

Come valorizzare i dati nell'era post-digital

di Severino Meregalli, Leonardo Maria De Rossi, Lorenzo Diaferia*

Il World Economic Forum (WEF) ha recentemente stimato che entro la fine del 2020 l'ammontare di dati memorizzati a livello mondiale dovrebbe raggiungere il volume di 44 Zettabytes, il che significa disporre di una quantità di byte quaranta volte superiore al numero di stelle presenti nell'universo.

Nel mondo dei big data e della digitalizzazione, i riferimenti immaginifici come quello del WEF sul trend di crescita del volume dei dati sono oramai diventati la norma. Centri di ricerca, vendor e pubblicitaria fanno a gara nel segnalare con similitudini e metriche

di forte impatto narrativo il tema della *data explosion*. Se è innegabile che ci troviamo di fronte a un trend inarrestabile e di dimensioni rilevanti, il tema fondamentale diventa sempre più capire come iniziare a usare questi dati ed estrarre valore dagli stessi. Su questo fronte, il DEVO Lab della SDA Bocconi School of Management ha condotto una serie di ricerche, l'ultima delle quali su stimolo e supporto di Google Italia (1).

LE TRE DOMANDE FONDAMENTALI

Lo studio ha coinvolto l'analisi di oltre 20 iniziative di valorizzazione dei dati condotte prevalentemente in Italia e in Europa.



© iStock - Z. Wei

© Egea SpA - TUTTI I DIRITTI SONO RISERVATI

* *Severino Meregalli* è Associate Professor of Practice di Information Systems presso SDA Bocconi School of Management e Coordinatore scientifico del laboratorio di ricerca Digital Enterprise Value and Organization (DEVO Lab). *Leonardo De Rossi* è Junior Lecturer del Knowledge Group Information Systems-IT presso SDA Bocconi School of Management ed è parte del gruppo di lavoro del laboratorio di ricerca DEVO Lab. *Lorenzo Diaferia* è Junior Fellow del Knowledge Group Information Systems-IT presso SDA Bocconi School of Management e fa parte del gruppo di lavoro del laboratorio di ricerca DEVO Lab.

Attraverso questa comparazione, sono emerse tre domande fondamentali per strutturare progetti meno basati sull'*hype* della crescita esponenziale dei dati e più focalizzati sulla generazione di un ritorno economico misurabile.

Le domande chiave in questa prospettiva sono:

- chi sono i principali attori per la buona riuscita del progetto?
- qual è la natura di questi attori? Si tratta esclusivamente di soggetti privati o può essere invece necessaria qualche forma di cooperazione pubblico-privato?
- nel corso del progetto, quali sono i passi da seguire per generare un tangibile valore economico?

LA COOPERAZIONE COME BASE PER LA PRODUZIONE DI VALORE

Fenomeni pervasivi come quello della *data exploitation* suggeriscono come le iniziative che ruotano intorno all'uso massiccio di dati arrivino a coinvolgere molteplici fonti e attori, che si combinano tra loro nei modi più disparati. In molti casi, la realizzazione di un'iniziativa di valorizzazione economica richiede forme di cooperazione e partnership all'interno di un vero e proprio «ecosistema dei dati», dove attori privati (aziende) e pubblici (enti e organizzazioni) possono trovare reciproci vantaggi.

Prendiamo l'esempio di Arup, società che opera nel campo dell'ingegneria e del design edile che da tempo utilizza dati pubblici nell'ambito dei suoi progetti su impatto ambientale, valutazione dei rischi urbanistici e per l'*urban planning*. Un secondo esempio è quello di Waze Connected Citizens program, progetto lanciato nel 2014 con l'obiettivo di tenere costantemente aggiornati gli utenti sulla viabilità urbana e che conta oggi più di 500 partner globali tra agenzie governative, città, enti no profit e operatori di primo soccorso.

Gli enti pubblici che partecipano all'iniziativa forniscono dati fondamentali per mantenere elevato il livello di servizio offerto dall'app di navigazione Waze. In cambio, le amministrazioni traggono vantaggi da un potente strumento che permette una più efficiente gestione dei network stradali e un miglior monitoraggio dei flussi dei viaggiatori.

