

Big data e giovani universitari europei. Una ricerca empirica

Simona Maria Cavagnero¹, Maria Adelaide Gallina² e Renato Grimaldi³

¹ Università di Torino, 10124 Torino, Italia

² Università di Torino, 10124 Torino, Italia

³ Università di Torino, 10124 Torino, Italia
renato.grimaldi@unito.it

Abstract.

Il contributo intende studiare la consapevolezza degli studenti universitari in merito alle “briciole digitali” che quotidianamente si generano utilizzando la Rete. I giovani hanno una grande familiarità con le tecnologie digitali, la comunicazione e la condivisione di informazioni personali. La ricerca che viene presentata ha rilevato il valore che i giovani attribuiscono ai propri dati ed è stata condotta sul territorio italiano ed europeo nell’anno accademico 2016-2017 (Torino, Biella, Genova, Palermo, Valencia-Spagna, Resita-Romania), coinvolgendo oltre mille studenti. Lo strumento d’indagine utilizzato è un questionario a risposta prevalentemente chiusa, mediante il quale si sono rilevate le azioni svolte in Rete, il possesso di strumenti infotelematici, il valore (misurato in euro) dei dati. Si è voluto controllare il valore che gli studenti attribuiscono alle loro informazioni quando vengono “cedute”, e al tempo stesso quello che essi ritengono di trarre dai servizi che vengono “resi” quando i loro dati sono intercettati da enti esterni. Sulla base di questi dati sono stati elaborati quattro profili di giovani universitari considerando il numero di attività svolte in Rete dichiarate e il numero di oggetti tecnologici posseduti. I risultati dimostrano la necessità educativa di formare le nuove generazioni all’uso consapevole della Rete e del valore dei propri dati che quotidianamente lasciano on-line. Cavagnero ha redatto la sezione 1, Gallina la 2 e Grimaldi la 3.

Keywords: Big Data, uso della Rete, consapevolezza, tracce digitali

1 Uso della Rete e Big data. Un quadro teorico

L’utilizzo diffuso della Rete genera quotidianamente una mole di dati considerevole, restituendo un profilo dei nostri “usi e costumi”, che viene immagazzinato da aziende a cui cediamo nostri dati per poter usufruire di diversi servizi, dalla navigazione per ricerca di informazioni all’utilizzo di App ai social network. A questo proposito il rapporto Censis del 2015 presenta dati relativi all’importanza dei social network nel

nostro Paese: il più utilizzato dagli italiani, risulta Facebook con una quota del 50,3%, seguito a ruota da Youtube 42%, Twitter 10,1% delle persone coinvolte nella ricerca, mentre Instagram il 9,8%. La Rete ha l'obiettivo di creare interazione ma spesso i 'profili utente' protetti da password e concepiti, ovviamente, come astratti e virtuali sono reali e concreti in quanto dietro a ognuno di essi vi sono molte delle azioni sociali compiute: basta pensare agli aggiornamenti del proprio profilo, alla condivisione di foto o di pensieri che riguardano la vita strettamente personale, agli acquisti, etc.

Tale mole potenzialmente infinita di informazioni è denominata Big Data, un insieme di dati in formato digitale ottenuti e elaborati attraverso dataset di ampie dimensioni, che registrano comportamenti individuali e collettivi degli attori sociali: desideri, opinioni, stili di vita, movimenti, relazioni [Giannotti, 2015], e si caratterizzano per *volume* dei dati, *velocità* e *varietà*, [Kitchin, 2014], oltre che *variabilità* e *complessità* [Kshetri, 2014].

