



23 GIUGNO 2016  
14.30 | 16.00

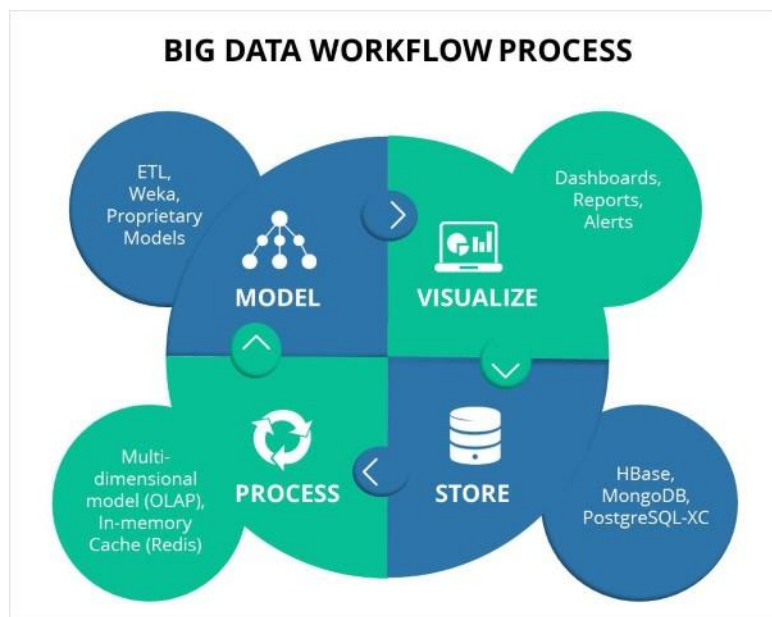
## NUOVE FONTI E DOMANDE

Big Data e archivi dell'Amministrazione finanziaria:  
metodologie innovative di analisi e integrazione,  
potenzialità e limiti



