

*Giuseppe Roccasalva

I big data e gli strumenti di visualizzazione analitica: interazioni e studi induttivi per le P.A.

Parole chiave: *big data, visualization analytic tools, story-telling, we-gov, inductive learning, business analytics.*

Abstract Il saggio presenta alcuni risultati di una collaborazione tra Politecnico di Torino e il CSI Piemonte (Società di servizi Informatizzati partecipata dalla Regione Piemonte). Sono stati selezionati e studiati diversi strumenti di visualizzazione dei dati scientifici (Gapminder, ManyEyes, Open eXplorer e Fineo) al fine di individuare quello più utile per una lettura induttiva di grandi quantità di dati informativi (*big data*). Lo sfruttamento intelligente dei dati digitali può portare a uno sviluppo conoscitivo ma anche a un profitto, le cui soglie di sfruttamento possono essere misurate in un sistema economico. Nell'irreversibile fenomeno di crescita dei dati digitali, la disciplina del "Data Visualization" diventa cruciale per accedere e comprendere informazioni complesse. Few, guru della comunicazione visiva, scrive che "scopriamo il mondo attraverso gli occhi"; le forme di comunicazione e interpretazione tradizionali dei dati hanno puntato sulla dimensione visuale per migliorare la comprensione e hanno permesso sia agli analisti sia agli utenti la sperimentazione di nuove interazioni ("story-telling").

Come urbanisti e cittadini, ci affidiamo alla vista che gestisce molti dei sensori (70%) legati alla percezione, alle mappe cognitive, agli errori e ai nuovi pensieri. L'ipotesi di fondo di questo articolo vuole generare delle riflessioni sui *Big Data* come strategia importante per le imprese pubbliche e private che intendono imparare a cambiare dalle informazioni digitali di cui oggi disponiamo. Attraverso l'uso di uno strumento analitico di visualizzazione dei dati informativi, si descrive un recente caso di studio in un contesto territoriale come quello dei nuovi consorzi amministrativi (Unione dei Comuni NordEst Torino). In questo esperimento torna a essere attuale la necessità di pianificare le scelte in modo sistematico anche cercando di utilizzare in modo nuovo e semplice i sistemi informativi territoriali già disponibili.

INTRODUZIONE

L'urbanistica, come molte altre discipline che hanno come sfera d'interesse il Territorio, non può più trascurare l'enorme quantità d'informazioni che sono generate anche in seguito al crescente utilizzo del *web*, dei *software* liberi e dei *social network*. Siano dibattiti *on-line*, *blog*, elenchi pubblici e classifiche, *database* settoriali o indagini statistiche, le informazioni generate e disponibili ci raggiungono consapevolmente e/o involontariamente attraverso una moltitudine di canali, generando così l'esigenza, in ogni ambito della conoscenza, di letture informative semplici e visualizzazioni coerenti. Questo articolo comincia definendo le principali caratteristiche dei cosiddetti "big data" e le opinioni autorevoli sul loro significativo sfruttamento nei settori pubblici e privati, grazie alle crescenti possibilità date dagli strumenti di visualizzazione analitica.

In seguito si propone una possibile curva di utilità focalizzata sul tipo di domanda d'informazioni (utenza) al fine di orientare opportunamente le scelte d'utilizzo degli strumenti di visualizzazione tra il settore pubblico e quello privato. Infine è descritto uno strumento ed il suo ruolo di interazione tramite una ricerca sperimentale svolta in un consorzio di sei comuni del nordest torinese.

I BIG DATA, IL LORO VALORE¹

I cosiddetti "big data" (BD) rappresentano il "buzzword" con cui si definisce oggi la mole sempre crescente di informazioni digitali che sono accumulate e che vanno solitamente da qualche *terabyte* alle decine di *petabyte* di dati racchiusi in un *dataset*. Per dare una dimensione significativa e intuibile, se otto *bit* (B) rappresentano la dimensione base dell'informatica (necessaria per creare una lettera o un numero) e 1 KB (*Kilobyte*) rappresenta una pagina di testo, un *megabyte* rappresenta una canzone e un *gigabyte* un film. Fin qui sono dimensioni comuni alla maggior parte degli utilizzatori del digitale mentre un *terabyte* o addirittura un *petabyte* sono misure che possono essere paragonate rispettivamente a tutti i testi posseduti da una grande biblioteca e le *mail* processate dal provider Google ogni ora.

Il crescere dei dati informativi mondiali immagazzinati cresce a elevata velocità (100% ogni due anni). Il McKinsey Global Institute riporta che più dell'80% dei settori economici americani sono costituiti da aziende che possiedono in media più dati della Biblioteca del Congresso. La gran parte dei dati sono rappresentativi o rapportabili alle città e ai suoi cittadini per cui offrirebbero numerose possibilità di approfondire le complesse dimensioni del territorio. Come scrive Batty², "Our need to understand how all these dimensions are coalescing, merging, complementing, and substituting for one another has never been more urgent. It constitutes a major challenge for planning and design in the near future" (*la nostra esigenza di capire come tutte queste dimensioni si stanno fondendo, integrando, completando e avvicinando l'una con l'altra non è mai stata così urgente. Nel prossimo futuro, costituirà una sfida importante per la pianificazione e per la progettazione*). Inoltre, a molti dati sono anche associate le relative dimensioni temporali, stimolando nei moderni analisti urbani una particolare propensione per la visualizzazione delle reti e dei flussi.

Nel 2020 si stima che circa trentacinque *zettabyte* di dati saranno immagazzinati tra dischi di massa e *server-web*. Una fonte che può efficacemente essere considerata al pari di altre risorse che più comunemente riteniamo tali (energia, materie prime, petrolio etc...).

¹ McKinsey Global Institute (MGI) reports.

² Mikael Batty, Editoriale 2012 di Environment and Planning.

Inoltre, negli ultimi anni lo sfruttamento di *database* ha stimolato lo sviluppo di applicazioni in grado di portare benefici tangibili alla collettività, risparmi per il funzionamento degli enti locali e nuove opportunità di *business* per il settore privato. I dati, infatti, rappresentando delle informazioni sono strettamente connessi alle dinamiche delle indagini di mercato, di ricerca, di sviluppo e se opportunamente letti, possono rivelare informazioni capaci di indicare dai trend alle abitudini individuali, dagli spostamenti urbani alle preferenze di consumo, dalle scelte urbanistiche ai prezzi di vendita degli immobili. Sono in grado di creare un valore significativo per l'economia mondiale, migliorando la produttività e la competitività delle imprese e del settore pubblico, creando un sostanziale *surplus* economico per i consumatori.

Il rapporto del McKinsey Global Institute mostra che i BD potrebbero far risparmiare solo alle Pubbliche Amministrazioni dei paesi dell'Ocse, tra 150 e 300 miliardi di euro, facendo leva su gestione ottimizzata dei dati elementari dei cittadini, sistemi di decisione automatici, identificazione automatica di frodi o errori e trasparenza della performance operativa. Si stima che la produttività del settore pubblico potrebbe aumentare del 20%, non solo per effetto della riduzione dei costi, ma anche per l'accresciuta qualità dei servizi offerti ai cittadini. L'assistenza sanitaria statunitense migliorerebbe qualità ed efficienza con un risparmio annuo di 300 miliardi di dollari riducendo la spesa sanitaria nazionale dell'8% e nelle aziende private si potrebbero incrementare i margini fino al 60% rispetto ai valori attuali.

