

## La Rete in Italia

### Marzo 2003: un nuovo sistema per la misurazione delle audience sul Web

#### Dati, studi e commenti sull'utilizzo della Rete in Italia

La Rete in Italia raggiunge in Marzo 10,5 milioni di utenti che accedono da casa o dal luogo di lavoro. Il resto del contributo al numero totale di utenti (prossimo agli 11 milioni) viene da scuola e università

Nel mese di Marzo 2003 gli utenti Internet italiani sono stati poco più di 10,5 milioni. Onetone Research intende per *utente Internet* un individuo con più di 14 anni di età che accede alla Rete autonomamente e abitualmente almeno una volta nel mese considerato.

La situazione del numero di utenti Internet italiani a Marzo 2003, in relazione al luogo di accesso, è la seguente:

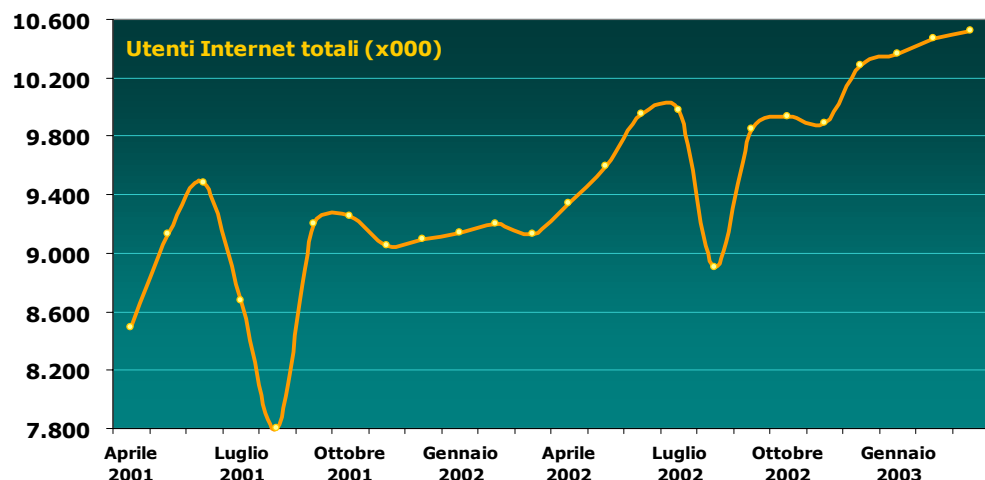
Utenti Res + Biz	Utenti Residenziali	Utenti Business	Utenti sia Res che Biz
10,52	8,74	4,36	2,58

Gli utenti residenziali coprono l'intero insieme degli utenti in misura pari a circa l'80%

Da qui si ricava che l'83% degli utenti Internet dispone di una connessione da casa, il 41% dall'ufficio.

Gli utenti complessivi possono essere qualche centinaio di migliaia di più rispetto al numero presentato. Si deve infatti includere l'insieme di coloro che accedono alla Rete da scuola, università, luoghi pubblici. Nonostante questo insieme sia piuttosto numeroso e composto da circa 1,5 milioni di individui, esso tuttavia è molto sovrapposto ai due citati in precedenza (casa e lavoro) e il suo contributo all'incremento della numerosità totale degli utenti italiani risulta marginale.

L'andamento del numero di utenti residenziali e lavoro negli ultimi 24 mesi. La tendenza alla crescita è evidente anche se limitata a circa il 9% annuo

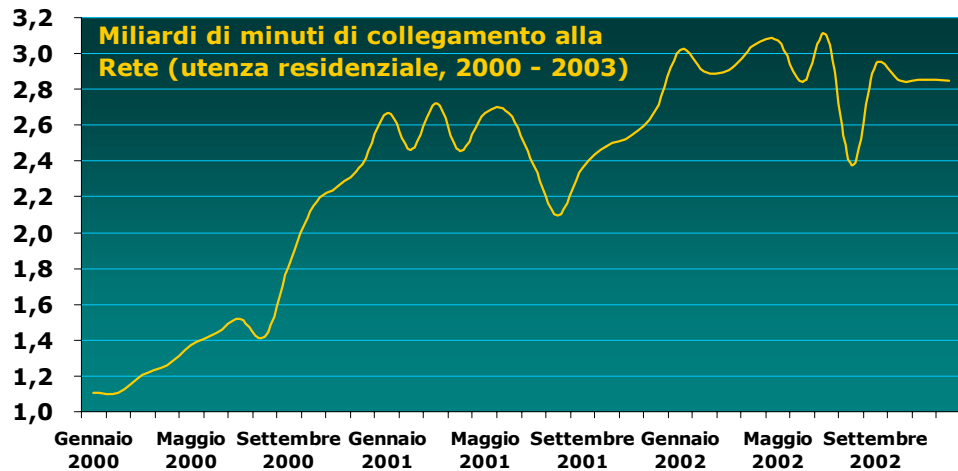


Andamento degli utenti Internet mensili in Italia (luogo di accesso: casa e lavoro) in migliaia. Fonte: Onetone Research, 2000-2003

Il tempo speso dagli utenti Internet residenziali in Italia negli ultimi 36 mesi ha seguito un andamento simile alla crescita del numero di utenti, con una forte impennata nel 2000, un graduale rilassamento della crescita nel corso del 2001 e sostanziale stabilità per tutto il 2002, come mostrato nella figura che segue:

per tutto il 2002, come mostrato nella figura che segue:

Il consumo di Internet cresce in modo significativo tra Novembre 2001 e Gennaio 2002. In particolare, il consumo di banda e il tempo speso online subiscono una vera e propria impennata



Andamento negli ultimi due anni del tempo speso online dall'utenza residenziale italiana (miliardi di minuti mensili)  
Fonte: Onetone Research, 2000-2003

Il consumo residenziale di Internet (in minuti di connettività) sviluppa un giro d'affari di dimensioni significative e ben superiori a quelle della pubblicità online

La media mensile di consumo nel 2003 è pari a circa 3 miliardi di minuti di consumo (si ricordi che questo è il contributo della sola utenza residenziale). Di qui si giunge alla stima del giro d'affari della connettività dial-up residenziale per il 2002, che in Italia stimiamo sia stato pari a 475 milioni di Euro. La cifra è circa otto volte il fatturato pubblicitario online nel 2002. L'85% circa di questa cifra è realizzata dai tre grandi ISP dial-up italiani residenziali (Wind, Seat-Tin.it e Tiscali).

