Statistics of Open Access Journals

Franco Toni – Biblioteca dell'Istituto Superiore di Sanità

Conference Institutional Archives for Research: Experiences and Projects in Open Access

> Istituto Superiore di Sanità Rome, 30 November – 1 December 2006

General frame

- Until few years ago e-journals statistical data were collected in an heterogeneous way and following customized criteria
- Results were rough, sectorial, not comparable or addable
- A great change in the last years due to several factors:
 - Exponential growth of references
 - Librarians need new tools in order to better manage use of resources
 - Great relevance of statistics in measuring and evaluating libraries performances
- Use of standards allows to obtain uniform and reliable results

Standards

- ISO Standards concerning statistics and performance indicators:
 - ISO 2789 International Library Statistics (4th edition, September 2006)
 - Principal goal is "to ensure conformity for those statistical measures that are frequently used by libraries"
 - Divided in 6 parts
 - Part 3 "terms and definitions" contributes to exactly identify and define terminology and definitions for each used item
 - Annex A "Measuring the use of electronic library services"

Standards

- ISO 11620 Library Performance Indicators (1998)
 - Now under revision to incorporate the content of TR 20983 (2003) with 15 new indicators on electronic library services
 - Indicators and data sets analyze not only quantity ma also quality of provided services
 - The list of indicators is "best seen as a menu of possible performance indicators that could be used in a range of library settings"
 - Each library "will need to decide which indicators are most appropriate to its particular situation"

The COUNTER Project

- **COUNTER = Counting Online Usage of Networked Electronic Resources** (launched and become operating in 2002)
- A successful example of cooperation between producers and users of information and documentation
- The aim is to achieve clarity, uniformity and simplicity in collecting and showing data
- First release of the Code of Practice:
 - Two reports for e-journals
 - Journal Report 1 Number of successful full-text article request by month and journals
 - Journal Report 2 Turnaways by month and journal
 - Three reports for the databases
 - No reports for other supports. E-books added only in a specific Code published last Spring

COUNTER evolution

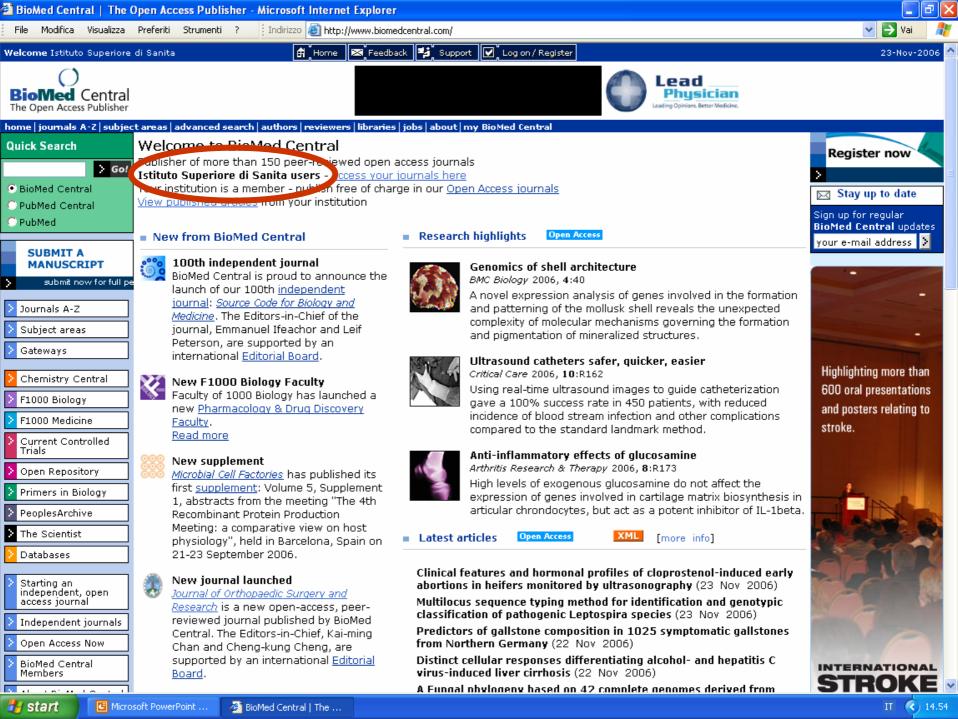
- Rapidly become a standard de facto
- More than 50 publisher are now (end of October 2006) COUNTER compliant.
- All the most important such as Elsevier, Blackwell, Wiley, Springer, NPG, etc. are included
- They guarantee to produce at least one of the COUNTER report (47 produce JR1)
- At the beginning of 2006 was published the release 2 of the Code of Practice with new important features such as the possibility to download data in excel or excel compliant (CSV) format

