



SAPIENZA
UNIVERSITÀ DI ROMA

Dai Big Data alle Trusted Smart Statistics. L'Istat e gli Istituti di statistica europei di fronte alla sfida posta dalle nuove fonti di dati.

Facoltà di Scienze Politiche, Sociologia e Comunicazione
Dipartimento CoRis Comunicazione, Ricerca Sociale e Marketing
Dottorato di ricerca XXXV Ciclo

Gerarda Grippo
Matricola 1841566

Tutor
Prof. Alberto Marinelli
Prof. Stefano Epifani

Co-tutor
Dott.ssa Monica Scannapieco

A.A. 2021-2022

Indice

Indice	3
Indice delle tabelle	5
Indice delle figure	5
Indice dei grafici	6
Introduzione	7
Capitolo 1 - I Big Data e il loro impatto nella costruzione di nuova conoscenza	10
1.1 Big data, tecnologie e trasformazione digitale: connessioni e relazioni in un nuovo ambiente.	10
1.2 Da small a big: cosa cambia? Partiamo dalle definizioni	13
1.3 Il valore dei big data	21
1.4 Big data e paradigmi che mutano.	26
1.5 Dai dati alla conoscenza. Nuove domande e nuove competenze.....	30
1.6 I dati da soli sanno parlare? E la correlazione basta?.....	36
Conclusioni.....	43
Capitolo 2 L'evoluzione della statistica	44
Introduzione.....	44
2.1. La statistica ufficiale e gli Istituti di statistica	45
2.2 Come cambia il modo di produrre informazione in Istat	50
2.3. Big Data e statistica ufficiale, un percorso che viene da lontano.	55
2.3.1 Le statistiche sperimentali	66
2.4 Fiducia e statistica ufficiale: una relazione possibile	70
Allegato 1 –Scheveningen Memorandum	75

Allegato 2 – Bucharest Memorandum.....	78
Capitolo 3 - Le conseguenze della trasformazione digitale in Istat: dal programma di modernizzazione al Centro per le <i>Trusted Smart Statistics</i>.	80
3.1 Il processo di digitalizzazione in Istat	80
3.2 <i>Born digital functions</i> in Istat: le <i>Trusted Smart Statistics</i>	83
3.3 Il carattere antropocentrico delle TSS.....	89
3.4 Il sostegno alle nuove iniziative: verso il sistema di produzione delle TSS.	92
3.5 Il Centro per le TSS in Istat	97
3.5.1. La roadmap	100
3.6 Implementazione di un <i>framework</i> per il monitoraggio	108
3.6.1. Area soft/human level	109
3.6.2. Area Metodi e tecniche – <i>Hard skills</i>	111
3.7 Implementazione della matrice	113
3.8 Il percorso fino alle TSS: il processo di modernizzazione in Istat. Cause e impatti. ...	118
3.9 La ricerca in Istat	122
3.10 Le infrastrutture per la ricerca.....	126
Allegato n. 3 Il Modello di Business Architecture dell'Istat	130
Capitolo 4 - Verso le <i>Trusted Smart Statistics</i>: i risultati dell'indagine tra gli Istituti Nazionali di Statistica all'interno del sistema statistico europeo.	131
4.1 La <i>Survey</i>	131
4.1.1 Istituti coinvolti	132
4.2 Lo strumento di rilevazione: struttura	132
4.3 Struttura del Questionario	138
4.3.1 I rispondenti	138
4.3.2 Prima sezione	140

4.3.3 Seconda Sezione – Investimenti realizzati	142
4.3.4 Terza sezione – Investimenti programmati	151
4.3.5 Quarta sezione – Investimenti auspicati	155
4.3.6 Sezione dati e strumenti usati	159
4.3.7 Sezione risultati raggiunti	163
Allegato n. 4 Il questionario.....	170
Conclusioni	180
Bibliografia.....	182
Ringraziamenti	191

Indice delle tabelle

Tabella 1 – Fonti dati Classificazione UNECE	19
Tabella 2 - Statistiche sperimentali Istat*	68
Tabella 3 – Progetti prioritari TSS 2022-2023	103
Tabella 4 – Livello di Maturità organizzativa e relativo punteggio	114
Tabella 5 – Livello di coerenza e relativo punteggio	114
Tabella 6 – <i>Framework</i> di monitoraggio: esempio su azioni trasversali – Soft Skills	115
Tabella 7- <i>Framework</i> di monitoraggio: esempio su azioni trasversali Hard Skills	117

Indice delle figure

Figura 1 - Piramide della conoscenza. Dal Dato alla conoscenza applicata.....	31
Figura 2 - Representations of T vs. Pi-shaped education.	34
Figura 3 - Le tappe del percorso verso le TSS	97
Figura 4 - Dimensioni organizzative che concorrono alla realizzazione del sistema TSS...	104

Indice dei grafici

Grafico 1 - funzione svolta dai rispondenti.....	140
Grafico 2 - Investimenti nel settore comunicazione e diffusione	146
Grafico 3 - Fonti usate per la produzione statistica	160
Grafico 4 - Tecniche di accesso ai dati.....	161
Grafico 5 - Tipologia di soggetti con cui gli stipulano accordi di collaborazione	161
Grafico 6 - Ambiti tematici delle statistiche sperimentali pubblicate (settembre 2020)	164
Grafico 7- Piattaforme più funzionali alla diffusione di TSS.....	165
Grafico 8 - Competenze più utili alla produzione di TSS	167
Grafico 9 - Competenze su cui gli Istituti stanno investendo.....	168
Grafico 10 - Vantaggi ottenuti dall'integrazione dei big data nei processi di produzione ..	169

Introduzione

Il presente lavoro traccia il percorso evolutivo che gli Istituti di statistica, e l'Istat in particolare, hanno fatto e stanno facendo per adeguarsi, accompagnare e raccontare i significativi mutamenti sociali che hanno caratterizzato gli ultimi anni e che sono inescindibilmente connessi con la trasformazione digitale.

La prima parte della tesi si focalizza sulla definizione di un *framework* teorico dedicato alle nuove fonti di dati e sul loro impatto su specifiche dimensioni della vita sociale, in particolare sulla loro capacità di contribuire alla costruzione di nuova conoscenza e di rispondere a bisogni informativi emergenti. I big data rappresentano una componente del più ampio processo di digitalizzazione che ha trasformato in dati quasi ogni aspetto dell'interazione sociale, economica e fisica tra individui, organizzazioni, oggetti e sistemi. I big data sono uno dei prodotti della nuova infrastruttura sociale rappresentata dal digitale, e saranno analizzati in un'ottica relazionale, ecologica e a *network* che terrà conto, cioè, delle interazioni, dei processi e degli impatti che le nuove fonti di dati determinano nei vari ambiti delle nostre vite e in particolare nelle pratiche di produzione e consumo delle statistiche ufficiali. Il lavoro contiene infatti un approfondimento sui processi evolutivi che i big data hanno innescato all'interno degli Istituti di statistica europei e in particolare all'interno dell'Istat, e sulle risposte che gli Istituti hanno dato a queste trasformazioni in termini di innovazione di prodotto e di processo.

Per quanto riguarda l'Istituto di Statistica Italiano presso il quale lavoro da oltre dieci anni, nelle pagine seguenti sarà riportata l'analisi della traiettoria evolutiva dei diversi paradigmi di ricerca di volta in volta adottati fino ad arrivare a descrivere, praticamente in *tempo reale*, i recentissimi investimenti in cui l'Istat è attualmente coinvolto e che sono finalizzati alla costruzione di un nuovo sistema di produzione, quello delle *Trusted Smart Statistics*. Le *Trusted Smart Statistics* rappresentano la nuova frontiera della statistica europea, sono prodotti statistici innovativi che inglobano le nuove fonti di dati e che sono effettivamente in

grado di intercettare e tradurre in informazione statistica i cambiamenti determinati dai processi di digitalizzazione. Quello riportato nei capitoli centrali è quindi un racconto delle diverse sfide che l'Istat ha affrontato e sta affrontando per mantenere intatta sia la sua rilevanza istituzionale in un ecosistema informativo che via via diventa sempre più affollato sia la *trust*, la fiducia che da sempre i cittadini nutrono nei confronti della statistica ufficiale.

Viene dato particolare rilievo ad alcune delle innovazioni introdotte in Istat per accogliere i cambiamenti richiesti dall'integrazione delle nuove sorgenti di dati all'interno dei processi di produzione statistica. Partendo dall'esperienza internazionale, l'Istat ha affrontato un profondo processo di modernizzazione che ha portato l'Istituto a riorganizzarsi per rispondere meglio alle nuove esigenze informative e conoscitive mantenendo integro il ruolo di eccellenza dell'Istituto in un contesto in cui sono molti gli attori anche non istituzionali che producono e diffondono informazione statistica. Si avvia un processo innovativo di infrastrutturazione della ricerca, la ricerca viene riconosciuta anche nello statuto, come strumento per contribuire al *“miglioramento della qualità delle informazioni statistiche e dei processi adottati per la produzione, sviluppo e diffusione della statistica ufficiale e all'introduzione nei processi suddetti dei risultati della ricerca metodologica e tematica”*¹. Negli ultimi anni, per capitalizzare e i risultati raggiunti in ambito europeo in tema di produzione statistica con i big data e andare verso la produzione di *Trusted Smart Statistics* l'Istat si dota di una *governance* e di una *Roadmap* per delineare la traiettoria evolutiva verso le nuove frontiere della statistica ufficiale. Uno dei prodotti di questo lavoro è l'implementazione di un *framework* di monitoraggio per seguire l'andamento dei servizi a supporto di questi nuovi prodotti statistici. Se il terzo capitolo rappresenta quindi un focus specifico sull'Istat e sulle innovazioni introdotte e qui solo citate, il quarto capitolo offre una panoramica più ampia sul Sistema Statistico Europeo. L'ultima parte della tesi, infatti, riporta i risultati di un'indagine effettuata presso gli istituti di statistica europei che sono già ampiamente stati coinvolti in progetti di sperimentazione statistica con l'uso dei big data e che, come l'Istat sono impegnati in un riadattamento del loro *business model* al fine di continuare ad integrare

¹ Art. 2 Statuto dell'Istituto Nazionale di Statistica.

ricerca e produzione e ad innovare per arricchire l'informazione statistica in termini di qualità, tempestività, granularità territoriale e continuare a soddisfare varie tipologie di *stakeholders* a cui la statistica ufficiale si rivolge.

La *survey* nasce con l'obiettivo di indagare, per ciascun Istituto Statistico del Sistema Statistico Europeo, *l'as-is* cioè lo stato dell'arte rispetto al riassetto dei propri sistemi di produzione resi necessari dall'integrazione delle nuovi fonti di dati nei processi produttivi e il *to-be* vale a dire il modello verso cui ciascun paese tende. Il confronto tra *l'as-is* e il *to be* mette in evidenza i gap da colmare nelle varie dimensioni organizzative analizzate. Come vedremo, lavoro da fare è ancora molto, le *Trusted Smart Statistics* rappresentano un grosso cambiamento culturale, per questo richiedono tempi più lunghi per essere totalmente assorbiti e intrecciati all'interno della cultura organizzativa degli Istituti, ma è uno sforzo a cui la comunità statistica non si può sottrarre e che anzi deve cogliere per rafforzare il legame di fiducia con la collettività e comunicare con maggiore forza il valore dei dati come bene pubblico e patrimonio comune.

Vivere dall'interno queste evoluzioni ed essere stata talvolta fortemente coinvolta nei processi di innovazione che l'Istat ha fatto per cogliere i cambiamenti dettati dalla trasformazione digitale, mi ha consentito di lavorare alla stesura della tesi con un livello di coinvolgimento profondo. L'aver potuto quasi verificare "sul campo" gli impatti di alcuni fenomeni scaturiti dalla trasformazione digitale e riportarli all'interno di questo lavoro è stata una grossa opportunità. Emerge con forza che quello degli Istituti di statistica non è un mondo inerte chiuso all'interno di rigidi criteri metodologici, è un mondo in evoluzione perenne perché in forte interconnessione con i sistemi sociali in cui è immerso, anche il presente lavoro è per la sua stessa natura *aperto*, c'è ancora spazio per raccontare come andrà a finire: la statistica ufficiale non si può fermare perché come ha affermato Mariana Kotzeva, Direttrice di Eurostat, all'apertura della XIII Conferenza di statistica nel luglio 2018 "*statistics follows the life*".

Capitolo 1 - I Big Data e il loro impatto nella costruzione di nuova conoscenza.

1.1 Big data, tecnologie e trasformazione digitale: connessioni e relazioni in un nuovo ambiente.

Il digitale ha trasformato le nostre vite a livello sociale, tecnologico ed umano e ci ha spinto a ripensare le strutture dei nostri sistemi economici e sociali, le nostre modalità di relazioni, le nostre abitudini, i nostri comportamenti. Il digitale, l'invenzione e lo sviluppo delle tecnologie dell'informazione e della comunicazione (d'ora in poi ICT) hanno fatto la differenza tra chi siamo e chi eravamo². Il periodo storico che stiamo vivendo, per le profonde trasformazioni in atto, è stato definito *quarta rivoluzione industriale*. Stiamo sicuramente vivendo l'inizio di un nuovo capitolo nello sviluppo umano favorito dalla crescente disponibilità e interazione tra innovazioni tecnologiche straordinarie ma, parlare di rivoluzione industriale è riduttivo perché rimanda ad una dimensione puramente tecnologica. La rivoluzione digitale, rispetto a quella industriale, ha fatto di più, ha creato un nuovo ambiente, nuove relazioni e nuove modalità di interazione, ha cambiato la nostra cultura in modo profondo e significativo, è una rivoluzione *ontologica*, ambientale, ecologica, in cui tutti noi abbiamo sviluppato atteggiamenti digitali e nuovi comportamenti grazie all'uso di tecnologie ormai pervasive. La rivoluzione digitale non è banalmente solo una rivoluzione tecnologica, non è legata all'avvento di una singola tecnologia (come il motore a scoppio, l'elettrificazione o l'energia nucleare per le rivoluzioni industriali). E non è stata neanche provocata solo dalle tecnologie digitali, ma dal modo in cui, attraverso queste ultime, è stata ripensata la struttura dei nostri sistemi economici e sociali. Per capirlo basti pensare al significato letterale della parola *digitalizzazione*. Il vocabolario Treccani definisce

² Floridi L., *La quarta rivoluzione: come l'infosfera sta cambiando il mondo*. Cortina Editore 2017

la digitalizzazione come una *conversione di grandezze analogiche in informazioni digitali, effettuata mediante un dispositivo, detto digitalizzatore o convertitore analogico-digitale*. La digitalizzazione è di fatto un'attività di traduzione in un linguaggio digitale dei processi in modo da poterli gestire in maniera automatica. Ma se l'attività è (apparentemente) banale, non si può dire lo stesso per i risultati che porta: agendo sull'automazione conduce ad una reingegnerizzazione dei processi, non a livello individuale o della singola organizzazione, ma a livello sociale. La rivoluzione digitale ha provocato un cambiamento dell'approccio, dei modi di lettura, di rappresentazione e di conoscenza del reale: le ICT stanno mutando la natura vera e propria e in tal senso il significato stesso della realtà. Le nuove tecnologie non sono da considerare una forza implacabile o un mero strumento che l'uomo può scegliere di usare ogni qualvolta lo desidera. Per capirne la portata bisogna comprendere come queste si interconnettano e influenzino le nostre esistenze³. Le tecnologie rappresentano di fatto nuovi sistemi di relazione con il mondo, sono diventate "forze ambientali", che creano e trasformano le nostre realtà e le nostre relazioni e il nostro stare al mondo. La capillarità e la pervasività dell'informatica, attraverso la diffusione dei computer, degli *smart devices*, lo sviluppo della rete, hanno conseguenze sui comportamenti delle persone, sul modo in cui esse comunicano tra di loro, si informano, costruiscono le loro convinzioni, ridefiniscono il loro orizzonte degli eventi, mettono cioè in atto una trasformazione che va ben oltre il meccanico processo di traduzione in forma digitale, la trasformazione digitale provoca una trasformazione di senso⁴ che sviluppa un inesorabile processo di rimediazione dei comportamenti individuali e collettivi. Si tratta di una trasformazione talmente profonda da richiedere un ripensamento radicale dei nostri concetti. La società, la politica, l'ambiente e, nel contesto di questo lavoro, aggiungiamo anche la statistica, sono ambiti che devono essere ridefiniti in seguito all'avvento del digitale. Per dirla con Floridi possiamo dire che la rivoluzione digitale *ri-ontologizza* la modernità (ne modifica cioè la natura) e la "*ri-epistemologizza*" (ne modifica la comprensione). Le tecnologie dell'informazione ridanno

³ Schwab K., *Governare la quarta rivoluzione industriale*, Franco Angeli, 2016.

⁴ Epifani S., *Sostenibilità digitale*, Digital Institute, 2020

un'ontologia al mondo, lo trasformano nella sua natura intrinseca in cui le barriere del mondo reale offline e quello online si stanno corrodendo e l'aspetto fisico e quello digitale della realtà sono ormai irreversibilmente integrati (Floridi, 2017)⁵.

La transizione dall'analogico al digitale e la crescita esponenziale di spazi informativi, in cui trascorriamo sempre più tempo illustrano con la massima evidenza il modo in cui le ICT stanno trasformando il mondo in un'*infosfera* vale a dire in uno spazio informativo, una sfera informativa che ci circonda costantemente, un ambiente - tipico dell'epoca digitale - costruito dalle informazioni in cui viviamo la nostra esperienza quotidiana. La costruzione, o meglio il popolamento di questo nuovo ambiente, è in parte garantito dalle nuove fonti di dati, da questo diluvio di dati generato, in maniera molto spesso inconsapevole, dai dispositivi che indossiamo e che ci circondano. Nell'*infosfera* ci muoviamo rimanendo sempre connessi: è un'esperienza nuova, che intreccia *offline* e *online*, e che possiamo definire *onlife*, viviamo costantemente immersi in un mare di dati: dobbiamo imparare a capire come leggerli e selezionarli per costruire la conoscenza. Quando i confini tra la vita *online* e quella *offline* vengono meno, e siamo continuamente connessi gli uni agli altri, circondati da oggetti intelligenti in grado di interagire con noi, allora possiamo dire di essere integrati nel mondo *onlife*. La modalità *onlife* è sempre più alla base delle nostre attività quotidiane e definisce il nostro modo di comprare, lavorare, imparare, curare la nostra salute, divertirci, coltivare le nostre relazioni; il nostro modo di interagire con il mondo informazionale, giuridico, finanziario, politico; e in maniera tragicamente attuale, persino il nostro modo di fare la guerra. In questo nuovo ambiente, siamo agenti interconnessi generatori di informazione, ciascuno di noi non è un'entità isolata ma un organismo informazionale interconnesso con altri agenti biologici e artificiali che si muovono in un mondo dominato dall'informazione. Il nostro modo di abitare il mondo e di conoscerlo si struttura attorno ad un universo connesso e quindi sempre più dipendente da questo stato di permanente interconnessione – reale o potenziale – che influisce e si riverbera su ogni aspetto della nostra vita, dal modo di

⁵ Floridi L., La quarta rivoluzione. Come l'infosfera sta trasformando il mondo. Cortina Editore, Collana Scienze e Idee, 2017

relazionarci gli uni agli altri, al pensare e realizzare nei contesti più disparati organizzativi, educativi, e anche informativi e dell'intrattenimento. Emerge, dunque, un terreno culturale in cui sono intrecciate in modo nuovo tecnologie, forme relazionali e apparato bio – cognitivo. I soggetti “connessi” non solo consumano informazione ed intrattenimento, ma partecipano, producono, distribuiscono, condividono nuove forme culturali. È quella che Castells chiama *self mass communication* per designare questa fase della società dell'informazione, ancora in atto, in cui, gli utenti o *users* della rete non sono più limitati partecipando solo come pubblico di un'opinione pubblica, ma sono essi stessi a produrre società, svincolandosi dalle istituzioni tradizionali. Il nuovo individuo è consapevole di essere “pubblico in pubblico”. Non siamo, dunque, più immersi in internet, ma una sua parte, siamo pubblici connessi, siamo attivatori e, grazie al ruolo delle tecnologie, anche diffusori della narrazione⁶.

La rete diventa un *dispositivo culturale*, “un luogo della società e come tale va pensata, più che come considerarla come spazio esterna ad essa”⁷. Diventa progressivamente difficile scindere i momenti relazionali svolti attraverso modalità tradizionali da quelli attuati attraverso processi tecnologicamente determinati. Ed è in questo nuovo ambiente che vanno contestualizzate le nuove fonti di dati e il loro ruolo all'interno dell'ecosistema informativo.

1.2 Da small a big: cosa cambia? Partiamo dalle definizioni

Etimologicamente la parola dati deriva dal latino *dare*, i dati sono quindi elementi grezzi che possono essere estratti da fenomeni misurati e registrati in vari modi. I dati si riferiscono a quegli elementi che vengono presi; estratti attraverso osservazioni, calcoli, esperimenti e registrazioni⁸. Tecnicamente, quindi, ciò che intendiamo come dati sono in realtà *capta* (dal

⁶ Boccia Artieri G., 2012, *La Rete: una realtà connessa e una prospettiva di mutazione per il potere simbolico della società*, introduzione in Avveduto S., (a cura di), *Scienza connessa: Rete Media e Social Network*, Gangemi editore spa, Roma.

⁷ Boccia Artieri G., *Introduzione al libro Scienza connessa Rete Media e Social* a cura di Sveva Avveduto, Gangemi Editori 2012.

⁸ Borgman, C., *Big Data, Little Data, No Data: Scholarship in the Networked World*, The MIT Press, Cambridge Massachusettes, London England 2017

latino *capere*, che significa “prendere”) cioè quelle unità di dati che sono state selezionate e raccolte dalla somma di tutti i dati potenziali⁹. Come afferma Jensen “*è uno sfortunato incidente della storia che il termine datum ... piuttosto che il termine captum ... sia arrivato a identificare il fenomeno-unità nella scienza*. La scienza, infatti, non si occupa di "ciò che è stato dato" dalla natura allo scienziato, ma di "ciò che è stato preso" o scelto nella natura dallo scienziato a seconda dell'obiettivo scientifico prefisso.

Queste considerazioni evidenziano che i dati raccolti attraverso la misurazione sono sempre una selezione dalla somma totale di tutti i possibili dati disponibili: ciò che si è scelto di prendere da tutto ciò che potrebbe essere potenzialmente preso. In quanto tali, i dati sono intrinsecamente parziali, selettivi e rappresentativi e i criteri distintivi utilizzati nella loro acquisizione hanno delle conseguenze. I dati non sono mai semplicemente solo dati; il modo in cui i dati sono concepiti e usati varia nelle varie fasi di acquisizione, analisi e interpretazione¹⁰. La parola dati designa qualcosa che si può registrare, analizzare e riorganizzare. La capacità di registrare le informazioni è una delle linee di demarcazione tra le società primitive e società avanzate: infatti quando sono diventati disponibili sistemi per registrare eventi e, in tal modo, trasmettere ed accumulare informazioni per un futuro consumo, le lezioni apprese dalle generazioni passate hanno iniziato a evolvere esponenzialmente e la società ha fatto ingresso nella storia¹¹. Col tempo si sono diffuse e sviluppate diverse modalità per *datizzare* un fenomeno cioè convertirlo in una forma quantitativa in modo da poterlo tabulare e analizzare. L'avvento dei computer e la messa a disposizione di strumenti di misurazione e archiviazione digitale hanno reso la datizzazione molto più efficiente e hanno facilitato enormemente l'analisi matematica dei dati che consente di farne emergere il valore occulto.

⁹ Kitchin, R.; Dodge M. *Code/Space: Software and Everyday Life*, MIT Press 2011

¹⁰ Kitchin R, McArdle G. *What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets*. *Big Data & Society*. June 2016. doi:[10.1177/2053951716631130](https://doi.org/10.1177/2053951716631130)

¹¹ Floridi L., *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*. Cortina Editore, Collana Scienze e Idee, 2017

Oggi tutte le azioni che compiano lasciano traccia e danno vita ad un enorme mole di dati: i cosiddetti big data, espressione sicuramente abusata di cui manca ancora una definizione rigorosa.

In assenza di definizioni vincolanti, quando si parla di big data si fa riferimento alla raccolta, all'analisi e all'accumulo di ingenti quantità di dati provenienti da fonti diverse che dopo un processo di raccolta, trattamento e analisi possono proficuamente contribuire a darci informazioni su diversi ambiti tematici. Nella metà degli anni '90, l'espressione big data fu usata per la prima volta da John Mashey, scienziato presso la Silicon Graphics, per riferirsi alla gestione e all'analisi di enormi set di dati. I primi tentativi infatti di definire i perimetri di questo *fenomeno* si sono focalizzati su alcune caratteristiche di questi dati, la prima è sicuramente rappresentata dal volume. L'idea era che il volume delle informazioni fosse talmente cresciuto da non risultare più compatibile con la memoria usata dai computer per la processazione, secondo questa visione i big data sono un immenso insieme di dati che non può essere acquisito, gestito ed elaborato da strumenti informatici *software hardware* "tradizionali" in un tempo tollerabile.¹²

Il Volume rappresenta una delle ormai famose 3V con cui nel 2001 Doug Laney tentò di dare una definizione di big data¹³ attribuendo loro tre caratteristiche principali:

- ✓ volume si riferisce alla dimensione dei dati;
- ✓ velocità si riferisce alla velocità di fornitura dei dati e al tempo entro il quale è necessario agire su di essi.
- ✓ varietà si riferisce all'eterogeneità dell'acquisizione dei dati, rappresentazione dei dati e interpretazione semantica.

Volume, Varietà e Velocità sottolineano l'abbondanza, la diversificazione e l'immediatezza dei big data, ma ne definiscono solo le principali dimensioni puramente quantitative. Ai fini del presente lavoro, le 3V sono veramente poco funzionali ad indagare le dimensioni

¹² Diebold, F. X., On the Origin(s) and Development of the Term 'Big Data' (September 21, 2012). PIER Working Paper No. 12-037, Available SSRN: <https://ssrn.com/abstract=2152421> or <http://dx.doi.org/10.2139/ssrn.2152421>

¹³ Laney D., *Application Delivery Strategies*, Meta Group, 2001

che qui ci interessano. Abbiamo infatti più volte definito i big data come un *fenomeno*, e continueremo a farlo nel prosieguo del lavoro, perché terremo conto dei loro impatti e della loro capacità di incidere su alcuni aspetti della società (e nel nostro caso anche degli Istituti Nazionali di statistica) e di modificarne alcune dimensioni. I big data non sono semplicemente dati che i nostri computer non sono in grado di gestire da un punto di vista computazionale: l'espressione big data designa delle cose che si possono fare solo su larga scala per estrapolare nuove indicazioni o creare nuove forme di valore, con modalità che possono modificare mercati, organizzazioni, relazioni tra cittadini e istituzione e altro ancora e mettono in discussione il nostro modo di vivere e interagire con il mondo.¹⁴

Negli anni il fenomeno è diventato più complesso ed accanto a dimensioni meramente quantitative, gli studiosi si sono sbizzarriti ed hanno integrato le 3V con numerose altre dimensioni che tenessero conto anche di aspetti qualitativi tra cui:

- ✓ esaustività (viene catturato un intero sistema, n = tutto, anziché essere campionato)¹⁵;
- ✓ granularità (Dodge e Kitchin, 2005);
- ✓ relazionalità (contenente campi comuni che consentono la correlazione tra diversi dataset)¹⁶;
- ✓ estensione (può aggiungere/modificare facilmente nuovi campi) e scalabilità (può espandersi rapidamente di dimensioni)¹⁷;
- ✓ veridicità (i dati possono essere disordinati, rumorosi e contenere incertezza ed errore)¹⁸;
- ✓ valore (è possibile estrarre molte informazioni e riutilizzare i dati) (Marr, 2014);

¹⁴ Mayer-Schönberger V., Cukier K., *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, Garzanti, 2013

¹⁵ Mayer-Schönberger V., Cukier K., *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, Garzanti, 2013

¹⁶ Boyd D., Crawford K. (2012) *Critical Questions for Big Data*, *Information, Communication & Society*, 15:5, 662-679, DOI: [10.1080/1369118X.2012.678878](https://doi.org/10.1080/1369118X.2012.678878) <https://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878>

¹⁷ Marz, N., Warren J., *Big Data: Principles and best practices of scalable realtime data systems*. Manning 2015

¹⁸ Marr, B., *Big data: The 5 vs everyone must know*. *LinkedIn Pulse* 6 2014

- ✓ variabilità (dati il cui significato può cambiare continuamente in relazione al contesto in cui vengono generati)¹⁹.

Emma Uprichard in un articolo del 2013²⁰ ricorda che molte altre v che sono state usate anche per descrivere i big data, tra cui: 'versatilità, volatilità, virtuosismo, vitalità, vigore, vitalità, vivacità... virilità... senza valore, simile a un vampiro, velenoso, volgare, violento e molto violento. La sociologa Deborah Lupton nel 2015 ha suggerito di eliminare le v-words per adottare le p-words per descrivere i Big Data: portentoso, perverso, personale, produttivo, parziale, predittivo, politico, provocatorio, *privacy*, polivalente, polimorfa e giocosa.

Questo elenco di parole, a nostro avviso, seppure contribuisca in qualche modo al dibattito critico sui big data, contiene parole che sembrano essersi aggiunte ogni qualvolta si ponesse un problema associato a questi dati o ne emergesse una nuova caratteristica, di fatto nonostante sia un elenco corposo non ci aiuta a definire i tratti ontologici dei dati stessi. In questo contesto tutte questo susseguirsi di aggettivi utili a descrivere la complessità di cui i big data sono portatori, sono stati citati perché ci aiutano a superare la visione meramente tecnologica che le 3 V attribuivano al fenomeno, ma di fatto non aiutano a far emergere la loro capacità di incidere sulle pratiche di produzione e di consumo di informazione.

Di sicuro la generazione di questi dati si realizza continuamente sia in ambienti *online* sia in ambienti *offline* (nell'infosfera cui abbiamo fatto riferimento prima) nell'ambito di attività *online* e *offline* che svolgiamo quotidianamente anche in assenza di interazione diretta con un dispositivo elettronico. Queste attività generano dati che possono fornire informazioni rilevanti sui comportamenti e sulle preferenze degli individui: quando accediamo a Internet prendiamo informazioni dalla rete, ma cediamo alla rete pezzi di noi.

Tutte le nostre attività *online* – a prescindere dal dispositivo utilizzato – generano una enorme quantità di dati, tipicamente personali, che alimentano una copiosa attività di acquisizione.

¹⁹ McNulty, E. *Understanding big data: the seven V's.* *Dataconomy*. URL: <http://dataconomy.com/2014/05/seven-vs-big-data> (2014).

²⁰ Uprichard E. *Focus: Big data, little questions?* Articolo apparso su Discover society, ottobre 2013

Basti pensare alla posta elettronica, ai social networks su cui si caricano foto, video testi, alla navigazione satellitare, alle piattaforme digitali, alle app, alle videocamere di sorveglianza, agli strumenti di pagamenti elettronici, ai dati di geolocalizzazione degli individui, insomma a tutti i dati non strutturati generati dagli utenti. Nella generazione di questi dati gli *smartphone* rivestono un ruolo centrale in quanto dispongono di numerosi dispositivi di *input* integrati in un unico strumento connesso ad internet e che accompagna l'utente in tutte le sue attività quotidiane e lascia tracce ovunque, non a caso questi dispositivi sono stati definiti da Pedreschi *protesi digitali*. Per comprendere quanto siamo ormai questi *devices* siano diventati un nostro prolungamento, riporto una citazione fatta nel 1991 da Mark Weiser che nel descrivere il concetto di *ubiquitous computing* scrisse "sono le tecnologie più importanti che tendono a sparire; diventano talmente parte integrante della vita quotidiana che a un certo punto non sono più distinguibili da questa". E sicuramente questa osservazione rientra tra le profezie che si sono autoavverate. I dispositivi digitali sono ormai componenti decisamente fondamentali nelle nostre auto, nei dispositivi elettronici di consumo e nella maggior parte degli elettrodomestici, sono integrati nell'abbigliamento e nei tessuti nonché nelle infrastrutture quali strade, semafori, ponti e palazzi. Le protesi digitali a nostra disposizione hanno cambiato la nostra connettività.

Adam Greenfield ha definito lo *smartphone* l'artefatto simbolo della nostra epoca. Quest'oggetto *proteiforme*, dice Greenfield, è diventato il mediatore universale e pressoché indispensabile della vita quotidiana. Pochissimi manufatti sono stati così onnipresenti quanto queste lastre luminose di policarbonato. Per molti di noi sono l'ultima cosa che guardiamo prima di addormentarci e la prima che prendiamo in mano al risveglio, Lo usiamo per incontrare persone, comunicare, divertirci, orientarci, comprare, vendere. Ci affidiamo agli *smartphone* per documentare i luoghi in cui andiamo, le cose che facciamo, le nostre frequentazioni; contiamo su di loro per riempire gli spazi vuoti, i momenti di pausa, e i silenzi che occupavano abitualmente la nostra vita.

Gli *smartphone* hanno di fatto modificato l'intero tessuto della vita quotidiana, riorganizzando completamente spazi e rituali di vecchia data e trasformandone altri in modo irriconoscibile²¹.

²¹ Greenfield, A., *Tecnologie radicali. Il progetto della vita quotidiana*. Einaudi, collana I Maverich, 2017

Attraverso l'uso di questo artefatto e in generale delle nostre protesi digitali, diamo vita ad una quantità di dati in grado di offrire nuovi sguardi e nuovi contenuti informativi su fenomeni emergenti e su noi stessi. La vera sfida è estrarre valore, non semplicemente economico, da queste sorgenti di dati.

Secondo una classificazione proposta dall'UNECE (Commissione economica per l'Europa delle Nazioni Unite) esistono tre tipi principali di fonti di dati che possono essere viste come big data:

- ✓ di origine umana (ad es. piattaforme social),
- ✓ mediata da processi (ad es. record)
- ✓ generati dalla macchina (ad es. misurazioni dei sensori).

In linea molto generale, i dati di origine umana sono i dati meno strutturati e i dati generati dalla macchina sono i dati più strutturati. I dati basati sui processi hanno caratteristiche miste di dati di origine umana e dati generati da macchine.

Tabella 1 – Fonti dati Classificazione UNECE

Source	Structure	Human influence
Human sourced	Loosely structured	Direct
Process mediated	Structured	Indirect (e.g., data entry activities)
Machine generated	Well structured	None

Fonti di informazioni di origine umana

Queste informazioni sono la registrazione delle esperienze umane, precedentemente registrate in libri e opere d'arte, e successivamente in fotografie, audio e video. Le informazioni di origine umana sono ora quasi interamente digitalizzate e archiviate ovunque, dai personal computer ai social network. I dati sono spesso non governati. Ciò include una vasta quantità di tipi di dati come: social network (Facebook, Twitter, LinkedIn,

ecc.), blog e commenti, ricerche in Internet sui motori di ricerca (Google, ecc.), video caricati in Internet (YouTube, ecc.), mappe generate dagli utenti, archivi di immagini (Instagram, Flickr, Picasa, Google Foto, ecc.), dati e contenuti da telefoni cellulari (sms, ecc.), e-mail e così via.

Fonti mediate dal processo

I processi aziendali registrano e monitorano gli eventi aziendali di interesse, come la registrazione di un cliente, la produzione di un prodotto e l'acquisizione di un ordine. I dati mediati dal processo così raccolti includono transazioni, tabelle di riferimento e relazioni, nonché i metadati che definiscono il contesto. I dati aziendali tradizionali sono la stragrande maggioranza delle informazioni gestite ed elaborate dalle tecnologie informatiche, sia nei sistemi operativi che di Business Intelligence (BI). I dati mediati dal processo sono generalmente strutturati e archiviati in sistemi di database relazionali. Ne sono un esempio: dati prodotti da enti e istituzioni pubbliche (cartelle mediche, ecc.) e dati prodotti dal settore privato (transazioni commerciali, archivi bancari/stock, e-commerce, carte di credito, ecc.).

Sorgenti generate dalla macchina

Le sorgenti generate dalle macchine sfruttano l'impressionante crescita del numero di sensori e macchine utilizzati per misurare e registrare gli eventi e le situazioni nel mondo fisico anche attraverso l'IoT—Internet-of-Things. L'*output* di questi sensori è costituito da dati generati dalla macchina e da semplici record di sensori a complessi registri di computer, è ben strutturato. Con la proliferazione dei sensori e la crescita dei volumi di dati, sta diventando una componente sempre più importante delle informazioni archiviate ed elaborate da molte aziende. La sua natura ben strutturata è adatta per l'elaborazione al computer, ma le sue dimensioni e velocità sono al di là degli approcci tradizionali. Esempi includono: dati da sensori fissi (sensori di automazione degli edifici, sensori meteo/inquinamento, sensori di traffico/telecamere web, sensori scientifici, video/immagini di sicurezza/sorveglianza, ecc.), dati da sensori mobili, ovvero per scopi di tracciamento o analisi (immagini satellitari, GPS,

posizioni di telefoni cellulari, dispositivi dell'auto, ecc.) e dati da sistemi informatici (file di registro, registri Web, ecc.)²².

1.3 Il valore dei big data

Sebbene non vi sia una univoca definizione di big data, si tratta infatti senza dubbio di una rivoluzione nel modo di produrre ed usare la conoscenza con un rilevante, immediato impatto sulla società (Mayer-Schönberger, Cukier 2013). *Big* non è affatto sinonimo di qualità, accumulare molti dati non vuol dire aumentare l'informazione, anche il matematico e fisico teorico francese Henri Poincaré ci ricordava che *“l'accumulazione dei dati non è scienza più di quanto un mucchio di mattoni non sia una casa”*. Troppe informazioni tendono a comportarsi come se fossero pochissime. L'errore sta nel considerare i big data come se fossero solo tanti dati, si tratta di dati qualitativamente nuovi sui comportamenti e le convinzioni delle persone, che offrono nuovi tipi di strumenti e mostrano nuovi tipi di attori. Big data è, infatti, una locuzione che indica lo sviluppo di tecnologie capaci di conservare, combinare ed analizzare enormi volumi di dati provenienti da fonti eterogenee. Nonostante l'espressione evochi in primo luogo l'aspetto dimensionale del fenomeno, il vero valore è rappresentato *«dalla nuova espansiva capacità di connettere, attraverso l'analisi algoritmica, datasets disparati, forgiando relazioni tra dati raccolti in differenti momenti e luoghi e per diverse finalità»*²³. Un tale uso delle informazioni permette di creare nuova conoscenza, fare previsioni e risolvere problemi, formulare e testare nuove ipotesi e migliorare la prevedibilità di modelli predittivi e scientifici, tanto che al fenomeno Big Data e alla *data analytics* è stato riconosciuto un ruolo importante per la crescita economica, lo sviluppo ed il benessere²⁴.