LE DIVERSE TIPOLOGIE DI DATI

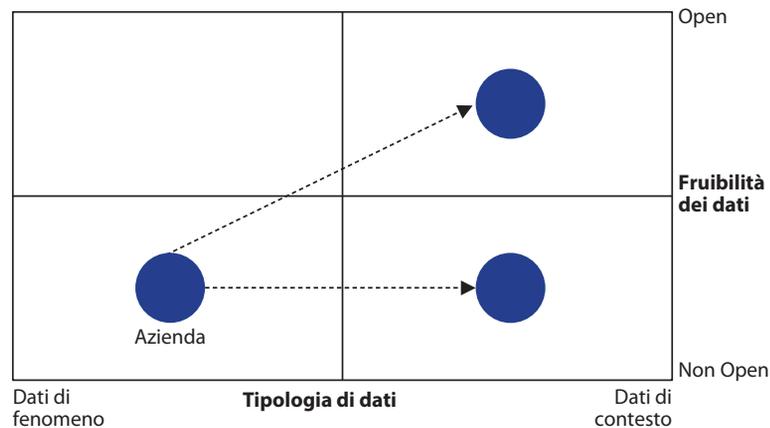
Da cosa scaturisce la necessità di realizzare delle filiere in cui sono coinvolti attori eterogenei? Ciò è spesso dovuto alla natura stessa dei dati richiesti per ottenere risultati concreti; ossia, di iniziative che non siano motivate dall'effetto moda, ma dalla reale tensione a ottenere valore tenendo conto di quanto serve per trarre un effettivo vantaggio. A questo proposito, una distinzione fondamentale è quella tra «dati di fenomeno», ossia quei dati che descrivono un evento specifico (per esempio, il volume di vendite di gelati in un bar), e i relativi «dati di contesto», fondamentali per un'interpretazione consapevole e non distortiva (per esempio, la temperatura nella zona di vendita dei gelati). A ciò si aggiunge un altro elemento: non tutti i dati sono fruibili allo stesso modo. Alcuni sono in formato open, ossia accessibili da fonti pubbliche o comunque aperte a tutti, altri, spesso la maggior parte, sono un patrimonio privato e custodito da aziende o altri enti (Figura 1).

Facciamo l'esempio di un'azienda che, partendo dall'utilizzo di dati di fenomeno di cui dispone, decida di avviare un progetto per generare valore economico. In questo caso, a seconda dei fini del progetto, può rendersi necessario un arricchimento dei dati di partenza con dati di contesto, complementari e fondamentali per generare risultati significativi. Le opzioni disponibili in tal senso sono due: rivolgersi ad altre aziende che vendono tali informazioni, oppure utilizzare piattaforme (se presenti) che permettano l'accesso a dati in formato

open. Laddove disponibili, i dati di contesto creano un valore più elevato rispetto a un utilizzo esclusivo di informazioni di fenomeno. Si può dire che i primi siano una vera e propria leva di valore e, in alcuni casi, senza di questi non è possibile generare valore (pensiamo ai dati di vendita senza dati sull'andamento congiunturale e sulle azioni dei concorrenti).

Che cosa significa tutto questo per aziende e regolatore pubblico? Se per le prime questo significa dare il giusto peso a tutte le categorie di informa-

FIGURA 1 | FRUIBILITÀ E TIPOLOGIA DEI DATI



zioni necessarie a instaurare un rapporto strutturato con gli attori in gioco per creare una filiera di contenuti ben allineati all'obiettivo, per chi si occupa di policy-making; invece, questo significa prestare particolare attenzione alla creazione di dati di contesto, stimolando l'emergere di soggetti, pubblici o privati, che favoriscano l'accessibilità in ottica open, promuovendo un quadro legislativo che incentivi anche il riutilizzo di open data pubblici.

COME ESTRARRE VALORE ATTRAVERSO PROCESSI STRUTTURATI

Per la realizzazione di un progetto dati è necessaria l'attività coordinata di alcune classi di attori chiave. Soffermarsi su questi soggetti consente di porsi le giuste domande per una sistematica individuazione dei protagonisti in gioco:

