers and marketing [2-9].

These issues formed the context in which researchers, "including millions of students, teachers, physicians, scientists, and other potential readers, who do not have access to a research library that can afford to pay for journal subscriptions" [10] attempt to access NOA journal articles.

The research on the medical web site then briefly described the structure of the web site, with its more than 125,000 registered users and 300,000 posts in electronic forums. The focus of that research was on the practice that allowed participants with no legitimate access to NOA journals to request required articles from other participants. These articles would be found by those participants who did have access to them, and the articles were then posted to the web site, so that they could be accessible by any person who visited the web site. This exchange occurred in a sub-forum of the web site named "Databases & Journals – Requests and Enquiries."

Over a six-month period, a total of 6,587 articles were requested, 5,464 were returned, and these articles had been viewed by others a mean of 4.47 times.

ACCESSING DATABASES

Although a researcher may subscribe individually to a journal or database of journals, most access is through a library or institution that that utilises databases of journals selected on a range of criteria [11]. The library has a license, and issues an access code (usually a username and password) to the user. Various models of access may exist, but users at a university usually receive the access code upon registration as a student or staff member. At public libraries, users receive the access code when they become members of the library. Each user has a unique access code, and that code grants access to a range of databases referencing a range of journals. The user does not need to physically be in

the library to access the library system – this can be done through a web browser, using the assigned access codes.

Given that the medical web site discussed in the previous research aims at sharing journal articles, surely a far more effective method of sharing would be to share these institutions' access codes? Rather than requesting and receiving individual articles, the sharing of access codes would allow users to access full databases, and therefore retrieve any number of journal articles at their leisure.

Logic would tell us that this is happening to some extent, but the actual extent is not known. How much sharing is happening, and which institutions are having their access codes shared has not yet been investigated.

THE SETTING

The setting for this research is the same medical web site described in the previous research [1]. On that web site, a sub-forum in the "Databases & Journals" forum is labelled "University Passwords." In that sub-forum, users who have institutions' access codes post those access codes into forum messages, and others who require them use those access codes to access journal articles for which they would otherwise have to pay. The last publicly-visible message posted to the forum was on 23 November 2008.

Similar to the previous research, the aim of this research was to investigate the sub-forum, report on the number of postings, the locations of the sites accessed, and to discuss the implications of the findings.

METHOD

A 6-month snapshot (25 May to 23 November 2008) of all messages in the forum "University Passwords" was taken (25 May, rather than 24 May, was chosen so as to coincide with the dates selected in the previous research). All the messages posted between those dates were read, and access codes for institutions were extracted.

An 'access code' was defined as a username and password combination, or a single user code, or a proxy server Internet Protocol (IP) address and port (if no username and/or password were required). An institution was defined as a university, or public library, or publisher's web site, or an online database web site.

Repeated access codes were ignored, so that only unique access codes were noted. For each access code, the access details (URL or IP address and codes themselves) were recorded. The author visited each site listed, in order to verify the URL and institution, and to determine the geographical location of the institution.

In 12 instances, where the discussion indicated that the access codes might still be valid, the author attempted to use the access codes.

The results were placed into an MS-Excel spreadsheet. From the data, the following was determined: total number of access codes supplied, total number of institutions affected, and the number of access codes from each institution and from each country. In addition, the access codes were analysed to find patterns that might have been used in their creation and discovery.

For the same reasons cited in the previous research, using Eysenbach and Till [12] as a guide, permission to search the forum was not required. In addition, between the time of the study and the time of preparing this article, it appears that the site has either gone permanently off-line, or is now behind firewalls. Nevertheless, the site and the data still serve as an example of this type of activity.

RESULTS

OVERALL RESULTS

In total, during the six-month period, 491 access codes from 248 institutions in 40 countries were supplied in the forum. (For the purposes of reporting, Hong Kong is treated separately from China).

Of the 248 institutions, 229 (92.3%) were libraries at educational institutions, 12 (4.8%) were publisher or database sites, 6 (2.4%) were public library systems, and 1 (0.4%) was a professional organisation.