²² Firmani, D., Mecella, M., Scannapieco, M. *et al.* On the Meaningfulness of “Big Data Quality” (Invited Paper). *Data Sci. Eng.* **1**, 6–20 (2016). <https://doi.org/10.1007/s41019-015-0004-7>

²³ Zook M., Barocas S., Boyd D., Crawford K., Keller E., Gangadharan SP, et Al. *Ten simple rules for responsible big data research*. *PLoS Comput Biol* 13(3): e1005399. <https://doi.org/10.1371/journal.pcbi.1005399>, 2017

²⁴ OECD *Data driven innovation: Big Data for Growth and Well Being*, OECD Publications, Paris https://read.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en#page302, 2015

Se non si riducono i big data a una questione di volume e di capacità computazionale si valorizza la dimensione fondamentale che li caratterizza: quella epistemologica che si occupa della modalità con cui si sviluppa la conoscenza scientifica. Luciano Floridi (2017) ci ricorda infatti che se teniamo conto solo del volume dei big data e delle infrastrutture tecnologiche utili a processarli, si insinuano nel ragionamento due valutazioni inesatte, la prima che il problema epistemologico risieda nel fatto che vi siano troppi dati e la seconda che la soluzione del problema epistemologico sia di natura tecnologica legata alla presenza di maggiori e migliori tecniche che riporteranno i big data entro i confini gestibili. Il problema epistemologico è differente e richiede una soluzione epistemologica. L'espressione big data è stata formulata dopo che espressioni di riferimento simili come "marea informativa" o "sovraccarico informativo" avevano preso a scomparire sebbene l'idea fosse rimasta la stessa. Questa idea fa riferimento alla sensazione di inadeguatezza per cui abbiamo ingerito più di quanto siamo in grado di masticare, come oche nutrite a forza il cui fegato intellettuale sta esplodendo. Si tratta di un errore. Sicuramente abbiamo assistito ad una palese crescita esponenziale di dati su un numero sempre più elevato di temi, ma lamentarsi di tale sovrabbondanza equivarrebbe a lamentarsi di un banchetto in cui sia offerto più di quanto saremo mai in grado di mangiare. I dati restano un bene e una risorsa da sfruttare. Stiamo diventando più ricchi di dati ogni giorno, ma ciò non può costituire il problema essenziale. Dal momento che il problema non risiede nella ricchezza crescente di dati a nostra disposizione, la soluzione evidentemente deve essere ripensata non può consistere solo nel numero di dati che siamo tecnologicamente più in grado di processare (Floridi, 2017), in sostanza la definizione di big data non può quindi esaurirsi nelle *p-words* o nelle *v-words* cui abbiamo fatto riferimento.

I big data rappresentano nuove fonti informative, sono una finestra aperta sull'uomo e sui nuovi comportamenti e significati generati dal digitale. Il valore dei big data non è solo meramente economico, non rappresentano semplicemente il *nuovo petrolio* come ha affermato nel 2006, il data scientist e matematico inglese, Clive Humby secondo cui così come il petrolio ha permesso lo sviluppo socio-economico mondiale tra il XIX ed il XX secolo, nel XXI secolo

sono le connessioni, le tecnologie ed i dati a svolgere questo importante ruolo. I dati non creano solo valore economico, ma creano significati, hanno un valore sociale perché creano un nuovo umanesimo digitale e informativo perché ci permettono anche di misurare fenomeni emergenti. I big data sono considerati un fenomeno sociale proprio perché hanno un grande potenziale informativo ma non possiedono una loro autonomia informativa. Hanno reso più quantificabile lo spazio sociale, rappresentano una lente nuova per poter leggere le tracce che ogni giorno, anche inconsapevolmente, lasciamo nelle nostre interazioni con la rete sta tanto nel fatto che facciamo azioni attraverso gli strumenti digitali, quanto piuttosto che qualsiasi cosa che facciamo lascia traccia. Abbiamo un potere conoscitivo che prima non avevamo, i big data possono creare nuovo potere informativo, la loro caratteristica più rilevante non è, come già stato detto, la loro dimensione, ma la misura in cui sono informativi. Però, pur producendo con le nostre azioni quotidiane una quantità enorme di dati, ci troviamo nella contraddittoria situazione in cui, mentre affoghiamo nei dati, moriamo di fame per conoscenza, siamo ricchi di dati e poveri di informazioni²⁵.

Il digitale, in quanto nuova infrastruttura sociale, ha dato vita a nuovi processi, fra questi quelli che riguardano il ciclo di vita dei big data fino all'estrazione da essi di quel valore che va a riempire il vertice della piramide della conoscenza (che vedremo nel paragrafo 5) , rispondendo a nuove esigenze informative e a nuovi *stakeholders*. Il vero big dei dati è l'essere carburante per le macchine della conoscenza che finora avevamo tenuto in garage²⁶. Gli ultimi eventi legati alla pandemia hanno fatto emergere la necessità di poter disporre di dati "parlanti" e di informazione affidabile elementi di ampia rilevanza nei processi di sviluppo dell'informazione e anche nei processi di *decision making*.

La questione con i big data è quindi una questione anche epistemologica e riguarda la possibilità di disporre di *pattern*, cioè strutture, di modelli di piccola scala. Oggi vengono generati così tanti dati e in modo così veloce, economico e potenzialmente su tutto, che la

²⁵ Giannotti F., Nanni M., Thanos C., Rauber A., *Big Data Analytics: towards a European research agenda*, White paper ERCIM, 2 gennaio 2015.

²⁶ Vespignani A., *L'algoritmo e l'oracolo*, Il Saggiatore, 2019.

difficoltà quindi è proprio quella di individuare dove reperire, in tali immensi database, nuovi *pattern* dotati di valore aggiunto e in che modo questi ultimi possano essere sfruttati per la creazione di ricchezza, per lo sviluppo delle vite umane e il progresso della conoscenza. Si tratta di un problema di potenza intellettuale e non computazionale (Floridi, 2017). I big data non possiedono una loro intrinseca autonomia informativa, ma richiedono attività di trattamento automatizzato, mediante algoritmi e altre tecniche avanzate al fine di individuare correlazioni di natura per lo più probabilistica, tendenze e/o modelli che ci permettono di risalire sempre più il triangolo della conoscenza che vedremo più avanti.

La disponibilità di questi dati e la loro potenziale capacità di creare valore conoscitivo, ci permette di affermare che il vero protagonista della rivoluzione che stiamo vivendo e a cui abbiamo fatto riferimento all'inizio, è l'informazione. Più che una nuova rivoluzione industriale, è in atto la *quarta rivoluzione* scientifica (dopo quella copernicana, darwiniana e psicoanalitica) ed è quella dell'informazione che è in grado di provocare, come le precedenti, un grande cambiamento del pensiero umano su di sé e sul mondo. I big data hanno dato e continueranno ad offrire un grosso contributo all'informazione: grazie ai big data, per la prima volta nella storia, la nostra società sta diventando misurabile, o comunque stanno diventando misurabili anche fenomeni emergenti che non possono essere misurati con gli strumenti tradizionali. Le attività umane, tutte ormai mediate dalle tecnologie dell'informazione e della comunicazione, lasciano tracce digitali. Questo è un fatto che sta cambiando prepotentemente il mondo. Invece di porre le macchine al centro, si sta creando un nuovo *umanesimo* digitale, che consiste nel mettere la società e le persone al centro dell'investigazione scientifica e del dibattito pubblico²⁷. Non sono le tecnologie al centro, ed è per questo che la definizione di rivoluzione industriale poco si addice al nostro lavoro, le tecnologie sono strumenti abilitanti, sistemi di relazione con il contesto esterno, al centro ci sono i dati che attraverso le tecnologie si riesce a *captare* (cfr. par. 2): le nostre attività sui cosiddetti social network, i nostri acquisti elettronici, le telefonate, gli spostamenti tramite GPS che rimandano il segnale della nostra presenza al satellite corrispondente. Sensori di

²⁷ Pedreschi D., *Explanaible AI. Aprire le scatole nere per un'intelligenza artificiale umana*. Gnosis 2/2019.

traffico automobilistico e droni rilevano dati e immagini senza bisogno della nostra partecipazione attiva. I big data sono quindi una componente di quell'infrastruttura sociale rappresentata dal digitale che comporta una trasformazione e una reingegnerizzazione non solo di processo ma soprattutto di significato.

In questo contesto la tecnologia rappresenta una parte quindi di una questione molto più ampia che riguarda un ripensamento in termini di processo complessivo di formazione, interpretazione e comunicazione della conoscenza.

I big data sono qui per restare e anche se i big data non si qualificano più come un "fenomeno emergente", continuano a suscitare interesse, a diventare più grandi e a determinare molti cambiamenti nella società e a rappresentare come suggeriscono Dryden e Hodge una *nuova risorsa naturale*²⁸. Molti sono gli studiosi che, a partire da Boyd e Crawford, vedono i big data come fenomeno sociale complesso basato su un'integrazione di fattori tecnologici, scientifici e culturali. Boyd e Crawford, in uno degli articoli più influenti sul tema in questione, hanno paragonato gli impatti provocati dai big data a quelli introdotti dal fordismo nei primi decenni del XX secolo quando Henry Ford ideò un nuovo sistema di produzione di massa che prevedeva l'uso di macchinari specializzati e prodotti standardizzati. Proprio come dal fordismo è emerso un nuovo modello di produzione e una nuova concezione del lavoro, dai big data emergono, secondo i due studiosi, nuovi ambiti di conoscenza, nuove modalità di comprensione delle relazioni umane e della comunità, i big data rappresentano una svolta computazionale nel pensiero e nella ricerca: così come ci ricorda Latour *"Cambia gli strumenti e cambierai l'intera teoria sociale che li accompagna"*.

I big data provocano un cambiamento radicale nel modo in cui pensiamo alla ricerca a livello di epistemologia ed etica, si ridefiniscono le domande chiave sulla costituzione della conoscenza, sui processi di ricerca. Du Gay e Pryke notano che *"gli strumenti contabili non aiutano semplicemente la misurazione dell'attività economica, ma modellano la realtà che misurano"* e che le modalità usate e per misurare un dato fenomeno influenzano il modo di pensare a

²⁸ Dryden, I. Hodge, D. J. Journeys in big data statistics, *Statistics & Probability Letters*, Volume 136, Pages 121-125, ISSN 0167-7152, <https://doi.org/10.1016/j.spl.2018.02.013> 2018.

quella realtà e all'oggetto stesso di misurazione. Allo stesso modo i big data puntano su nuovi ambiti di ricerca, nuove domande, metodi di conoscenza e definizioni di vita sociale.

1.4 Big data e paradigmi che mutano.

Thomas Kuhn nel 1962 sosteneva che la scienza si trasforma periodicamente da un paradigma dominante (un modo accettato di interrogare il mondo e sintetizzare la conoscenza) a un altro. Per paradigma Kuhn intendeva un insieme consolidato di conoscenze, valori e principi che sistematicamente viene sfidato e destabilizzato da una nuova serie di idee, conoscenze, valori e principi, raggiungendo infine un punto di svolta in cui quest'ultima sostituisce la prima. Un cambiamento di paradigma si verifica, secondo Kuhn, quando il modo dominante della scienza non può spiegare fenomeni particolari o rispondere a domande chiave. Nel 2007 Jim Gray, un noto informatico che ha lavorato anche per Microsoft e IBM che ha vinto il premio Turing assegnato ai grandi matematici, durante un suo discorso al *Board* della *Computer Science Telecommunication* (riportato da Hey et al., 2009)²⁹ ha affermato che – contrariamente a quanto detto da Kuhn - le transizioni tra i paradigmi possono essere fondate anche sui progressi nella produzione di dati e sullo sviluppo di nuovi metodi analitici. In riferimento ai cambiamenti introdotti dalle nuove fonti di dati, Tony Hey, Stewart Tansley e Kristin Tolle parlano esplicitamente di un nuovo paradigma epistemologico, il *quarto*³⁰ in grado di produrre nuova conoscenza, chiamata *eScience* che secondo Gray non è né *sensata esperienza*, né certa dimostrazione e neppure simulazione, è il luogo dove “*IT meets scientists*” è la sintesi cioè tra tecnologia dell'informazione e scienza.

²⁹ Hey, T., Tansley, S. Tolle, K. Jim Grey on eScience: *A transformed scientific method*, in Hey, T., Tansley, S. and Tolle, K. (eds) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research; Redmond, Washington. xvii-xxi, 2009.

³⁰ Greco P., Il quarto paradigma, *Micron*, n. 23 ottobre 2013, consultabile al link: https://www.arpa.umbria.it/resources/docs/micron%2023/06-MICRON_23.pdf

La *eScience* è, secondo Gray il quarto paradigma della scienza. Gray sostiene che siamo entrati in una nuova era nella produzione della conoscenza scientifica. I dati digitali e i loro metodi hanno contribuito non solo al passaggio dalla causazione alla correlazione, ma hanno addirittura definito un nuovo modo di produrre conoscenza: la scienza esplorativa fondata sull'idea che grandi moli di dati possano essere facilmente trasformate in una nuova forma di conoscenza scientifica. Jim Gray sostiene che il quarto paradigma consiste nel *navigare in un mare sconfinato di dati alla ricerca (anche) di ordine e regolarità che non vediamo e che le teorie non prevedono*. Si tratta di una navigazione interdisciplinare capace di generare nuova conoscenza. Siamo in una fase di transizione epistemologica, abbiamo a disposizione non solo una gran massa di dati in ogni settore ma anche la possibilità tecnica di navigarci dentro anche con il pilota automatico (gli algoritmi). Sarebbe un peccato perdere o ridurre fortemente questa opportunità³¹.

Qualche anno dopo il discorso di Gray, Lazer ed altri, nell'ormai famoso articolo *Life in the network, the coming age of computational social science* hanno identificato i dati di Internet come il cuore della Computational Social Science (CSS): per gli autori le nostre vite immerse nella rete rappresentano una svolta per la ricerca sociale. L'enorme disponibilità di dati e di tracce digitali disseminate nelle discussioni sulle varie piattaforme, negli scambi di email e messaggi al cellulare, nelle transazioni con carte di credito, nelle immagini catturate da telecamere di videosorveglianza, sono al cuore di un radicale cambiamento epistemologico e metodologico. Stiamo assistendo all'emergere di un nuovo campo disciplinare, sottolineavano Lazer e gli altri, in cui confluiscono sociologia, statistica, informatica, matematica, economia, scienze politiche. Per questo motivo, affermavano gli autori, big data e tecniche di analisi computazionale dovrebbero diventare pane quotidiano per gli scienziati sociali³².

³¹ Greco P. *Il quarto paradigma*, pubblicato il 3/09/2012 su www.scienzainrete.it

³² Pitrelli N., *Big data e metodi digitali per la ricerca in comunicazione della scienza: opportunità, sfide e limiti*, Journal of Science Communication 16(02)2017

Nello stesso anno in cui Gray tenne il discorso diventato famoso, i sociologi Savage e Burrows pubblicarono sulla rivista *Sociology* un articolo dal titolo *The coming crisis of empirical sociology*, nel quale mostrarono preoccupazione sui possibili rischi connessi all'emergere e al diffondersi di nuove forme di dati in formato digitale: secondo gli studiosi, il processo di crescente digitalizzazione in atto avrebbe messo a rischio il ruolo della sociologia come forma pura di sapere e marginalizzato la ricerca accademica, spostando l'interpretazione dei fenomeni sociali dalle mani dei sociologi a quelle di altri esperti, in primis i professionisti dell'informatica, la cui attività di esplorazione del web sarebbe andata a vantaggio soprattutto di coloro che erano interessati a disporre di queste informazioni per fini commerciali. L'ampia disponibilità in termini quantitativi di questi dati avrebbe comportato ricadute sulle possibilità di condurre studi attraverso le tecniche di ricerca tradizionali, a favore di analisi di secondo livello condotte attraverso nuove metodologie. Anche Emma Uprichard, in un articolo già citato, ha addirittura affermato che l'avvento dei big data avrebbe comportato addirittura un *genocidio metodologico* perché avrebbe sminuito quasi del tutto il ruolo degli scienziati sociali e della ricerca qualitativa. Secondo Uprichard l'unica cosa che possono fare i big data è fornirci una "*polaroid*" una fotografia di uno specifico momento e aiutarci potenzialmente a coltivare una sana "*immaginazione sociologica*", ma questo non basta per aiutarci ad affrontare i grandi problemi sociali del mondo di oggi e ad apportare modifiche sostanziali a loro domani. Questo tipo di preoccupazioni si è rilevato infondata, infatti, come è emerso nel corso degli anni successivi, tutte le tecniche che hanno avuto e stanno avendo grande rilevanza nella ricerca digitale "*hanno un'origine interdisciplinare e possono essere qualificate come tecniche mixed, nel senso che combinano elementi computazionali ed elementi sociologici*"³³, cioè elementi tecnologici ed epistemologici.

Le tecniche native digitali non sono per niente aliene alla sociologia, ma anzi, le tecniche incorporate nelle piattaforme e nei dispositivi digitali sono comunque costruite sulla base di principi metodologici consolidati e duraturi³⁴.

³³ Marres N., *Digital Sociology. The reinvention of social research*. Polity Press 2017, p 104

³⁴ Amaro E., Aragona B, *Per un'epistemologia del digitale: note sull'uso di big data e computazione nella ricerca sociale*, Quaderni di Sociologia, 81- LXIII | 2019.

Le scienze sociali computazionali permettono di avviare interessanti esplorazioni del nostro tempo: misurano alcune dinamiche sociali, lavorando sui big data e sulle numerose tracce che lasciamo nel nostro girovagare fra i vari ambienti del web. Numeri percorsi e intrecci se ben modellati, danno la possibilità di identificare ed accedere, in maniera empirica, ai meccanismi di azione e interazione sociale, partendo dal “micro” e arrivando al “macro” permettendo un’analisi della società contemporanea che supera i limiti della teoria e delle speculazioni³⁵. Nell'era dei big data le nuove tipologie di analisi daranno origine ad un'ondata di nuove scoperte, utili previsioni, noteremo dei collegamenti che in precedenza non avevamo notato. Capiremo complesse dinamiche tecniche e sociali che erano sempre sfuggiti alla nostra comprensione nonostante tutti gli sforzi, ma soprattutto queste analisi non causali ci aiuteranno a interpretare il mondo chiedendoci cosa e non perché. I big data ci aiutano ad avvicinarci maggiormente alla realtà rispetto a prima, quando eravamo legati a doppio filo agli small data e all’accuratezza. (Schonberger, Cukier2013).

Il digitale è un ambiente nuovo che crea nuovi processi e ne istituzionalizza altri. La diffusione di questi dati digitali abbiamo visto non richiede solo soluzioni tecnologiche necessarie per processarli, ma richiede quindi nuovi paradigmi o, per dirla con Batini, la considerazione di una nuova categoria culturale: la *datacy* che prevede la comprensione dell’impatto che i dati, unitamente alla pervasività delle tecnologie digitali, hanno sulla società. La *datacy* presuppone l’elaborazione di un nuovo quadro normativo per regolare lo sfruttamento e l’accesso ai dati, i processi di diffusione, lo scambio e l’uso dei dati, in generale richiede una nuova *governance* dei dati. In questo senso, parlare di *datacy*, del suo significato, dei modelli e degli strumenti atti ad analizzarla, diventa fondamentale nel disegno di un progetto e di una strategia di crescita digitale del paese³⁶. Abbiamo detto che il digitale rende possibili nuovi processi e i big data rende possibili nuovi processi conoscitivi che possono

³⁵ Quattrociochi, W. Vicini A. *Liberi di crederci. Informazione, internet e post-verità*, Codice editore, collana Le Scienze, 2018

³⁶ Batini, C., *Datacy, perché una scienza per studiare l’impatto dei dati sulla società*. <https://www.agendadigitale.eu/cittadinanza-digitale/datacy-perche-una-scienza-per-studiare-limpatto-dei-dati-sulla-societa/> Articolo pubblicato su [agendadigitale.eu](http://www.agendadigitale.eu) il 19/03/2018.

essere istituzionalizzati attraverso l'adozione di nuovi paradigmi. Riprenderemo questo concetto più avanti per chiarirlo e per renderlo maggiormente "visibile" adattandolo al funzionamento degli Istituti di Statistica che hanno cominciato ad integrare le nuove fonti nei processi di produzione statistica per continuare a fornire statistiche ufficiali a supporto dei processi di *decision making*.

1.5 Dai dati alla conoscenza. Nuove domande e nuove competenze.

Williams Edgar Demings, noto statistico statunitense, già nel 1942 sosteneva *"i dati non vengono raccolti per scopi museali, ma come base per fare qualcosa. Se non si deve fare nulla con i dati, allora non serve a nulla raccogliarli. Lo scopo ultimo della raccolta dei dati è fornire una motivazione o un suggerimento per un'azione"*. Lo stesso Demings affermava che *"senza dati sei solo un'altra persona con un'opinione"*. Per far sì che i dati servano a qualcosa dobbiamo però estrarne valore conoscitivo e farli diventare utili.

Tutti i dati a nostra disposizione, che siano essi quantitativi, qualitativi, strutturati e non ecc..., hanno una caratteristica comune: formano la base o il fondamento di una piramide della conoscenza: i dati precedono l'informazione, che precede la conoscenza, che precede la comprensione e la saggezza (come appare nella figura 1).³⁷

³⁷ Kitchin R., *The Data Revolution. Big Data, open Data, Infrastructure and their Consequences*, SAGE, 2014

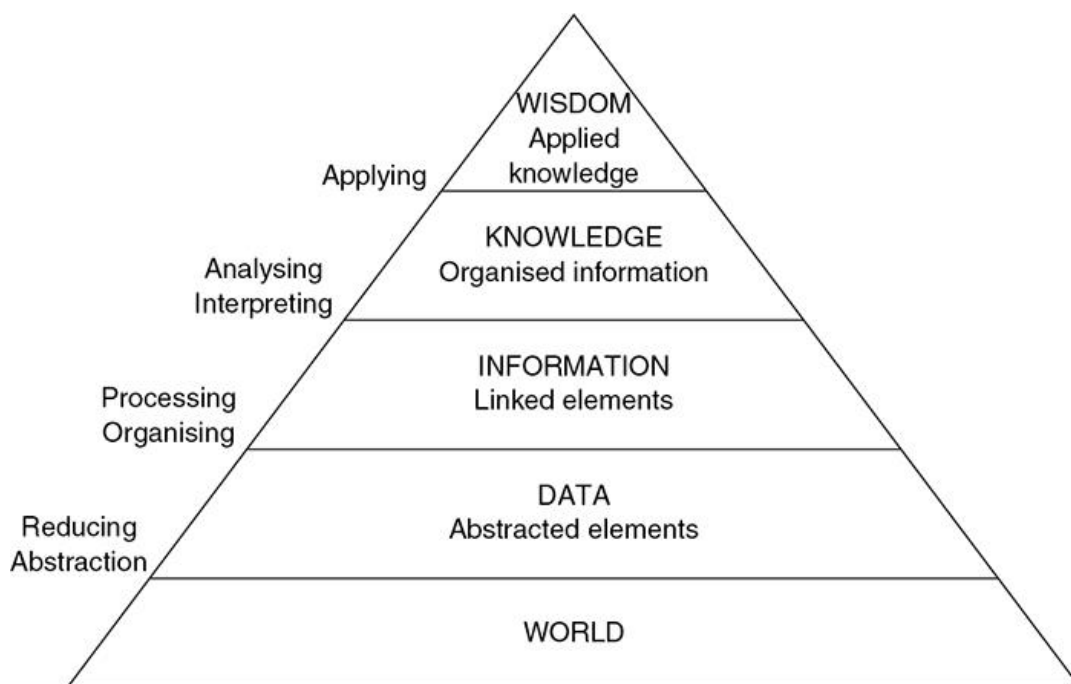


Figura 1 - Piramide della conoscenza. Dal Dato alla conoscenza applicata.

Avere tanti dati a disposizione non vuol dire avere tanta informazione, come emerge dalla piramide qui sopra i dati hanno bisogno di essere processati, analizzati e interpretati per creare nuove forme di conoscenza. Weinberger dichiara: "*L'informazione rappresenta per i dati ciò che il vino rappresenta per il vigneto: il delizioso estratto e distillato*"³⁸.

I dati sono registrazioni dell'esperienza umana, una sorta di grande osservatorio con cui guardare la società, offrono capacità di conoscenza, decisione e pianificazione purché siamo in grado di porre loro nuove domande di ricerca per evitare risposte che non cambiano la nostra visione del mondo, che non contribuiscono a modificare né il nostro universo conoscitivo, né il sistema informativo generale. I dati sono un *asset*, il loro valore risiede nella capacità di renderli utili fornendo nuovi strumenti di autoconsapevolezza. L'avvento dei big data, infatti, non determina semplicemente un salto di scala dal piccolo al molto grande, ma implica una riflessione sulla misura in cui questi dati possono essere informativi.

³⁸ Weinberger, D, *La stanza intelligente: La conoscenza come proprietà della rete*, Torino, Codice Edizioni, 2012

Quando si parla di portata epistemologica o della capacità dei dati di creare nuovi significati e nuovi processi si fa riferimento alla loro capacità di riformulare le domande sulla costituzione della conoscenza. L'esplorazione dei big data è un esercizio per porre domande nuove e migliori, nonché un'opportunità per sfidare il nostro pensiero convenzionale sulla raccolta dei dati e sulla produzione di statistiche. Claude Lévi-Strauss ci ricorda che *“la mente scientifica non fornisce tanto le giuste risposte, quanto le giuste domande”*. Senza scomodare un filosofo e antropologo della levatura di Levi Strauss, se pensiamo al romanzo di Douglas Adams del 1979 *“Guida galattica per autostoppisti”*, dobbiamo evitare che i dati rispondano “42” alle nostre domande. 42 è infatti la risposta che un'intelligenza artificiale definita Pensiero Profondo ha dato, dopo aver riflettuto per sette milioni e mezzo milioni di anni “alla vita, all'universo, a tutto quanto” giustificando la sua risposta con questa motivazione: *“ho controllato molto approfonditamente e questa è sicuramente la risposta. Ad essere sinceri penso che voi non abbiate mai saputo qual è la domanda”*.

A distanza di più di 40 anni dalla risposta di Pensiero Profondo, per evitare di non riuscire a sfruttare appieno le potenzialità offerte dalle nuove fonti dati e della mente scientifica, è arrivato il momento di porre loro le domande giuste, al fine di ottenere nuova conoscenza e migliorare la qualità dell'informazione³⁹.

Per far fronte alla complessità di questi dati e ricavarne le risposte che possono dare a nuove domande, servono nuovi modelli di analisi e nuovi paradigmi che consentano di non rimanere vittime dell'*information overflow*, dell'infodemia ovvero una sovrabbondanza di informazione a cui il nuovo ambiente digitale ci espone. Fino a quando non vengono elaborati e interpretati tramite processi di analisi, i big data sono generalmente inutili poiché il primo sguardo non rivela informazioni. Serve la *scienza dei dati* che è l'insieme di principi metodologici e tecniche multidisciplinari volti a estrarre informazione da dati non strutturati o provenienti dalle fonti più disparate attraverso un approccio multidisciplinare finalizzato a comprendere e raccontare la realtà da diversi punti di vista sui vari ambiti tematici. Maggiore è la complessità dei dati più difficile diventa il compito di immagazzinarli, gestirli

³⁹ Epifani S., *Sostenibilità digitale* Digital Transformation Institute, 2020.

in database, collegarli fra loro e ottenere *patterns* interessanti. La varietà dei dati, ancor più che il loro volume e la loro velocità, sta costringendo ad una revisione radicale non solo dei procedimenti di raccolta e di immagazzinamento, ma anche delle metodologie di analisi⁴⁰. Mettere in condizione i dati di raccontarci delle storie individuali o collettive è un'operazione complessa e la scienza dei dati consiste in questa ricerca di senso, ci aiuta a estrarre dai dati i nuovi significati, le nuove pratiche, le nuove relazioni prodotte dal digitale. Tre sono i fattori, derivati dalla trasformazione digitale, che hanno determinato lo sviluppo della data science: la grandissima disponibilità di dati; i progressi nell'analisi dei dati grazie a tecniche di *machine learning* ed apprendimento automatico; i progressi nella capacità computazionale delle infrastrutture di calcolo ad alte prestazioni. I tre elementi, giunti alla maturità insieme negli ultimi anni, sono un mix esplosivo: i big data forniscono la massa critica di esempi concreti da cui imparare a generalizzare; l'analisi è in grado di produrre modelli predittivi e modelli comportamentali da questi dati; piattaforme di elaborazione scalabili consentono di inserire dati ed eseguire analisi⁴¹. Per passare dai dati all'informazione sono sicuramente importanti le tecnologie abilitanti che consentano di «ottenere dal trattamento di questi dati, grazie ad algoritmi che sappiano interrogare la macchina in modo da avere da essa la risposta voluta o la informazione ricercata, una quantità ancora più sterminata di nuovi dati, che consentano nuove conoscenze ed analisi relative ai fenomeni naturali ed ai comportamenti umani»⁴², ma sono importanti le competenze che sappiano progettare e supportare questi nuovi approcci inferenziali, servono competenze multidisciplinari che hanno portato allo sviluppo della data science. La consapevolezza della necessità di questo approccio multidisciplinare è ben descritta nella figura n. 2 che contiene il modello Pi-shaped, aggiunge un'altra competenza verticale, relativa alle capacità statistiche e computazionali necessarie per affrontare l'analisi di (big) data. Tale competenza è utile in tutti i campi, comprese le

⁴⁰ Arbia, G., *Statistica, nuovo empirismo e società nell'era dei big data*, Edizione Nuova Cultura, 2018.

⁴¹ Pedreschi D., *The black Box Society: come fanno le imprese a fidarsi dei loro prodotti*, articolo apparso su il Sole 24Ore il 23 luglio 2018 e disponibile al link <https://www.ilsole24ore.com/art/the-black-box-society-come-fanno-imprese-fidarsi-loro-prodotti-AEc7FIJF>

⁴² Pizzetti, F. M., *Privacy e il Diritto Europeo alla Protezione dei Dati Personali. Dalla Direttiva 95/46 al nuovo Regolamento europeo*, Torino, Giappichelli, 2016

discipline umanistiche: si pensi all'importanza di un corretto approccio alla scienza dei dati nelle scuole di giornalismo, i cui studenti dovrebbero essere formati a comprendere non solo la verità fattuale delle notizie riportate, ma anche la loro rilevanza statistica.⁴³

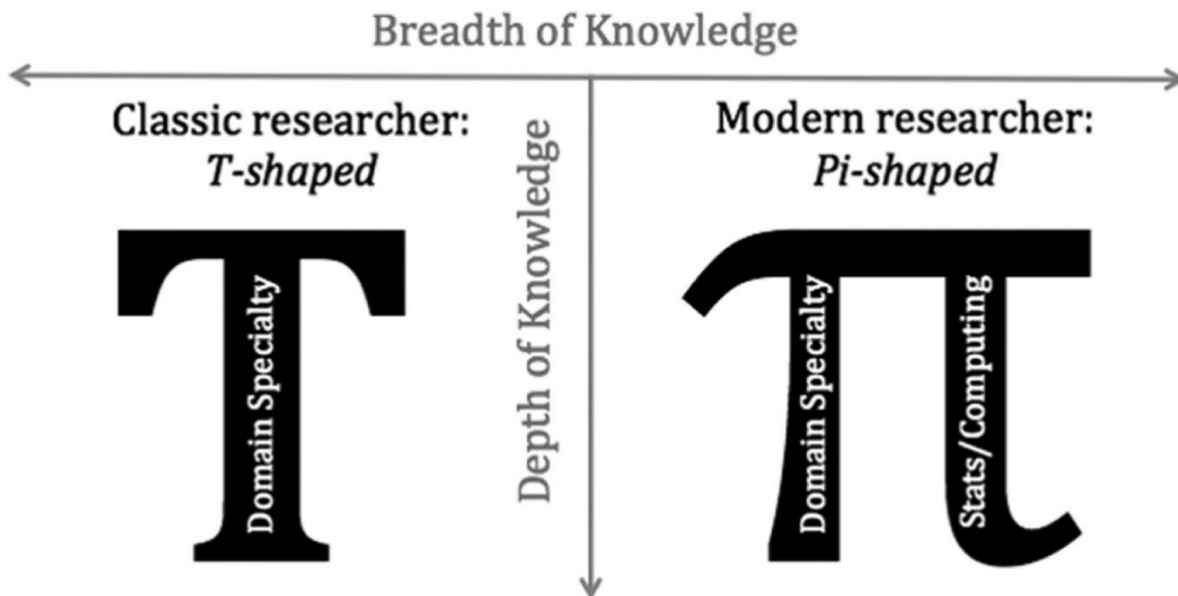


Figura 2 - Representations of T vs. Pi-shaped education.

Quindi come emerge dalla figura, per diventare un *modern researcher* e quindi adeguarsi a un mondo in continua evoluzione, alle specifiche competenze di carattere informatico rappresentate dalla I base della T e da *soft skills* rappresentate nella parte superiore della T, è necessario affiancare un'altra I rappresentata dalla statistica per dare vita a quel nuovo paradigma necessario per scalare la piramide della conoscenza.

L'informatica non è un mero fornitore di tecnologia contribuisce all'infrastruttura della scienza dei dati, in quanto si occupa degli aspetti di: acquisizione, rappresentazione, pulizia e integrazione dei dati; modellazione, metodi di intelligenza artificiale e *machine learning* ecc... D'altra parte, la statistica fornisce i pilastri per l'analisi dei dati: un approccio rigoroso alla scienza dei dati richiede una profonda conoscenza della teoria e dei metodi dalla

⁴³ Ceri S., *On the role of statistics in the era of big data: A computer science perspective*, Statistics & Probability Letters, Volume 136, 2018

statistica. Per dirla con le parole di Gray che abbiamo analizzato prima, la data science è, al pari della eScience, il luogo dove *IT meets statistics*. Ormai sono disponibili sempre più spesso insieme enormi di dati e programmi di analisi facili da usare, questo potrebbe far pensare che non è più necessario l'apprendimento dei metodi statistici e delle teorie sottostanti. È una conclusione estremamente ingenua. Gli insiemi di dati più grandi e la moltiplicazione degli studi scientifici sempre più complessi non ci sollevano dal compito di imparare la statistica, rendono ancora più difficile trarre conclusioni appropriate. L'abbondanza di dati ci impone una consapevolezza ancora maggiore di quanto sia significativo ciò che si riscontra⁴⁴.

L'informatica e la statistica dovrebbero abbracciare insieme la rivoluzione della scienza dei dati che realizza una formazione trasversale, applicabile a tutte le scienze: alla gestione aziendale quando si considera il valore economico dell'informazione; alla sociologia se ne consideriamo il valore sociale; alla biologia e alla medicina quando si considerano le scienze della vita, e così via. Il risultato principale della scienza dei dati è la creazione di valore e il beneficiario è l'intera umanità.

Come vedremo nel prosieguo del lavoro, quando sposteremo il focus sulle pratiche all'interno degli Istituti Nazionali di statistica (d'ora in poi INS nel testo), la scienza dei dati ha contribuito molto alla creazione di valore informativo, ed è l'ambito su cui gli Istituti di statistica stanno investendo di più per continuare a perseguire la propria *mission* e a "servire la collettività" attraverso la produzione di statistica ufficiale. Queste riflessioni sul ruolo della data science sono confermate anche dai risultati del questionario somministrato agli Istituti di statistica europei per conoscere gli investimenti fatti, in corso d'opera e programmati, sul tema risorse umane nell'ambito dell'integrazione dei big data nei processi di produzione statistica. I risultati sono illustrati nel Cap 4 del presente lavoro.

Avere tanto carburante (i dati) e un motore potente (la tecnologia) non bastano a vincere la gara: il talento del pilota rimane un ingrediente indispensabile⁴⁵

⁴⁴ Spieghealter D., *L'arte della statistica*, Einaudi I Maverick, 2019

⁴⁵ De Mauro, A, Impariamo a conoscere i big data. <https://www.apogeonline.com/articoli/impariamo-a-conoscere-i-big-data-andrea-de-mauro/> 15 novembre 2019.

A mio avviso la competenza dei pilota è rappresentata dalla multidisciplinarietà insita nella data science, nel dialogo tra comunità di ricerca diverse, tra esperti con varie competenze. Questo approccio è cruciale anche per individuare nuovi metodi di analisi. L'integrazione dei dati da sola non basta ma deve essere accompagnata da una integrazione di competenze e diverse comunità professionali se vogliamo far parlare i dati e indagare fenomeni sociali in diversi settori e dare impulso ad un vero ecosistema della conoscenza.

Contrariamente a quanto affermato in maniera del tutto fideistica da Chris Anderson, i dati da soli non ci dicono nulla.

1.6 I dati da soli sanno parlare? E la correlazione basta?

"When you're trying to reach a goal, data not only tells you if you're succeeding, but it also suggests which activities you should do more of in order to improve your results." Ho scelto di aprire questo paragrafo con queste parole di Bill Gates per sottolineare l'importanza di poter disporre di dati "parlanti" vale a dire di dati che effettivamente possano supportare i processi di *decision making* per tutte le tipologie di *stakeholders*. Mettere i dati in condizioni di parlare non è un'impresa semplice.

Chris Anderson era redattore della rivista Wired e proprio sulle pagine di Wired nel 2008 pubblicò un articolo⁴⁶ volutamente provocatorio in cui proponeva un cambio di paradigma che renderebbe, qualora applicato, non più necessario il metodo scientifico classico: non è più necessario partire dalle ipotesi, verificarle sperimentalmente, raccogliere dati e formulare una teoria, ma basta partire proprio dai dati stessi: sui *dataset* a disposizione, l'intelligenza artificiale applica una serie di algoritmi, fino a trovare quello che meglio descrive il comportamento del fenomeno: Anderson propone, di fatto, la fine della teoria. *"Con dati sufficienti, i numeri parlano da soli"* senza che gli esseri umani pongano domande intelligenti. *"In presenza di dati massicci - questo approccio della scienza - ipotesi, modello, test - diventa*

⁴⁶ Anderson, C. *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*. Wired, 23/06/2008 http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

obsoleto...ora c'è un migliore approccio. I petabyte ci permettono di dire che la correlazione è sufficiente e sostituisce la causalità. La scienza può avanzare anche senza schemi coerenti".