ALCUNE DIMENSIONI LIMITE

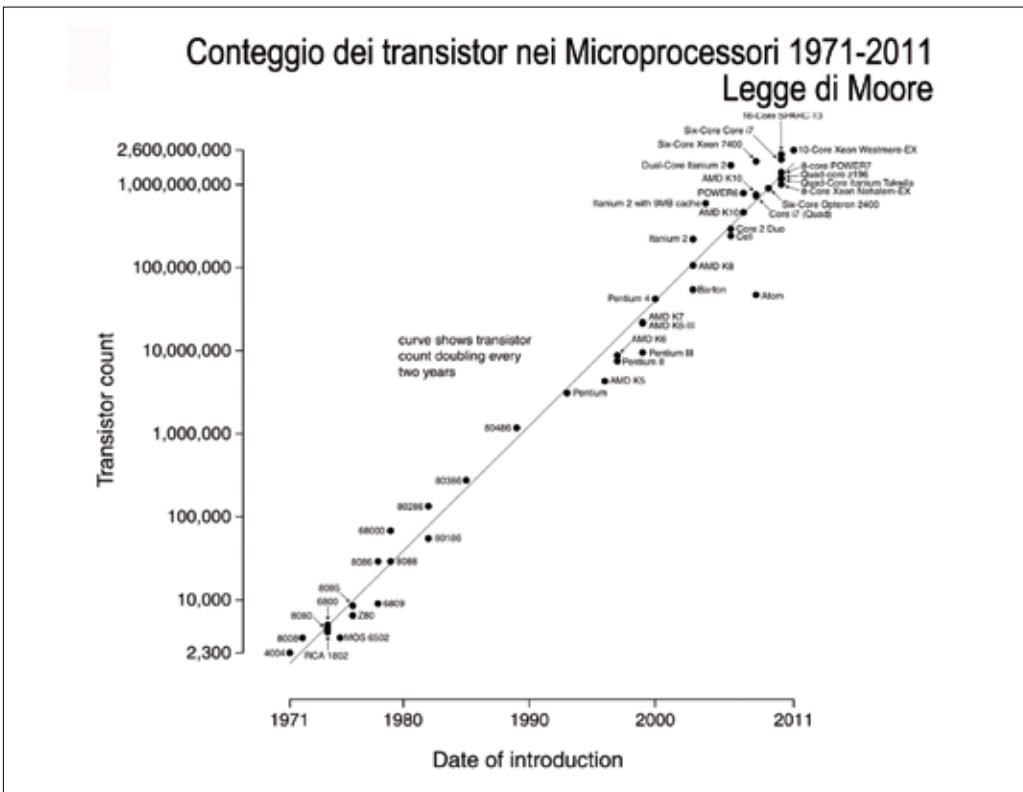
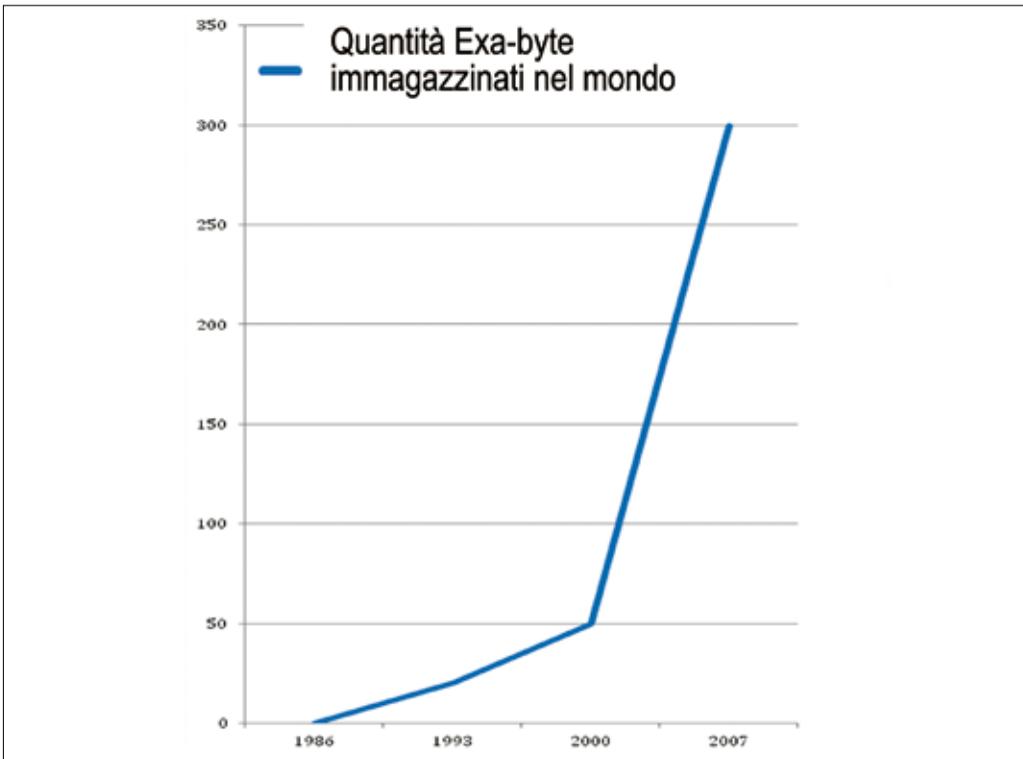
Un'affermazione comune agli studi inter-disciplinari sull'uso dei BD, riguarda il fatto che la crescita dei dati informativi non diminuisce da anni. Numerosi gruppi di ricerca hanno studiato la quantità totale di dati generati, immagazzinati e consumati nel mondo, indicando negli anni futuri una tendenza di crescita esponenziale (cfr. Figura 1). Nel 2010 sono stati memorizzati sulle unità disco aziendali circa sette *exabyte* di dati, mentre i consumatori ne tengono in memoria sui propri PC solo sei *exabyte*³. Le ragioni di questa esplosione sono legate ai costi della tecnologia che abbassandosi ha allargato l'utenza di consumatori ma anche allo sviluppo di strumenti di dialogo digitale (*blog, wiki, social network*) e allo sviluppo aperto e collaborativo di *software* e applicativi gratuiti. Nel 2008 le famiglie americane erano raggiunte-producevano giornalmente 3,6 *zettabyte* di dati informativi (34 *gigabyte* per persona al giorno); in questo quadro le parole scritte ammontano solo allo 0,1% del totale di *byte*⁴. Se ci si domanda quale limite ha la crescita dei BD, alcune semplici considerazioni possono aiutarci a capire. Nell'assioma noto come la Legge di Moore⁵ si notò che da quando erano stati inventati i circuiti integrati, il numero di *transistor* presenti su un unico *chip* era raddoppiato di anno in anno seguendo una curva esponenziale di crescita. Se si considera la curva di Moore, si apprezza una certa armonia con l'andamento della crescita del dato immagazzinato ogni anno nel mondo (cfr. Figura 1)

³ Fonte: McKinsey Global Institute (MGI) reports.

⁴ Fonte : Università della California a San Diego in *The Economist*.

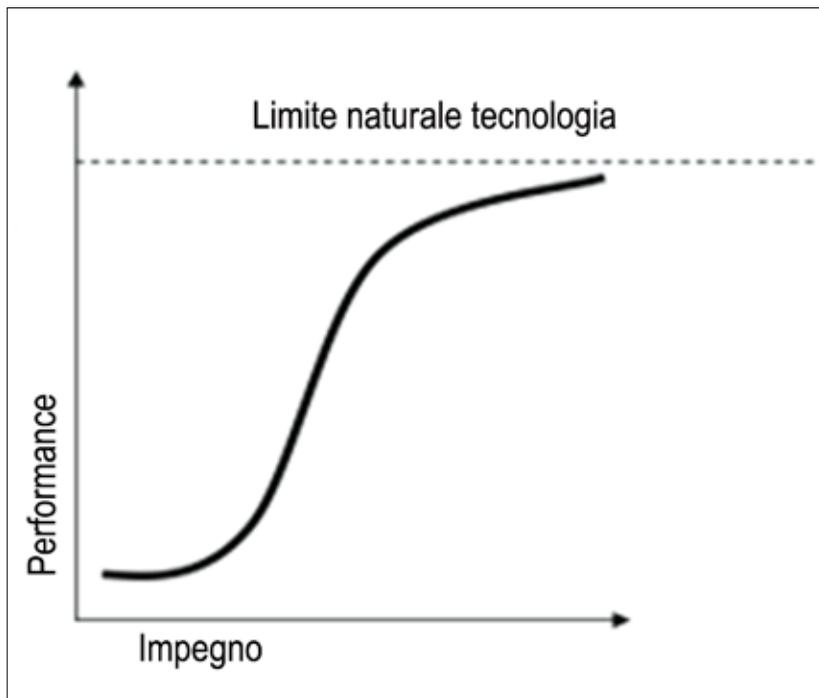
⁵ Gordon Moore, 1965.

Figura 1 Curva di crescita dei dati informativi (fonte MGI) e Legge di Moore



Nel ragionare sul futuro dei BD, si ritiene che la quantità di dati digitali possa avere un andamento simile alla curva tecnologica a S che descrive il limite generico che ha una tecnologia rispetto all'aumentare del suo impiego (cfr. Figura 2). Secondo Schilling “Nella fase iniziale, il miglioramento della *performance* è lento perché i principi base della tecnologia sono stati compresi in modo ancora parziale. Quando però i ricercatori e l'organizzazione nel suo complesso hanno acquisito una conoscenza più approfondita della tecnologia, il miglioramento comincia a essere più rapido. A un certo punto, però, il rendimento delle risorse e delle energie impegnate per lo sviluppo della tecnologia comincia a decrescere. E la tecnologia si avvicina al proprio limite naturale”⁶.

Figura 2 Curva tecnologica a S.



LA DEFINIZIONE DI “PROFITABILITÀ”

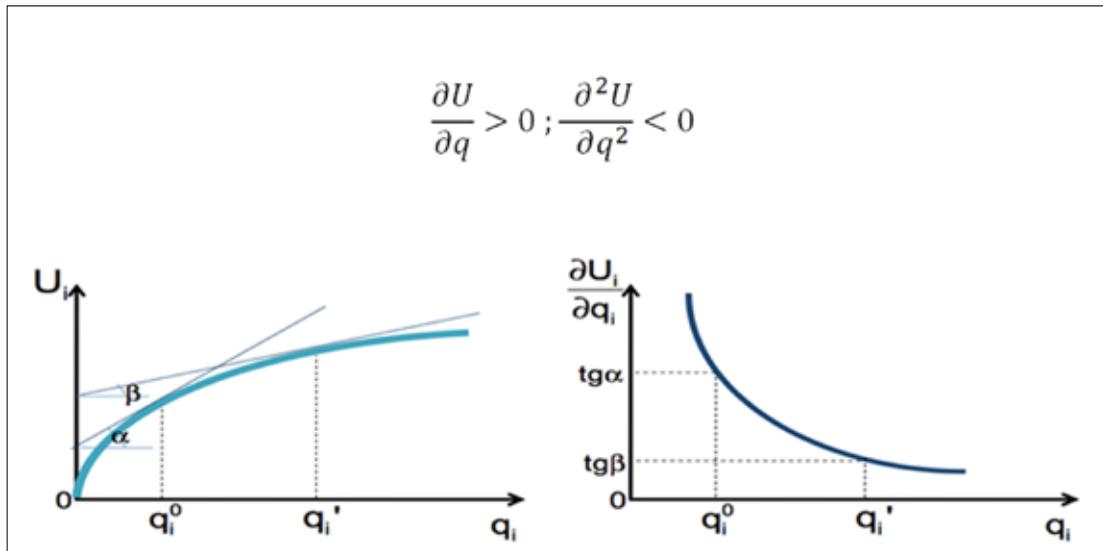
Riferendoci al pensiero degli economisti di scuola neoclassica e keynesiana⁷, in un sistema economico in cui c'è l'interazione di operatori funzionali (famiglia, impresa, pubblica amministrazione etc...), è possibile considerare la domanda in modo prioritario rispetto alla produzione di beni e servizi?