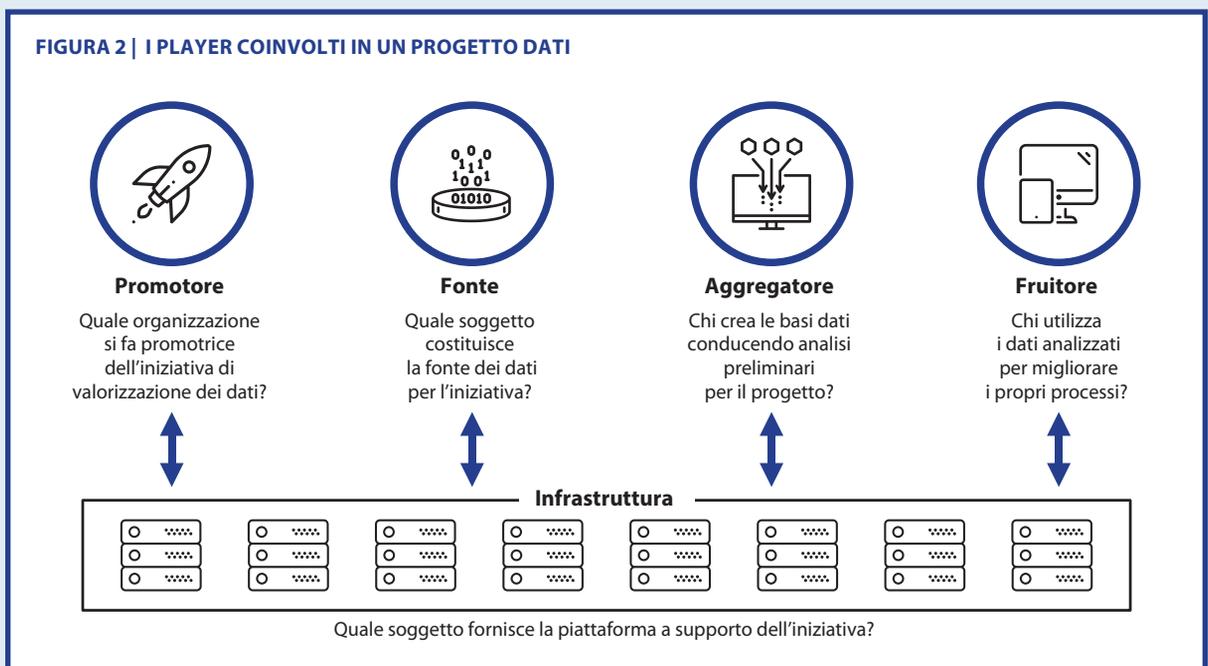
- il *promotore* dell'iniziativa rappresenta il vero sponsor del progetto, che svolge una funzione di attivatore e coordinatore degli altri soggetti coinvolti;
- la *fonte* è il soggetto, fisico o umano, che con le proprie attività genera i dati necessari per supportare l'iniziativa e realizzare l'obiettivo prefissato;
- l'*aggregatore* riunisce le basi dati fondamentali per il progetto, si occupa della loro organizzazione e delle attività di analisi preliminare;
- il *fruitore* rappresenta l'utilizzatore finale, che sfrutterà i risultati del progetto per migliorare i propri processi, in modalità statica (*una tantum*) o dinamica (continuativa);
- il *fornitore dell'infrastruttura* svolge invece un ruolo trasversale e abilitante per la valorizzazione dei dati.

È questo soggetto, infatti, a fornire la piattaforma necessaria per supportarne le attività.

La presenza e la cooperazione sinergica tra le cinque categorie di attori descritte si rivela cruciale. Se il promotore di un progetto è il vero soggetto attivatore dell'iniziativa, è altresì chiaro che questo player centrale non potrà non prevedere l'intervento degli altri quattro ruoli chiave per la buona riuscita dell'iniziativa, che, per realizzarsi, dovrà trovare risposta ai seguenti interrogativi (Figura 2): da dove provengono i dati e chi ne è responsabile? Chi aggrega le basi dati, spesso provenienti da diverse fonti? Chi fornisce l'infrastruttura abilitante? Chi saranno gli utilizzatori delle informazioni finali? Quali sono le loro esigenze e priorità?

Tra questi ruoli, va sottolineato quello trasversale, rispetto alle fasi del processo, dell'infrastruttura tecnologica, che abilita la gestione e le analisi dei dati. Se è vero che il processo di valorizzazione passa per un'attenta valutazione dei soggetti che dovranno presidiare queste cinque aree chiave, è altrettanto vero che spesso tali ruoli non possono essere ricoperti dalla stessa organizzazione, né da organizzazioni della medesima natura. Anche da qui nasce la necessità di prevedere e incentivare forme di cooperazione tra soggetti pubblici e privati per fornire, a seconda dei casi, i dati e le capacità necessari.

Perché le iniziative di valorizzazione dei dati abbiano successo, non solo è necessario che diverse classi di attori portino il proprio apporto e le proprie competenze al progetto, ma anche che le attività di tali soggetti siano coordinate e organizzate attraverso un processo strutturato.



