Usage Statistics and Demo Services

Leslie Carr, University of Southampton

http://id.ecs.soton.ac.uk/people/60

Interoperable Repository Project, JISC

Problem

- We would like to be able to find out
  - How much usage is my paper getting?
  - How much usage is my repository getting?
  - How much use are Open Access repositories getting in the UK?
  - How often are *Nature*'s articles being read from repositories?
- But repository & service usage stats are uncommon and incomparable
  - Full texts vs metadata
  - Web pages vs eprints
  - Internal operations vs public data

Usage Statistics

- Usage of what? By whom?
- Who wants to know? Why?

- There are three kinds of lies:
  lies, damned lies and statistics

Mark Twain quoting Benjamin Disraeli

Usage Statistics

- Other websites want to know about *hits* or *downloads* or *visits*
- By contrast, researchers would really like to know
  - *How many people have read, understood and engaged with my work?*
- Unfortunately, we only have web logs as evidence

What? Web Logs?

- Everyone knows that web logs are unreliable and incomplete
  - Many downloads are missed
  - Downloads are not tied to individuals, but hosts
- Only citation analysis is guaranteed and ‘proper’

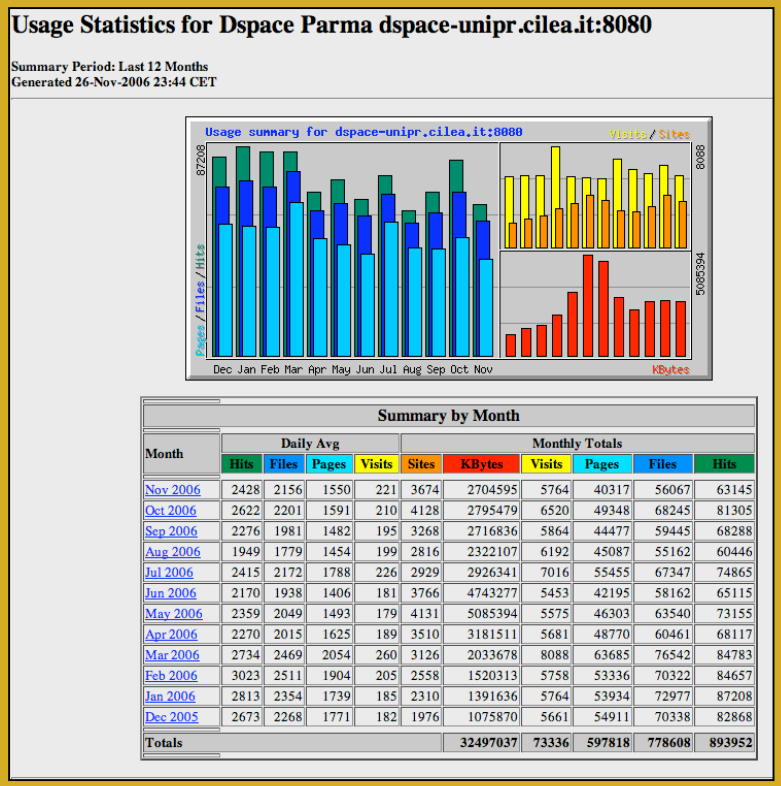
- But both are auditable evidence of some type of use of the article

Asides

- Well, perhaps we should have questionnaires that are filled out after download?
- Perhaps readers should rate, review or provide feedback to articles after they have read them
- Or perhaps we should use cookies to track individual researchers?
- Or perhaps the convenience of the user is more important than the whims of the author?
- Or perhaps not…