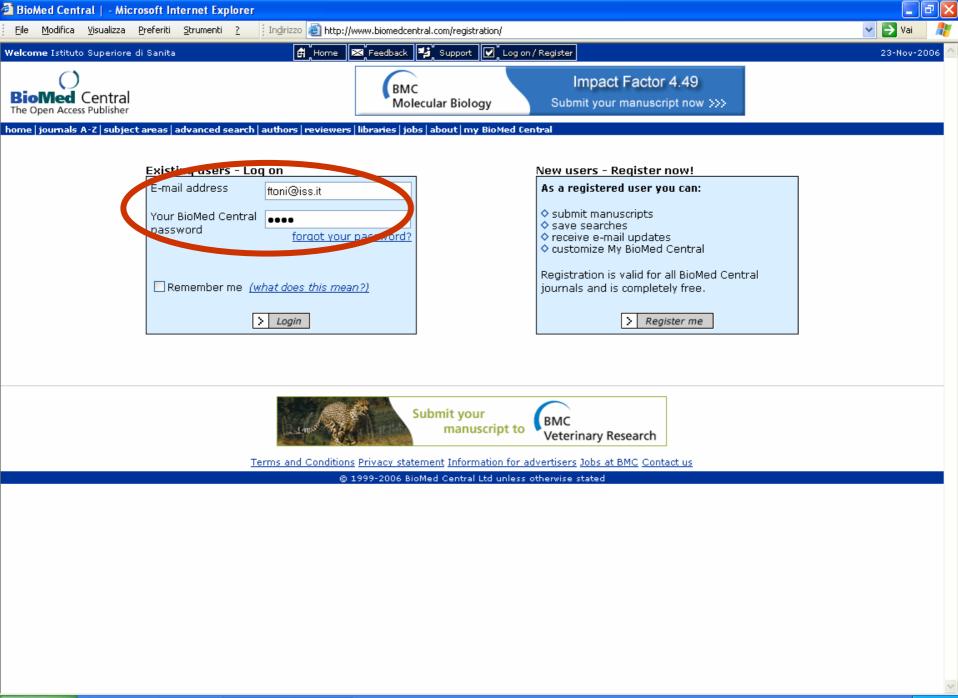
Standards

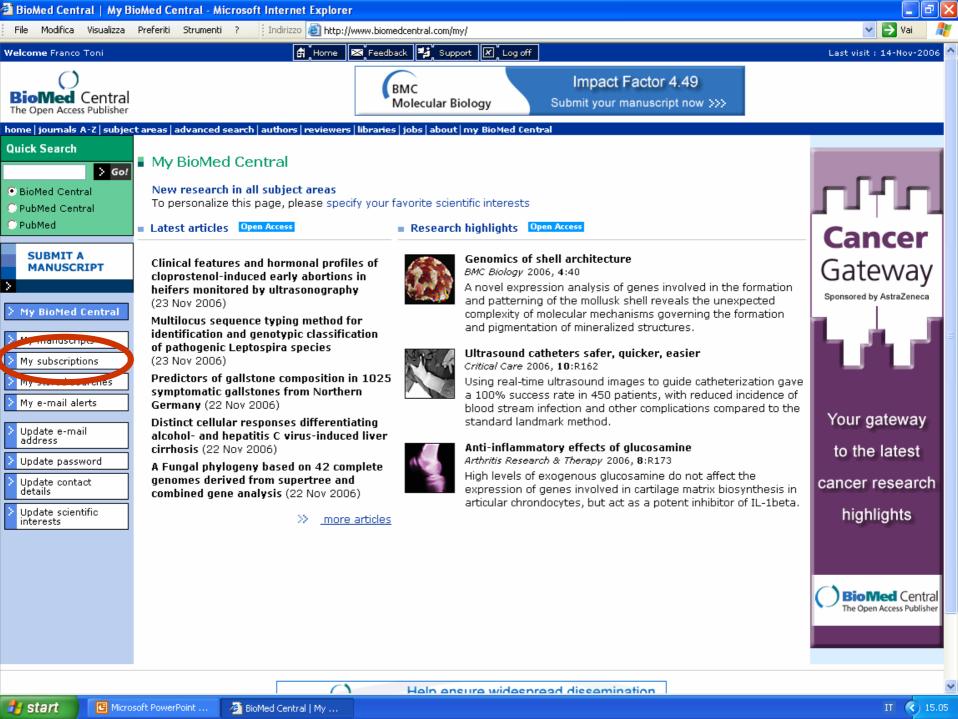
- NISO initiative called SUSHI (Standardized Usage Statistics Harvesting Initiative) whose aim is:
 - "to develope a standard data container for moving Project COUNTER usage statistics into a digital repository. Currently COUNTER reports are received in Excel spreadsheets with data from each individual service provider in separate files. The short-term solution will be the design of an automated request and response protocol based on a web services model for downloading XML versions of the data into an electronic repository"

