

BIG DATA NEL MONDO ASSICURATIVO E PREVIDENZIALE

Un campo da esplorare ed una strada per lo sviluppo sostenibile

Big Data: cosa sono

- Sono i **flussi di informazione** presi nel loro complesso e che caratterizzano ciascun soggetto all'interno d'una collettività organizzata.
- Sono composti da **dati strutturati** (generalmente stipati in appositi archivi che rappresentano caratteristiche stabili) e **dati non strutturati** (che rappresentano eventi estemporanei e, generalmente, non organizzati).
- Sono di **grandi dimensioni** (appunto «Big») e **polimorfi sia nel tempo (aggiornamento e frequenza) che nello spazio (tipologia)**.

Big Data: come si riconoscono

- **Volume**: sono generalmente di grandi dimensioni, dal terabyte (10^{12} byte) allo zetabyte (10^{21} byte).
- **Varietà**: sono di svariati formati ed altrettanto svariate strutturazioni e tipologie (record, post, tweet, SMS, MMS, video, voce, ...)
- **Velocità**: hanno svariate frequenze di aggiornamento (istantanee, quotidiane, mensili, ...).

Big Data: come si riconoscono

Spesso viene aggiunta una ulteriore variabile: la «**Accuratezza**» o «**Affidabilità**». Un post su Facebook su un servizio erogato può essere eccessivamente negativo o positivo (e non è chiaro neanche in che misura), mentre i dati memorizzati per erogare il servizio devono essere ritenuti certi, quindi affidabili. I dati affidabili e strutturati gestiti direttamente dall'azienda sono detti «**Dati Master**», quelli non gestiti dall'azienda ma che si scatenano dalla fruizione dei «Dati Master» sono «**Dati collegabili ai dati master**» e sono il vero oggetto dei «Big Data».

UNA DEFINIZIONE

Una definizione interessante di Big Data è stata fornita da Alexander Jaimes, ricercatore presso Yahoo Research: «**i dati siamo noi**». Il senso è quello che sono tutti i soggetti attivi di una collettività a generare dati poiché **tutti interagiscono e reagiscono scambiandosi informazione** (quindi, «flussi», quindi, «Big Data»).

UNA CARATTERISTICA

L'informazione è come la frutta: si deteriora subito e dà i suoi benefici solo se viene consumata. Per esempio, l'informazione su un volo di linea vale solo finché l'aereo non è partito e per fruirne occorre prendere l'aereo... Da questo semplice concetto, i Data Warehouses e la Business Intelligence: **implementare i metadati per servirsene in tempo utile e sfruttare il vantaggio che ne deriva.**

DATA WAREHOUSE E BIG DATA

Un Data Warehouse è fondamentale per approcciare il fenomeno dei Big Data, ma non basta. Innanzitutto, deve mutare ruolo e passare da semplice «magazzino» dove un dato c'è se qualcuno l'ha riposto a strumento sul quale costruire la politica di gestione dell'informazione in funzione della missione aziendale. Questo è il passaggio concettuale dal **DWH** (Data WareHouse = magazzino) all'**EDWH** (Enterprise DWH) a sottolineare il fatto che **l'impresa appare e si caratterizza attraverso l'informazione che gestisce e governa.**

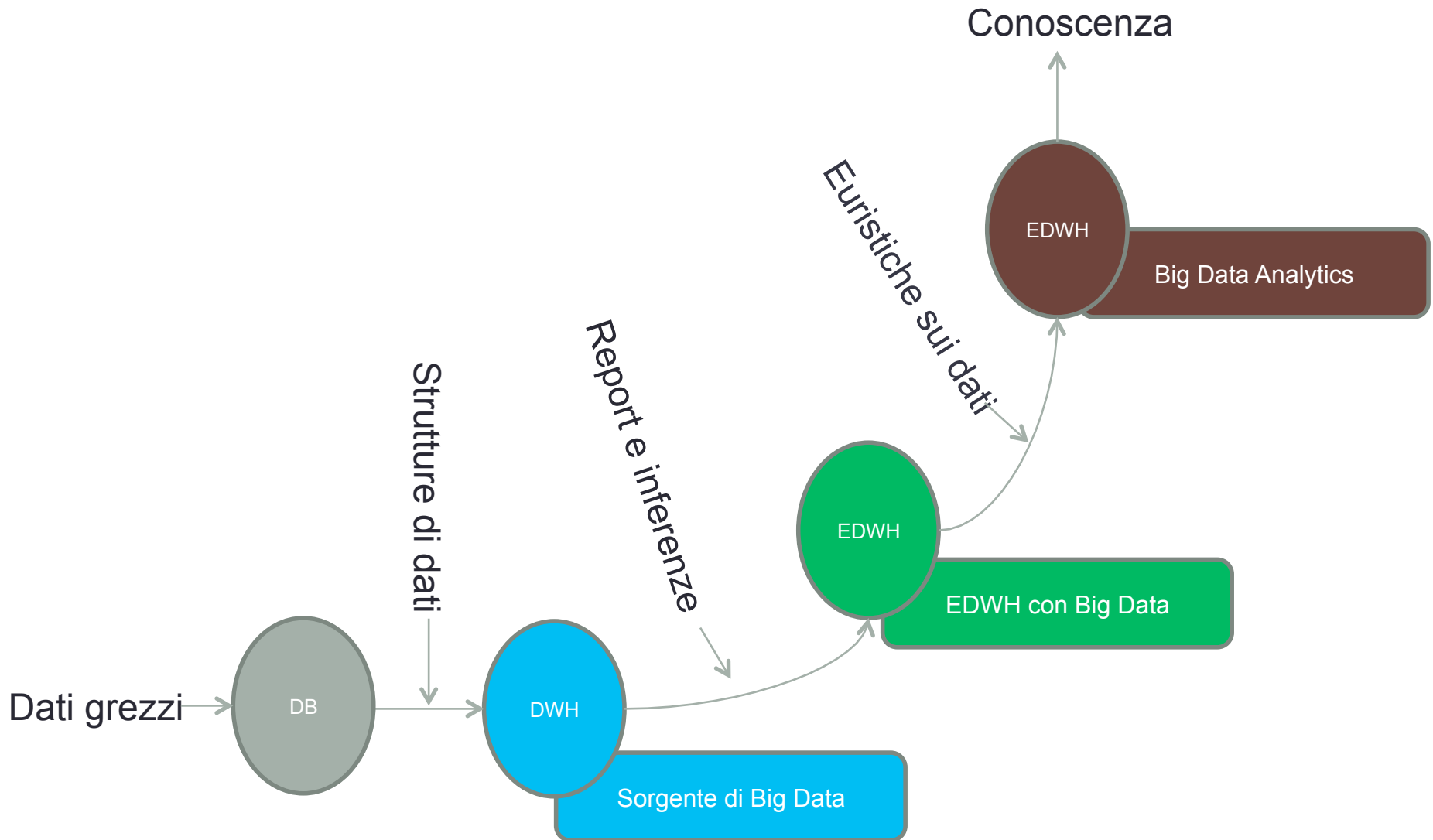
DATA WAREHOUSE E BIG DATA

In un EDWH trovano posto tutte le euristiche sui dati aziendali in funzione della propria missione. Se la missione aziendale è di tipo competitivo, le euristiche saranno orientate a generare previsioni per conquistare «prima» (e col minimo sforzo) un mercato (**Business Intelligence**). Se la missione è di tipo istituzionale/sociale, le euristiche saranno orientate al benessere della collettività da perseguire col minimo impegno di risorse pubbliche (**Business + Social Intelligence**).

DATA WAREHOUSE E BIG DATA

Le euristiche presenti in un EDWH formano una rete semantica che, utilizzata ai vari livelli aziendali, realizza la c.d. «Business Analytics». C'è un mutuo scambio concettuale ed implementativo tra i portali web dell'azienda ed il suo EDWH: **nei portali web sono implementate le applicazioni che trasformano i dati, negli EDWH sono implementati i significati dei dati in funzione della missione aziendale. I significati dei dati (rete semantica) forniscono indicazioni su come devono evolversi i portali web e i portali web danno indicazioni su quali euristiche implementare negli EDWH.** Il concetto ricorda molto da vicino una celebre frase dell'eminente fisico J.A. Wheeler sulla Relatività Generale: *«la materia dice allo spazio come curvarsi e lo spazio dice alla materia come muoversi»*. In entrambi i casi è un processo continuo e mutevole che cresce esponenzialmente all'aumentare della dimensione, quindi, delle «**tre V**» che caratterizzano i Big Data. **Si stima che nel breve-medio periodo almeno il 50% delle aziende di grandi dimensioni investiranno su applicazioni basate sul Data Warehousing (DWH appliances) in vista di gestire Big Data.**