Visto l'uso pervasivo della Rete e dei social sites è importante favorire una riflessione in merito alla consapevolezza che i giovani hanno delle tracce digitali che lasciano quotidianamente in Rete. A questo proposito si è cercato di misurare il valore che gli studenti universitari attribuiscono ai propri dati che vengono generati attraverso l'utilizzo della Rete. L'ambito spazio-temporale dell'indagine ha come riferimento l'anno accademico 2016-2017 e il coinvolge diversi grappoli di popolazioni di studenti universitari appartenenti a corsi di laurea provenienti da Italia, Romania e Spagna. L'esito non fornisce un campione casuale e rappresentativo in senso statistico, ma una raccolta di dati che se da una parte non ci permettono di fare inferenze verso popolazioni differenti da dove si è scesi sul campo, dall'altra consentono di far emergere relazioni interne alla realtà esplorata. Tale indagine si è svolta attraverso un questionario strutturato a risposta prevalentemente chiusa e sono stati indagati diversi aspetti: le azioni svolte in Rete, il possesso di strumenti infotelematici, soffermandosi in particolare sulla consapevolezza in merito ai propri dati lasciati in Rete. Una parte dello strumento di indagine è stata utilizzata per misurare il valore che gli studenti attribuiscono ai loro dati quando vengono "ceduti" in Rete, e al tempo stesso il valore che essi ritengono di trarre dai servizi che vengono "resi" quando le loro informazioni sono intercettate da enti esterni, come ad esempio l'uso delle mappe di Google per la geo-localizzazione. Per ciascuna delle preferenze (politiche, religiose, sessuali, sportive, alimentari, della moda, del tempo libero, automobilistiche) sono stati rilevati questi due valori con le seguenti due domande (che abbiamo ripetuto per ciascuna delle preferenze indagate, qui riferiti alle sole "politiche") che per brevità chiameremo rispettivamente OUT e IN:

[politiche_OUT]

Qual è il valore minimo per cui saresti disposto a comunicare le tue preferenze *politiche* sapendo che potranno essere cedute a una organizzazione interessata (che le terrà riservate)? (Esprimi tale valore in euro; nel caso tu ritenga che le tue preferenze non abbiano valore, scrivi: 0) _____

[politiche_IN]

Sapendo che l'organizzazione interessata alle tue preferenze *politiche*, avendo acquisito i tuoi dati, potrà inviarti informazioni o materiale di tuo interesse, quanto ritieni possa valere tale servizio? (Esprimi tale valore in euro; nel caso tu ritenga che non abbia valore, scrivi: 0)

Al fine di delineare alcuni profili degli intervistati si prenderanno in considerazione i principali dati emersi, soffermandoci su 5 variabili indici generate come sintesi attraverso le valutazioni delle preferenze, delle attività svolte, degli oggetti posseduti, e l'età. Come si evince dalla Tabella 1 il campione, costituito da 1.091 studenti universitari, mostra un'età media di 22,9 anni, una media di 1.107,2 euro come valore complessivo delle proprie preferenze cedute (OUT) e una media di 688,2 euro del valore attribuito alle proprie preferenze nel caso in cui vengano restituiti servizi (IN). Gli studenti svolgono in media 7,7 attività in Rete, e possiedono in media 3 oggetti tecnologici.

Tabella 1. Statistiche delle principali variabili cardinali.

Variabili	N	Minimo	Massimo	Media	Deviazione std.
Valore complessivo in euro delle proprie preferenze nel caso vengano cedute (OUT)	883	0	80000	1107,2	4832,5
Valore complessivo in euro delle proprie preferenze nel caso vengano restituiti servizi (IN)	908	0	80000	688,2	3541,2
Differenza tra valutazione dei dati ceduti e servizi ricevuti (OUT-IN)	840	-80000	80000	336,9	5065,9
N attività in Rete	1091	0	18	7,7	2,9
N oggetti tecnologici posseduti	1091	0	7	2,9	1,0
Età (anni)	1047	17	51	22,9	4,2

