

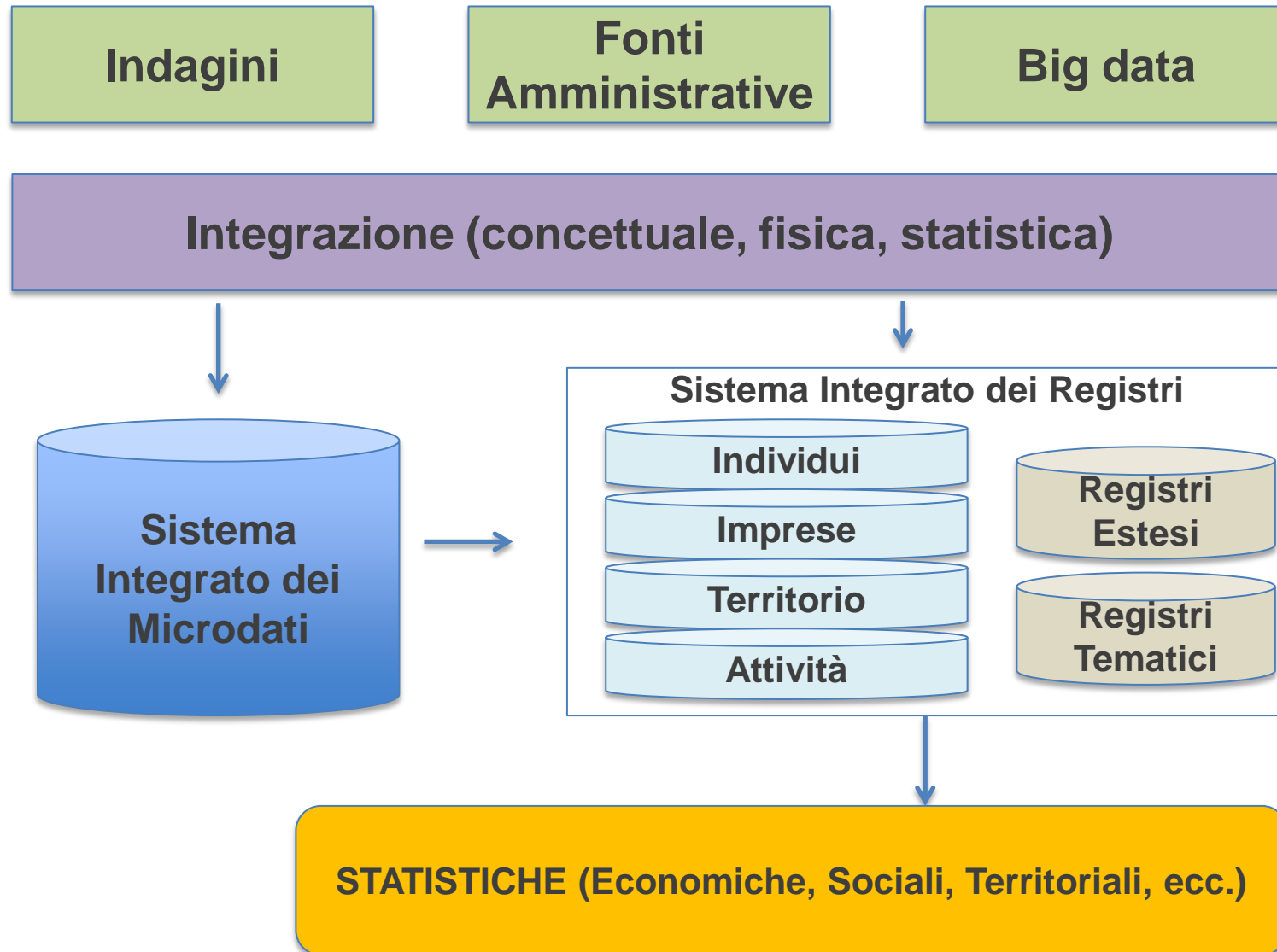
**IRPET – Regione Toscana**

Laboratorio: Nuovi strumenti informativi e Big Data a supporto delle POLITICHE  
del TERRITORIO e della CITTÀ

# **Le iniziative dell'Istat in tema di Big data**

Firenze, 13 giugno 2017

# Verso l'integrazione delle fonti



- 1. Reti sociali**
- 2. Imprese**
- 3. Internet delle cose / sensori**

## Opportunità e benefici:

Incremento di efficienza con una riduzione delle risorse.