Il mese di Marzo sulla Rete è stato caratterizzato dal forte balzo dei consumi avvenuto in occasione dello scoppio della guerra in Iraq. Alcuni siti di informazione hanno in certi casi decuplicato il volume di traffico e moltiplicato per diverse volte il numero di utenti e visite giornaliere.

Onetone Research e Red Sheriff sono stati in grado di seguire l'andamento dei consumi giornalieri grazie al nuovo sistema di misurazione costruito negli ultimi mesi: il "Cookie Panel".

I dati di alcuni siti Web misurati con il sistema censuario di Red Sheriff nel mese di Marzo. Il mese è stato particolarmente interessante per l'attenzione che si è riversata sul Web in seguito agli avvenimenti internazionali

Marzo 2003	Browser unici	Pagine viste	Visite	Visite / giorno	Totale Minuti
MonRif.Net	497.236	8.187.041	1.086.298	2,18	9.921.522
ISole24Ore.com	728.195	16.423.206	2.661.478	3,65	47.019.445
Repubblica.it	2.347.239	113.088.073	15.206.509	6,48	209.089.499
TGCom.it	980.445	40.777.002	4.303.221	4,39	42.530.168
Virgilio.it	12.295.334	583.085.550	63.064.049	5,13	519.227.337

Rilevazione del mese di Marzo 2003 (fonte: Red Sheriff).  
I dati si riferiscono ai volumi di consumo dell'intero mese

I dati presentati nelle tabelle sono ottenuti con il metodo censuario, cioè "contando" le richieste di contenuti e servizi sui siti Web effettuate dai browser. Le richieste sono registrate e contate da Red Sheriff, una ad una.

Il sistema di misurazione censuario di Red Sheriff (<http://www.redsheriff.it>) è utilizzato da molti grandi editori e portali, in Italia e in altri dieci paesi nel mondo.

Il metodo di misurazione garantisce la perfetta rappresentazione dei consumi perché questi sono *contati* e non *stimati* con metodi statistici. Per questo motivo, il metodo viene definito *censuario*.

Il nuovo sistema del Cookie Panel combina l'approccio censuario e quello campionario e conserva i vantaggi di entrambe le metodologie

Il punto di partenza dell'analisi del traffico del Web con il sistema del Cookie Panel è proprio questo: dal momento che è possibile ottenere il dato preciso del volume di consumo su un sito Web con il sistema di Red Sheriff, questo dato deve essere considerato più attendibile di quello ottenuto con un panel meterizzato.

Marzo 2003	Browser unici	Pagine viste	Visite	Visite / giorno	Totale Minuti
MonRif.Net	30.281	264.098	35.042	1,16	320.050
IlSole24Ore.com	57.973	529.781	85.854	1,48	1.516.754
Repubblica.it	252.218	3.648.002	490.533	1,94	6.744.829
TGCom.it	82.441	1.315.387	138.814	1,68	1.371.945
Virgilio.it	1.278.726	18.809.211	2.034.324	1,59	16.749.268

Rilevazione del mese di Marzo 2003 (fonte: Red Sheriff).  
I dati si riferiscono ai volumi di consumo del **giorno medio**

Per Onetone Research questa assunzione rappresenta una novità rivoluzionaria<sup>1</sup>.

L'esperienza di tre anni di attività ha portato Onetone Research e ripensare dalle fondamenta il modo di produrre i dati di audience e consumo del Web

Per tre anni, Onetone Research ha condotto una ricerca continuativa sul traffico e il consumo dei siti Web in Italia. Dal Marzo 2000 fino alla fine del 2002 il sistema adottato è stato fondato sull'analisi di un panel di utenti Internet rappresentativo dell'utenza residenziale.

Il panel veniva posto sotto osservazione mediante l'installazione di un software (il *meter*).

Il meter registrava il consumo del Web prodotto da ogni panelista e spediva con regolarità i dati a Onetone Research per la successiva analisi.

Questo field di ricerca, il primo realizzato in Italia, ci ha consentito di seguire lo sviluppo della Rete, l'evoluzione del consumo dei servizi online, la crescita delle audience dei siti Web.

Per primi abbiamo introdotto un sistema di misurazione omogeneo, basato su metriche finalmente utili per l'analisi del traffico e la pianificazione pubblicitaria: contatti lordi e netti, visite, GRP, classificazione degli interessi e dei contenuti, quota di attenzione dedicata ai diversi servizi Web etc...

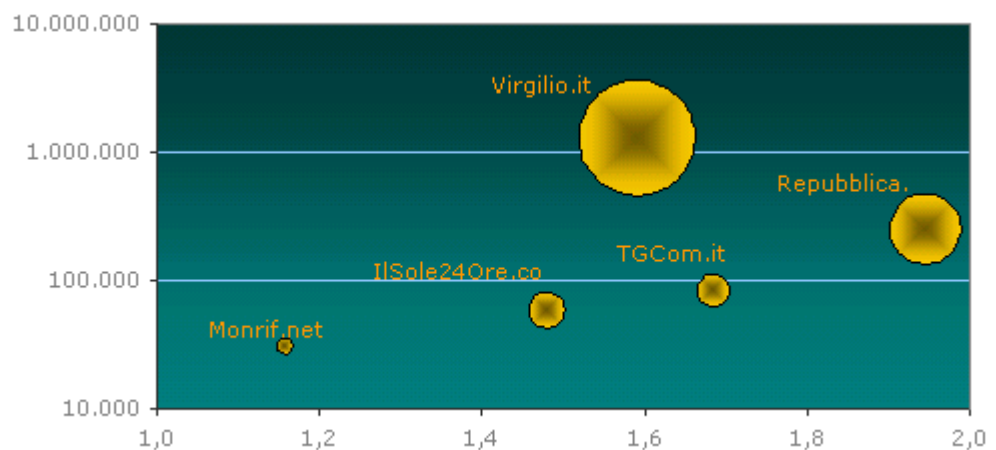
<sup>1</sup> Le pagine che seguono trattano argomenti di natura statistica e metodologica che rischiano di tediare il lettore in ragionamenti poco comprensibili e destinati ai soli addetti ai lavori. Questo stile rappresenta un'eccezione alla regola: i prossimi numeri della newsletter, così come quelli passati, saranno più leggeri e colorati.