DAI DATI GREZZI ALLA CREAZIONE DI INFORMAZIONI IN FUNZIONE DELL'OBIETTIVO DEL PROGETTO

La rielaborazione DEVO Lab del tradizionale modello della data value chain rende evidenti i passi che è necessario percorrere e presidiare per passare dai dati grezzi alla creazione di informazioni. Le fasi chiave sono quattro:

- *generazione*, risponde all'interrogativo di come vengono generati i dati che alimentano il processo;
- *organizzazione e gestione qualità*, riguarda il processo di *data cleaning* e di aggregazione dei dati grezzi attraverso il supporto di una piattaforma;
- *analisi e visualizzazione* è il passaggio cruciale per l'estrazione di informazioni che verranno poi impiegate nei *processi decisionali attraverso opportune analisi e modalità di visualizzazione*;
- *applicazione e utilizzo*, coinvolge l'utilizzo finale di quanto realizzato per l'obiettivo interno o esterno prefissato (Figura 3).

Nei progetti di valorizzazione dei dati, a guidare la value chain deve essere l'obiettivo finale che determina cosa serve per la generazione di valore economico. È in funzione di questo obiettivo che la catena del valore e gli attori in campo si devono articolare e integrare in un progetto unitario.

ESEMPI DI ORGANIZZAZIONE DEGLI ATTORI INTORNO ALLA DATA VALUE CHAIN

Il progetto sperimentale lanciato dal Comune di Milano, con l'obiettivo di contribuire a un più efficace marketing territoriale per la città, costituisce un chiaro

esempio di come obiettivi, attori e processi siano tre elementi complementari da allineare (Figura 4).

L'obiettivo del progetto, attualmente in fase di sviluppo, è realizzare uno strumento che permetta di ottenere dati puliti e coerenti per indirizzare in modo più mirato le attività di marketing e comunicazione della città di Milano. Per fare ciò, sono imprescindibili dati che permettano di conoscere meglio gli «utilizzatori» della città, comprendendone comportamenti e abitudini, che vadano oltre a quelli identificabili tramite dati pubblici (per esempio dati ISTAT, permanenze in hotel ecc.). Attraverso tali informazioni, si profila la possibilità di uno strumento di matching tra comportamento della città e audience a livello globale.

Questo ambizioso obiettivo richiede, oltre al Comune di Milano, il coinvolgimento di diversi partner privati, fondamentali come aggregatori delle basi dati generate dai comportamenti dei singoli individui. Attualmente, tra i principali figurano SEA, ATM, Vodafone, Tim, MasterCard, Google, Fiera Milano, Ernst & Young. Perché lo strumento sia in grado di fornire un quadro non distortivo ed efficace di fenomeni e comportamenti si renderà senz'altro necessaria una graduale espansione di tali soggetti. Tutti gli attori, pubblici e privati, presidiano diverse fasi della catena del valore descritte in Figura 5. Per questo, capacità chiave per il promotore dell'iniziativa è la selezione, l'allocatione e l'orchestrazione dei partner nel modo più funzionale possibile all'obiettivo finale.

Per tornare a un contesto più aziendale, una simile necessità di allineamento di attori e processi si ritrova in molti altri casi.

FIGURA 3 | LA RIELABORAZIONE DEVO LAB DELLA DATA VALUE CHAIN



Come vengono generati i dati?

I dati che alimentano il processo di valorizzazione vengono generati dalle diverse fonti coinvolte. A seconda delle situazioni, possono avere diverse caratteristiche (per esempio: IT, IoT) e provenire da un singolo soggetto o da più parti, pubbliche o private.

Come vengono aggregati i dati?

Tipicamente i dati grezzi, spesso provenienti da più fonti e con caratteristiche non omogenee, necessitano di subire un processo di *data cleaning* e aggregazione attraverso il supporto fornito da una piattaforma di gestione.

Come vengono analizzati i dati?

La fase di analisi e visualizzazione è cruciale per l'estrazione di informazioni a supporto dei processi di decision-making e avviene tramite la definizione di opportuni algoritmi e il supporto di piattaforme per l'analisi e dashboard di visualizzazione.

Come vengono utilizzati i dati?

I dati adeguatamente aggregati, puliti, analizzati e visualizzati sono pronti per essere utilizzati per realizzare l'obiettivo prefissato, internamente (per esempio: progetto big data aziendale) o esternamente (predisposizione di portali open data) al soggetto promotore.