Statistiche tempestive e rilevanti.

Miglioramento delle indagini esistenti.

Statistiche rispondenti alle richieste degli utenti.

Rilevazioni impraticabili con le indagini standard.

Minori costi.

## Rischi e costi:

Rappresentatività/Coverage dei Big Data rispetto al target della Statistica Ufficiale.

Qualità del dato non controllata.

Informazione che può essere troppo granulare (e non informativa).

Stabilità e proprietà dei dati.

Investimenti in infrastrutture IT (on-premise Hadoop cluster).

Skill informatici/statistici.

Privacy.

## Le attività in corso

Stato delle attività	Denominazione	Tempistica risultati
Progetti già in produzione	Scanner data prezzi al consumo e <i>web scraping</i> siti per stima alcuni particolari beni	2016
Progetti sperimentali maturi per la produzione	Stima di popolazioni tramite <i>mobile phone</i>	2017
	Frame sugli agriturismi	
	Google trends e stima dell'occupazione	
	<i>Web scraping</i> per l'indagine ICT nelle imprese	
	<i>Web scraping</i> per cattura altre informazioni aziendali	Fine 2017

# Le attività in corso

Stato delle attività	Denominazione	Tempistica risultati
Progetti in sperimentazione	Studio flussi di traffico e incidentalità per mezzo di sensori sul traffico e Google traffic	2017
	Immagini satellitari sul suolo	
	Usò internet e social media per stima fiducia consumatori	
	Domanda turistica (da mobile phone)	Fine 2017
Nuove attività da sperimentare	Frame strutture ricettive (nuove forme)	2018
	Utilizzo di dati satellitari per la descrizione dei viaggi del trasporto marittimo	2018
	Reti, strutture e servizi del trasporto aereo	
	Smart meters per la misurazione della qualità e correzione del registro di popolazione	
	Usò di big data per nuovi leading indicators e per modelli per stime anticipate e nowcasting	
Utilizzo di big data per indicatori di Smart Cities/ BES/SDGs		

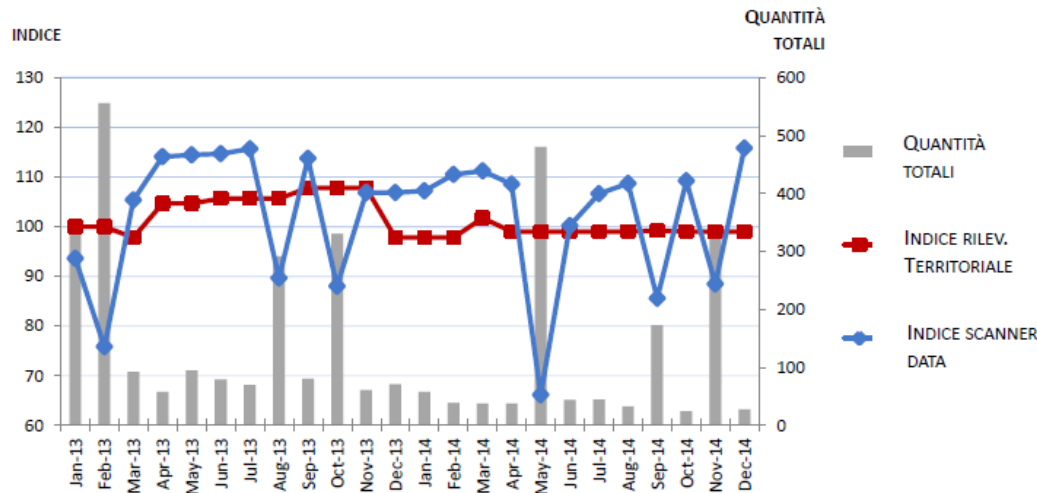
# Prezzi al consumo: Scanner data e web scraping

- **colmare il gap informativo** (promozioni, e-commerce);
- rispondere all' **articolazione della domanda** di informazione statistica sui prezzi al consumo, soprattutto a livello territoriale.

Dalla fine del 2013, tavolo informale con **ADM** e **GDO** per l'acquisizione degli **scanner data**.



Raccolta dati ad alta frequenza attraverso «internet robots»).



Indice dei prezzi al consumo di singola referenza del **caffè tostato** e quantità vendute nel mese.

*Comparazione tra indice scanner data e indice calcolato sulla base dei dati della rilevazione territoriale.*

Gen. 2013 – Dic. 2014

**Uso di dati anonimizzati di telefonia mobile** nel processo di stima di flussi di popolazione intercomunale, utilizzando i cosiddetti **call data record (cdr)** forniti dalle compagnie telefoniche.