Repository Usage Statistics

- Web-based repository site statistics

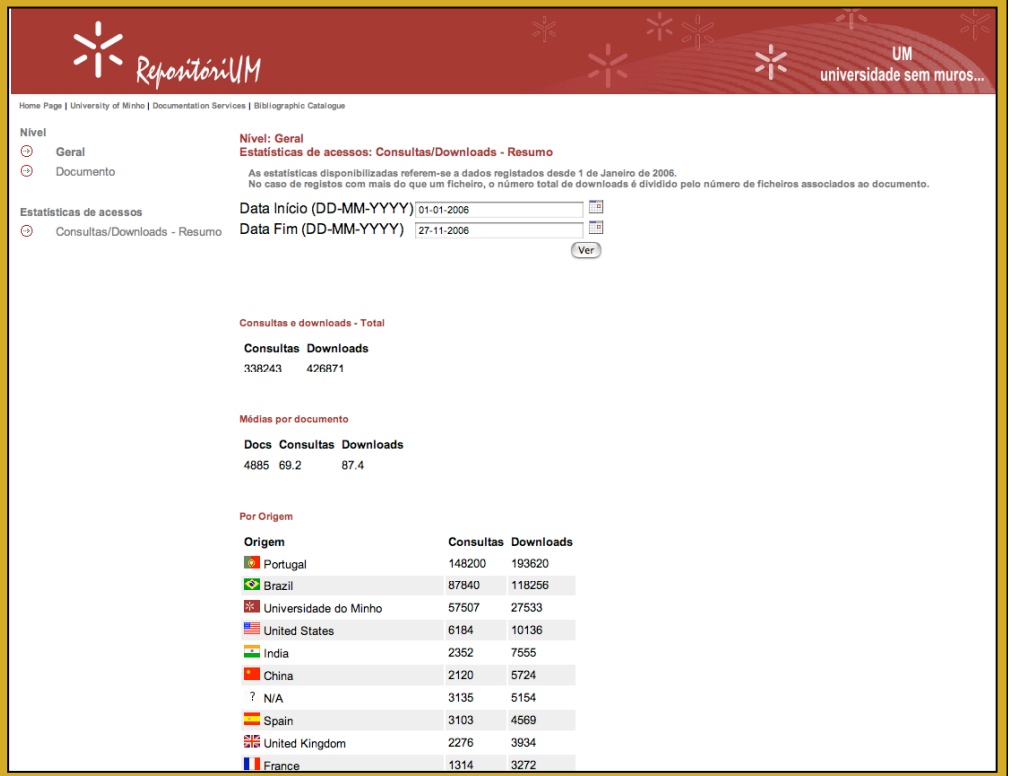


Usage Statistics for Dspace Parma dspace-unipr.cilea.it:8080

Summary Period: Last 12 Months
Generated 26-Nov-2006 23:44 CET

Summary by Month

Month	Daily Avg				Monthly Totals					
	Hits	Files	Pages	Visits	Sites	KBytes	Visits	Pages	Files	Hits
Nov 2006	2428	2156	1550	221	3674	2704595	5764	40317	56067	63145
Oct 2006	2622	2201	1591	210	4128	2795479	6520	49348	68245	81305
Sep 2006	2276	1981	1482	195	3268	2716836	5864	44477	59445	68288
Aug 2006	1949	1779	1454	199	2816	2322107	6192	45087	55162	60446
Jul 2006	2415	2172	1788	226	2929	2926341	7016	55455	67347	74865
Jun 2006	2170	1938	1406	181	3766	4743277	5453	42195	58162	65115
May 2006	2359	2049	1493	179	4131	5085394	5575	46303	63540	73155
Apr 2006	2270	2015	1625	189	3510	3181511	5681	48770	60461	68117
Mar 2006	2734	2469	2054	260	3126	2033678	8088	63685	76542	84783
Feb 2006	3023	2511	1904	205	2558	1520313	5758	53336	70322	84657
Jan 2006	2813	2354	1739	185	2310	1391636	5764	53934	72977	87208
Dec 2005	2673	2268	1771	182	1976	1075870	5661	54911	70338	82868
Totals						**32497037**	**73336**	**597818**	**778608**	**893952**



RepositóriUM

UM universidade sem muros...

Home Page | University of Minho | Documentation Services | Bibliographic Catalogue

Nível
- Geral
- Documento

Estatísticas de acessos
- Consultas/Downloads - Resumo

Nível: Geral
Estatísticas de acessos: Consultas/Downloads - Resumo

As estatísticas disponibilizadas referem-se a dados registados desde 1 de Janeiro de 2006.
No caso de registos com mais do que um ficheiro, o número total de downloads é dividido pelo número de ficheiros associados ao documento.

Data Início (DD-MM-YYYY) 01-01-2006
Data Fim (DD-MM-YYYY) 27-11-2006
Ver

Consultas e downloads - Total

Consultas	Downloads
338243	426871

Médias por documento

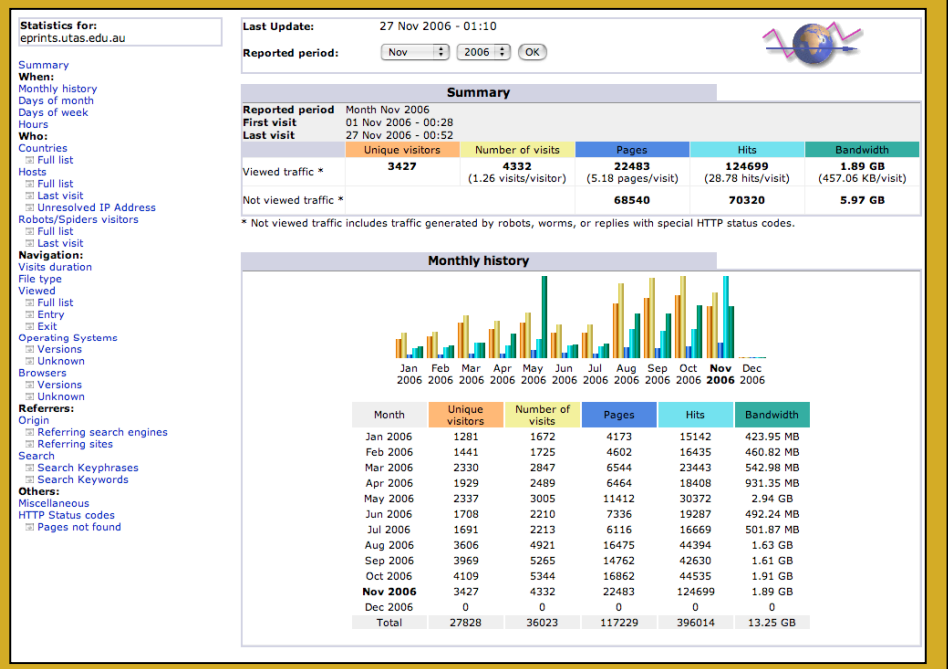
Docs	Consultas	Downloads
4885	69.2	87.4

Por Origem

Origem	Consultas	Downloads
Portugal	148200	193620
Brazil	87840	118256
Universidade do Minho	57507	27533
United States	6184	10136
India	2352	7555
China	2120	5724
N/A	3135	5154
Spain	3103	4569
United Kingdom	2276	3934
France	1314	3272

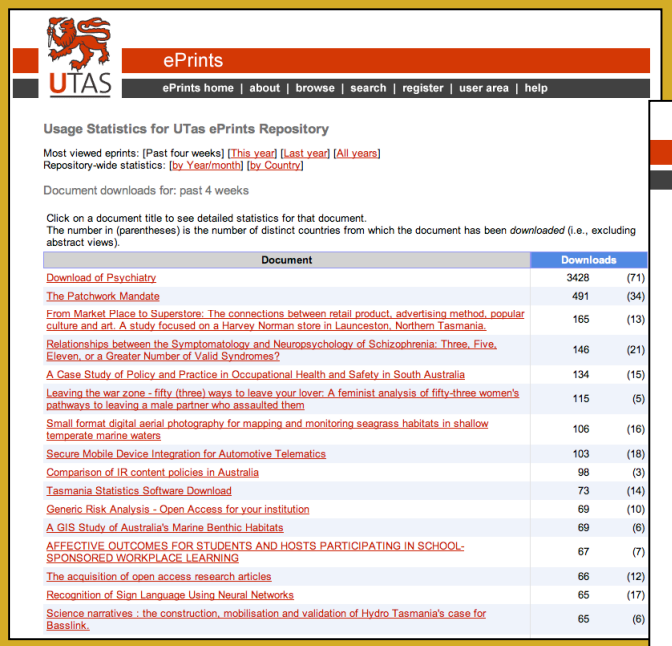
Repository Usage Statistics

- Web-based repository site statistics



Individual Articles

- More difficult to report about individual articles



Answering Questions

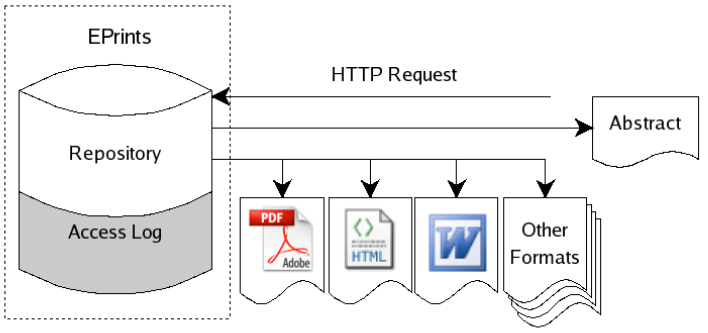
- Interoperable Repository Statistics (IRS, JISC) project
  - Help understand access patterns
  - Who is downloading what? Why?
  - What affects the different rates of downloads?