IL PERCORSO EVOLUTIVO

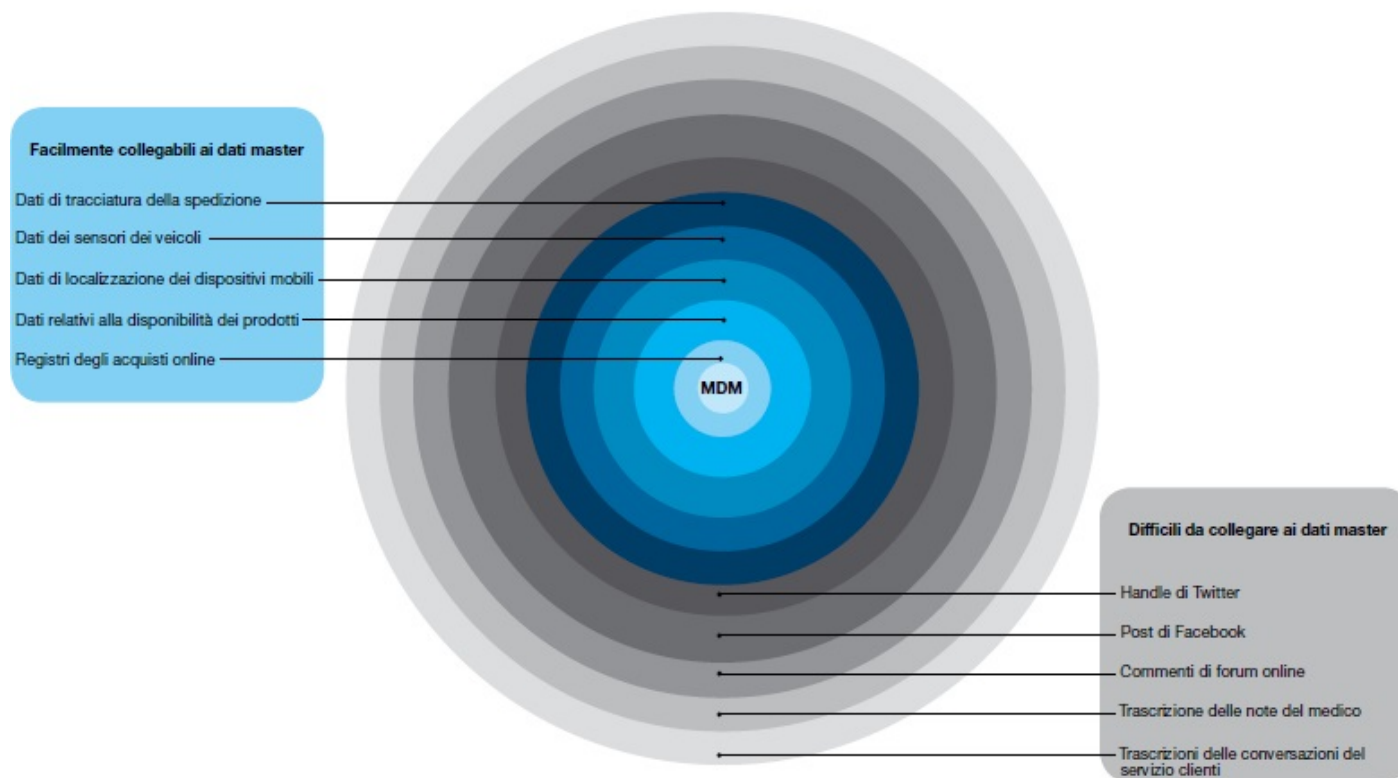


LA VELOCITA' DELLE DECISIONI

Decidere velocemente non è nuovo nel mondo delle aziende, degli enti e di tutte le collettività strutturate. **Per questo motivo si immagazzinano dati e si cercano euristiche per prevedere fenomeni.** In generale, si è costantemente alla ricerca di sistemi per **trasformare nel più breve tempo possibile dati in informazioni ed informazioni in decisioni.** Con lo scenario attuale, però, le cose si complicano a causa delle «tre V» che, a loro volta sono frutto della fortissima interconnessione tra tutti i componenti di una collettività (e tale fenomeno è sospinto dal dilagare del Web Mobile ed al suo sempre maggior utilizzo anche in applicazioni che, solo fino a qualche anno fa erano inconcepibili «extra moenia»). Cresce il bisogno di velocità per catturare informazione memorizzandola e strutturandola per poi ricavare, dalla struttura, euristiche chiave che a mo' di neuroni arricchiscono il «cervello» di tutta l'organizzazione. **Il livello di aggiornamento e la coerenza diventano un fatto vitale: il loro non governo può essere disastroso perché porta a decisioni errate.** Ben si comprende, quindi, come l'allineamento della basi dati, l'elaborazione delle queries e la restituzione dei risultati necessitino di tecnologie, architetture e applicazioni ottimizzate e dedicate. Da qui, però, nascono costi che una collettività non può permettersi se non li riesce a bilanciare con benefici strutturali. Ne consegue la necessità di una oculata politica di investimento che ha come ulteriore variabile i risultati ottenibili attraverso le euristiche memorizzate negli EDWH pubblici, ove è necessario che tali strutture esistano.

LA VELOCITA' DELLE DECISIONI

Trasformare nel più breve tempo possibile dati in informazioni ed informazioni in decisioni passa attraverso l'analisi di collegabilità tra l'informazione non strutturata (e meno affidabile) e quella strutturata (e più affidabile) che rappresenta il «core business» dell'azienda.



FENOMENI E NUMERI

Il concetto di Big Data è recente ed è frutto di un Web dilagante ed ubiquo che ormai impatta sulla società in maniera strutturale mettendo in evidenza fenomeni nuovi. Le recenti sommosse che hanno caratterizzato il Maghreb e, di recente, la Turchia sono state accompagnate (e, secondo studi effettuati, scatenate) dai Social Network che hanno fortemente accelerato la circolazione delle idee. Secondo un recente studio di Martin Hilbert, dell'University of Southern California: «**Viviamo in un mondo in cui le economie, la libertà politica e la crescita culturale dipendono sempre più dalle nostre capacità tecnologiche**». L'anno in cui la totalità delle informazioni immagazzinate in forma elettronica ha superato quelle su tutti supporti analogici è stato il 2002, ma già nel 2007 esse rappresentavano il 94% della «memoria» dell'umanità. I ricercatori hanno stimato che gli strumenti di archiviazione digitale e analogica in possesso dell'umanità siano in grado di immagazzinare **295 exabyte** (10^{18} byte) di informazione. Sempre nel 2007 sono stati scambiati attraverso le tecnologie broadcast, come la televisione e il GPS, ben **19 zettabyte** (10^{21} byte) di informazione. E' come se ogni abitante del pianeta leggesse ogni giorno 174 quotidiani. Questo sembra non entrarci nulla col tema in questione, **ma ci dice quanto è vasto il problema e, soprattutto, che siamo di fronte a qualcosa di totalmente nuovo.**

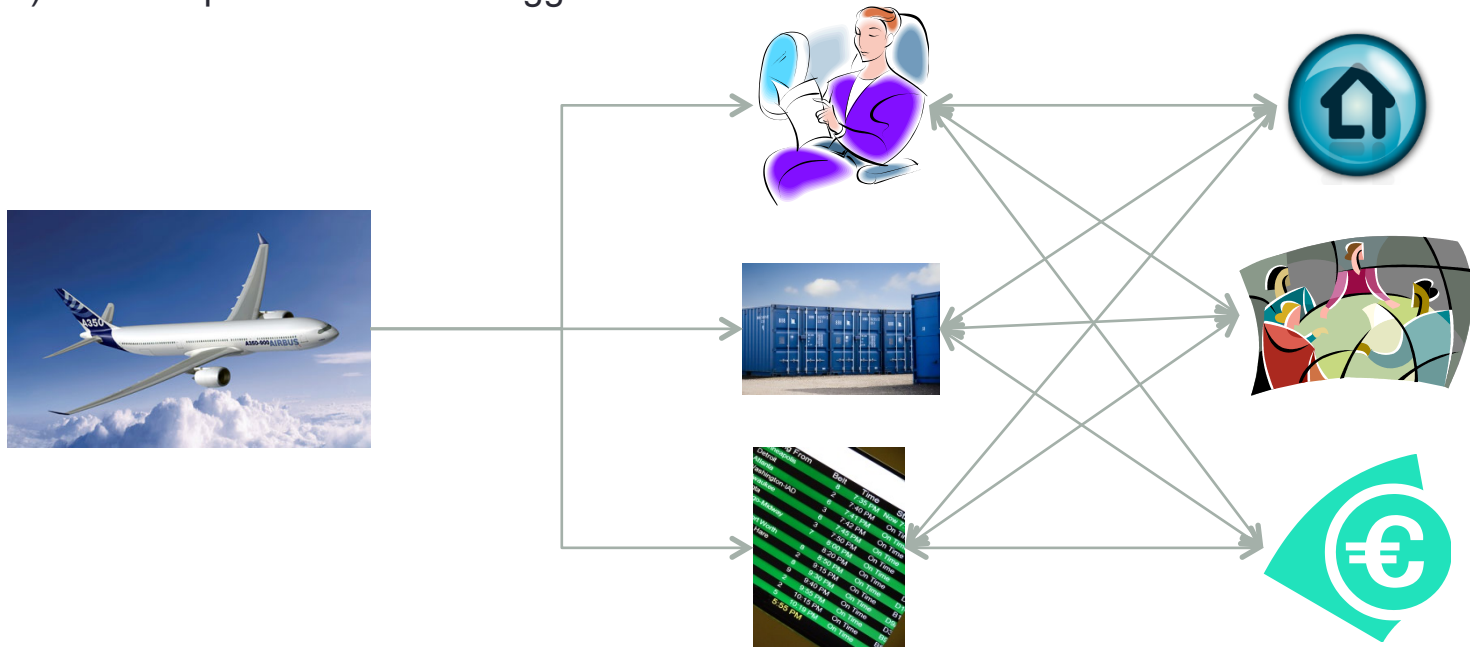
FENOMENI E NUMERI: PERCHE'?

Perché questi numeri enormi? **Perché il Data Warehousing e la Business/Social Intelligence hanno spostato l'asse dell'attività di analisi dai problemi di strutturazione dei dati a quelli che riguardano i fenomeni che i dati rappresentano.** Per fissare le idee, facciamo l'esempio di una compagnia aerea che gestisce voli in ambito nazionale ed internazionale. Semplificando, le variabili da prendere in esame sono gli aeroporti, gli orari, gli aeromobili, i posti, le capacità delle stive, i rapporti con le altre compagnie ed i servizi aeroportuali. In genere, è così che nasce una «Base Dati» ed il problema, detta X_i la cardinalità di ciascuna variabile, ha una complessità $N = X_1 X_2 \dots X_k$ se k è la totalità delle variabili. In tale caso si dice che **la complessità del problema di strutturare i dati dei voli è di tipo polinomiale.**



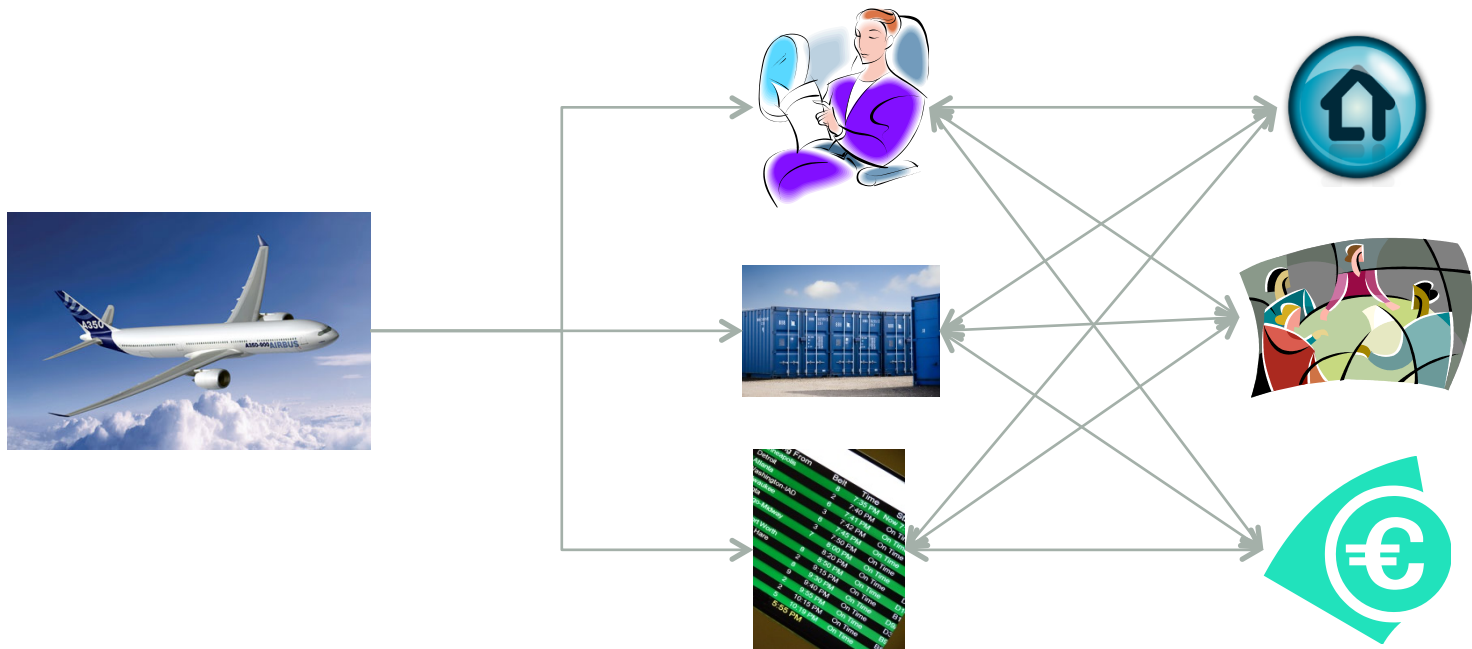
FENOMENI E NUMERI: PERCHE'?

Torniamo all'esempio della compagnia aerea, ma sotto un altro punto di vista: **cosa succede nello svolgimento di ciascun volo? Quante variabili sono da prendere in esame?** Rispetto a prima, il problema è molto diverso: un passeggero può perdere l'aereo, una merce può essere smarrita o destinata male, qualcuno non ottiene ciò che voleva (e protesta), qualcun altro ha una perdita economica, qualcun altro ne «approfitta» etc.. Il bello che **un solo evento ne può scatenare a catena molti altri**. Adesso ogni evento è una variabile e ciascuna può rappresentare un qualcosa che va bene (= 1) o male (= 0). Il numero di variabili sarà un certo **N** (molto più grande del precedente) e **ciascun esito complessivo di «bonus/malus» comporterà un costo (positivo o negativo)**. Il problema, adesso, **ha complessità 2^N , diventando, quindi, esponenziale**. Big Data, sostanzialmente, vuol dire, se non tenerne traccia, analizzare fenomeni che ricadono in questo (enorme) universo per ricavarne vantaggio.