2 Una descrizione del campione dei giovani universitari e delle loro tracce digitali

Un obiettivo che ha caratterizzato questo lavoro di ricerca è stato quello misurare il valore che gli studenti universitari danno ai loro dati. In particolare è stata posta attenzione alle loro preferenze politiche, religiose, sessuali, sportive, alimentari, della moda, del tempo libero, automobilistiche, che in questa sezione abbiamo sommate insieme. Rispetto a una precedente ricerca svolta su questo tema [Grimaldi, Cavagnero, Gallina, 2016] è stato controllato

il valore che gli studenti attribuiscono ai loro dati quando vengono “ceduti”, ma al tempo stesso è stato rilevato il valore che essi ritengono di trarre dai servizi che vengono “resi” quando le loro informazioni sono intercettate da enti esterni. Il campione è composto prevalentemente da soggetti di genere femminile (75,6%), come conseguenza della scelta dei corsi di studio oggetto della ricerca. La Tabella 2 evidenzia come non sussistano grandi differenze di genere rispetto alle variabili indagate; solamente in merito al valore complessivo attribuito alle proprie preferenze cedute (OUT) si nota una leggera differenza tra femmine (1.045 euro) e maschi (1.301 euro) e in merito al numero di attività in Rete svolte (7,5 per le femmine e 8,4 per i maschi).

Tabella 2. Comparazione delle medie delle principali variabili cardinali per genere dell'intervistato.

Genere		Valore complessivo in euro delle proprie preferenze nel caso vengano cedute (OUT)	Valore complessivo in euro delle proprie preferenze nel caso vengano restituiti servizi (IN)	Differenza tra valutazione dei dati ceduti e servizi ricevuti (OUT-IN)	N attività in Rete	N oggetti tecnologici posseduti	Età (anni)
femmine	Media	1045	684	296	7,5	2,9	22,9
	N	665	681	631	813	813	789
	Dev. std.	5149	3813	5678	2,7	1	4,2
maschi	Media	1301	710	462	8,4	2,9	22,7
	N	217	224	208	262	262	255
	Dev. std.	3716	2582	2428	3,1	1,1	3,6
Totale	Media	1108	690	337	7,7	2,9	22,8
	N	882	905	839	1075	1075	1044
	Dev. std..	4835	3547	5069	2,8	1	4,1

La Tabella 3 compara invece le medie delle principali variabili indagate in riferimento alla nazione di residenza dell'intervistato. Sebbene dalle analisi statistiche, in particolare con il calcolo dell'Eta quadro, non emergano relazioni importanti, possiamo comunque notare alcune specificità. Infatti la media del valore complessivo delle preferenze “cedute” (OUT) è superiore in Spagna (1.776 euro) rispetto all'Italia e alla Romania (rispettivamente 1.029 e 847 euro); la media del valore complessivo delle proprie preferenze nel caso in cui vengano restituiti servizi (IN) vale 844 euro per la Spagna, 710 per

l'Italia e 435 per la Romania. Pare dunque che gli studenti di Valencia attribuiscono un valore maggiore ai loro dati.

Tabella 3. Comparazione delle medie delle principali variabili cardinali per nazione dell'intervistato.

Nazione		Valore complessivo in euro delle proprie preferenze nel caso vengano cedute (OUT)	Valore complessivo in euro delle proprie preferenze nel caso vengano restituiti servizi (IN)	Differenza tra valutazione dei dati ceduti e servizi ricevuti (OUT-IN)	N attività in Rete	N oggetti tecnologici posseduti	Età (anni)
Italia	Media	1029	710	193	8,1	3	22,6
	N	643	657	607	804	804	772
	Dev. std.	3708	4100	4023	2,8	1	3,8
Romania	Media	1776	435	1459	6,6	2,5	24,7
	N	121	131	114	166	166	154
	Dev. std.	9791	1041	10104	2,9	1,2	6,2
Spagna	Media	847	844	-4	6,3	2,8	22,4
	N	119	120	119	121	121	121
	Dev. std.	1300	1275	419	2,8	1	2,1
Totale	Media	1107	688	337	7,7	2,9	22,9
	N	883	908	840	1091	1091	1047
	Dev. std.	4833	3541	5066	2,9	1	4,2

Infine portando l'attenzione sulle differenze per corso di studio frequentato, si può notare (Tabella 4) che gli studenti di ingegneria siano maggiormente consapevoli del valore delle preferenze "cedute" (1.566 euro); ritengono invece scarsamente importante il valore dei servizi "ricevuti" (381 euro). Inoltre i futuri ingegneri sono quelli che svolgono il maggior numero di attività in Rete: 9,2 rispetto a 7,8 degli studenti di scienze della vita e 7,5 di studi umanistici. Non ci sono differenze significative sul numero di oggetti tecnologici posseduti; tutti ne possiedono mediamente 3.