  - Initial survey told us that in-depth stats were important.
  - Augmenting COUNTER stats is a useful objective

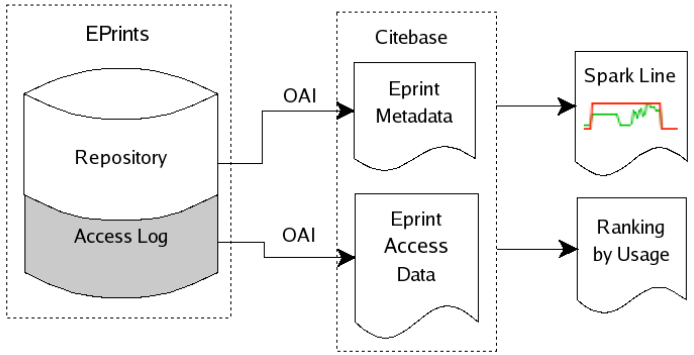
Proposal

- Shared usage collection procedure
  - Agreed policies and interpretation
    - Privacy, longevity, spider-filtering, multi-downloading
  - Interoperable data streams
  - Knowledge Exchange EU project
- Sample statistics services
  - shared between repositories
- Sample statistics plugins
  - local to a repository

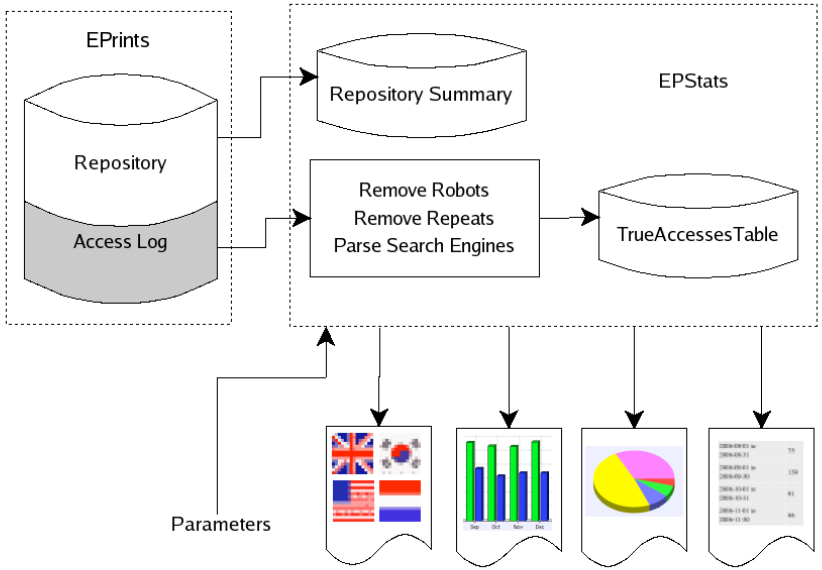
IRS Infrastructure



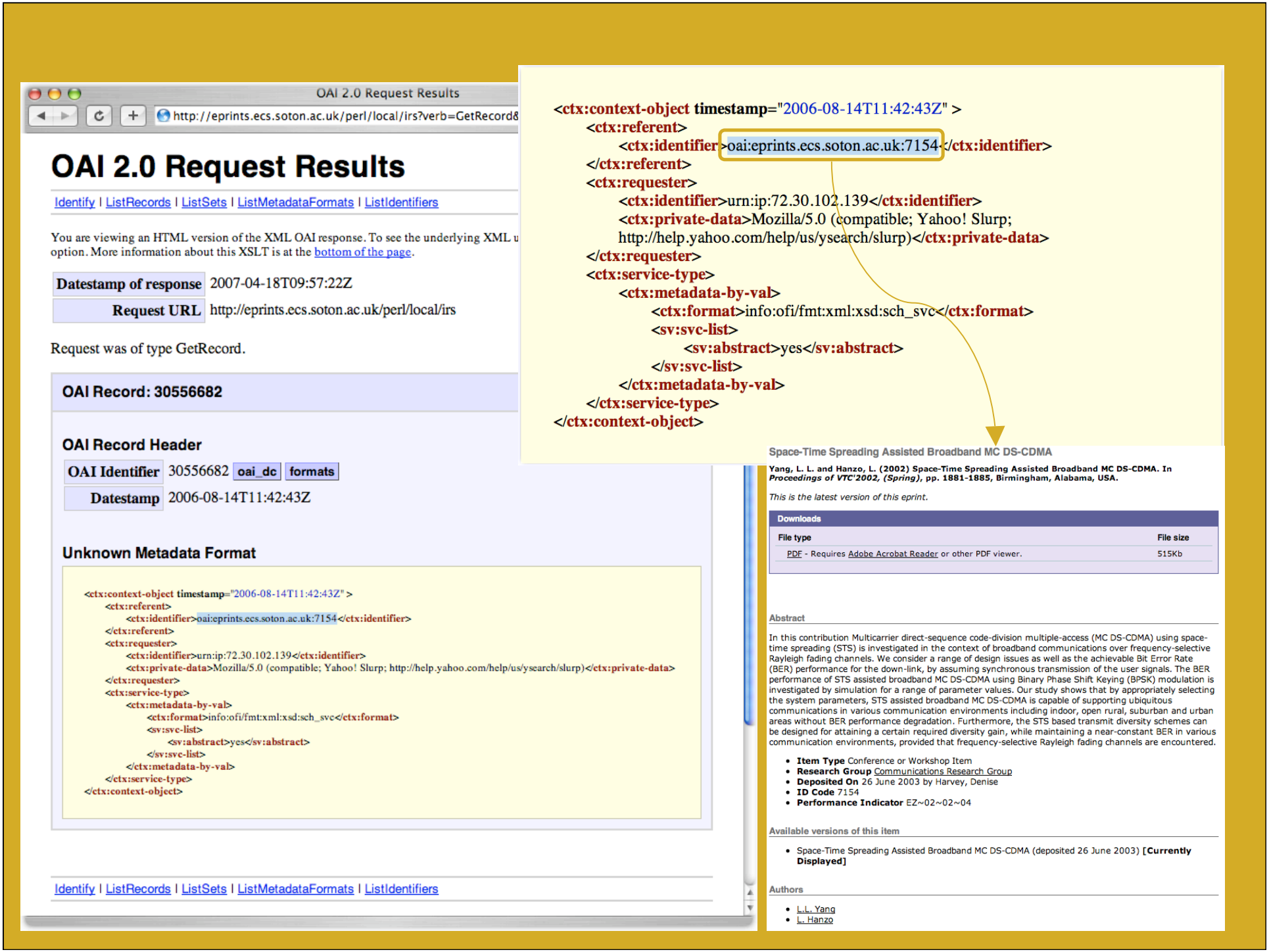
EPrints logs every access to an document or an abstract page.