COUNTRIES

Table 1 shows the total number of institutions affected, broken down by country. For purposes of reporting, the publisher / databases are separated from the other sites.

Figure 1

Table 1: Total number of institutions (n=248) affected, listed by country, in decreasing rank order.

Rank	Country	Total Unique Institutions	Perc.
1	USA	133	53.6
2	Publisher / Data Base	12	4.8
3	Canada, Taiwan	10	4.0
5	Australia	8	3.2
6	China	7	2.8
7	South Africa	6	2.4
8	United Kingdom	5	2.0
9	Poland, Spain	4	1.6
11	France, Saudi Arabia, Turkey	3	1.2
14	Brazil, Germany, Hong Kong, Indonesia, Ireland, Japan, Korea, Lebanon, Malaysia, Netherlands, Portugal, Sweden	2	0.8
26	Algeria, Argentina, Bahrain, Belgium, Colombia, Denmark, India, Israel, Italy, Lithuania, Namibia, New Zealand, Norway, Puerto Rico, Switzerland, UAE	1	0.4

By far, the country with the largest number of affected institutions was the USA, with a total of 133 institutions. The next largest category was Publishers and Data Bases' direct sites, 12 of whom were affected. It is also obvious from Table 1 that almost the entire globe is affected. Interestingly, however, Lithuania is the only state from the former Soviet Union that is listed. There is no clear reason for this. Language is unlikely to be the cause, as many of the sites from Japan, Korea, China and elsewhere, were not in English.

ACCESS CODES

Table 2 shows the total number of unique access codes shared, broken down by country.

Figure 2

Table 2: Total number of unique access codes (n=491) shared, listed by country, in decreasing rank order.

Rank	Country	Total Unique Access Codes	Perc.
1	USA	333	67.8
2	Publisher / Data Base	25	5.1
3	Australia	22	4.5
4	Canada	12	2.4
5	Taiwan	11	2.2
6	South Africa	10	2.0
7	China	7	1.4
8	France, Spain, United Kingdom	6	1.2
11	Poland	4	0.8
12	Japan, Saudi Arabia, Turkey	3	0.6
15	Bahrain, Brazil, Germany, Hong Kong, Indonesia, Ireland, Korea, Lebanon, Malaysia, Netherlands, Portugal, Sweden, Switzerland	2	0.4
28	Algeria, Argentina, Belgium, Colombia, Denmark, India, Israel, Italy, Lithuania, Namibia, New Zealand, Norway, Puerto Rico, UAE	1	0.2

While there are minor differences in the country rankings between Table 1 and Table 2, an expected high Spearman Rank Correlation of 0.95 indicates a strong relationship between the number of institutions affected and the number of access codes shared.

INSTITUTIONS

The aim of this paper is not to embarrass institutions and publishers. As a result, individual institution names are not published here. What can be shown, however, is the number of access codes shared per institution, so that the reader can see how the codes were spread across the range of institutions, and that some institutions were far more affected than others.

Table 3 shows the number of access codes shared per institution, in rank order, indicating the number of institutions in each rank.

Figure 3

Table 3: Table showing the number of access codes (n=491) per institution, in rank order, indicating the number of institutions in each rank

Rank	Institutions	Unique Access Codes	Percentage
1	1	59	12.1
2	1	30	6.1
3	1	20	4.1
4	1	12	2.5
5	1	11	2.2
6	1	10	2.0
7	1	9	1.8
8	1	8	1.6
9	3	6	1.2
12	6	5	1.0
17	1	4	0.8
19	11	3	0.6
30	28	2	0.4
57	191	1	0.2

At this point, it must be remembered that Table 3 lists only the number of access codes shared, not the number than might be known and not shared. In the case of the first institution, experiments indicate that the number of access codes available is far higher, at a theoretical maximum of 9,999 (this institution is discussed in a little more detail below).

In several cases, access was not to a single library or university, but an entire library system. In one instance, one system had access to "35 public and community college libraries joined together."

Of the 12 sites tested by the author, the access codes for three were still valid at the end of March 2009. By May 10th 2009, one of these had expired; by September 10th 2009, the second had expired, but the third was still functioning.