Ci siamo interrogati su come sfruttare i cosiddetti PSI (*public sector information*) come una risorsa economica immateriale dalle grandi potenzialità. Si è cercato di abbozzare il concetto di “profitabilità” dei dati, ossia in che misura e in quale forma essi possano creare valore economico. Ci siamo avvalsi della teoria microeconomica per rappresentare l'utilità di un determinato bene *i*-esimo (in questo caso i dati digitali) attraverso una funzione crescente, in cui ogni unità addizionale consumata aumenta il grado di soddisfazione-utilità *U* ma a rendimenti decrescenti.

⁶ Schilling M. A. (2009), *Gestione dell'innovazione*.

⁷ Ci si riferisce a un approccio alla disciplina economica che non è basato sulla determinazione dei prezzi, produzione e reddito attraverso il modello di Domanda e Offerta. L'economia Keynesiana secondo Bruno Jossa, ha spostato l'attenzione dell'economia dalla produzione di beni alla domanda, osservando come in talune circostanze la domanda aggregata è insufficiente a garantire la piena occupazione. Di qui la necessità di un intervento pubblico.

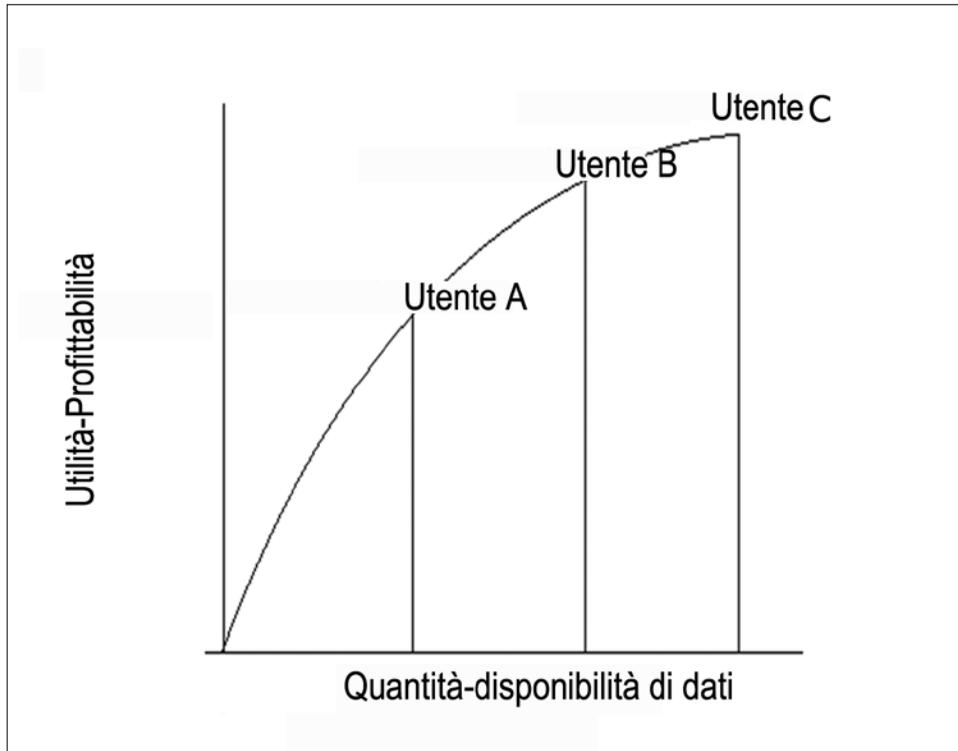
Figura 3 Fonte: liberamente adattato da Urban Economy di O'Sullivan A.



La suddivisione in due prospettive differenti, una economica-imprenditoriale l'altra economico-sociale, ci ha permesso di capire in termini generalizzati come e con quale intensità i BD possano interagire in sistemi di riferimento differente per obiettivi, contesti e soggetti promotori. E' necessario per ultimo definire l'utente come un individuo che ha la possibilità di accedere ai dati digitali e abbia dei livelli accettabili di refrattarietà all'innovazione (*digital divide experience*). A tal proposito sono state ipotizzate delle categorizzazioni rispetto a delle griglie di valutazione specifiche, la cui descrizione si è scelto di tralasciare per questo saggio.

La prima prospettiva, rifacendosi al grafico, mostra l'andamento dei dati digitali rispetto alla loro utilità-performance da un lato economico-imprenditoriale. In questo caso possiamo categorizzare tre tipi di utenti-consumatori, l'utente A, B e C. Essi possiedono un differente comportamento nell'interazione con i dati digitali suddiviso in:

- (0;A) → utente/consumatore che utilizza i dati per comunicare, es. *chatting, twitting*; un tipo di utente che conosce vari strumenti per scambiare dati e informazioni digitali ai fini della comunicazione (bassa profittabilità);
- (0;B) → utente/consumatore che utilizza i dati per conoscere es. *web-searching*; un tipo di utente che riesce ad apprendere attraverso la ricerca di dati e farne delle semplici analisi (media profittabilità);
- (B;C) → utente/consumatore che utilizza i dati per progettare *business*. es. programmatore di *app, data-scientist*, imprenditore od operatore pubblico esperto; un tipo di utente che è in grado di capire in maniera profonda i dati digitali creandone della ricchezza economica (alta profittabilità).

Figura 4 Curva di utilità-profitabilità rispetto alla disponibilità-qualità di dati informativi

La seconda prospettiva è descritta nel secondo grafico a rendimenti decrescenti e ci piace definirla economico-sociale. La curva esemplifica l'utilità rispetto al costante aumento di dati digitali e, se calibrata, può servire per prefigurare l'uso dei BD a partire dalla domanda – sia essa prodotta da consumatori dei servizi pubblici o da utenti privati. A tal fine, sono stati prefigurati scenari alternativi di utilizzo di BD pubblici per cogliere nelle curve le criticità e opportunità per avviare un adeguato sfruttamento dei dati. A titolo di esempio si può rilevare che il crescere dei dati potrebbe:

- creare maggiore confusione nel distinguere i dati di valore, infatti, un set di dati troppo grande complica di molto la selezione, il loro apprendimento e diffusione tanto da non diventare utile per la società e addirittura cominciare a diventare un peso, uno scarto da contenere o da trasferire per far posto alle esigenze di nuova informazione e comunicazione;
- recare problemi alla comunità rispetto a una possibile violazione dei consolidati livelli di “privacy”;
- rendere inefficaci le attuali strategie che si appoggiano sull'utilizzo delle informazioni digitali imputate o raccolte.

Inoltre, se guardiamo i due grafici insieme, si può notare la sintonia che viene anche descritta dalla cosiddetta legge di Moore. Quest'ultima, infatti, definisce la velocità con cui cresce la capacità tecnologica ma ricorda anche che quest'ultima ha un limite naturale mentre, in teoria, la crescita dei dati non ha limite.

LA VISUALIZZAZIONE E L'APPRENDIMENTO

Intorno al fenomeno BD sono nati gruppi di ricerca e di studio strettamente legati al ramo del Marketing e Business Intelligence; tuttavia, questa ricerca si interroga sulle sinergie con le attività analitiche legate alla Pianificazione. “*Scopriamo il mondo attraverso i nostri occhi*” scrive Few⁸, uno dei personaggi più noti nel campo della Visual Communication e del Data Visualization. La comunicazione visiva di tipo scientifico si è sviluppata divenendo sempre più sofisticata. Ad esempio, si sono moltiplicate le forme di rappresentazione per i grafici, che sono solitamente più adatti a comprendere *trend* ed eccezioni, a differenza delle tabelle che rivelano solo liste di valori specifici. I grafici riescono anche a creare relazioni tra valori, dando loro dimensioni, forma, colore e oggi anche dinamicità. Tutto ciò influisce sul modo in cui le persone percepiscono e assimilano le informazioni. Tra i ricettori presenti nel corpo umano, il 70% risiede negli occhi. La percezione visiva porta le informazioni al nostro cervello in maniera immediata. Colin Ware⁹ spiega che “L’occhio e la corteccia visiva del cervello formano un processore in parallelo... la percezione e la cognizione sono così strettamente interrelate che parole come ‘capire’ e ‘vedere’ diventano sinonimi”. I modi con cui si presentano dei dati possono renderli chiari o anche invisibili. Capire come funziona la percezione e strutturare adeguatamente la visualizzazione d’informazioni può quindi permetterci di rafforzare la comunicazione e l’apprendimento. “Data Communication and Visualization”, ovvero lo studio della percezione per rappresentare informazioni sta diventando sempre più apprezzato per i benefici che porta a tutti i settori economici. La visualizzazione dei dati, negli ultimi anni, è diventata oggetto di studio e di ricerca in tutte le discipline scientifiche poiché tutte utilizzano e studiano informazioni e dati. Tra gli apporti di maggior rilievo, si rileva la capacità analitica “aumentata” che il Data Visualization permette. I *software* di analisi grafica ci permettono di non rappresentare i dati solo graficamente, ma anche di interagire con queste forme di visualizzazione, cambiando la natura della loro visualizzazione, ci permettono inoltre di entrare in livelli di dettaglio più bassi e specifici, filtrare ciò che non è rilevante, evidenziare sottoinsiemi di dati attraverso più grafici simultanei. Mentre inizialmente i geografi si focalizzavano sulla descrizione dei processi complessi, oggi la visualizzazione è molto focalizzata sull’analisi e induzione alla comprensione (*inductive learning*¹⁰) dei sistemi, offrendo un più sofisticato strumento che in tempo reale può permettere l’esplorazione dei *pattern* attuali e del passato (es. dati del traffico in *real time* permettono di tracciare schemi di congestione, preferenze di rotta durante i giorni lavorativi, etc...).