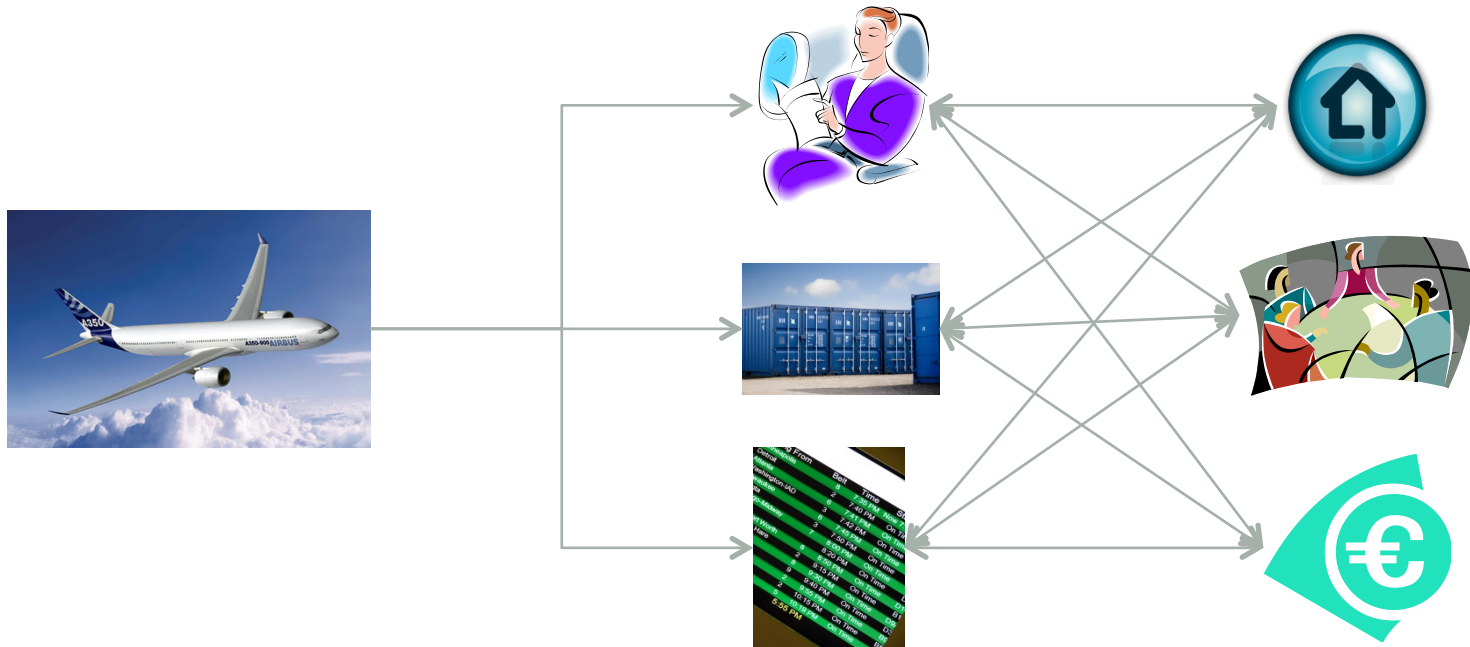
FENOMENI E NUMERI: PERCHE'?

Non è pensabile né tantomeno proponibile che un problema di complessità 2^N (esponenziale) possa essere memorizzato su un supporto e trattato nella sua interezza. Si tratta di una classe di problemi detti «NP-completi» per i quali **sicuramente** la soluzione può essere **verificata** in un tempo polinomiale avendo le giuste informazioni, non **trovata** nel medesimo tempo (in tal caso si tratterebbe di un problema di classe computazionale «P»). La differenza è sottile, ma evidente: mettere in ordine un insieme di numeri **ha soluzione** in un tempo polinomiale, trovare a quali fattori primi corrisponde un numero intero molto grande **non può essere risolto se non enumerando tutte le possibili soluzioni** (e su questa difficoltà, al momento, si fonda la sicurezza delle transazioni finanziarie online) ed **è il tempo di ricerca d'una soluzione ad essere polinomiale, non la determinazione della soluzione stessa.**



FENOMENI E NUMERI: PERCHE'?

Proprio dall'emergere di detti problemi è scaturita la strategia di utilizzare i Data Warehouses. **In essi potevano essere riposte le euristiche sui dati gestiti a mo' di risultati parziali per indagare su problemi complessi e poco appariscenti al fine di ricavarne un vantaggio competitivo.** Una euristica, in effetti, è un passo in più nella conoscenza delle giuste informazioni per affrontare un problema squisitamente «NP». **Non è**, come spesso viene interpretata, **una risposta per risolvere un problema estemporaneo.** Parallelamente, è giusto dire che un Data Warehouse **non è solo un sistema per effettuare query complesse in breve tempo**, ma un luogo dove i modi per convergere alla soluzione che rappresenta l'ottimo della propria strategia di business (o missione aziendale che dir si voglia) si stratificano e si riutilizzano per continuare ad essere competitivi (nel caso privato) o per continuare a garantire benessere collettivo (nel caso pubblico), prevedendo e fronteggiando in tempo utile eventuali situazioni di crisi.



UN ESEMPIO DI «BIG DATA SOURCE»

Esempi di **sorgenti di «Big Data»** ce ne sono molti ed emergono e si sviluppano con una rapidità sinora impensata. L'esempio seguente è un caso di una azienda che vende form di input online ed ha utilizzato i Social Network per veicolare il suo business. Come è evincibile in figura, si tratta dell'azienda «JotForm» che ha messo un premio di 5000 \$ per incentivare la realizzazione di nuove app che utilizzano la sua piattaforma.



The image shows a screenshot of an email from JotForm. At the top left is the JotForm logo, which consists of a stylized pencil icon followed by the text "JotForm". The main content of the email is as follows:

We have some exciting news to share today!

We have recently released an API to build apps for JotForm. We are now celebrating our new API with a **competition for software developers**. The grand prize is **\$5,000**.

<http://developers.jotform.com/competition>

Our goal is to encourage developers to **build great apps** for JotForm. You will be able to get these apps on a new JotForm Apps section in our site.

Help us get more apps for JotForm! Here is how you can help: If you know any software developers please forward this email to them. If you use any of the sites below please share & help us spread the news!

Below this text are five social media icons: Twitter (blue square with white bird), Facebook (blue square with white 'f'), Google+ (red square with white 'g+'), YouTube (orange square with white 'Y'), and Reddit (blue square with white alien head).

You are receiving this e-mail because you have a JotForm account: fsculco.
If you no longer want to receive emails from JotForm please [click here to unsubscribe](#).
E-mail not displaying correctly? [View it in your browser](#).

UN ESEMPIO DI «BIG DATA SOURCE»

Senza entrare nei dettagli, utilizziamo questo esempio per calcolare come l'iniziativa può diffondersi a dismisura in poche ore: detto **N** il numero di utilizzatori dei Social Network interessati e supponendo che ognuno abbia un numero di «amici» superiore o uguale a 1 (il che garantisce la propagazione ad almeno 2 soggetti) ed inferiore od uguale a un certo **A** ($\leq N$), basta che in un giorno si realizzino **X** inoltri in cascata per raggiungere subito un numero **C** di contatti elevatissimo. **X** inoltri sono la profondità di un albero che ad ogni nodo ha una ramificazione **R** compresa tra 2 ed **A**. In formule:

$$2^{\uparrow X} \leq C = R^{\uparrow X} \leq A^{\uparrow X}$$

$$2 \leq R \leq A$$

E' evidente quanto la propagazione sia rapida ed esponenziale, così come s'addice ad un fenomeno di «Big Data».

INPS COME «BIG DATA SOURCE»

In tale caso si vuole evidenziare solo l'aspetto collegato ai Social Network (e già dà il sentore del grande impatto), ma come sorgente di «Big Data» non è il solo aspetto. Quelli a contenuto ancora più alto e rilevante sono quelli che derivano dagli incroci dei dati amministrati con quelli delle altre PA (e/o altri soggetti di pubblico interesse) al fine dello svolgimento dei compiti istituzionali (ISTAT, Comuni, Regioni, Banche, Consulenti del Lavoro, Patronati, Associazioni di categoria etc.). **Come sarà evidente in seguito, l'impatto va ben oltre quello numerico o del successo competitivo di «Business», ma diventa una fonte preziosa per l'intero Paese in tema di giusto investimento di risorse su previdenza e welfare.**

INPS COME «BIG DATA SOURCE»

Facebook

Risultati raggiunti

Pagina Facebook	Data pubblicazione	N. «Mi piace»
Riscattare la Laurea	maggio 2011	5.800
Utilizzare i Buoni Lavoro	novembre 2011	4.360
Gestire il Lavoro Domestico	febbraio 2012	3.280
Pensioni: Sistema Contributivo	dicembre 2012	2.370

dati aggiornati al 20/08/2013

INPS COME «BIG DATA SOURCE»

YouTube Risultati

Dal lancio del progetto - dicembre 2012 - ad oggi, il canale YouTube Inps ha registrato

32.550 visualizzazioni video

308 iscritti

Il video più visto è stato quello dedicato all'operazione **CUD 2013 Online** (febbraio 2013) che ha ricevuto più di 17.200 visualizzazioni.



CUD 2013 Online

17.253 visualizzazioni 5 mesi fa

dati aggiornati al 20/08/2013



INPS COME «BIG DATA SOURCE»

Twitter Risultati

Ad oggi il profilo Twitter @Inps_it:

- ha pubblicato **128 tweet**
- segue **26 following**
- ha raggiunto **860 follower**

Il 16 luglio scorso la presentazione alla Camera dei Deputati della Relazione Annuale del Presidente dell'Inps è stata seguita in diretta *live-tweeting* tramite l'hashtag **#Rapporto_Annuale_INPS**

dati aggiornati al 20/08/2013



INPS COME «BIG DATA SOURCE»

Social Media Map

I contenuti pubblicati sui diversi canali social rinviano ad approfondimenti, informazioni o servizi disponibili sul [portale istituzionale](#), che rimane il principale punto di riferimento per la comunicazione dell'Istituto sul web.



Gli unici contenuti portati «fuori» dal portale istituzionale sono i video della sezione *Inps Comunica* che popolano anche il canale YouTube, per essere poi ri-distribuiti sulle altre piattaforme social.

INPS COME «BIG DATA SOURCE»

Social Network	Attivazione	Contatti/follower
Facebook	Maggio 2011	Ca. 108 al giorno
You Tube	Dicembre 2012	Ca. 133 al giorno
Twitter	Luglio 2013	Ca. 28 al giorno

Dati aggiornati ad Agosto 2013

INPS COME «BIG DATA SOURCE»



- Il progetto **INPS Open Data** ha risposto all'esigenza di fornire ai cittadini i dati dell'Istituto in maniera semplice e trasparente.