Tabella 4. Comparazione delle medie delle principali variabili cardinali per corso di studi frequentato dall'intervistato.

Corso di studi umanistici, scienze della vita, ingegneria		Valore complessivo in euro delle proprie preferenze nel caso vengano cedute (OUT)	Valore complessivo in euro delle proprie preferenze nel caso vengano restituiti servizi (IN)	Differenza tra valutazione dei dati ceduti e servizi ricevuti (OUT-IN)	N attività in Rete	N oggetti tecnologici posseduti	Età (anni)
umanistici	Media	1035	739	209	7,5	2,8	23
	N	694	712	656	853	853	827
	Dev. std.	4484	3771	4933	2,7	1	4,3
scienze della vita	Media	1220	627	452	7,8	3,0	22,4
	N	99	99	97	108	108	104
	Dev. std.	3838	3277	1746	3,1	1,2	2,8
ingegneria	Media	1566	381	1202	9,2	3,1	22,6
	N	84	92	82	111	111	107
	Dev. std.	7905	1467	8027	3,2	1,1	4,4
Totale	Media	1107	690	335	7,7	2,9	22,9
	N	877	903	835	1072	1072	1038
	Dev. std.	4848	3551	5080	2,8	1	4,2

3 I profili degli intervistati: Quiet, Effective, Collector e Smart

Una sezione del questionario ha misurato le attività che i giovani universitari coinvolti nella ricerca compiono in Rete (di cui abbiamo già visto alcuni risultati), mediante una classificazione di 18 dimensioni: sottoscrivere tessere in supermercati/negozi, fare acquisti on-line, fare vendite on-line, rispondere a sondaggi (sia telefonici, sia on-line), iscriversi a siti per ottenere informazioni, iscriversi a siti per ottenere App gratuite, postare su social network, pubblicare su blog, mettere "mi piace" su Facebook anche su post di tipo sensibile (politico, religioso, etc.), postare foto/video, fornire la registrazione del proprio stato emotivo (es. "sono triste" su Facebook), consentire la registrazione del proprio stato personale fisico (es. peso, battito cardiaco, pressione sangue, ecc.), consentire la geolocalizzazione del proprio cellulare/smartphone (ad es: per registrare o farsi guidare in un percorso a piedi, in bicicletta o in auto, o per segnalare la propria posizione su

Facebook), usare sistemi di condivisione Cloud (es. Dropbox, Google Drive, etc.), usare App per car sharing, usare App per bike sharing, uso di consolle per videogame in modalità interattiva.

Tali attività sono numerose, in media 7 per intervistato, con un 3% che dichiara fino a tre attività e un 5% che afferma di compiere dalle 12 alle 18 delle attività proposte. Oltre alle attività svolte in Rete il questionario ha indagato il numero di oggetti tecnologici posseduti (smartphone, tablet, computer fisso, computer portatile, smartwatch, elettrodomestico che si può comandare via Internet) e anche in questo caso si rileva una capillare diffusione di tali strumenti, con una media di 2,8 strumenti posseduti a persona. L'utilizzo di strumenti sempre più sofisticati come smartphone o tablet diminuisce drasticamente la necessità di avere più strumenti tecnologici in quanto singolarmente in grado di adempiere alla maggioranza delle funzioni normalmente utilizzate con singoli strumenti. I dati raccolti in merito alle attività svolte in Rete e al possesso degli oggetti tecnologici hanno permesso di svolgere delle operazioni statistiche che hanno portato a delineare quattro tipi di studenti: *Quiet* (basso numero oggetti-basso numero attività, 24,8%), *Effective* (basso numero oggetti-alto numero attività, 16,2%), *Collector* (alto numero oggetti-basso numero attività, 23,1%), *Smart* (alto numero oggetti-alto numero attività, 35,8%)(Tabella 5). La tipologia ottenuta indica come la maggioranza degli studenti universitari si caratterizzino per avere sia un alto numero di oggetti tecnologici sia per svolgere un elevato numero di azioni in Rete.