Più volte, presentando i dati ottenuti dal panel, ci confrontavamo con altri dati in possesso degli editori dei maggiori siti Web italiani che affermavano di rilevare risultati talvolta molto diversi dai nostri.

Noi stimavamo i dati, loro li "contavano". Noi cercavamo da un campione di utenti di capire come stavano le cose, loro le stesse cose le vedevano mentre accadevano, le misuravano con una precisione per noi impossibile.

La dialettica tra il dato censuario e quello campionario ha attraversato il mercato della Rete dalla sua nascita e non accenna a risolversi. Il Cookie panel, come vedremo, produce la riconciliazione dei dati

D'altra parte, noi potevamo vedere il consumo del Web dalla parte dell'utente mentre loro lo vedevano solo dalla parte del sito. Quindi, la nostra analisi aggiungeva un valore importante: studiare gli utenti e i target, le sovrapposizioni dei siti, l'attenzione dedicata ai diversi servizi sul Web, indipendentemente dal sito sul quale venivano consumati.



Rilevazione di Marzo 2003 (fonte: Red Sheriff, **giorno medio**)  
I dati sono riprodotti nel classico piano copertura/volumi di Onetone Research (in ordinata è presente una scala logaritmica)

In linea di principio, dato campionario su panel e dato censuario sul sito sono due approcci perfettamente complementari ed entrambi utili. Spesso, tuttavia, i due dati finiscono per diventare "antagonisti"

Nonostante questa complementarità dell'analisi, il confronto tra dati panel e dati interni dell'editore sulle metriche di consumo (pagine viste, visite, tempo speso) era spesso frustrante, talvolta doloroso, alcune volte lacerante.

Noi dovevamo anche accettare che da questo confronto sui dati di consumo il nostro panel ne usciva sempre sconfitto.

L'analisi dei dati prodotti da un panel è un'analisi statistica che, a partire dall'osservazione di un campione di utenti, estende tale comportamento all'intero universo che si intende rappresentare (in questo caso l'intera utenza residenziale). Inoltre, il nostro panel

<sup>1</sup> I nostri dati indicano in circa 98% la copertura dell'utenza attualmente *cookizzata* da Red Sheriff in Italia.

<sup>2</sup> La privacy dei *cookie panelist* è rispettata in modo più severo rispetto ai panelisti dei panel meterizzati: ai panelisti del cookie panel non vengono infatti richiesti i dati anagrafici.

misurava solo l'utenza residenziale mentre il dato interno degli editori misurava tutto il consumo, indipendentemente dal luogo dal quale questo veniva generato.

E dire che la misurazione mediante panel meterizzati è un sistema consolidato: è quello adottato da Auditel in Italia e, in generale, è il sistema di misurazione degli ascolti televisivi in tutto il mondo.

La Rete è un sistema interattivo e quindi misurabile. Per la prima volta nella storia dei mezzi, gli editori conoscono meglio di ogni altro i volumi di consumo dei contenuti sul loro sito

Oggi appare evidente che l'approccio del panel meterizzato, ereditato dall'industria televisiva, nella misurazione del Web presenta due grandi, enormi, limiti:

1. al Web si accede da diversi luoghi di utilizzo: non solo da casa ma anche dal luogo di lavoro, da scuola o da università, dai luoghi pubblici. La rappresentazione dei consumi del Web non può prescindere dalla corretta considerazione dei contributi al consumo che derivano da tutti questi luoghi;
2. gli editori, a differenza delle TV, sanno con precisione come avviene il consumo sui siti Web delle loro Properties. Il dato che stima il consumo non corrisponde mai al dato rilevato con metodo censuario sul sito. Questo produce una dialettica tra due dati che dovrebbero misurare la stessa cosa ma si trovano a dare descrizioni diverse degli stessi fenomeni. Ma, tra il dato statistico e quello censuario, chi vince? Tra i dati di un sondaggio e quelli del voto, quali prenderesti in considerazione per stabilire che vince le elezioni?

Al problema del punto 1. abbiamo cercato di rispondere adottando lo stesso metodo (panel meterizzato) anche al luogo di lavoro. Il progetto di rappresentare l'utenza che accede dal luogo di lavoro è stato svolto nell'autunno del 2000 e il panel costruito era arrivato a dimensioni decisamente ragionevoli (700 individui) per produrre dei dati accettabili.

Il modello del panel meterizzato si applica con buoni risultati all'universo dell'utenza residenziale. Lo stesso modello non funziona quando si tratta di fornire la rappresentazione dell'utenza lavoro

Niente da fare.

Il problema non era la numerosità.

Certo, averne un po' di più di poche centinaia avrebbe aiutato un po' nella stabilità dei dati nel tempo. Il vero problema, emerso in modo del tutto evidente, è che il panel meterizzato ha bisogno dell'installazione di un software per funzionare e questo è difficile in almeno due casi sui tre nell'ambiente di lavoro.

Quindi ci siamo trovati di fronte a una drammatica alterazione della rappresentatività dell'universo. Questa dipendeva dalla metodologia utilizzata (la necessità di installare un meter) e non dalla numerosità del panel.