The Eprint Metadata and Access Log are made available via OAI.
Citebase makes use of this to provide Spark Lines and Usage Ranking.

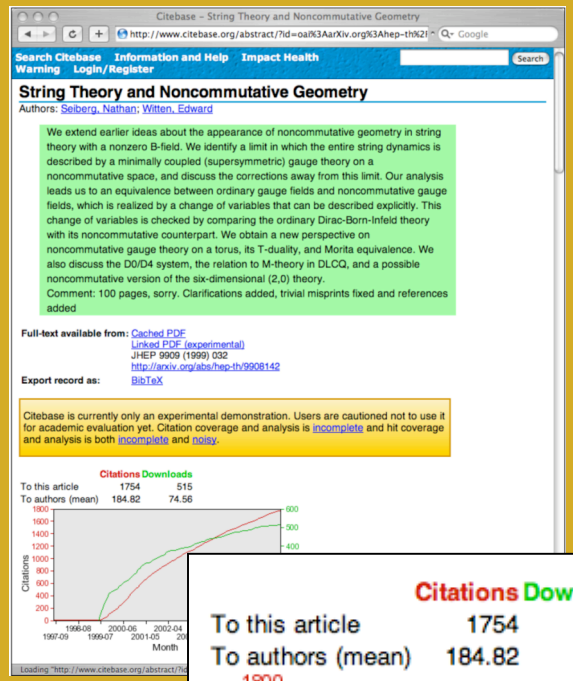


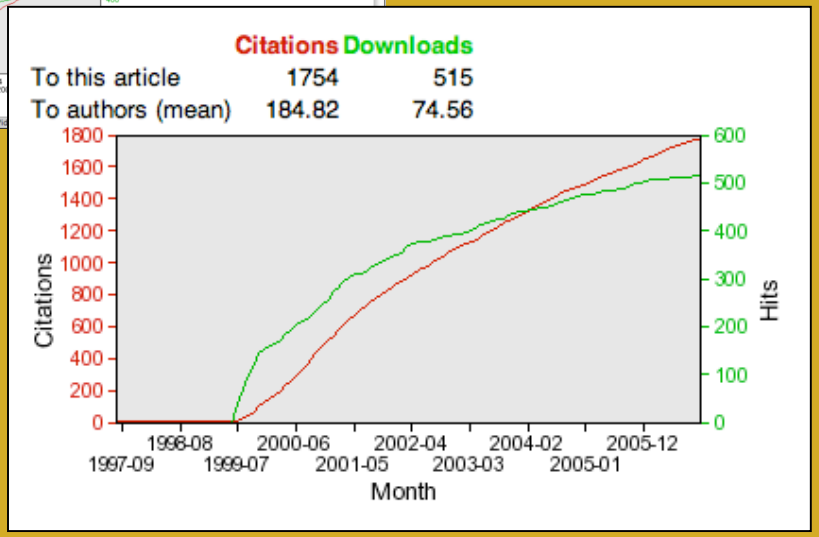
EPStats uses the data from EPrints to provide a wide variety of graphs and tables
to assist with analysis of repository statistics.


OAI 2.0 Request Results
http://eprints.ecs.soton.ac.uk/perl/local/irs?verb=GetRecord&
OAI 2.0 Request Results
Identify | ListRecords | ListSets | ListMetadataFormats | ListIdentifiers
You are viewing an HTML version of the XML OAI response. To see the underlying XML u
option. More information about this XSLT is at the bottom of the page.
Datestamp of response 2007-04-18T09:57:22Z
Request URL http://eprints.ecs.soton.ac.uk/perl/local/irs
Request was of type GetRecord.
OAI Record: 30556682
OAI Record Header
OAI Identifier 30556682 oai_dc formats
Datestamp 2006-08-14T11:42:43Z
Unknown Metadata Format
<ctx:context-object timestamp="2006-08-14T11:42:43Z" >
<ctx:referent>
<ctx:identifier>oai:eprints.ecs.soton.ac.uk:7154</ctx:identifier>
</ctx:referent>
<ctx:requester>
<ctx:identifier>urn:ip:72.30.102.139</ctx:identifier>
<ctx:private-data>Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.com/help/us/ysearch/slurp)</ctx:private-data>
</ctx:requester>
<ctx:service-type>
<ctx:metadata-by-val>
<ctx:format>info:ofi/fmt:xml:xsd:sch_svc</ctx:format>
<sv:svc-list>
<sv:abstract>yes</sv:abstract>
</sv:svc-list>
</ctx:metadata-by-val>
</ctx:service-type>
</ctx:context-object>
Identify | ListRecords | ListSets | ListMetadataFormats | ListIdentifiers
<ctx:context-object timestamp="2006-08-14T11:42:43Z" >
<ctx:referent>
<ctx:identifier>oai:eprints.ecs.soton.ac.uk:7154</ctx:identifier>
</ctx:referent>
<ctx:requester>
<ctx:identifier>urn:ip:72.30.102.139</ctx:identifier>
<ctx:private-data>Mozilla/5.0 (compatible; Yahoo! Slurp;
http://help.yahoo.com/help/us/ysearch/slurp)</ctx:private-data>
</ctx:requester>
<ctx:service-type>
<ctx:metadata-by-val>
<ctx:format>info:ofi/fmt:xml:xsd:sch_svc</ctx:format>
<sv:svc-list>
<sv:abstract>yes</sv:abstract>
</sv:svc-list>
</ctx:metadata-by-val>
</ctx:service-type>
</ctx:context-object>
Space-Time Spreading Assisted Broadband MC DS-CDMA
Yang, L. L. and Hanzo, L. (2002) Space-Time Spreading Assisted Broadband MC DS-CDMA. In Proceedings of VTC'2002, (Spring), pp. 1881-1885, Birmingham, Alabama, USA.
This is the latest version of this eprint.
Downloads
File type
File size
PDF - Requires Adobe Acrobat Reader or other PDF viewer.
515Kb
Abstract
In this contribution Multicarrier direct-sequence code-division multiple-access (MC DS-CDMA) using space-time spreading (STS) is investigated in the context of broadband communications over frequency-selective Rayleigh fading channels. We consider a range of design issues as well as the achievable Bit Error Rate (BER) performance for the down-link, by assuming synchronous transmission of the user signals. The BER performance of STS assisted broadband MC DS-CDMA using Binary Phase Shift Keying (BPSK) modulation is investigated by simulation for a range of parameter values. Our study shows that by appropriately selecting the system parameters, STS assisted broadband MC DS-CDMA is capable of supporting ubiquitous communications in various communication environments including indoor, open rural, suburban and urban areas without BER performance degradation. Furthermore, the STS based transmit diversity schemes can be designed for attaining a certain required diversity gain, while maintaining a near-constant BER in various communication environments, provided that frequency-selective Rayleigh fading channels are encountered.
Item Type Conference or Workshop Item
Research Group Communications Research Group
Deposited On 26 June 2003 by Harvey, Denise
ID Code 7154
Performance Indicator EZ~02~02~04
Available versions of this item
Space-Time Spreading Assisted Broadband MC DS-CDMA (deposited 26 June 2003) [Currently Displayed]
Authors
L.L. Yang
L. Hanzo