In questo tipo di scienza la ricerca è guidata dai dati, la costruzione di modelli per spiegare la realtà sociale non è più fondamentale, la conoscenza può avanzare anche senza teorie di riferimento. Come osserva Croll, con la scienza esplorativa si sovverte il modo in cui sono formulate le ipotesi: «prima raccogliamo i dati e poi formuliamo le ipotesi»⁴⁷, in opposizione al metodo ipotetico deduttivo. In questo approccio è insito un pregiudizio tipico del dataismo e cioè che la realtà è conoscibile attraverso la mera lettura dei dati: non ci si preoccupa di come si costruisce la conoscenza perché la conoscenza si produce automaticamente nel momento in cui si leggono i dati, ma i dati non sono autoevidenti. Secondo questa *teoria* la società deve abbandonare almeno in parte la sua ossessione per la causalità, in cambio di correlazioni semplici: non dovrà chiedersi perché ma solo cosa. Questo nuovo modo di affrontare i problemi ribalta secoli di prassi consolidate e mette in crisi il nostro approccio istintivo alle decisioni e alla comprensione della realtà. Il fenomeno dei big data segna l'inizio di una trasformazione radicale sintetizzata in queste parole «*Big data is about what, not about why. We don't always need to know the cause of phenomenon: rather, we can let data speak for itself*» (Mayer-Schonberger e Cuckier, 2013). In realtà il metodo scientifico può essere arricchito da immensi database, ma non sostituito da esso.

Anderson sottostimava un aspetto essenziale che era ben chiaro a Platone: la conoscenza è qualcosa in più dell'informazione poiché richiede spiegazione e comprensione e non soltanto verità e correlazioni. Il gioco della conoscenza sarà sempre vinto da chi, come dice Platone in uno dei suoi celebri dialoghi "*sa come domandare e come rispondere alle domande*" in modo critico e dunque sa quali siano i dati utili e rilevanti meritevoli di essere raccolti e curati per sfruttare i loro preziosi *pattern* vale a dire le piccole e decisive configurazioni e correlazioni stabili da trovare nell'immenso database a nostra disposizione (Floridi, 2017).

La vera informazione, come ci insegna Pensiero Profondo, è nella domanda e la novità introdotta dai big data è che i dati possono parlare di più e quindi rispondere a più domande,

⁴⁷ Croll A., *Big Data is our generation's civil right issue, and we don't know it*, in Croll A. *Big Data now: Current perspectives from O'Reilly Media*, Sebastopol (CA), O'Reilly Media, pp. 55-59. 2012

affinché i dati possano “raccontare una storia” è necessario creare le condizioni per permettere loro di parlare.

La scienza sociale deve fornire interpretazioni, ma l’interpretazione deve sempre attingere ai fatti, come mostra la piramide della conoscenza rappresentata nella Figura n. 1. Secondo Charles Taylor, la sociologia, la scienza politica, o anche la filosofia e la teoria sociale provano a formulare il miglior resoconto, (*best account*) della nostra situazione, della nostra formazione sociale, dato tutto ciò che sappiamo al riguardo. E per questo miglior resoconto bisogna avvalerci di idee filosofiche, di prove empiriche che sono necessarie in forma quantitativa e qualitativa, di teorie. È molto importante tenere a mente che un “miglior resoconto” non può basarsi su un singolo insieme di dati: l’idea di avere un completo controllo metodologico è un mito e un’illusione. Se c’è un fenomeno che vogliamo comprendere, per esempio l’accelerazione sociale, non potrà mai esserci un singolo insieme di dati da cui possiamo dedurlo⁴⁸.

In altre parole, è possibile pensare a nuove epistemologie che non scartano o rifiutano quelle tradizionali, ma che piuttosto usano un approccio metodologico interdisciplinare tipico della scienza basata sui dati all'interno di un nuovo frame epistemologico che consente agli scienziati sociali di trarre preziose informazioni dalle nuove fonti di dati.

Il valore dei big data è dato dagli *insight* che emergono, cioè dall’interpretazione del fenomeno, che possa indicare delle scelte e/o direzioni operative. I pattern emergenti dai big data sono la base per costruire e validare i modelli, sono un punto di partenza per validare nuove analisi: i processi che portano i dati ad essere rilevanti e quindi a modelli e a decisioni comportano la creazione di altri dati, migliori. I big data non implicano passività intellettuale né possono fare a meno di modelli e concettualizzazioni.

I dati digitali sono stati presentati ed accolti come un cambiamento rivoluzionario del modo di conoscere, che ha segnato il passaggio da una ricerca che persegue la causazione, ad una ricerca che insegue la correlazione⁴⁹, occorre passare dall’apprendimento per correlazioni, al

⁴⁸ Visentin C. *Teoria sociale per un futuro aperto*. Intervista a Hartmut Rosa, pubblicato su Pandora Rivista, 3/2020

⁴⁹ Calude, CS, Longo, G. *The Deluge of Spurious Correlations in Big Data Found Sci* 22, 595-61
<https://doi.org/10.1007/s10699-016-9489-4> - 2017

ragionamento causale, da modelli decisionali *black box* a modelli interpretabili. È improbabile che i big data possano offrire di per sé una forma più alta di intelligenza e conoscenza, la sola accumulazione di dati senza una solida comprensione teorica impedisce di fatto una loro interpretazione (Boyd, Crawford, 2012), e fa sì che questi dati digitali rimangano, come sosteneva già nel 1959 Mills, semplicemente “*un insieme di frammenti confusi di realtà umana*”. Mayer-Schonberger e Cuckier ritengono che pensare che l’analisi dei big data non abbia bisogno di modelli concettuali è del tutto irragionevole l’analisi dei big data si basa su delle teorie, non possiamo eluderle. Esse condizionano i nostri metodi e i nostri risultati. Il condizionamento parte dal modo in cui selezioniamo i dati. Ciò che scegliamo influenza ciò che scopriamo come hanno spiegato Boyd e Crawford in *Six provocation for big data*.

In generale, “co-relazione” denota fenomeni che si relazionano in modo covariante: “Co-relazione” è essenzialmente “coincidenza”, cioè cose che accadono insieme. Una correlazione può essere una relazione tra due variabili: se un valore di variabile aumenta, anche l'altra aumenta (o diminuisce) allo stesso modo. Ad esempio, "all'aumentare della temperatura, aumentano anche le vendite di gelato" o "quando il tempo diventa più freddo, diminuiscono i costi dell'aria condizionata". Ma non bisogna confondere correlazione con causalità. Questa correlazione non va intesa in senso causale. Le correlazioni possono essere utili per il loro potenziale potere predittivo: usare una variabile o agire sul suo valore per prevedere o modificare il valore dell'altra. Ma la correlazione non implica causalità: ci possono essere centinaia di possibili spiegazioni per una correlazione: il numero è limitato solo dall’immaginazione e ingegno nel pensare possibili ragioni per una relazione tra due variabili. Ciò che una correlazione non ti dice è perché due cose tendono ad andare insieme. Ma ricordando quanto ci hanno insegnato Lazarsfeld e la sua scuola sulle correlazioni spurie (1955), interpretare ogni correlazione come significativa può portare a commettere errori molto gravi. I big data ci aiutano a rispondere a nuove domande e sviluppare nuovi indicatori per ottenere correlazioni in tempo reale e in questo modo rompono con il metodo tradizionale per la ricerca della causalità. Lavorare con i big data implica la ricerca di modelli e correlazioni che potrebbero non dirci perché qualcosa sta accadendo, ma piuttosto avvisarci che sta accadendo (Mayer-Schönberger e Cukier 2014).

Sono numerose le correlazioni che si possono osservare nella realtà sociale, ma molte di queste sono attribuibili al caso, e non dimostrano un legame tra fenomeni (Amaturo, Aragona, 2019).

I big data sono solo una parte della storia. Se si vuole veramente capire la causalità, si devono fare degli esperimenti. E se esegui continuamente esperimenti, puoi migliorare continuamente il tuo sistema. Nel 1910, Henry Ford e i suoi colleghi erano in fabbrica ogni giorno, per mettere a punto la catena di montaggio, un metodo di produzione che ha rivoluzionato la produzione. Negli anni '70, il ronzo nella produzione era *kaizen*, il termine giapponese per continuo miglioramento. Ora abbiamo *computer kaizen*, in cui la sperimentazione può essere completamente automatizzata. Proprio come la produzione di massa ha cambiato il modo in cui i prodotti sono stati assemblati e il miglioramento continuo ha cambiato il modo in cui è stata eseguita la produzione, la sperimentazione continua migliorerà il modo in cui ottimizziamo i processi aziendali nelle nostre organizzazioni.⁵⁰

I big data si riferiscono discorsivamente a un cambiamento qualitativo nel significato dei dati, non solo nella quantità di dati (avvicinandosi all'eshaustività) ma anche nella sua qualità (legata per esempio alla tempestività e alla granularità territoriale), i big data trasformano la nostra realtà quotidiana e la nostra relazione immediata con le cose che ci circondano. Questa "dataficazione" della vita quotidiana è al centro dei Big Data: un modo di accedere alla realtà portando in superficie interazioni e relazioni e rendendole visibili, leggibili e quindi governabili, piuttosto che cercare di comprendere le leggi nascoste della causalità. Si ritiene quindi che i dati generino un diverso tipo di conoscenza: più simile alla traduzione o all'interpretazione dei segni piuttosto che a quella della comprensione delle catene di causalità⁵¹. Secondo Cukier e Mayer-Schoenberger la dataficazione definisce le capacità delle piattaforme di trasformare in dati aspetti del mondo mai quantificati prima, non solo informazioni demografiche e di profilazione fornite volontariamente dai clienti o sollecitate

⁵⁰ Varian H. L. *Beyond Big Data*, Business Economics 49, 27–31. doi:10.1057/be.2014.1. 2014

⁵¹ Chandler D. A World without Causation: Big Data and the Coming of Age of Posthumanism. *Millennium*. 2015;43(3):833-851. doi:[10.1177/0305829815576817](https://doi.org/10.1177/0305829815576817)

attraverso sondaggi (online) ma anche metadati comportamentali ricavati in maniera automatica dagli *smartphone*.

Con l'avvento del quarto paradigma il processo di produzione modernista della conoscenza appare come meno mediato da apparati concettuali. Come sottolinea Rob Kitchin, i big data sono unici in quanto la loro costruzione non fa parte di un processo consapevole di produzione della conoscenza, sono intesi come generati dalla vita complessa o dalla realtà stessa nelle tracce di dati lasciate dalle nostre impronte digitali mentre svolgiamo la nostra vita quotidiana. I dati non vengono generati avendo in mente una domanda o uno scopo specifico e sono principalmente un sottoprodotto o un effetto collaterale di attività intraprese in campi con record generati e archiviati tecnologicamente.¹⁵ L'analisi viene dopo che i dati sono stati raccolti e archiviati, non prima di questo. Tuttavia, il fatto che i dati non siano generati consapevolmente, guidate dal desiderio di testare teorie di modelli, è visto come un vantaggio piuttosto che un problema: "L'analisi dei Big Data consente un approccio epistemologico completamente nuovo per dare un senso al mondo; invece di testare una teoria analizzando dati rilevanti, la nuova analisi dei dati cerca di ottenere informazioni "nate dai dati".

È cambiato molto rispetto a prima quando l'approccio tradizionale ai dati, o almeno quello pre-big data, si basava sul presupposto di indipendenza, tra realtà misurata e processo di misurazione: i dati erano ritenuti in grado di abbracciare «oggettivamente» il fenomeno analizzato. Insieme all'obiettività c'era il pragmatismo; i dati non erano necessari solo per la conoscenza, ma per scopi amministrativi e politici. Ogni dato era prezioso e affidabile solo se utile nella pratica: un mezzo per un fine⁵². Come sarà esplicitato meglio quando applicheremo tutto questo *framework* teorico al funzionamento dell'Istituto Nazionale di Statistica, i big data non segnano assolutamente la "fine della teoria" ma trasformano radicalmente il nostro modo di interpretare la realtà. Questo cambiamento richiede un lungo

⁵² Aragona, B., Felaco, C., & Marino, M. *The Politics of Big Data Assemblages*, *Partecipazione e Conflitto*, 11(2), 448-471. Retrieved from <http://siba-ese.unisalento.it/index.php/paco/article/view/19550> (2018)

periodo di adattamento; metterà in discussione processi, pratiche e significati, ma lo straordinario valore che crea lo renderà non solo prezioso ma anche inevitabile.

I numeri non hanno modo di parlare da soli, siamo noi a parlare per loro, li carichiamo di significato e di nostri apparati concettuali e schemi interpretativi. Molti di questi dati sono ingannevoli contengono errori ed incompletezze o sono assolutamente irrilevanti rispetto ai fenomeni che vogliamo conoscere sono in buona sostanza inutilizzabili al fine di costruire un insieme informativo affidabile che ci porti ad una conoscenza dei fenomeni e dunque ad una base obiettiva sulla quale fondare le nostre scelte individuali e collettive. (Arbia 2018)

Questi big data, specialmente in un ambito come quello di alcune scienze sociali, non possono e non devono essere considerati come alternativi rispetto alla teoria; in altri termini la loro abbondanza non dispensa dall'utilizzare la teoria al fine di ricavare "vera" conoscenza da essi. Seneca nelle "Lettere morali a Lucilio", afferma che: "*Ignoranti quem portum petat nullus suus ventus est.*" Nessun vento è favorevole per il marinaio che non sa a quale porto vuole arrivare. (Epistulae morales ad Lucilium, LXXI), I dati sono il vento, mentre la teoria è il timone. Nell'ambito di quelle scienze sociali che studiano sotto profili anche quantitativi le interazioni tra esseri umani, con i big data si assiste a un passaggio abbastanza repentino dalla quasi-bonaccia precedente, cioè una relativa penuria di dati utili, ad un vento fortissimo, se non addirittura una tempesta. L'illusione è che la potenza stessa del vento – l'abbondanza straordinaria che è intrinseca ai big data- renda possibile lasciarsi andare ad esso senza un timone che aiuti ad imprimere la direzione. Quel timone è invece rappresentato dalla teoria, che permette di sfruttare la potenza dei dati per arrivare al porto desiderato, cioè alla conoscenza della struttura che sta dietro i dati stessi: in altri termini è la teoria che aiuta a identificare i rapporti causali tra i diversi fenomeni. In effetti, la teoria aiuta a decidere quali siano –nel mare magnum dei big data- le variabili su cui concentrarsi e quali le variabili da tralasciare.⁵³

⁵³ Pugliesi, R., *Big data e scienze sociali. Una nuova sfida alla teoria*. Dossier Economia Digitale I Copernicani, 10/2018, 52-59. Consultabile al link: <http://www.fondazionecomunica.org/wp-content/uploads/2018/10/CopernicaniDossier-1-Economia-digitale.pdf>

Conclusioni

Questo primo capitolo fornisce quindi una chiave di lettura per affrontare più consapevolmente i capitoli successivi perché esplicita il contesto in cui anche i produttori di statistica ufficiale si muovono, i loro sforzi sia per intercettare il cambiamento e per portarlo all'interno dei propri *business model*.

Ciascuno dei temi trattati è stato scelto pensando alle pratiche tipiche degli Istituti di Statistica. Questa panoramica sul fenomeno sociale rappresentato dalle nuove fonti di dati è stata fatta prediligendo infatti quegli aspetti che potessero in qualche modo aiutarmi ad articolare una lettura critica dei cambiamenti già prodotti dal digitale negli Istituti di Statistica e in Istat in particolare. Questo capitolo rappresenta il filo conduttore dei capitoli a venire che affronteranno i temi qui trattati calandoli nella realtà dei Istituti di statistica.

Capitolo 2 L'evoluzione della statistica

Introduzione

“Arriva infatti sempre un momento in cui il mondo che fino a ieri era familiare appare irriconoscibile: fenomeni che si presentavano vasti e intensi scompaiono, soggetti che erano protagonisti di primo piano impallidiscono e si perdono, mentre ne appaiono e se ne affermano di nuovi, sconosciuti fino ad allora. Cambiano le strutture familiari, cambiano le forme di lavoro, emergono come importanti nuove competenze. La storia della statistica ufficiale è anche quella di riforme degli strumenti di rilevazione, dei modelli di analisi e delle classificazioni. Le chiavi di lettura dei fenomeni, le classificazioni dei modi con cui possono presentarsi, le categorie che ci permettono di descriverli e metterli in relazione con altri, debbono essere sempre aggiornate, in modo condiviso, e la sfida per i sistemi statistici è farlo attraverso processi decisionali collettivi agili, quanto più possibile rapidi”⁵⁴

Ho scelto le parole con cui l'ex Presidente dell'Istat prof. Giorgio Alleva ha aperto la XIII Conferenza di Statistica nel luglio del 2018, perché sintetizzano l'approccio migliore per definire la statistica ufficiale e per delinearne gli impatti e le evoluzioni. In questo capitolo saranno descritte le risposte che la statistica ufficiale e gli Istituti di statistica hanno dato al cambiamento del contesto esterno e come abbiamo modificato i paradigmi della ricerca per valorizzare le trasformazioni in atto. Mariana Kotzeva, Direttrice Generale di Eurostat all'apertura della XIII Conferenza Nazionale di Statistica nel luglio del 2018, per sottolineare la capacità degli istituti di statistica europei di cogliere le sfide della trasformazione digitale ha sostenuto: *“Le statistiche ufficiali misurano la vita e, quando la vita cambia, cambiano anche le statistiche ufficiali. Statistics follows the life”*.

Man mano che le società e le economie diventano più globalizzate e complesse, le statistiche ufficiali, al fine di misurare adeguatamente le transazioni sociali ed economiche, diventano inevitabilmente più complesse. La statistica ufficiale deve saper cogliere i cambiamenti in

⁵⁴ Alleva G., *Relazione di apertura, Dall'incertezza alla decisione consapevole: un percorso da fare insieme* Atti della 13a Conferenza nazionale di statistica Roma, 4-6 luglio 2018

atto nella società offrendo nuovi parametri e metriche per comprendere i fenomeni più attuali, la globalizzazione, la competitività, le migrazioni, lo sviluppo sostenibile.

La pervasività delle tecnologie comporta anche una ridefinizione delle modalità di accesso e uso delle risorse dell'informazione, della conoscenza, della comunicazione nella sfera personale e in quella pubblica. Queste trasformazioni interessano una serie sempre più estesa di attività della vita quotidiana, in cui le tecnologie online si innestano senza soluzione di continuità nella nostra esperienza del mondo, mentre l'emergenza sanitaria ha impresso loro una importante accelerazione. Con il *lockdown* abbiamo preso atto che le tecnologie digitali rappresentano *ubiquitous interfaces* che abilitano e strutturano l'esperienza di una pluralità di contesti⁵⁵. Alla luce di questo, anche i sistemi statistici devono rivedere i propri *business model* e ridisegnare relazioni con utenti e *stakeholders* per mantenere e accrescere i livelli di fiducia che i cittadini nutrono verso le istituzioni che rilasciano statistiche ufficiali e permettono una conoscenza profonda della società e delle sue caratteristiche.

2.1. La statistica ufficiale e gli Istituti di statistica

La statistica ufficiale rappresenta un pilastro indispensabile della democrazia, svolge un ruolo chiave per l'avanzamento della conoscenza, per facilitare i processi di *decision making* e risponde ai bisogni di cittadini, decisori, ricercatori e media. È dunque un bene pubblico capace di dare risposte a domande cruciali, evolve col mutare dei fenomeni.

Da quando esistono, le statistiche ufficiali si sono adattate agli eventi economici, politici, sociali e ambientali per rimanere rilevanti. Questo processo di evoluzione continua ancora oggi a un ritmo accelerato.⁵⁶

La missione delle statistiche ufficiali è fornire una rappresentazione quantitativa della società, dell'economia e dell'ambiente per scopi di interesse pubblico, per la progettazione e la valutazione delle politiche e come base per informare il dibattito pubblico. In altre parole,

⁵⁵ Marinelli A. 122 Scuolademocratica n.s./2021

⁵⁶ MacFeely S., The Continuing Evolution of Official Statistics: Some Challenges and Opportunities, *Journal of Official Statistics*, Vol. 32, No. 4, 2016, pp. 789–810, <http://dx.doi.org/10.1515/JOS-2016-0041>

le statistiche ufficiali forniscono alla società la conoscenza di se stessa⁵⁷: la missione di alto livello delle statistiche ufficiali nelle società democratiche quindi consiste nel fornire una visione quantitativa della società alla società stessa, includendo in tal modo i responsabili delle decisioni e il pubblico in generale. La produzione delle moderne statistiche ufficiali si basa su un sistema di metodi scientifici, regolamenti, codici, pratiche, principi etici e impostazioni istituzionali che è stato sviluppato negli ultimi due secoli a livello nazionale parallelamente allo sviluppo degli stati moderni. Al centro di ogni sistema statistico c'è spesso un unico Istituto di statistica o un sistema di Istituti che producono e rilasciano statistiche ufficiali a supporto dei cittadini e dei *decision makers*⁵⁸.

Le statistiche ufficiali rivestono da sempre un ruolo fondamentale nella società e il loro apporto alla conoscenza del complesso contesto del mondo attuale risulta sempre più centrale. Le organizzazioni statistiche mirano a fornire informazioni su tutti gli aspetti importanti della società agendo secondo standard scientifici sempre più elevati e condivisi a livello internazionale. Come indicato nell'art. 338 del Trattato sul funzionamento dell'Unione europea⁵⁹ (Carta dei diritti fondamentali) e come riportato nei vari regolamenti e codici sulla qualità della statistica la statistica ufficiale è caratterizzata da specifici fattori di qualità quali imparzialità, affidabilità, obiettività, indipendenza scientifica, efficienza economica e riservatezza statistica e non comporta oneri eccessivi per gli operatori economici. La statistica ha avuto inizio come attività pratica volta alla soluzione di problemi pratici della vita. Le sue origini si perdono nella notte dei tempi e non possono essere precisate con esattezza. Nel passaggio dalla piccola comunità umana alla società sempre più organizzata, è aumentato il bisogno di conoscere quei fatti che costituivano le manifestazioni essenziali del gruppo sociale e perciò le statistiche diventavano sempre più frequenti e meno imperfette⁶⁰.

⁵⁷ Letouzé, E and Jütting, J *Big data and human development: Towards a new conceptual and operational approach*. DATA-POP Alliance White Paper. Disponibile al seguente link https://paris21.org/sites/default/files/WPS_OfficialStatistics_June2015.pdf 2015

⁵⁸ Zuliani A, L'evoluzione dell'informazione statistica, le indagini di opinione e le sfide della statistica ufficiale, Istat 2015 disponibile su <https://www.istat.it/it/files/2011/05/Zuliani.pdf>

⁵⁹https://www.senato.it/application/xmanager/projects/leg18/file/repository/relazioni/libreria/novita/XVII/Trattato_sull_unione_europea.pdf

⁶⁰Leti G., Cerbara L. *Elementi di statistica descrittiva*, Il Mulino Manuali, 2009

Già nel 1826 Melchiorre Gioia nel suo *“Filosofia della statistica”* definiva la statistica come *“quella somma di cognizioni relative ad un paese, che nel corso giornaliero degli affari possono essere utili a ciascuno e alla maggior parte dei suoi membri od al governo, che ne è l’agente, il procuratore o il rappresentante”*. Tale definizione, per quell’epoca, è da considerarsi *“rivoluzionaria”*, in quanto introduceva una concezione *“democratica”* della statistica ufficiale. Gioia, infatti, sosteneva che il lavoro degli Uffici produttori delle statistiche ufficiali, inteso come funzione pubblica, deve essere utile e al servizio della Società assegnando al cittadino un ruolo prioritario rispetto al governo che ne è il rappresentante (Biggeri L.,1989) La Storia ci insegna come in realtà, nei quasi duecento anni che ci separano dallo scritto di Gioia, questa concezione della statistica ufficiale si sia scontrata con la esistenza di Stati sovrani, con un forte potere centrale, nei quali la funzione statistica era posta al servizio del *“principe”*, che normalmente preferiva tenere celate le motivazioni della propria condotta e le stesse statistiche commissionate, tanto che le rilevazioni servivano soprattutto a fornire strumenti più di controllo che di intervento sociale. L’auspicio di Gioia è però divenuto una esigenza reale nelle società democratiche e con ordinamento *“pluriclasse”* dove la statistica ufficiale è un bene pubblico. Nel nostro ordinamento costituzionale l’informazione statistica tout court è vista come uno strumento essenziale per la realizzazione dei fondamentali scopi sociali e politici indicati nella Costituzione e, in definitiva, per lo sviluppo di una società democratica (Sandulli A. M., Baldassarre A., 1971). È il decreto legislativo del 6 settembre 1989 n. 322⁶¹ di istituzione del Sistema Statistico Nazionale (Sistan) che definisce i dati elaborati nell’ambito del Programma Statistico Nazionale (Psn) come patrimonio della collettività, e ad essi, contraddistinti da un elevato grado di imparzialità ed attendibilità, si attribuiscono particolari effetti giuridici.

Tutti ora riconosciamo che l’informazione statistica rappresenta un supporto indispensabile per decisioni razionali non solo da parte dei governi e delle amministrazioni pubbliche, ma anche degli operatori economici, delle parti sociali, dei media e dei cittadini, e sappiamo che

⁶¹ Norme sul sistema statistico nazionale e sulla riorganizzazione dell’Istituto Nazionale di statistica, ai sensi dell’art. 24 della legge 23 agosto 1998, n. 400

costituisce per questi ultimi uno strumento di partecipato controllo democratico. E proprio per questo condividiamo l'idea che ogni società democratica necessita di informazioni statistiche ufficiali imparziali, affidabili, trasparenti, accessibili e pertinenti, capaci di descrivere le sue condizioni e i cambiamenti che avvengono in essa e rispettose della tutela della *privacy*.⁶² Le statistiche ufficiali sono prodotte secondo principi, regole e standard stabiliti che mirano collettivamente al raggiungimento di indipendenza, professionalità, trasparenza e qualità lungo l'intero processo. Il Codice italiano delle statistiche ufficiali, emanato dal Comstat (l Comitato di indirizzo e coordinamento dell'informazione statistica organo di governo del Sistema statistico nazionale) con la direttiva n.10 del 17 marzo 2010, è stato concepito in perfetta coerenza con quello europeo, *il Code of practice on european statistics*, è l'esito di un processo condiviso di analisi e di consultazione che ha coinvolto i principali attori, centrali e locali, del Sistan. Il Codice Italiano ha come obiettivo fondamentale la promozione di una cultura comune della qualità, a sostegno della quale sono indicati principi e parametri di valutazione degli assetti istituzionali, dei processi e dei prodotti della statistica. È definita ufficiale la produzione statistica contenuta nel Programma statistico nazionale, nei programmi statistici regionali e, in generale, quella prodotta dagli Enti ed Uffici del Sistema statistico nazionale intendendo, rispettivamente, gli enti ed organismi pubblici di informazione statistica (di cui all'art. 2, comma 1, lettera h, del decreto legislativo 6 settembre 1989, n. 322) e gli Uffici di statistica costituiti ai sensi del medesimo decreto legislativo n. 322 del 1989 e della legge 28 aprile 1998, n. 125.

In Italia l'ente produttore di statistica ufficiale è l'Istat, istituito nel 1926 con la legge n. 1162, che attribuisce all'Istituto Centrale di Statistica le funzioni fino ad allora svolte dalla Divisione di statistica generale del Ministero dell'Agricoltura. Viene così rafforzata l'idea della statistica come strumento indispensabile per conoscere la realtà sociale ed economica del Paese e matura la convinzione che solo l'accentramento graduale di tutti i servizi statistici può soddisfare queste esigenze conoscitive. La storia dell'Istat è la storia di un'Italia in

⁶² Biggeri L., *I principi e caratteristiche della statistica ufficiale tra passato e futuro*, Atti della VII Conferenza nazionale di statistica, Istat, 2004

evoluzione, il racconto della crescita del Paese, l'Istat con le sue statistiche accompagna l'evoluzione del paese, fornendo dati su fenomeni sociali via via emergenti. Negli anni Novanta, per esempio in una società in forte cambiamento, lo studio quantitativo dei fenomeni sociali ed economici rappresenta un approfondimento necessario e uno stimolo per orientare le decisioni pubbliche e private, nel 1993, viene realizzato il primo "Rapporto annuale sulla situazione del Paese" che negli anni diventa una delle pubblicazioni di punta dell'Istat. Nel 1996 l'Istituto si apre al web, nasce così il sito istituzionale. Dal 1998 l'Istat inizia a diffondere, secondo scadenze e criteri stabiliti dal Trattato di Maastricht, i dati relativi agli indicatori di convergenza, necessari per valutare l'andamento dell'economia del Paese. Gli anni Duemila del ventunesimo secolo portano l'impronta del progresso tecnologico. Il sito web diventa lo strumento principe per la comunicazione e diffusione dei dati. Nel 2010 l'Istat rilascia I.Stat, il *datawarehouse* generale delle statistiche prodotte dall'Istituto. I.stat offre un patrimonio informativo statistico completo e omogeneo. L'accesso al datawarehouse è libero e gratuito. Sempre nel 2010 sono aperti i canali Istat sui social network più conosciuti (Twitter, YouTube, SlideShare), per garantire una diffusione tempestiva e virale dei contenuti Istat in rete. Anche il versante della produzione è sempre più caratterizzato da forti innovazioni tecnologiche e metodologiche. Quasi tutte le indagini realizzate presso le imprese offrono la possibilità di compilare i questionari direttamente online, con un significativo risparmio di tempo e risorse. Analogamente per le indagini presso le famiglie, tradizionalmente legate a tecniche di intervista diretta con questionario cartaceo, si sta andando verso tecniche di acquisizione digitale con interviste assistite da computer.

2.2 Come cambia il modo di produrre informazione in Istat

Nel XX secolo il paradigma dominante che guida la produzione di statistica ufficiale era *survey data based statistics*, i dati provenivano fundamentalmente da indagini, vale a dire una raccolta di dati seguendo un processo produttivo complesso che ha come obiettivo la produzione di informazione statistica cioè la produzione di descrizioni riassuntive di carattere quantitativo riguardanti un fenomeno collettivo.

Lo strumento per raccogliere le informazioni è rappresentato dal questionario nel caso delle indagini dirette (i dati si raccolgono direttamente presso la popolazione oggetto di studio al fine di conoscere un dato fenomeno) o da fonti amministrative nel caso di indagini indirette.

Con riferimento alla popolazione da esaminare le indagini possono essere totali o censuarie nel caso in cui venga osservata l'intera popolazione oggetto di studio o campionaria quando si osserva una parte, vale a dire un campione della popolazione oggetto di studio.

Per la produzione statistica, la diffusione delle tecniche campionarie ha di fatto determinato un salto di qualità. Si introduce quindi un'innovazione metodologica data da un insieme interdipendente di regole che vengono denominate disegno di campionamento: un procedimento attraverso il quale, da un insieme di unità costituenti l'oggetto dello studio, si estrae un numero ridotto di casi scelti con criteri tali da consentire la generalizzazione all'intera popolazione dei risultati ottenuti. La rilevazione campionaria consente di arrivare ad una stima del fenomeno, e viene scelta quando è impossibile effettuare una rilevazione totale; quando si vogliono ridurre costi e/o tempi di un'indagine. Con le indagini campionarie una parte delle unità della popolazione di riferimento viene selezionata per far parte del campione, le unità selezionate si sottopongono ad osservazione per ottenere informazioni statistiche su certe caratteristiche (variabili) della popolazione e per analizzare le relazioni, semplici e complesse, che aiutino ad interpretare il fenomeno oggetto di studio.

Le indagini, sia campionarie sia censuarie, comportano una serie di problematiche legate alla bassa copertura delle liste di campionamento, ai tassi di risposta (totale e parziale) sempre

più bassi, all'errore di misura, alla necessità di un *trade-off* fra tempestività e granularità e non da ultimo il *burden* di risposta e i costi elevati. All'interno del Sistema Statistico ci si è interrogati su come superare queste difficoltà e gli Istituti si sono mossi sempre di più in un'ottica di integrazione di più fonti, tra cui in prima battuta quelle amministrative da integrare con i risultati delle indagini tradizionali. L'integrazione costituisce ormai sempre più un orizzonte di riferimento per gli Istituti di Statistica, soprattutto alla luce dei vantaggi che può portare in termini di riduzione del *burden* statistico per i rispondenti e per l'abbattimento dei costi anche se la *popolazione amministrativa* può non coincidere quella statistica e sicuramente emergono delle variabili che non possono essere rilevate con fonte amministrativa.

I tradizionali modelli di rilevazione basati sull'acquisizione diretta dei dati dalle fonti (cittadini e imprese) attraverso le rilevazioni e i censimenti sono messi in discussione per l'impatto, a volte troppo invasivo, sui rispondenti, che influisce sull'abbassamento dei tassi di risposta. Emerge, pertanto, la spinta a realizzare e a utilizzare prevalentemente registri statistici, derivati da fonti amministrative con meccanismi di alimentazione a carattere continuo basati esclusivamente su flussi telematici. Il primato dei dati delle indagini rispetto alle fonti di dati esistenti era insostenibile per molte ragioni. Per gli Istituti di statistica, è appropriato e legittimo raccogliere dati se e solo se questi dati non possono già essere ottenuti da fonti esistenti di qualità soddisfacente. Questo pone le basi per un nuovo modello di *business* completamente diverso che presuppone cambiamenti fondamentali nei compiti di un istituto statistico: nuovi strumenti metodologici, (ad esempio, il *record linkage*), adattamenti della *governance* statistica (ad esempio l'adeguamento del quadro normativo per l'accesso ad alcuni dati amministrativi sensibili) e cambiamenti nella gestione della qualità e nella comunicazione agli utenti. Si potrebbe quasi arrivare a dire che l'integrazione tra indagine e dati amministrativi non sia altro che la sostituzione di una voce del questionario con un dato simile proveniente da un registro. È stato dimostrato che questa impressione è troppo semplicistica e non realistica. È l'intero processo statistico che deve essere rivisto perché è molto lontano dal disegno classico, in cui un'indagine censuaria tradizionale

("bussare alla porta") ogni dieci anni a un disegno moderno (completamente integrato), in cui una regolare fonte amministrativa si integra con un'indagine campionaria assicurando dati più tempestivi.⁶³ L'Istat per esempio ha sostituito il Censimento decennale della Popolazione con il Censimento Permanente della Popolazione e delle Abitazioni allo scopo di rilasciare informazioni continue e tempestive sulle principali caratteristiche socio-economiche della popolazione dimorante abitualmente in Italia. Il Censimento Permanente della Popolazione prevede l'integrazione dei dati provenienti dalle fonti amministrative con quelli delle rilevazioni campionarie annuali, da misurazioni costose realizzate ogni dieci anni a un monitoraggio continuo del livello e delle caratteristiche della popolazione, fortemente basato sul riutilizzo a fini statistici dei dati già in possesso della pubblica amministrazione, ma anche sull'integrazione con le fonti statistiche dirette.

Il nuovo orizzonte della statistica europea vede un rafforzamento della logica d'integrazione delle informazioni prodotte e un forte uso dei registri statistici per fornire nuovi e più approfonditi strumenti anche a sostegno delle politiche d'intervento: un approccio olistico finalizzato a guadagnare in qualità ed efficienza.

Il Sistema statistico europeo (d'ora in poi SSE nel testo) ha spinto infatti verso l'integrazione delle fonti non solo quelle amministrative, nella *Vision2020* che ha guidato i processi di modernizzazione di gran parte degli Istituti europei, gli Istituti di statistica dichiarano infatti *"...basiamo i nostri prodotti e servizi statistici sia su indagini tradizionali che su fonti più recenti, inclusi dati amministrativi, geospaziali e, ove possibile, big data. Nuove fonti di dati integrano quelle esistenti e ci aiutano a migliorare la qualità dei nostri prodotti. Lavoreremo insieme per ottenere l'accesso a nuove fonti di dati, creare metodi e trovare la tecnologia adatta per utilizzare nuove fonti di dati per produrre statistiche europee in modo affidabile"*.

In vista di queste considerazioni si rafforza e sviluppa maggiormente un approccio *Register based statistics* basato sulle fonti amministrative seguendo l'esempio dei paesi del Nord Europa pionieri di questo tipo di approccio. La Danimarca ha effettuato nel 1981 il primo

⁶³ Radermacher W. J., *Official Statistics 4.0 Facts for People in the 21. Century*, Tesi di Dottorato, 2020

censimento integrando i dati da indagine con gli archivi amministrativi. L'Istat con l'approvazione del Programma di modernizzazione (cfr. Capitolo 3 del presente lavoro) che ha comportato una reingegnerizzazione dei processi produttivi, si spinge verso il progressivo uso di dati da fonte amministrativa e verso la costruzione del SIR, il sistema integrato di registri, un sistema basato sull'uso di dati di fonte amministrativa e statistica organizzate in un sistema di archivi, che permette una descrizione più dettagliata dei fenomeni in una prospettiva multi-dominio. Per i sistemi statistici ufficiali, a mano a mano che gli strumenti di raccolta, analisi, diffusione si affinano, emergono lacune conoscitive. Una produzione basata sempre più sull'integrazione di più fonti e non sulle tradizionali indagini campionarie rappresenta anche una sfida metodologica, in relazione a nuovi approcci inferenziali e di comunicazione dell'incertezza nelle stime elaborate dagli INS⁶⁴. Per questo motivo il Sistema integrato dei registri deve essere garantito da *framework* metodologici solidi per permettere l'integrazione delle fonti e migliorare e aumentare il potere informativo a costi non crescenti, le indagini statistiche diventano ausiliarie alle fonti amministrative e mantengono nello stesso tempo la loro centralità perché assicurano la completezza, la pertinenza e l'accuratezza dell'informazione prodotta.

La strategia *European Statistical System Vision 2020* è di fatto elaborata in risposta a quattro sfide che il documento con cui Eurostat descrive la strategia Vision 2020 sintetizza così:

- ✓ la rivoluzione dei dati (che è stata ampiamente descritta nel capitolo precedente);
- ✓ nuovi fenomeni: la globalizzazione ha cambiato il mondo. Ne è emersa una nuova e complessa realtà deve essere catturata dalle statistiche ufficiali nelle sue varie sfaccettature. Sono emersi nuovi fenomeni importanti nella nostra società: crisi economiche, finanziarie e politiche, nuove forme di organizzazione della produzione economica, nuove tendenze demografiche globali, le questioni dello sviluppo sostenibile. Questi fenomeni devono essere letti attraverso nuovi prodotti statistici in grado di servire molteplici scopi e diverse tipologie di *stakeholders*;

⁶⁴Ricciato, F. et al. *'Trusted Smart Statistics: Motivations and Principles* 1 Jan. 2019 : 589 – 603.

