

Lies, damn lies, and usage statistics?¹

Peter van Laarhoven p.j.b.m.van.laarhoven@rug.nl
Ismail Fahmi i.fahmi@rug.nl

Gebruiksgegevens van elektronische tijdschriften

Net als alle surfgedrag op het web kan ook het gebruik van elektronische tijdschriften in detail worden geregistreerd door de aanbieders van de informatie. Dat levert de laatste jaren een stortvloed aan gegevens op -de zogenaamde usage statistics- die maandelijks aan de UB wordt toegestuurd. Dat is nuttige en gewenste managementinformatie, maar het is veel werk om die te verzamelen en de interpretatie ervan is minder simpel dan het wel eens lijkt.

De elektronische tijdschriften die de RUG in licentie heeft, staan opgeslagen op de webserver van de uitgever of van de 'platforms' aan wie de uitgever de opslag uitbesteed hebben. De *server logs* daarvan bevatten gedetailleerde gegevens over het gebruik, zoals welk document op welk tijdstip vanaf welk werkstation is opgevraagd en van welke webpagina men binnenkwam. Met de alomtegenwoordige *cookies* kunnen bovendien complete sessies gereconstrueerd worden.

Al enkele jaren geleden hebben de bibliotheken zich de mogelijkheden van deze gegevens ge-

realiseerd en als klanten van de uitgever eisen geformuleerd aan de omgang ermee. Die eisen hadden zowel betrekking op de privacybescherming van de gebruikers als de levering van gegevens aan de bibliotheek. De bibliotheken verwachtten met de gebruiksgegevens hun uitgaven en de verdeling ervan -over disciplines en faculteiten- beter te kunnen verantwoorden.

Veel uitgever waren aanvankelijk nogal terughoudend met het leveren van het gevraagde materiaal. Zij vreesden dat er al te snel conclusies getrokken zouden kunnen worden uit gegevens die weinig gestandaardiseerd waren en een grotere hardheid suggereerden dan waarschijnlijk gerechtvaardigd was. De bibliotheken deden die twijfels overigens.

Het gegeven waarop de aandacht zich van meet af aan geconcentreerd heeft, is het 'gebruik' van de full-text van artikelen. Andere gegevens -zoals het aantal zoekacties, de raadpleging van inhoudsopgaven, attenderingsprofielen- lijken hooguit interessant voor diepgaander onderzoek

of voor evaluatie van de geboden interfaces. Ten aanzien van dat full-textgebruik bestond aanvankelijk grote zorg over de kwaliteit van de gegevens. Hoe wordt precies geteld? Wat als een gebruiker een paar muiskliks achter elkaar geeft? Wat als men eerst de html-versie van een artikel oproept, om het te screenen, om vervolgens de pdf-versie af te drukken voor het echte leeswerk? Wat als de browser een lang pdf-document in stukken ophaalt? Wat als een tijdschrift voor de instelling online beschikbaar is op meer dan één platform? Ondanks deze bezwaren begonnen steeds meer uitgever gebruiksstatistieken te leveren, elk op hun eigen manier.

Project COUNTER

Tegen deze achtergrond ging in 2002 COUNTER² van start, een kleine non-profitorganisatie waarin uitgever, bibliotheken en intermediairs samenwerken aan standaarden voor gebruikstatistieken van elektronische informatiebronnen. In een jaar of drie tijd heeft COUNTER grote vooruitgang geboekt. Het voornaam-



ste resultaat is de voorbereiding en publicatie van diverse *codes of practice*, die gedetailleerd voorschrijven hoe gebruik geteld moet worden en in welke vorm de gegevens aan de klant beschikbaar moeten worden gesteld. Behalve een *Code of Practice for Journals and Databases* -die in april 2005 zijn tweede release heeft gekend- is een *Code of Practice for Books and Reference Works* in voorbereiding.

De COUNTER-standaard bevat op dit moment:

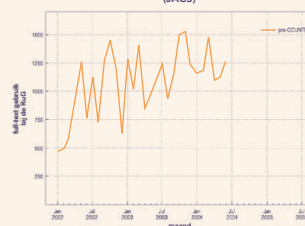
- een kleine set van gestandaardiseerde tellingen en tabellen;
- een protocol voor onafhankelijke controle op de tellingen van uitgevers, waarmee uitgevers een soort accountantsverklaring van *COUNTER compliance* kunnen verwerven;
- een XML-DTD voor gebruiksrapporten, waarmee op den duur de bewerking van de gegevens gemakkelijker kan worden.