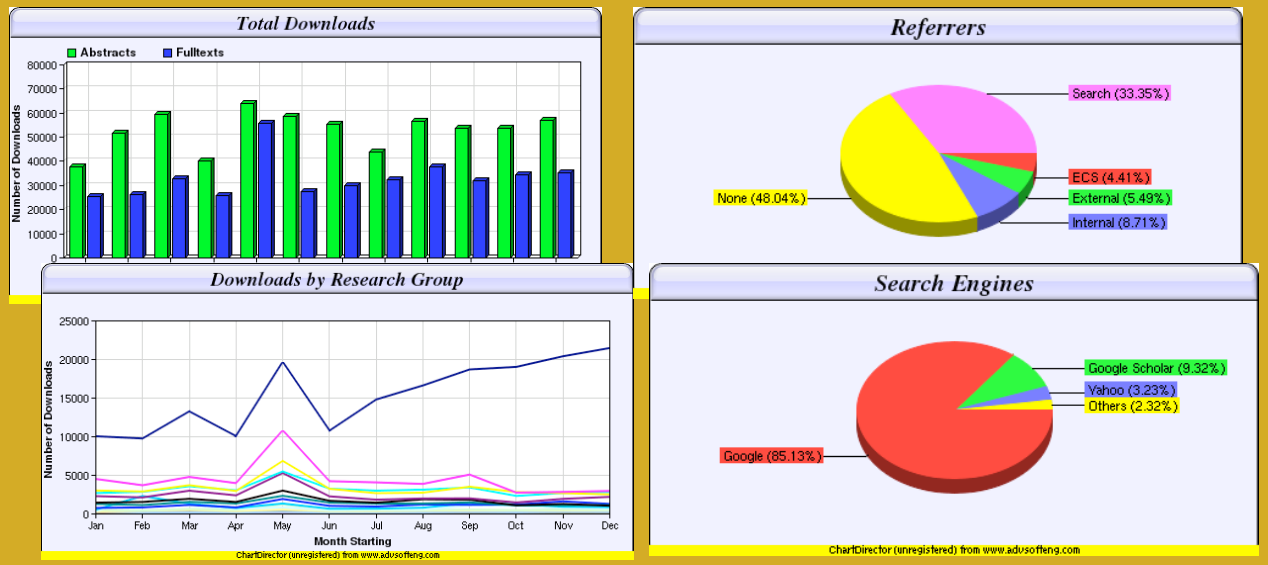
External Services

- *An external service can use stats for its own purposes*
- *or export visualisations & analyses back to repositories*