Potenzialità:

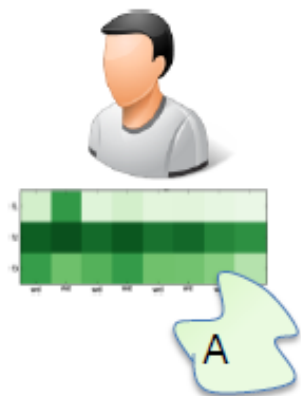
- aumentare **l’efficienza dei sistemi urbani** e promuovere la loro integrazione;
- anticipare la **domanda sociale di infrastrutture e servizi di trasporto**;
- Produrre statistiche sul **turismo**.

Attori coinvolti: Istat, CNR, Università di Pisa.

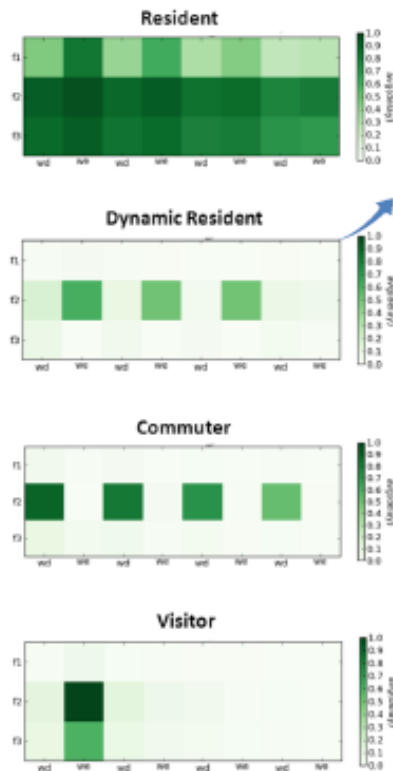


# “Persons and places”

## PROFILO DI CHIAMATA INDIVIDUALE



ALGORITMO DI CLASSIFICAZIONE



RESIDENTI STATICI



RESIDENTI DINAMICI

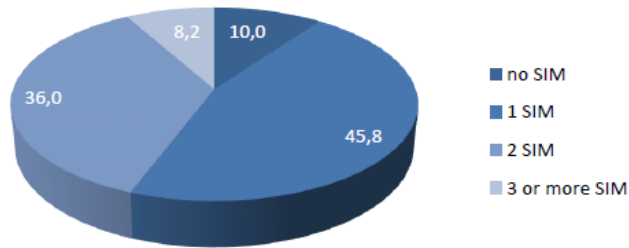


PENDOLARI



VISITATORI

## Calibrazione:



GEOGRAPHICAL AREA OF RESIDENCE	TRAVELLING SIM IN ITALY		DOMESTIC TRIPS		Calibrators to be applied to CdRs/DDRs
	ths	%	ths	%	
North-West	7.138	19,9	11.365	24,1	1,6
North-East	12.757	35,6	15.633	33,2	1,2
Centre	8.755	24,5	10.518	22,3	1,2
South	5.770	16,1	7.587	16,1	1,3
Islands	1.371	3,8	1.989	4,2	1,5
<b>ITALY</b>	<b>35.790</b>	<b>100,0</b>	<b>47.093</b>	<b>100,0</b>	<b>1,3</b>

**Turismo domestico:** basato sull' «*home anchor point*»

**Turismo straniero:** basato sul *roaming*.

- **Informazione più fine e accurata.**
- **Possibilità di descrivere i comportamenti nel tempo e lo spazio.**
- **Costi ridotti.**
- **Possibilità di monitorare zone remote (parchi) e esercizi ricettivi informali.**