TECNICHE PIÙ DIFFUSE DI VISUALIZZAZIONE ANALITICA DEI DATI

Nel lavoro di ricerca in collaborazione con il CSI Piemonte, sono stati valutati diversi strumenti e tecniche di visualizzazione dei dati complessi al fine di cominciare a scrivere un libro bianco per lo sfruttamento dei dati informativi locali. In particolare ricordiamo tre strumenti:

- Google public data Explorer

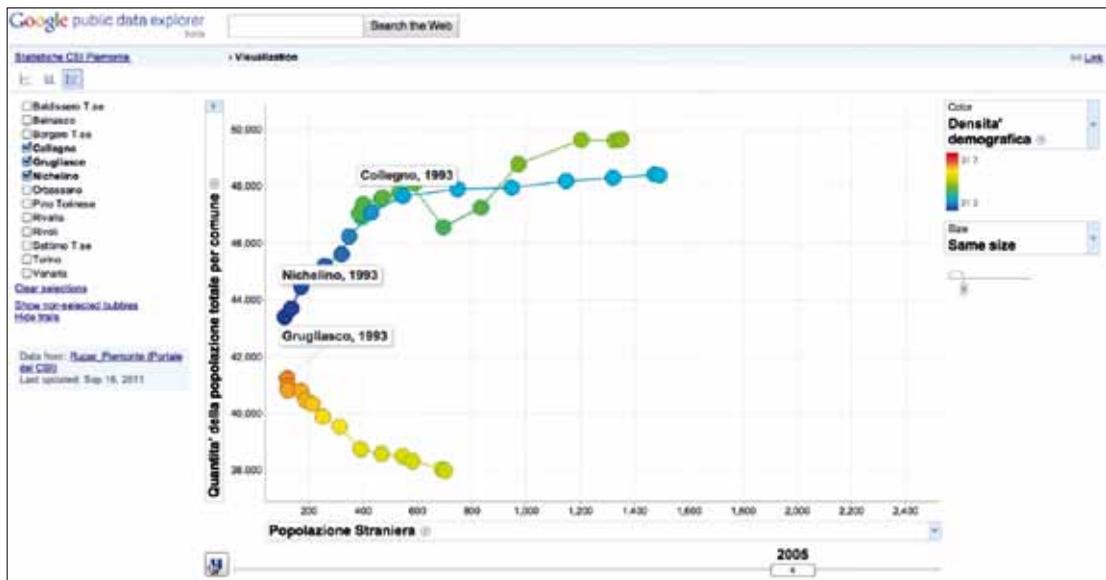
E’ uno strumento per la visualizzazione di grandi *dataset*. Possiede tre viste: un grafico a linee, un istogramma animato e un diagramma animato a bolle. L’interfaccia con cui interagisce l’utente permette di selezionare uno degli indicatori del *database* (programmabile solo con conoscenze di DSPL Dataset Publishing Language) e scegliere uno tra i grafici disponibili (grafico a linee, istogramma animato e diagramma animato a bolle). Le serie temporali dei dati permettono il confronto tra gli indicatori e i colori possono essere personalizzabili in base alle esigenze di analisi.

⁸ Autore di tre libri sulla visualizzazione dati, consulente presso importanti aziende come Apple, Microsoft, Siemens, Tableau Software ed eBay e lecturer presso Harvard, Stanford e Oxford.

⁹ Direttore del *Data Visualization Research Lab* all’università del New Hampshire.

¹⁰ Corsico, F. Roccasalva, G., (2005) *Visu-an-alyse indicators of urban quality. The crucial role of forecasting Scenarios in sustainable decision making processes*, International Conference for Integrating Knowledge and Practice, Life in The urban Landscape, Gothenburg.

Figura 5 Visualizzazione del tasso di popolazione straniera nei comuni dell'Area Metropolitana Torinese.



- ManyEyes

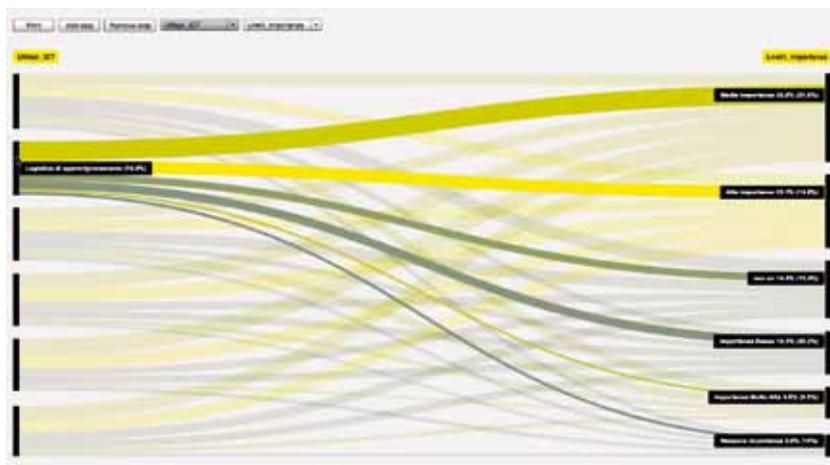
Si tratta di un visualizzatore che incrocia coppie di dati, programmato da IBM, e che permette di creare e condividere grafici di dati. La capacità di gestire grandi quantità di dati anche in tempo reale e la resa grafica di alta qualità, rendono questo strumento utile per gli esperimenti su grandi masse di dati anche non strutturate.

Oltre ai classici grafici permette un'interrogazione trasversale (*drill down*) della tassonomia. Anche in questo caso una terza dimensione è personalizzabile attraverso la scelta dei colori.

- Fineo

Si tratta di una applicazione *web* che visualizza flussi continui di dati (energia, denaro, traffico) al fine di rappresentare la relazione tra le dimensioni dei dati della tassonomia. Questa innovativa applicazione è stata programmata da Density Design, un centro di ricerca del Politecnico di Milano con il fine di portare la visualizzazione a nuovi stadi utili per la conoscenza. Le percentuali sono calibrate automaticamente in funzione del *dataset* immesso.

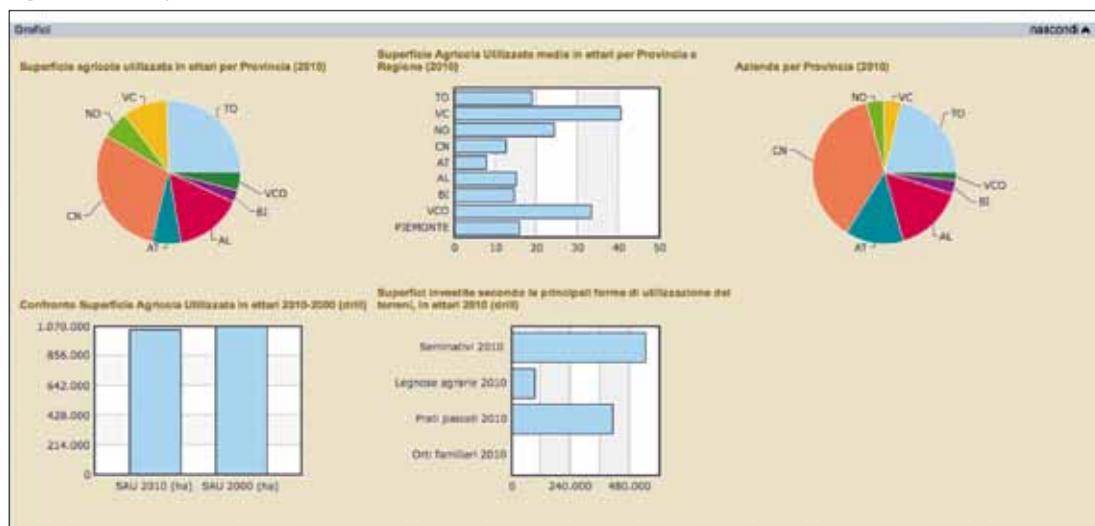
Figura 6 Visualizzazione dei flussi di utilizzo dell'ICT in Piemonte nei diversi settori economici.