OPEN DATA

- In linea con la Direttiva del Parlamento Europeo 2003/98/CE, il **15 Marzo 2012** INPS pubblica sul proprio portale una sezione dedicata agli Open Data.
- Cosa offre al cittadino?
 - Ricerca full-text e semantica (navigazione a faccette)
 - Download dei dataset in formati standard (XML, CSV, XLS) e semantici (OWL – Linked Open Data)
 - Descrizione dei dataset attraverso metadati standard (Dublin Core)
 - Pubblicazione dell'ontologia di riferimento dell'Istituto
 - **API – Application Programming Interface**
- Tutti i dati possono essere **riutilizzati**, come definito dalla licenza **IODL 2.0** con cui vengono pubblicati

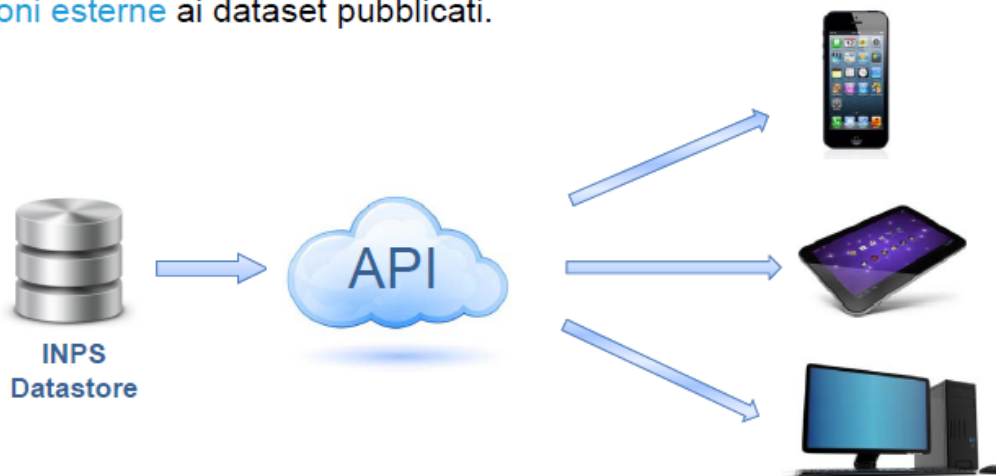


INPS COME «BIG DATA SOURCE»



- Cosa sono?

Un'interfaccia utilizzabile da sviluppatori software per accedere mediante **applicazioni esterne** ai dataset pubblicati.



- Le informazioni prelevate mediante API possono quindi essere **riutilizzate** negli ambiti più diversi:

- ✓ Realizzazione di applicazioni mobile (apps)
- ✓ Analisi statistiche
- ✓ Sistemi di Knowledge Extraction (data mining)
- ✓ ...

INPS COME «BIG DATA SOURCE»



- Servizi offerti:
 - ✓ Catalogo dataset (<http://serviziweb2.inps.it/odapi/catalog>) – Fornisce l'elenco degli identificativi di tutti i dataset disponibili
 - ✓ Download metadati (<http://serviziweb2.inps.it/odapi/catalog/49>) – Permette l'estrazione di tutti i metadati di uno specifico dataset (nell'esempio, il #49)
 - ✓ Bulk (<http://serviziweb2.inps.it/odapi/bulk>) – Il bulk contiene l'intero datastore INPS, ovvero dati e metadati di tutti i dataset. In questo caso il formato utilizzato è **JSON-LD** (JSON for Linking Data) che permette di accorpate i dati con i relativi metadati
 - ✓ Status (<http://serviziweb2.inps.it/odapi/status>) – Servizio che permette agli sviluppatori di accedere ad informazioni utili quali:
 - Ultima data di aggiornamento del bulk
 - Numero di dataset presenti
 - ...

INPS COME «BIG DATA SOURCE»



- Dataset pubblicati: 322, nei formati xls, csv, xml, owl. Altri 99 in lavorazione.
- Download effettuati (18 mesi): 750.000
- Da fuori Italia: 10%
- Indice gradimento utenti: 7/10

INPS COME «BIG DATA SOURCE»



- CITAZIONI/INTERVENTI:
 - EPSI: citazione sul proprio sito *The INPS data release is a good step forward, and creates appetite for more.* “Interestingly the **INPS** data website explicitly refers to the EU PSI Directive as the reason for publishing data for re-use.”
 - Partecipazione, su invito, a:
 - Webinar su Open Data tenuto dal Formez,
 - Forum PA-Open Data 2012, relazione “Open Data INPS”
 - ICT Torino 2012, “OpenData Inps, un modello di sviluppo per l’OpenGovernment nella PA.
 - Forum PA 2013, Barcamp Open Data.
 - Webinar su Big Data, organizzato dal Formez.
 - nel corso della presentazione da parte della FISAL del bando Open Data Regione Lazio, gli Open Data INPS sono stati definiti “**i migliori per qualità tecnica fra tutti gli Open Data pubblicati in Italia**”.
 - Partecipazione alla stesura delle Linee Guide Nazionali Open Data redatte da DigitPA.
 - Partecipazione alla stesura delle Linee Guide Nazionali per la valorizzazione del patrimonio informativo pubblico, redatte da Agenzia per l’Italia Digitale.
 - Open Data INPS menzionati nel corso di convegni collaterali al recente G8.
- EPSI: citazione sul sito per il progetto API

INPS «BUONE PRATICHE»

The screenshot displays the INPS Intranet interface. At the top, there is a browser window with the URL <http://spi-inps01/inpsdwh>. Below the browser, the website header includes the INPS logo, the word "Intranet", and a search bar with the text "CERCA". A navigation menu contains the following items: ISTITUTO, DIREZIONE GENERALE, SEDI, PROCESSI, INFORMAZIONI, SERVIZI, UTILITÀ, and I MIEI LINK. A red stamp with the text "IN PROGRESS" is overlaid on the navigation menu.

The main content area features the INPS logo and the text "Il portale di accesso al Data Warehouse di INPS". A search bar labeled "Ricerca Report" is positioned to the right. Below the header, there is a sidebar on the left with a "Menu principale" and "Aree attivate" sections. The main content area is titled "BENVENUTO IN I.NPS" and contains a "Benvenuto in I.NPS" section with a map of Italy and a "Cosa è I.nps?" section. The map shows the regions of Italy, with a legend indicating three categories: "Importo INPS 2013" (Import INPS 2013), "meno di 10 miliardi di Euro" (less than 10 billion Euro), "tra 10 e 20 miliardi di Euro" (between 10 and 20 billion Euro), and "più di 20 miliardi di Euro" (more than 20 billion Euro). The "Cosa è I.nps?" section explains that I.nps is the portal for the INPS Data Warehouse, designed to manage and provide access to all information extractable from the data archives for the purpose of supporting, at all levels, the individual Centers of Responsibility in the performance of institutional activities and in tactical and strategic decisions.

The sidebar on the left contains the following menu items:

- Menu principale
 - Pagina iniziale
 - Presentazione e finalità
 - Contatti
 - Abilitazione alla reportistica
- Aree attivate
 - Area Ammortizzatori Sociali
 - Area Aziende, Lavoratori dipendenti e Fondi speciali
 - Area Interventi a sostegno delle fasce deboli
 - Area Autonomi e Gestione separata
 - Area Invalidità
 - Area Pensioni
 - Area Lavoratori Domestici, contributi individuali
 - Area Bilanci e Contabilità
 - Area Controllo di gestione e Business Intelligence
 - Area Vigilanza e Risk management
 - Area Risorse
 - Area Logistica
 - Area Sinergie
 - Area Portali
 - (Internet/Intranet/Extranet) e Gestione Documentale
- Cruscotti
- Report combinati
- Report personali

INPS «BUONE PRATICHE»

The screenshot shows a web browser window displaying the INPS Data Warehouse portal. The browser's address bar shows the URL `http://spi-inps01/inpsdwh#`. The page header includes a navigation menu with items like 'ISTITUTO', 'DIREZIONE GENERALE', 'SEDI', 'PROCESSI', 'INFORMAZIONI', 'SERVIZI', 'UTILITÀ', and 'I MIEI LINK'. The main content area is titled 'Il portale di accesso al Data Warehouse di INPS' and features a search bar labeled 'Ricerca Report'. A sidebar on the left contains a 'Menu principale' with categories like 'Pagina iniziale', 'Presentazione e finalità', 'Contatti', and 'Aree attivate'. The 'Aree attivate' section lists various service areas, with 'Osservatorio ISEE' highlighted. The main content area displays the 'Area Interventi a sostegno delle fasce deboli' section, which includes a 'Data Warehouse' icon and a detailed text block explaining the ISE and ISEE indicators and their application in determining social assistance benefits.

Menu principale

- Pagina iniziale
- Presentazione e finalità
- Contatti
- Abitolazione alla reportistica

Aree attivate

- Area Ammortizzatori Sociali
- Area Aziende, Lavoratori dipendenti e Fondi speciali
- Area Interventi a sostegno delle fasce deboli
- Osservatorio ISEE**
- Area Autonomi e Gestione separata
- Area Invalidità
- Area Pensioni
- Area Lavoratori Domestici, contributi individuali
- Area Bilanci e Contabilità
- Area Controllo di gestione e Business Intelligence
- Area Vigilanza e Risk management
- Area Risorse
- Area Logistica
- Area Sinergie
- Area Portali
- Area (Internet/Intranet/Extranet) e Gestione Documentale

Cruscotti

- Report combinati
- Report personali

OSSEVATORIO ISEE

Area Interventi a sostegno delle fasce deboli

Per la richiesta di **prestazioni assistenziali legate al reddito o di servizi di pubblica utilità** è prevista la valutazione della situazione economica del richiedente, con riferimento al suo nucleo familiare: a tal fine sono calcolati due indici: l'ISE (indicatore della situazione economica) e l'ISEE (indicatore della situazione economica equivalente).

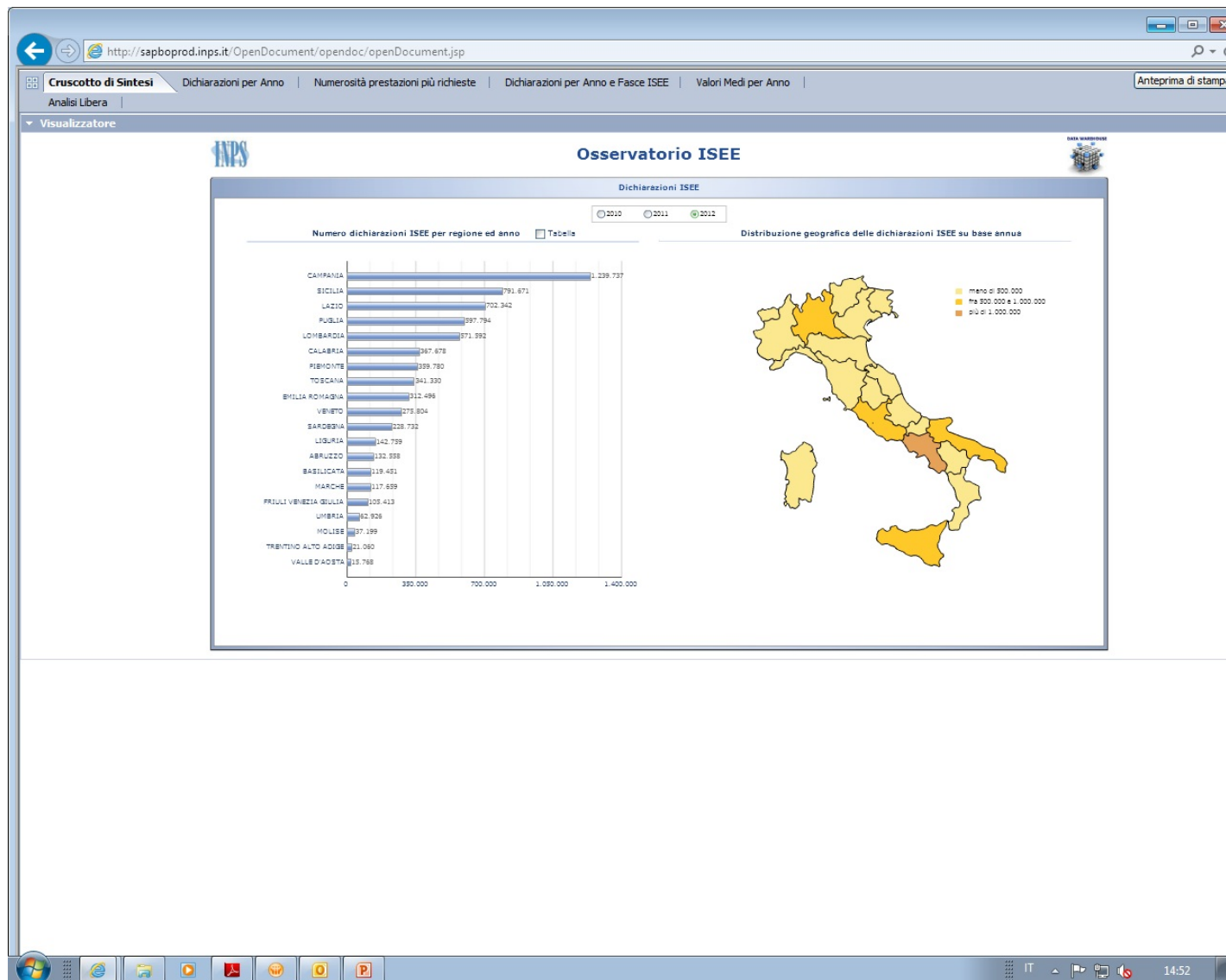
L'**ISE** è determinato dalla somma dei redditi e del venti per cento del patrimonio; l'**ISEE** scaturisce invece dal rapporto tra l'ISE e il parametro desunto dalla scala di equivalenza.