Maria Teresa Monteduro | Dipartimento delle Finanze

# Big data un nuovo percorso metodologico



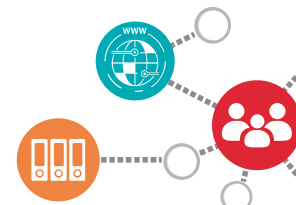
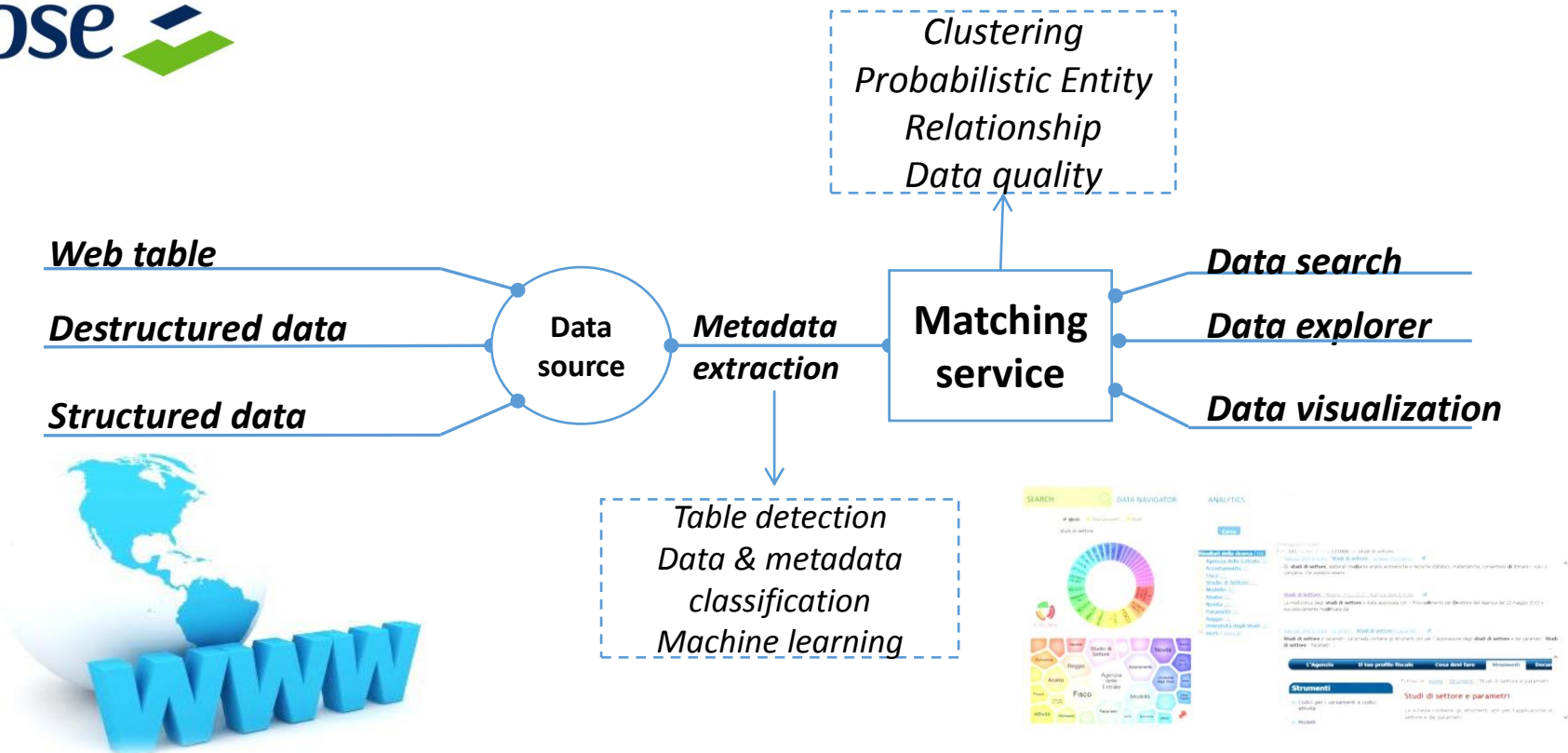
- Velocità di creazione di banche dati correlate e accesso alle informazioni
  - Le nuove tecnologie e metodologie offrono supporti utili alla creazione di nuove banche dati correlate con tempi sempre più brevi e strumenti semplici per l'analisi, come la Visual Analytics
- Raccolta e ricerca di informazioni puntuali
  - Oggi è possibile raccogliere informazioni dall'Amministrazione finanziaria, dalla rete Internet, correlarle, aggregarle e effettuare delle ricerche mirate.
- Nuovi modelli organizzativi di gestione delle informazioni
  - I nuovi modelli organizzativi di informazioni aggregate offrono grandi volumi, velocità, varietà e veridicità, di supporto alle nuove analisi



# Progetto Big Data Analytics

Il progetto, in fase sperimentale, ha l'obiettivo di creare un Big Data Analytics in grado di ricercare e analizzare informazioni di tipo fiscale ed economico sia strutturate, sia destrutturate, come ad esempio documenti economico/fiscali, note metodologiche, norme, fonti informative presenti su Internet in particolare Opendata.

Gli analisti possono effettuare elaborazioni in modalità cluster-metasearch, sulle informazioni, strutturate e destrutturate, raccolte tramite la rete Internet e nel proprio patrimonio informativo; selezionare le informazioni di interesse e utilizzare degli strumenti per effettuare le analisi.

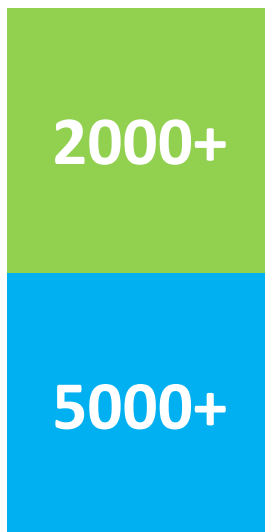


## I numeri ad oggi

Ad oggi sono stati «indicizzati» documenti strutturati e destrutturati sia interni che esterni al proprio patrimonio informativo.