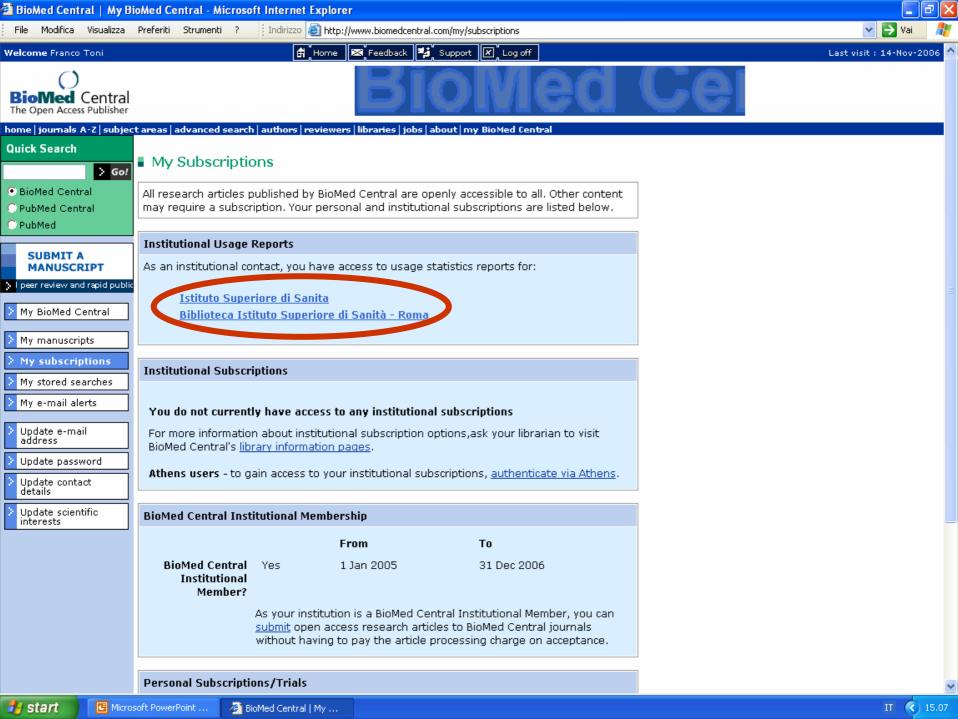
Open Access Journals and Statistics

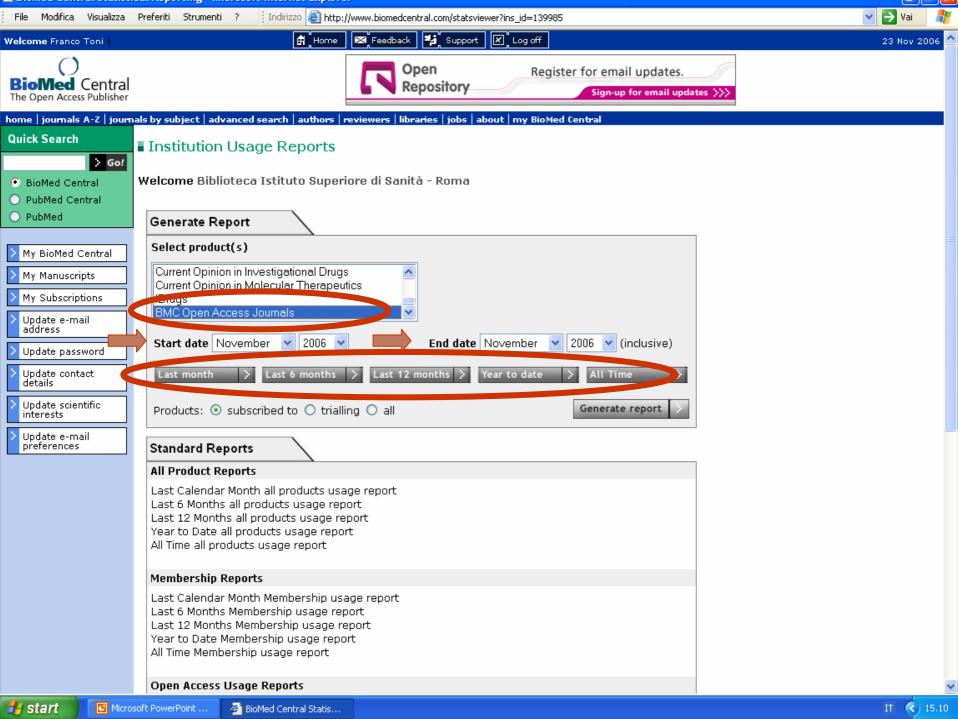
- An OA resource is hardly compatible with a gathering of statistical data
- Producers and distributors of OA contents haven't the need to recognize the complete track of accesses on their servers
- Among the 4 main access sources to OA e-journals DOAJ, PubMed Central, Plos and BioMed Central – only this last provides users with data about accesses and downloads