Inoltre, tra le più ricorrenti tecniche di visualizzazione, ricordiamo il “Dashboard” (cruscotto). Questo tipo di comunicazione visiva combina le informazioni necessarie per monitorare un aspetto (nella sua complessità di dati) su un unico schermo. Se propriamente utilizzati, i cruscotti offrono una foto molto efficace di quello che accade nei processi di ogni genere, dando un livello di consapevolezza che non potrebbe essere facilmente raggiunta dai rapporti di analisi tradizionali.

Nei sistemi informativi di gestione, il *dashboard* è un sistema di visualizzazione progettato per essere facilmente comprensibile. I cruscotti consolidando le metriche e i *report* sui programmi, sui processi di *business* o sull'infrastruttura tecnologica, possono essere strumenti preziosi per *manager* e “decision-makers”.

Figura 7 Un esempio di *dashboard* utilizzato dal CSI Piemonte.



Il fallimento-successo di un *dashboard* si potrebbe sintetizzare in poche regole di base¹¹. I cruscotti non sono utili quando:

- lasciano l'interpretazione e visualizzano i dati in maniera troppo aggregata (indice anziché indicatore);
- presentano troppi dati da leggere allo stesso tempo;
- sono prodotti da analisti che leggono i dati ma che non vivono-lavorano nel contesto in cui sono prodotti;
- sono sviluppati esternamente, da persone che non hanno un'esperienza dell'azienda.

Al contrario, un buon *dashboard* deve:

- avere un grafico che mostri la tendenza della metrica;
- interpretare le tendenze e fornire un contesto al fenomeno studiato;
- far comprendere gli impatti per sollecitare le scelte;
- suggerire le azioni da intraprendere e i passi da seguire.

Un altro strumento di analisi interessante è l'OLAP (Online Analytical Processing)¹². Si tratta di un

¹¹ Tratto dagli studi dell'ingegnere Leonardo Bellini esperto in *web marketing and communication* e docente presso l'università della Sapienza di Roma.

¹² Fonte : "The OLAP report" o "The BI verdict", curato da Nigel Pendse.

sistema che raccoglie una serie di tecniche per l'analisi interattiva e veloce di grandi quantità di dati. Questo strumento è utile alle aziende per analizzare i risultati delle vendite o l'andamento dei costi di acquisto delle merci, per il *marketing*, per misurare il successo di una campagna pubblicitaria o per verificare i dati di un sondaggio. La creazione di un *database* OLAP consiste nell'eseguire una fotografia di informazioni (ad esempio quelle di un *database* relazionale) in un determinato momento e trasformare queste singole informazioni in dati multidimensionali. Per esempio un archivio di clienti che utilizza il servizio pubblico di raccolta rifiuti può essere raggruppato per città, provincia, regione; questi clienti possono essere relazionati con le esigenze di ogni area o quartiere ed ogni richiesta (anche quelle eccezionali) possono essere raggruppate per categoria. Il calcolo delle possibili combinazioni di queste aggregazioni forma una struttura OLAP che, potenzialmente, potrebbe contenere tutte le risposte per ogni singola combinazione lavorativa. Un sistema OLAP permette di studiare un'ingente quantità di dati, di vederli da prospettive diverse e aiuta i processi decisionali. In particolare si può:

- analizzare il totale di un *dataset* in conformità a dimensioni differenti e analizzare aggregazioni trasversali (*slicing*);
- focalizzare ed estrarre una sezione avente particolare interesse per l'analista (*dicing*);
- scomporre il *dataset* nelle sue determinanti all'interno della stessa gerarchia (*Drill-down*);
- scomporre il *dataset* nelle sue determinanti passando da una gerarchia ad un'altra (*Drill-across*).

Un'altra tecnica di comunicazione che ha catturato l'immaginazione di molti è la visualizzazione geo-spaziale. Oggi molte delle informazioni che le imprese devono monitorare e comprendere sono legate a località geografiche. Spesso le informazioni rispetto ai servizi, vendite o eventi possono essere comprese solo se si riesce a vedere dove effettivamente avvengono o dove esiste una domanda e quindi una utilità. Ad esempio, la capacità di vedere l'andamento dei ricavi di vendita su mappa, aggiunge una dimensione di interpretazione "critica".

Infine, il miglior espediente informatico che si sta sviluppando oggi in molti tra gli strumenti analizzati è l'utilizzo di animazioni (il movimento di oggetti nei grafici) per mostrare i cambiamenti di una determinata variabile nel corso del tempo o rispetto a un'altra variabile. Alcuni dei migliori lavori su grafici dinamici sono stati sviluppati da Rosling e dalla sua fondazione, Gapminder. Nel 2006 è stato lanciato il *software* Trendalyzer, un tipo di applicativo che permette di creare grafici interattivi e dinamici e che va a cancellare la componente statica, predominante in tutti gli altri tipi di grafici. Nel Marzo 2007 Google ha acquistato Trendalyzer da Gapminder, riconoscendo il valore ed efficienza del lavoro dei ricercatori che hanno contribuito alla costruzione di tale *software*.

LO STRUMENTO SCELTO

Nello sperimentare le applicazioni sopradescritte, Open Explorer è senz'altro uno tra gli strumenti informatici più completi e sviluppati che racchiude il concetto di visualizzazione scientifica multilivello, i cui utilizzi collaudati sono molteplici e vanno dalla Medicina alla Pianificazione fino alla Fisica. Sviluppato alla Norrköping University (SE) dal *team* di Mikael Jern, questo strumento risponde alle esigenze d'induzione alla comprensione (*inductive learning*) che rappresentano il filo ordinatore degli applicativi descritti e analizzati in questo articolo. Inoltre, questo strumento aggiunge una semplificazione nell'implementazione e personalizzazione delle proprie analisi, grazie ad un'interfaccia *on-line* (*vislet*) gestita dal gruppo NcomVA (Norrköping communication and Visual analytics) che trasforma e rende subito utilizzabili i dati immessi. La validità degli studi di Jern è stata riconosciuta da premi internazionali e dagli autorevoli utenti che hanno utilizzato i suoi brevetti (ISTAT, EuroSTAT, WorldBank, Commissione Europea, Provincia di Kensington e Chelsea, The Economist).

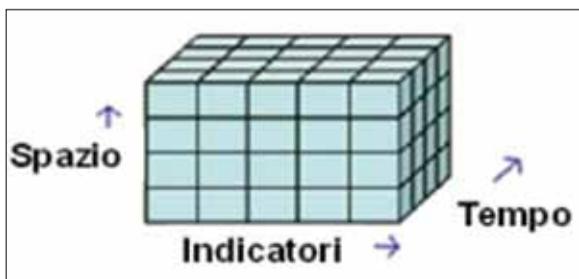
Open Explorer, è primariamente un visualizzatore geografico di dati che possono essere imputati o derivare da un PSI ed essere in formati tra i più comuni (Pc-Axis, eXplorer Unicode e SDMX).

I file di dati da importare devono essere realizzati utilizzando M. Excel e simili e devono essere opportunamente strutturati rispetto alla posizione dei valori e dei metadati necessari. Le animazioni presenti sono utilizzabili attraverso Adobe Flash Player. Il *layout* è personalizzabile e si ha la possibilità di avere un unico grafico per pagina o suddividere la pagina fino a tre viste collegate.

Da un punto di vista percettivo, l'interfaccia principale con cui l'utente interagisce è suddivisa in quattro *box* dentro ai quali vengono effettuate e visualizzate le analisi rispetto agli indicatori di base. L'interfaccia si compone della mappa tematica, il diagramma di regressione (*scatter plot*), l'istogramma (o grafico a coordinate parallele) e l'area dei percorsi di analisi (*content browser*) all'interno del quale c'è spazio per la cosiddetta *story-telling* (ovvero per le opinioni e idee degli utilizzatori a cui è stato fornito lo strumento).

Open Explorer si basa su GAV Flash, ovvero un modello concettuale di dati ottimizzato per la gestione spazio-temporale e multivariata di indicatori. Questo modello può essere rappresentato attraverso un cubo di dati con le seguenti tre dimensioni: spazio (città-territori), tempo (serie storica) e indicatori.

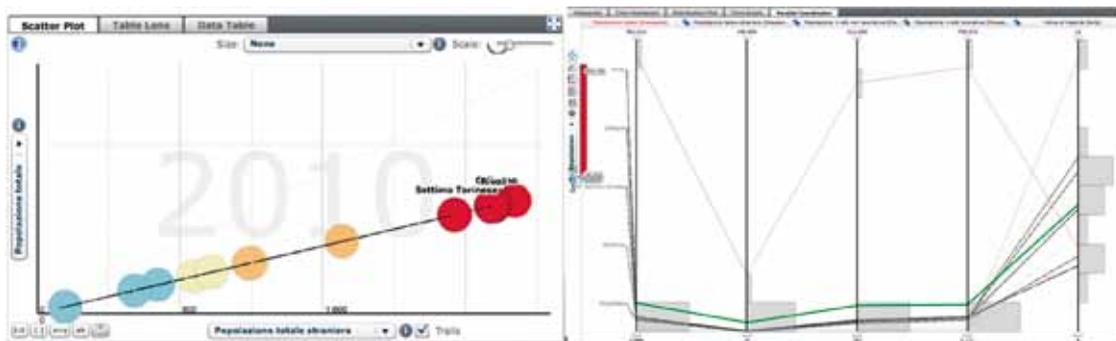
Figura 8 Modello concettuale di cubo dei dati applicato a Open eXplorer.