Installare un meter in una grande azienda, in un ministero, in una banca è onestamente un'impresa troppo difficile se si vuole procedere in modo serio.

Ci si scontra con problemi sindacali, limitazioni tecniche legate alla configurazione delle reti locali delle grandi aziende, politiche di installazione dei software sui PC dei dipendenti, lunghi processi burocratici che spesso arrivano a coinvolgere lo stesso vertice aziendale.

Non è realistico pensare di svolgere questo lavoro in almeno trecento aziende, in tempi accettabili e con costi contenuti.

Non si può far passare un panel SOHO per un panel capace di rappresentare tutti gli utenti dal luogo di lavoro: o si è in malafede oppure non si è capito di cosa si sta parlando

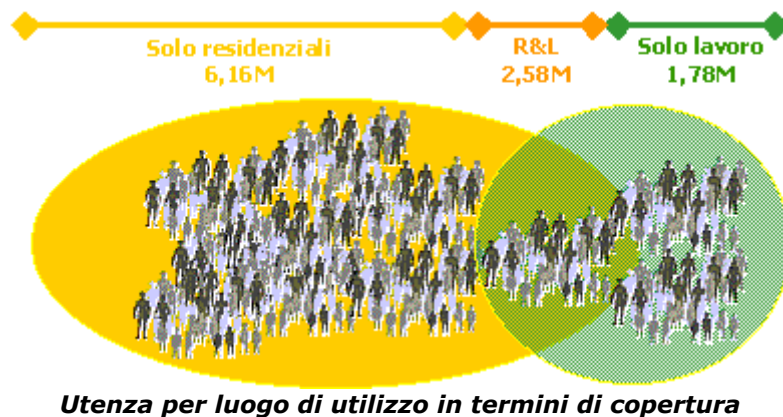
Sono però esattamente questi i luoghi nei quali si produce gran parte del consumo dal luogo di lavoro, cioè le grandi aziende, le banche, i ministeri etc...

Il mercato SOHO (Small Office, Home Office) composto da piccoli artigiani, professionisti, studi e imprese individuali, ha uno stile di fruizione e consumo del Web paragonabile a quello che si sviluppa da casa. In tali contesti è relativamente facile installare un meter ma, alla fine, la presunta operazione di costruzione di un panel di utenti dal luogo di lavoro si configura in realtà come una banale estensione dell'universo residenziale al solo universo SOHO.

Noi oggi pensiamo che non si possa affermare in buona fede di poter costruire un panel meterizzato rappresentativo degli utenti dal luogo di lavoro.

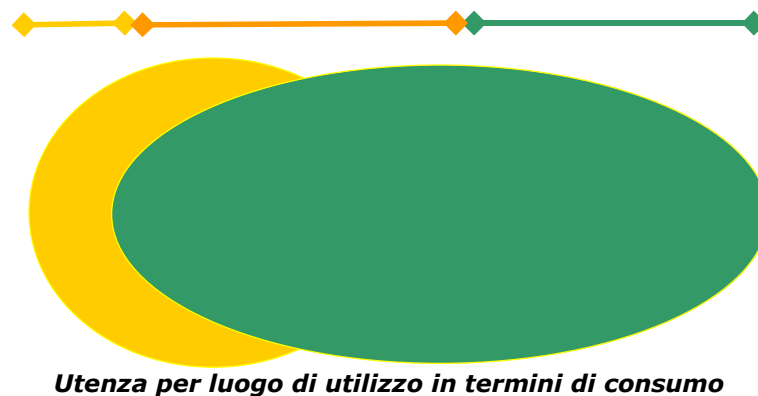
**Audience & copertura:**  
l'utenza residenziale risulta prevalente nel totale dell'universo degli utenti Internet (la sua numerosità è doppia rispetto all'utenza lavoro)

Inoltre, l'utenza residenziale è ben rappresentabile in un panel meterizzato



**Audience & consumo:**  
l'utenza lavoro è responsabile del 60-80% dei consumi Web.

Gran parte di questo segmento di utenza non può essere rappresentato con un panel meterizzato



Come, poi, dire di poter misurare con questo sistema le scuole, le università, gli *Internet caffè*?

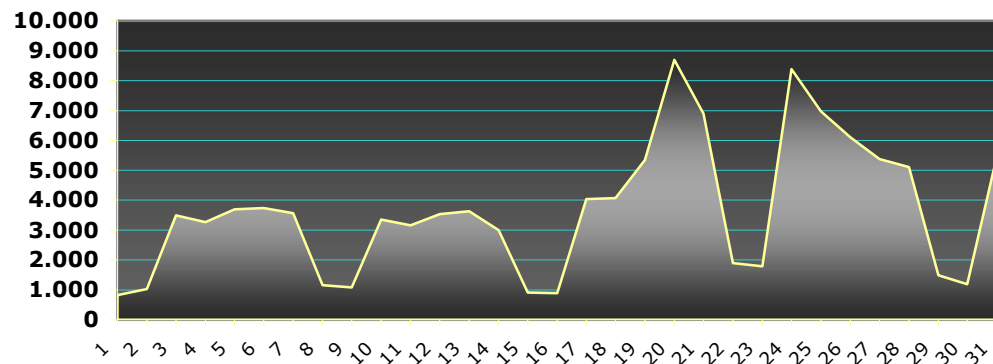
A questa obiezione, uno spirito semplice risponde dicendo che l'utenza residenziale è maggioritaria nell'insieme complessivo totale dell'utenza del Web. Quindi, proseguirebbe il nostro personaggio, misurare l'utenza residenziale pari all'84% del totale significa partire da una ricerca che, se proprio non spiega tutto, almeno spiega il comportamento della gran parte dell'utenza.

Come si nota graficamente nelle due figure precedenti, *in termini di copertura* è vero che l'utenza residenziale è assolutamente maggioritaria.