Il **nucleo** di riferimento è composto, in linea generale, dal dichiarante, dal coniuge e dai figli, nonché da altre persone con lui conviventi e da altre persone a suo carico ai fini IRPEF, con alcune eccezioni e particolarità; il **reddito** "medio" è quello complessivo ai fini IRPEF sommato al reddito delle attività finanziarie, con una detrazione in caso di residenza del nucleo in un'abitazione in locazione; per **patrimonio** si intende sia quello immobiliare che mobiliare, con l'applicazione di una franchigia; la **scala di equivalenza** prevede i parametri legati al numero dei componenti il nucleo familiare e alcune maggiorazioni da applicare in casi particolari, quali la presenza di un solo genitore o di figli minori o di componenti con handicap, lo svolgimento di attività lavorativa da parte di entrambi i genitori ecc.

INPS informa

Open Data

INPS «BUONE PRATICHE»



INPS «BUONE PRATICHE»

The screenshot shows a web browser window displaying the INPS Data Warehouse portal. The browser's address bar shows the URL `http://spi-inps01/inpsdwh#`. The page header features the INPS logo and the text "Il portale di accesso al Data Warehouse di INPS". A search bar labeled "Ricerca Report" is visible, along with a user greeting "Benvenuto Sculco Luigi".

The main content area is titled "CONTATTI" and contains the following text:

Contatti

Sono disponibili due canali di contatto strutturato (oltre ai consueti canali telefonici):

- Aiuto sul sistema I.nps (malfunzionamenti, interpretazione dei contenuti, utilizzo, scopi etc).
- Proposta e richiesta di nuovi report o osservatori.

Entrambi sono implementati attraverso un apposito form che memorizza il richiedente, l'oggetto, la struttura di provenienza al fine di utilizzare dette informazioni per meglio catalogare i contenuti di I.nps, giacché quest'ultimo si pone l'obiettivo di rappresentare e rendere esplicita la conoscenza su tutta l'informazione elaborata dall'INPS e sui fenomeni ad essa connessi finalizzandola al miglioramento del ruolo dell'Ente sulla collettività amministrata (Social Intelligence) e sul miglioramento delle sue performances (Business Intelligence).

The left sidebar contains a "Menu principale" with categories like "Pagina iniziale", "Presentazione e finalità", "Contatti", "Abitolazione alla reportistica", "Area attivate", "Cruscotti", "Report combinati", and "Report personali". The right sidebar features logos for "INPS Contact Center informa" and "Open Data".

INPS «BUONE PRATICHE»

The screenshot shows a web browser window displaying the INPS Data Warehouse portal. The browser's address bar shows the URL `http://spi-inps01/inpsdwh#`. The page header includes the INPS logo and the text "Il portale di accesso al Data Warehouse di INPS". A search bar labeled "Ricerca Report" is visible, along with a user greeting "Benvenuto Sculco Luigi".

The main content area is titled "RICHIESTE INVIATE" and features a sub-header "Richieste inviate" with a "Data Warehouse" icon. Below this is a table with the following data:

Id	Oggetto richiesta	Data richiesta	Stato
Richieste in stato Inviata (3)			
41	estrazione dati relativi alle domande di indennità	24/09/2013 0.00.00	Inviata
42	Test	25/09/2013 0.00.00	Inviata
43	Richiesta PWD	25/09/2013 0.00.00	Inviata

Below the table, it indicates "Oggetti trovati: 3" and provides navigation controls: "<< < 1 di 1 > >>".

On the right side of the page, there are logos for "INPS Contact Center informa" and "Open Data".

The left sidebar contains a "Menu principale" with the following items:

- Pagina iniziale
- Presentazione e finalità
- Contatti
 - Richiesta attività
 - Richiesta assistenza
 - Richieste inviate**
- Abilitazione alla reportistica

Below the menu is the "Aree attivate" section, listing various service areas such as "Area Ammortizzatori Sociali", "Area Aziende, Lavoratori dipendenti e Fondi speciali", "Area Interventi a sostegno delle fasce deboli", "Area Autonomi e Gestione separata", "Area Invalidità", "Area Pensioni", "Area Lavoratori Domestici, contributi individuali", "Area Bilanci e Contabilità", "Area Controllo di gestione e Business Intelligence", "Area Vigilanza e Risk management", "Area Risorse", "Area Logistica", "Area Sinergie", and "Area Portali (Internet/Intranet/Extranet) e Gestione Documentale".

At the bottom of the sidebar, there are sections for "Cruscotti", "Report combinati", and "Report personali", each with a help icon and a right-pointing arrow.

BIG DATA NEL MONDO DELLA PREVIDENZA E DELLE ASSICURAZIONI

Previdenza, Banche ed Assicurazioni condividono, sotto aspetti differenti, un comune driver: **non tutti vanno in pensione, liquidano il conto corrente, o hanno un sinistro contemporaneamente**. Da ciò deriva una forbice tra «raccolta» ed «impiego» sulla quale il settore privato (Banche ed Assicurazioni) costruisce il proprio utile mentre il settore pubblico la impegna totalmente nelle prestazioni erogate generando **stabilità e benessere**. In ambo i casi sussiste un problema di sostenibilità insito nell'attività stessa. Leggere bene i fenomeni socio-economici, pertanto, è di fondamentale importanza per i suddetti comparti e, giunti allo stato attuale, non si può più prescindere dall'analisi di Big Data secondo criteri di Business e Social Intelligence. I motivi sono tanti, ma, soprattutto, **sono sospinti da una normazione e da una tecnologia che rende possibili cooperazioni prima impensabili e che oggi, a ben guardare, rappresentano opportunità non indifferenti sul versante della crescita sostenibile:**

- **Analisi di sostenibilità.**
- **Liberazione della conoscenza.**
- **Contrasto alle frodi.**
- **Smart-government vs e-government.**

BIG DATA NEL MONDO DELLA PREVIDENZA E DELLE ASSICURAZIONI

I driver di Previdenza, Banche ed Assicurazioni li spingono a dover analizzare problemi che per loro natura:

- **Sono proiettati nel futuro (breve, medio e lungo periodo).**
- **Sono squisitamente esponenziali (NP-completi) perché dipendono dalle azioni e dalle reazioni di una vasta pluralità di soggetti (decisore politico incluso).**

Il caso della Previdenza (e del Welfare), poi, è ancor più critico, vuoi perché è un tema pubblico nella stragrande maggioranza degli Stati, vuoi perché dalla Previdenza e dal Welfare nasce la «Social Security» e quest'ultima **si traduce direttamente in una sorta di «tranquillità istituzionale» che dà ad una collettività la prerogativa di costruire il proprio futuro ricercando buone soluzioni.** Per fare un esempio, in una casa ben solida e funzionante si ha il tempo per pensare ad abbellirla o dotarla dei migliori comfort, in una casa cadente dove dal tetto piove, l'acqua è a singhiozzo e manca la luce non c'è certo il tempo per pensare alla carta da parati! Questa «tranquillità», dunque, **è una «energia decisionale» da mettere in campo nei momenti di crisi e rappresenta «l'oggetto di business» degli enti di previdenza e dell'INPS, nel caso di specie.**

BIG DATA NEL MONDO DELLA PREVIDENZA E DELLE ASSICURAZIONI

Già da tempo grandi gruppi bancari ed assicurativi si sono dotati di Data Warehouses, non foss'altro perché anch'essi, alla luce delle recenti normazioni nazionali ed UE, sono entrati nel mercato dei Fondi Pensione, implementando il c.d. «Terzo Pilastro» del sistema previdenziale.

Anche l'INPS s'è dotato d'un Data Warehouse ben prima che convergessero su di esso INPDAP ed ENPALS. Inizialmente, ciò è avvenuto non per strategia di business (l'INPS non opera in regime di concorrenza) ma per esigenze strategiche ed operative connesse col suo ruolo di principale ente di previdenza e per far fronte ai molteplici compiti cui era chiamato, senza escludere quello di «principale consulente» del decisore politico in materia di politiche pubbliche sulla previdenza e sul welfare.

La recente crisi economica e la convergenza in INPS di INPDAP ed ENPALS hanno introdotto nuovi fattori che nei settori su descritti stanno asseverando ancora di più la necessità di governare il fenomeno «Big Data».

Sul versante assicurativo e bancario s'è presa ormai coscienza che **non si può pensare di fare utile semplicemente utilizzando la forbice tra «raccolta» ed «impiego»**. Il mercato è globalizzato, la concorrenza agguerrita e, soprattutto, **il cliente ha molte informazioni per scegliere, preferendo chi fa «impieghi» che garantiscono stabilità di valore nel tempo (sostenibilità...)**.

BIG DATA NEL MONDO DELLA PREVIDENZA E DELLE ASSICURAZIONI

Sul versante previdenziale e del welfare, invece, s'è altrettanto presa coscienza (vedi le numerose riforme) che **non si può procedere senza una profonda razionalizzazione dell'intero comparto da perseguire non più con tagli o allungamenti dell'età pensionabile, bensì attraverso l'attento monitoraggio di quei meccanismi che ne garantiscono la sostenibilità nel tempo**. Anche in tale caso, la recente crisi economica ha messo in luce un sistema produttivo poco attrattivo per gli investimenti, una burocrazia a volte troppo opprimente, una normativa previdenziale assai complicata che nel tempo ha ingenerato notevoli disomogeneità di trattamento tra varie fasce della società. A tutto questo si aggiunge quello che è un po' più noto grazie ai mass-media: l'allungamento della speranza di vita e l'invecchiamento della popolazione che è senz'altro vero, ma non è l'unico problema del sistema previdenziale italiano. E' noto che il nostro sistema si regge sul meccanismo della ripartizione, dove la popolazione attiva ed occupata attraverso la contribuzione paga quella collocata a riposo. E' altrettanto noto che l'INPS non eroga solo pensioni, ma anche prestazioni (quali CIG e Disoccupazione), che sono comunque direttamente collegate al predetto meccanismo, e prebende assistenziali quali l'Invalidità Civile. Alla crisi, l'INPS ha retto molto bene grazie al suo poderoso sistema informativo ed alla capillarità della sua diffusione. Ma non c'è motivo per vantarsene: nel periodo 2010-2012 è stato autorizzato oltre un miliardo di ore di CIG e molte altre uscite ci sono state per altre forme di sussidio (vedi [RAPPORTO ANNUALE INPS 2012](#)).