- ✓ il costo delle statistiche: le informazioni statistiche di alta qualità non sono economiche. C'è necessità di investire nella produzione di statistiche di alta qualità o continuarle con budget ridotti;
- ✓ il futuro dell'Europa: l'evoluzione delle statistiche ufficiali come bene pubblico riflette l'evoluzione della società. In Europa la fase evolutiva delle statistiche comporterà l'integrazione a diversi livelli dei dati amministrativi e di nuove fonti.

In sostanza possiamo senz'altro affermare che la crescente integrazione dei dati amministrativi con i dati provenienti da indagini per produrre informazione statistica sui fenomeni demografici, sanitari ed economici rappresenta la risposta che il Sistema Statistico Europeo dà alle sfide e ai mutamenti avvenuti all'esterno degli Istituti stessi. Le fonti amministrative presentano numerosi vantaggi: consentono di produrre dati statistici con costi aggiuntivi limitati; sono aggiornate; sono validate dal procedimento amministrativo al quale si riferiscono e che sostengono; contengono informazioni individuali, riferite a singole persone, famiglie e imprese, che possono essere riaggregate secondo le esigenze e collegate con altre; non aggravano il carico statistico sui rispondenti. Tuttavia, sussistono delle criticità quali il disallineamento temporale, l'uso di classificazioni e concetti non coerenti tra loro, o la presenza di differenti tipologie di errori, che rendono tale processo nient'affatto scontato, nè automatico. Da qui i consistenti investimenti in solide infrastrutture metodologiche che i singoli Istituti in termini, per esempio, di informazione ausiliaria sulla popolazione per garantire la coerenza ed aumentare l'efficienza delle stime, disegni *multiple frames* per ridurre l'errore di copertura, per aggiustare distorsione da mancata risposta, modelli di imputazione più sofisticati, *sample coordination* per ridurre il disturbo statistico, elaborazione di modelli per piccole aree per aumentare la granularità.

Indagini (compresi i censimenti) e registri amministrativi sono stati quindi per decenni gli unici tipi di dati di *input* disponibili per le statistiche ufficiali. Pertanto, qualsiasi aspetto della produzione statistica, le metodologie di trattamento, modelli di *governance* dei dati, regolamenti, pratiche organizzative e così via sono stati adattati a questo tipo di dati. Negli ultimi due decenni, a seguito della digitalizzazione, smartificazione e dataficazione delle

nostre società (Cukier e Mayer-Schoenberger, 2013), una ricchezza di nuovi tipi di fonti di dati digitali diventano disponibili e quindi di grande interesse per la comunità statistica.

2.3. Big Data e statistica ufficiale, un percorso che viene da lontano.

Le evoluzioni degli ultimi hanno fatto emergere fenomeni non misurabili attraverso gli strumenti della statistica tradizionale (per esempio gli effetti che la globalizzazione e la digitalizzazione hanno sul sistema delle imprese, sulla loro competitività, sulla loro capacità di creare reddito e lavoro, le nuove forme di lavoro, di professioni e competenze; popolazioni particolarmente difficili da identificare e studiare statisticamente caratterizzate da un elevato grado di vulnerabilità). Davanti alla complessità da cui siamo circondati, gli Istituti di statistica non possono fare altro che puntare su ricerca e innovazione e migliorare la loro capacità di produrre e rilasciare informazione statistica. Il motore dei processi di miglioramento continuo in cui l'Istat è coinvolto, è dato dalla crescita della domanda che chiede sostegno in diversi ambiti, nuovi e specifici, a supporto di strategie anche di carattere globale (per esempio gli indicatori per il monitoraggio degli Obiettivi dello sviluppo sostenibile dell'Agenda 2030). La risposta degli INS è il potenziamento dell'offerta statistica, il miglioramento della sua rilevanza e tempestività, attraverso innovazione, ricerca e sperimentazione.

Una sfida quindi che investe processi ma che riguarda anche i tematismi, i prodotti e quindi i contenuti, la statistica pubblica si deve ampliare, cogliere le novità, i nuovi fenomeni e le nuove esigenze informative. I dati che più di tutti riescono a supportare la risposta a queste esigenze e a dare un significativo contributo per affrontare queste sfide e per indagare nuove dimensioni sociali ed economiche sono quelli provenienti dalle nuove fonti, come è già stato ampiamente argomentato nel capitolo precedente. La statistica ufficiale ha già da alcuni anni puntato al rinnovamento delle proprie fonti dati, considerando in particolare l'opportunità di utilizzare anche i big data come possibili fonti in combinazione, alternativa o ausilio ai dati tradizionali. Gli istituti di statistica quindi hanno iniziato a sperimentare un nuovo

paradigma di ricerca e produzione e ad integrare nei loro processi di produzione statistica anche i big data per offrire risposte più puntuali, con un livello di precisione spaziale e temporale maggiore di quanto offerto dalle tradizionali rilevazioni, e favorire la “scoperta” di nuovi elementi e contribuire all’avanzamento della conoscenza, grazie alla capacità di mettere in relazione dati provenienti da fonti diverse.

Per i sistemi statistici è un passaggio quasi obbligato, dovuto anche al fatto che il ruolo dei dati è cambiato: dalla computabilità e precisione degli *small data* si passa all’incomputabilità e alla ridondanza, alla diversità dei big data. Con il modello basato sulle indagini era del tutto razionale privilegiare l'esattezza nell'età degli *small data* perché potendo raccogliere solo un quantitativo limitato di informazioni la loro qualità doveva essere il più alta possibile, nel mondo degli *small data* sia le indagini sul rapporto di causalità sia le analisi di correlazione partivano da un'ipotesi che veniva poi messa alla prova per essere confermata oppure confutata. Ora dal dato come fotografia di un sistema passato, al dato come base su cui costruire modelli predittivi. I dati prima erano usati come una fotografia immobile del presente, adesso invece siamo di fronte ad un flusso continuo di dati che nutre algoritmi, rafforza la capacità predittiva e crea delle fotografie, delle mappe del futuro⁶⁵ (Vespignani, 2019). Dal dato che descrive una relazione precisa di causa-effetto al dato che evidenzia una correlazione in base a *pattern* emergenti. L'enorme quantitativo di dati aumenta la difficoltà, ma soprattutto l'importanza, di filtrare e di pulire i dati, per trasformare questa grande mole in informazione e in sapere e quindi in conoscenza, attraverso l'applicazione di modelli. L'estrazione del valore dai dati deve necessariamente interagire con un insieme di componenti quali il contesto, le relazioni, l'interdisciplinarietà e quindi con la complessità.

A generazioni di studenti sono stati inflitti corsi di statistica aridi, basati sull'apprendimento di un insieme di tecniche da applicare a varie situazioni privilegiando la teoria matematica rispetto alla comprensione del motivo per cui si usano certe formule e delle difficoltà che insorgono quando si cerca di usare i dati per rispondere ad alcune domande. Le cose stanno cambiando, come ho avuto modo di scrivere anche nel capitolo precedente, la scienza dei

dati e l'alfabetizzazione ai dati richiedono una strategia basata sui problemi, in cui l'uso di problemi statistici specifici è soltanto un elemento di un intero ciclo di indagine⁶⁶.

Così come per l'integrazione delle fonti amministrative nei processi di produzione statistica, trattare dati di grandi dimensioni e soprattutto di estrema varietà di formato e di velocità di generazione obbliga ad una attenta valutazione del quadro metodologico, organizzativo e tecnologico all'interno del quale si muove tradizionalmente la statistica ufficiale. In particolare, la domanda che si sono poste le organizzazioni statistiche nazionali/ufficiali è stata in che modo le fonti di big data possono essere efficacemente sfruttate "statisticamente", con l'obiettivo di produrre ulteriori informazioni, aumentare la qualità delle fonti già disponibili, favorire la riduzione del carico sui rispondenti e dei costi correlati alla raccolta dei dati, accogliendo in questo specifiche raccomandazioni di organizzazioni internazionali (quali Eurostat e Nazioni Unite).

L'esigenza di disporre di dati sempre più abbondanti e di qualità superiore, se da una parte si sta quindi rivelando determinante per soddisfare esigenze di conoscenza dei fenomeni sociali ed economici di una società sempre più complessa e per rendere sostenibile il costo delle indagini – soggette a forti limitazioni dalle pressioni di bilancio e dai costi associati agli oneri di risposta – dall'altra comporta la non facile risoluzione di questioni oltre che sul fronte tecnico, anche su quello metodologico ed organizzativo. Sia per dati statistici raccolti mediante indagini tradizionali con diretto contatto con le unità selezionate sia per i dati amministrativi, le metodologie e i processi sono da tempo consolidati e standardizzati a livello internazionale. Per i dati originati dall'uso di dispositivi digitali si rende necessaria la definizione e risoluzione diverse problematiche legate alla loro classificazione, e al loro trattamento lungo il ciclo di vita dei dati statistici (raccolta, validazione, analisi, trattamento e uso). Volume, varietà e velocità dei dati, che come abbiamo ampiamente visto hanno rappresentato le prime caratteristiche specifiche e comuni dei big data, mantengono anche in ambito statistico peculiarità tali da condizionarne l'uso, cambiando il concetto stesso di

⁶⁶ Spiegelhalter D., *L'arte della statistica. Cosa ci insegnano i dati*. Einaudi I Maverick, 2020

archivio e base informativa, non più rappresentati solamente da data base rigidamente strutturati ma aperti a nuovi formati, tipicamente non relazionali e possibilmente da trattare con nuovi approcci (*Map&Reduce*, *text-mining*, *extreme-scale visual analytics*, ecc.). In questo contesto, anche il tema della qualità dei dati assume una specifica caratterizzazione. Per i big data è sensato parlare di qualità “specifica per le fonti”. In particolare, una fonte Big come i sensori avrà una qualità specifica che dipende dal fatto che i dati da sensori sono spesso mancanti, soggetti a rumore o ad effetti di calibrazione degli strumenti di misura. I dati da *social media* sono invece scarsamente strutturati e spesso non accompagnati da metadati che ne consentano una corretta interpretazione. In aggiunta alla specificità di fonte, c’è anche una caratterizzazione della qualità delle fonti big per dominio. Se il dominio di interesse è quello della statistica ufficiale, alcune dimensioni di qualità particolarmente rilevanti sono la “rappresentatività” di una fonte, fondamentale per poter produrre stime affidabili, l’accuratezza in termini di qualità intrinseca dei valori acquisiti e l’affidabilità della fonte⁶⁷. In questo nuovo paradigma, diventa fondamentale garantire la qualità delle statistiche per preservare la credibilità e la reputazione degli istituti che rilasciano statistica ufficiale che, se vogliono produrre dati affidabili da nuove fonti, devono introdurre metodi, concetti e strumenti per garantire la qualità e affrontare ambienti multi-sorgente.

I sistemi statistici si sono formati in un mondo predigitale in cui i dati erano una merce scarsa, molto costosa da acquisire su larga scala, quindi, gran parte delle risorse nel passato è stata dedicata alla raccolta di dati. Le indagini (compresi i censimenti) e i registri amministrativi sono stati per decenni gli unici tipi di dati di *input* disponibili per le statistiche ufficiali. Pertanto, qualsiasi aspetto della produzione statistica, le metodologie di elaborazione, i modelli di *governance* dei dati, le normative, le pratiche organizzative e così via sono stati adattati a questo tipo di dati. Nella comunità statistica ufficiale, il termine big data è spesso usato per fare riferimento collettivamente a origini dati non tradizionali, come termine *wrapper* per tutti i tipi di dati diversi da sondaggi e dati amministrativi, ma è opportuno usare

⁶⁷ Barcaroli G., De Francisci S., Righi P., Scannapieco M., Big data: si apre un’enorme possibilità per la statistica ufficiale, pubblicato il 28 gennaio 2016. <https://www.forumpa.it/pa-digitale/big-data-si-apre-unenorme-opportunita-per-la-statistica-ufficiale/>

invece il termine “nuovi dati digitali”⁶⁸, perché di fatto rispecchiano nuovi comportamenti resi possibili dai processi messi in moto dalla trasformazione digitale. Per integrare questi cambiamenti all’interno dei propri processi gli uffici di statistica sono chiamati oggi a ripensare il modo in cui operano per riaffermare il loro ruolo nella moderna società democratica. La questione in gioco è notevolmente più ampia e più profonda del semplice adattamento dei processi esistenti per abbracciare i cosiddetti Big Data. Questo nuovo approccio assegna un ruolo cruciale all’integrazione di fonti tradizionali e nuove, con particolare attenzione alle tecnologie digitali; richiede anche un cambio di paradigma rispetto al modello *legacy* di produzione di statistiche ufficiali nella metodologia, dalle tradizionali stime basate su indagini campionarie a un sistema applicabile a un ambiente multi-sorgente. Tale complessa evoluzione deve essere guidata da una visione globale a livello di sistema e il SSE ha accolto questa sfida. La consapevolezza che le nuove sorgenti di dati rappresentino difatti un’enorme opportunità per la statistica ufficiale, non solo in termini di abbattimento di costi ma anche e soprattutto di arricchimento della produzione statistica e di aumento della tempestività dei prodotti statistici, si è subito consolidata all’interno del Sistema Statistico europeo. Infatti nel 2013 Eurostat adotta il *Memorandum di Scheveningen*⁶⁹ (Allegato 1) che rappresenta una fase fondamentale nel percorso di adozione di nuovi paradigmi di ricerca e produzione: da un modello «tradizionale» basato sull’acquisizione diretta dei dati ad un modello Integrato, basato sull’uso integrato di dati amministrativi, indagini e Big data. Il *Memorandum* formalizza la necessità per tutti gli istituti di statistica europei di considerare le fonti di big data come nuove fonti per le statistiche ufficiali, da integrare con dati amministrativi e con i risultati delle indagini tradizionali. Parte così una densissima fase di sperimentazione con i progetti ESSnet Big Data I e II che apre una prospettiva strategica di uso di grandi masse di dati per finalità statistiche, sia in forme

⁶⁸ Ricciato, F., Wirthmann, A., & Hahn, M., *Trusted Smart Statistics: How new data will change official statistics*. Data & Policy, 2, E7. doi:10.1017/dap.2020.7, 2020

⁶⁹ European Statistical System - ESS, Directors General of the National Statistical Institutes - DGINS. 2013. Scheveningen Memorandum. Big Data and Official Statistics. Scheveningen – The Hague, The Netherlands, 25th – 27th September 2013: DGINS Conference.
https://ec.europa.eu/eurostat/cros/system/files/SHEVENINGEN_MEMORANDUM%20Final%20version.pdf

integrate con altre fonti, sia come nuova fonte, in sostituzione di quelle usate tradizionalmente. Gli ESSnet sono stati lanciati in occasione di un meeting DGINS a Palermo nel settembre 2002, durante il quale i Direttori degli Istituti hanno espresso la necessità di trovare sinergie, armonizzazione e diffusione delle migliori pratiche nel Sistema Statistico Europeo. La proposta venuta fuori dal meeting è stata quella di creare uno strumento dedicato: i progetti dei Centri e delle Reti di Eccellenza (Cenex, ora ESSnet), per mettere insieme competenze distribuite tra i diversi Istituti al fine di sviluppare azioni specifiche a beneficio dell'intero sistema. Un progetto ESSnet è infatti "*una rete di diverse organizzazioni ESS volte a fornire risultati che saranno vantaggiosi per l'intero Sistema Statistico europeo*". L'obiettivo è sfruttare le sinergie della cooperazione di alcuni Stati membri al fine di condividere competenze e risparmiare sui costi nella risoluzione di problemi comuni di interesse europeo. Il trasferimento dei risultati e delle conoscenze ai partner non partecipanti a beneficio dell'intero SSE è una caratteristica essenziale dei progetti ESSnet.

Lo *Scheveningen memorandum* apre quindi ufficialmente una prospettiva strategica di utilizzo di grandi masse di dati per finalità statistiche e dà avvio a complesse attività di sperimentazione e innovazione per esplorare il potenziale dei big data e avviare lo studio di metodologie adeguate a sfruttare le potenzialità dei big data e di definizione di complessi strumenti di valutazione della qualità dei dati utilizzati e degli indicatori prodotti.

Gli istituti di statistica europei per strutturare e inserire in un *framework* coerente le iniziative avviare e da avviare, elaborano in maniera condivisa il *Big data Action plan e roadmap*⁷⁰, documento strategico che contiene la risposta delle statistiche europee al processo di datificazione in atto nonché le azioni da intraprendere per mettere in atto quanto stabilito dal *Memorandum di Scheveningen*. La *Roadmap* prevede lo sviluppo di progetti Pilota per acquisire esperienza nell'uso dei big data nel contesto di statistiche ufficiali e prevede investimenti in

⁷⁰ European Statistical System, Big Data Action Plan and Roadmap, Task Force composed of representatives of national statistical offices, the OECD, the UNECE, DG CONNECT, the JRC and academic expert, European Statistical System Committee, 22nd meeting on 26 Sep 2014 in Riga.

ognuno dei settori produttivi e trasversali che contribuiscono al funzionamento di un Istituto di statistici⁷¹ tra cui:

- ✓ *Governance* e policy
- ✓ Metodologie
- ✓ Infrastrutture IT
- ✓ Comunicazione e diffusione
- ✓ Risorse umane
- ✓ Ambito giuridico e legale

Con l'approvazione della *Roadmap* nel corso del 22° Meeting dell'European Statistical System Committee a Riga il 26 settembre 2014, si dà avvio ai Progetti ESSnet Big data I e ESSnet Big Data II. Gli ESSnet sono progetti all'interno dei quali cooperano e collaborano tutti gli istituti facenti parte del Sistema statistico europeo e all'interno dei quali l'Istat ha sempre avuto un ruolo di punta.

ESSnet Big Data I: *dall'esplorazione allo sfruttamento* è un progetto all'interno del intrapreso congiuntamente da 22 istituti *partner* con l'obiettivo, come suggerisce la sua denominazione, di integrare i big data nella produzione corrente di statistiche ufficiali. L'ESSnet si snoda attraverso vari *pilot* ovvero progetti pilota ciascuno dei quali esplora il potenziale di fonti di big data selezionate e provvede a costruire applicazioni concrete. La prima edizione dell'ESSnet Big Data ha avuto inizio a febbraio 2016 e si è conclusa nel maggio 2018; era composto da 10 work packages (d'ora in poi WP nel testo) ,di cui due trasversali (uno di coordinamento e l'altro di diffusione dei risultati) e otto orientati ai contenuti:

- ✓ **WP1 *Web Scraping job vacancies***: lo scopo di questo progetto pilota è dimostrare, mediante stime concrete, quali approcci (tecniche, metodologiche, ecc.) sono più

⁷¹ Come vedremo nel capitolo successivo, questo approccio sistemico, basato su ciascuna delle leve organizzative e produttive che contribuiscono al normale funzionamento di un'organizzazione statistica, sarà lo stesso che guiderà l'elaborazione di un altro documento strategico: la Roadmap verso il sistema di produzione Trusted Smart Statistics.

adatti per produrre stime statistiche nel campo delle offerte di lavoro e in quali condizioni questi approcci possono essere usati nel sistema statistico europeo.

- ✓ **WP2 *Webscraping enterprise characteristics*** con l'obiettivo di indagare se le tecniche di *web scraping* (cioè tecniche di estrazione di dati da siti web, o altre tecniche tipo text mining e inferenza) possono essere utilizzate per raccogliere, elaborare e migliorare le informazioni generali sulle imprese.
- ✓ **WP3 *Smart meters***, si concentra sui contatori intelligenti, e più in particolare sui contatori elettrici che possono essere monitorati continuamente a distanza ed essere usati per statistiche sul consumo di energia, ma eventualmente anche su alloggi, spesa delle famiglie, impatto ambientale e produzione di energia.
- ✓ **WP4 *AIS Data Automatic Identification System***, Sensori rilevazione navi con l'obiettivo di esaminare se le misurazioni in tempo reale delle posizioni delle navi, misurate dal sistema di identificazione automatica AIS, possono essere usate per migliorare la qualità e la comparabilità interna delle statistiche esistenti e per sviluppare nuovi prodotti statistici.
- ✓ **WP5 *MND Mobile Network data*** Dati di telefonia Mobile con l'obiettivo di indagare e consentire l'accesso ai dati della telefonia mobile estremamente promettenti per l'arricchimento della statistica ufficiale ma con una serie di difficoltà da un punto di vista tecnologico, istituzionale e legale (privacy);
- ✓ **WP6 *Early estimates*** avente l'obiettivo di studiare più fonti di dati di diverso tipo al fine di produrre stime iniziali a fini statistici;
- ✓ **WP7 *Multiple domains*** per comprendere come una combinazione di diversi tipi di fonti (big data, dati amministrativi e dati di indagini statistiche) possa arricchire la produzione nei domini statistici della popolazione, del turismo, dell'agricoltura;
- ✓ **WP8 *Methodology*** si è concentrato sulla metodologia, sull'IT e sugli aspetti di qualità del progetto ed è partito una fase successiva quando cioè i primi 7WP avevano già prodotto i primi risultati.

Da questa fase di sperimentazione ne parte successiva legata all'ESSnet Big Data II che comprende questa volta 28 partner con l'obiettivo di sviluppare ulteriori progetti pilota per esplorare il potenziale di fonti di big data e per implementare applicazioni concrete. I progetti partono nel novembre 2018 e si concludono nel Giugno 2021 e sostanzialmente continuano i lavori cominciati nell'ESSnet Big data I con l'obiettivo di⁷²:

- ✓ Preparare l'implementazione di prototipi di successo nei seguenti settori:
 - offerte di lavoro online
 - Caratteristiche aziendali
 - *Smart meters*: Energia intelligente
 - AIS Monitoraggio delle navi
 - Processo e architettura
- ✓ Sviluppare nuovi progetti pilota su
 - Fatturazione elettronica
 - Osservazione della Terra
 - Dati di telefonia mobile
 - Statistiche del turismo innovativo
 - Metodologia e qualità
- ✓ Studiare applicazioni innovative nel dominio delle Trusted Smart Statistics

I primi promettenti rilasci di statistiche sperimentali e i risultati ottenuti dagli ESSnet Big Data spingono Eurostat a fare un passo in più: il 12 ottobre 2018 i responsabili degli uffici di statistica in seguito al 104° convegno DGINS (Directors General of National Statistical Institutes) dal titolo "*The European percorso verso Trusted Smart Statistics*", firmano il "*Bucharest Memorandum on Official Statistics in a datafied society (Trusted Smart Statistics)*"⁷³ (Allegato n.

⁷² Eurostat: ESSnet Big Data II Grant agreement Workpagage Coordination and communication, Derivable A4 Final version, 29 June 2021

⁷³European Statistical System - ESS, Directors General of the National Statistical Institutes - DGINS. 2018. Bucharest Memorandum on Official Statistics in a Datafied Society (Trusted Smart Statistics), Bucharest, 10th and 11th October 2018. 104thDGINS Conference

2). Con l'approvazione di questo memorandum, la *governance* degli Istituti di statistica europei ha riconosciuto che gli sviluppi rapidi e trasformativi dell'Internet delle cose e, più in generale, la pervasività delle tecnologie digitali nella vita quotidiana come strumenti di relazione con il mondo, rappresentano sfide, opportunità, minacce e responsabilità fondamentali per il Sistema statistico europeo. In sostanza, il tema della conferenza DGINS si è concentrato sul ruolo futuro delle statistiche ufficiali in una società prevalentemente digitale. Il Memorandum di Bucharest riconosce i risultati raggiunti finora e identifica le priorità per sviluppare ulteriormente il sistema statistico europeo.

Il Memorandum di Bucharest come vedremo in maniera diffusa nel prossimo capitolo, ha introdotto il concetto di *Trusted Smart Statistics* per valorizzare ancora di più il contributo dei big data in termini di validità, accuratezza e affidabilità degli *output*. Ciò aggiunge importanza alla missione di alto livello delle statistiche ufficiali e degli Istituti di statistica. Le mutevoli caratteristiche dei nuovi dati richiedono un profondo ripensamento dei modelli di produzione statistica. Per adempiere alla stessa missione e rispettare gli stessi principi che li hanno qualificati per decenni, i sistemi statistici devono adattare i loro processi produttivi rendendoli funzionali alle nuove fonti di dati alle tecnologie digitali. Più che un processo fisiologico di aggiornamento delle componenti esistenti, si tratta di un aggiornamento radicale a livello di sistema nel funzionamento dei sistemi statistici. Per descrivere questo salto evolutivo ed etichettare l'evoluzione delle statistiche ufficiali nella nuova società basata sui dati, il Sistema Statistico Europeo come vedremo meglio nel capitolo successivo, ha adottato l'espressione *Trusted Smart Statistics* ed ha individuato la direzione di marcia e cominciato a compiere i primi passi e sta guidando questa complessa evoluzione con una visione globale a livello di sistema.

Siamo quindi di fronte ad un ulteriore e fondamentale cambiamento di paradigma reso possibile sia dalla disponibilità di tecnologie mobili ma anche dai nuovi comportamenti che questa disponibilità ha favorito: l'interazione continua con le *app mobili* è considerata al giorno d'oggi piuttosto normale. Non si tratta solo di sfruttare le nuove tecnologie digitali,

ma anche di fare leva sui nuovi e nuovi comportamenti digitali. Non è una questione solo di disponibilità di dati, sono cambiati atteggiamenti e comportamenti⁷⁴, i cittadini hanno sviluppato attraverso l'uso quotidiano delle tecnologie nuovi comportamenti digitali. Il digitale, ovvero l'ecosistema, l'infrastruttura in cui noi tutti ci muoviamo, e in cui si muovono anche gli Istituti di statistica, ha determinato nuovi comportamenti che la statistica non può ignorare perché sono una rappresentazione, *una tomografia al microscopio* della nostra società e rendono possibili predizioni un tempo impensabili.⁷⁵ Cambiano così anche le modalità nella rilevazione dei dati, basti pensare al passaggio da indagini tradizionali alle *smart survey* che sono indagini che usano dispositivi *smart* per rilevare dati sia da questionario sia soprattutto da sensori raccolti in modo passivo o attivo. In altre parole, se "*siamo i nostri strumenti*" come sostenevano Boyd e Crawford, il cittadino con lo *smartphone* oggi è un soggetto diverso dal cittadino del secolo scorso (senza *smartphone*), con comportamenti diversi ma anche con una visione e aspettative diverse rispetto a dati e statistiche. E questo cambio di comportamenti, punti di vista e aspettative influenzano il ruolo e la relazione con la Statistica ufficiale⁷⁶ (Ricciato et al. 2020)

I progetti ESSnet non si sono fermati al rilascio di *derivable* tecnici all'interno del sistema Statistico Europeo, ma hanno proseguito la loro fase di sviluppo e implementazione dando vita a statistiche sperimentali, migliorando i processi produttivi ad essi connessi e spingendo verso nuovi progetti Europei tra cui ricordiamo l'ESSNet *Trusted Smart Surveys*, l'ESSnet *MND Mobile Data Network*, ESSnet *WIN Web Intelligence Network*, legato alla produzione di statistiche con i dati *IaD Internet as Data Source* (l'Istat è coinvolto con il progetto *OJA Online Job advertising* proveniente dal WP1 e il progetto *OBEC online enterprises characteristics* da WP2). In questo momento è in fase di valutazione la partecipazione dell'Istat a un bando Europeo su "*TSS multi-MNO: Development, implementation and demonstration of a reference processing pipeline for the future production of official statistics based on multiple Mobile Network*).

⁷⁴ Kennedy H., *How People Feel About What Companies do with Their Data is Just as Important as What They Know About It*. Impact of Social Sciences Blog. 2018

⁷⁵ MacFeely S., *The Continuing Evolution of Official Statistics: Some Challenges and Opportunities*, Journal of Official Statistics, Vol. 32, No. 4, pp. 789–810, <http://dx.doi.org/10.1515/JOS-2016-0041>, 2016

I risultati di alcuni di questi WP legati agli ESSnet sono stati ulteriormente sviluppati, alcuni nel 2020 inseriti nei progetti prioritari nella *Roadmap 2021-2024* verso il sistema di produzione delle *Trusted Smart Statistics* che sarà oggetto di approfondimento nel capitolo successivo, e altri fin da subito hanno permesso il rilascio, in molti Istituti europei delle prime statistiche sperimentali che hanno segnato un vero e proprio punto di svolta nel nuovo paradigma di produzione statistica.

2.3.1 Le statistiche sperimentali

Per essere pienamente sfruttati a fini statistici, i big data richiedono importanti investimenti in denaro e tempo e, soprattutto, in competenze. Ed è per questo motivo che ricerca e innovazione nell'ultimo decennio si sono ulteriormente rafforzati e per sostenere la strategia big data dell'Istat e, più in generale, dell'intero processo di modernizzazione dell'Istat⁷⁷. Le prime attività condotte dell'Istat di sperimentazione ed esplorazione dell'utilizzo dei big data per migliorare la qualità delle statistiche correnti o per sviluppare nuovi indicatori sono state orientate dal *Big data Committee* un Comitato formato da esperti internazionali e nazionali, oltre che da membri Istat, con un programma di sperimentazioni su diversi temi. Un avanzamento di rilievo è stato realizzato a partire dal 2018 con la messa a regime dell'utilizzo di *scanner data* (cioè i dati che si ricavano dai lettori dei codici a barre sui prodotti dei supermercati) come fonte che affianca le rilevazioni Istat per il calcolo mensile dell'inflazione, con riferimento ai prodotti alimentari e per la cura della persona e della casa e soprattutto con la pubblicazione delle prime statistiche sperimentali. L'Istituto, nell'ottica di selezionare gli usi più promettenti e sostenibili dei big data per incorporarli nella produzione ufficiale, dopo le prime sperimentazioni avviate in ambito europeo, sta da anni investendo sull'uso dei dati di telefonia mobile per rilevare i flussi di mobilità e turismo, sui dati generati dagli *Smart meters* per i consumi energetici, sulle informazioni sul traffico ottenibili da *webcam* per il rischio d'incidentalità e per le statistiche sul traffico, e sull'uso di immagini satellitari per

⁷⁷ Istat. Big Data Committee – Annual Report ISTAT, Chapter 1 –Alleva Giorgio, 2018, <https://www.istat.it/it/files//2018/09/Big-data-committee.pdf>

statistiche sull'uso e la copertura del suolo, la stima delle aree verdi e la stima delle produzioni agricole⁷⁸.

Con il rilascio delle statistiche sperimentali, l'Istat capitalizza la partecipazione ai progetti ESSnet e gli investimenti fatti in ricerca e innovazione e nella infrastrutturazione delle attività di ricerca.

Non esiste una definizione di statistica sperimentale condivisa nella comunità statistica, quella che maggiormente ne rispecchia la natura e le caratteristiche è quella che ne dà l'Ufficio Statistico Inglese secondo cui le statistiche sperimentali sono statistiche che *in the testing phase and not yet fully developed*⁷⁹, sono appunto in fase di sperimentazione e non ancora pienamente sviluppate. Sono prodotti statistici non ancora pronte per essere diffuse in maniera ufficiale perché basate su un impianto metodologico ancora non maturo, ma sono prodotti che fanno "parlare i dati", diffondono l'informazione in maniera più tempestiva, con maggiore dettaglio territoriale e su diversi ambiti tematici. Sono definite sperimentali in quanto non hanno raggiunto la piena maturità e sono soggetti a ulteriore sviluppo o miglioramento, ciò significa che non soddisfano ancora tutti i rigorosi requisiti per essere statistiche ufficiali. Nonostante i limiti, le statistiche sperimentali hanno un alto potenziale in quanto sono in grado di colmare lacune informative, con una tempestività⁸⁰ maggiore e di suggerire nuove analisi e nuovi indicatori.

Le statistiche sperimentali sono pubblicate in un'area specifica del sito istituzionale⁸¹ e sono contrassegnate da un logo dedicato per renderle inequivocabilmente riconoscibili rispetto alla statistiche ufficiali. Tutte le statistiche sperimentali prodotto dagli Istituti Europei hanno un'area dedicata sui rispettivi siti istituzionali per evitare possibili confusioni rispetto ai prodotti di statistica ufficiale. Anche Eurostat raccoglie in una sezione denominata *Experimental Statistics Hub* i prodotti sperimentali rilasciati dai singoli istituti⁸². Ciascuna

⁷⁸ Alleva G., Focus: Pubblica Amministrazione: Digitalizzazione, Informazione e Dati, In Nuova Etica Pubblica Rivista dell'Associazione Etica PA, Anno 6, n. 11 – luglio 2018

⁷⁹ www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/guidetoexperimentalstatistics

⁸⁰ La tempestività si riferisce allo sfasamento tra la data di pubblicazione e il periodo a cui si riferiscono le statistiche

⁸¹ <https://www.istat.it/it/statistiche-sperimentali>

⁸² <https://ec.europa.eu/eurostat/web/experimental-statistics/overview/ess>

statistica sperimentale è accompagnata da una specifica nota metodologica, che ne descrive appunto l'impianto metodologico e le modalità di implementazione. Per ciascuna delle Statistiche sperimentali pubblicate sul sito è pubblicato l'invito a rilasciare dei *feedback*.

Il *feedback* degli utenti sottolinea la natura antropocentrica di questi prodotti e dà centralità ai cittadini: gli utenti possono svolgere un ruolo importante nel promuovere il passaggio dalle statistiche sperimentali a quelle ufficiali, manifestando un particolare interesse per una specifica statistica sperimentale, in quanto ne hanno bisogno per attività di ricerca e/o per accelerare e favorire un processo decisionale.

Per quanto riguarda l'esperienza Istat, una statistica sperimentale dopo un periodo medio di tre anni dalla prima pubblicazione può:

- ✓ venire regolarmente aggiornato con i dati più recenti disponibili e rimanere sperimentale;
- ✓ diventare statistica ufficiale;
- ✓ rimanere una statistica *one-shot*.

Il passaggio a statistica ufficiale, così come l'approvazione di una statistica sperimentale, è compito del Comitato per la Ricerca (una delle infrastrutture per la ricerca di cui si parlerà nel capitolo successivo) che ha funzioni di *governance* sulle attività di Ricerca e innovazione condotte in Istituto. Per diventare ufficiale una statistica sperimentale deve soddisfare una serie di criteri e provvedere al rafforzamento e al miglioramento delle metodologie e delle affidabilità delle stime, della copertura geografica; della pertinenza e della tempestività e deve ricevere *feedback* positivi da parte degli utenti.

Nella tabella che segue si riportano le statistiche sperimentali disponibili a settembre 2020⁸³

Tabella 2 - Statistiche sperimentali Istat*

⁸³ Carciotto A., Signore M., Improving relevance: Istat experience on experimental statistics, Statistical Journal of the IAOS 37 (2021) 593–601 593 DOI 10.3233/SJI-200764 IOS Press

Categoria	Prodotti	Data rilascio
CLASSIFICAZIONI NON-STANDARD	Famiglie per gruppo sociale	Aprile 2018
	Modalità di internazionalizzazione delle imprese	Aprile 2018
	Classificazioni delle generazioni	Aprile 2018
	Classificazione dei sistemi locali	Aprile 2018
	Classificazione dei comuni secondo le ecoregioni d'Italia	Dicembre 2018
NUOVI INDICATORI	Digitalizzazione delle imprese e indicatori di <i>performance</i> economica	Marzo 2020
	Comportamenti di impresa e sviluppo sostenibile	Marzo 2020
	Vent'anni di occupazione dal lato della qualificazione del lavoro	Marzo 2019
	Indicatori di disuguaglianza retributiva nelle piccole imprese	Novembre 2018
	A misura di Comune	Agosto 2018
ANALISI E QUADRI INTERPRETATIVI	I percorsi museali in Italia	Dicembre 2019
	Popolazione insistente per studio e per lavoro	Marzo 2020
	Conti economici integrati e ambientali nel turismo	Marzo 2019
SPERIMENTAZIONI CON BIG DATA	<i>Social Mood on economy Index</i>	Ottobre 2018
	Modalità di uso dei siti web da parte delle imprese	Giugno 2018
	Uso di <i>Open Street map</i> per il calcolo dell'incidentalità stradale	Luglio 2019

*(mia elaborazione a partire dalla Tabella 1 contenuta in Carciotto Signore)

Tra questi primi prodotti rilasciati, una grossa eco ha avuto la produzione di un indice sperimentale, denominato "*Social Mood on Economy Index*" che l'Istat diffonde e aggiorna trimestralmente ormai dal 2018. L'indice fornisce misure ad alta frequenza - basate su una complessa metodologia di filtraggio, classificazione e aggregazione - sull'evoluzione del *sentiment* italiano sull'economia, derivate da campioni di *tweet* pubblici in lingua italiana, catturati in *streaming*. Si tratta di un indicatore che ha un valore informativo in sé e se messo in relazione ad altre serie economiche o sociali anche a più bassa frequenza.

Da questa prima statistica, a distanza di qualche mese dall'inizio della guerra in Ucraina, è stata sviluppata un'ulteriore analisi "*La Guerra in Ucraina nei tweet italiani*" basata su dati

dei *social media*, e in particolare su dati *Twitter*, che ha analizzato le conversazioni sulla guerra Russia-Ucraina dei cittadini intervenuti su *Twitter*, prendendo in considerazione i *tweet* relativi ai periodi dal 24 febbraio all'8 marzo (le prime due settimane di guerra), dall'8 al 20 maggio e dall'1 al 15 giugno. Questo lavoro costituisce un ulteriore prototipo per la produzione di *output* statistici basati su dati testuali in contesti socioeconomici, ed è stato inserito nell'ultimo Rapporto Annuale Istat 2022 presentato l'8 luglio a Palazzo Montecitorio, dal Presidente dell'Istat Gian Carlo Blangiardo⁸⁴. L'iniziativa si inquadra perfettamente nella strategia generale dell'Istituto di migliorare la qualità dell'informazione rilasciata e diffusa attraverso l'uso delle nuove fonti e prevede la sperimentazione e la messa in produzione di innovazioni tecnico-metodologiche applicate ai processi di produzione statistica basata sull'uso di fonti di dati di tipo testuale attraverso la definizione di metodi per la rappresentazione e l'analisi di testi e di strumenti per la navigazione, visualizzazione e analisi assistita di informazione testuale.

Le statistiche sperimentali quindi sono la modalità con cui gli Istituti portano in produzione i loro investimenti in ricerca e innovazione sulle nuove fonti di dati, ma rappresentano anche uno strumento con cui la comunità statistica vuole avvicinarsi ai propri *stakeholders* offrendo informazione fruibile in maniera più agile e rafforzando il rapporto di fiducia tra il mondo istituzionale e scientifico e la società civile.