Steeds meer uitgevers schakelen op dit moment over op gebruiksstatistieken volgens de COUNTER-standaard, zij het nog lang niet allemaal. De UB verzamelt dit materiaal voor de RUG, ordent het en stelt het beschikbaar voor belanghebbenden³.

De JACS als voorbeeld

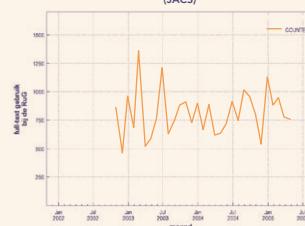
Om het belang van een standaard als COUNTER te illustreren, kunnen we het RUG-gebruik aanhalen van de tijdschriften van de American Chemical Society (ACS), waarop we sinds begin 2002 een licentie hebben. Tot medio 2004 leverde de ACS eigen gebruiksgegevens, daarna schakelde men over op *COUNTER compliant*-gegevens; voor een flinke periode zijn beide tellingen beschikbaar.

figuur 1a Journal of the American Chemical Society (JACS)



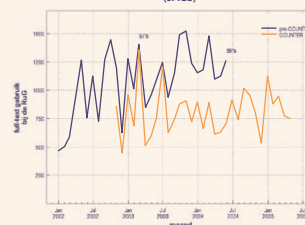
Figuur 1a geeft het RUG-gebruik weer van de *Journal of the American Chemical Society*, het vlaggenschip van de ACS, volgens de 'oude' telling. Na de snelle acceptatie van de online JACS in 2002 lijkt het gebruik gekenmerkt door een onmiskenbaar stijgende trend naar ca. 1.250 full-textraadplegingen per maand - afgezien van de 'seizoenseffecten' die we bij alle tijdschriften zien.

figuur 1b Journal of the American Chemical Society (JACS)



Figuur 1b geeft het gebruik van de JACS weer volgens de COUNTER-telling. Een trend is daarin niet meer waarneembaar en het gebruik komt gemiddeld op ca. 750 raadplegingen per maand, waarmee de JACS overigens een van onze meest gebruikte tijdschriften blijft.

figuur 1c Journal of the American Chemical Society (JACS)



In figuur 1c zijn beide tellingen naast elkaar gepresenteerd, waaruit niet alleen is af te lezen dat het gebruik volgens COUNTER flink lager uitkomt, maar vooral dat er geen vaste relatie tussen beide tellingen is; de COUNTER-telling

belooft tussen de 56% en 97% van de oude ACS-telling.

Zijn we er met COUNTER?

COUNTER heeft de eigenlijke tellingen gestandaardiseerd en de medewerking van uitgevers gestimuleerd. Maar daarmee zijn we er niet. In minstens twee opzichten blijven we, ook als alle uitgevers COUNTER-gegevens zouden leveren, met diverse management- en onderzoeksvragen zitten.

- COUNTER beperkt zich vooralsnog bewust tot een minimumstandaard, met tabellen die met opzet sober en globaal zijn gehouden en waarmee niet is na te gaan:

- welke gebruikersgroepen welke tijdschriften en artikelen gebruiken;
- hoeveel verschillende artikelen geraadpleegd worden;
- uit welke jaargangen de geraadpleegde artikelen afkomstig zijn.

- Ook COUNTER kan het bezwaar niet wegnemen dat wel *lack of context* is genoemd: gebruiksstatistieken registreren slechts 'aan de oppervlakte' muiskliks waarmee gebruikers door een online-tijdschrift navigeren en documenten oproepen. Daarmee is niet na te gaan:

- wat het gebruiksdoel is: blijven, diepgaande studie, citatie, onderwijs, etc.;
- wat de waarde van een artikel voor de gebruiker is, en daarmee ook
- wat de waarde van een tijdschriftlicentie voor de instelling is.