BIG DATA NEL MONDO DELLA PREVIDENZA E DELLE ASSICURAZIONI

Sempre sul versante previdenziale, **il periodo 2009-2012 è stato quello in cui praticamente tutti i servizi erogati dall'INPS sono stati telematizzati e questo fatto, riducendo il problema dell'input ed implementando meglio gli archivi gestionali, ha consentito una disponibilità di informazioni unica nel suo genere, vuoi a livello nazionale, vuoi a livello UE.**

Da questa enorme disponibilità, derivante dalle funzioni che è chiamato ad esplicitare (e dal modo di esplicitarle) è naturale che si approdi, anche in INPS, sulla necessità di governare il fenomeno dei «Big Data» affinché questo fatto divenga una concreta risorsa per la collettività amministrata.

ANALISI DI SOSTENIBILITA'

Un sistema è sostenibile se in ogni istante le entrate coprono le uscite. Nel caso previdenziale (come anche in quello bancario/assicurativo) la differenza tra entrate e uscite (o meglio, quella tra «raccolta» e «impieghi») deve contenere un margine positivo che, nell'impreditoria privata garantisce l'utile e nella previdenza pubblica rappresenta la copertura del rischio socio-economico.

Il monitoraggio della predetta differenza è un tema di «Big Data» perché vanno analizzati gli impatti degli «impieghi» sulla «raccolta» e viceversa. Per fissare le idee, un sistema previdenziale che eroga più prestazioni (impieghi) di quanto incassa (raccolta) sarà prima o poi costretto ad alzare il margine di raccolta (aumento dei contributi) o a raffreddare gli impieghi (diminuire strutturalmente in valore e/o numero le prestazioni). Simmetricamente, un sistema assicurativo che liquida troppi sinistri sarà costretto ad aumentare i premi (raccolta) ed a raffreddare in numero e valore la liquidazione dei sinistri. Qualcosa di affine si può dire per il sistema bancario.

Orbene, **«raccolta» ed «impieghi» sono rivolti ad una pluralità di soggetti a loro volta operanti singolarmente all'interno d'un sistema socio-economico con meccanismi di mutua azione e reazione (complessità 2^N).** Senza euristiche su tali meccanismi è impossibile fare buone proiezioni di sostenibilità, senza contare che la rapidità di variazione degli attuali contesti rende non sufficiente i riscontri in coincidenza dei bilanci economici.

LIBERAZIONE DELLA CONOSCENZA

E' quello che spesse volte va sotto il nome di «Open Data». Cosa sono, in effetti? **Sono la mappa funzionale dell'attività di un ente evidenziata e resa disponibile nei limiti consentiti dalla legge. Il senso è quello «matematico» del termine: un ente accetta in input un insieme X di dati sulla collettività amministrata (la domanda) e restituisce un insieme $Y=F(X)$ di dati (la risposta) come servizi alla collettività stessa.** In Y (quindi in F) la sua attività ed il suo impatto sulla società a partire da obiettivi di legge, strategici e/o operativi fissati, rielaborati e consuntivati a cadenze regolari.

Una simile visione (che sintetizza in maniera estrema l'attività d'una istituzione) rende evidente l'importanza del DWH (anzi, dell'EDWH) in INPS: **esso è cassaforte e vetrina di tutta l'informazione processata. In esso, in altri termini, nasce e s'evolve la relazione funzionale $Y=F(X)$ che, per sua mole e natura, va trattata con logiche di «Big Data» ed in sinergia con altre istituzioni.**

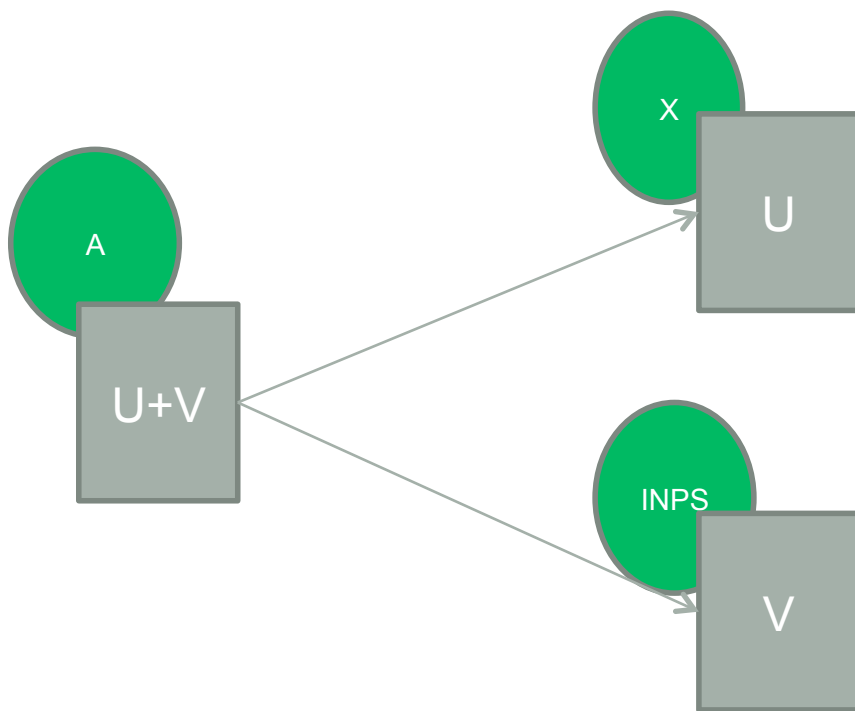
Nel caso di istituti privati, il problema degli «Open Data» non è granché sentito, ma ciò non toglie che le logiche che stanno alla base continuano a rimanere valide, seppur «intra moenia».

CONTRASTO ALLE FRODI

In una moderna collettività strutturata, la moneta ha un valore stabile e si comporta esattamente per come viene intesa nell'accezione comune: **un liquido incompressibile**. Come tale, **essa trabocca dove ce n'è troppa e viene risucchiata dove ce n'è poca per ristabilire l'equilibrio**. Quando ciò non avviene si creano delle spinte di pressione (positiva o negativa) che prima o poi emergeranno. **Le frodi diffuse, prima di essere cattiva cultura o retaggi d'un passato non roseo, sono soprattutto espressione di pressioni dovute a squilibri insiti in un contesto**. Non è semplice scovarle e la telematica non aiuta granché quando i volumi sono alti e le variabili in gioco sono molte. Banche ed Assicurazioni operano all'interno di contesti predefiniti eppure, non ne sono immuni. L'INPS, purtroppo, opera in un contesto aperto e «serve lo Stato» in moltissimi modi ed in moltissimi contesti, ecco perché è così esposto alle frodi (che spesso popolano le pagine di cronaca). **Le frodi sono un problema di «Big Data»? La risposta è Sì perché spesso derivano da un sapersi sapientemente nascondere nelle pieghe d'un sistema complesso dopo averne individuato un punto debole attraverso il quale «parassitarlo»**. Le frodi non sempre appaiono all'interno di un contesto operativo, a volte emergono solo quando si mettono a confronto dati appartenenti ad altri contesti più o meno collegati al primo. Ad esempio, se il tasso di sopravvivenza della popolazione novantenne è del 5% in un determinato luogo (dato per es. ISTAT) un tasso di pensioni pagate a novantenni pari al 20% in quel luogo è un campanello d'allarme...

CONTRASTO ALLE FRODI: UN CASO

La recente normativa (D. Lgs. 82/2005 e s.m. e .i. – Codice dell'amministrazione digitale) incentiva notevolmente la sinergia interenti basata sullo scambio B2B di dati per fini istituzionali. **Questo fatto ha interessanti risvolti sul tema «Big Data» ed, è, tra l'altro, una fonte preziosa di recuperi finanziari, oltre che un modo per asseverare strutturalmente la Legge.** Partiamo da un esempio concreto: una azienda A che versa all'INPS e ad un ente X (es. Comune) una somma **U+V** a titolo di contributi/imposte.



CONTRASTO ALLE FRODI: UN CASO

Si ipotizzi che per un qualsivoglia motivo, la quota **U** in favore dell'ente X venga aumentata di una piccola quantità **u**. Se la precedente dinamica aveva trovato un suo equilibrio e se il messaggio sull'aumento non mette bene in luce la sua necessità, **l'azienda A lo verserà «obtorto collo» e sarà probabile che, pur versandolo ad X, ritenendolo «opportuno», tenti di «scomutarlo» dalla quota V destinata all'INPS.**

In generale, nei rapporti contributivi **una somma viene divisa un molte voci e queste ultime sono memorizzate in archivi spesso diversi, alimentati da conti di gestione altrettanto diversi con complicate sequenze di movimenti (spesso dovute a complicate normazioni e/o convenzioni che si sono stratificate nel tempo)**. A ciò potrebbe aggiungersi l'ulteriore complicazione che la quota **V** può riferirsi ad una condizione marginale controllata indirettamente e in modo confuso (a titolo d'esempio, una riscossione per delega d'una quota contributiva scaturente da una obsoleta convenzione con una associazione datoriale in via di trasformazione/soppressione).

Inutile dire che molti sono i fattori che possono rendere difficoltoso l'individuare «dall'interno» l'anomalia e la cosa migliore **sarebbe quella di controllare simili cose sfruttando le possibilità di incrocio suffragate dalla normazione attuale** (in particolare, l'art. 50 del Codice dell'amministrazione digitale)

CONTRASTO ALLE FRODI: UN CASO

Mettendo, difatti, a raffronto (cioè, incrociando) i dati dell'INPS con quelli di X, si noterebbe che l'aumento percentuale del rapporto non è quello che ci si aspetta. Per logica dovrebbe essere:

$$U + u/V - U/V = u/V$$

mentre, in realtà è:

$$U + u/V - u - U/V = V(U + u) - U(V - u) / (V - u)V = VU + Vu - UV + Uu / (V - u)V = u/V \cdot U + V/V - u$$

In buona sostanza, sarebbe molto più semplice effettuare raffronti con l'esterno (rapporti B2B) che non «scavare» all'interno alla ricerca di un problema che, magari per sua stessa natura, dipende da fattori esogeni. **Questa è la lungimiranza del Codice dell'amministrazione digitale che può estrinsecarsi attraverso la potenza dell'EDWH e di processi di analisi secondo logiche di «Big Data».** Va detto, però, che questo approccio non è stato gran che adottato dal comparto pubblico per svariati motivi, **ma adesso, in pieno momento di ristrettezze, è indubbio che sia diventato una necessità, se non altro per asseverare il principio che ciascun contribuente debba essere messo nelle condizioni di pagare il giusto.**

CONTRASTO ALLE FRODI: ALTRO CASO

Il caso precedente era piuttosto semplice e ve ne sono altri molto più complessi. Non si vuole qui farne un elenco, ma è bene metterne in luce un altro di cui l'INPS è spesso vittima e che si basa sottilmente sul meccanismo della ripartizione, nonché sulla forbice tra «raccolta» e «impiego».

E' noto che **la contribuzione per unità di tempo è molto più piccola della prestazione corrispondente per unità di tempo poiché si presuppone che il numero di percettori sia sempre minore del numero di contribuenti (rapporto di sostenibilità).**

Se un soggetto X riesce a configurarsi fraudolentemente come contribuente (per esempio, attraverso una c.d. «cartiera»), questo potrebbe far versare la quota di contribuzione al soggetto percettore Y **affinché quest'ultimo, spesso dietro illecito compenso al soggetto X, possa percepire, fraudolentemente, una prestazione.** La forbice tra contributo unitario e prestazione unitaria garantisce la fraudolenta utilità del meccanismo, consentendo un illecito guadagno sia ad X che ad Y. Il problema sociale che ne scaturisce è che **il contribuente X non è associato ad alcuna produzione di ricchezza, ma la prestazione, oltre ad essere un costo sociale immediato (specialmente se è della tipologia «a sostegno del reddito») dà vita ad una contribuzione (a volte figurativa) che non è assolutamente ripagata da ricchezza prodotta.** Il valore del contributo unitario è più basso di quello della prestazione non solo per il meccanismo della ripartizione, ma anche per il fatto che ci si attende che la ricchezza unitaria prodotta consenta un investimento adiuvante nella copertura della prestazione unitaria.