2.4 Fiducia e statistica ufficiale: una relazione possibile

Nell'ambito della società dell'informazione gli Istituti di Statistica non appaiono più come produttori esclusivi di dati, ma come uno dei diversi soggetti che concorrono a tale processo, le statistiche sono diventate una *commodity*. Contestualmente a questo, abbiamo assistito ad un cambiamento radicale del modello di *business* del mondo dell'informazione e della comunicazione. Da sempre la competizione ha per oggetto l'attenzione degli utenti, ma prima questo avveniva a partire da una posizione privilegiata dei tradizionali soggetti *istituzionali*: esisteva un numero limitato di produttori di dati, oggi la situazione è cambiata.

⁸⁴ Istat, Rapporto Annuale 2022. La situazione del Paese, 2022

Ci sono molti più soggetti che competono per l'attenzione dell'utente. Il dibattito sull'infodemia, tende a concentrarsi sulla quantità delle informazioni che circolano, ignorando il fatto che è un intero modello di *business* ad aver soppiantato quello precedente: è aumentata la quantità di contenuti prodotti e la velocità con cui essi vengono diffusi. L'informazione prodotta genera meno attenzione e la qualità complessiva del dibattito pubblico diminuisce⁸⁵. Gli studi dell'Agcom, Agenzia per la telecomunicazioni e la concorrenza, dimostrano che i fenomeni patologici di disinformazione tendono ad annidarsi lì dove il sistema dell'informazione fallisce⁸⁶. L'Istat in questo panorama complesso, gode di una posizione assolutamente privilegiata: è innanzitutto il principale produttore di statistica ufficiale e quindi di informazione a supporto dei cittadini e dei decisori pubblici e la qualità dei prodotti diffusi è indiscussa perché garantita da rigidi codici sulla qualità delle statistiche ufficiali sia italiani⁸⁷ sia europei⁸⁸. Le statistiche ufficiali rappresentano un'infrastruttura dell'informazione "certificata" perché garantita dal rispetto di criteri stringenti ed è principalmente ciò che differenzia la statistica ufficiale dalle altre informazioni disponibili: le statistiche ufficiali sono il risultato di un processo di natura scientifica, il valore dei dati è nei metodi statistici che garantiscono servizi di qualità.⁸⁹ È nella qualità che si gioca la partita per assicurarsi la fiducia dei cittadini. In questi ultimi due anni, l'informazione statistica ha costituito uno strumento ancora più indispensabile per capire le condizioni del paese, del territorio e della collettività: gli eventi più recenti ci hanno mostrato che il mondo coinvolto nella pandemia si è trovato all'improvviso precipitato nell'epoca digitale e ha scoperto la centralità dei dati e della *data science*. La stessa lotta alla pandemia ha improvvisamente fatto comprendere a tutti, quanto siano importanti i dati, le informazioni essenziali relative ai

⁸⁵ Quattrocioni W., Scienza, dati e decisione. Intervista a Walter Quattrocioni – pubblicato sul numero 3 del 2020 di Pandora Rivista dedicato alle piattaforme.

⁸⁶ AGCOM, Autorità per le garanzie nelle comunicazioni, *News vs fake news nel sistema dell'informazione*, <https://www.agcom.it/documents/10179/3744102/Allegato+22-11-2018/3aff8790-8039-4456-8f9a-dae2497289a4?version=1.0>, novembre 2018.

⁸⁷ Sistan Codice italiano per la qualità delle statistiche, Direttiva n. 12/Comstat Pubblicata nella G.U. n. 23 (Serie Generale) del 29 gennaio 2022.

⁸⁸ Commissione europea, Eurostat, *Codice delle statistiche europee : per le autorità statistiche nazionali ed Eurostat (autorità statistica dell'UE)*, Ufficio delle pubblicazioni, 2018, <https://data.europa.eu/doi/10.2785/948366>

⁸⁹ Radermacher W. J., The future role in official statistics, in *Power from Statistics: data, information and knowledge*, Eurostat 2018.

fenomeni che si vogliono conoscere, analizzare e imparare a contrastare. Le statistiche sono e devono rimanere un modo per offrire conoscenza sulle nostre società, sono strumento di razionalità e di valorizzazione del processo decisionale. Nel contesto della post-verità sono un punto di riferimento per raccontarci quello che siamo e da dove veniamo, sono una bussola che ci permette di osservare, valutare e orientarci, sono un elemento politico, non una politica in sé, devono contribuire a razionalizzare il dibattito piuttosto che strumentalizzarlo. La comunità statistica deve interagire con il pubblico e cooperare in modo ancora più intenso con i vari utenti e *stakeholder*, siano essi decisori pubblici o privati, giornalisti, ricercatori o cittadini.

Gli investimenti che l'Istat ha sostenuto in maniera incisiva per potenziare l'informazione statistica arricchendola di contenuti, misurando i cambiamenti e le trasformazioni in atto in campo economico e sociale e contribuendo a rafforzare e accrescere la fiducia nei dati, vanno in questa direzione. Nell'era della rivoluzione digitale, in un contesto in cui si moltiplicano i dati e i produttori di informazione, è fondamentale comunicare il valore dei dati per rafforzare il ruolo degli istituti di Statistica e migliorarne il posizionamento nel mondo dell'informazione.

“Cambia la comunicazione, che ha raggiunto una velocità mai conosciuta prima, che non chiude mai, 24 ore su 24, ed è diretta, da punto a punto, spesso anonima, concisa e spesso brutale e riduttiva, polarizzata, affidata più all'immagine che al ragionamento, e capace di generare eco dalla portata sconosciuta e incontrollata. Discernere e selezionare segnali informativi sovrabbondanti, a volte, sembra un'impresa disperata. Si diventa bulimici, infobesi. Si ingeriscono con superficialità quantità esagerate di dati senza ordine, gerarchia, senso. Si rilanciano, si dimenticano. Si perde l'orientamento. E mentre i fenomeni diventano sempre più complessi e globali, saltano i monopoli dell'interpretazione e dell'autorevolezza, e si generano reazioni di rigetto, in parte anche causate da quella che ormai si chiama affaticamento da fatti (fact fatigue): rigetto delle istituzioni, delle evidenze e dell'expertise (basti pensare al diffondersi delle pratiche mediche fai-da-te); l'orgoglio dell'ignoranza”⁹⁰.

⁹⁰ Alleva G., *Relazione di apertura, Dall'incertezza alla decisione consapevole: un percorso da fare insieme* Atti della 13a Conferenza nazionale di statistica Roma, 4-6 luglio 2018

In questo ulteriore momento dell'intervento del Prof. Alleva alla XIII Conferenza nazionale di statistica si introducono molti elementi fondamentali da considerare se si vuole valorizzare il ruolo degli INS e della statistica ufficiale e il loro contributo all'avanzamento della conoscenza e al rafforzamento della comunicazione scientifica. Oltre a riadattare il sistema di statistica ufficiale sviluppato per il mondo analogico al mondo digitale, gli INS devono adattarsi ai nuovi *business model* del mondo dell'informazione: sono cambiate le modalità di reperire le informazioni. Le audience oggi consumano informazioni in modi molto differenziati lasciando una traccia di dati dei loro percorsi⁹¹. I dati in tutte le forme sono sempre più centrali in tutte le nostre vite. Questo sta creando una grande opportunità per il sistema statistico, per fornire informazioni su ambiti strategici per i decisori pubblici. L'aumento dei dati e anche delle piattaforme e delle modalità in cui questi dati circolano comporta anche una possibile diffusione di informazioni non affidabili che rendono ancora più necessario un sistema statistico sicuro che possa garantire la qualità dell'informazione rilasciata e diffusa. Ora l'informazione che circola sulle piattaforme è disintermediata, abbiamo a disposizione troppe informazioni e non abbiamo tempo di processarle e non c'è chi lo fa per noi.

Quando, come cittadini, siamo posti di fronte per esempio ai risultati di una sperimentazione scientifica o a una nuova teoria economica o sociale, spesso non abbiamo le competenze per controllarne da soli la veridicità, la fiducia nelle competenze degli esperti, nella fonte dei dati e nel sistema informativo è quindi un elemento essenziale. Il presidente della Royal Statistical Society britannica David Spiegelhalter, nel suo discorso di insediamento nel 2017, ha sostenuto che il vero problema dell'informazione non sono solo le fake news, ma anche la manipolazione e l'inappropriata interpretazione dei dati attraverso pratiche discutibili di interpretazione e comunicazione. Al cuore del problema, secondo Spiegelhalter, c'è il rapporto tra truth e trust, tra verità e affidabilità.

Nei processi di diffusione dell'informazione e in quelli di comunicazione, più che di *fact-checking* converrebbe parlare di *trust building*: si tratta di offrire delle verità e non solo e non

⁹¹ Van Dijck J., Poell T., De Waal M., *The Platform Society – Public Values in a Connective World*, Oxford University Press, New York 2018

sempre di verificare quelle presunte che sono già in circolazione. L'ingranaggio fondamentale della serietà informativa, ricorda Lorusso, è l'affidabilità che non può essere irrazionale moto di adesione. L'affidabilità è credibilità, fiducia nella persona che parla (prima che nella verità dei fatti) e "corrispondenza" ai saperi consolidati (dove corrispondenza vorrei che qui fosse intesa non come rispecchiamento, adesione, ma come adeguatezza, capacità di relazionarsi, rispondere e corrispondere a sapere consolidato già diffuso). Dire che la verità ha a che fare con meccanismi fiduciari non significa buttarla alle ortiche significa focalizzarsi e lavorare sui criteri da adottare per discriminare i soggetti affidabili. La verità è *matter of trust* molto più che *matter of fact* anche se la fiducia non è cieca, non è irrazionale, non è immotivata: è giudizio e discernimento⁹². In un contesto di questo tipo, l'Istat gode di una posizione assolutamente privilegiata, la vera sfida è saper sfruttare le opportunità che questo posizionamento offre. Un'informazione più vicina e più rispondente ai bisogni dell'utente accelera il processo di rafforzamento della fiducia.

Il problema dell'infodemia e delle *fake-news* non può assolutamente ricondursi al solo *fact checking* e alla mera definizione di una serie di algoritmi potenzialmente in grado di stabilire se una notizia sia vera o meno (chi mai può stabilire quale sia la verità?). La ricerca delle possibili soluzioni al problema va affrontata su un piano socio-culturale. Le infrastrutture tecnologiche - considerate da qualcuno come soluzioni - vanno intese come componenti di un'architettura più ampia in grado di accompagnare e supportare il cambiamento offrendo fiducia, un'infrastruttura sociotecnica con componente *hardware, software e humanware*.⁹³

Il problema della disinformazione trova terreno fertile in un contesto sociale caratterizzato da una diffusa mancanza di fiducia verso le istituzioni, i produttori dei dati e il sistema informativo. È necessario un articolato e complesso impegno pubblico e privato nell'educazione alla società digitale e al pensiero e alla valutazione critica. Il migliore antidoto contro la cattiva informazione risiede in una rinnovata fiducia nel sistema

⁹² Lorusso, A., *Postverità, Laterza, Roma-Bari 2019.*

⁹³ Ricciato F., Giannakouris K., Wirthmann A., Hahn M. *Trusted Smart Surveys: a possible application of Privacy Enhancing Technologies in Official Statistics*, conference paper presentato alla 50 Conferenza Scientifica della Società Italiana di statistica a Roma, 2020

informativo, nella crescita della qualità dell'informazione, a partire da quella pubblica. La sfida da vincere è quella di ridurre la distanza tra la domanda e l'offerta di informazione di qualità ed questa la direzione che deve percorrere anche la statistica ufficiale. Per far fronte alla crescente quantità di dati, le persone devono essere più consapevoli della qualità e dell'affidabilità dei dati. La qualità dovrebbe diventare il fattore decisivo nella scelta di una fonte di dati. È qui che le statistiche ufficiali possono distinguersi dagli altri produttori di dati.

Gli Istituti di statistica dovrebbero sviluppare strategie di promozione per pubblicizzare i propri punti di forza, assicurandosi che si colleghino in modo efficace ed efficiente con i propri utenti e costruendo la propria reputazione come fornitori di dati affidabili⁹⁴.

Allegato 1 –Scheveningen Memorandum

Scheveningen Memorandum

Big Data and Official Statistics

⁹⁴ Von Oppeln-Bronikowski S, Hagenkort-Rieger S., João Santos M., *New trends in communication: Branding and content marketing*, in *Power from Statistics: data, information and knowledge*, Eurostat 2018 pag. 117

The DGINS

CONSIDERING

1. Recent innovations in the information and communication technologies have been leading to an increasing degree of digitization of economies and societies at all levels that offer new opportunities for the compilation of statistics.
2. The use of Big Data for statistical purposes challenges the European Statistical System to effectively address a variety of issues.
3. The demand for timely and cost efficient production of high quality statistical data increases, as well the need for new solutions to declining response levels.
4. Official statistics should incorporate as much as possible all potential data sources, including Big Data, into their conceptual design.
5. The distinguishing aspect of many Big Data sources is that they are not confined to national borders and, as such, represent unique opportunities for collaboration at European level as well as on global level.
6. Many European initiatives have a link to Big Data, including the European Commission's ambition for developing a strategy for the European data value chain, the on-going EU Data Protection reform and the Horizon2020 program.
7. The implementation of new methods of production of European statistics represents an objective of the European Statistical Programme 2013-2017⁹⁵ and aims at efficiency gains and quality improvements, including increased timeliness.

The DGINS

1. Acknowledge that Big Data represent new opportunities and challenges for Official Statistics, and therefore encourage the European Statistical System and its partners to effectively examine the potential of Big Data sources in that regard.

⁹⁵ Regulation (EU) No 99/2013 of the European Parliament and of the Council of 15 January 2013 on the European statistical programme 2013-17, OJ L 39, 9.2.2013, p. 12–29

2. Recognise that Big Data is a phenomenon which is impacting on many policy areas. It is therefore essential to develop an 'Official Statistics Big Data strategy' and to examine the place and the interdependencies of this strategy within the wider context of an overall government strategy at national as well as at EU level.
3. Recognise that the implications of Big Data for legislation especially with regard to data protection and personal rights (e.g. access to Big Data sources held by third parties) should be properly addressed as a matter of priority in a coordinated manner.
4. Note that several NSIs are currently initiating or considering different uses of Big Data in a national context. There is a momentum to share experiences obtained from concrete Big Data projects and to collaborate within the ESS and beyond, on a global level.
5. Recognise that developing the necessary capabilities and skills to effectively explore Big Data is essential for their integration into the European Statistical System. This requires systematic efforts like appropriate training courses and establishing dedicated communities including academics for sharing experiences and best practice.
6. Acknowledge that the multidisciplinary character of Big Data requires synergies and partnerships to be effectively built with experts and stakeholders from various domains including government, academics and owners of private data sources.
7. Acknowledge that the use of Big Data in the context of official statistics requires new developments in methodology, quality assessment and IT related issues. The European Statistical System should make a special effort to supports these developments.
8. Agree on the importance of following up the implementation of this memorandum by adopting an ESS action plan and roadmap by mid-2014 that should be further integrated into the Statistical Annual Work Programmes of Eurostat.

Allegato 2 – Bucharest Memorandum

BUCHAREST MEMORANDUM

on Official Statistics in a Datafied Society (Trusted Smart Statistics)

104th DGINS Conference, Bucharest, 10th and 11th October 2018

As adopted by the European Statistical System Committee (ESSC) meeting on the
12th October 2018

Considering

1. That society and economy are being rapidly and fundamentally transformed by pervasive digital technologies and the process of datafication.
2. That official statistics need to remain relevant in the future and continue to satisfy users with high quality statistics.
3. The progress made in implementing the ESS Vision 2020 and the [Scheveningen Memorandum on big data and official statistics](#).
4. The ESSC's adoption on 8 February 2018 of the [smart statistics and big data business case](#), which describes the main achievements and sets out the strategic orientations for big data and smart statistics for the period 2018-2020.
5. That joining forces and developing expertise within the ESS has resulted in synergies and achievements that would not have been possible at an individual level.
6. That the integration of multi-source and heterogeneous data serving multiple statistical domains provides opportunities to produce new relevant statistics and make efficiency gains by sharing sources, methods, software and services.
7. That new data sources represent a unique opportunity to produce new and improve existing statistics within a collective collaborative framework.
8. The development of smart systems based on the expansion of the world-wide network of interconnected devices (Internet of Things), which embed smart technologies enhanced by artificial intelligence.
9. That evidence-based policymaking requires trusted statistics based on the principles of official statistics as set out in the [European Statistics Code of Practice](#), maintaining high quality and sound methodology is of utmost importance in this process.
10. The [General Data Protection Regulation](#), which sets out new principles and lays down requirements, such as privacy by default and privacy by design.
11. That addressing global statistical challenges and opportunities in a digital world requires effective cooperation with the UN and other international statistical bodies, and partnerships with national, European and global stakeholders.

The Directors-General of national statistical institutes and of Eurostat

1. Welcome and embrace the opportunity for official statistics in a datafied society and economy.
2. Encourage the ESS to implement practical and mature cases of using 'big data-enhanced' statistical products and develop experimental statistics on new phenomena.
3. Agree that the variety of new data sources, computational paradigms and tools will require amendments to the statistical business architecture, processes, production models, IT infrastructures, methodological and quality frameworks, and the corresponding governance structures, and therefore invite the ESS to formally outline and assess such amendments.
4. Recognise the need to further develop the legal framework at European and national level to reduce obstacles to the access, use and integration of heterogeneous data for sustainably producing Trusted Smart Statistics.
5. Recognise the importance of privacy by design approaches and encourage the ESS to explore the potential of privacy-preserving computation technologies, such as secure multiparty computation, within a wider framework of Trusted Smart Statistics.
6. Encourage the creation of dedicated smart statistics communities to ensure that knowledge and skills are shared, and that achievements and further developments are implemented in a sustainable way. These communities could be organised within specific application and subject domains, e.g. smart cities, m-health, smart agriculture, citizen science, etc.
7. Acknowledge that exploring the application of artificial intelligence to official statistics and the use of smart technologies require that new standards be adopted, which can ensure interoperability between various data systems and adherence by design to ethical, legal, transparency, and quality principles.
8. Assign high priority to developing and maintaining the necessary skills to cover technical, scientific, organisational, communication and change-management capabilities.
9. Agree on the need to engage in partnerships with key stakeholders, such as citizens, data protection authorities, central banks, industry, private data holders, relevant interest groups, standardisation bodies and the scientific and academic community.
10. Underline the need for European and international coordination, and invite national statistical institutes, Eurostat, the UN and other international statistical bodies to develop effective collaborative platforms to ensure synergies, alignment, streamlining of activities and communication initiatives at European and global levels.
11. Agree on the importance of following up the implementation of this memorandum by adopting ESS action plans and roadmaps (by mid-2019) that should be further integrated into the European Statistical Program (ESP) and the Annual Planning and Monitoring procedure of Eurostat.

Capitolo 3 - Le conseguenze della trasformazione digitale in Istat: dal programma di modernizzazione al Centro per le *Trusted Smart Statistics*.

3.1 Il processo di digitalizzazione in Istat

Nel capitolo precedente si è discusso dell'evoluzione del modo di fare statistica e di come gli istituti di statistica abbiano saputo riportare all'interno dei propri *business model* gli impatti determinati dalla trasformazione digitale. Il contesto esterno ha agito e ha dato un senso ai nuovi modi di fare statistica. Sebbene il presente lavoro non sia un trattato giuridico, è a mio avviso indispensabile tracciare anche l'evoluzione del contesto normativo italiano che ha accompagnato e sostenuto il processo di digitalizzazione della Pubblica Amministrazione (d'ora in poi PA nel testo) per capire anche l'impatto che l'evoluzione normativa ha avuto sul funzionamento complessivo dell'Istat in tema di orientamento all'innovazione di prodotti e di processi. Le norme fanno parte di quel contesto esterno che costituisce, con la sua presenza invisibile, una cornice di riferimento da cui non si può prescindere per non cadere nella trappola del solo determinismo tecnologico. Il percorso evolutivo delle norme che hanno favorito e accompagnato l'evoluzione digitale in Istat, rispecchia i diversi contesti sociali in cui queste norme si sono via via innestate esattamente in coerenza con la definizione che del diritto ha dato il famoso giurista e accademico Feliciano Benvenuti considerandolo *"la materia viva che producono i rapporti sociali, continuamente in movimento"*⁹⁶.

Nonostante già a partire dagli anni '90 in Italia la separazione tra il mondo digitale e la PA comincia a farsi più labile, fino ad una quindicina di anni fa le proposte innovative contenute nelle norme che spingevano la PA verso l'uso di strumenti digitali, sono state applicate in misura estremamente marginale. Probabilmente fino ai primi anni Duemila, il contesto esterno e quello interno alle PA in generale non erano sufficientemente maturi da favorire

⁹⁶ F. Benvenuti, *Il nuovo cittadino. Tra libertà garantita e libertà attiva*, cit., 119, 1994

l'integrazione del digitale nel funzionamento delle PA. Cambiamenti così impattanti di fatto, non si realizzano solo grazie a una previsione normativa, condizione sicuramente necessaria ma non sufficiente a determinare di per sé un cambiamento reale. Gli sforzi di riforma infatti non hanno in gran parte raggiunto i risultati sperati ed annunciati anche perché spesso hanno enunciato principi senza poi indicare strumenti efficaci e forme di accompagnamento alle riforme stesse. Le innovazioni digitali proposte infatti avevano una natura meramente amministrativa e non coinvolgevano le singole amministrazioni con un approccio sistemico, ma individualistico e verticale. Massimo Severo Giannini, studioso di chiarissima fama e ex Ministro per la Funzione Pubblica già nel 1979 trasmise alle Camere l'indimenticato "Rapporto sui principali problemi della Amministrazione dello Stato" in cui evidenziava l'*impreparazione* della P.A. rispetto al processo tecnologico ed alla sua evoluzione e già coglieva l'aspetto più significativo che ha poi caratterizzato l'informatica negli anni a venire, ovvero il fatto che *"i sistemi informativi non servono più alle amministrazioni per fatti di gestione interna, ma servono proprio per amministrare, si proiettano cioè sempre più verso l'esterno"*⁹⁷. Il digitale, nella visione di Giannini, rappresentava già allora un'ineludibile infrastruttura per garantire il normale funzionamento della PA.

Bisogna aspettare il 2005 e la legge 15/2005⁹⁸ per vedere una significativa spinta in avanti nel processo di digitalizzazione della PA. La Legge citata introduce un importante articolo, l'art 3bis alla L.241/90 sull'uso della telematica che recita *"Per conseguire maggiore efficienza nella loro attività, le amministrazioni pubbliche incentivano l'uso della telematica, nei rapporti interni, tra le diverse amministrazioni e tra queste e i privati"*. Questo articolo connette per la prima volta il principio di efficienza della PA all'incentivazione dell'uso della telematica nei rapporti interni tra le diverse amministrazioni e tra queste e i privati. Si tratta sicuramente di un momento significativo, ma il vero passaggio evolutivo importante nel percorso verso una

⁹⁷De Ruvo V, Il decreto delegato sui sistemi informativi automatizzati delle amministrazioni pubbliche, Informatica e diritto, XX annata, Vol. III, 1994, n. 1, pp. 121-131

⁹⁸ Modifiche ed integrazioni alla legge 7 agosto 1990, n. 241, concernenti norme generali sull'azione amministrativa. La legge 241/1990 è uno dei capisaldi nel processo di innovazione della PA. Essa ha introdotto nella PA i principi del *New Public Management*: efficacia, efficienza, economicità, pubblicità e trasparenza

digitalizzazione consapevole arriva con l'adozione del Codice per l'Amministrazione digitale (d'ora in poi CAD) che collega alla digitalizzazione tutti gli aspetti legati all'impianto organizzativo delle singole amministrazioni. L'art. 2 alla voce "Finalità e campo di applicazione" recita *"Lo Stato, le Regioni e le autonomie locali assicurano la disponibilità, la gestione, l'accesso, la trasmissione, la conservazione e la fruibilità dell'informazione in modalità digitale e si organizzano ed agiscono a tale fine utilizzando con le modalità più appropriate e nel modo più adeguato al soddisfacimento degli interessi degli utenti le tecnologie dell'informazione e della comunicazione"*.

L'espressione *"...si organizzano per agire"* e la conseguente forte connessione con gli aspetti organizzativi, come vedremo nel prosieguo del lavoro è fondamentale per il nostro caso di studio e ci guiderà per analizzare il percorso evolutivo intrapreso dall'Istat verso il sistema di produzione delle *Trusted Smart Statistics*. Più avanti sarà introdotto anche il concetto di sostenibilità organizzativa legata al nuovo sistema di produzione per sostenere la centralità data alla dimensione organizzativa.

All'interno della PA e quindi anche degli Istituti Nazionali di Statistica, la digitalizzazione rappresenta il processo con il quale gli Istituti agiscono e si organizzano per agire al fine di esercitare le proprie funzioni e prestare servizi pubblici (che nel caso dell'Istat coincidono con il rilascio e la diffusione di statistiche ufficiali) tramite ICT, ovvero con le modalità rese possibili dalle tecnologie dell'informazione e delle comunicazioni e dalle nuove fonti dati. Accanto alle loro attività tradizionali quindi l'Istat, così come le altre PA, può affiancare, sempre facendo leva sull'ICT, le cosiddette *born-digital functions*⁹⁹, nuovi processi, nuove attività resi possibili dall'uso più invasivo di strumenti tecnologici e dall'integrazione, nei processi di produzione, delle nuove fonti dati ampiamente descritte nel Capitolo 1.

⁹⁹Rossa, S.. The digitalisation of public administration and born-digital functions: a modern "Janus Bifrons"? *Smart Governments, Regions and Cities*, 207, 2020

3.2 *Born digital functions in Istat: le Trusted Smart Statistics*

Con l'avvento della digitalizzazione, l'attività conoscitiva "tradizionale" dell'Istat, viene fortemente innovata e rafforzata grazie alle tecnologie digitali. L'uso di nuovi strumenti tecnologici e di nuove analisi, l'introduzione di strumenti di Intelligenza Artificiale (pensiamo alla *Big Data Analysis*, alle tecniche di *machine learning*, di *deep learning*...) dimostrano che il processo di digitalizzazione per gli INS diventa abilitante per consentire loro di agire e di organizzarsi per continuare a rilasciare statistiche tradizionali e affiancare a queste, nuovi *output* innovativi al fine di offrire informazione statistica *potenziata*. Continuità e innovazione vanno di pari passo nella produzione di statistica. La prima garantisce la comparabilità dei fenomeni e una base informativa stabile nel tempo; la seconda rende conto delle trasformazioni e delle nuove tendenze in atto e garantisce il potenziamento dell'offerta informative attraverso nuovi prodotti tra cui emergono sicuramente le *Trusted Smart Statistics* (d'ora in poi TSS nel testo).

Le TSS indicano infatti l'apertura della Statistica ufficiale a nuovi tipi di dati, fonti e tecnologie per la produzione statistica. La rete del Sistema Statistico Europeo da tempo sta investendo in questa direzione, realizzando diversi progetti sperimentali quali: *ESSnet Big Data I*, *ESSnet Big Data II*, *ESSnet Towards Trusted Smart Statistics*, *ESSnet Smart Surveys*. Un progetto ESSnet è: "Una rete di diverse organizzazioni dell'SSE volta a fornire risultati che saranno utili all'intero SSE". I progetti ESSnet devono servire gli interessi dell'intero Sistema Statistico Europeo (d'ora in poi SSE nel testo) e devono essere in linea con il programma statistico quinquennale, finanziato da Eurostat. L'obiettivo è sfruttare le sinergie derivanti dalla cooperazione di alcuni Stati membri per condividere le competenze e risparmiare sui costi nella risoluzione di problemi comuni di interesse europeo. E' evidente che l'uso delle nuove fonti di dati è ormai da molti anni al centro dell'agenda europea del network d'istituti nazionali di statistica coordinato da Eurostat e ha richiesto in tutti gli istituti nazionali di statistica una fase di sperimentazione e innovazione per lo studio di metodologie adeguate a sfruttarne le potenzialità, tenendo adeguatamente conto dei problemi di rispetto della *privacy*.

All'interno del Sistema statistico Europeo a seguito del *Memorandum di Scheveningen* (cfr Cap. 2) e del conseguente documento di pianificazione *Big Data Action Plan and Roadmap*, un elevato numero di INS ha intrapreso iniziative relative allo studio dei vantaggi e delle sfide dell'utilizzo di nuove fonti di dati e di tecniche di data science nelle statistiche ufficiali. Durante il periodo coperto dal programma statistico europeo 2014-2020, gli INS, tra cui l'Istat, hanno cooperato, all'interno degli ESSnet Big data I e II, per condurre progetti, sviluppare conoscenze e capacità tecniche, avviare piani di attuazione, produrre statistiche sperimentali e sviluppare strategie nazionali. Eurostat ha contribuito attivamente a queste attività sia nella fase di progettazione che in quella di esecuzione nel quadro dell'innovazione nelle statistiche ufficiali. Nell'ottobre 2018, è stato adottato il Bucharest Memorandum on "*Official Statistics in a datafied society (Trusted Smart Statistics)*" che indica nel dettaglio gli aspetti definatori e il percorso europeo verso le Trusted Smart Statistics. Il Memorandum di Bucharest ha quindi contribuito a valorizzare e a formalizzare il contributo dei big data in termini di validità, accuratezza e affidabilità degli *output*.

L'espressione *Trusted Smart Statistics* è stata proposta da Eurostat per rappresentare l'evoluzione dalle statistiche tradizionali ed è stato adottato ufficialmente dal Sistema statistico europeo proprio nel memorandum di Bucharest durante la 104° conferenza dei DGINS (*Directors General of the National Statistical Institutes*). Il termine fa riferimento a sistemi di produzione statistica multi-fonte e multi-*output* che usano tecnologie innovative volte integrare in modo flessibile le nuove fonti big data nella produzione statistica. L'affidabilità delle statistiche che è richiamata nella denominazione *trusted* è strettamente legata alla affidabilità dell'istituzione produttrice e si basa su criteri che vanno da rispetto delle norme che presidiano i trattamenti dati e la *privacy*, alle infrastrutture che permettono il trattamento dei dati, alle caratteristiche metodologiche e alle garanzie di qualità dei processi di elaborazione. Occorre non perdere di vista il fatto che i big data non rientrano naturalmente nel quadro di qualità stabilito delle statistiche ufficiali perché il meccanismo di generazione dei dati delle sorgenti di big data non ricade sotto il controllo degli INS ed è in genere sconosciuto. Ciò è molto diverso da quanto accade con le indagini statistiche, il cui meccanismo di generazione dei dati è progettato all'interno degli istituti stessi attraverso il

campionamento probabilistico, ma anche con le fonti di dati amministrativi, il cui meccanismo di generazione dei dati è quanto meno noto agli INS. Il richiamo ad un concetto così importante come la *trust* è dovuto ai cambiamenti “*di significato e di senso*”¹⁰⁰ introdotti dalla trasformazione di cui si è parlato. Storicamente infatti, gli INS hanno sempre avuto il controllo dell’intero processo di produzione statistica perché avveniva al loro interno, dalla raccolta diretta dei dati dei rispondenti sino alla diffusione delle statistiche prodotte, la gestione interna dei processi ha permesso agli INS di poter garantire la veridicità, la qualità, la robustezza dei dati raccolti, delle metodologie applicate e di tutti gli standard necessari affinché le statistiche prodotte potessero esser definite ufficiali. Da quando si producono statistiche con le nuove fonti dati, si usano dati raccolti e detenuti da soggetti esterni agli INS (e.g. operatori di telefonia mobile...). È però necessario continuare a mantenere gli stessi livelli di qualità e le stesse caratteristiche che consentono di poter assicurare l’ufficialità delle statistiche prodotte e la fiducia che gli utenti hanno rispetto al ruolo Istituzionale degli INS e alle statistiche da loro rilasciate. In un contesto in cui aumenta la quantità di informazioni a disposizione degli utenti, è solo il riconoscimento della qualità dei dati e del ruolo istituzionale di chi li produce che può consentire agli utilizzatori e ai *decision makers* di orientarsi in un ecosistema dell’informazione che diventa sempre più affollato.

Perché venga mantenuto questo livello di *trust*, è indispensabile un adeguamento dell’intero processo di produzione statistica per assicurare il controllo su quei processi che avvengono al di fuori degli INS e garantire gli stessi livelli di qualità e rilevanza. La disponibilità di nuovi strumenti metodologici e tecnologici e di nuove fonti rappresenta una straordinaria occasione per la produzione dei dati, trasformando la statistica ufficiale da “prodotto di nicchia” a servizio facilmente accessibile a tutti. Assistiamo ad un vero e proprio cambiamento di paradigma in questo settore, con la presenza nel mercato di nuovi soggetti in grado di elaborare quantità infinite di dati, sono emersi nuovi produttori di dati in grado di competere con gli Istituti di Statistica, almeno in certi ambiti, lo sviluppo e la diffusione dell’ICT, inoltre, hanno abbattuto la spesa per la produzione dell’informazione e altri

¹⁰⁰ Epifani S., Sostenibilità digitale, Digital Institute, 2020

soggetti, pubblici e privati, sono ora in grado di raccogliere, elaborare e comunicare dati statistici come mai in precedenza. La cosiddetta datificazione della società, come già ampiamente descritto nelle pagine precedenti, ha portato anche alla proliferazione dei *data player* sia nel settore pubblico che privato esterni al Sistema della statistica ufficiale. Nel mondo *predatification* gli INS erano i monopolisti esclusivi di un'attività quasi esclusiva, dove l'unica alternativa alle statistiche ufficiali era l'assenza di statistiche. Oggi, gli Istituti di Statistica sono uno dei produttori di dati all'interno del complesso ecosistema informativo. Diversi attori del settore privato e pubblico stanno producendo nuovi dati e offrendo punti di vista alternativi su fenomeni emergenti. La statistica ufficiale, nella competizione con altri produttori di dati, deve riuscire a mantenere integro il suo ruolo istituzionale e continuare a produrre statistiche ufficiali anche con l'ausilio dei nuovi dati, garantendo gli stessi livelli di qualità, credibilità, precisione e affidabilità. In altre parole, gli INS devono affrontare una duplice sfida: sfruttare l'enorme disponibilità di dati prodotti e raccolti esternamente e competere con attori nuovi, con nuovi produttori di statistiche mantenendo lo stesso elevato grado di qualità delle informazioni statistiche prodotte e rilevanza del proprio ruolo. Per fare questo gli INS devono intrecciare i big data con la loro cultura organizzativa. I big data sono *qui per restare*¹⁰¹ perciò non si può non cogliere la sfida e le opportunità che offrono e le TSS rappresentano il primo risultato di questo intreccio tra nuove sorgenti di dati e cultura organizzativa. Le TSS rappresentano di fatto la risposta degli Istituti di statistica ai cambiamenti in corso dentro e fuori gli Istituti stessi, sono una delle *born digital function* realizzabili grazie alla trasformazione digitale. Le TSS sono prodotti con cui gli Istituti continuano a realizzare la loro missione tradizionale ma con uno sguardo proiettato al futuro, rispondono meglio a bisogni conoscitivi vecchi ed emergenti e a *stakeholders* tradizionali e nuovi.

Le TSS rappresentano di fatto un cambiamento culturale, una nuova visione, sono il risultato dell'introduzione dell'innovazione nei processi, rappresentano l'attuazione e

¹⁰¹ Vincent JL, Creteur J. Big data are here to stay! *Anaesth Crit Care Pain Med.* 2019 Aug;38(4):339-340. doi: 10.1016/j.accpm.2019.05.001. Epub 2019 May 9. PMID: 31078797

l'istituzionalizzazione di nuovi processi e prodotti resi possibili da tutto ciò che è accaduto fuori dagli Istituti stessi, e per questo necessitano dell'assunzione di paradigmi innovativi per essere accettati allo stesso modo di come è accettata la statistica ufficiale. Il concetto di *Trusted Smart Statistics* implica alcune modifiche sostanziali al paradigma tradizionale di produzione della Statistica ufficiale dovute principalmente al fatto che sia le fonti dei dati, sia le procedure di elaborazione possono essere esterne agli Istituti di statistica. Questo passaggio, unito alla natura privata delle fonti di dati, può mettere in discussione il tradizionale sistema di fiducia che sta alla base del modello di produzione delle statistiche ufficiali. Quindi, siccome i principi e gli obiettivi della statistica ufficiale rimangono gli stessi, definiti in appositi codici della Statistica Ufficiale sia italiani sia europei, è necessario un diverso insieme di strumenti tecnici, organizzativi e legali per garantire il rispetto dei principi. Per arrivare a raggiungere questi risultati, gli INS devono prevedere un adeguamento dei propri paradigmi organizzativi, metodologici e di ricerca che favorisca la produzione di *Trusted Smart Statistics*.

Nelle pagine precedenti abbiamo già visto come gli ultimi decenni siano stati caratterizzati da profonde trasformazioni che hanno determinato rilevanti cambiamenti grazie alla crescente disponibilità e interazione tra innovazioni tecnologiche straordinarie. La digitalizzazione ha dato un forte impulso alla produzione di dati e al processo di datificazione della società. La diffusione di dispositivi *smart* in molti ambiti della vita quotidiana ha portato a generare dati sempre più granulari dal punto di vista territoriale e temporale e che rappresentano fonti sempre più appetibili per soggetti pubblici e privati.

La trasformazione di senso¹⁰² di cui abbiamo ampiamente parlato nel primo capitolo, causata dalla rivoluzione digitale e riscontrabile in vari contesti sociali politici ed economici ha provocato anche negli Istituti di statistica profondi cambiamenti di processo che richiedono un ripensamento, una ridefinizione radicale di concetti e modalità di funzionamento.

¹⁰² Epifani S., Sostenibilità digitale, Digital Institute, 2020

Il mondo degli INS è erroneamente percepito come statico, chiuso nelle proprie regole, nei propri *framework* caratterizzati da solidi criteri di qualità per diffondere i dati in differenti domini. Non è così, cambia il mondo e cambia anche il modo di fare statistica. Si è passati da un mondo *pre-datification* in cui gli sforzi degli INS si concentravano sulla raccolta dati in particolare su indagini e censimenti che rappresentavano la maggiore fonte di dati, ad un mondo caratterizzato da enormi giacimenti di dati in cui le autorità statistiche devono *distillare* dati affidabili da molte fonti.

Il nuovo modello della statistica europea vede un rafforzamento della logica d'integrazione delle informazioni prodotte e un forte uso dei registri statistici e dei big data per fornire nuovi e più approfonditi strumenti anche a sostegno delle politiche d'intervento: un approccio olistico finalizzato a guadagnare in qualità ed efficienza.