Ma quando si passa a considerare il consumo, cambia tutto. L'utenza dal lavoro diventa responsabile di almeno il 60% dei consumi e su alcuni siti di informazione (soprattutto informazione finanziaria) questa quota può salire oltre l'80%.

A prova di quanto appena detto basta guardare il grafico del numero di pagine viste per giorno del sito Web di *repubblica.it*:

Il consumo (pagine viste) di un sito di informazione come *repubblica.it* cade di circa 4 volte nei giorni festivi rispetto ai giorni feriali



Pagine viste (x1000) per giorno su *repubblica.it* nel mese di Marzo 2003 (fonte: Red Sheriff, per gentile concessione di *repubblica.it*). I giorni di fine settimana sono ben individuabili guardando la caduta dei consumi.

I limiti del panel meterizzato: rappresentatività dell'universo lavoro e difficoltà a rilevare correttamente i volumi di consumo

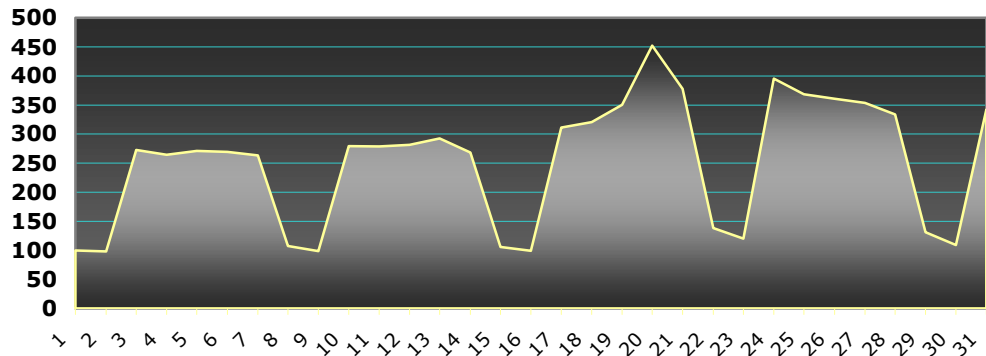
Le conclusioni del primo problema esposto in precedenza sono due:

1. per ottenere un quadro completo del consumo del Web non si può non considerare il contributo dell'utenza "non residenziale"
2. il profilo e il consumo dell'utenza non residenziale non possono essere rappresentati correttamente da un panel meterizzato

Per questi due motivi Onetone Research si è sempre limitata a dare i risultati della misurazione dell'utenza solo residenziale, senza la presunzione di poter tracciare anche quella dal luogo di lavoro, pur avendoci provato per diversi mesi.



Il numero di lettori di un sito di informazione come [repubblica.it](http://repubblica.it) cade di circa 2,5 volte nei giorni festivi rispetto ai giorni feriali. Il consumo pro-capite è diverso nei giorni feriali rispetto a quelli festivi.



Browser unici (x000) per giorno su [repubblica.it](http://repubblica.it) nel mese di Marzo 2003 (fonte: Red Sheriff, per gentile concessione di [repubblica.it](http://repubblica.it)). Anche qui i giorni di fine settimana sono ben individuabili.

Gli stessi motivi che ci portano a riconoscere i limiti dell'approccio del panel meterizzato sono alla base della convinzione che il metodo del Cookie Panel sia la soluzione convincente al problema della misurazione del consumo dei mezzi digitali e interattivi.

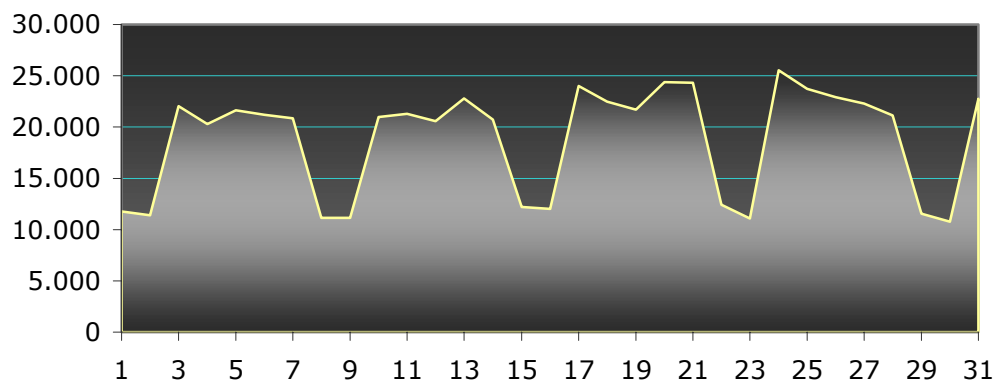
Il Cookie Panel è il sistema di misurazione che acquisisce i dati di consumo censuari e arricchisce i risultati con l'analisi di profilo degli utenti

A questo punto è necessario capire meglio cosa intendiamo per Cookie Panel.

Il termine si riferisce al fatto che la nuova ricerca è basata su un campione di cookies che viene seguito nel tempo (nella terminologia delle ricerche, un campione stabile nel tempo è detto panel).

Si tratta di un *Cookie Panel* perché l'universo rappresentato è quello dei cookies italiani e l'unità campionaria è il cookie.

Il consumo (pagine viste) del maggior portale italiano, [virgilio.it](http://virgilio.it): la caduta dei consumi (pagine viste) non è vistosa come nel caso di [repubblica.it](http://repubblica.it) ma è pur sempre significativa: il traffico è dimezzato nei giorni festivi rispetto ai giorni feriali



Pagine viste (x000) per giorno su [repubblica.it](http://repubblica.it) nel mese di Marzo 2003 (fonte: Red Sheriff, per gentile concessione di Matrix SpA). I giorni di fine settimana sono ben individuabili guardando la caduta dei consumi.

