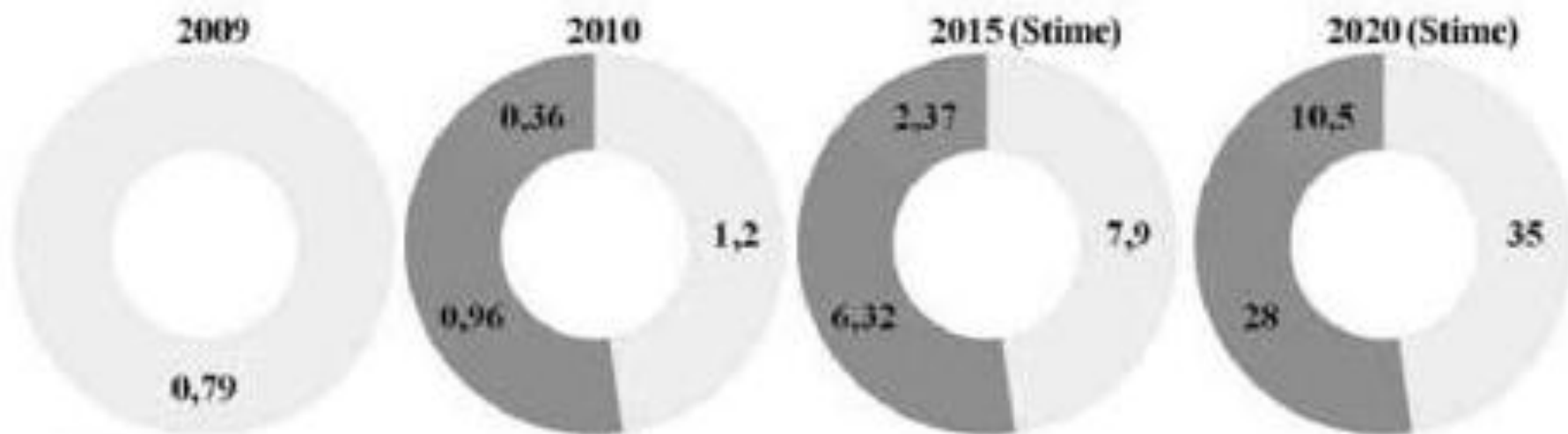


Cenni sui big data

Big Data

- Rappresentano uno dei fattori evolutivi nel mondo dell'analisi dei dati e della BI.
- La rete è una condizione necessaria per raccogliere e condividere i dati e grazie alle nuove tecnologie di raccolta, storage ed elaborazione e alla maggior capacità analitica e interpretativa di cui le imprese possono dotarsi, oggi le informazioni trattate si aggirano nell'ordine dei Petabytes fino a spingersi ai Zetabytes e i volumi sono in continuo aumento.

SIAMO OLTRE La *Computer Sciences Corporation* nel 2013 ha effettuato delle stime sui volumi dei dati raccolti nel mondo dal 2009 al 2020 e ne emerge che sono stati generati ben 35 Zetabytes da Imprese e Pubbliche Amministrazioni.