SOURCE OF THE CODES

While it appeared that some users were sharing their own legitimate access codes, in almost all cases of code-sharing, the discussion led one to believe that the codes had been obtained by users patiently 'hacking' at the access pages, looking for patterns or other weaknesses. Sometimes, users obtained codes from other similar sharing sites (a practice frowned upon by the site's community).

NATURE AND DESCRIPTION OF THE ACCESS CODES

Access codes were accompanied by the name of the institution and / or a hypertext link (usually through an "ezproxy" server) into the library service or to the institution's home page with instructions on how to access the library page. IP address access codes supplied a link to the proxy service, and information regarding changes required to settings in the user's Internet browser. Usually the code was a username and password, but IP addresses and bar code numbers were also used.

PROCESS OF SHARING

There was a common sequence of events with the sharing. A code would be shared, followed by a period in which it was used, and then the code would 'die.' The death of an access code appeared to occur when the Information Technology (IT) support staff at the institution became aware of the fact that it was being used by multiple users, and disabled it. From the discussion in the forum, there appeared to be a constant fight between the code-sharers and the institutions' IT support staff in the process of finding, sharing and disabling access codes.

With this in mind, it is crucial to note that the number of times an institution features on the list of unique accesses shared is not an indication of the number of times the access code has been used, as some institutions do not know that they have been compromised. As evidenced by the author's ability to access some institutions for several months, some access codes allow access for a long time.

SECURITY

An important consideration is the security of the institutions affected. Although all the library systems have some type of log-in process, in many instances, the security was poor, and the systems lent themselves to easy cracking. Some examples of poor security were:

IT support staff frequently used very simple and obvious codes for testing purposes (even repeated for the username and password, or with no password at all). Typical examples that occurred are familiar to anyone who has worked on electronic systems: "12345," "tester," "test123," "testuser," "testpass," "demo," "helpdesk," "library," "guest," and "admin." This use of obvious codes or simple English words occurred even at universities where the web sites were not created in English (e.g. China).

Many institutions did not use a username and password combination, but simply a single code number, frequently 5 or 6 characters. Given the small number of digits, and the number of people who will have legal access to the site, with a little patience, it is very easy to guess a valid access code of this type.

Some universities tried to disguise this, by using 'systematic' methods for setting up usernames and passwords, and, once the method had been found, a large number of usernames and passwords could be discovered. For example, one university began the username with the number "294241041" then attached a 5-digit number (e.g. 60683) to that, to make "29424104160683" as a username. That 5-digit number was also the password for that username. This meant that, with enough patience, it became very easy to discover new usernames and passwords. A second university used "000xyzw" followed by a short series of numbers – again, easily cracked with a little patience. A third used a sequence of running numbers (e.g. from 13035531 to 13035540). A positive development, however, was that it appeared from the discussion in the forum, that at least two universities changed their password systems after they had been discovered.

In one instance, a valid username (although not the password) was supplied in the institutions' on-screen documentation. Because the password was only 4 numerical characters, it was obviously discovered very quickly.

In 17 instances, simply using the proxy IP address and port (usually 8080) was good enough – there was no username and password. (These instances are usually known as "Free Proxies.")

The worst security, however, was at a university in the USA that had a simple running sequence of 4 digits (no further password required). A total of 59 of these access codes were shared, but the implications were clear that others existed. On experimentation with this site, the author randomly entered 3 sets of 4-digit codes (1767, 1232, 1961) not listed

in the forum, and all three granted access to the site's databases. From there, the author was able to access an article (one of his own) still under copyright (from Medical Teacher 2005). It appears that most, if not all, 4-digit numbers will grant access to this site. The site has access to some 40 databases, including EBSCOHost, Academic Search Premier, ERIC and Medline.

DISCUSSION

The previous research described the way in which users can access online NOA journal articles by using forums on a medical web site to receive copies of those articles from users with legitimate access. This paper has shown how the sharing of information is expanded into the sharing of library and data base access information, thereby allowing users much wider accessibility to journal articles.