Open eXplorer consente quindi di visualizzare contemporaneamente i dati sia nella loro dimensione spaziale sia temporale con il supporto di più visualizzazioni collegate. La barra del “play” avvia la componente temporale ed è presente al di sotto delle quattro aree distinte e permette una visualizzazione dinamico-temporale degli indicatori selezionati in simultanea sulla finestra geografica, sull'istogramma e sul diagramma di regressione.

Gli strumenti di navigazione della mappa tematica sono anche strumenti di analisi. Lo Scatter Plot è un grafico che consente la visualizzazione contemporanea di quattro dimensioni sullo stesso territorio: una abbinata alla Mappa (colore della bolla), due sugli assi dello Scatter Plot stesso (asse X, asse Y) ed una relativa all'ampiezza della bolla.

Figura 9 Esempio di Scatter plot con le linee tracciate dall'andamento della serie storica; esempio di grafico a coordinate parallele (PCP).



Il Grafico a Coordinate Parallele (PCP) permette di visualizzare contemporaneamente, con linee e istogrammi, diverse dimensioni, anche alternative rispetto a quanto rappresentato su Mappa tematica e Scatter Plot, seguendone l'evoluzione nel tempo in modo dinamico e mostrando la media del fenomeno misurato dagli indicatori selezionati. Il PCP è una tecnica di geo-visualizzazione utilizzata per identificare *trend*, individuare gruppi (*cluster*) e valori anomali (*outliers*) a supporto di analisi di approfondimento sulle relazioni esistenti fra territori e fra indicatori.

È disponibile un'ampia personalizzazione della rappresentazione dei dati sulla Mappa tematica, che permette all'utente di realizzare efficaci visualizzazioni di dati spaziali e di confrontare storicamente e dinamicamente i dati informativi territoriali. E' possibile assegnare la scala cromatica desiderata all'indicatore da rappresentare sulla Mappa scegliendo tra le otto scale predefinite. La scala può anche essere regolata dinamicamente spostando con il cursore gli estremi degli intervalli oppure cliccando sulla singola etichetta e definendo quindi il valore desiderato.

E' evidente che questo strumento permetta molteplici visualizzazioni della stessa distribuzione di un dato, affidando, come scritto in precedenza, un ruolo critico alla percezione visiva che renderebbe la stessa sequenza di informazioni più o meno comprensibile in funzione della struttura scelta per la comunicazione.

IL CASO STUDIO

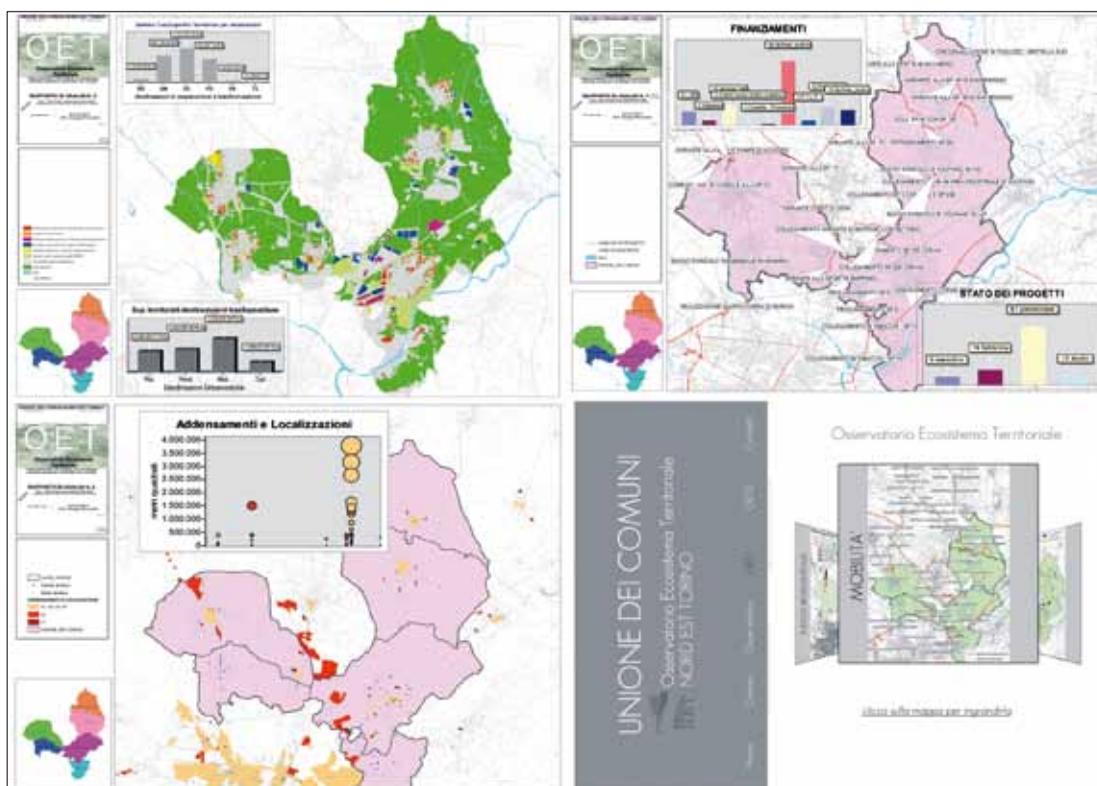
Al fine di avere un quadro delle esperienze applicative, sono stati studiati alcuni casi che hanno utilizzato metodi interattivi di visualizzazione come quelli fin qui descritti. In Italia, con questa ottica, grandi dimensioni di dati sono stati resi disponibili sulla base degli esperimenti prodotti dall'ISTAT in collaborazione con il DPS del Tesoro. Il progetto ha permesso di comunicare i principali dati statistici su base regionale per tutto il territorio italiano. Recentemente, a dimostrazione dell'interesse crescente sul tema, una società di consulenza a partecipazione pubblica, acquistando alcuni degli strumenti di visualizzazione sopra descritti, sta studiando la migliore struttura comunicativa di dati economici per supportare ed indirizzare in Italia i lavori del Ministero dello Sviluppo economico.

Di seguito verrà introdotto il caso studio in cui è stato applicato uno tra questi strumenti per un nascente ente di servizi pubblici aggregati: l'Unione dei Comuni NordEst Torino (Unione NET). Quest'ultimo rappresenta un consorzio di sei Amministrazioni dell'Area Metropolitana di Torino (circa 120 mila abitanti su un territorio di 140 Km²) che hanno siglato un protocollo di intenti sulla base della possibilità che la legislazione italiana suggerisce (Art. 32 del D. Lgs. 18/08/2000, n°267), per quegli enti che vogliono provvedere in modo collegiale e quindi più efficace alla gestione di servizi e scelte concernenti il loro mandato.

A marzo 2012, l'Unione NET, su iniziativa di un comune consorziato, ha incaricato uno studio sulle numerose e differenti fonti di dati già disponibili nei processi lavorativi dei loro uffici tecnici per ragionare sugli scenari comuni di programmazione territoriale. Lo studio, durato sei mesi, ha definito con gli enti dell'Unione gli obiettivi e cominciato a costruire alcune letture ed applicazioni. L'obiettivo principale che questo studio ha proposto è di costruire una base digitale per conoscere, comprendere e divulgare informazioni territorializzate e favorire, in modo semplice ed intuitivo, economie di scala in ambito sociale e pubblico. I sei comuni coinvolti dall'esperimento hanno un buona domanda di informazioni e comunicazioni internet avendo anche predisposto un *wifi* gratuito per la popolazione. La curva di domanda è pertanto potenzialmente ampia e variegata, comprendendo un generico canale informativo per i cittadini, uno più specifico per le attività pubbliche e uno ancora sottoutilizzato per le iniziative private. Tuttavia, gli uffici tecnici non possiedono conoscenze omogenee sull'uso dei dati né strutturati livelli di gestione e mantenimento delle informazioni geografiche. Inoltre, si è notato un diverso livello di quantità e qualità dei dati che è risultato difficile da omogeneizzare e pertanto si è preferito selezionare per questo esperimento dei dati meno dettagliati ma omogenei come

quelli predisposti da enti sovralocali. Il CSI Piemonte ha aderito alla proposta di ricerca fornendo i primi dati rispetto alla composizione demografica delle principali città dell'*hinterland* torinese. Il CSI Piemonte possiede e raccoglie un'enorme quantità di dati per conto degli Enti sovralocali principalmente (1300 *database* alfanumerici, 1400 *database* spaziali e 160 *database* a supporto delle Amministrazioni locali). Inizialmente, si è preferito strutturare delle mappe specifiche che contenessero già delle informazioni elaborate per poi scegliere quali dati semplici possono essere resi disponibili per le visualizzazioni e studiarne gli usi rispetto alla domanda che proviene dal canale delle attività pubbliche (attività autorizzative, di analisi e di verifica). Le prime elaborazioni sono state delle cosiddette "mappe esperte", prodotte su base GIS e che avevano l'obiettivo di fare ordine, "mosaicare" le informazioni geografiche per tutti e sei i comuni e renderle coerenti. Con l'utilizzo delle funzioni tipiche dello *spatial analysis*, sono stati aggregati alcuni indici che distinguessero le condizioni attuali rispetto alle scelte di trasformazione, per cercare significative informazioni che potessero indirizzare i dibattiti del Consiglio dei Comuni. I temi su cui ci si è concentrati sono stati le decisionalità urbanistiche consolidate e in atto, le infrastrutture ultimate e in progetto, la quantità delle aree commerciali attualmente disponibili e licenziate. In particolare (cfr. Figura 10), sul tema delle trasformazioni urbanistiche, sono stati distinti i terreni per destinazione d'uso e per grado di trasformabilità (consolidato, moderata trasformazione e nuovo impianto). Sul tema commerciale sono stati distinti gli addensamenti sia per superfici sia per categoria commerciale tra quelli già in essere e quelli in costruzione. Sul tema trasportistico sono state distinte la natura del finanziamento e lo stato di avanzamento dei progetti per le infrastrutture in trasformazione.