Meer gegevens en resultaten

Bijna alle uitgevers behandelen de minimumstandaard die COUNTER is feitelijk tevens als maximum-

Faculty	full-text usage		attributable to		
	2003		print subscription	2000	'big deal' extra
			own subscription	X-access	
Arts & Humanities	1%	2691	37%	45%	19%
Economics, Business & Geography	3%	6500	41%	30%	30%
Psychology, Sociology & Education	5%	11537	55%	18%	27%
Medicine	40%	95980	47%	9%	44%
Science	36%	85532	43%	12%	45%
Other (faculties, facilities & remote)	15%	34892	45%	12%	43%
overall		237132	45%	12%	43%

Tabel 1. De 'gebruikswinst' van de Elsevier-licentie in 2003

standaard. Soms laten zij daarvoor nuttige gegevens vallen die ze tevoren wel verstrekten. Zo leverde OVID, het platform waarop we de medische tijdschriften van Lippincott, Williams & Wilkins in licentie hebben, tot eind 2004 gedetailleerde sessiegegevens over de full-textraadpleging per artikel, maar schakelde met ingang van 2005 en zonder overlap over op de standaard COUNTER-tabellen.

Voor meer gegevens zijn we daarom aangewezen op samenwerking met uitgevers. Zo was Elsevier bereid ons gedetailleerde gegevens te leveren over het gebruik per artikel sinds begin 2003. Eerste resultaten daarvan hebben we onlangs gepresenteerd op de LIBER-conferentie in Groningen⁴.

rank	usage 2004	U of Groningen authored	U of Groningen cited	journal ISSN	title	article publication date
1	358	-	X	0361-3682	Accounting, Organizations and Society	March 2003
2	317	-	-	0361-3682	Accounting, Organizations and Society	February 2002
3	259	-	-	1044-5005	Management Accounting Research	December 1999
4	222	X	-	1044-5005	Management Accounting Research	June 2000
5	183	X	-	1044-5005	Management Accounting Research	September 2001
6	114	X	-	1044-5005	Management Accounting Research	December 2003
7	105	X	X	0021-9150	Atherosclerosis	January 2004
8	93	X	-	0168-0102	Neuroscience Research	June 2004
9	85	-	X	0167-9236	Decision Support Systems	May 2001
10	81	X	-	1567-1356	FEMS Yeast Research	May 2004

Enkele van onze bevindingen:

- De *big deal* met Elsevier heeft in alle geledingen van de RUG geleid tot veel full-textgebruik van tijdschriften waar vóór 2001 geen abonnement op bestond (43% van het totale full-textgebruik in 2003) of waarvoor men naar een andere bibliotheeklo-

catie van de universiteit moest (12%; zie tabel 1).

- Veruit de meeste Elsevier-artikelen (52%) worden in de loop van een jaar maar één keer geraadpleegd, niet meer dan 1% van de artikelen wordt meer dan 10 keer gebruikt; het gemiddelde lag voor 2004 op 2,26 raadplegingen van ruim 150.000 verschillende artikelen. Voor de tijdschriften van Lippincott, Williams en Wilkins (LWW) speelt dit nog sterker: 71% van de artikelen is één keer gebruikt, bijna geen artikelen meer dan vijf keer (zie tabel 2). De lage toegangsdrempels tot een compleet fonds lijken daarmee te leiden tot laagdrempelig gebruik van talloze verschillende artikelen.
- Van de 10 meestgebruikte Elsevier-artikelen uit 2004 hebben er 7 een RUG-auteur of medeauteur, en gaat het dus om 'eigen werk' (zie tabel 3), bij LWW zijn dit er 6 van de 10.

Tabel 2. Top-10 van Elsevier-artikelen in 2004

Tabel 3. Gebruiksfrequentie per artikel in 2004

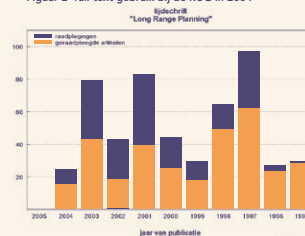
# of uses in 2004	Elsevier		LWW	
	abs.	%	abs.	%
1 x	78014	52%	7871	71%
2 x	37835	25%	1919	17%
3 x	12809	9%	690	6%
4 x	8079	5%	311	3%
5 x	4139	3%	273	2%
6 x	2823	2%	28	0%
7 x	1719	1%		
8 x	1269	1%		
9 x	830	1%		
10 x	615	0%		
11 x - 20 x	1934	1%		
> 20 times	477	0%		
# of articles used	150543	100%	11092	100%
total fulltext usage	340723		17064	