Se misurare il Web fosse solo stabilire i volumi di consumo dei siti, basterebbe il sistema di Red Sheriff

In realtà, il problema della misurazione dei consumi dei siti Web potrebbe considerarsi risolto in modo brillante con la metodologia di Red Sheriff. Nella figura presentata qui sopra si può seguire l'andamento delle pagine Web richieste su [Virgilio.it](http://Virgilio.it) in tutti i giorni del mese di Marzo 2003. Il sistema adottato è quello censuario, le



pagine sono contate, i dati sono precisi e aggiornati ogni giorno: dov'è quindi il problema? Cioè, perché dobbiamo ragionare sui cookies e introdurre il Cookie Panel quando abbiamo già un sistema preciso di misurazione?

Misurare il Web in modo completo è possibile quando ai volumi di consumo si associano le informazioni di profilo degli utenti e si sviluppa un sistema in grado di seguire gli utenti su più siti

Vediamola da un altro punto di vista.

Le indagini di consumo e audience sui mezzi servono essenzialmente al mercato della comunicazione e della pubblicità.

Chi opera in questa industria non può limitarsi a conoscere i volumi di consumo di un sito Web: sapere che un sito fa 100 o 200 milioni di pagine in un mese è un dato abbastanza interessante ma non è il parametro con il quale si decide se acquistare spazi pubblicitari sul primo o sul secondo sito, quali formati adottare, quale pressione di comunicazione esercitare.

Chi comunica ragiona essenzialmente *per target*: (1) definisce un insieme di individui obiettivo della comunicazione, il target, (2) elabora una creatività, (3) lavora perseguendo alcuni obiettivi definiti con il cliente, pianificando la campagna ed eventualmente, (4) misurandone l'efficacia.

Per fare questo deve analizzare i dati di misurazione del mezzo ottenuti *dal punto di vista del fruitore* del mezzo. Non basta il dato di consumo sul singolo sito.

La disponibilità di dati di misurazione che partono dal punto di vista dell'utente Internet, ovunque esso sia, consente di seguire un target su tutta la Rete per capire quali siti visita, con quale frequenza e quale intensità. Mettendo insieme le caratteristiche di un target e i dati di tutti i siti, il pianificatore ha la possibilità di produrre il giusto mix di comunicazione, bilanciando nel modo più efficace formati e creatività, siti da coinvolgere e pressioni da applicare.

Per misurare il Web non si può rinunciare all'approccio del panel: ben campionato e reclutato, stabile, profilato in modo adeguato agli obiettivi dell'analisi

Per queste ragioni un campione di utenti stabile nel tempo, un panel, è la scelta di gran lunga preferibile per produrre i dati di misurazione di un mezzo: ciò è valido anche nel caso di un mezzo interattivo, quindi misurabile, come Internet.

Tuttavia, come abbiamo visto in precedenza la misurazione mediante un panel meterizzato non è soddisfacente. Perché lo dovrebbe essere il Cookie Panel? Per due motivi:

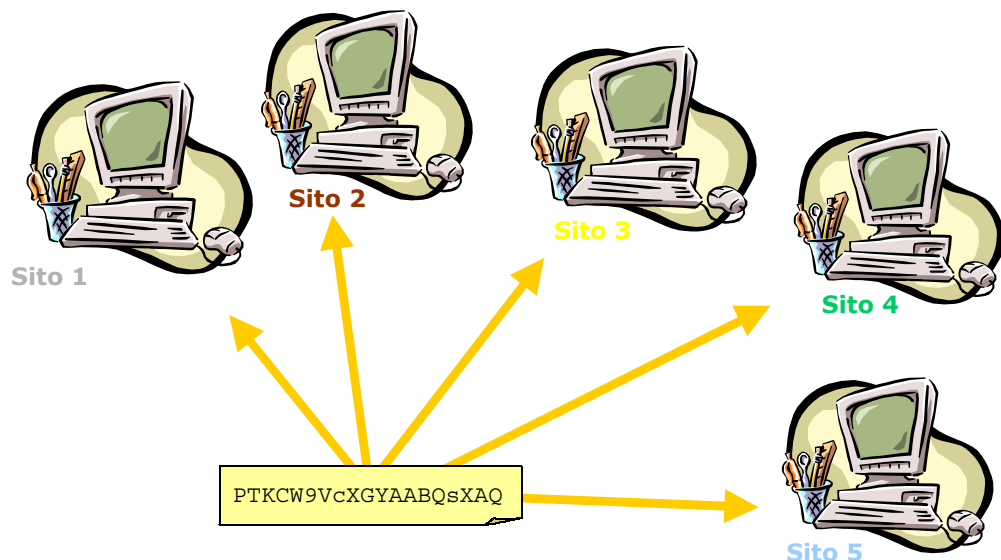
1. il Cookie Panel riproduce il massimo livello di rappresentatività possibile.
2. Il Cookie Panel assicura la piena aderenza al dato di consumo misurato sul versante dei siti Web

Il primo motivo si spiega con il fatto che il cookie di Red Sheriff è già installato sul browser di (quasi) ogni utente italiano<sup>1</sup>. Red

Sheriff misura diversi grandi siti italiani (portali, editori) e la quota di utenti che non interagisce mai con tali siti è decisamente minoritaria.

La conseguenza è che Red Sheriff con ottima approssimazione è in grado di seguire tutti i cookies italiani, dal momento che ne rileva l'interazione con i siti Web sotto misurazione.

Il cookie di Red Sheriff *diffonde* per il Web e interagisce con i siti sotto misurazione. In questo modo il cookie lascia una traccia e questa viene successivamente utilizzata per produrre i dati di consumo e per qualificare l'audience dei diversi siti per caratteristiche sociodemografiche



L'interazione del cookie con i diversi siti misurati da Red Sheriff lascia una traccia del passaggio: raccogliendo le diverse tracce dei cookies profilati si costruisce il database del Cookie Panel.