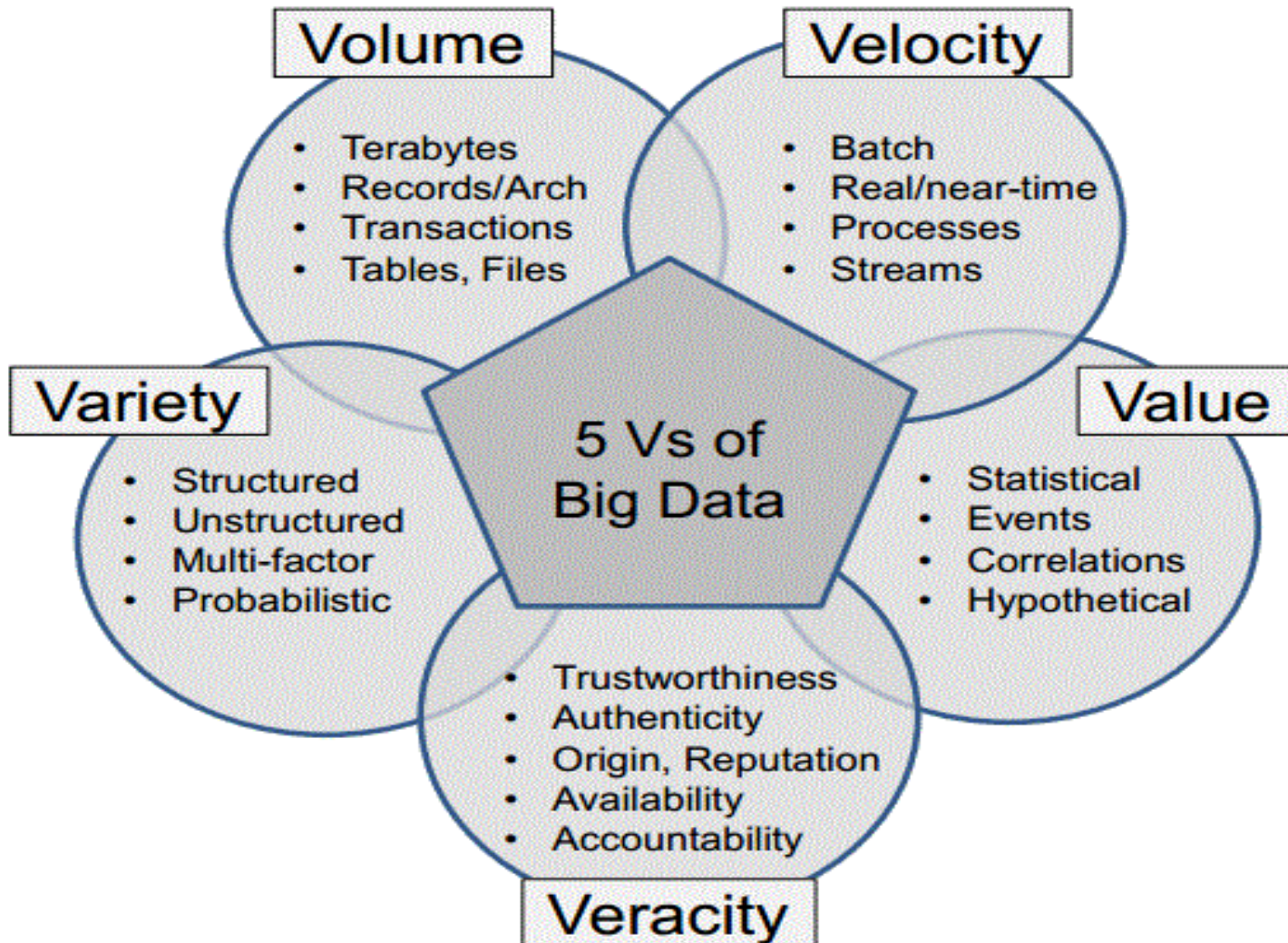


- Totale dati raccolti da imprese e amministrazioni pubbliche
- Dati gestiti dalle imprese
- Dati raccolti dalle imprese

Big Data & Open Data-Stime e dati raccolti dal pubblico e dal privato tra il 2009 e il 2020.

Fonte: <http://www.csc.com/>

Le 5 V dei Big Data



Secondo la TDWI (The Data Warehousing Institute), per riconoscere i Big Data è essenziale definire 3 requisiti base:

- **Volume:** indica la crescita esponenziale di informazione disponibile
- **Velocità:** è riferita alla velocità di raccolta e analisi dei dati
- **Varietà:** rappresenta l'eterogeneità dei dati a disposizione

1. Volume

Kilobyte (KB)	10^3
Megabyte (MB)	10^6
Gigabyte (GB)	10^9
Terabyte (TB)	10^{12}
Petabyte (PB)	10^{15}
Exabyte (EB)	10^{18}
Zetabyte (ZB)	10^{21}
Yottabite (YB)	10^{24}

Considera la dimensione in termini di Bytes dei database utilizzati per archiviare i dati aziendali. Ad oggi, però non è stata definita una soglia che distingua tra ciò che è Big Data e ciò che non lo è.

Dando uno sguardo alla tabella possiamo notare i multipli dei Bytes.

2. Velocity È la velocità con la quali i dati si generano, si raccolgono, si aggiornano e si elaborano. Riferita alla necessità di ridurre i tempi di gestione e analisi, poiché il dato può diventare obsoleto in pochissimo tempo.

3. Variety

Che può essere intesa come molteplicità di fonti o come eterogeneità di formato dei dati. In prima istanza, infatti si possono avere dati generati da diverse fonti interne o esterne. Questi dati poi possono avere diversi formati (*database, testo, video, immagini, audio eccetera*) riconducibili a tre categorie di dati; quelli strutturati, semi-strutturati e non strutturati

dati strutturati : es.

01 Paolo Verde

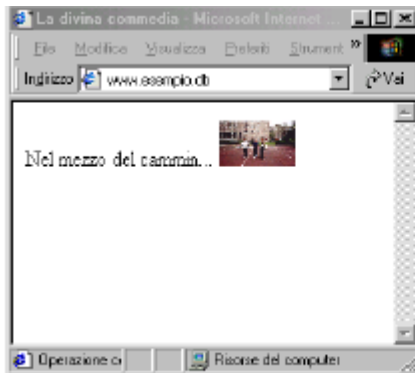
02 Mario Gialli

03 Grazia Bianchi

3. Variety

Che può essere intesa come molteplicità di fonti o come eterogeneità di formato dei dati. In prima istanza, infatti si possono avere dati generati da diverse fonti interne o esterne. Questi dati poi possono avere diversi formati (*database, testo, video, immagini, audio eccetera*) riconducibili a tre categorie di dati; quelli strutturati, semi-strutturati e non strutturati

dati semi strutturati



```
<html>
<title>La divina commedia</title>
Nel mezzo del cammin...
<img/></html>
```

DATI CON STRUTTURA PARZIALE

www.cs.unibo.it/~montesi/

3. Variety

Che può essere intesa come molteplicità di fonti o come eterogeneità di formato dei dati. In prima istanza, infatti si possono avere dati generati da diverse fonti interne o esterne. Questi dati poi possono avere diversi formati (*database, testo, video, immagini, audio eccetera*) riconducibili a tre categorie di dati; quelli strutturati, semi-strutturati e non strutturati
dati non strutturati es:



4. Value

Valore: riguarda le perdite di informazione dovute all'estrazione di conoscenza

5. Veracity

Veridicità: rappresenta l'inaffidabilità dei dati a disposizione

Privacy-by design

- È necessario un cambio nel concetto di proprietà del dato personale che ponga il singolo individuo al centro restituendogli trasparenza e diritti. In una parola democratizzare i Big Data.
- La grande sfida sta nel progettare ecosistemi per i dati personali che diano a tutti la possibilità di gestire le proprie informazioni personali e l'inter-scambio con le entità esterne, persone ed istituzioni, promuovendo trasparenza e trust.