Figura 10 Elaborazioni analitico-geografiche (mappe esperte) di dati informativi sui tre temi inizialmente individuati per lo studio. Rispettivamente in senso orario, analisi sulle decisionalità urbanistiche (slp per destinazione), analisi sulle autorizzazioni commerciali (addensamenti, slp per categoria di commercio), analisi delle infrastrutture (stato dei progetti, tipologia di finanziamento). L'immagine in basso a destra è una pagina del sito internet di divulgazione della mosaicatura dei PRGC, del quadro delle infrastrutture e delle analisi sul commercio

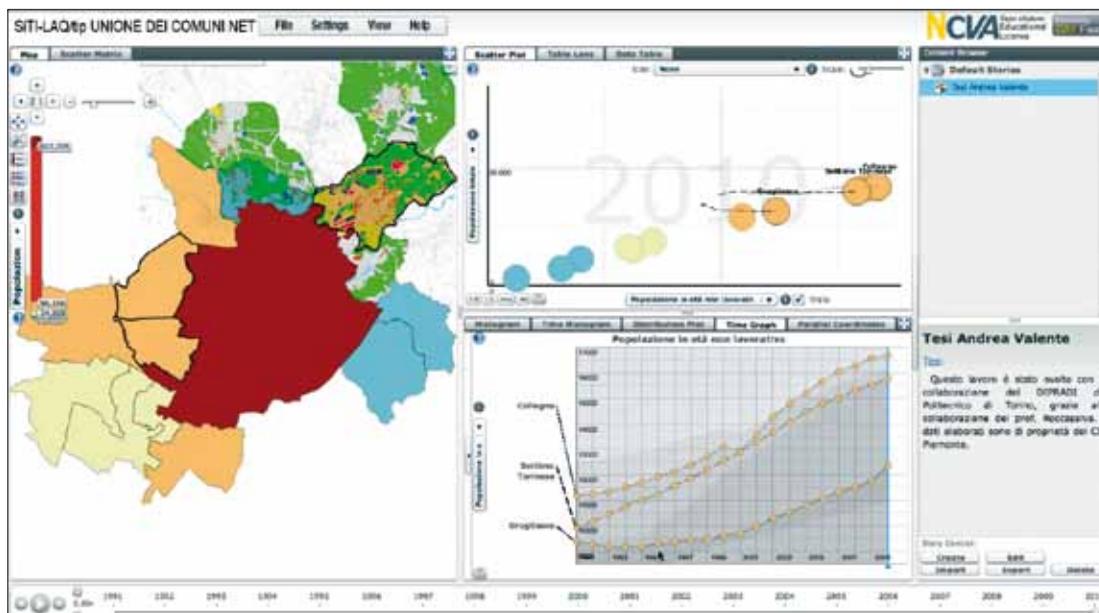


Al fine di studiare la quantità e qualità di informazioni da selezionare, sono state pubblicate le “mappe esperte” su un portale che permette la visualizzazione e divulgazione dei contenuti mantenendo alcune delle peculiarità delle visualizzazioni scientifiche (cfr. Figura 10).

Parallelamente, è stato fornito ai rappresentanti delle Amministrazioni una prima bozza di un sistema di visualizzazione di dati e delle informazioni emerse dalle visualizzazioni analitico-geografiche attraverso lo strumento Open Explorer (cfr. Figura 11). L'osservazione sull'uso dello strumento ha portato lo studio a fare scelte di semplificazione e aggiustamenti, azioni di supporto e a trarre alcune considerazioni sulle scelte future da compiere; di seguito si riporta una lista delle principali scelte adottate nel corso dello studio.

- È stato proposto un numero limitato di indicatori tra quelli aperti e disponibili poiché era difficile, in un primo approccio alla visualizzazione interattiva, poter gestire più di 7 categorie di dati. In questo caso i dati statistici, seppur noti, innescano agilmente il dialogo tra temi tecnici e questioni sociali, trasferendoli a livelli di conoscenza più ampi della sola base geografica. Il tasso di disoccupazione, la multietnicità o la scolarità di un territorio sono temi trasversali che permettono un insieme di ragionamenti, deduzioni ma anche semplicistiche interpretazioni quando vengono dettagliate su un comune o addirittura su di un quartiere.
- L'interattività e l'interrogabilità del dato è stata limitata ai soli dati statistici, lasciando le analisi e le “mappe esperte” come base della visualizzazione. Questo ha permesso di considerare la geografia elaborata in precedenza come una condizione data e spingere gli utenti alla verifica locale di eventuali criticità. Le mappe sull'infrastruttura sottolineano in rosso le strade non ancora costruite o migliorate, così come le informazioni sui nuovi supermercati e sulle nuove aree residenziali sono dati su cui il singolo utente non interagisce ma può eventualmente suggerire delle anomalie.
- Nel corso dell'osservazione sull'uso dello strumento, è stato prodotto un piccolo *tutorial* che spiegasse le forme fondamentali d'interazione con lo strumento. Questa azione è stata resa necessaria perché si è osservata una discreta refrattarietà iniziale all'uso dello strumento, soprattutto nelle funzioni aggiuntive e più importanti come le visualizzazioni sincroniche dei dati. Si è notato che nelle simulazioni di utilizzo del *software*, gli utenti si limitavano ad osservare le mappe esperte e gli indicatori automaticamente rappresentati senza interagire con le funzioni di base della visualizzazione, o proporre delle interpolazioni, o semplicemente osservare la trasformazione di un dato nel tempo in una specifica porzione del territorio.
- È stato aperto un canale di ascolto delle osservazioni e richieste degli utenti. Lo strumento non è autosufficiente, sia nell'imputazione sia nella gestione remota. È necessario che venga fornito un reale o virtuale ausilio per l'utenza. Alcune di queste funzioni di supporto potrebbero essere automatizzate perché risultano ricorrenti.

Figura 11 Screen shot dello strumento di geovisualizzazione analitica su cui si è impostata la sperimentazione



Tra i principali comportamenti degli utenti, è stato notato che non c'era una vera e propria ricerca di significati nascosti tra i dati ma piuttosto una voglia di classificazione: si è osservata la volontà di distinguere il proprio Comune attraverso la combinazione dei dati informativi che meglio presentassero quel territorio rispetto agli altri. Nonostante siano stati inizialmente forniti un numero limitato di indicatori, gli utenti hanno prodotto una numerosa quantità di personalizzazioni delle visualizzazioni.

Inoltre, nell'esperimento, la "story-telling" è risultata tra le funzioni più apprezzate ed utilizzate. Discutere dei dati che si visualizzano è fondamentale per pesare la familiarità che gli utenti hanno sia con l'argomento visualizzato sia con il metodo di visualizzazione. L'utente ha mostrato di aver bisogno di contestualizzare ciò che ha capito o scoperto per poterlo fissare e confermare anche comunicandolo per iscritto agli altri. Le storie più lunghe e concrete sono costruite dagli analisti che inducono alla riflessione e alla scoperta di nuove criticità. Si ritiene che la *story-telling* sia utile non solo per rafforzare la comprensione ma anche per la disseminazione delle informazioni statistiche. Il prossimo obiettivo sarà quello di testare i contenuti con un pubblico più ampio e attraverso tematiche ed indicatori più sensibili. Si tenterà di favorire anche un'interazione nell'imputazione e correzione del dato iniziale, ampliando e migliorando BD legati al territorio. Inoltre, sarà necessario studiare gli effetti prodotti dall'aumentato grado di accessibilità ad informazioni tecniche così da formulare una valutazione di profittabilità più precisa, con una curva che leghi i dati informativi più utilizzati alla domanda.

CONSIDERAZIONI E CONCLUSIONI

Questo articolo s’inserisce in un programma di ricerca e di sviluppo che ha lo scopo di indagare lo sfruttamento intelligente dei BD, focalizzando sull’aspetto della visualizzazione e della comprensione dei dati nelle analisi conoscitive per la programmazione territoriale. In primo luogo si è mostrato il potenziale economico che i dati digitali possiedono, cercando di individuare, in un sistema economico normale, le classi di domanda. Successivamente sono state definite alcune tra le più innovative applicazioni e tecnologie attualmente utilizzate nell’ambito della “visual-data communication”. Infine si è posta l’attenzione su un caso sperimentale ancora in fase di sviluppo, sfruttando alcuni tra gli strumenti selezionati.