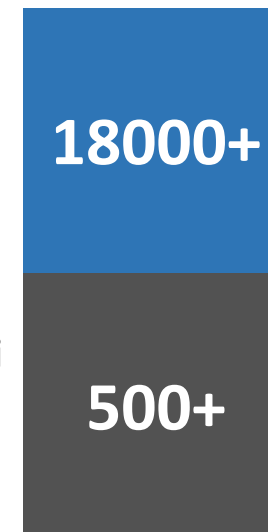
Un processo con le seguenti caratteristiche:

- Aggiornamento costante e quotidiano
- Controlli di qualità
- Catalogazione e clustering



**Oltre 2000 documenti  
provenienti da più di 20  
università**

**Oltre 5000 documenti  
provenienti da oltre 220 fonti  
istituzionali e private (*Banca  
d'Italia, Unioncamere, Istat,  
CNEL, Sole 24 Ore, etc.*)**



**Oltre 18000 archivi  
provenienti da Open Data  
(*ACI, dati.gov.it, EUROSTAT,  
INPS, ISPRA, Comuni, etc.*)**

**Oltre 500 banche  
dati strutturate  
provenienti dalle  
varie Istituzioni**



# Dalla centralità delle informazioni alla centralità delle meta- informazioni

**Pubbliche  
Amministrazioni**



**Altre fonti dati (web e  
non solo)**



*Raccolta dei metadati e/o  
dati strutturati e  
destrutturati*

un processo virtuoso di conoscenza  
dei metadati



*Servizio di interscambio  
metadati e gestione  
richieste informazioni*

**Pubbliche  
Amministrazioni**



*Utilizzo di strumenti di data  
explorer, data visualization,  
data search e analytics  
Richiesta informazioni*



# I limiti dell'utilizzo dei Big Data Analytics

Metodologie, strumenti e ambiti applicativi richiedono determinati e soprattutto «nuovi» requisiti, che si scontrano con dei limiti da fronteggiare.



- **Competenze**
  - Le competenze non sono ancora mature e sono nati recentemente i percorsi formativi, in particolare quelli universitari, per Data Scientist.
- **Sicurezza**
  - Le tecnologie e in particolare le politiche di sicurezza devono garantire l'autorizzazione agli accessi e garanzie rispetto a particolari minacce.
- **Disponibilità**
  - Non sempre le informazioni utili alle analisi sono disponibili per problemi di riservatezza o assenza degli stessi dati.
- **Privacy**
  - E' necessaria una metodologia che dia supporto all'organizzazione delle informazioni tale da garantire la riservatezza delle informazioni dei contribuenti (es. analizzando le informazioni necessarie solo per quel contesto di analisi e creando dei modelli aggregativi).



## Nuove analisi e nuovi modelli di integrazione

**Analisi del rischio dei contribuenti:** l'utilizzo di dati esogeni insieme ai dati dell'Amministrazione finanziaria potrebbe aiutare a definire meglio i profili dei contribuenti, per stabilire i profili di evasione, effettuare analisi di forecasting, per prevedere il gettito fiscale rilevando anomalie o errori, individuare i modelli comportamentali tramite l'uso degli alberi decisionali.