Tabella 5. Profilo dell'intervistato definito mediante una tipologia basata sul numero di oggetti tecnologici posseduti e le attività in Rete svolte.

Tipi	Frequenza	Percentuale
<i>Quiet</i> (basso n. oggetti - basso n. attività)	271	24,8
<i>Effective</i> (basso n. oggetti - alto n. attività)	177	16,2
<i>Collector</i> (alto n. oggetti - basso n. attività)	252	23,1
<i>Smart</i> (alto n. oggetti - alto n. attività)	391	35,8
Totale	1091	100

La Tabella 6 consente di comparare le principali variabili-indice utilizzate in questo contributo, in base al profilo dell'intervistato. Come si può osservare – prendendo come riferimento i valori OUT – i valori più elevati si registrano in

corrispondenza del tipo *Effective* e del tipo *Smart*, caratterizzati dall'aver un alto numero di attività in rete. Lo stesso risultato lo vediamo con i valori IN. Questi risultati inducono a pensare che chi svolge molta attività in Rete ha una maggiore consapevolezza del valore dei propri dati, mentre ciò non si riscontra direttamente dal numero di oggetti tecnologici posseduti. In sintesi è simile ed elevata la consapevolezza del valore dei propri dati del tipo *Effective* e del tipo *Smart*, mentre è simile e più bassa la consapevolezza di *Quiet* e *Collector*.

Tabella 6. Comparazione delle medie delle principali variabili cardinali per profilo dall'intervistato.

Tipi		Valore complessivo in euro delle proprie preferenze nel caso vengano cedute (OUT)	Valore complessivo in euro delle proprie preferenze nel caso vengano restituiti servizi (IN)	Differenza tra valutazione dei dati ceduti e servizi ricevuti (OUT-IN)	N attività in Rete	N oggetti tecnologici posseduti	Età (anni)
<i>Quiet</i> (basso n. oggetti - basso n. attività)	Media	910	358	547	4,9	1,8	23,8
	N	214	217	200	271	271	258
	Dev. std.	5013	884	5097	1,8	0,4	5,4
<i>Effective</i> (basso n. oggetti - alto n. attività)	Media	1451	916	550	9,4	1,9	23,1
	N	137	146	129	177	177	171
	Dev. std.	7524	3357	7232	1,4	0,2	3,9
<i>Collector</i> (alto n. oggetti - basso n. attività)	Media	884	640	340	5,7	3,5	22,6
	N	210	216	202	252	252	242
	Dev. std.	3411	2995	1612	1,4	0,6	4
<i>Smart</i> (alto n. oggetti - alto n. attività)	Media	1237	836	110	10,1	3,6	22,3
	N	322	329	309	391	391	376
	Dev. std.	3976	4815	5440	2	0,8	3,1
Totale	Media	1107	688	337	7,7	2,9	22,9
	N	883	908	840	1091	1091	1047

	Dev. std.	4833	3541	5066	2,9	1	4,2
--	-----------	------	------	------	-----	---	-----

Le micro informazioni che lasciamo nella Rete sono un vero e proprio DNA digitale che consente di ricostruire un profilo della persona che frequenta lo spazio digitale. Informazioni che fanno gola ovviamente a esperti di marketing, psicologi, antropologi, sociologi e politologi, che se non sapranno orientarsi in queste nuove frontiere delle scienze sociali rischiano di essere respinti ai confini di questi nuovi scenari. Nel convegno AIS di Verona 2016 il professor Giorgio Alleva, presidente dell'Istat, ha ricordato ai sociologi presenti che alla ricerca di personale esperto in campionamento si è ormai sostituita la richiesta di esperti in Big Data in quanto le nuove indagini si stanno orientando in questa direzione.

