How Much is Too Much?

Excessive Downloading of Articles At Macquarie
WHEN?

March – April 2007

Focus of presentation is on the above date, but Previous incidents of excessive downloading had occurred in late January 2007, December 2006, and July 2005.
WHICH JOURNALS?

• Nature
• Applied Physics Letters
• Annual Review of Plant Biology and Annual Review of Anthropology

• Earlier downloads were from Modern Language Journal (JSTOR) and a Science Direct biology title.
In the Beginning ...

• Vendor’s server detects excessive downloading
• Usually, access is blocked (Nature and AIP) to all users of the proxy server or the IP range
• Email sent by vendor to Library contact advising of them of incident
• Email/phone queries from Library users about lack of access
FIRST RESPONSE

• Request precise time of incident from vendor
• Request EZproxy log files from ITS
• Analyse log files to identify ID No. of user
• Deactivate user’s account
• Advise vendor that the offending account has been stopped while investigation proceeds.
• Ask them if access can be restored
SECOND RESPONSE

PHASE 1

• Examine closely the log file pattern

• Always shows abnormal activity indicating use of robots or special scripting e.g. an article per second being downloaded in the case of Nature journal

• Sometimes there have been attempts to hide the user ID number
SECOND RESPONSE

PHASE 2

• Contacted user identified and explained circumstances

• Asked them if they saved their password on their PC (usually have auto-complete ticked), given it to others, what type of ISP account (cable, ADSL, wireless), PC security software, Windows updates loaded, and research areas?

• Most were distance students
USER’S RESPONSE

• Generally very concerned (Access is blocked!)
• Usually not that IT literate and unaware that the IE Auto-Complete options are selected
• Sometimes have no Internet security software activated and no knowledge of Windows updates
• Not studying or enrolled in the subject area of the downloads
Copy of the notices given to users before accessing content on Nature

Description of current procedures to detect and address suspicious behaviour

Description of what happened in this case and what was done with the content

Statement of steps being undertaken to address the issue.
Access only restored:
1. After email responding to questions to local Australian representative
2. Followed by phone conversation with Tokyo Asia Pacific manager
3. And a second response to questions, this time on a Macquarie University letterhead, faxed and then posted.
• Most database licences include a provision stating that universities using their product do not allow, and take steps to prevent, excessive downloading.

• Nature Journal’s request, essence, was for Macquarie to investigate and implement detection and blocking by the licensee rather than the licensor.
IMPACT OF EXCESSIVE DOWNLOADING

• Loss of access for over a week
• Distorted usage statistics

2007 Feb Mar April
May

APL 42 18663 3239 74
Nature 747 62,316 697 1254

• Need to investigate technical solutions and update such as using newer EZproxy version

• Improved online signage of user
PROGRESS SO FAR ...

• EZproxy software not updated yet. This is part of a broader project. The server and software are not located in the Library.

• Still investigating what the settings should be, if excessive downloading is to be detected and stopped by the University.

• Need to have more explicit advice which users see when accessing an electronic resource.
DEFINING “TOO MUCH”

What is “Too Much”?
1. When a university’s server detects specific software in use such as robots?
2. When the total amount a user downloads exceeds a specific figure?
3. When the speed at which the downloading occurs exceeds a pre-specified limit?
A COMMON SOLUTION?

An agreed response to a shared problem?