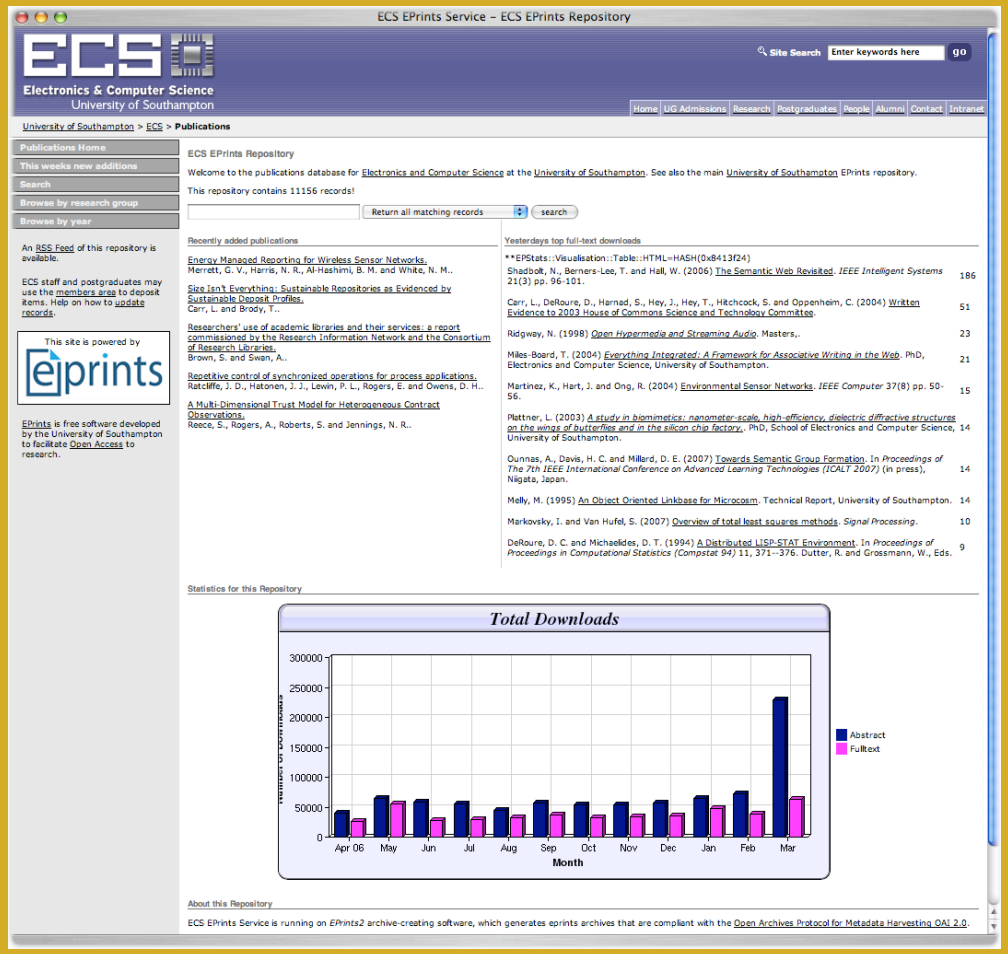
IRS Repository Stats



…or the repository can do its own analyses

- High level, aggregated over whole repository, over a year
- Comparisons between communities
- Summaries of most downloaded papers

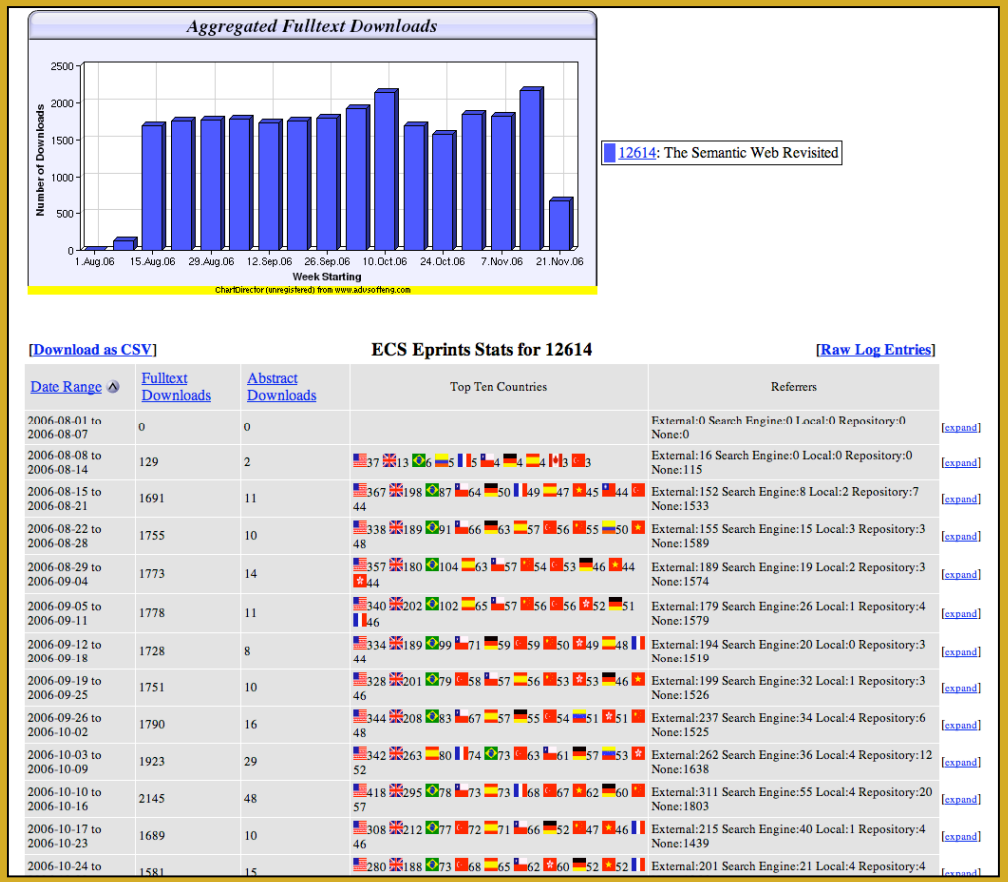
IRS Repository Stats



- Repository home page
  - Recent deposits
  - Yesterday's most downloaded (what's hot?)
  - Graph of all downloads over year (how successful is the whole school?).
- Policy is to minimise divisive comparison
- Art to finding interesting stats and graphs

IRS Repository Stats (2)

- Focus on single paper
  - usage over time
  - summary of sources and referers

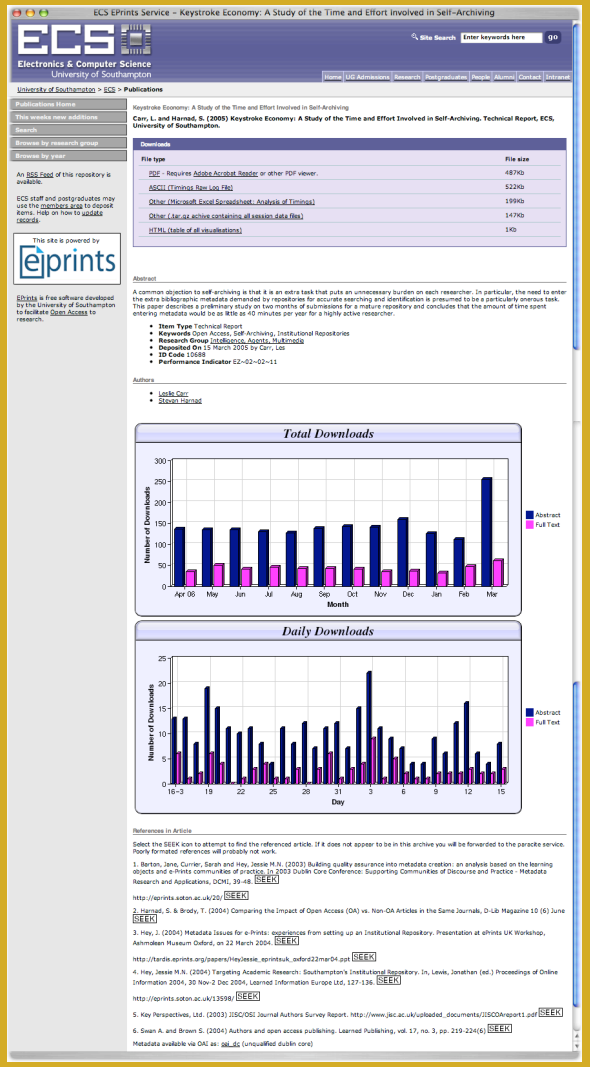


[Download as CSV] **ECS Eprints Stats for 12614** [Raw Log Entries]