Si tenga conto che parallelamente al Web esiste il Deep Web, profondo e oscuro, non accessibile ai normali motori di ricerca e che si stima essere 500 volte quello in chiaro [Casi, 2016], ma potrebbe essere addirittura 7.000 volte tanto. In queste memorie nascoste sono custoditi gran parte dei Big Data, un territorio in gran parte inesplorato che può essere attraversato da quanti abbiano competenze tecnologiche sufficienti per violarlo. La scarsa consapevolezza in tale ambito è dunque un rischio da non sottovalutare; occorre che soprattutto le nuove generazioni siano consapevoli che da grandi quantità di loro micro-dati è possibile estrarre informazioni con tecniche di data mining [Cannavò, Frudà, 2007], conoscenze e previsioni che riguardano le loro abitudini e comportamenti anche futuri. Una scarsa consapevolezza del valore delle informazioni che si depositano più o meno consapevolmente in Rete possono infatti costituire un terreno fertile su cui si sviluppano pratiche di *cyberbullismo*. Il problema va quindi letto attentamente e affrontato a diversi livelli (politico, culturale, scolastico) investendo le risorse necessarie allo scopo. Helen Nissenbaun e Solon Barocas [2014] sottolineano come i Big Data rappresentino non solo una evoluzione tecnologica ma un vero e proprio paradigma di pensiero, un modo di pensare alla conoscenza attraverso i dati per poter prendere decisioni e razionalizzare l'azione. In questa prospettiva evidenziano le implicazioni etiche e sociali dei Big Data e la necessità di riflettere sul concetto di privacy, proponendone una definizione più ampia rispetto alle semplici accezioni di anonimità e consenso informato. Diventa quindi indispensabile formare a un uso critico degli strumenti tecnologici, in modo da sviluppare una consapevolezza informata in merito al valore dei propri dati.

Bibliografia

1. Cannavò L., Frudà L. (a cura di), *Ricerca sociale. Dall'analisi esplorativa al data mining*, Carocci, Roma (2007).
2. Casì F., *Le potenzialità dei Big Data nella mutazione del paradigma culturale contemporaneo*, Corso di laurea magistrale in Filosofia, Dipartimento di Filosofia e Scienze dell'Educazione, Università di Torino, tesi di laurea (relatore prof. Maurizio Mori)(2016).
3. Grimaldi R., Cavagnero S.M., Gallina M.A., *Usa della rete e consapevolezza delle tracce digitali. Una ricerca tra studenti universitari*, in «Sociologia e Ricerca Sociale», XXXVII, n. 109, pp. 204-218 (2016).
4. Grimaldi R., Cavagnero S.M., Gallina M.A., *Big data e tracce digitali. Una ricerca internazionale tra studenti universitari*, in Trincherò R., Parola A. (a cura di), *Educare ai processi e ai linguaggi dell'apprendimento*, FrancoAngeli, Milano (2017).
5. Giannotti F., *Big Data e social mining: i dati, a saperli ascoltare, raccontano storie*, in E. Tasso, A. Mola, A. Cortesi, A. Candiello (a cura di), *Misurare l'innovazione digitale. Gli indicatori di successo delle politiche di innovazione territoriale*, Edizioni Ca' Foscari, Firenze (2015).
6. Kitchin R., *Big Data, new epistemologies and paradigm shifts*, «Big Data & Society», 1(April-June), 1–12 (2014).
7. Kshetri N., *The emerging role of Big Data in key development issues: Opportunities, challenges, and concerns*, «Big Data & Society», 1(2), (2014).
8. Mosca G., *Big Data: tutto quello che c'è da sapere sulla professione del Data Scientist*, Il Sole 24Ore (2014), <http://www.ilsole24ore.com/art/tecnologie/2014-10-27/big-data-tutto-quello-che-c-e-sapere-professione-data-scientist--144450.shtml?uid=ABgFVE7B>, ultimo accesso 2017/03/31.
9. Nissenbaum H., Barocas S. (2014), *Big Data's end run around anonymity and Consent*, in Lane J., Stodden V., Bender S., Nissenbaum H., *Privacy, Big Data and the Public Good. Frameworks for engagement*, Cambridge Books Online, <https://www.nyu.edu/projects/nissenbaum/papers/BigDataEndRun.pdf>.