ETHICS AND FINANCE

Similar to the issues raised in the previous research, questions of ethics and finance are raised by these results. What are the ethical considerations of this web site, and others like it, especially considering the prominence of medical ethics in the field of medical practice? Unlike the previous research, there is simply no way to measure the financial impact of access code-sharing, but, given the number of members on the website, the impact is likely to be significant.

DATABASE SECURITY

The results indicate that, while some of the access codes were obtained by legitimate users' sharing their personal information, a large amount of information is obtained through very simple, albeit patient, hacking processes. Given that the publishers have a great deal invested in their publications, and that the institutions pay large amounts to access those publications, it is surely incumbent on those institutions to ensure that their security is tightened – whatever the definition of tight security might be, access through a simple 4- or 5-digit number cannot be considered secure.

INSTITUTIONAL SECURITY

But the problem of the database security has far wider implications for the institutions - that of the overall security of the institutions' other systems. If this lack of security is applied to systems to which institutions have a contractual obligation to maintain, what about the security of other online systems at the university? Given university resource constraints, it is likely that the IT support staff who are responsible for the security of the library databases are also responsible for the security of other online systems at the university. Are learning management systems, student records, performance records, staff confidential information, or any of their Internet and Intranet systems kept 'secure' in a similar fashion? If so, how much of this is being accessed (and possibly edited) illegally?

Given the implications of these questions, it is surely necessary to determine not only the wider extent of the sharing, but also the extent to which institutions are aware of this practice, and whether or not they are taking steps to ensure that only legitimate users have access to the materials.

THE WAY FORWARD

Finally, and again, in a vein similar to a conclusion reached in the previous research, it may be time for the publishers to realise that their current systems, run on such a large scale, across such a range of institutions (many of whom struggle for funding) are simply not secure, and probably never will be. Whether the solution lies in fighting harder to secure their rights, or moving towards a greater degree of openness, remains to be seen.

CONCLUSION

This paper has examined the sharing of institutional access codes to databases of non-open access journals on a medical web site. It has found that a large number of access codes, from across the world, are being shared. The prime issues raised are the ethics of this sharing and the extremely poor security implemented at the institutions affected, and indicates that new models of access to medical and other journals are urgently required.

References

1. Masters K: Opening the non-open access medical journals: Internet-based sharing of journal articles on a medical web site. Internet Journal of Medical Informatics; 2009; 5(1): http://tinyurl.com/kmoajournals (Accessed 28/10/2009).
2. Antelman K: Do open access articles have a greater research impact? College & Research Libraries News; 2004; 65(5): 372-82.
3. Björk B-C: Open access to scientific publications - an analysis of the barriers to change? Information Research; 2004; 9(2): http://InformationR.net/ir/9-2/paper170.html (Accessed 07/01/2009).
4. Butler D: Who will pay for open access? Nature; 2003; 425(6958): 554-5.
5. Crawford BD: Open-access publishing: Where is the value? The Lancet; 2003; 362: 1578-80.

6. May C: The Academy's new electronic order? Open source journals and publishing Political Science. European Political Science; 2005; 4(1): 14-24.
7. Plutchak TS: Embracing open access. J Med Libr Assoc; 2004; 92(1): 1-3.
8. Swan A, Brown S: Authors and open access publishing. Learned Publishing; 2004; 17: 219-24.
9. Swan A, Brown S: Open access self-archiving: An author study. 2005;.http://www.jisc.ac.uk/uploaded_documents/Open%20Access%20SelfArchiving-an%20author%20study.pdf. Joint Information Systems Committee (Accessed 08/01/2009).
10. Brown PO, Eisen MB, Varmus HE: Why PLoS Became a Publisher. PLos Biol; 2003; 1(1): e36.
11. Nisonger TE: Electronic journal collection management issues. Collection Building; 1997; 16(2): 58-65.
12. Eysenbach G, Till JE: Ethical issues in qualitative research on internet communities. BMJ; 2001; 323: 1103-5.

Author Information

Ken Masters, PhD, ITHealthEd