## Uso delle immagini del satellite Sentinel-2

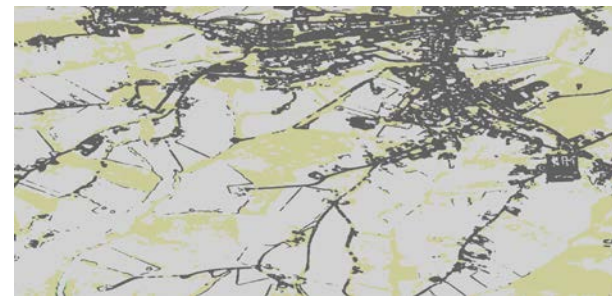
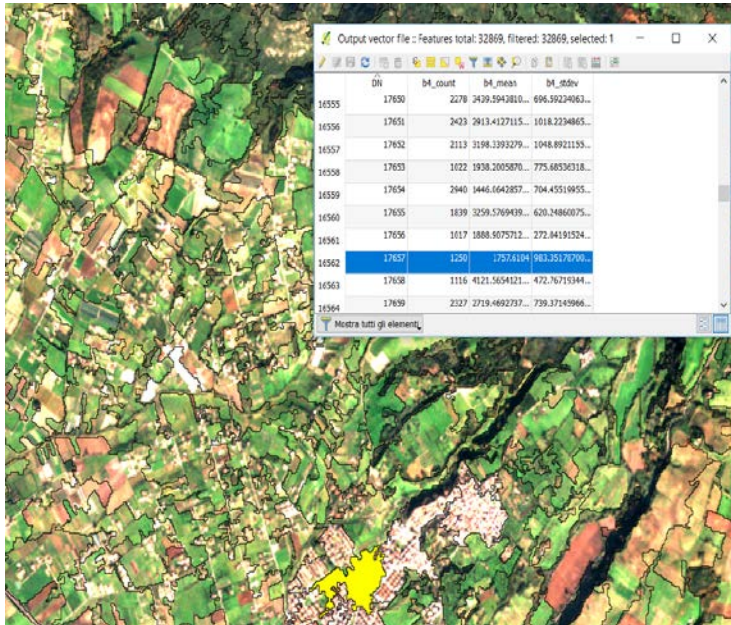
### Opportunità:

- Esiste una domanda di informazione (Urban sprawl, uso del suolo, coltivazioni) soprattutto a livello europeo.
- Dati disponibili gratuitamente.
- Capacità computazionali e software non sono più un ostacolo (CINECA Virtual Data center).
- Solida comunità scientifica attorno al *Copernicus user forum*.
- Informazione statistica da includere nella Space Economy strategy.

# Immagini satellitari

## Metodologia:

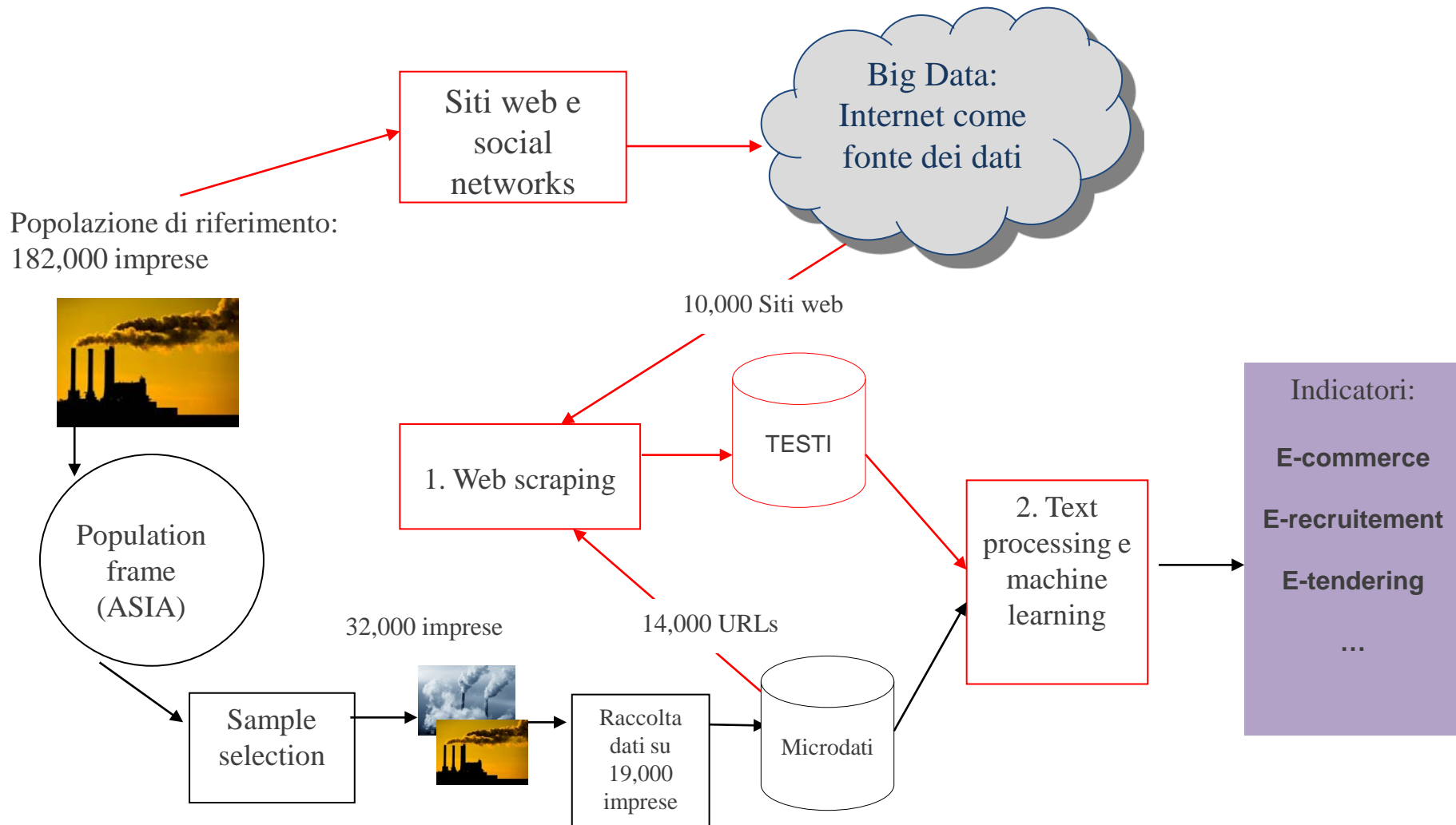
1. **Pre-trattamento dei dati** (Correzione mappe, mappe vettoriali, raster).
2. **Classificazione.**
3. **Valutazione dell'accuratezza.**
4. **Risultati** (Mappe e statistiche su uso del suolo).