FIGURA 4 | LA DATA VALUE CHAIN E LE CLASSI DI ATTORI COINVOLTI NEL PROGETTO DEL COMUNE DI MILANO

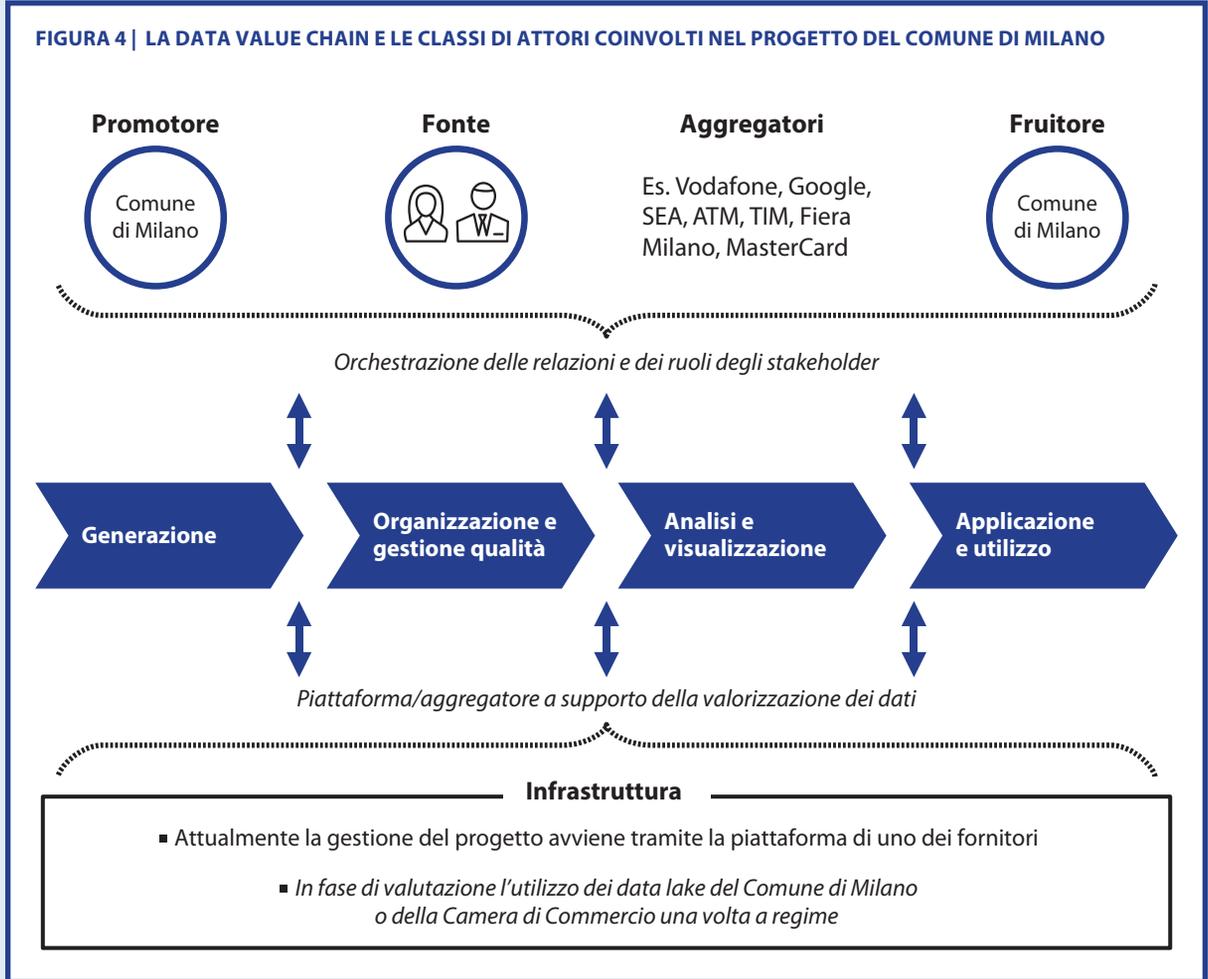


FIGURA 5 | LE FASI DELLA DATA VALUE CHAIN NEL PROGETTO DEL COMUNE DI MILANO



Per esempio, nel 2015 Trenitalia ha iniziato il progetto Dynamic Maintenance Management System (DMMS) che combina sensori *IoT*, data analytics e *in-memory computing* per la manutenzione predittiva. Gli obiettivi del progetto sono una riduzione dell'8 per cento dei costi di manutenzione della flotta Trenitalia entro il 2020/2021. A ciò si aggiungono impatti positivi, attualmente difficilmente quantificabili, sugli utilizzatori finali dei servizi (per esempio diminuzione dei guasti, maggiore disponibilità di treni, maggiore puntualità). A oggi, è stato prodotto un risparmio osservabile sui costi di mantenimento di circa il 4 per cento, per un valore economico corrispondente di 16 milioni di euro. Per la realizzazione di questo obiettivo, Trenitalia si avvale della collaborazione di SAP, AlmamivA e Scuola Superiore Sant'Anna. Questi soggetti forniscono infrastrutture, competenze e servizi fondamentali nelle fasi di organizzazione, gestione qualità, analisi e visualizzazione dei dati IT (allestimento dei treni, manutenzioni subite) e *IoT* (tra le 1000 e le 5000 segnalazioni al secondo per ogni treno) di Trenitalia.

Per porsi sempre più nella condizione di estrarre effettivo valore dai dati, da marzo 2018 Acea sta conducendo il progetto Effective Dashboard, rivolto a fornire uno strumento efficace e un punto di accesso unico ai dati per supportare le decisioni di business. Tale progetto si basa sulla valorizzazione di tre categorie di dati: quelli interni, provenienti dalle funzioni di business e dai sistemi real time di campo; quelli di contesto, come notizie o dati meteorologici; quelli provenienti dai social media di Acea. Nella fase di organizzazione e controllo qualità e come supporto infrastrutturale, l'azienda si avvale di partner esterni (Talend, Cloudera, SAS, Qlik). L'iniziativa sta organizzando e valorizzando il grande volume di reportistica aziendale per fornire informazioni sempre aggiornate su dati interni (relativi per esempio alle divisioni HR, amministrativa,

manutenzione) e di contesto (situazione meteorologica, notizie inerenti al business ecc.). Già oggi, i costi associati alle estrazioni e alla reportistica risultano significativamente diminuiti grazie ad una maggiore facilità nel reperire informazioni autonomamente in modalità self service da parte di top e middle management.