Di per sé, un cookie è l'identificativo di un browser, nulla più. Il Cookie Panel viene costruito associando ai cookie uno o più profili individuali degli utilizzatori del browser. Allo stesso modo, un utente che usa più browser viene individuato come un solo individuo a cui sono associati più cookies

Un cookie, visto da vicino, è un codice incomprensibile formato da lettere e numeri. Tuttavia, il cookie è legato in modo univoco al browser che l'utente utilizza sul Web ed è sempre lo stesso. Questo significa che il cookie può essere riconosciuto come uno stesso individuo anche a distanza di tempo, anche su siti diversi.

L'idea di base del Cookie Panel è di individuare un certo numero di cookies (da alcune migliaia fino a circa 50.000) da mettere sotto osservazione nella loro interazione con i diversi siti in modo da seguire nel tempo la loro navigazione. Questa osservazione avviene mediante la tecnologia di Red Sheriff ed è alternativa all'utilizzo del meter.

Il cookie, quindi, rimpiazza il meter del panel meterizzato. La differenza è che mentre i meter bisogna installarli, i cookie ...sono già installati.

Il Cookie Panel rispetta alla radice l'anonimato dei partecipanti: un ulteriore vantaggio rispetto ai panel meterizzati

Per fare di un Cookie Panel uno strumento di ricerca da cui ottenere profili degli utenti, target e dati utili per la pianificazione pubblicitaria è necessario "dare un volto" ai cookies selezionati.

In realtà non un volto in senso anagrafico<sup>2</sup> ma solo alcuni dati di profilo che sappiano qualificare l'audience dei siti e individuare i target di utenti (per esempio, sesso, età, titolo di studio, luogo di residenza etc...).

Il reclutamento di un cookie nel Cookie Panel è un'operazione semplicissima e non invasiva

Questa operazione è detta profilazione: un cookie profilato entra a far parte del Cookie Panel.

Tutto ciò avviene mediante la dichiarazione dei dati di profilo e l'azione di un solo click dell'utente: un minuto di tempo, nessun software da installare, nessun permesso da chiedere se non l'autorizzazione al trattamento dei dati personali (in ogni caso, anonimi).

Finito questo passaggio abbiamo creato il Cookie Panel. Abbiamo quindi risolto il primo problema dei panel meterizzati, la limitata rappresentatività. Infatti, il Cookie Panel è perfettamente rappresentativo sia dell'utenza che accede alla Rete dal luogo di residenza che dal luogo di lavoro.

Il Cookie Panel è il sistema che assicura la massima rappresentatività possibile. Tuttavia, l'ottima rappresentatività non è sufficiente a produrre la riconciliazione tra dato campionario e dato censuario

Supponiamo ora di avere un Cookie Panel ben rappresentativo e di dimensioni ragguardevoli, per esempio 10.000 individui.

Questa numerosità ci assicura che il dato di consumo rilevato con il Cookie Panel sia aderente a quello censuario rilevato con Red Sheriff sui siti Web?

No.

Arriviamo così al secondo problema, quello della discrepanza tra dato censuario ottenuto con il sistema del conteggio delle pagine e dato statistico, ottenuto con la proiezione dei dati del panel a tutto l'universo di indagine.

Anche in presenza di un panel ben rappresentativo (come può essere considerato il Cookie Panel), il dato di consumo ottenuto dall'analisi è affetto da errore.

Non si tratta però di un errore sistematico originato alla fonte dei dati, come potrebbe essere quello derivante dalla debole rappresentatività del panel meterizzato.

Si tratta piuttosto di un errore statistico, dovuto al fatto che i dati ottenuti con campioni di individui sono sempre accompagnati da una distribuzione di probabilità associata al risultato. Il dato ottenuto non è da intendere in senso puntuale ma in un intervallo di valori. Tale intervallo può essere decisamente ampio nel caso dei panel che misurano il comportamento di individui.

Chi vince tra dato censuario e dato campionario? E' necessario ammettere la superiorità del dato censuario e ponderare il dato campionario sul dato censuario.

Perché?

Ricordate i sondaggi politici prima delle elezioni o gli Exit Poll prodotti durante il giorno delle elezioni. I risultati sono presentati al pubblico citando una "forchetta", un intervallo di valori. L'ampiezza della forchetta è generalmente una funzione del numero di intervistati: maggior il numero di intervistati, maggiore la confidenza dei risultati, ovvero minore l'ampiezza della forchetta.

Lo stesso vale per un panel, sia esso meterizzato o misurato tramite cookie. I dati prodotti analizzando il campione di utenti saranno sempre affetti da errore.

Dato medio e  
varianza: troppo  
spesso ci si  
dimentica dei  
fondamenti della  
statistica

Quando diciamo che in Ottobre 2002 il numero di utenti residenziali di Virgilio.it è stato pari a 4.644.000 (fonte: Onetone Research) in realtà stiamo dicendo che potrebbe essere tra 4,55 milioni 4,7 milioni. Questo assumendo che il panel sia rappresentativo, altrimenti il dato deve essere messo in discussione nel suo complesso.

Quando diciamo che il numero di pagine viste di Virgilio.it prodotto dall'utenza residenziale in Ottobre 2002 è stato pari a 195,9 milioni di pagine viste, anche qui stiamo dando un dato che dovrebbe essere accompagnato dalla sua forchetta. Quanto è grande questa forchetta? Qui le cose si complicano un po'.

E' possibile  
mostrare  
matematicamente  
la ragione per la  
quale i panel  
meterizzati faticano  
a rappresentare con  
precisione i  
consumi di utilizzo  
dei siti Web

Il dato delle pagine viste, in generale il dato di consumo su un sito, è il risultato di un prodotto tra pagine viste medie per utente (PV/U) e numero complessivo di utenti (UU):

$$PV = PV/U \times UU$$

Il numero è il prodotto di due dati che sono affetti da errore (cioè "ballano" in una forchetta).

Il numero di utenti (UU), per esempio, può variare in una forchetta ampia 250mila unità (il 4% del suo valore). Anche il numero di pagine viste per utente (PV/U) presenta la sua varianza.