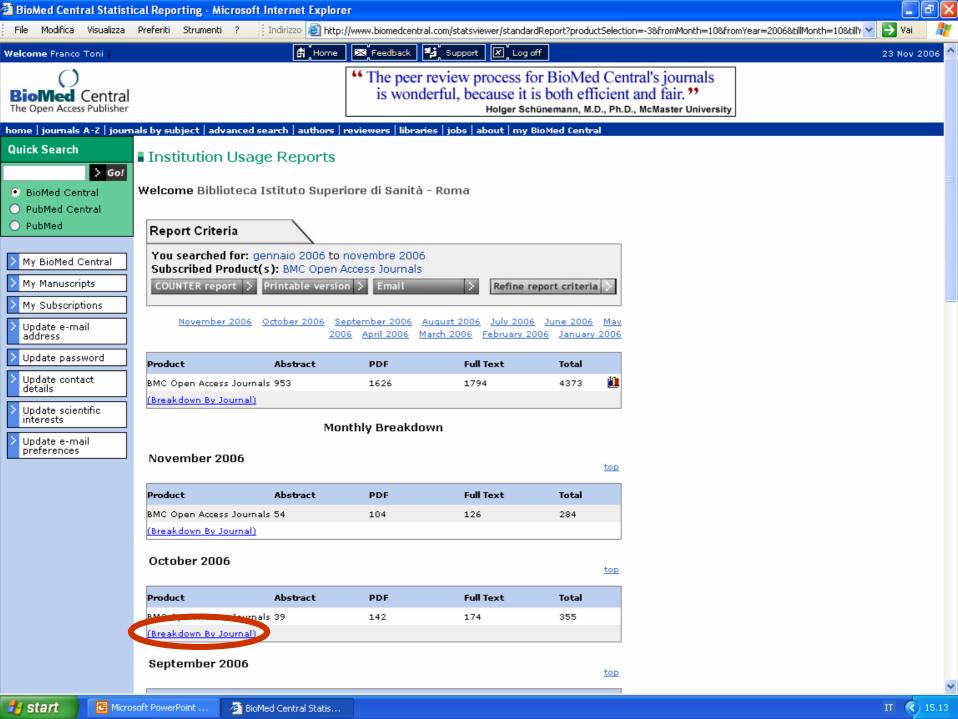










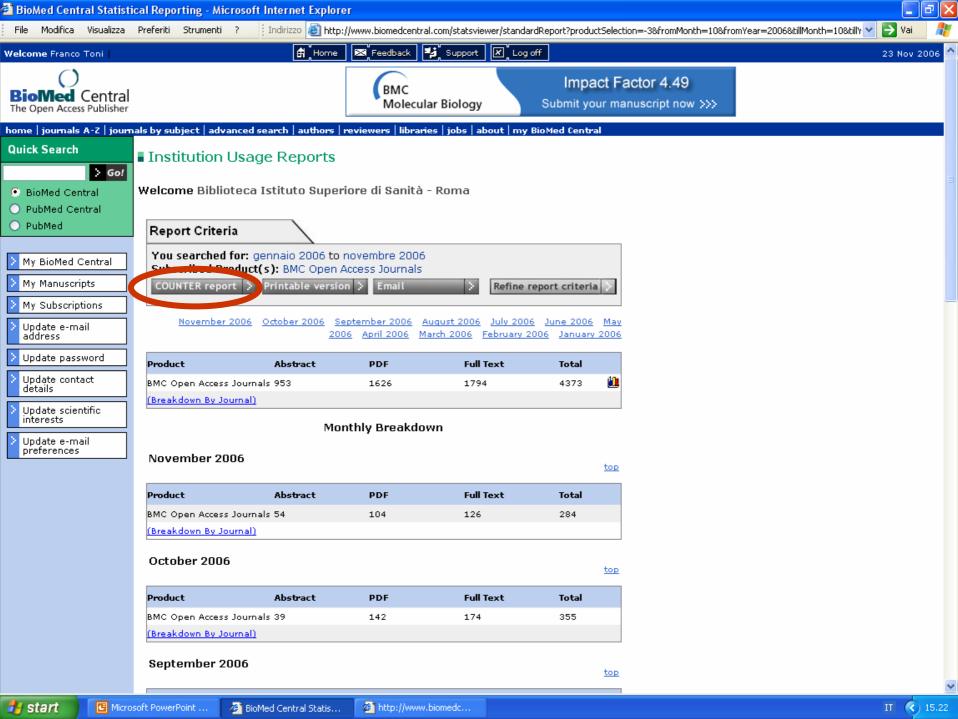


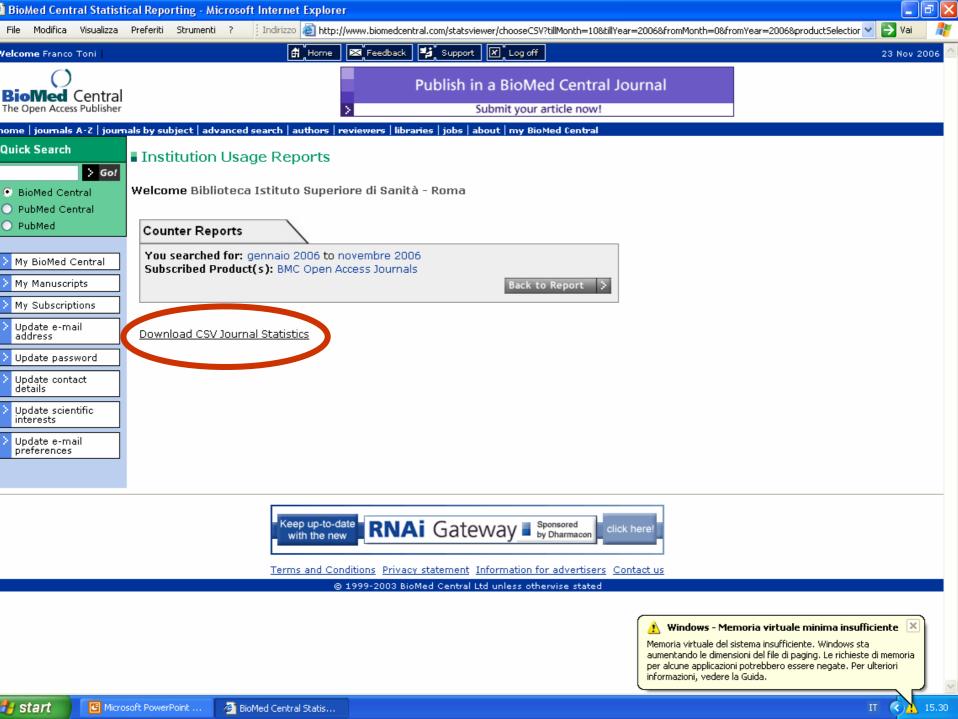
(Breakdown By Journal)

top

October 2006

Product	Abstract	PDF	Full Text	Total
AIDS Research and	2	2	0	4
Therapy	2	2	0	4
Annals of General Psychiatry	0	0	1	1
BMC Biochemistry	0	0	1	1
BMC Bioinformatics	3	12	16	31
BMC Biology	2	3	0	5
BMC Biotechnology	0	0	1	1
BMC Cancer	1	5	7	13
BMC Cell Biology	0	1	1	2
BMC Clinical Pharmacology	0	0	1	1
BMC Complementary and Alternative Medicine	0	1	2	3
BMC Evolutionary Biology	4	6	2	12
BMC Gastroenterology	0	1	2	3
BMC Genomics	0	2	8	10
BMC Health Services Research	2	2	0	4
BMC Immunology	0	1	2	3
BMC Infectious Diseases	5	26	18	49
BMC Microbiology	1	5	3	9
BMC Neurology	0	0	1	1
BMC Neuroscience	0	0	1	1
BMC Pediatrics	0	1	1	2
BMC Public Health	1	8	14	23
BMC Veterinary Research	3	7	4	14
Behavioral and Brain Functions	0	1	3	4
Biology Direct	0	2	0	2
Cancer Cell International	0	5	7	12
Cardiovascular Diabetology	0	1	1	2
Cerebrospinal Fluid Research	0	0	1	1
Chiropractic & Osteopathy	0	1	3	4
Clinical Practice and Epidemiology in Mental Health	0	2	5	7
Cost Effectiveness and	n	n	3	3