CONTRASTO ALLE FRODI: ALTRO CASO

La contribuzione versata senza ricchezza prodotta, quindi, riduce il valore economico delle risorse a disposizione del meccanismo di ripartizione per il futuro, portando il sistema verso l'insostenibilità.

Il governo di una simile situazione non è affatto semplice e non è risolubile attraverso continue riforme restrittive senza incidere su cause strutturali. Queste ultime, però, **vanno cercate nelle dinamiche sociali ed economiche, oltre che nel contrasto materiale alle devianze. Inutile dire che siamo un'altra volta di fronte ad un problema di «Big Data» che passa attraverso l'analisi di variabili interne ed esterne (Social + Business Intelligence).**

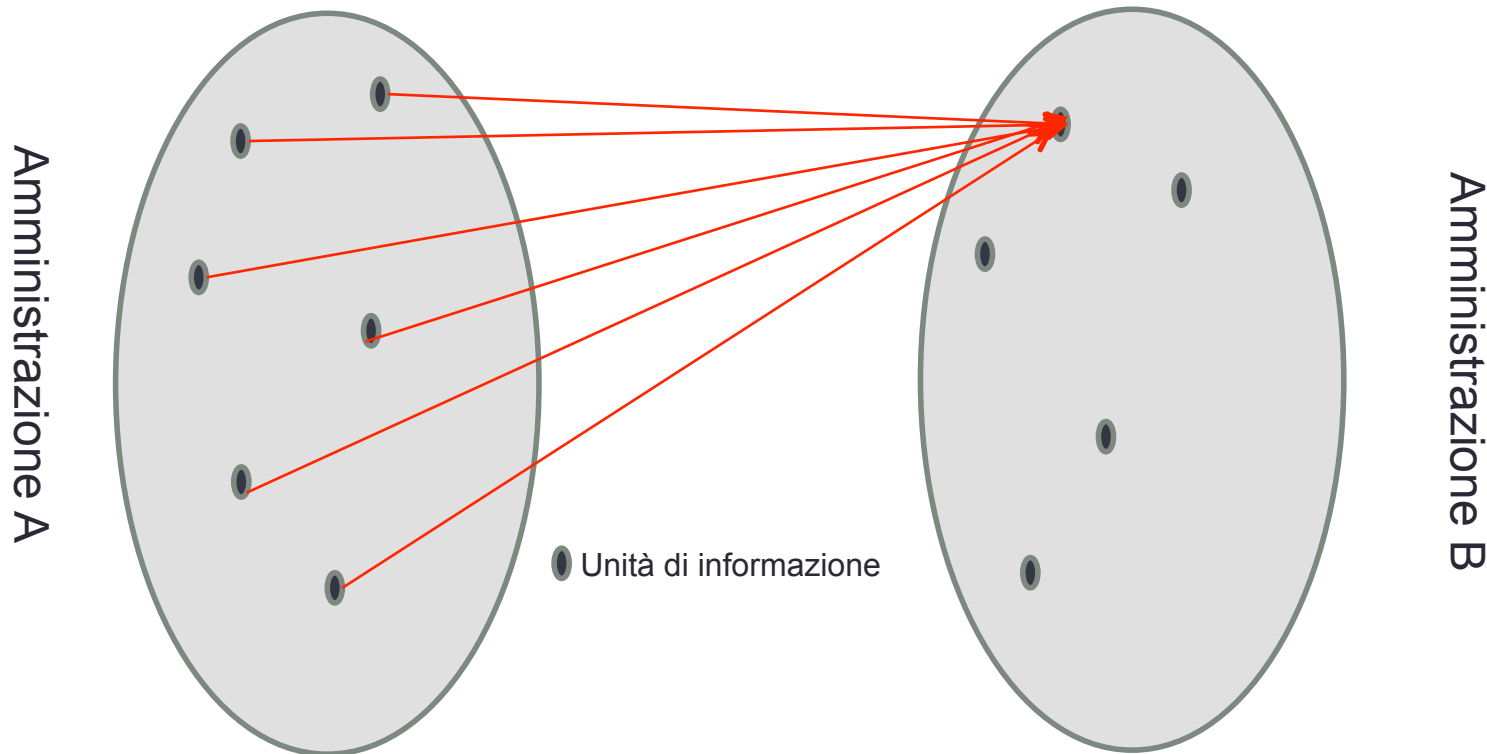
UNA TESI

A tal punto è possibile avanzare una tesi che suffraga pienamente l'utilizzo dei «Big Data»:

- 1. Tutti gli archivi delle Pubbliche Amministrazioni mappano valori direttamente od indirettamente collegati al valore intrinseco della moneta (valore economico).**
- 2. Archivi trattati in maniera indipendente da Pubbliche Amministrazioni distinte sono comunque collegati da relazioni funzionali dirette ed inverse (perché sono moneta, nella sua accezione di fluido incompressibile).**
- 3. Frodi diffuse nascondono spesso disequilibri socio-economici o desideri di illecito arricchimento e si palesano nelle dinamiche che seguono le relazioni funzionali di cui al punto 2.**

UNA IMPLICANZA DELLA TESI

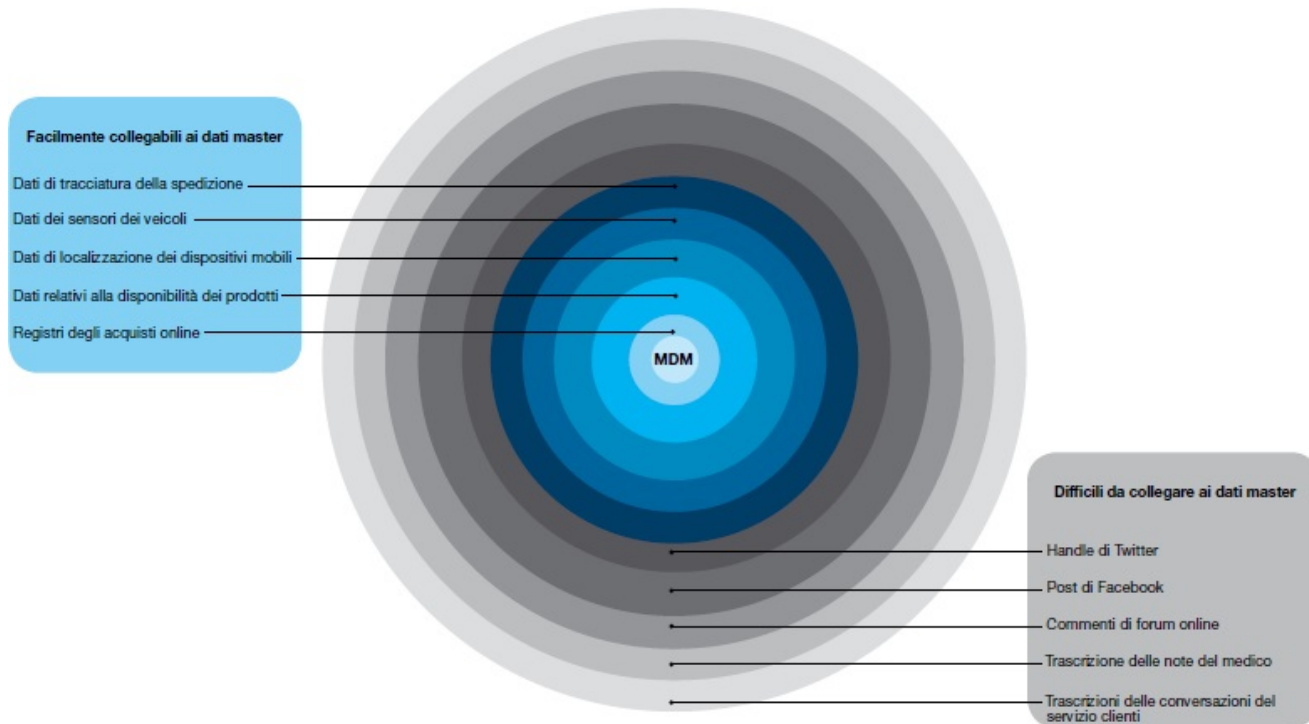
Se gli archivi della Pubblica Amministrazione sono tutti interdipendenti e se ciascuno di essi è suddiviso in unità di informazione logicamente legate a processi distinti (assimilabili ai campi), le funzioni che li collegano e le loro inverse vanno pensate (e costruite) come **la mappatura dei singoli contributi (positivi, negativi o nulli) che ciascuna unità di informazione dell'archivio di partenza (amministrazione A) dà a ciascuna unità di informazione dell'archivio d'arrivo (Amministrazione B).**



UNA IMPLICANZA DELLA TESI

Ciascun «contributo» dato da una unità di informazione di A, detta X_i , ad una di B, detta Y_j , se non nullo e preso per tutte le unità di informazione di A, rappresenta un impatto di X_i su Y_j e rappresenta l'immagine, tramite una relazione funzionale F da A a B, di X_i su Y_j .

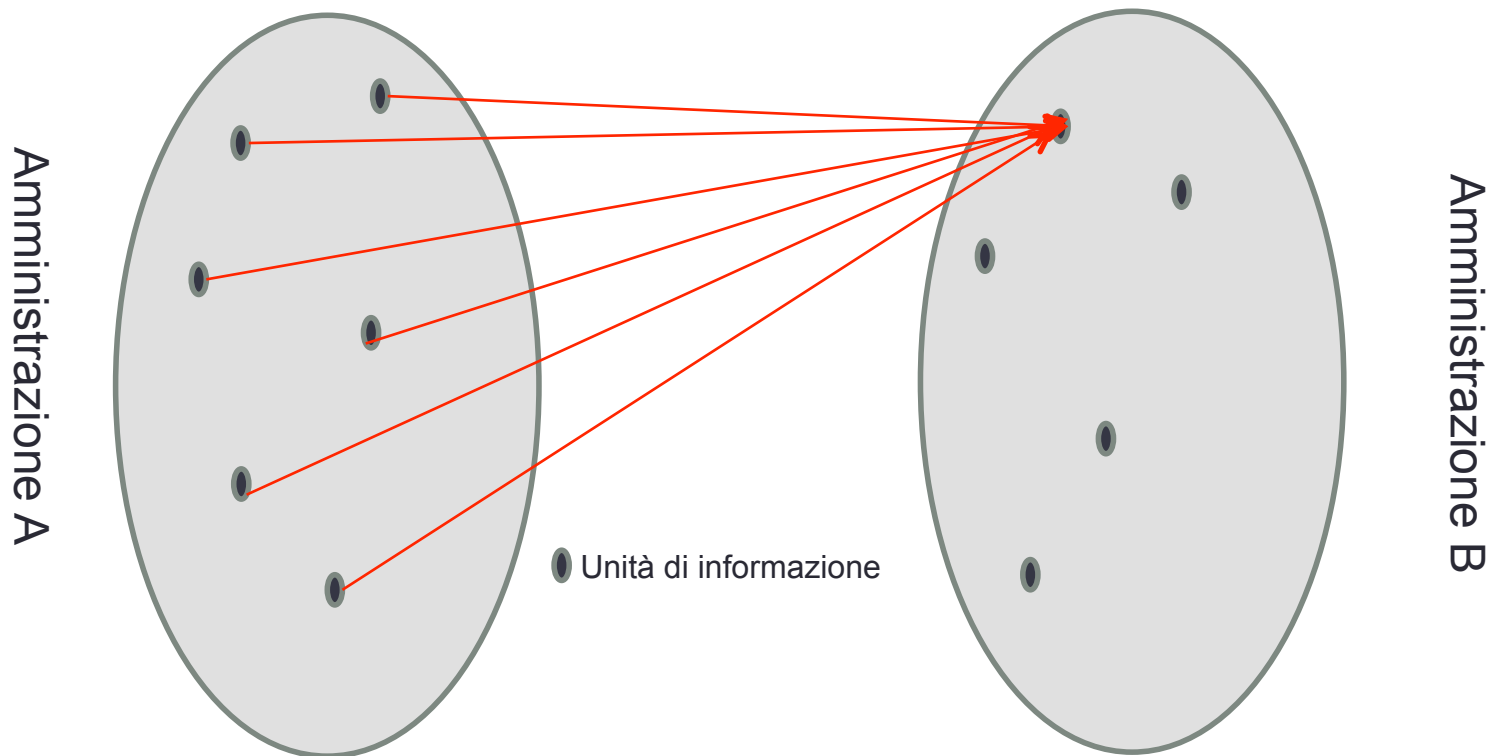
Per definizione, punti di A che danno contributo ad un punto di B hanno la stessa immagine tramite F (regola per «costruire» la relazione funzionale). La funzione F rappresenta un collegamento tra i dati «master» di A e i dati «master» di B.