## Prospettive informative:

- Indicatori agro-ambientali possono essere prodotti più facilmente e in forma armonizzata.
- Soddisfazione richieste Eurostat su uso del suolo.
- Produzione di indicatori armonizzati per gli SDGs.
- Il Sendai framework richiede statistiche “*near real time*” ad es. i danni all’agricoltura in caso di eventi estremi.

# Web scraping per l'indagine ICT nelle imprese

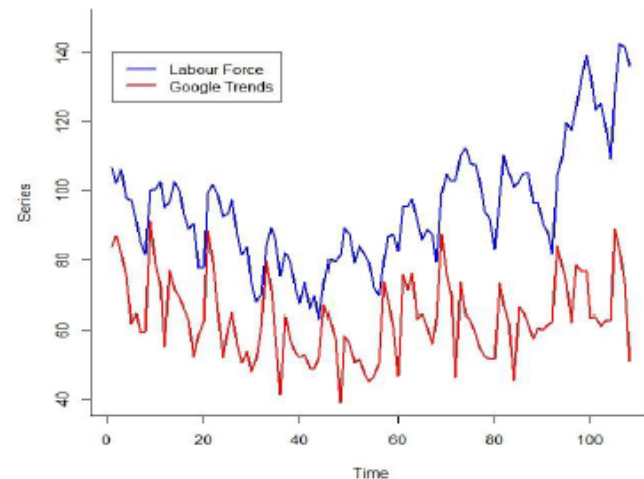


**Google trends** può essere utilizzato per migliorare le stime prodotte dall'Istat sul mercato del lavoro in termini di previsioni e *nowcasting*.

- Si avvicina il ciclo dei dati a quello delle **decisioni**.
- Si amplia la capacità di **dettaglio territoriale** degli indicatori sul lavoro.
- Si attenua il trade-off tra **accuratezza** e **tempestività**.

Tasso di disoccupazione mensile  
(RFL);  
Offerta di lavoro (Google trends)

*indice 2004=100*



# Progetto TREND2 sulla micro e piccola impresa



## Trend2

Osservatorio sulla Micro e Piccola Impresa

### Fonte:

Bilanci delle imprese fino a 20 addetti depositati presso CNA.

### Metodologia:

Stima delle variazioni tendenziali per panel mobili.

### Output:

Stima delle variazioni tendenziali trimestrali di fatturato, investimenti e retribuzioni a 6mesi dal periodo di riferimento;

Dettaglio provinciale per ATECO a 2 digit (*non per tutti i settori, non per tutte le provincie*).



- **Integrazione delle fonti prospettiva a venire per la statistica ufficiale**
- **Big data avranno un ruolo sempre più rilevante**
- **Garantire qualità del dato**
- **Sviluppo Public Private Partnerships**
- **Privacy**
- **Skills**