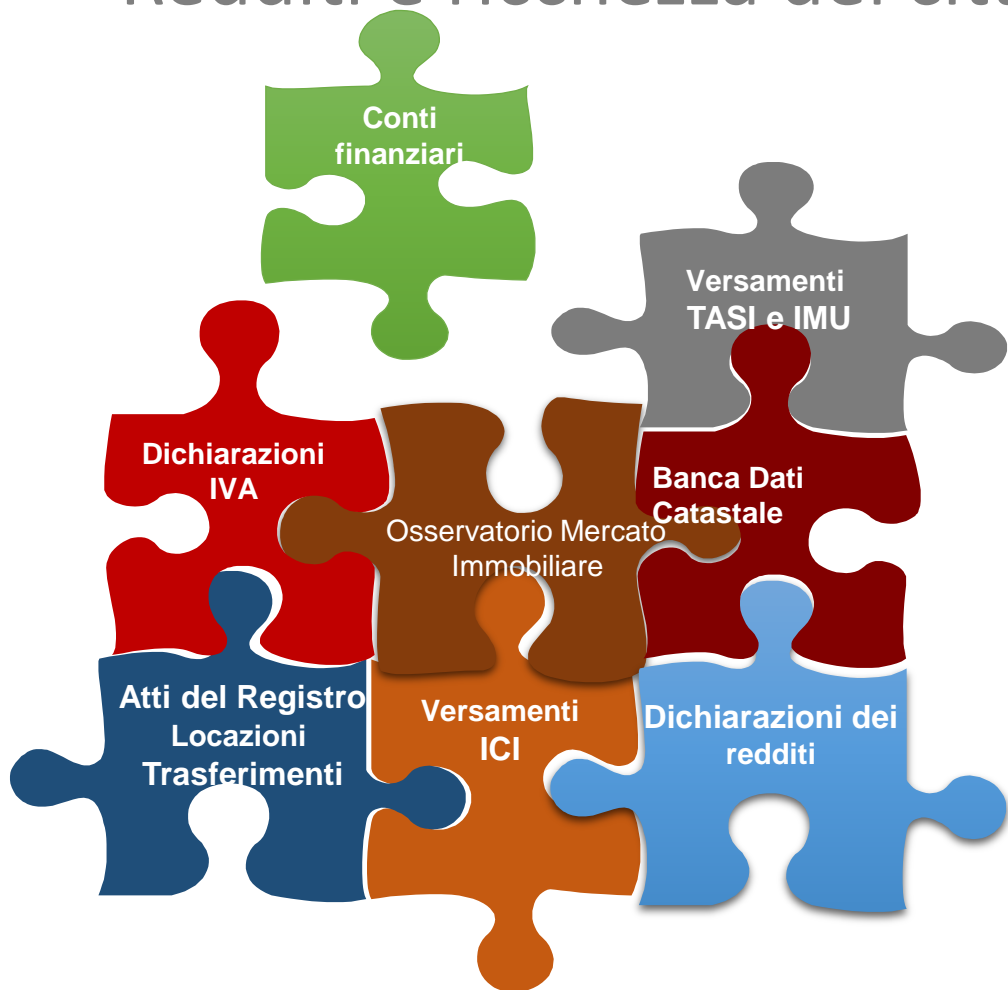
**Valutazione di politiche economiche degli enti locali:** con sistemi di raccolta automatica delle informazioni e di analisi testuali (*text mining, NLP, deep learning*) potrebbero essere analizzate le delibere comunali ad esempio quelle riguardanti l'IMU e la TASI per le seconde case.

**I Fabbisogni standard e analisi predittive:** l'incrocio di informazioni di un determinato periodo e territorio possono essere utili per le analisi predittive.

**Redditi e ricchezza dei cittadini:** incrociare le banche dati utili a svolgere analisi sui redditi e ricchezza dei cittadini.



## Redditi e ricchezza dei cittadini



La banca dati dei redditi e della ricchezza immobiliare potrebbe essere estesa alle variabili “esogene” dell’Amministrazione finanziaria. Un progetto che completerebbe la costruzione della banca dati immobiliare, integrata con le dichiarazioni dei redditi e con i versamenti delle imposte, al fine di effettuare analisi sui redditi e ricchezza dei cittadini, individuando gli eventuali nodi di correlazione con i suoi “metadati” (es. *famiglia, immobili, titoli, saldi C/C*) e migliorando sia la qualità delle statistiche nazionali sui redditi e ricchezza sia le valutazioni sulle *policies* e i loro effetti.

Come superare i limiti:

- Sicurezza: i dati dovrebbero essere integrati in un sistema di security policy.
- Disponibilità: i dati dovrebbero essere resi disponibili tramite adeguamenti normativi.
- Privacy: i dati dovrebbero essere anonimizzati, aggregati e trattati secondo il “*codice di deontologia e di buona condotta per i trattamenti di dati personali a scopi statistici*”.





Grazie

Maria Teresa Monteduro

Dipartimento delle Finanze