Le nuove fonti dati permettono di produrre nuovi *output* per supportare per esempio più efficacemente i decisori pubblici: il carattere della tempestività e della granularità temporale e spaziale delle TSS può consentire ai *policy makers* di prendere decisioni basate su dati molto più ricchi rispetto a quelli prodotti finora dalle sole statistiche tradizionali. I cosiddetti *real world data* sono capaci di assicurare maggiore tempestività informativa e offrire un significativo apporto alla qualità dell'informazione statistica in termini di accuratezza, danno ai decisori pubblici la possibilità di ancorare le proprie scelte sulla base di dati più aggiornati e aderenti alla realtà fattuale, garantiscono loro un accesso più tempestivo ai dati in diversi settori.

Per i *decision makers* poter usare nei loro processi decisionali i nuovi prodotti statistici è un'opportunità estremamente significativa proprio per la *funzione conoscitiva potenziata* a cui questi nuovi *output* assolvono. Le TSS offrono infatti una conoscenza su diversi ambiti sociali a un livello territoriale tale da avvicinare le istituzioni ai cittadini. I nuovi prodotti statistici, pur non raggiungendo il livello di solidità metodologica tipico della statistica ufficiale, permettono la conoscenza di elementi relativi a un territorio che possono aiutare sicuramente i *policy makers* a governare meglio, a rispondere in maniera più efficace alle esigenze dei

cittadini e a rafforzare quel legame tra istituzioni e società civile che negli ultimi tempi si è via via deteriorato.

La disponibilità di statistiche *real time* e comparabili a livello europeo è diventata cruciale per consentire ai *policy maker* di prendere decisioni e sostenere misure correttive efficaci e tempestive, come avvalorato dalla recente crisi pandemica da Covid-19, la quale ha dimostrato l'utilità di queste fonti di dati anche come strumento di contrasto al diffondersi di *fake news* e della disinformazione.

3.3 Il carattere antropocentrico delle TSS

Le TSS, in quanto prodotti che si basano sull'uso di dati prodotti in gran parte dai cittadini in maniera anche inconsapevole attraverso le loro pratiche quotidiane ridanno un ruolo centrale al cittadino. Questi nuovi prodotti possono essere considerati la proiezione del concetto di PA come "arena aperta"¹⁰³ proposto dal giurista e accademico italiano Sabino Cassese. Il concetto di "arena aperta" ridà un ruolo dinamico al cittadino nel suo rapporto con la PA e che rimanda ad un approccio non statico, ma dinamico e dialogativo tra cittadino e Stato.

Nell'arena aperta si supera il paradigma bipolare e statico e si riscontra una forte e crescente apertura delle amministrazioni verso i cittadini che sono chiamati a partecipare e a collaborare ai processi delle PA. La partecipazione, la trasparenza e la collaborazione sono i principi chiave dell'*Open Government*, che permette di realizzare un inedito rapporto di fiducia tra la collettività e la PA, di consolidare un processo di avvicinamento e una relazione aperta, incentrata su innovative forme di interazione *bidirezionali* tra cittadino e Stato. L'integrazione delle nuove fonti di dati nei processi di produzione in Istat e negli Istituti di

¹⁰³ Cassese, S. *L'arena pubblica. Nuovi paradigmi per lo Stato*. Rivista trimestrale di diritto pubblico Numero 3, 2001

statistica offre una grande opportunità di rafforzare questa *bidirezionalità* e di avvicinare l'Istat al cittadino e il cittadino all'informazione statistica.

I cittadini da semplici fruitori di dati e informazioni prodotte verticalmente dagli INS diventano, attraverso l'uso *online* e *offline* dei loro *devices* produttori di dati in un'ottica di scambio bidirezionale e di *citizen science*. La *citizen science* considera i cittadini anche come parte attiva nella produzione dei dati (i cosiddetti *citizen data*) e valorizza il loro coinvolgimento nei processi di ricerca scientifica e di produzione di informazione e conoscenza. La *citizen science* riguarda così qualsiasi attività che coinvolge i cittadini nella ricerca scientifica e quindi ha il potenziale per riunire la scienza, i decisori politici e la società nel suo insieme in modo efficace: le TSS riassumono tutto questo. Gli studi più recenti sulla comunicazione strategica come strumento per rafforzare la reputazione delle istituzioni sostengono che “...l'evidenza empirica suggerisce che la partecipazione alla vita civile e sociale è associata a livelli di fiducia più elevate nelle istituzioni”¹⁰⁴. Il futuro delle statistiche ufficiali non dipende solo dal lavoro con le nuove tecnologie digitali, dalle fonti di dati e dall'invenzione di metodi, ma anche da un ripensamento delle relazioni con i cittadini dall'instaurazione di nuove interazioni con i cittadini stessi¹⁰⁵ che permettano di alimentare quella *social trust* che, in sociologia, è potenza collante dei gruppi e motore invisibile del progresso sociale. In generale è molto difficile spiegare alle persone che potrebbe esserci una discrepanza tra la loro esperienza personale e ciò che mostrano le statistiche ufficiali. Per arrivare alla *human story* di ciascuno dei cittadini, per far sentire i dati più vicini, potrebbero essere utili dati più dettagliati, suddivisi per area geografica o sottogruppi, per questo le nuove fonti e di conseguenza le TSS rappresentano un'opportunità imperdibile per tutti gli Istituti di statistica ¹⁰⁶.

¹⁰⁴ Hassain, J. *Disinformation in Democracies: Improving Societal Resilience to Disinformation* (2022). Riga: NATO Strategic Communications Centre of Excellence

¹⁰⁵ Ruppert, E. (2018). *Sociotechnical Imaginaries of Different Data Futures: An Experiment in Citizen Data*. 3e Van Doornlezing. Rotterdam, NL: Erasmus School of Behavioural and Social Sciences.

<https://www.eur.nl/sites/corporate/files/2018-06/3e%20van%20doornlezing%20evelyn%20ruppert.pdf>

¹⁰⁶ Porter T., *Statistics in the digital era*, in *Power from Statistics: data, information and knowledge*, pag.12, Eurostat 2018

Queste nuove relazioni prevedono la co-produzione di dati – i *citizen data* - in cui i cittadini sono di fatto impegnati nella produzione statistica, possono fornire informazioni utili alla produzione statistica con minor disturbo e maggiore accuratezza. I cittadini possono partecipare dare un concreto contributo a specifiche fasi del processo di produzione dell'informazione statistica, nell'era pre-datification o nell'era dei prodotti statistici "tradizionali" i cittadini erano solo rispondenti, fornivano risposte a questionari progettati a monte dagli INS. Nel mondo digitale i cittadini da semplici rispondenti diventano coproduttori di dati anche inconsapevolmente attraverso le nuove pratiche quotidiane scaturite dalla rivoluzione digitale: *data from society e data for society*. Per suggellare questa collaborazione, gli istituti di statistica dovrebbero definire un patto sociale con i cittadini che permetta loro di prendere i dati dai cittadini per restituirli loro sotto forma di informazione utile un patto che includa i cittadini nel processo statistico e definisca la finalità della raccolta dati e i benefici che i fornitori dei dati possano ricavarne, un patto fondato su regole definite che metta il cittadino al centro e ne rispetti i diritti, prima tra tutti quello alla *privacy*. Il coinvolgimento dei cittadini è fondamentale per creare fiducia negli istituti di statistica e nella statistica ufficiale, ed è proprio nel coinvolgimento dei cittadini che si concretizzano i principi chiave dell'*Open Government* già citati partecipazione, collaborazione e trasparenza. Con le TSS Si sostanziano quindi nuove modalità di partecipazione, collaborazione-sussidiarietà orizzontale che producono un duplice vantaggio: il coinvolgimento del cittadino diminuisce il contenzioso, dall'altro le PA, coinvolgendo i cittadini nelle loro pratiche correnti, aumentano il proprio consenso fra i cittadini, aumentano cioè l'*accountability*, la *trust* e rafforzano il proprio ruolo istituzionale.

I dati prodotti dai cittadini diventano lo strumento a disposizione dei cittadini stessi per una reale collaborazione con la pubblica amministrazione. Le *smart surveys*, le TSS rappresentano l'evoluzione delle statistiche tradizionali anche da un punto di vista sociale perché rappresentano la concretizzazione di quell'umanesimo digitale di cui abbiamo parlato nel Capitolo 1: cambia il ruolo del cittadino, che diventa utente, *prosumer*, partecipante attivo e collaboratore della PA e strumento attraverso cui si concretizzano gli obiettivi dell'*Open*

Gouvernement. Le TSS diventano il prodotto di uno scambio basato sulla fiducia tra cittadino e Istituto di Statistica.

3.4 Il sostegno alle nuove iniziative: verso il sistema di produzione delle TSS.

Dopo l'adozione del Memorandum di Bucharest è iniziata una riflessione in Istat su come governare questo processo di innovazione. Il rilascio dei primi *output* con l'uso dei big data, le statistiche sperimentali, il confronto tra i diversi INS nel Sistema statistico europeo all'interno dei progetti EssNet, ha permesso di maturare la consapevolezza che l'uso delle nuove fonti di dati, non richiede soltanto capacità strettamente tecniche e infrastrutture informatiche più potenti, ma richiede investimenti nei diversi settori di cui si compongono le singole organizzazioni. Si è consolidato a livello europeo il convincimento che l'evoluzione della statistica ufficiale verso analisi più complete, celeri, trasversali ma allo stesso tempo accurate e solide, possa essere realizzata non solo attraverso l'uso delle fonti amministrative ma anche dei dati generati da fonti di varia natura, offerti da diversi soggetti che concorrono alla fornitura di servizi digitali e di telecomunicazione, ai quali si dovrà affidare un ruolo complementare rispetto alle tradizionali fonti utilizzate. Poiché nella maggior parte dei Paesi le nuove fonti di dati integreranno, ma non sostituiranno, quelle preesistenti, le componenti consolidate dei sistemi statistici non saranno eliminate, ma piuttosto incrementate da quelle nuove. In altre parole, il percorso evolutivo del sistema statistico assumerà la forma di un *incremento sistemico*. Per analogia, possiamo pensare alle fonti di dati come a un carburante e al sistema statistico come a un motore: il nuovo carburante non può essere alimentato nel motore preesistente e i sistemi statistici devono sviluppare un nuovo tipo di motore, con principi operativi diversi da quelli preesistenti, adattati alle peculiarità del nuovo *data fuel*. Il futuro sistema statistico sarà una macchina multicarburante con due motori, le TSS rappresentano la modalità con cui il SSE chiama il nuovo motore che deve essere sviluppato.

Le TSS, e tutti i processi di produzione e a supporto necessari a realizzarle, all'interno degli Istituti di statistica rappresentano quindi un cambiamento culturale, perciò necessitano di paradigmi innovativi, di capacità di visione a lungo termine e di nuovi approcci per essere accettati e inglobati all'interno dei meccanismi di funzionamento degli Istituti stessi. Il processo di progettazione del TSS deve partire dalla considerazione che le condizioni operative attorno alle statistiche ufficiali sono cambiate. Come in altri sistemi socio-tecnologici, il cambiamento non riguarda solo l'*hardware* e livelli *software*, ma anche livello umano (processi, organizzazione, normative e pratiche). La disponibilità di nuove fonti di dati digitali fa parte di un cambiamento più ampio nel modo in cui le persone percepiscono, usano e, in generale, si relazionano ai dati e alle tecnologie digitali. Ciò ha un impatto su ciò che i cittadini e gli utenti si aspettano dalle statistiche ufficiali. In altre parole, la continua evoluzione delle statistiche ufficiali non comporta solo nuove opportunità tecniche (ciò che le statistiche ufficiali possono fare ora che prima non era possibile) ma anche nuove opportunità sociali, aspettative e doveri (quello che le stesse statistiche dovrebbero fare ora che prima non era dovuto) che, più o meno direttamente, comprendono non solo quali statistiche devono essere fornite, ma anche come vengono prodotte¹⁰⁷.

L'integrazione delle nuove fonti ha quindi forti ricadute su vari aspetti sia metodologici, sia organizzativi sia giuridici. Nuove fonti di dati richiedono nuovi approcci metodologici (i) per trasformare i dati grezzi in informazioni e concetti statistici, (ii) per usare dati che non sono stati progettati e raccolti con fini statistici, (iii) per superare i problemi di selettività, (iv) per integrare le nuove fonti di dati con quelle tradizionali.

In un sistema di produzione di *Trusted Smart Statistics* l'elemento generale di qualità dei processi di produzione statistica ha una ineludibile centralità, ma deve essere modificato perché occorre condividere con i titolari dei dati l'esecuzione del trattamento dei dati per gli scopi concordati, con algoritmi concordati, programmi affidabili e tutelando la *privacy* con

¹⁰⁷ Ricciato, F., Wirthmann, A., & Hahn, M., *Trusted Smart Statistics: How new data will change official statistics*. Data & Policy, 2020

opportuni metodi di calcolo. Tenendo conto dell'ampia gamma di parti interessate, l'adozione di processi completamente trasparenti, l'uso di serie comuni di strumenti di garanzia della qualità e l'efficace comunicazione al pubblico sono essenziali per dimostrare l'impegno della statistica ufficiale a continuare a essere una fonte affidabile di statistiche.

Eurostat riassume le principali proprietà delle TSS nei seguenti criteri¹⁰⁸.

- ✓ Fonti di dati multiuso - Per utilizzare in modo efficace queste nuove fonti di dati e gli enormi investimenti associati al loro sfruttamento, il loro sviluppo deve mirare a una progettazione multiuso per promuovere usi multipli e garantire la scalabilità (orizzontale) degli approcci. Pertanto, invece di perseguire la raccolta, l'elaborazione e l'analisi dei dati specifici del dominio, nuove fonti di dati sono (pre) elaborate e trasformate in dati intermedi da utilizzarsi (possibilmente in combinazione con altri) per derivare indicatori statistici finali per diversi domini applicativi;
- ✓ Statistiche multi-fonte - Nuovi dati statistici e indicatori devono essere sviluppati sulla base dell'integrazione di più fonti di dati, comprese le combinazioni di fonti di dati tradizionali (dati di indagine, dati amministrativi);
- ✓ Organizzazione multilivello del flusso di lavoro dell'elaborazione dei dati - I nuovi dati sono spesso generati come sottoprodotto di processi ad alta intensità tecnologica, lo sviluppo di nuovi approcci metodologici richiede il contributo di esperti provenienti da discipline che sono al di fuori del campo di conoscenza tradizionale della Statistica ufficiale (ad esempio, *Data scientist*, esperti di elaborazione di immagini).
- ✓ Quadri metodologici modulari - Essendo spesso i nuovi dati un sottoprodotto di processi tecnologici soggetti a modifiche dovute all'evoluzione delle tecnologie e / o dell'uso, si possono determinare rotture di serie o modifiche di formati e delle strutture di dati generati da diverse fonti e / o in diversi paesi. Questi aspetti pongono

¹⁰⁸Eurostat Item 1b. Beyond ESS Vision 2020: Innovation actions implementing the multiannual statistical programme 2021-2027, 39th Meeting of the European Statistical System Committee, 7 February 2019

ulteriori sfide per lo sviluppo di metodologie di elaborazione. La chiave per affrontare tali sfide è sviluppare quadri metodologici altamente modulari, in cui ogni modulo può essere evoluto o sostituito senza richiedere modifiche al resto del flusso di lavoro di elaborazione.

- ✓ Possibilità di utilizzo dei dati anche senza piena condivisione dei dati - Quando i dati sono riservati per l'azienda, altamente sensibili alla *privacy* o quando il volume e la velocità dei dati di *input* sono molto grandi, il focus del sistema statistico dovrebbe essere l'estrazione delle informazioni di *output* desiderate, ovvero gli indicatori statistici finali. L'acquisizione dei dati di *input* dovrebbe essere vista semplicemente come un compito ausiliario non sempre necessario. In alcune applicazioni, il trattamento dei dati di *input* grezzi può essere lasciato alle aziende che li producono, specie quando riguardano dati personali o ogni volta che tale approccio è conveniente dal punto di vista dell'ottimizzazione delle risorse. In questi casi, tutti i moduli di elaborazione, compresi quelli eseguiti presso la sede di origine, devono essere (co) sviluppati, o almeno rivisti e accettati, dall'Istituto di statistica.

Dalle caratteristiche sopra indicate emergono anche i ruoli delle strutture organizzative coinvolte. In un quadro così complesso, per sostenere la fiducia nei nuovi prodotti statistici occorre sviluppare un insieme coerente di disposizioni legali, organizzative, metodologiche e tecniche al fine di garantire un elevato livello di affidabilità e qualità durante l'intero processo. Mentre il codice delle statistiche europee e il codice deontologico nazionale sono i riferimenti insieme ai principi di gestione della qualità per garantire le condizioni necessarie allo sviluppo di statistiche affidabili, in una situazione di questo tipo, il ripensamento della struttura organizzativa che si rende necessario e a cui abbiamo fatto riferimento prima, deve rifarsi al principio di *sostenibilità organizzativa*. Come sappiamo sostenibilità vuol dire soddisfare i bisogni della generazione presente senza compromettere quelli della generazione futura. La sostenibilità organizzativa implica quindi una necessaria reingegnerizzazione dei processi organizzativi e trasversali, indispensabile per sostenere il percorso degli altri INS verso il Sistema di produzione TSS, e un riadeguamento di tutte le

componenti organizzative ai nuovi paradigmi di funzionamento, senza compromettere la qualità, la rilevanza e la robustezza degli *output* prodotti e il normale funzionamento dei singoli Istituti.

In una struttura complessa così com'è quella degli Istituti di Statistica e nel nostro caso dell'Istat, che garantiscono il rilascio e la diffusione di statistica ufficiale a sostegno di varie tipologie di *stakeholders*, è necessario riorganizzare i processi di ogni singola dimensione organizzativa dell'Istituto: il cambiamento di una leva necessariamente ha impatto su tutte le altre dimensioni coinvolte. La complessità dell'adeguamento organizzativo per garantire l'implementazione del nuovo sistema di produzione consiste proprio nel far evolvere in maniera coerente e coordinata tutti le componenti e orientarne gli assestamenti e gli investimenti verso un obiettivo comune. È necessario adottare un approccio sistemico e olistico che faccia muovere tutte le dimensioni coinvolte nella medesima direzione, agire su una singola leva con operazioni che impattano su una sola componente non permette di raggiungere l'obiettivo. È solo l'approccio multidisciplinare e interdisciplinare che può veramente spingere l'Istituto verso il nuovo Sistema di produzione. La proiezione organizzativa del concetto di sostenibilità organizzativa in Istat è rappresentato dal Centro per le TSS.

3.5 Il Centro per le TSS in Istat

In Istat l'evoluzione verso il sistema di produzione delle TSS è stato lungo e articolato e sempre coerente con la programmazione Eurostat. Nella figura sottostante sono riportati i passaggi fondamentali compiuti dall'Istat in concomitanza con quelli compiuti in ambito Europeo.

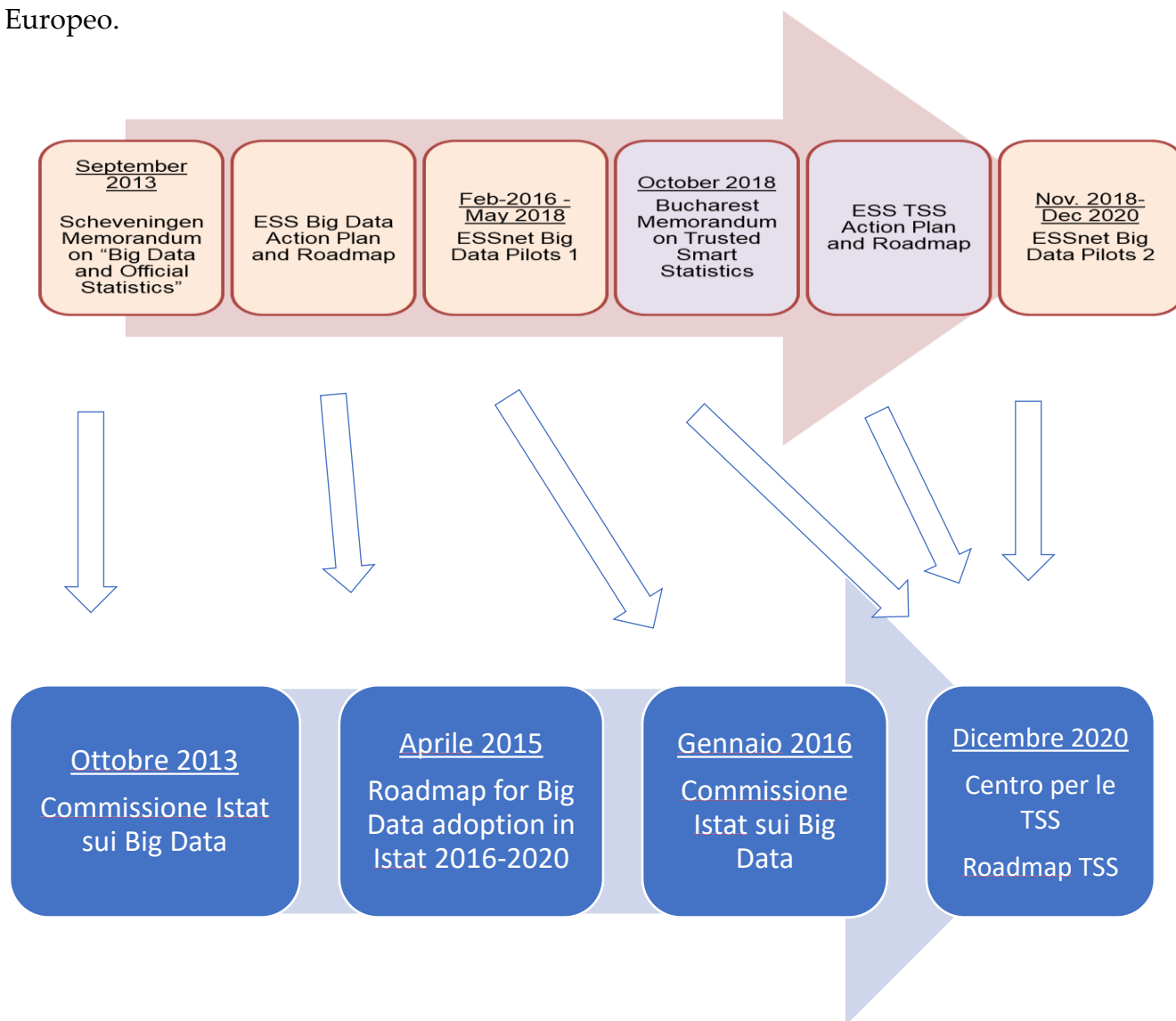


Figura 3 - Le tappe del percorso verso le TSS¹⁰⁹

Dopo l'adozione a ottobre 2018 da parte dei Presidenti degli INS europei nell'ambito del DGINS del *Bucharest Memorandum su Trusted Smart Statistics*, è iniziata una riflessione in Istat

¹⁰⁹ Scannapieco M, Righi A, Presentazione alla XIV Conferenza Nazionale di Statistica, Come cambia la produzione statistica con i Big Data: il Centro e la Roadmap per le TSS, all'interno della sessione Trusted Smart Statistics: sfide e opportunità, 30 novembre 2021

su come governare questo processo d'innovazione. È stata privilegiata una soluzione agile che garantisca una partecipazione interdipartimentale e trasversale perché l'avvio di sperimentazioni sulle nuove fonti non può che essere multidisciplinare. Il modello organizzativo scelto infatti sfrutta al meglio l'interdisciplinarietà delle competenze presenti nell'Istituto e permette di superare la frammentazione organizzativa e di favorire le collaborazioni tra strutture dell'Istituto.

Da un punto di vista meramente organizzativo, l'Istat ha costituito un Centro per le TSS composto da uno *Steering Committee* di cui fanno parte i vertici dell'Istituto e da una segreteria tecnico scientifica che rappresenta un organismo di gestione operativa e di supervisione con compiti specifici di sperimentazione, industrializzazione e messa in produzione di TSS. La Segreteria è il punto di raccordo e monitoraggio di tutte le attività di natura organizzativa, tecnica, metodologica e del passaggio in produzione delle TSS, ha compiti di coordinamento delle attività realizzate nell'ambito delle Aree tematiche del Programma strategico sulle Nuove fonti. Supporta e monitora il processo di Analisi Strategica delle TSS a responsabilità dello *Steering Committee*.

Lo *Steering Committee* è composto dal Direttore Generale, dai Direttori del Dipartimento di produzione e del Dipartimento dei servizi trasversali per lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, dai Direttori Centrali dei servizi metodologici, dei servizi informatici e della Raccolta Dati.

Il Centro per le TSS è stato istituito il 14 dicembre 2020 con una delibera a firma del Presidente. La delibera sintetizza e riporta le esperienze a livello internazionale, europeo e nazionale sul fronte integrazione dei big data nella statistica ufficiale. Di seguito i riferimenti più significativi che sintetizzano anche quanto contenuto nelle pagine precedenti e che sintetizzano le motivazioni alla base della costituzione del Centro: (i) *“la rilevanza per la statistica ufficiale del tema dell'utilizzo delle fonti di Big Data come riconosciuto a livello internazionale sia dall'High Level Group for the Modernization of Official Statistics dell'UNECE, sia dallo Scheveningen Memorandum on Big data and Official Statistics dell'European Statistical System”*, (ii) *“l'adozione da parte dei Presidenti degli Istituti di statistica europei e delle*

organizzazioni statistiche nell'ambito del DGINS a ottobre 2018 del Bucharest Memorandum su Trusted Smart Statistics; (iii) "l'avvio della pubblicazione di Statistiche sperimentali realizzate a partire da Big Data sul sito istituzionale dal giugno 2018"; "l'inquadramento delle fonti Big Data a supporto della produzione di statistica ufficiale nell'ambito del Programma Strategico dell'Istat PG3 – Nuove fonti".

Il Centro è un'organizzazione strategica flessibile in grado di intercettare le esigenze e aderire facilmente ai cambiamenti in atto, sia dal punto di vista conoscitivo che tecnologico, visti i repentini cambiamenti determinati dalla crisi pandemica nella società e nell'economia. Il Centro ha il compito, così come espressamente previsto dalla delibera di costituzione, di favorire la creazione di un nuovo sistema di produzione per le TSS che sia completamente integrato con l'infrastruttura di acquisizione, produzione e diffusione di dati dell'Istituto, ed in particolare con il Sistema Integrato dei Registri e le indagini correnti. Opera, inoltre, per fornire soluzioni di calcolo in un perimetro istituzionale più esteso, che include le PA nell'ottica di una Strategia Dati Nazionale. Il Centro recepisce e valorizza le esperienze internazionali realizzate in ambito Eurostat e UNECE, e si pone come strumento operativo per realizzare e infrastrutturare le azioni sul fronte Big Data e/o TSS che fanno parte dello *European Statistical Program 2021-2027*¹¹⁰. Predisporre e monitora la *Roadmap* triennale per le TSS e fornisce indicatori di monitoraggio per la Pianificazione Strategica e Operativa.

Le motivazioni che stanno alla base della costituzione del Centro, nonché le attività ad esso assegnate, sono assolutamente coerenti con le evoluzioni del contesto esterno e con la programmazione prevista in ambito Eurostat per il 2020-2027¹¹¹ e UNECE che prevede di fatto di valorizzare e capitalizzare i risultati raggiunti con i progetti ESSnet Big Data I e II. L'Istat dal 2013 ha assunto un ruolo di punta in Europa per sperimentazione e messa in produzione di nuovi *output*, anche attraverso collaborazioni internazionali, accademiche e *partnership*. L'Istat ora è in una fase di passaggio da una fase di *playground*, necessaria a capire

¹¹⁰ <https://ec.europa.eu/eurostat/web/european-statistical-system/programmes-and-activities/statistical-programmes>

¹¹¹ Eurostat Strategic Plan 2020-2024

le potenzialità e i limiti delle fonti Big Data e i metodi necessari a trattarle, ad una fase di utilizzo maturo di tali fonti con individuazione di nuovi prodotti.

3.5.1. La roadmap

Attraverso un processo partecipativo fra tutte le strutture dell'Istituto, all'interno del Centro TSS è stato redatto un documento strategico chiamato "*Roadmap per la produzione delle Trusted Smart Statistics. Anni 2021-2024*" che guida la realizzazione delle azioni operative necessarie per costruire il percorso dell'Istituto verso il nuovo sistema di produzione. Il documento rappresenta una riflessione strategica sui progetti prioritari da realizzare in ambito TSS e sulle necessarie attività trasversali a supporto dei progetti stessi. La *Roadmap* si sviluppa quindi su due livelli distinti: (i) un livello *strategico* elaborato secondo un approccio *top-down* che ha coinvolto direttamente i Dirigenti delle strutture di produzione i quali, sulla base della programmazione europea contenuta nello *Strategic Plan* di Eurostat 2020-2024¹¹² e sulle base di quanto programmato a livello di Istituto, hanno evidenziato alcuni progetti strategici da realizzare con l'uso di nuove fonti ritenute particolarmente promettenti e con impatto su specifiche tematiche di specifica rilevanza; (ii) un livello *operativo* che ha visto maggiormente coinvolte le strutture trasversali, con un approccio questa volta *bottom up*, che ha fatto emergere quali fossero le azioni a supporto necessarie a implementare i progetti prioritari definiti "dall'alto".

Con un approccio *top down* sono stati quindi individuati nuovi *output*, ai quali sono stati associati obiettivi e azioni da porre in essere per garantirne la realizzazione in materia di:

- ✓ investimenti IT e metodologici
- ✓ processi organizzativi
- ✓ partnership e regole (*privacy...*)

¹¹² https://ec.europa.eu/info/system/files/estat_sp_2020_2024_en.pdf

Le azioni segnalate negli ambiti di cui sopra sono azioni trasversali oggetto di uno specifico monitoraggio, reso possibile grazie all'implementazione di un *framework* dedicato che sarà analizzato nelle pagine seguenti.

La selezione dei progetti prioritari della *Roadmap* ha coinvolto tutto l'Istituto. In prima battuta è stata realizzata una prima ricognizione dei progetti relativi alle TSS in atto_in Istituto e presenti nel Programma Statistico Nazionale¹¹³. Le attività in corso sono state ricondotte a cinque principali filoni per tipologia di fonti (*Web intelligence, Smart systems, Smart surveys, Smart personal data, Web relations*) che i Direttori delle strutture di produzione hanno selezionato e prioritizzato in base a 20 criteri decisionali appartenenti a 4 domini

1. Arricchimento della produzione statistica
 - 1.1. Arricchimento informativo
 - 1.2. Caratteristica della fonte
 - 1.3. Potenzialità per la statistica ufficiale
 - 1.4. Stadio di sviluppo dell'*output*
2. Avanzamenti metodologici e di ricerca
3. Avanzamenti tecnologici
4. Avanzamenti nella raccolta dati

Ognuno di questi macro criteri individuati è stato spaccettato in una serie di caratteristiche e dimensioni legati anche ai vantaggi ottenibili tramite le TSS e più volte emersi anche nel presente lavoro. Tutti questi elementi hanno alimentato una griglia per la cui compilazione si è messo in moto un processo partecipativo che ha richiesto la collaborazione di tutti i settori e che è stata poi validata dai Direttori delle strutture di produzione che sono stati invitati ad attribuire punteggi a ogni dimensione/criterio inserito nella griglia e riportato nella Tabella 3. .

¹¹³ Il Programma statistico nazionale (Psn) è l'atto normativo che, in base all'art. 13 del d.lgs. n. 322 del 1989 e successive integrazioni, stabilisce le rilevazioni statistiche di interesse pubblico affidate al Sistema statistico nazionale e i relativi obiettivi informativi.

Tabella 3 - Domini e criteri di selezione

MACRO-CRITERI per la prioritizzazione delle trusted smart statistics		CRITERI con PUNTEGGI (1- basso, 2- medio, 3- alto, NR- non so /non risponde)
ARRICCHIMENTO DELLA PRODUZIONE STATISTICA	Arricchimento informativo	Fenomeno emergente ad elevata priorità per utenti/policy makers
		Disponibilità di nuove informazioni non desumibili da altre fonti (rilevazioni dirette o archivi amministrativi) già utilizzate da ISTAT
	Caratteristiche della fonte	Granularità dei dati rispetto alle fonti disponibili
		Tempestività dei dati rispetto alle fonti disponibili
		Affidabilità e continuità della fonte
		Grado di copertura della fonte rispetto al fenomeno oggetto di studio
	Potenzialità per la statistica ufficiale	Potenzialità di diventare un prodotto statistico ufficiale rispetto a realizzazione di un prodotto intermedio / avanzamento conoscitivo
		Generazione di Statistiche sperimentali
	Stadio di sviluppo dell'output	Generato Statistica Sperimentale (2) o Produzione Statistica ufficiale (3)
		Progetti a elevato grado di maturità
AVANZAMENTI METODOLOGICI E DI RICERCA		Integrabilità nel SIR (impatto sul sistema di produzione dei registri)
		Integrabilità nel sistema delle indagini correnti (es. Smart survey)
		Non necessita di investimenti metodologici ad hoc (costo dell'investimento metodologico)
		Coerenza con progetti internazionali
AVANZAMENTI TECNOLOGICI		Utilizzo della piattaforma disponibile (costo dell'investimento IT hw)
		Non necessita di sviluppi applicativi ad-hoc (costo dell'investimento IT applicativo)
AVANZAMENTI RACCOLTA DATI E DIREZIONE GENERALE		Riduzione del carico informativo sui rispondenti tramite la riduzione dei dati acquisiti da rilevazioni dirette
		Vantaggio in termini di costi di accesso/acquisizione delle nuove fonti rispetto ai costi sostenuti per la raccolta dati
		Facilità di definire accordi istituzionali e commerciali
		Progetti che ci impegnano contrattualmente

L'esito di questa ricognizione ha fatto rientrare nella programmazione 2022-2023 i seguenti progetti:

Tabella 3 – Progetti prioritari TSS 2022-2023

TIPOLOGIA	FONTE	NUOVI PROGETTI
Smart System	Transazioni elettroniche	Statistiche sulla digitalizzazione delle transazioni elettroniche nella Pubblica Amministrazione
	Sensori	Statistiche sul trasporto navale e l'uso dei porti per mezzo di Automatic Identification Systems (AIS)
Web Intelligence	Generic scraping	Informazioni derivate dai siti delle imprese per l'arricchimento del Registro delle imprese ASIA
		Statistiche sulle classificazioni emergenti per le imprese con uso delle informazione dei siti delle imprese
	Ad hoc scraping	Statistiche con uso di Scanner data per indici dei prezzi e cambiamenti nei comportamenti di spesa
	Immagini satellitari	Statistiche con uso di immagini satellitari sulla copertura dei suoli e la riforestazione

Il monitoraggio di questi progetti e quindi il loro stadio di realizzazione non sarà qui trattato in quanto riguarda ambiti prettamente statistici e informatici che esulano dal perimetro del presente lavoro che nelle pagine successive si occuperà del monitoraggio dei servizi trasversali a supporto del sistema di produzione delle TSS.

È stato fatto riferimento ai progetti prioritari innanzitutto perché trattare di servizi trasversali senza agganciarli ad uno specifico *output* non ci permette di capire l'importanza del supporto delle azioni trasversali e della integrazione e collaborazione tra i vari settori in prima battuta

per il rilascio dei singoli prodotti e poi per l'implementazione del nuovo sistema di produzione.

Nella *Roadmap* quindi sono stati inseriti i progetti prioritari scaturiti da questo processo di selezione basato sugli impatti che i singoli progetti possono avere sulla produzione di statistica ufficiale e sono stati individuati tutte le azioni necessarie a realizzare i progetti individuati afferenti alle dimensioni riportate nella figura.

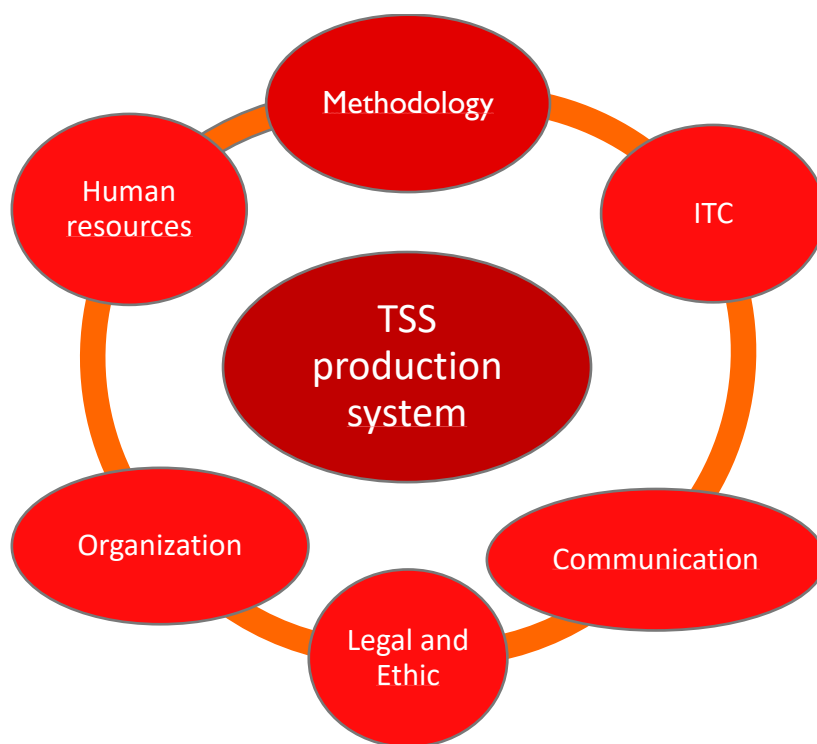


Figura 4 - dimensioni organizzative che concorrono alla realizzazione del sistema TSS

Una volta definiti i progetti di produzione, le singole strutture sono state chiamate ad individuare, ciascuno per il proprio ambito di competenza, le azioni necessarie a realizzarli. È quindi emerso un catalogo di attività il cui monitoraggio consente di verificare lo stato di realizzazione di ciascun nuovo *output* statistico di cui si parlerà a partire dal paragrafo successivo.

3.5.1. Un esempio di attività trasversale: la *privacy*

In questo paragrafo riportiamo a titolo di esempio il riferimento ad uno degli ambiti chiaramente emersi in fase di rilevazione delle attività che già nella letteratura scientifica sul tema TSS e nei documenti interni agli INS emerge come fondamentale nella traiettoria evolutiva degli Istituti verso il sistema TSS: quello della *privacy* legata sia all'accesso alle fonti detenute da *providers* esterni agli INS sia alla tutela del cittadino.

La breve trattazione che segue descrive la logica con cui ciascun ambito organizzativo, tra quelli inseriti nella Figure n. 4, seppure con diversi livelli di approfondimento ha, all'interno della *Roadmap*, definito (i) il proprio contesto di riferimento e quindi il livello di partenza, (ii) l'obiettivo da raggiungere, (iii) le attività da realizzare.