Date Range	Fulltext Downloads	Abstract Downloads	Top Ten Countries	Referrers	
2006-08-01 to 2006-08-07	0	0		External:0 Search Engine:0 Local:0 Repository:0 None:0	[expand]
2006-08-08 to 2006-08-14	129	2	37 13 6 5 5 4 4 4 3 3	External:16 Search Engine:0 Local:0 Repository:0 None:115	[expand]
2006-08-15 to 2006-08-21	1691	11	367 198 87 64 50 49 47 45 44 44	External:152 Search Engine:8 Local:2 Repository:7 None:1533	[expand]
2006-08-22 to 2006-08-28	1755	10	338 189 91 66 63 57 56 55 50 48	External:155 Search Engine:15 Local:3 Repository:3 None:1589	[expand]
2006-08-29 to 2006-09-04	1773	14	357 180 104 63 57 54 53 46 44 44	External:189 Search Engine:19 Local:2 Repository:3 None:1574	[expand]
2006-09-05 to 2006-09-11	1778	11	340 202 102 65 57 56 56 52 51 46	External:179 Search Engine:26 Local:1 Repository:4 None:1579	[expand]
2006-09-12 to 2006-09-18	1728	8	334 189 99 71 59 59 50 49 48 44	External:194 Search Engine:20 Local:0 Repository:3 None:1519	[expand]
2006-09-19 to 2006-09-25	1751	10	328 201 79 58 57 56 53 53 46 46	External:199 Search Engine:32 Local:1 Repository:3 None:1526	[expand]
2006-09-26 to 2006-10-02	1790	16	344 208 83 67 57 55 54 51 51 48	External:237 Search Engine:34 Local:4 Repository:6 None:1525	[expand]
2006-10-03 to 2006-10-09	1923	29	342 263 80 74 73 63 61 57 53 52	External:262 Search Engine:36 Local:4 Repository:12 None:1638	[expand]
2006-10-10 to 2006-10-16	2145	48	418 295 78 73 73 68 67 62 60 57	External:311 Search Engine:55 Local:4 Repository:20 None:1803	[expand]
2006-10-17 to 2006-10-23	1689	10	308 212 77 72 71 66 52 47 46 46	External:215 Search Engine:40 Local:1 Repository:4 None:1439	[expand]
2006-10-24 to	1581	15	280 188 73 68 65 62 60 52 52	External:201 Search Engine:21 Local:4 Repository:4	[expand]

IRS Repository Stats (2)

- Focus on single paper
  - usage over time
- Problem:
  12,000 eprints * n graphs of whole year analyses = too many hours daily regeneration



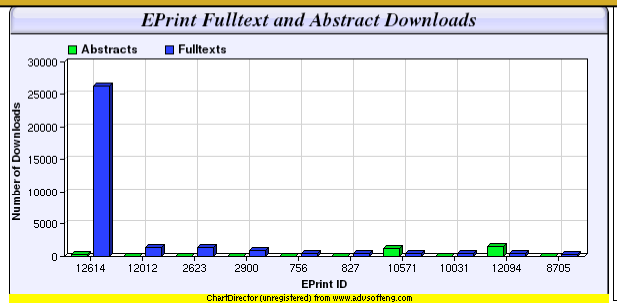
IRS Repository Stats (3)

2006-08-31 17:32:09	12614	F	Colt telecom	213.229.146.15		None			
2006-08-31 17:43:35	12614	F	Covad Communications	h-72-244-103-25.lsanca54.covad.net		None			
2006-08-31 17:49:21	12614	A	Cyprus International University, Lefkosa, KKTC	marge.levent.com		External			http://www.w3.org/2001/sw/
2006-08-31 17:49:21	12614	F	Cyprus International University, Lefkosa, KKTC	marge.levent.com		External			http://www.w3.org/2001/sw/
2006-08-31 17:49:59	12614	F	Connected by GTS Poland	217.153.131.234		None			
2006-08-31 17:50:19	12614	F	EgyNet	84.36.84.162		None			
2006-08-31 17:59:53	12614	F	o2 Germany GmbH & Co. OHG	pat-97.sg.de.o2.com		External			http://www.w3.org/2001/sw/
2006-08-31 18:01:48	12614	F	CHINANET fujian province network	218.5.5.48		None			
2006-08-31 18:11:04	12614	F	Illinois Century Network	st-209-7-87-12.gesd.n-cook.k12.il.us		None			
2006-08-31 18:16:44	12614	F	Matsuyama Canon OA System Corporation	giant.cr-office.co.jp		None			

- Download details
  - Every article has a different story, and different usage patterns
  - *Why is this article popular?*

Begin with High Level

- Which items are most downloaded?
- *why not ask which are LEAST downloaded??*



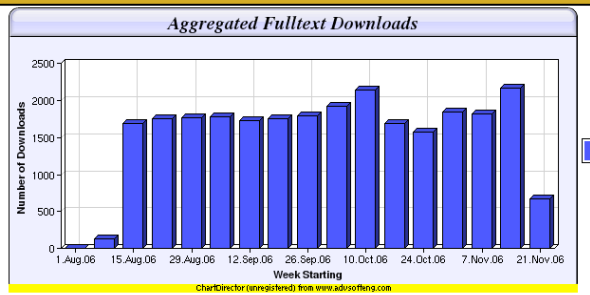
12614: The Semantic Web Revisited
12012: Mutualism, Parasitism, and Evolutionary Adaptation
2623: Free at Last: The Future of Peer-Reviewed Journals
2900: Implementing Peer Review on the Net: Scientific Quality Control in Scholarly Electronic Journals
756: Data Structures for Image Processing in C
827: Open Hypermedia and Streaming Audio
10571: M-grid: Using Ubiquitous Web Technologies to create a Computational Grid
10031: A study in biomimetics: nanometer-scale, high-efficiency, dielectric diffractive structures on the wings of butterflies and in the silicon chip factory.
12094: Opening Access by Overcoming Zeno's Paralysis
8705: For Whom the Gate Tolls?