- Van de 10 meestgebruikte Elsevier-artikelen uit 2004 zijn er 7 afkomstig uit een drietal Bedrijfskundetijdschriften. Qua ouderdom lopen deze artikelen uiteen van december 1999 tot december 2003. Waarschijnlijk zijn deze artikelen gebruikt in het onderwijs, en is het full-textgebruik tot stand gekomen doordat de docent in Nestor links naar de artikelen heeft opgenomen (overigens de aanbevolen route om artikelen te vrijwaren van de afdracht van reproductie).
- Niet meer dan eenderde van de gebruikte Elsevier-artikelen is gebruikt in hetzelfde jaar als waarin ze gepubliceerd zijn. Je zou kunnen betogen dat de oudere artikelen al betaald en 'verworven' zijn in eerdere jaren en dat de kosten-per-gebruik daarom veel hoger zijn dan het lijkt als je niet kijkt naar de publicatiedatum van de artikelen.

Figuur 2 full-text gebruik bij de RUG in 2004



- Oudere artikelen blijken vaker gebruikt te worden dan doorgaans verwacht wordt. Van de geraadpleegde Elsevier-artikelen is gemiddeld zo'n 15% op het moment van gebruik tussen 5 en 10 jaar oud. Er zijn op dat punt aanzienlijke verschillen tussen disciplines (Geneeskunde meer 'jonge' artikelen, Bedrijfskunde meer oudere) en tussen tijdschriften. Een boeiend maar extreem geval is het tijdschrift *Long Range Planning*, waarvan het gebruik nagenoeg gelijkmatig over een periode van 10 jaar is gespreid (zie *figuur 2*). Kennelijk kijken planners regelmatig achterom. Let wel dat we van Elsevier tot nu toe alleen de artikelen van 1995 en later beschikbaar hadden. Met de onlangs afgesloten nieuwe licentie krijgen we toegang tot alle Elsevier-artikelen uit alle jaren, waardoor het gebruiksaandeel van oudere artikelen nog zal toenemen.
- Ook uit andere bron weten we dat oudere online beschikbare artikelen verrassend vaak gebruikt worden. Onze licen-

tie op JSTOR Arts & Sciences I bijvoorbeeld, een pakket historische jaargangen van 100 toptijdschriften, genereerde in 2004 bijna 31.000 full-textraadplegingen; topper daarbij is de *American Economic Review* met 2.064 raadplegingen. Of het archief van de *American Chemical Society*, dat 30% van het full-textgebruik aan zich trekt.

Voorzichtigheid geboden

Onze nadere analyses laten zien dat achter de schijnbaar eenvoudige COUNTER-tellingen nogal gevarieerde gebruikspatronen schuil kunnen gaan. Die nuances de veronderstelde 'hardheid' van de gebruiksgegevens, zonder dat ze al veel zicht geven op de 'echte' gebruiksdoelen en gebruikswaarde voor de RUG. We zullen ons daar, naast de periodieke verzameling en ordening van de COUNTER-gegevens, zeker mee bezig blijven houden.



noten

- 1 Een variant op de uitspraak van Mark Twain (die hem weer leende van de Britse premier Disraeli): "There are three kinds of lies: lies, damned lies, and statistics".
- 2 COUNTER is een acroniem voor Counting Online Usage of Networked Electronic Resources en heeft zijn homepage op www.projectcounter.org/
- 3 De UB-dienst voor gebruikstatistieken heeft de voorlopige naam SGAVOT gekregen, Statistische GebruiksAnalyse van Online Tijdschriften. De homepage is <http://sgavot.ub.rug.nl/>. Deze is alleen binnen de RUG raadpleegbaar.
- 4 Zie pagina 15 voor meer informatie over de LIBER conferentie. Onze LIBER-presentatie is te vinden op <http://liber.ub.rug.nl/presentations/Laarhoven.ppt>.