La scelta di questo ambito organizzativo è stata dettata dal fatto che nel presente lavoro è stato fatto riferimento più volte alla visione antropocentrica delle TSS, ai big data come nuovi misuratori sociali, all'importanza del digitale come nuova infrastruttura sociale, alla necessità per gli Istituti di statistica di adeguare le varie dimensioni riportate in figura 4 per accogliere il sistema TSS: la *privacy* come vedremo riassume tutti questi aspetti e fa emergere l'assoluta urgenza di un avanzamento normativo sul versante accesso, uso e integrazione delle nuove fonti.

Eurostat e il Sistema statistico europeo, in considerazione del rapido processo di trasformazione digitale che ha provocato i fenomeni di cui si è parlato ampiamente nei capitoli precedenti, hanno orientato i loro sforzi a approfondire le prospettive e le possibilità correlate all'uso dei big data per la produzione della statistica ufficiale e a incoraggiare lo sviluppo di statistiche sperimentali su nuovi fenomeni. Parallelamente a ciò, emerge la necessità della definizione di uno specifico quadro giuridico a livello europeo e nazionale che riduca gli ostacoli all'accesso, uso e integrazione di dati eterogenei. Il Sistema Statistico Europeo riconosce e sottolinea come tale cambiamento esiga un approccio metodologico basato sul principio di *privacy by design* in linea con il *General Data Protection Regulation* (d'ora in poi GDPR nel testo), il nuovo regolamento europeo in materia di protezione dei dati

personali entrato in vigore nel 2018. Attualmente non esiste un approccio armonizzato alla condivisione dei dati *Business-to-Government* negli Stati membri e ciò espone al rischio che le iniziative messe in campo dai singoli Istituti, siano soggette a regole e pratiche amministrative disomogenee. Eurostat evidenzia quindi l'importanza di una regolamentazione uniforme sulla condivisione dei dati tra imprese e amministrazioni e sulla definizione di principi chiari, volti ad assicurare un solido sistema di tutele e garanzie per gli interessati e gli operatori economici in materia di riservatezza. Di fatto in diversi paesi europei taluni istituti di statistica nazionali sono stati realizzati dei lavori sperimentali che prevedono l'uso dei Big Data ed in particolare dei dati di telefonia mobile. In Italia sui problemi di *privacy* legati all'uso delle nuove fonti il Garante per la protezione dei dati personali (d'ora in poi Garante nel testo) si è più volte pronunciato. In linea di massima il Garante aderisce alla posizione di Eurostat nel dibattito comunitario sull'uso dei Big data a fini di statistica ufficiale. Se infatti la necessità quanto meno di sperimentare l'uso di queste nuove fonti risulta più che comprovata ed urgente per una produzione statistica adeguata al contesto storico di riferimento, d'altro canto essa non può prescindere da una disciplina omogenea a livello europeo che tenga anche conto delle rilevanti implicazioni sui diritti e le libertà fondamentali degli interessati sotto il profilo della protezione dei dati e della riservatezza, in linea con quanto stabilito dalla Carta dei diritti fondamentali dell'Unione Europea e dal Regolamento¹¹⁴.

L'accesso a dati personali, con particolare riferimento ai dati di telefonia mobile, gestiti da enti privati, inerenti ad aspetti particolarmente intimi della vita degli interessati (come le comunicazioni personali e gli spostamenti sul territorio) potrebbe risultare del tutto imprevedibile e sproporzionato, se non corroborato da specifiche misure di trasparenza e

¹¹⁴ Sotto altro profilo, il Garante ha avuto già modo di esprimersi in merito all'uso di dati provenienti da fonti privati, quelli relativi ai consumi energetici per scopi censuari, sottolineando come l'intromissione nella dimensione più intima e privata degli individui, ossia quella domestica, esiga che vengano individuate specifiche misure a tutela degli interessati anche per prevenire violazioni del divieto di ricadute amministrative sarà necessario, in vista della sistematizzazione dell'uso dai dati *Business-to-Government* nel panorama delle fonti utilizzabili per scopi di statistica ufficiale, individuare a livello comunitario garanzie adeguate a salvaguardare la legittima pretesa di riservatezza vantata da ogni individuo (inteso come utente/contraente) rispetto ad indebite ingerenze dello Stato nella sua vita privata.

bilanciamento¹¹⁵. Il Centro TSS ha cercato di affrontare queste criticità messe in luce dal Garante perché connesse al fatto che le TSS delineano un approccio all'acquisizione delle informazioni e alla generazione di conoscenza del tutto innovativo, che ne rende problematico l'inquadramento nell'ambito del sistema normativo che attualmente regola lo sviluppo e la produzione della statistica ufficiale. L'attuale sistema normativo, infatti, prende essenzialmente a riferimento il paradigma tradizionale del processo statistico rispetto al quale le TSS presentano un'evidente differenziazione strutturale, sotto molteplici e rilevanti profili: tipologia e natura delle fonti e dei dati trattati, tecniche e metodi di raccolta ed elaborazione dei dati, moduli organizzativi e soggetti coinvolti. La riflessione in Istat sugli aspetti normativi, giuridici e deontologici concernenti le TSS deve partire dalla consapevolezza che anche a livello europeo manca una regolamentazione giuridica specifica. Questa carenza è resa evidente, in primo luogo, dalla mancanza nell'ordinamento europeo e in quello nazionale di una definizione normativa di TSS. L'assenza di una definizione normativa delle TSS - che corrisponde all'assenza di una definizione normativa anche dei big data - è strettamente collegata alle oggettive difficoltà che in questa fase ancora si riscontrano nel precisare, anche sul piano strettamente tecnico-statistico, i contenuti tipici delle attività in cui le TSS si sostanziano. Queste attività sono infatti caratterizzate, oltre che da una intrinseca novità e complessità, da una spiccata tendenza ad evolvere rapidamente per adattarsi ai continui mutamenti del contesto di riferimento, tecnologico e non. Tuttavia, la formalizzazione a livello normativo della nozione di *Trusted Smart Statistics* costituisce il presupposto necessario per l'adozione di una regolamentazione che ne legittimi pienamente l'implementazione da parte dell'Istat. L'accurata perimetrazione definitoria offrirebbe le coordinate per progettare i contenuti regolatori più appropriati per la disciplina di aspetti che altrimenti risulterebbero sprovvisti di adeguata copertura giuridica. Questi aspetti riguardano in particolare la raccolta e alla successiva elaborazione dei dati, la protezione dei

¹¹⁵ Garante per la protezione dei dati personali, doc. web n. 9802796] Parere sui lavori statistici IST 02834 Studio dei Mobile Network Data a fini statistici e IST 02829 La violenza raccontata dai social - 9 giugno 2022 Registro dei provvedimenti n. 235 del 9 giugno 20202, disponibile al link <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9802796#note>

dati personali e la tutela della riservatezza. Sebbene l'uso delle nuove fonti pone ovviamente problemi legati alla *privacy*, bisogna segnalare innanzitutto che l'Istat ha fatto in qualche modo da "apripista" e pioniere nello sviluppo della riflessione sulla tematica dell'uso dei big data per finalità pubbliche e che attualmente svolge un ruolo di punta nei progetti UNECE legati all'*Input privacy preserving* ovvero all'implementazione di tecnologie che supportano i fornitori di dati e gli utenti a condividere le informazioni in modo sicuro, utilizzando protocolli che consentono loro di produrre dati di *output* utili senza "vedere" i dati di *input*. Tutte i prototipi e le sperimentazioni necessarie a rafforzare strumenti di *privacy by design* e *privacy by default*, rappresentano strumenti di rafforzamento della *trust* nei confronti dell'Istituto che rilascia le informazioni statistiche, dei nuovi prodotti statistici e dell'infrastruttura digitale in generale. In questo filone si inserisce perfettamente la sentenza n. 20 del 2019 della Corte Costituzionale «[i]l diritto alla protezione dei dati di carattere personale non è una prerogativa assoluta ma va considerato alla luce della sua funzione sociale e va temperato con altri diritti fondamentali, in ossequio al principio di proporzionalità». Il rimando alla *funzione sociale* conferma nuovamente la centralità del ruolo del cittadino e la visione antropocentrica più volte richiamata e che sta alla base di questi nuovi prodotti statistici, questo è il principio guida che dovrebbe orientare l'adeguamento delle varie componenti del nuovo sistema di produzione in un'ottica *win-win*: aumenta il coinvolgimento del cittadino e aumenta la fiducia verso i dati e verso il ruolo istituzionale degli Istituti che li rilasciano.

3.6 Implementazione di un *framework* per il monitoraggio

L'individuazione delle azioni a supporto necessarie per la realizzazione dei progetti prioritari inseriti nella *Roadmap* e riportati nella tabella n 3 è stata curata dalla segreteria tecnica organizzativa, anch'essa composta secondo un criterio interdisciplinare e interdipartimentale. Ciascuna struttura coinvolta nella Segreteria Tecnica ha in prima battuta, elaborato quindi un insieme di azioni funzionali allo sviluppo dei primi progetti prioritari selezionati nei seguenti settori:

- ✓ metodologia
- ✓ informatica

- ✓ legale (*privacy*, etica...)
- ✓ risorse umane
- ✓ comunicazione e diffusione
- ✓ organizzazione

Al fine di monitorare l'avanzamento dei progetti di produzione e delle necessarie azioni a supporto, in quanto membro della Segreteria Tecnica del Centro, a diretto supporto del responsabile degli aspetti tecnici e metodologici del Centro stesso, ho elaborato una matrice che rappresenta di fatto un vero e proprio *framework* di monitoraggio utile a rendicontare allo *Steering Committee* lo stato di avanzamento raggiunto e soprattutto a fare emergere aree di criticità e/o di miglioramento e aree maggiormente consolidate.

Tutte le azioni che ciascuna struttura organizzativa ha segnalato come necessarie alla programmazione strategica contenuta nella *Roadmap*, sono quindi state usate per alimentare la matrice e aggregate in due grandi blocchi. All'interno del primo gruppo ricadono azioni "*Soft/Human level*" riconducibili al supporto organizzativo in tutte le sue possibili declinazioni, il secondo – "*Metodi e Tecnologie/Hard Skills*" contiene azioni maggiormente infrastrutturali relative alla costruzione e al consolidamento delle infrastrutture informatiche, tecniche e metodologiche.

I due blocchi soprariportati contengono *cluster* di attività, all'interno dei quali ricadono azioni facenti capo a strutture organizzative diverse, la cui realizzazione rende necessaria la collaborazione e l'interazione tra più strutture. L'uso della matrice permette di rendere maggiormente visibile la necessità di collaborazione tra le strutture e quindi rappresenta anche uno strumento di sensibilizzazione sull'imprescindibile esigenza di interazione tra strutture. Per comprendere meglio quanto detto si riportano di seguito le attività che le Direzioni dell'Istituto hanno segnalato come indispensabili alla creazione del nuovo sistema di produzione delle TSS e che ricadono nei blocchi che per semplicità definiremo *Soft* e *Hard*

3.6.1. Area *soft/human level*

In questa area rientrano le azioni riconducibili ai vari ambiti legati al supporto organizzativo e nello specifico:

- ✓ Area Risorse umane: investimenti su nuove competenze, con specifico riferimento al rafforzamento dell'area *data science*.
- ✓ Ambito legale:
 - individuazione di nuove forme di accordi, partnership e contratti per l'accesso alle fonti e per rafforzare l'ecosistema della ricerca.
 - Elaborazione di strumenti di *privacy by design* e *by default*
 - Sviluppo di codici etici per l'uso dei dati e degli strumenti di intelligenza artificiale usati nel trattamento dati
- ✓ Ambito Pianificazione e *performance*: definizione di azioni di: supporto organizzativo finalizzate al sistema TSS, verifica delle coerenze per l'inserimento delle attività contenute nella *Roadmap* nella pianificazione di Istituto legata alla ricerca e all'innovazione, individuazione di indicatori di monitoraggio utili a misurare il contributo di ciascuna struttura al raggiungimento degli obiettivi programmati in ambito TSS.
- ✓ Area Comunicazione: individuazione di strumenti innovativi per la comunicazione dei nuovi prodotti e progettazione di un nuovo paradigma di comunicazione finalizzato a valorizzare e a diffondere i nuovi *output* statistici: se le statistiche ci aiutano a capire meglio il mondo che ci circonda e se effettivamente migliorano le decisioni (e quindi le nostre vite) non sono solo questioni di metodologia scientifica. Il fattore decisivo qui è se le statistiche, come un linguaggio, sono comprese da coloro per i quali l'informazione è rilevante.
- ✓ Area Terza Missione: individuazione di nuovi *stakeholders* e di nuove domande informative.
- ✓ Strategia nazionale dati – Consolidamento della *stewardship* dell'Istat-

3.6.2. Area Metodi e tecniche – *Hard skills*

Le azioni infrastrutturali che ricadono in questa area sono ricollegabili ai seguenti ambiti:

- ✓ metodologico: individuazione di nuovi metodi, tecniche, strumenti e *standard* (finalizzati per esempio alla riduzione del rischio di *bias* delle statistiche derivate da big data, ed in particolare quello indotto della selettività, non-rappresentatività delle sorgenti);
- ✓ Informatico: Individuazione di moderne infrastrutture IT e consolidamento di quelle esistenti.
- ✓ Raccolta dati: industrializzazione e uso di nuove fonti (*web scraping*, immagini satellitari...).

Tutte queste azioni - e le relative attività operative connesse ad esse - sono state inserite nel *framework* di monitoraggio al fine di valutarne il grado di implementazione e di misurare il livello di maturità complessivo del sistema TSS in Istituto.

Per fare questo, il Centro TSS si è avvalso, riadattandola ai propri fini, della *Matrice di Maturità* organizzativa, uno strumento di valutazione realizzata in ambito UNECE nel 2020 dal *Team on Training, Competencies and Capacity Development di UN Global Working Group on Big Data*.¹¹⁶

Questa Matrice di Maturità, rappresenta uno strumento di autovalutazione per aiutare gli istituti di statistica a comprendere fino a che punto sono stati sviluppati le infrastrutture e le azioni necessarie ad integrare le nuove fonti nei processi di produzione statistica. Aiuta l'organizzazione a comprendere il proprio livello di "maturità organizzativa" e a identificare le aree di criticità e quelle di miglioramento. Dai risultati di questa matrice poi può derivare un piano di sviluppo per programmare azioni future: la *Maturity Matrix* può essere usata per definire il livello *target* a cui l'istituto mira.

¹¹⁶ Task Team of the UN Committee of Experts on Big Data and Data Science for Official Statistics, Big Data Maturity Matrix UNECE 2020

La matrice analizza quattro settori organizzativi che coincidono in gran parte con quelli identificati in Istat e usati per il nostro *Framework* e sono:

- ✓ **Settore giuridico e definizione di policy:** si riferisce alla misurazione del consolidamento del quadro giuridico e delle politiche associate all'accesso ai dati e alla condivisione dei dati. Questo settore include le salvaguardie a tutela della *privacy* e la riservatezza dei big data e dei processi per analizzarli.
- ✓ **Infrastruttura IT:** l'esistenza di un'infrastruttura IT per il trattamento e analisi dei big data in un ambiente sicuro.
- ✓ **Risorse umane:** si riferisce ad azioni di rafforzamento delle competenze con specifico riferimento alla data science.
- ✓ **Application:** applicazioni concrete data science.

Per ciascuno di questi settori, sono considerati 4 stadi di maturità organizzativa:

Livello 1: *Pre-foundation*: questo livello in genere si riferisce ad un'organizzazione che ha appena iniziato il proprio percorso in uno specifico settore e che si trova in una fase di riflessione e di analisi di contesto.

Livello 2: *Foundation*: le organizzazioni che hanno raggiunto questo livello stanno cominciando a sviluppare strategie e *framework* dedicati.

Livello 3: *Practitioner*: nelle organizzazioni che si attestano a questo livello sono già stati definiti i *framework* di riferimento e avviati i progetti di consolidamento del settore.

Livello 4: *Expert*: questo è il livello di maturità più alto, le organizzazioni che lo raggiungono hanno anche superato la fase di sperimentazione e sono in grado di offrire formazione e tutoraggio anche all'esterno.

La *Maturity Matrix* fornisce anche specifiche indicazioni su quali sono le azioni necessarie per raggiungere ciascuno dei livelli e pertanto fornisce anche una sorta di traiettoria evolutiva ideale per passare dal livello di partenza a quello *Expert*.

Questo strumento si innesta perfettamente sul nostro *framework* di evoluzione/orientamento e monitoraggio del sistema TSS, perché permette l'emersione di punti di forza e di debolezza e così aiuta ad individuare anche le azioni necessarie a raggiungere gli obiettivi fissati nella *Roadmap*: il *framework* diventa anche un vero e proprio strumento di gestione della *performance* organizzativa. La misurazione del grado di maturità di ciascuna area è stato integrato con un ulteriore indicatore che definisce invece il grado di coerenza di ciascuna delle azioni inserite dalle singole direzioni all'interno del monitoraggio rispetto agli obiettivi inseriti nella *roadmap*. Questo indicatore misura la performance ovvero il contributo delle singole azioni rispetto al raggiungimento degli obiettivi dichiarati nella *roadmap*. Ne deriva che le azioni indicate nel monitoraggio che, pur contribuendo agli obiettivi generali dell'Istituto, non sono funzionali agli obiettivi indicati nella *roadmap* e che non impattano sulla traiettoria evolutiva verso la costituzione del Sistema di produzione delle TSS, non alimentano la misurazione di questo indicatore.

La matrice e i relativi indicatori, sono stati progettati con l'intento di superare un approccio verticale legato alla misurazione della singola attività, la matrice rappresenta uno strumento di gestione della performance e non un catalogo di azioni separate le une della altre. E infatti già a partire dalla definizione delle azioni, è stata adottata una visione sistemica che permettesse di valutare il contributo di ciascuna azione e le necessarie interazioni con le altre attività per al raggiungimento degli obiettivi contenuti nella *Roadmap*.

3.7 Implementazione della matrice

Ad ognuno dei cluster indicati sopra sono state attribuiti una o più azioni e a ciascuna di queste è stata attribuito un rating a seconda del livello di maturità raggiunto:

Tabella 4 – Livello di Maturità organizzativa e relativo punteggio

Livello	Rating
Pre-foundation	1
Foundation	2
Practitioner	3
Expert	4

Anche all'indicatore relativo al grado di coerenza è stato attribuito un punteggio (riportato nella tabella sottostante) a seconda del grado di coerenza dell'azione rispetto agli obiettivi strategici contenuti nella *Roadmap*.

Tabella 5 – Livello di coerenza e relativo punteggio

Livello	Rating
Nessuna coerenza	1
Basso livello di coerenza	2
Medio livello di coerenza	3
Alto livello di coerenza	4

Le interlocuzioni con le persone coinvolte nelle diverse linee di attività ha permesso di implementare la matrice e di far emergere sia i settori sui quali è necessario investire maggiormente sia i settori più consolidati sui quali occorre lavorare per mantenere e, auspicabilmente, rafforzare il livello di maturità raggiunto. A titolo di esempio si riportano di seguito i risultati del primo monitoraggio effettuato e i *rating* assegnati a ciascuna azione.

Per il livello *soft* la tabella seguente riporta il monitoraggio effettuato sull'area relativa alle risorse umane, la scelta di questa specifica area è stata fatta perché più volte nel presente lavoro è stata evidenziata l'importanza di investimenti per rafforzare le competenze delle risorse umane per spingere e sostenere gli INS verso la produzione di statistiche innovative.

Tabella 6 – Framework di monitoraggio: esempio su azioni trasversali – Soft Skills

Human level/soft skills						
Macro aree di intervento	Obiettivi della macroarea contenuti nella Roadmap	Azioni inserite nel monitoraggio e legate agli obiettivi nella Roadmap	Strutture titolari	Strutture in collaborazione	Grado di raggiungimento obiettivi e stato maturità del sistema (Uso Big data Matrix) 1 Base, 2 Intermedio, 3 Avanzato, 4 Esperto	Coerenza delle azioni segnalate con Progetti Roadmap (1: nessuna coerenza, 2 Bassa; 3 Media; 4 Alta)
Investimenti su nuove competenze	Investimento su nuove risorse umane e nuove competenze a sostegno del nuovo sistema di produzione	Adeguamento della Banca Dati Competenze	Direzione delle risorse Umane	Direzione per le Metodologie e il disegno dei processi strategici (per la definizione delle competenze in area Data science)	2	4
		Adeguamento offerta formativa	Direzione delle risorse Umane	Direzione per le tecnologie Informatiche	2	4
		Ampliamento aree concorsuali	Direzione delle risorse Umane		2	4
		Partnership Università, EpR...	Direzione per gli Affari Legali	Direzione delle risorse Umane e strutture coinvolte di volta in volta	1	4
		Attivazione di nuovi Tirocini	Direzione delle risorse Umane	Direzione delle risorse Umane e strutture coinvolte di volta in volta	3	3
		Estensione e completamento della formazione per l'utilizzo delle infrastrutture	Attivazione dei canali con la Dir.Risorse Umane per la analisi e predisposizione degli interventi formativi	Direzione per le tecnologie Informatiche	Direzione delle risorse Umane	2

Questo primo monitoraggio fa emergere innanzitutto un grado di coerenza molto alto delle attività programmate rispetto all'obiettivo finale, segnala che le singole strutture hanno maturato un elevato livello di consapevolezza in merito alle esigenze del nuovo paradigma di produzione e che le *lessons learned* grazie alle esperienze europee stanno orientando al meglio specifiche scelte operative dell'Istituto. Emerge con forza la necessità che il capitale umano disponibile presso gli Istituti di Statistica si adegui ai cambiamenti per rigenerarsi velocemente e per gestire le tecnologie disponibili e i nuovi dati e affrontare nuove sfide informative. A fronte un elevato livello di coerenza, il grado effettivo di realizzazione delle attività appare meno incoraggiante in alcuni settori specifici. Le motivazioni sono riconducibili a diversi fattori. Il primo fra tutti è una resistenza culturale al cambiamento, le risorse sono fortemente impegnate nelle stringenti attività legate alle attività correnti e difficilmente sono disponibili risorse da dedicare allo sviluppo di progetti innovativi. L'Istituto rispetto a quanto rilevato nel quadro riportato, ha già fatto investimenti mirati a migliorare alcuni degli aspetti segnalati come maggiormente carenti. Sono infatti stati calendarizzati *webinar* di formazione e diffusione sulle TSS, stati programmate nuove forme di reclutamento e di riqualificazione del capitale umano disponibile.

Per quanto riguarda l'ambito più strettamente tecnologico, di seguito si riportano due macro aree di intervento nelle quali ricadono azioni assolutamente fondamentali per la realizzazione del nuovo sistema di produzione delle TSS: quelle relative all'implementazione di nuove infrastrutture IT e al consolidamento di quelle esistenti e la definizione di policy e standard tecnici.

Anche da questo quadro emerge l'importanza della collaborazione tra le strutture e nel caso della Macro Area Policy e standard tecnici emerge anche la rilevanza del contributo che l'Istat dà all'interno dei progetti UNECE in termini di definizione e realizzazione di policy a tutela della *privacy* (la cui importanza è già stata ampiamente descritta nelle pagine precedenti) in termini di (i) rassegna di casi d'uso relativi alla *privacy in input*, (ii) *framework* concettuale per la loro descrizione standardizzata, (iii) software prototipali per eseguire funzionalità di

Private Set Intersection with Analytics e per addestrare modelli *machine learning* in modalità *privacy preserving*, (iv) attività di sviluppo di tecniche e di ambienti tecnologici per il trattamento della *privacy in input*.

Dal monitoraggio in questa area emerge che l'identificazione di strumenti e tecniche utili ad assicurare la *privacy in input* nelle fasi di trattamento delle nuove fonti è non solo coerente ma ha anche raggiunto un alto livello di realizzazione. In generale risulta che gli aspetti più tecnici sono a un livello più consolidato di maturità e che bisogna lavorare maggiormente sugli aspetti trasversali, rafforzando l'infrastruttura organizzativa investendo anche qui in processi innovativi e nuovi paradigmi.

Tabella 7 - Framework di monitoraggio: esempio su azioni trasversali *Hard Skills*

Metodi e tecnologie/ <i>Hard skills</i>						
Macro aree di intervento	Obiettivi della macroarea contenuti nella Roadmap	Azioni inserite nel monitoraggio e legate agli obiettivi nella Roadmap	Strutture titolari	Strutture in collaborazione	Grado di raggiungimento obiettivi e stato maturità del sistema	Coerenza delle azioni segnalate con Progetti Roadmap
Nuove infrastrutture e consolidamento di quelle esistenti	Consolidamento degli investimenti in infrastrutture realizzati	Acquisizione, aggiornamento e messa in esercizio di piattaforme a supporto delle nuove fonti e big data	Direzione per le tecnologie Informatiche	Direzione per le Metodologie e il disegno dei processi strategici	2	3
Policy e standard tecnici	<i>Privacy in input</i>	Progetto UNECE <i>Input Privacy Preservation</i> :	Direzione per le metodologie e i disegno dei processi statistici	Direzione per la raccolta dati	3	3

3.8 Il percorso fino alle TSS: il processo di modernizzazione in Istat.

Cause e impatti.

I risultati finora raggiunti in ambito TSS sono frutto di un processo interno di modernizzazione e di adeguamento dei modelli organizzativi e di produzione al contesto europeo e internazionale. Dall'aprile del 2016, l'Istat ha impresso un radicale cambiamento ai processi di produzione, in linea con la *Vision2020*, la strategia europea per la modernizzazione della statistica ufficiale, e le raccomandazioni dell'*High-level Group for the Modernization of Official Statistics*¹¹⁷ dell'UNECE (d'ora in poi HLG- MOS nel testo), aprendo nuove prospettive per l'intero Sistema statistico nazionale.

I gruppi HLG-MOS hanno il compito di supervisionare lo sviluppo di *frameworks* e la condivisione di informazioni, strumenti e metodi a sostegno della modernizzazione delle organizzazioni statistiche. L'obiettivo è migliorare l'efficienza dei processi di produzione statistica e la capacità di produrre risultati che soddisfino meglio le esigenze degli *stakeholders*, garantendo massima convergenza e coordinamento tra tutti i paesi partecipanti. Il gruppo HLG MOS a cui l'Istat ha partecipato, ha definito una *roadmap* comune per la modernizzazione degli istituti nazionali e internazionali di statistica basata sugli sviluppi della *Business Architecture* (d'ora in poi BA nel testo) per assicurare la piena condivisione di informazioni e il coordinamento tra i paesi partecipanti alla *Conference of European Statisticians*.

La BA (Allegato 3) ha svolto un ruolo centrale in un programma così complesso come quello della modernizzazione degli Istituti Nazionali di statistica. Il contesto internazionale ed europeo è composto da differenti istituzioni che si occupano di produzione di statistiche ufficiali, al loro interno spesso caratterizzate da una molteplicità di modelli organizzativi. Dal momento che risulta essenziale migliorare lo scambio di sinergie reciproche, così come la collaborazione tra i diversi paesi, diventa sempre più importante individuare linguaggi comuni e condividere strategie di produzione. Ciò anche al fine di rispondere più

¹¹⁷ <https://unece.org/statistics/networks-of-experts/high-level-group-modernisation-statistical-production-and-services>

tempestivamente ed efficacemente alle esigenze provenienti sia dal SSE, sia dalle organizzazioni internazionali.

La BA è un modello di riferimento finalizzato a ottimizzare i processi di lavoro all'interno di un'organizzazione e renderli più efficienti. Tale modello riguarda sia le attività statistiche, sia le funzioni organizzative e strategiche, sia quelle inerenti le capacità. La BA è la parte principale e strategica della cornice più generale rappresentata dall'Enterprise Architecture, che individua i diversi elementi che compongono un'organizzazione e le modalità attraverso le quali interagiscono tra loro, fornendo un quadro chiaro, coerente e realizzabile di ciò che è necessario per conseguire gli obiettivi da raggiungere. La BA rappresenta quindi un linguaggio comune, indispensabile per intraprendere percorsi di innovazione congruenti e condivisi, che riguarda sia le attività statistiche, sia le funzioni organizzative e strategiche, nonché le capacità e le competenze; copre tutte le attività intraprese per la produzione di *output* statistici, tra cui la parte concettuale, la progettazione, le risorse informative e le attività applicative.

È un modello integrato di rappresentazione dei processi e delle attività che costituisce un linguaggio comune, indispensabile per intraprendere percorsi di innovazione congruenti e condivisi. A livello europeo e internazionale, l'Istat ha contribuito attivamente allo sviluppo di un modello generalizzato di BA, considerato come uno strumento di riferimento per ottimizzare i processi di lavoro all'interno di una organizzazione statistica e renderli più efficienti. La BA guida il cambiamento culturale connesso al processo di modernizzazione, privilegiando l'armonizzazione e la standardizzazione, nonché facilitando il superamento delle eterogeneità di soluzioni procedurali, metodologiche e tecnologiche, come anche l'adozione di standard e l'eliminazione delle ridondanze nei dati e nelle applicazioni. La BA copre tutte le attività intraprese dall'Istituto per la produzione di *output* statistici, tra cui la parte concettuale, la progettazione, le risorse informative e le attività applicative, ed è caratterizzata da quattro aree omogenee rispetto alle attività svolte e alla natura delle informazioni trattate e/o dei servizi che impattano su tali informazioni. Le quattro aree, chiamate *Business Lines*, sono caratterizzate da specifici gruppi di attività, a loro volta

articolate in singole azioni e sono definite in modo da garantirne sia l'indipendenza dalla struttura organizzativa, sia la loro stabilità rispetto a futuri processi di riorganizzazione: (i) Strategia; (ii) Supporto generale; (iii) Capacità; (iv) Produzione ¹¹⁸.

La BA ha quindi guidato gli adeguamenti organizzativi del programma di modernizzazione che, nella sua fase di progettazione e di sviluppo, ha tenuto conto anche di quei cambiamenti del contesto esterno che hanno avuto un impatto significativo sull'evoluzione della domanda e dell'offerta di informazione di statistiche ufficiali. Sono aumentate le aspettative degli utenti, la crescente domanda d'informazione statistica ha cominciato a caratterizzarsi ormai per un'ampiezza senza precedenti, sia di carattere tematico (economico, sociale, ambientale, etc.), sia di dettaglio territoriale (da fenomeni globali a tendenze micro territoriali); questa domanda è stata però spesso soddisfatta da un'abbondanza di informazioni, anche non strutturate.

Nuovi impulsi provenienti dall'esterno, per lo più riconducibili alle nuove tecnologie, hanno completamente riscritto le regole della produzione e della comunicazione a tutti i livelli. La straordinaria abbondanza di informazioni, la facilità e la velocità con le quali queste possono essere acquisite, elaborate e distribuite, nonché la presenza sul mercato di operatori in grado di svolgere attività statistica a fianco di quelli istituzionali spingono gli INS a rivedere i propri business model per rimanere competitivi all'interno di un ormai affollatissimo ecosistema della conoscenza. L'Istat per decodificare i cambiamenti provenienti dal mondo esterno ha disegnato e poi attuato un percorso di evoluzione, mettendo al centro di questo cambiamento l'innovazione di processo e di prodotto come fondamento dell'attività istituzionale e ha identificato nuovi linguaggi attraverso cui porsi nei confronti degli utenti.

Il programma di modernizzazione dell'Istat si muove lungo due direttrici. La prima direttrice è l'opportunità di fare parlare tra loro la pluralità delle fonti a disposizione e cogliere, così, relazioni tra dati e fenomeni che sarebbero rimaste "invisibili". La parola chiave, qui, è

¹¹⁸ Barcaroli G., Falorsi P.D., Mignolli N., Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session STS048)

integrazione e lo strumento è la costruzione di un'infrastruttura di base (il Sistema integrato dei Registri), che sfrutti a pieno le informazioni derivanti dagli archivi amministrativi e le integri con quelle tratte dalle tradizionali rilevazioni campionarie e dalle nuove fonti, come i big data. L'obiettivo è "conciliare" la documentazione tratta dagli archivi, sui diversi soggetti di interesse (individui e famiglie, unità economiche e istituzioni, unità geografiche e territoriali), con i dati su caratteristiche, eventi, comportamenti e atteggiamenti rilevati con le indagini o dall'analisi delle nuove fonti.

La seconda direttrice è di carattere organizzativo. Il disegno ha previsto l'accentramento delle funzioni trasversali relative alla raccolta e acquisizione dei dati, metodologie e tecnologie per il loro trattamento e alla comunicazione al fine di giungere ad un maggiore livello di standardizzazione ed efficienza dei processi produttivi.

Il modello organizzativo adottato ha come obiettivo quello di superare la frammentarietà organizzativa e il modello *stovepipe* che prevedeva veri e propri silos organizzativi non integrati fra loro e che ha generato inevitabili duplicazioni e ridondanze con un notevole spreco di risorse¹¹⁹.

Il programma di modernizzazione dell'Istituto ha rappresentato un significativo progetto di *change management* che ha implementato una struttura organizzativa più snella per sostenere in maniera più efficace i processi di produzione statistica ed è stato guidato in ciascuna delle sue fasi, da un forte orientamento all'innovazione e alla ricerca.

¹¹⁹ Il programma di Modernizzazione in Istat, 2016 Istat <https://www4.istat.it/it/dodicesima-conferenza/programma/005-modernizzazione>

3.9 La ricerca in Istat

Il nuovo modello organizzativo scaturito dal programma di modernizzazione, per il suo forte orientamento alla ricerca e alla innovazione, ha sicuramente rappresentato un contesto favorevole anche al recepimento del Decreto Legislativo (d'ora in poi Dlgs nel testo) 218/2016¹²⁰ che all'art. 1 riconferma la presenza dell'Istat tra gli Enti pubblici di Ricerca (d'ora in poi EpR) e riconferma il ruolo centrale della ricerca. Il Decreto ha lo scopo di innovare la disciplina degli EpR operando uno sganciamento di questi enti dalle procedure previste per la maggior parte delle PA, al fine di allinearli quasi completamente al sistema delle autonomie universitarie. L'intervento normativo si inserisce nel generale processo di riordino e semplificazione della PA voluto dalla L. 124/2015¹²¹ e improntato ai principi di efficienza, efficacia, trasparenza dell'attività amministrativa e di accessibilità delle informazioni.

Il Dlgs 218/2016 contribuisce a rendere le procedure e le normative più snelle e consone alle peculiarità degli scopi istituzionali degli EpR, a superare alcune criticità legate alla scarsa autonomia statutaria e all'insufficiente livello di premialità per i ricercatori e tecnologi e a garantire un allineamento al contesto europeo e alla Carta europea dei ricercatori¹²². La semplificazione gestionale in un ambito avanzato e competitivo com'è quello della ricerca scientifica, è un elemento importantissimo in quanto conferma la peculiarità degli EpR

¹²⁰ Decreto legislativo 25 novembre 2016, n. 218, recante "Semplificazione delle attività degli enti pubblici di ricerca ai sensi dell'articolo 13 della legge 7 agosto 2015, n. 124". Pubblicato nella Gazzetta Ufficiale, Serie generale, n. 276 del 25 novembre 2016. Entrata in vigore: 10 dicembre 2016.

¹²¹ Legge 7 agosto 2015, n. 124 Deleghe al Governo in materia di riorganizzazione delle amministrazioni pubbliche Pubblicata nella gazzetta Ufficiale n. 187 del 13 agosto 2015.

¹²² La Carta Europea dei Ricercatori è una Raccomandazione della Commissione Europea che contiene un insieme di principi generali e requisiti che specificano il ruolo, le responsabilità e i diritti dei ricercatori e delle persone che assumono e/o finanziano i ricercatori. Scopo di tale Carta è garantire che la natura dei rapporti tra ricercatori e datori di lavoro o finanziatori favorisca esiti positivi per quanto riguarda la produzione, il trasferimento, la condivisione e la diffusione delle conoscenze e dello sviluppo tecnologico, e sia propizia allo sviluppo professionale dei ricercatori.

rispetto al resto della PA e inquadra la ricerca pubblica in un sistema di regole più snello e più appropriato a gestirne le specificità dei tempi e delle esigenze del settore.

Il Dlgs. 218/2016 introduce il ruolo di valutazione dell'ANVUR - Agenzia nazionale di valutazione del sistema universitario e della ricerca (art. 17, commi 1 e 2). La missione dell'Istituto rimane quella stabilita dal Dlgs 322/1989¹²³ di *“servire la collettività attraverso la produzione e comunicazione di informazioni statistiche, analisi e previsioni di elevata qualità; realizzate in piena autonomia e sulla base di rigorosi principi etico-professionali e i più avanzati standard scientifici; allo scopo di sviluppare un'approfondita conoscenza della realtà ambientale, economica e sociale dell'Italia ai diversi livelli territoriali e favorire i processi decisionali di tutti i soggetti della società”*. In questo contesto, l'Istat si configura quindi come un ente di ricerca rivolto alla produzione di dati e analisi, organizzato secondo un ben definito modello di produzione. In linea di massima, Programma statistico europeo (Pse) e il Programma statistico nazionale (Psn) - rispettivamente adottati con atti del Consiglio e del Parlamento europeo e del Presidente della Repubblica - stabiliscono cosa produrre, come produrre, è stabilito dal *Code of Practice* europeo e dal Codice della qualità della statistica ufficiale, con vigilanza rispettivamente da parte di Eurostat e della Commissione per la Garanzia dell'informazione statistica (Cogis).

I programmi e i codici di riferimento citati, richiedono un miglioramento costante dei livelli di qualità dei dati prodotti e l'adozione nei propri processi dei più avanzati standard scientifici. Ciò vuol dire che l'Istituto deve mantenere e migliorare continuamente i propri metodi di raccolta, elaborazione e analisi dei dati, sviluppando al suo interno le competenze necessarie per lo sfruttamento delle nuove tecnologie e metodologie statistiche. La ricerca non è quindi il fine principale dell'Istituto, ma rappresenta lo strumento essenziale per mantenere ed accrescere il livello di informazione statistica prodotta e diffusa e la sua qualità nonché quella dei processi adottati per sua produzione.

¹²³ Dlgs. 322/1989 Norme sul Sistema statistico nazionale e sulla riorganizzazione dell'Istituto nazionale di statistica, ai sensi dell'art. 24 della legge 23 agosto 1988, n. 400

L'Istat, nel ridefinire il modello di produzione con il programma di modernizzazione, ha rimesso al centro l'importanza di introdurre in ciascuna delle dimensioni organizzative la ricerca e l'innovazione. L'Istituto quindi ha avviato una serie di attività volte a valorizzare ricerca e innovazione legate soprattutto alle opportunità offerte dalla trasformazione digitale e dalle nuove fonti di dati. La centralità della ricerca trova riscontro nella pianificazione strategica destinata a rafforzare l'attività di ricerca che viene dunque riconosciuta come un fattore di crescita dell'Istituto e del personale da perseguire in maniera strutturata e organica.