Il numero complessivo di pagine viste si trova così ad avere lui stesso una forchetta, determinata dalla composizione dei due errori relativi delle due variabili che compongono il prodotto:

$$\Delta PV = \text{RadQ} [ \Delta(PV/U)^2 + \Delta(UU)^2 ]$$

Il numero delle pagine viste si porta con sé gli errori dei due da cui proviene da cui è prodotto.

Nei casi migliori, PV può presentare una varianza del 10-15% ma in molti casi la sua varianza è ben superiore.

La distribuzione di  
consumo del Web  
mostra un profilo  
simile a quella di  
molti altri casi.  
Metà del consumo  
Web è generata da  
una piccola quota di  
utenti. Ciò rende  
più inaffidabili i  
risultati dei volumi  
di consumo ottenuti  
con un panel

Questi esempi sono ottenuti considerando un panel (cookie o meterizzato nell'esempio non fa differenza) di 2.000 unità. All'aumentare della numerosità, le forchette tendono a diminuire di ampiezza e il dato diventa più accurato.

Si tenga però presente quest'altra componente, che indebolisce ulteriormente l'affidabilità del dato di consumo prodotto da un panel.

Dall'analisi dei dati del panel di Onetone Research emerge come il 12-15% dell'utenza di un sito Web sia responsabile del 50% delle Pagine Viste.

Questa distribuzione è tipica del consumo di molti beni e servizi (birra, sigarette, pedaggi autostradali etc..) e non è una sorpresa ritrovarla anche sul Web.

La conseguenza è che in un panel di 2.000 individui, solo 240-300 di essi sviluppano la metà dei volumi di consumo.

La varianza di tali dati, quindi, è circa tre volte superiore a quella che si ottiene assumendo un panel composto da 2.000 individui.

Due conclusioni dalla precedente analisi: il consumo del Web deve essere ricavato con metodo censuario

Questa breve digressione spiega due cose:

1. un panel ben rappresentativo può dare indicazioni sulla copertura dei fenomeni osservati (quindi: reach & utenti) e sul profilo dell'utenza. Non può dare indicazioni altrettanto precise sui volumi di consumo;
2. anche in presenza di un panel ben rappresentativo e numeroso, i dati di consumo dei siti Web ottenuti con un panel possono essere notevolmente diversi da quelli rilevati con metodo censuario sugli stessi siti Web.

La soluzione non può che essere una: ponderare i dati di consumo del panel con i dati di consumo più precisi e affidabili ottenuti da Red Sheriff con il metodo censuario.

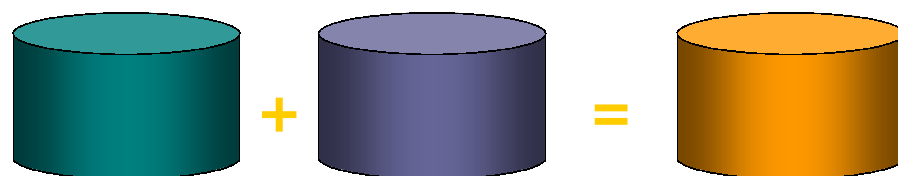
La ponderazione allinea i dati panel con i dati del sito: fine delle dialettiche e delle lacerazioni del mercato. Un solo dato, accettato tanto dagli editori quanto dai pianificatori della pubblicità online

Per questo motivo possiamo dire che il nuovo metodo del Cookie panel introduce un secondo vantaggio dopo il primo della rappresentatività.

Lo abbiamo chiamato la "riconciliazione dei dati": si tratta di riconciliare i risultati dell'approccio statistico con quelli dell'approccio campionario.

Allo stesso tempo, si tratta di riconciliare le ragioni metodologiche dell'editore che misura i propri dati interni con quelle dell'istituto che misura gli stessi dati dall'esterno, con un panel. Nei nostri propositi questa riconciliazione dei dati dovrebbe anche riconciliare le posizioni spesso divergenti tra editore ed istituto.

Il passaggio metodologico della riconciliazione dei dati di consumo prevede che i dati prodotti dal panel (e pertanto affetti da errore statistico) vengano ponderati sui dati di consumo ottenuti sui medesimi siti Web con il sistema di Red Sheriff (non affetti da errore statistico perché ottenuti con metodo censuario).



**Dati di consumo  
forniti da Red  
Sheriff**

**+**  
**Dati di profilo  
forniti da Onetone  
Research**

**=**  
**Dati riconciliati del  
Cookie Panel**

In questo modo si può creare una perfetta aderenza tra i dati censuari e quelli statistici, senza perdere la ricchezza dell'analisi su panel.

Il Cookie Panel  
ottiene due risultati  
positivi: la miglior  
rappresentatività  
possibile  
dell'universo di  
indagine unita alla  
riconciliazione dei  
dati di consumo

Il nuovo sistema consente quindi di prendere il meglio dei due approcci di misurazione:

1. il dato quantitativo è ottenuto dalla rilevazione di Red Sheriff sui siti degli editori. Il dato è di origine censuaria e quindi va considerato più attendibile del dato panel;
2. il dato di profilo, oltre alla possibilità di seguire il consumo del Web dal versante dell'utente, è ottenuto dall'analisi del Cookie Panel.

Il Cookie panel è  
già oggi il sistema  
migliore e potrà  
resistere anche  
all'evoluzione futura  
della Rete

Il Cookie Panel è basato su una metodologia solida e a nostro parere rappresenta l'alternativa da tempo ricercata alla misurazione instabile e insoddisfacente dei panel meterizzati. Inoltre, il sistema di misurazione è capace di reggere e ben rappresentare il consumo di Internet anche nel prossimo futuro, quando la Rete sarà utilizzata non solo dai PC ma anche da altri device digitali.

Su di essi installare un meter potrà essere un problema. Un buon cookie, invece, lo troveremo sempre dappertutto.

I dati di Marzo 2003 del Cookie Panel sono disponibili su:  
<http://research.onetone.it/newsletter/2003/cookpan.zip>

**Marzo 2003**

<http://research.onetone.it/>

**Stiamo misurando il Web**