[Download as CSV] **ECS Eprints Stats for Top 10** [Raw Log Entries]

Eprint ID(s)	Fulltext Downloads	Abstract Downloads	Top Ten Countries	Referrers	
12614	26244	239	4954 3075 1266 987 968 942 781 740 715 712	External:3179 Search Engine:485 Local:38 Repository:93 None:22688	[isolate]
12012	1453	68	1080 117 79 36 25 14 13 12 12 11	External:26 Search Engine:1285 Local:6 Repository:2 None:202	[isolate]
2623	1435	20	1195 106 54 18 13 7 4 4 4 3	External:4 Search Engine:1277 Local:6 Repository:5 None:163	[isolate]
2900	984	56	749 67 47 16 15 15 13 10 9 7	External:17 Search Engine:855 Local:4 Repository:5 None:159	[isolate]
756	514	58	194 90 25 19 18 17 12 12 11 10	External:12 Search Engine:470 Local:5 Repository:10 None:75	[isolate]
827	444	10	155 60 56 17 14 12 11 9 9 8	External:18 Search Engine:357 Local:0 Repository:1 None:78	[isolate]
10571	435	1299	1028 362 45 43 36 33 20 15 13 12	External:941 Search Engine:29 Local:43 Repository:355 None:366	[isolate]
10031	411	62	113 35 31 30 25 22 15 14 14 14	External:17 Search Engine:308 Local:0 Repository:11 None:137	[isolate]
12094	388	1538	1288 157 105 58 55 36 29 25 20 19	External:396 Search Engine:76 Local:21 Repository:261 None:1172	[isolate]
8705	371	20	194 48 20 19 11 11 10 9 7 4	External:120 Search Engine:32 Local:1 Repository:7 None:231	[isolate]
			Totals		
12614 12094 12012 10571 10031 8705 2900 2623 827 756	32679	3370	10180 4462 1302 1019 989 981 922 907 858 786	External:4730 Search Engine:5174 Local:124 Repository:750 None:25271	

Focus on Single Paper

- Consistently high levels of download over period of statistics
- *Some papers have a brief popularity*



[Download as CSV] **ECS Eprints Stats for 12614** [Raw Log Entries]

Date Range	Fulltext Downloads	Abstract Downloads	Top Ten Countries	Referrers	
2006-08-01 to 2006-08-07	0	0		External:0 Search Engine:0 Local:0 Repository:0 None:0	[expand]
2006-08-08 to 2006-08-14	129	2	37 13 6 5 5 4 4 4 3 3	External:16 Search Engine:0 Local:0 Repository:0 None:115	[expand]
2006-08-15 to 2006-08-21	1691	11	367 198 87 64 50 49 47 45 44 44	External:152 Search Engine:8 Local:2 Repository:7 None:1533	[expand]
2006-08-22 to 2006-08-28	1755	10	338 189 91 66 63 57 56 55 50 48	External:155 Search Engine:15 Local:3 Repository:3 None:1589	[expand]
2006-08-29 to 2006-09-04	1773	14	357 180 104 63 57 54 53 46 44 44	External:189 Search Engine:19 Local:2 Repository:3 None:1574	[expand]
2006-09-05 to 2006-09-11	1778	11	340 202 102 65 57 56 56 52 51 46	External:179 Search Engine:26 Local:1 Repository:4 None:1579	[expand]
2006-09-12 to 2006-09-18	1728	8	334 189 99 71 59 59 50 49 48 44	External:194 Search Engine:20 Local:0 Repository:3 None:1519	[expand]
2006-09-19 to 2006-09-25	1751	10	328 201 79 58 57 56 53 53 46 46	External:199 Search Engine:32 Local:1 Repository:3 None:1526	[expand]
2006-09-26 to 2006-10-02	1790	16	344 208 83 67 57 55 54 51 51 48	External:237 Search Engine:34 Local:4 Repository:6 None:1525	[expand]
2006-10-03 to 2006-10-09	1923	29	342 263 80 74 73 63 61 57 53 52	External:262 Search Engine:36 Local:4 Repository:12 None:1638	[expand]
2006-10-10 to 2006-10-16	2145	48	418 295 78 73 73 68 67 62 60 57	External:311 Search Engine:55 Local:4 Repository:20 None:1803	[expand]
2006-10-17 to 2006-10-23	1689	10	308 212 77 72 71 66 52 47 46 46	External:215 Search Engine:40 Local:1 Repository:4 None:1439	[expand]
2006-10-24 to	1581	15	280 188 73 68 65 62 60 52 52	External:201 Search Engine:21 Local:4 Repository:4	[expand]

Unusually the downloads come mainly from external links, not search engines.

Zoom in on Downloads

2006-08-31 17:32:09	12614	F	Colt telecom	213.229.146.15		None			
2006-08-31 17:43:35	12614	F	Covad Communications	h-72-244-103-25.lsanca54.covad.net		None			
2006-08-31 17:49:21	12614	A	Cyprus International University, Lefkosa, KKTC	marge.levent.com		External			http://www.w3.org/2001/sw/
2006-08-31 17:49:21	12614	F	Cyprus International University, Lefkosa, KKTC	marge.levent.com		External			http://www.w3.org/2001/sw/
2006-08-31 17:49:59	12614	F	Connected by GTS Poland	217.153.131.234		None			
2006-08-31 17:50:19	12614	F	EgyNet	84.36.84.162		None			
2006-08-31 17:59:53	12614	F	o2 Germany GmbH & Co. OHG	pat-97.sg.de.o2.com		External			http://www.w3.org/2001/sw/
2006-08-31 18:01:48	12614	F	CHINANET fujian province network	218.5.5.48		None			
2006-08-31 18:11:04	12614	F	Illinois Century Network	st-209-7-87-12.gesd.n-cook.k12.il.us		None			
2006-08-31 18:16:44	12614	F	Matsuyama Canon OA System Corporation	giant.cr-office.co.jp		None			

- Each access shown by
  - Requester DNS, Company Name, Country
  - Referer Type, Search Engine, Query Terms, Link
- This paper's full text is linked directly from W3C
- Only 4-25% of requesters can be identified as researchers

Every Paper Has a Story

- A different community accesses it through different methods
  - Some papers have abstracts downloaded more than full texts
  - Others vice versa
  - Some papers are exclusively downloaded from Google Scholar

Concluding Remarks

- Stats are incredibly important, but under utilised
- International agreement on baselines are key
- Need to deliver flexible, user-oriented stats
  - Also need to deliver other stats profiles for COUNTER and funding bodies