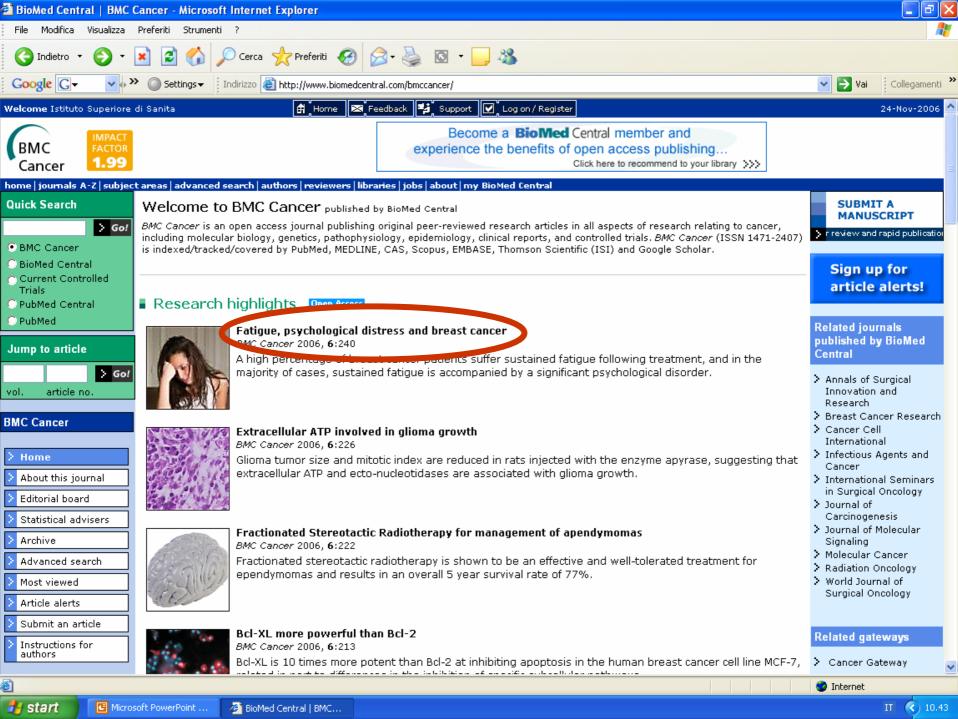


BMC in the Istituto Superiore di Sanità

2005 1500 downloads

2006 4200 downloads (10 months)

- Negative influence of BMC interface on statistical data
- Ph. M. Davis J.S. Price, eJournal interface can influence usage statistics: implications for libraries, publishers, and Project COUNTER, Journal of American Society for information science and technology, 57 (9), 1243-1248, 2006





Open Access Journals and Statistics

- 2 possible ways to obtain statistics about references or access on a web site:
 - Starting from the log files of the user's server
 - Using registrations on the receiver server
- Case 1 scant useful: we can monitor only the output address and if the receiver is an aggregator or a distributor we cannot recognize the final destination
- Case 2: only BMC performs this task

The IP address

- The only way to reconstruct for a certainty the user's pathway is through the recognition of IP address by the receiver server
- The user identification through IP is firstly a need of commercial publishers, and not of OA providers
- OA providers main goal is to measure how many times a specific resource is accessed and not who is consulting it
- Distributors (i.e. Plos and PMC) are be able in computing data in order to define indicators such as impact factor or the average of accesses on a title but not about quantity and type of researches and downloads institution by institution.
- DOAJ has recently introduced the possibility to search at article level on 1/3 of its 2400 journals. The DOAJ server logs the track of user's pathway and these records could be used also by the institutions that generate the requests in order to obtain statistical data

Some final considerations

- Measurement and evaluation of library performances and services for users are become a basic item in order to define strategies and policy of acquisitions.
- Statistics hold an increasingly value allowing a prompt check of how and how much a resource is used.
- The lack of statistical data about OA e-journals deprives librarians of an essential tool for the evaluation of all products available on the net
- Gunther Eisenbach (G. Eisenbach, *Citation advantage of open access article*, PLoS Biology, 4,5, May 2006, e157) demonstrates that OA articles are more cited and read compared with non open published on the same journal (PNAS in this case)
- It should be very important to verify the effective use of OA journals compared with the commercials in the same field or discipline
- If the results of this analysis were in favour of OA titles, this would become a relevant factor for promoting Open Access initiative and for a deep reduction of library expenses for journals

Thank you for your attention

ftoni@iss.it