L'IMPORTANZA DEI DATI DI CONTESTO IN FORMATO OPEN

Le tre esperienze italiane descritte in precedenza mostrano, nonostante i diversi ambiti applicativi e stati di avanzamento, la chiara necessità per i promotori delle iniziative di trattare in modo organico e coordinato:

- l'organizzazione del flusso della data value chain in funzione di un obiettivo chiaro;
- la valutazione delle tipologie di dati necessarie e della relativa natura interna o esterna, privata o pubblica;
- l'assegnazione di tutti i ruoli fondamentali lungo la catena, secondo le competenze disponibili.

Il rapporto tra promotore di un'iniziativa, fonti dati, aggregatori, fruitori e infrastruttura a supporto necessita infatti di essere inserito in un quadro di cooperazione più ampio. Un aiuto in questo senso può giungere dall'incentivazione a una più facile integrazione tra dati di fenomeno e di contesto tramite adeguate politiche pubbliche che incoraggino la creazione, il mantenimento e il riutilizzo di dati in formato open di alta qualità.

In questo contesto, un aiuto può poi giungere da quelle aziende che, per vocazione digitale e disponibilità di infrastrutture, si trovano a maneggiare grandi quantità di dati contestuali. Queste si possono candidare, o vanno comunque stimolate, a fungere da aggregatori o infrastrutture di base per progetti dati. Senza questa componente la trasformazione dei dati di fenomeno in valore risulterà spesso vana.

(1) Per ulteriori informazioni riguardo ai casi analizzati nel contributo e alla strumentazione sviluppata si rimanda al tool online accessibile al seguente link che espone in maniera dettagliata il progetto di ricerca DEVO Lab: http://devolab.sdbocconi.it/Ricerca_valorizzazione_dati/.

IN SINTESI

- I dati non sono tutti uguali. La loro natura, di fenomeno o di contesto, e fruibilità, open o non open, rendono fondamentale la cooperazione tra diverse classi di attori, sia pubblici sia privati.
- In un progetto legato all'uso dei dati devono intervenire molti attori. La chiara identificazione di ruoli e responsabilità e l'allineamento di tutti i player in un processo strutturato sono pilastri fondamentali per la realizzazione del progetto.
- I dati di contesto in formato open sono importati abilitatori: la loro creazione e il loro mantenimento possono incentivare i progetti dati.