UNA IMPLICANZA DELLA TESI

Detto N il numero di unità di informazione di A ed M il numero di unità di informazione di B , il numero totale di funzioni dirette ed inverse tra A e B è dato da:

$$M \uparrow N + N \uparrow M$$



UNA IMPLICANZA DELLA TESI

Per due amministrazioni con solo 10 unità di informazione ciascuna, tale numero è già:

$$10^{10} + 10^{10} = 2 \cdot 10^{10}$$

Cioè **venti miliardi di modi secondo i quali possono «farsi del bene» o «farsi del male» a seconda dei comportamenti propri e di chi fruisce dei loro servizi (o li determina)**. Ecco il perché siamo di fronte ad un problema squisitamente di «Big Data».

Ciascuna funzione è una potenziale «euristica» nell'EDWH e da ciascuna euristica è possibile determinare comportamenti virtuosi o frodi. Non è assolutamente pensabile mapparle tutte, vuoi perché sono veramente tante, vuoi perché alcune logicamente possono non aver senso nell'economia complessiva. Di certo, sotto di esse si nasconde il funzionamento interrelato dell'intero Paese, funzionamento che in molti casi è sicuramente da rivedere, semplificare, ottimizzare e razionalizzare per superare a pieno l'attuale crisi (che ha tutta l'aria di prefigurare una transizione epocale di modello socio-economico). Vale la pena, dunque, investire su tale percorso quando le condizioni lo impongono ed il mondo previdenziale pubblico e del welfare, oltre ad averne bisogno, **troverebbe nell'INPS un grande banco di prova ed una vera «miniera» di risultati.**

VANTAGGI DELLA «BIG DATA ANALYTICS»

Trovare le euristiche è già difficile, **se poi le si usa «al momento» dimenticandosene nel futuro (scarsa attenzione al riuso) non si realizza un granché se non eventi «spot» che sfociano nell'autoreferenziale**. Se invece vengono stratificate e valorizzate nel tempo secondo una policy guidata dalla missione aziendale esse costituiscono un valore che si incrementa nel tempo. Ecco perché l'utilizzo/esposizione intelligente degli «Open Data», ecco perché l'utilizzo di EDWH per enti che svolgono funzioni ad altissimo impatto sociale (Business + Social Intelligence) ed ecco perché questi ultimi sono naturali candidati per implementazioni concrete di analisi di «Big Data».

Il meccanismo della progressiva ricerca e stratificazione delle euristiche porta ad un risultato molto auspicabile nel convulso contesto attuale: il passaggio dall'e-government allo smart-government.

E-GOV vs SMART-GOV

- **E-government significa utilizzare la telematica per semplificare l'accesso ai servizi e per ridurre il problema dell'input (trasferimento da un supporto cartaceo ad uno elettronico).**
- **Smart-government significa anticipare le scelte e i desideri di una collettività per conservare energie decisionali da impegnare nella ricerca di una crescita duratura e non minata da tensioni interne od internazionali.**

E' naturale che il suddetto passaggio si realizza solo e soltanto quando le euristiche sull'informazione conosciuta (grazie all'EDWH ed all'analisi «Big») consentono di fare previsioni e strategie d'azione attendibili sul medio-lungo periodo.

I RISCHI

Secondo Darin Bartik, executive director di Dell Software, la maggior parte dei progetti relativi ai «Big Data» vanno incontro al fallimento. Ma la strada per il successo esiste e passa attraverso decisioni su temi di grandissimo respiro. Tra le principali cause, le seguenti:

- I dipartimenti IT e quelli amministrativi non sono allineati sui problemi da risolvere, si limitano ad affrontare le questioni da un punto di vista puramente tecnologico.
- L'accesso ai dati è spesso limitato e altrettanto spesso non si riesce ad accedere ai dati che servono per implementare le euristiche.
- Molte delle tecnologie, approcci e discipline relative ai «Big Data» sono nuovi, e c'è spesso resistenza al cambiamento. Tutti condividono che è giusto, anzi, necessario tuffarsi nella ricerca di euristiche ed «incroci», poi si ricade in timori e ritrosie perché, in effetti, analizzare «Big Data» significa entrare molto a fondo nei temi. D'altronde, i volumi sono enormi ed un errore di valutazione, anche piccolo, può essere disastroso. La condivisione «a parole» è assolutamente insufficiente.

L'APPROCCIO CONSIGLIABILE

Approcciare in maniera giusta il problema, dunque, è fondamentale da un lato per convergere al successo di un progetto «Big Data», dall'altro per mettere in campo le giuste condivisioni interne (ed esterne). Vero che la storia ce lo sta imponendo, ma è pur vero che non è detto (anzi è sconsigliato) che bisogna «agire con l'ascia» senza prevedere la giusta gradualità. **L'approccio «Big Data» nel mondo previdenziale pubblico presuppone analisi sistemiche (con conseguenti azioni) che travalicano le mura dell'INPS ed atterrano spesso volte sul tavolo del decisore politico.** Indubbio, quindi, l'uso della giusta gradualità che una Tecnostruttura deve incarnare (non dimentichiamo che la PA è l'espressione del Potere Esecutivo, cioè quello che «applica» le Leggi e che, quindi, ne decide la gradualità).

Una delle regole maestre, se si intende avviare un progetto di «Big Data» ritenuto necessario è **non chiamarlo o targarlo col termine «Big Data», ma rammentare, con concretezza, il problema pratico e reale per il quale si avvia il progetto.** «Big Data» è una metodica, non una bacchetta magica, uno slogan o qualcosa di «rivoluzionario»: **è il giusto epilogo di organizzazioni che, per loro natura lavorano basandosi su proiezioni, utilizzando grandi volumi di dati polimorfi e provenienti da più fonti.**

L'APPROCCIO CONSIGLIABILE

In sintesi:

- **Iniziare dai problemi che si desidera risolvere.** Non affrontare un problema enorme, ma partire da un progetto piccolo relativo a una questione specifica per procedere, poi, in modo incrementale (i DWH lo consentono). Redigere una lista delle domande e non perdere di vista l'obiettivo preoccupandosi troppo della tecnologia. Assicurarsi che il team sia circoscritto (non voluminoso) e che tutti gli stakeholder concordino sull'obiettivo.
- **Avere il consenso dall'alto.** Una volta identificato il problema da risolvere, occorre il giusto consenso per accedere ai dati necessari ed attivare le euristiche.
- **Avere un team fortemente collaborativo con skill adeguato.** Dal punto di vista ideale ci vuole qualcuno che comprenda il machine learning (gestione formale della conoscenza, perché le euristiche sono collegabili in rete semantica), abbia le competenze scientifiche ed analitiche necessarie, e sia in grado di lavorare con i dati per ottenere i risultati richiesti (query).

CONCLUDENDO...

Pur sembrando la contraddizione di quanto sinora detto, **non c'è granché di nuovo nei «Big Data» e se si comprende questa loro, tutto sommato, «normalità» li si può utilizzare come un normale «attrezzo IT»**. I «Big Data» altro non sono che rappresentazioni di fenomeni a complessità esponenziale anziché a complessità polinomiale (anche se molto grande). Il problema che è insito in loro è che mostrano comportamenti governabili più con l'analisi di domini continui (non numerabili) anziché di domini discreti (numerabili). L'indice di questo comportamento sta nell'esponenziale contrapposto al polinomiale (che in matematica rappresenta proprio il confine tra continuo e discreto). Nelle comuni Basi Dati, fenomeni caotici sono rovinosi perché significano la negazione del modello adottato, nei «Big Data» possono essere molto frequenti perché in un sistema a complessità esponenziale piccole variazioni possono generarne enormi. Tutto qui, dunque! Le nuove tecnologie che si stanno affacciando all'orizzonte sono, sostanzialmente, riedizioni più capaci, intelligenti e veloci di quelle attuali. Quel che è importante del mondo «Big Data» e che attraverso di esso possono essere trovate nuove soluzioni per problemi che tanto nuovi, purtroppo, non sono...

CONCLUDENDO...

C'è chi afferma che il fenomeno «Big Data» è stato «gonfiato» senza comprenderne appieno tutte le implicazioni del caso. E' stato come per il cloud e, come per il cloud, i principi di base dei big data stanno cambiando radicalmente il modo in cui le aziende saranno in grado di reagire o di anticipare le opportunità di business (Peter Prowse, General Manager for Data Centre Solutions di Dimension Data - Australia). Forse è vero, di sicuro vale una «proporzione profonda»: **i «Master Data» stanno ai numeri razionali come i «Big Data» stanno ai numeri irrazionali**. Questo fatto «matematico» motiva un approccio collaborativo al fenomeno: **più sorgenti da analizzare si hanno a disposizione e più è facile «avvicinarsi» ad una buona soluzione per un problema di business, o meglio, di sostenibilità socio-economica**. In generale nel vecchio continente, ma ancor di più in Italia, dove una legge abilita la collaborazione tra pubbliche amministrazioni (art. 50 D. Lgs 82/2005 – Codice dell'amministrazione digitale) questa opportunità è reale e rappresenta, ad avviso di chi scrive, il vero valore aggiunto dei «Big Data» che l'attuale crisi ci impone di cogliere.

RIFERIMENTI

- [Ecco quant'è tutta l'informazione del mondo](http://www.lescienze.it) – www.lescienze.it - 11 febbraio 2011
- [The World's Technological Capacity to Store, Communicate, and Compute Information \(Martin Hilbert, Priscila López\)](http://www.sciencemag.org) – www.sciencemag.org -10 febbraio 2011
- [Da Facebook ai batteri, come crescono le reti complesse](http://www.lescienze.it) – www.lescienze.it -13 settembre 2012
- [UNIVERSITA' STATALE DI MILANO: Non sono sei ma quattro i "gradi di separazione" su Facebook](http://www.lescienze.it) – www.lescienze.it - 22 novembre 2011
- [La borsa crollerà? La risposta è in Google Trends](http://www.lescienze.it) – www.lescienze.it - 26 aprile 2013
- [Quantifying Trading Behavior in Financial Markets Using Google Trends](http://www.nature.com/srep/) - www.nature.com/srep/ - 25 Febbraio 2013
- [Bilancio Sociale 2011 INPS](http://www.inps.it) – www.inps.it
- [Rapporto annuale 2012 INPS](http://www.inps.it) – www.inps.it