Il caso applicativo, anche se ha prodotto modesti risultati nell’ambito delle scelte locali, ha dato diversi spunti sulle modalità d’uso degli strumenti di visualizzazione analitica e sull’interattività tra il dato e la domanda d’informazione. Il caso studio ha valore rappresentativo per tutte le realtà consorziate (pubbliche e private) che in questo periodo vogliono provare a trarre beneficio dalle risorse informative in loro possesso. In questo articolo, si è scelto di non approfondire le considerazioni e le discussioni che ci sono state nella costruzione dei contenuti delle visualizzazioni per concentrare l’attenzione sul ruolo scientifico dell’interazione con gli strumenti analitici. Le attività e le scelte fatte nel caso applicativo sull’Unione dei Comuni, a seguito delle visualizzazioni interattive di dati, non sono in sé importanti quanto il modo con cui si compongono e implementano gli strumenti, le possibili visualizzazioni geo-spaziali, la conoscenza aperta dei dati disponibili, fino all’uso interattivo della *story-telling*. A tal proposito, gli esperti che hanno lavorato attorno a questo studio concordano e propongono che in futuro ci si dovrà concentrare su:

1. Espandere il ruolo della visualizzazione:
 - Indagare e programmare nuovi ed innovativi metodi e strumenti per visualizzare, esplorare, animare, presentare, pubblicare e condividere i propri dati (aperti, pubblici o costruiti per la disseminazione). In futuro, la visualizzazione dei BD non avrà più il compito di descrivere le informazioni (non sarà più possibile o necessario) ma dovrà studiare la complessità, accettando anche l’incertezza¹³.
 - Adattare gli strumenti di visualizzazione analitica ad un uso su internet così da raggiungere un bacino di utenza più ampio oltre che una maggiore quantità di dati.

2. Migliorare il lavoro di analisi dei fenomeni sociali:
 - Aumentare le possibilità di raggiungere informazioni e riscontri in *real time* così da produrre scelte efficaci, condivise ed incentivanti. Un esempio pratico è quello in atto in molte Amministrazioni, relativo alle visualizzazioni analitiche dei dati sul trasporto pubblico in cui utenti e analisti contribuiscono a ridurre congestioni, a comprendere e programmare le rotte legandole alle preferenze.
 - Dare maggiore rilievo alla funzione di *Story-telling*. Le ricerche sulla visualizzazione analitica mostrano una forte concentrazione sull’esplorazione e presentazione del dato piuttosto che sugli strumenti che supportano la pubblicazione della conoscenza acquisita. Jern suggerisce di studiare un modo per registrare e immagazzinare le interazioni digitali degli utenti nel loro processo analitico di visualizzazione. Si propone di permettere di memorizzare delle “viste interattive” durante il processo di analisi.

¹³ Roccasalva G., et al. (2012), *The Future of Cities and Regions: simulation, scenario, Governance and Scale*, Springer Geography .

Oggi, l'uso intelligente dei BD è uno degli argomenti di sperimentazione più influenti sia in termini di crescita della ricchezza (individuale o collettiva) sia per l'aumento della competitività territoriale. Le Pubbliche Amministrazioni (PA) sono le più grandi generatrici e raccogliatrici nonché utilizzatrici di dati e informazioni digitali. Le PA possiedono i dati su diversi aspetti dell'economia, mantengono statistiche sulle licenze commerciali, sulla popolazione e sul suo andamento nel tempo (vita, morte, matrimoni).

Dal 2009 i cittadini e le organizzazioni non governative di tutto il mondo stanno facendo sempre maggiore pressione affinché ci sia libero accesso ai dati informativi (Open Data) a livello nazionale, statale e municipale. Inizialmente, questa richiesta è stata accolta dal presidente degli Stati Uniti, Barack Obama, al suo primo giorno in carica, con la pubblicazione su internet di tutti i dati pubblici in possesso delle Amministrazioni. In Europa, paesi come la Svezia hanno già posto basi solide per il cambiamento da un'economia di servizi a un'economia digitale. Da un punto di vista legislativo, la Direttiva 2003/98/CE del Parlamento Europeo ha definito per gli Open Data (OD) delle regole minime in merito alle informazioni raccolte dalle PA, limitandosi a condizioni di trasparenza, condizioni di riutilizzo e di non discriminazione tra gli utenti. L'adozione italiana della Direttiva europea è stata lenta, ha avuto un momento d'infrazione ma alla fine si è definitivamente avviata con la legge nazionale 96/2010 e con una prima legge regionale e l'apertura del primo portale di Open Data (OD).¹⁴

Negli ultimi anni l'accesso ai *database* ha stimolato lo sviluppo di applicazioni in grado di portare benefici alla collettività attraverso risparmi nei servizi pubblici, nuove opportunità di *business* per il settore privato ma anche favorendo le esigenze e l'aggregazione dei singoli nelle scelte di consumo (prezzi benzina, assicurazioni etc...). L'OD sono parte di quella "conoscenza aperta" che tutti stiamo costruendo e a cui tutti abbiamo diritto. E' probabile che i "nativi digitali" supereranno gli ostacoli nel reperimento delle informazioni, non perché più esperti degli attuali decisori pubblici e privati ma perché più armonizzati con la tecnologia. Tuttavia la capacità critica e quindi la comprensione passa attraverso il modo di accedere al dato. L'accessibilità ai dati informativi è la nuova frontiera della collaborazione multidisciplinare di studiosi delle scienze sociali ed informatiche, analisti territoriali ed economici ma anche di amministratori e cittadini. Una buona accessibilità permette di costruire strumenti, relazioni, programmi ma anche esprimere le scelte e produrre cambiamenti. L'accessibilità degli OD del settore pubblico è un campo difficile, come dimostra la sperimentazione descritta in questo articolo, soprattutto se si considerano le varie forme di domanda e la scarsa dotazione attuale di strumenti interattivi di visualizzazione. Non basta mettere in rete un dato informativo per definirlo "open"; questo, come scrive Berners-Lee¹⁵, deve essere strutturato per poter essere aggiornato da tutti, in un formato esportabile da tutti e soprattutto deve poter essere riconoscibile attraverso un opportuno collegamento da tutte le altre fonti di dati ad esso riconducibili. Gli strumenti così studiati e composti potranno generare una significativa quantità di benefici diretti ed indiretti al settore economico, sociale e al territorio.

Ringraziamenti

Si ricorda che le attività iniziali sugli strumenti di visualizzazione sono state condotte con Giuliana Bonello del CSI Piemonte attraverso il lavoro di tesi di primo livello dell'Ing. Andrea Valente. Inoltre si ringrazia il Prof. Mikael Jern della Norrköping University per i suggerimenti e la collaborazione nella stesura dell'articolo.

¹⁴ Maggio 2010 pubblicazione di www.dati.piemonte.it

¹⁵ Fondatore del termine "world wide web" e del linguaggio ipertestuale HTML.

Bibliografia

- Baack, S.,(2011) *A new style of News Reporting: Wikileaks and Data-driven Journalism*, Anno 2011.
- Batty M.,(2012) *Editorial, Environment and Planning B: Planning and Design*, volume 39, pages 413 – 415.
- Bellini L., *Le guide di DML: marketing automation*, Digital Marketing Lab, Milano.
- Corsico, F. Roccasalva, G., (2005) *Visu-an-alyse indicators of urban quality. The crucial role of forecasting Scenarios in sustainable decision making processes*, International Conference for Integrating Knowledge and Practice, Life in The urban Landscape, Gothenburg.
- Few, S.,(2007) *Data visualization, past, present and future.*, Cognos Innovation Center, Gennaio 2007.
- Few, S.,(2010), *Visual communication*, IBM Cognos Innovation Center, Settembre 2006.
- Jern M., et al., (2011), *A web-enabled visualization toolkit for geovisual analytics*,in *Information Visualization*, pag. 1-21 SAGE.
- Jossa, B., (2005) *Macroeconomia*, Padova, Cedam.
- O'Sullivan A., (2011) *Urban Economics* Kindle Edition.
- Pendse N., (2001) *The BI Verdict o OLAP report*, Business Application Research Center.
- Roccasalva G., et al. (2012), *The Future of Cities and Regions: simulation, scenario, Governance and Scale*, Springer Geography.
- Schilling M. A. (2009), *Gestione dell'innovazione*, McGraw-Hill, Milano.
- Ware, C., (2004) *Information visualization: perception for desing*, Morgan Kaufmann Publishers, Second Edition.
- Ware, C. (2008). *Visual Thinking for Design*. Morgan Kaufman.

Emerografia

- CSI Piemonte (2009), *Bilancio Sociale 2009*.
- Il Sole 24ore, (maggio 2011), *Il dato è tratto*.
- Il Sole 24ore, (maggio 2011) *Il Big Bang dei dati*.
- McKinsey Global Institute (Report 2011) *Big data: The next frontier for innovation, competition, and productivity*.
- The Economist (2010) *Data, data everywhere*. (A special report on managing information).

Sitografia

- Sito per le informazioni sull'applicativo Google Explorer URL:<http://www.google.com/publicdata/home> data di accesso: 01/03/2011.
- Primo esperimento di visualizzazione di database italiani sulla base di InfoVis technology. URL: <http://www.ncomva.se/flash/projects/italy/>
- Sito per l'informazioni sull'applicativo manyEyes, URL: <http://www-958.ibm.com/software/data/cognos/manyeyes/>
- Sito informativo per l'applicazione Fineo, URL: <http://www.densitydesign.org/>
- Sito per i dati statistici generali del CSI Piemonte URL: <http://www.densitydesign.org/>
- Sito della società produttrice di Open eXplorer URL:<http://www.ncomva.org>
- Sito della McKinsey, URL:<http://www.mckinsey.com/>