Updating Proxy Server with Dealer Domains
Yosef Branse
Younes and Soraya Nazarian Library
University of Haifa

IGeLU Conference
September 2011
A couple of caveats

- Not a specifically SFX solution, rather a way to deal with a general problem of which SFX is one element
- I know virtually **nothing** about proxies!
Background

- Access to electronic resources from off-campus requires authentication
- Proxy server (Squid) maintained by University computing division
- Authorized users (staff, students) connect to University network and log in with username and password
- Connections to vendor sites “routed” through remote.pac
The Problem

- New resources must be recorded in the proxy server, identified by domain name of dealer site
- Original method: email from librarians or library tech support to proxy maintainer, as the need arose
- Not efficient, and susceptible to error
The Solution

- Periodically assemble list of recently added domains
- Send list automatically by email to proxy manager
Sources of domain information

- Aleph catalog – MARC tag 856 in bibliographic records for ejournals and etexts - retrieved via p_ret01, p_ret_03
- SFX – URLs of target resources – retrieved via Export Tool (run from shell script)
- Online databases – Excel file maintained locally by Reference librarian
Processing data

- Merge lists from three sources into unified file
- Remove all text following ‘/’, leaving only domain name
- Exclude items from within University domain
- Exclude domains with no ‘.’ delimiter
- Convert from upper to lower case, for uniformity
- Sort list, eliminating duplicates
Processing data (2)

- Compare list with version saved from previous run – Linux ‘diff’ utility produces list of new items
- Review list with text editor, make changes if necessary
- Approve forward of list by email to proxy administrator
- Save master list as control for next run
List of recently added domains

asia-studies.com
atilf.fr
authorities.loc.gov
beazley.ox.ac.uk
biblioline.nisc.com
bodd.cf.ac.uk
bookreviews.org
brillonline.nl
brs.leeds.ac.uk
c-spanvideo.org
cabi.org
catchword.com
chabadlibrary.org
charleston.publisher.ingentaconnect.com
charlestonco.com
data.un.org
dictionary.oed.com
Refinement – multisite domains

journal1.myvendor.com
journal2.myvendor.com
journal3.myvendor.com
journal4.myvendor.com

myvendor.com
Generate ‘inverted’ list of domains

<table>
<thead>
<tr>
<th>com.chadwyck.eebo</th>
<th>il.ac.huji.geobase</th>
</tr>
</thead>
<tbody>
<tr>
<td>com.chadwyck.fiaf</td>
<td>il.ac.huji.gr.webmaster</td>
</tr>
<tr>
<td>com.chadwyck.iipa</td>
<td>il.ac.huji.hebrew-terms</td>
</tr>
<tr>
<td>com.ingentaconnect.hermia</td>
<td>il.ac.huji.hebrew-treasures</td>
</tr>
<tr>
<td>com.isiknowledge.admin</td>
<td>il.ac.huji.igdc</td>
</tr>
<tr>
<td>com.isiknowledge.apps</td>
<td>il.ac.huji.jpress</td>
</tr>
<tr>
<td>com.metapress.ergonomics</td>
<td>il.ac.tau.ariela</td>
</tr>
<tr>
<td>com.oed.dictionary</td>
<td>il.ac.tau.ariela1</td>
</tr>
<tr>
<td>com.proquest.pqdtopen</td>
<td>il.ac.tau.genesis</td>
</tr>
<tr>
<td>com.proquest.search</td>
<td>il.ac.tau.www2</td>
</tr>
<tr>
<td>com.umi.roquest</td>
<td>il.co.bizportal.lxsrv1</td>
</tr>
<tr>
<td>com.umi.wwwlib</td>
<td>il.co.digitaler.tamrurim</td>
</tr>
<tr>
<td>edu.fau.faujsa</td>
<td>org.oclc.firstsearch</td>
</tr>
<tr>
<td>edu.iastate.public</td>
<td>uk.ac.cf.bodd</td>
</tr>
<tr>
<td>edu.princeton.cogsci</td>
<td>uk.ac.jisc</td>
</tr>
<tr>
<td>edu.uwm.lib.geobib</td>
<td>uk.ac.leeds.brs</td>
</tr>
<tr>
<td>edu.uwm.lib.leardo</td>
<td>uk.ac.ox.beazley</td>
</tr>
<tr>
<td>gov.ed.eric</td>
<td>uk.org.cornuco</td>
</tr>
<tr>
<td>gov.loc.authorities</td>
<td>uk.org.fihr</td>
</tr>
<tr>
<td>gov.nih.nlm.gateway</td>
<td></td>
</tr>
<tr>
<td>gov.usda.nal.agricola</td>
<td></td>
</tr>
<tr>
<td>hk.hku.lib.sunzi1</td>
<td></td>
</tr>
</tbody>
</table>
Generate ‘inverted’ list of domains (2)

Technique is of limited utility, since domains from same vendor may occur in lists generated at different dates.

Ideally, a program would check for domains already recorded in proxy.

BUT – probably not worth the effort.
Does proxy-update scheme work?

Since introduced in late 2007, and run twice each month, few or no incidents of resources being inaccessible due to proxy issues
Is this useful for you?

Depends on, e.g.:
- the nature of your site
- sources of information about dealers’ domains
- proxy server in place
- institution’s access policies
- technical facility in setting up and running script
- liaison with proxy management
Thank you!

Yosef Branse
University of Haifa
Younes and Soraya Nazarian Library
Information Systems Department
jody@univ.haifa.ac.il
972-4-8240288
972-52-8666430

http://aleph.haifa.ac.il/www/public/Igelu_2011/yb/Proxy_Updates.pps