L'Istat ha compiti di ricerca in alcuni ambiti specifici, funzionali al miglioramento sistematico dei processi di produzione dei dati, al rilascio di nuova informazione statistica, ad analisi e approfondimenti per migliorare la conoscenza dei fenomeni economici e sociali. Questa linea di indirizzo può essere declinata in due grandi filoni: la ricerca statistico-metodologica e la ricerca tematica. La ricerca in ambito statistico-metodologico ha come obiettivo il miglioramento dei metodi e degli strumenti usati per elaborare i dati lungo le varie fasi del processo di produzione, garantendo l'introduzione continua di soluzioni innovative da applicare nei processi correnti per l'ottimizzazione di uno o più aspetti della qualità: ad esempio l'accuratezza degli stimatori o la tempestività nel rilascio dei dati. L'attività di ricerca in ambito tematico costituisce uno strumento per l'arricchimento conoscitivo dei fenomeni, in termini quantitativi e qualitativi, grazie al carattere integrato e multidisciplinare delle analisi, allargando così l'offerta informativa dell'Istituto; permette, inoltre, di individuare lacune informative in grado di generare un successivo miglioramento delle fonti e un incremento della rilevanza dell'informazione statistica, contribuendo così anche ad aumentare la capacità di risposta dell'Istituto ai cambiamenti e alle nuove esigenze informative degli utenti esterni. Le aree prioritarie di ricerca in ambito statistico-metodologico o tematico, nonché l'impulso al miglioramento delle infrastrutture per la ricerca esistenti o l'impianto di nuove infrastrutture, saranno oggetto di revisione periodica e sottoposte alla preventiva approvazione del Comitato per la Ricerca e inserite in opportuni progetti strategici.

L'innovazione e la ricerca condotte nell'Istituto sono finalizzate al miglioramento sistematico dei complessi processi di produzione raggiungibile solo grazie ad azioni coordinate, integrate, programmate e di tipo collettivo, alle quali devono contribuire risorse appartenenti ad ambiti disciplinari differenti (statistici, metodologi, economisti, sociologi, esperti in *information technology*, etc)

Come stabilito dal Piano per la ricerca, documento strategico che orienta e struttura tutte le attività di ricerca dell'Istituto, il riferimento per inquadrare la funzione della ricerca è quello del cosiddetto *triangolo della conoscenza*, perno della strategia dell'Unione europea per sostenere l'occupazione, le riforme economiche e la coesione sociale nel contesto di un'economia basata sulla conoscenza, adottata nel consiglio europeo di Lisbona nel marzo 2000 (la cosiddetta strategia di Lisbona)¹²⁴, i cui elementi costitutivi sono (i) la ricerca finalizzata allo sviluppo di metodi, tecniche e approcci conoscitivi di comprovata efficienza e qualità, (ii) la formazione sui risultati della ricerca, (iii) l'innovazione, ossia l'introduzione dei risultati all'interno dei processi di produzione.

La ricerca per essere rilevante deve avere ricadute di rilievo nell'attività produttiva, i risultati della ricerca devono essere inseriti nei processi di produzione ed essere quindi funzionali al perseguimento della *mission* di Istituto: la ricerca è quindi rilevante se è motivata e guidata dalle attività svolte nei processi di produzione. Alla luce di quanto sopra, appare necessario che le attività di ricerca vengano opportunamente inquadrare e governate, per assicurare che i risultati delle attività di ricerca, siano essi di carattere metodologico e/o tematico, possano costituire un fattore di miglioramento strutturale e progressivo. Al fine di assicurare che la ricerca condotta in Istituto costituisca un'effettiva attivazione del triangolo della conoscenza,

¹²⁴ Il 23 e 24 marzo 2000, i capi di stato e di governo dei quindici Stati Membri dell'Unione, riuniti in sessione straordinaria a Lisbona, varano la Strategia di Lisbona (Lisbona I). La Strategia è finalizzata a creare i presupposti affinché l'economia europea possa diventare, entro il 2010, l'area più competitiva del mondo, sostenendo l'occupazione, le riforme economiche e la coesione sociale, nel contesto di una "nuova economia" basata sulla conoscenza e sull'investimento in capitale umano

l'Istituto si è dotato di una serie di infrastrutture che consentano di governare, di coordinare e di rendere pertinenti, le proposte di ricerca.

3.10 Le infrastrutture per la ricerca

Il miglioramento sotto il profilo dell'efficienza che il programma di modernizzazione si è proposto di perseguire ha permesso di liberare risorse, umane e monetarie, da valorizzare in nuovi progetti di ricerca. Lo sviluppo di questi progetti necessita di crescente attenzione e impegno: è necessario che l'Istituto sia sempre più attrezzato per affrontare le nuove opportunità offerte dal nuovo modello organizzativo che permette di dedicare maggiori risorse alle attività di ricerca. Una delle innovazioni organizzative di maggiore impatto e che ha richiesto notevoli sforzi da parte di tutte le strutture dell'Istituto, è stata quella legata all'infrastrutturazione delle attività di ricerca. A partire dal 2017, al fine di assicurare la qualità e il coordinamento delle attività di ricerca, in Istituto è stato istituito un sistema di infrastrutture per la ricerca monitorato da un organismo di *governance* - il *Comitato per la Ricerca*, e costituito da due organismi con prevalenti funzioni di indirizzo e sostegno scientifico - il *Comitato Scientifico per la ricerca tematica* e il *Comitato Consultivo per le metodologie statistiche (Advisory Board)*, dai Laboratori Tematici e dal Laboratorio Innovazione.

Il **Comitato per la Ricerca** ha il compito di assicurare la coerenza negli indirizzi e nel coordinamento delle attività di ricerca, tematica e metodologica. Ha funzioni di indirizzo, controllo di coerenza e di uniformità negli approcci delle diverse iniziative di ricerca in Istat, incluso il contributo alla definizione della programmazione strategica e al monitoraggio dell'attività di ricerca svolta in Istituto, e di proposta di specifiche *policy* in ambiti specifici legati alla ricerca.

Il **Comitato Consultivo per le metodologie statistiche (Advisory Board)** ha il compito di fornire sostegno ai progetti di innovazione metodologica dell'Istat, assicurando che essi possiedano le necessarie caratteristiche di qualità, di congruenza e di allineamento con lo stato corrente della ricerca a livello nazionale e internazionale. Il Comitato assicura il

referaggio della componente metodologica dei progetti, sia durante la fase di ideazione, sia in corrispondenza dei principali snodi decisionali del loro processo di realizzazione, svolgendo il tutoraggio di specifici progetti e suggerendo azioni di alta formazione metodologica per il personale Istat coinvolto nei progetti. Le attività del Comitato Consultivo per le Metodologie Statistiche hanno contribuito a rafforzare i progetti di ricerca, consentendone anche la diffusione al di fuori dell'Istat (ad esempio tramite pubblicazioni e/o convegni), la valutazione dei risultati dei progetti da parte del Comitato ha facilitato l'implementazione dei metodi nei processi di produzione correnti dell'Istituto

L'*Advisory Board* è composto da nove membri, tra professori universitari o esperti che hanno prestato servizio presso Istituti nazionali di statistica sia italiani che stranieri che si riuniscono in due *meeting* annuali, in cui sono discussi in sessione plenaria gli aspetti metodologici di specifici progetti strategici dell'Istat, identificati a valle del processo di programmazione annuale e triennale.

I Laboratori per la ricerca tematica sono dedicati alla ricerca economica e ambientale e alla ricerca demografica e sociale, hanno il ruolo di promuovere e coordinare un programma strutturato di attività di ricerca, in accordo con le aree tematiche di interesse per l'Istituto, che determini:

- il miglioramento del livello delle pubblicazioni scientifiche e istituzionali;
- l'arricchimento conoscitivo dei fenomeni, in termini quantitativi e qualitativi, grazie all'integrazione e alla multidisciplinarietà delle analisi;
- il rafforzamento della capacità di gestione e risposta a richieste di analisi tematiche;
- l'interazione tra i ricercatori dell'Istituto valorizzando le competenze specifiche;
- la valorizzazione del capitale umano e di crescita professionale, con impatti positivi in termini di motivazione e crescita della soddisfazione per il lavoro.

Il Laboratorio Innovazione¹²⁵

Il Laboratorio Innovazione (d'ora in poi LabInn nel testo) è l'altra componente del sistema dei laboratori, di cui l'Istat si è dotato per migliorare la propria capacità di innovare, nei processi e nei prodotti, in modo da rispondere in maniera efficace all'evoluzione della domanda di informazione statistica.

Il LabInn offre l'opportunità di dedicare del tempo alla ricerca e mette a disposizione uno spazio fisico e infrastrutture informatiche di elevate prestazioni, utili a testare le proprie idee in uno spazio dedicato. L'Istat ha effettuato un grosso investimento in termini tecnologici per garantire al LabInn un'infrastruttura informatica più potente rispetto a quella usata nei processi di produzione statistica corrente

In particolare, il LabInn offre la possibilità a *team* di ricercatori di sperimentare idee e progetti innovativi che afferiscono principalmente ai seguenti ambiti:

- utilizzo di nuove fonti di dati,
- miglioramento dei processi statistici,
- *output* innovativi: nuove tecniche di navigazione, scoperta e visualizzazione dell'informazione, combinazione tra diverse sorgenti di dati, *open data*, *linked open data*;
- utilizzo di nuove tecnologie e metodologie ICT.

Il LabInn è stato inaugurato nel marzo 2018 e ha permesso lo sviluppo di diversi progetti interdisciplinari che hanno coinvolto decine di colleghi appartenenti a diverse strutture organizzative e che hanno dato all'Istituto un importante contributo in termini di innovazione e di miglioramento dei processi di produzione statistica.

L'interdisciplinarietà dei team e la trasversalità dei progetti caratterizzano i gruppi che hanno lavorato e che lavorano allo sviluppo dei singoli progetti, in perfetta coerenza con quanto stabilito dall'articolo 3 dello Statuto dell'Istat che ci ricorda che "...l'attività di ricerca si realizza

¹²⁵ <https://www.istat.it/it/ricerca-in-istat/organizzazione/laboratorio-innovazione#:~:text=Il%20Laboratorio%20per%20l'Innovazione,e%20dell'innovazione%20in%20Istituto.>

attraverso azioni programmate, alle quali cooperano in modo integrato gruppi di ricercatori e tecnologi con competenze e professionalità differenti”¹²⁶.

Il funzionamento del LabInn, così come delle altre infrastrutture per la ricerca, è monitorato dal Comitato per la Ricerca, che sceglie anche i progetti da ammettere al laboratorio dopo un iter di valutazione che coinvolge vari settori dell’Istituto e che tiene conto soprattutto dell’impatto dei progetti sul miglioramento della produzione statistica e della coerenza di ciascun progetto con la *mission* dell’Istituto.

L’Istat, inoltre, in coerenza con altre esperienze di eccellenza in altri Istituti di Statistica europei e internazionali, ha anche garantito uno spazio fisico dedicato alle attività di sviluppo dei progetti innovativi rafforzando il ruolo della ricerca come valore fondante e strumento di crescita strategica dell’Istituto e del personale. La ricerca garantita dal LabInn è, così come previsto dall’art. 2 dello statuto dell’Istat, *“finalizzata al miglioramento della qualità delle informazioni statistiche e dei processi adottati per la produzione, sviluppo e diffusione della statistica ufficiale e all’introduzione nei processi suddetti dei risultati della ricerca metodologica e tematica”*.

Il LabInn e le altre infrastrutture per la ricerca rappresentano una realtà consolidata in Istituto, strumenti capaci di favorire ed agevolare la costituzione di una rete di conoscenza, ricerca e innovazione che contribuisce a rafforzare il circuito virtuoso in cui l’Istituto è inserito grazie ad accordi partnership, convenzioni con altri enti e organismi di ricerca a livello nazionale ed internazionale.

¹²⁶ Statuto dell’Istat approvato dal Consiglio dell’Istituto con deliberazione n. CDXLIV del 7 dicembre 2017 e modificato con deliberazione n. CDXCV del 13 dicembre 2019. <https://www.istat.it/it/files/2017/02/STATUTO-ISTAT.pdf>

Allegato n. 3 Il Modello di Business Architecture dell'Istat

Strategia									
Definisci visione e obiettivi		Governa		Orienta e collabora					
<ul style="list-style-type: none"> • Interpreta le direttive e i fattori chiave nazionali e internazionali • Definisci la visione e i valori dell'Istituto • Adatta il valore e le strategie dell'Istituto sulla base delle aspettative degli utenti • Definisci gli obiettivi dell'Istituto • Comunica valori e obiettivi 		<ul style="list-style-type: none"> • Sviluppa le strategie per raggiungere gli obiettivi dell'Istituto • Definisci le priorità del <i>portfolio</i> statistico • Definisci le priorità del <i>portfolio</i> delle capacità • Alloca le risorse (budget) • Sviluppa e sostieni l'eccellenza statistica e professionale all'interno 		<ul style="list-style-type: none"> • Sviluppa e sostieni le relazioni strategiche nazionali e internazionali • Sviluppa e sostieni l'eccellenza statistica all'esterno • Promuovi le collaborazioni internazionali e tra Enti • Garantisci le risorse per il <i>portfolio</i> statistico e delle capacità 					
Capacità				Supporto generale					
Pianifica il miglioramento delle capacità	Sviluppa le capacità	Gestisci le capacità	Supporta l'utilizzo delle capacità	Gestisci le attività e la performance	Gestisci le finanze	Gestisci le risorse umane	Gestisci l'IT	Gestisci l'informazione e le conoscenze	Gestisci i fornitori di dati e gli utenti
<ul style="list-style-type: none"> • Identifica i miglioramenti delle capacità, ad alto impatto o di altro tipo • Proponi progetti di miglioramento delle capacità, incluse le infrastrutture condivise • Gestisci programmi di miglioramento delle capacità 	<ul style="list-style-type: none"> • Conduci attività preliminari di ricerca • Definisci i requisiti dettagliati delle singole capacità • Progetta le soluzioni relative alle singole capacità • Sviluppa e diffondi le soluzioni relative alle singole capacità, incluse le infrastrutture condivise • Gestisci progetti di sviluppo delle singole capacità 	<ul style="list-style-type: none"> • Cura la manutenzione delle capacità, incluse le infrastrutture condivise • Promuovi l'utilizzo delle capacità • Valuta l'adeguatezza delle capacità 	<ul style="list-style-type: none"> • Promuovi l'utilizzo delle capacità in fase di progettazione • Supporta l'utilizzo delle capacità in fase di produzione • Supporta l'utilizzo delle capacità anche all'esterno dell'Istituto 	<ul style="list-style-type: none"> • Gestisci la performance delle attività • Gestisci il cambiamento • Gestisci il quadro normativo e la conformità a questo • Gestisci i beni materiali, incluse le infrastrutture logistiche 	<ul style="list-style-type: none"> • Gestisci la contabilità • Gestisci appalti e contratti 	<ul style="list-style-type: none"> • Gestisci la performance delle risorse umane • Gestisci e sviluppa le competenze • Valorizza i talenti • Gestisci il reclutamento • Pianifica il fabbisogno 	<ul style="list-style-type: none"> • Gestisci i servizi IT • Gestisci la sicurezza dei sistemi IT e dell'informazione 	<ul style="list-style-type: none"> • Gestisci la documentazione e gli archivi • Gestisci le conoscenze • Gestisci gli standard e i diritti dell'informazione 	<ul style="list-style-type: none"> • Gestisci le relazioni pubbliche • Gestisci le relazioni con i media • Consulta gli stakeholder • Gestisci il supporto agli utenti
Produzione									
Sviluppa			Realizza						
Specifica le esigenze	Progetta	Costruisci	Raccogli	Elabora	Analizza	Diffondi			
<ul style="list-style-type: none"> • Identifica le esigenze • Consulta gli stakeholder e conferma le esigenze • Definisci gli obiettivi di output • Identifica i concetti • Verifica la disponibilità dei dati • Prepara il <i>business case</i> 	<ul style="list-style-type: none"> • Progetta gli output • Definisci le variabili • Progetta la fase di raccolta • Progetta la lista di campionamento e il campione • Progetta le fasi di elaborazione e di analisi • Progetta i sistemi di produzione e i <i>workflow</i> 	<ul style="list-style-type: none"> • Assembla e configura le componenti di sistema • Configura i <i>workflow</i> • Effettua il test del sistema di produzione • Effettua il test del processo di produzione statistico • Perfeziona il sistema di produzione 	<ul style="list-style-type: none"> • Crea la lista di campionamento e seleziona il campione • Organizza la raccolta • Esegui la raccolta • Completa la raccolta 	<ul style="list-style-type: none"> • Integra i dati da più fonti • Classifica e codifica • Controlla e convalida • Correggi e imputa i dati mancanti • Deriva nuove variabili e unità • Calcola i pesi • Calcola gli aggregati • Produci i file di microdati validati 	<ul style="list-style-type: none"> • Prepara gli output preliminari • Convalida gli output • Interpreta e spiega gli output • Applica tecniche di protezione della riservatezza • Perfeziona gli output 	<ul style="list-style-type: none"> • Aggiorna i sistemi di output • Metti a punto i prodotti di diffusione • Gestisci il rilascio dei prodotti di diffusione • Promuovi i prodotti di diffusione 			
Gestisci									
Pianifica			Monitora			Correggi			
<ul style="list-style-type: none"> • Garantisci l'approvazione e il finanziamento del progetto • Pianifica le attività di progetto, la tempistica, il budget e le risorse • Pianifica le metriche e gli obiettivi di qualità e performance 			<ul style="list-style-type: none"> • Monitora qualità e performance del progetto • Monitora budget e rispetto delle scadenze • Identifica rischi e problemi emergenti • Riporta gli stati di avanzamento del progetto 			<ul style="list-style-type: none"> • Sviluppa azioni e strategie di correzione • Aggiorna il piano del progetto • Comunica le azioni di correzione e i nuovi risultati attesi 			

Fonte: Istat nell'ambito dello Statistical Network Business Architecture Project Team, 2014

Capitolo 4 - Verso le Trusted Smart Statistics: i risultati dell'indagine tra gli Istituti Nazionali di Statistica all'interno del sistema statistico europeo.

4.1 La Survey

Nel precedente capitolo abbiamo ripercorso le azioni messe in campo dagli Istituti di Statistica in Europa per cogliere le opportunità messe a disposizione dalle nuove fonti di dati, aumentare la rilevanza e la reputazione della statistica ufficiale, rispondere alle crescenti esigenze degli *stakeholders* e arricchire la loro offerta informativa in termini di qualità, tempestività e granularità territoriale e capacità di cogliere nuovi fenomeni non misurabili con le indagini tradizionali.

È emerso che gli Istituti di Statistica Europei si trovano in un momento di passaggio da una fase di *"playground"*, necessaria a capire le potenzialità e i limiti delle fonti big data e a studiare i metodi necessari a trattarle, ad una fase di utilizzo maturo di tali fonti con l'individuazione di nuovi prodotti: le *Trusted Smart Statistics* che rappresentano quindi la naturale evoluzione delle statistiche ufficiali nel nuovo mondo dei dati. Abbiamo già detto che loro realizzazione comporta cambiamenti radicali nel complessivo sistema di funzionamento degli Istituti Nazionali di Statistica e nelle loro relazioni con utenti e partner esterni in quanto il cambiamento non ha riguardato solo le fonti dati ma il contesto generale in cui gli Istituti si muovono e che è stato descritto nel primo capitolo.

Per capire a che punto fossero gli INS è stato realizzato un questionario (Allegato n. 4) con l'obiettivo di raccogliere informazioni utili a descrivere la fase di passaggio appena citata tenendo conto degli investimenti fatti dagli Istituti per accogliere e valorizzare i cambiamenti di contesto, quelli programmati e quelli auspicati per portare l'Istituto per andare verso un uso più maturo e strutturato dei big data, verso il sistema di produzione delle *Trusted Smart Statistics*.

Il questionario è stato somministrato a un campione di Istituti di Statistica facenti parte del Sistema Statistico Europeo.

4.1.1 Istituti coinvolti

Il campione di riferimento è costituito dagli Istituti Nazionali di Statistica coinvolti nei progetti ESSnet Big Data I e II. I progetti ESSnet sono progetti promossi da Eurostat, l'ufficio statistico dell'Unione Europea, per favorire la cooperazione, in termini di competenze ed esperienze, fra gli Istituti Nazionali di Statistica degli Stati membri. Gli ESSnet sui Big Data hanno avuto come specifico obiettivo l'implementazione di iniziative che prevedessero l'integrazione dei big data nella produzione di statistiche ufficiali, attraverso progetti pilota che hanno, in prima battuta, esplorato il potenziale delle fonti di big data selezionate, per arrivare alla costruzione e all'implementazione di applicazioni concrete. I progetti cominciati nel 2016 e proseguiti per tutto il 2021, in applicazione del Memorandum di Bucharest, hanno coinvolto in diversa misura 28 Istituti nazionali di statistica (cfr. Capitolo 2).

4.2 Lo strumento di rilevazione: struttura

Il questionario, composto complessivamente da 23 domande, ha l'obiettivo di fare il punto sullo stato dell'arte delle attività in corso e degli investimenti programmati per promuovere l'integrazione dei big data all'interno dei processi di produzione statistica dei singoli Istituti coinvolti. Al fine di avere una panoramica complessiva sui risultati raggiunti e su quelli programmati, il questionario indaga le diverse dimensioni della struttura organizzativa e mira a rilevare investimenti già fatti, investimenti programmati e investimenti auspicati per ciascuno dei seguenti settori:

- ✓ metodologico,
- ✓ informatico,
- ✓ organizzativo,
- ✓ risorse umane,
- ✓ comunicazione e diffusione,
- ✓ legale.

I settori organizzativi coinvolti rispecchiano le diverse dimensioni che sono state descritte nei capitoli precedenti come leve che compongono un sistema complesso e su cui investire parallelamente per l'implementazione dei nuovi sistemi di produzione statistica. I settori oggetto di questa indagine sono gli stessi che abbiamo riscontrato nel *Big Data Action Plan and Roadmap* elaborata nel 2018 in ambito Eurostat a valle dell'approvazione del Memorandum di Bucarest e nella *Roadmap Istat* verso le *Trusted Smart Statistics 2021-2024* elaborata in Istat nel 2020. Questa scelta trova riscontro in quanto abbiamo specificato nei capitoli precedenti, i nuovi dati non si riducono solo ad un problema di tecnologie e di risorse computazionali, ma investono tutte le componenti delle strutture organizzative secondo il concetto di sostenibilità organizzativa che abbiamo introdotto nelle pagine precedenti. L'esperienza già vissuta all'interno dei progetti europei, nonché la specifica esperienza in Istat, hanno dimostrato che incidere su una sola leva della struttura organizzativa non aiuta gli Istituti a rilasciare *output* statistici innovativi. È necessario investire su ciascuna componente della struttura organizzativa per poter sostenere l'avvio di un nuovo paradigma di produzione. Il questionario indaga i vari aspetti soprariportati perché, come è stato già detto, quello delle TSS, è un sistema complesso che comporta interazioni importanti tra le diverse dimensioni che lo compongono. Non si può quindi operare su una singola leva senza che questa operazione comporti cambiamenti anche sulle altre. Di qui la necessità di raccogliere informazione non solo su aspetti puramente tecnologici, le TSS non dipendono solo da infrastrutture informatiche particolarmente performanti, ma rappresentano un cambiamento culturale e chiamano in causa tutte le dimensioni analizzate. Il sistema TSS sarà realizzato in tutti gli Istituti quando tutte le dimensioni dovranno evolvere parallelamente, alla stessa velocità e nella stessa direzione in coerenza le une con le altre. Gli Istituti di statistica devono ridefinire infrastrutture organizzative, metodologiche, informatiche per continuare a fornire informazione statistica di altissima qualità sfruttando la potenza informativa dei big data.

Il questionario è strutturato in quattro parti:

1. la prima parte raccoglie informazioni sulle caratteristiche professionali del rispondente e sugli orientamenti generali degli istituti sull'uso delle nuove fonti.

2. La seconda parte ha l'obiettivo di analizzare lo stato dell'arte su (i) le sperimentazioni attuate e su quelle in fase di implementazione, (ii) gli investimenti fatti dagli Istituti intervistati nei diversi settori coinvolti, non solo in termini di metodi e strumenti, infrastrutture informatiche (*hard guarantees*) ma anche in termini organizzativi e di *governance* (*soft guarantees*), per favorire l'uso delle nuove fonti di dati nella produzione statistica. La divisione tra aspetti *hard* e *soft* ha guidato, come abbiamo visto, anche l'implementazione del nostro *framework* di monitoraggio presentato nel capitolo precedente.
3. La terza parte del questionario, si focalizza su un'analisi delle percezioni, su quegli investimenti da attuare in termini organizzativi ed infrastrutturali necessari a favorire la produzione di TSS. Si chiede ai ricercatori coinvolti direttamente sui processi indagati, quali siano, a loro parere, gli investimenti necessari e più funzionali a favorire uno sfruttamento più efficiente dei big data e a valorizzarne l'uso nella produzione statistica corrente.
4. La sezione *Data and tools used* è finalizzata a raccogliere informazioni relative alla tipologia di dati usati per ottenere i risultati raggiunti, agli strumenti alle collaborazioni necessari a raggiungerli.

I risultati della *survey* ci restituiscono quindi da una parte un'analisi dell'*as-is*, una fotografia del posizionamento dei singoli istituti nel percorso verso la produzione più strutturata di *output* statistici con l'uso di big data e dall'altra un possibile scenario del *to-be*, grazie alle segnalazioni e alle indicazioni che i ricercatori maggiormente coinvolti nei processi studiati, riportano nei loro contributi. La sezione n. 2, finalizzata all'analisi dell'*as-is* e la sezione n. 3 finalizzata invece all'analisi del *to-be* ci aiutano quindi a misurare il *gap* da colmare per raggiungere l'obiettivo e permettere ai singoli istituti di realizzare i nuovi sistemi di produzione delle TSS. Dal confronto dei risultati delle due sezioni emergono utili suggerimenti sugli investimenti necessari a consolidare i risultati già raggiunti e mettere in condizioni gli Istituti di realizzare i nuovi sistemi di produzione, che sono sintetizzati nelle conclusioni di questo capitolo.

Il questionario è stato implementato sulla piattaforma Google Moduli, per favorire la facilità di compilazione e agevolare la raccolta dei risultati. Ai rispondenti è stata offerta una duplice possibilità di risposta: sia *on line* sulla piattaforma Google Moduli – a cui accedevano tramite un link contenuto nella mail di presentazione della *survey* (riquadro n. 1), sia compilando la versione in formato *word* inviata in allegato alla mail di presentazione. La versione in formato *word* si è resa necessaria nei casi in cui il ricercatore coinvolto in prima persona, non essendo a conoscenza di tutte le informazioni richieste nelle sezioni del questionario, ha avuto la necessità di uno strumento più funzionale alla raccolta dei dati presso le diverse strutture organizzative del proprio ente di appartenenza.

Nella mail di presentazione riportata nel riquadro 1, oltre al link per la compilazione del questionario e/o al questionario in formato *word*, vengono riportate informazioni relative a: (i) il proponente del progetto di ricerca cui il questionario fa riferimento si inserisce. Nella mail si segnala quindi che il questionario è finalizzato anche alla redazione di una Tesi di Dottorato svolta in collaborazione tra l'Università La Sapienza, Dipartimento Coris "Comunicazione Ricerca Sociale e Marketing" e la Direzione Metodologica dell'Istat presso cui la proponente svolge la propria attività lavorativa, (ii) il tema di ricerca e i suoi macroobiettivi, (iii) l'obiettivo del questionario con il link al modulo online; (iv) la richiesta di materiale utile all'approfondimento del tema indagato, (v) le motivazioni per le quali il rispondente è stato direttamente coinvolto nell'indagine.

Dear colleague,

I work at the Italian Institute of Statistics (ISTAT) and, at the same time, I'm attending a PhD at La Sapienza University of Rome.

With the tutoring of ISTAT, I'm working on a research project on the impact of Big Data on Official statistics and on the production of Trusted Smart Statistics (TSS). Part of my research concerns the comparative analysis between the different approaches and tools used and the results obtained in European countries with the integration of Big Data in statistical production processes. I'm writing to you because your statistical office has participated in ESSnet Big Data, so I believe that your contribution is essential to deepen the topic I'm studying.

I ask you to fill in the questionnaire that you find at this link <https://docs.google.com/forms/d/1bkxguKmj7-qdgwGaYSk-DzRFsUPMYe5HrXERpm81-Ss/edit> and to send me any material (links, reports, documents ...) that can help me to have a more complete vision of the path that INS are generally taking to improve TSS production. As you can see, the questionnaire focuses on three levels: (i) the investments already made by your institution, (ii) those planned and (iii) those that you think could improve the TSS production.

Your contribution will be valuable, the link will be active for three weeks until October 30th, 2021. If you aren't the person who can answer the questionnaire, please forward it to a colleague who can, this will help me get a more complete picture of my research field.

Your contribution will be valuable, the link will be active for three weeks until October 29th, 2021.

Thanks for your very important help!

Greetings

Riquadro n.1 Mail di presentazione del questionario

Il link al questionario, attivo dal 15 settembre 2021 per 5 settimane, contenuto nella mail di presentazione, rimanda in prima battuta, ad una pagina - riportata nel riquadro n. 2 – con un ulteriore approfondimento dello strumento di rilevazione, del tema di ricerca e della finalità

del questionario. La pagina sottolinea la partecipazione al progetto da parte della DCME Direzione per la metodologia e il disegno dei processi statistici dell'Istat.

Dear colleague,

This questionnaire is an integral part of a Doctoral Thesis at La Sapienza University of Rome, Department of Communication and Social Research (CoRiS), in collaboration with the Directorate for Methodology and Design of Statistical Processes of the National Institute of Statistics (ISTAT). The cognitive objective of the research project is to detect the contribution provided by Big Data to the improvement of statistical production, the development of Trusted Smart Statistics in European Statistical Institutes, as well as to the investments already made, and those in the implementation phase who develop statistical products with the use of new data sources. Therefore - considering your specific experience on the subject - your contribution, together with that of other representatives of the National Statistical Institutes in Europe, will offer useful information for the development of a comparative analysis between European countries in order to highlight best practices and possible suggestions on future developments. We thank you for the precious help you will want to give to the research and we ask you to answer all the questions in order to provide us, in each of the investigated areas, with a more complete view on the results being achieved and the tools used to reach them, together with the strategic objectives of long term shared by the Board of your Institute. Where possible, please report any links to useful materials for further information on the individual items investigated. All data will be processed exclusively for scientific research purposes and in compliance with current legislation on *privacy*.

We appreciate and thank you for your collaboration. Greetings

Riquadro n. 2 – Presentazione questionario

4.3 Struttura del Questionario

4.3.1 I rispondenti

Il questionario è stato inviato a ricercatori opportunamente selezionati tra i referenti dei progetti ESSnet Big Data I e/o II e maggiormente coinvolti nelle attività legate alla sperimentazione e alle prime implementazioni di prodotti statistici con l'uso dei big data. I ricercatori individuati hanno avuto e hanno esperienze di lavoro dirette nell'ambito indagato, collaborano con il *board* che coordina le attività legate alla ricerca e all'innovazione e sono stati in qualche modo coinvolti anche nella produzione delle prime statistiche sperimentali rilasciate. I rispondenti, quindi, oltre ad avere competenze di tipo tecnico legate al settore indagato grazie alle esperienze maturate in ambito nazionale ed internazionale, hanno anche una visione di insieme sulla strategia adottata e da perseguire dall'Istituto di appartenenza perché conoscono il contesto europeo più in generale e gli orientamenti del SSE in questo ambito. I ricercatori, al fine di compilare nella maniera più esaustiva e corretta il questionario, hanno a loro volta coinvolto nella compilazione altri colleghi meglio inseriti in specifici settori, in modo da fornire risposte più complete e esaustive.

I questionari sono stati inviati ai ricercatori dei seguenti paesi:

1. Austria
2. Belgio
3. Bulgaria
4. Cipro
5. Croazia
6. Danimarca
7. Estonia
8. Finlandia
9. Francia
10. Germania
11. Gran Bretagna
12. Grecia
13. Irlanda
14. Lettonia
15. Lituania

16. Lussemburgo
17. Norvegia
18. Olanda
19. Polonia
20. Portogallo
21. Repubblica Ceca
22. Romania
23. Slovacchia
24. Spagna
25. Svezia
26. Svizzera
27. Ungheria

Hanno risposto

1. Bulgaria
2. Ireland
3. Estonia
4. The Netherlands
5. Slovenia
6. Slovakia
7. Sweden
8. Austria
9. France
10. France
11. Romania
12. Finland
13. Germany
14. Germany Other National Authority
15. United Kingdom
16. Italia

4.3.2 Prima sezione

La prima parte del questionario raccoglie informazioni generali sui rispondenti e sulle funzioni svolte nell'Istituto di appartenenza.

Come si può vedere dalla figura n. 1 il la maggior parte dei rispondenti dichiara di svolgere funzioni riconducibili alla *data science*, o funzioni metodologiche, o funzioni di *data analyst*, pochi i ricercatori impegnati anche in settori informatici e statistici, tra i rispondenti troviamo *data engineer*, *Research and development project manager*, un dirigente di struttura e un dirigente del Dipartimento di Ricerca e Sviluppo.

Emerge l'assenza di esperti di dominio, questo rafforza quanto già detto relativamente alla necessità di un approccio multidisciplinare nell'uso e nell'analisi delle nuove fonti dati. Deve anche far riflettere l'assenza tra i rispondenti di esperti di comunicazione e diffusione, questo mette una assoluta necessità di affiancare a *data scientist* anche esperti di comunicazione e diffusione in grado di valorizzare e divulgare i nuovi *output* statistici e ribadire il ruolo degli INS come attori istituzionali nell'ambito della divulgazione e comunicazione scientifica. Questa necessità emergerà, come vedremo, anche in altre sezioni del questionario.

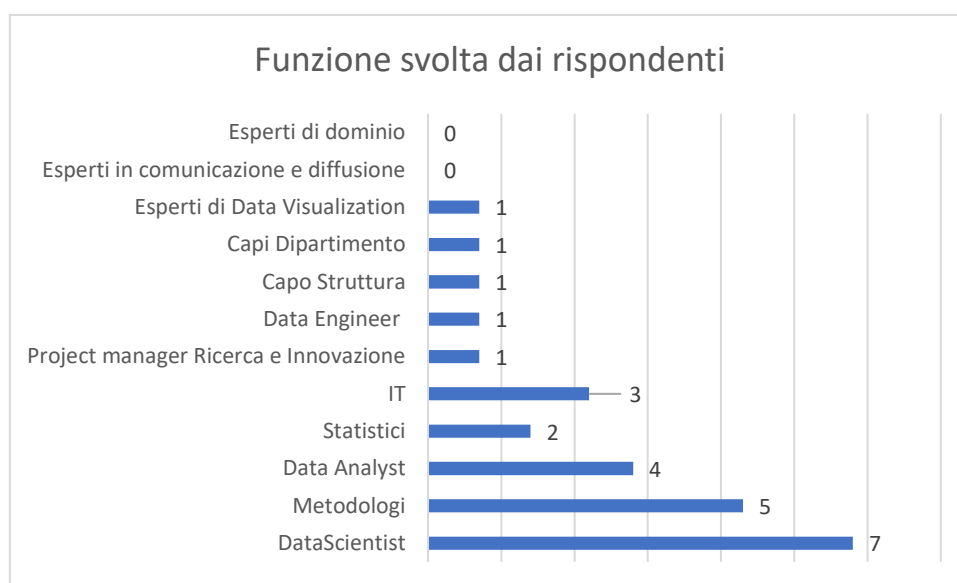


Grafico 1 - funzione svolta dai rispondenti

In questa fase di passaggio da una fase sperimentale ad una fase di produzione vera e propria, non viene ancora colta l'opportunità di coinvolgere anche il settore comunicazione e diffusione, come vedremo in seguito, sono anche altri i settori non ancora inseriti nei processi di trasformazione del *business model* degli Istituti. Ciò crea un disallineamento e un disequilibrio tra le varie componenti organizzative degli INS.

Differenti sono le strutture di appartenenza dei rispondenti, questo è dovuto chiaramente alla diversa articolazione organizzativa dei singoli istituti, si registra comunque una maggiore concentrazione dei rispondenti nei settori dedicati alla ricerca e all'innovazione e nelle strutture metodologiche, questo è dovuto al fatto che si tratta di settori trasversali caratterizzati da interazioni molto forti con gli altri settori sia trasversali sia di produzione all'interno degli INS.

Alla domanda n. 7 "*Il tuo Istituto ha fatto investimenti per cogliere le opportunità offerte dai Big Data*", tutti gli Istituti coinvolti nell'indagine rispondono positivamente. Alcuni hanno dato avvio al processo di integrazione dei big data nelle fasi di produzione statistica, fin dal 2010/2011 seppure in maniera episodica e frammentaria. I progetti ESSnet Big Data I e II hanno reso più strutturali queste iniziative facilitando la costituzione di team dedicati a specifici ambiti e tematiche di interesse europeo. In questo modo ciascun istituto è riuscito ad allinearsi alle politiche degli altri Istituti di statistica europei e a dare maggiore coerenza complessiva a sforzi e investimenti.

La partecipazione ai progetti europei ha spinto gli istituti coinvolti a dotarsi di una *governance* in grado di progettare e implementare azioni sistematiche e strutturate idonee a indirizzare e coordinare le attività finalizzate all'integrazione dei big data nei processi di produzione statistica, la metà degli Istituti si è dotato di un organismo di *governance* dedicato alle nuove funzioni. Da questo primissimo dato, emerge la volontà delle organizzazioni di rendere compatibile la propria struttura organizzativa col il nuovo modello di produzione.

4.3.3 Seconda Sezione – Investimenti realizzati

Dopo una prima parte più generale, la seconda sezione contiene domande che scendono ad un livello di dettaglio più granulare per entrare maggiormente nel merito del tema indagato. Le domande sono più specificamente relative agli investimenti che l'Istituto ha già fatto all'interno dei diversi settori coinvolti nei processi di produzione statistica con l'uso delle nuove fonti.

I rispondenti sono chiamati quindi a segnalare, per ciascuno dei settori preso in considerazione, gli investimenti già implementati in modo da fornire un complessivo stato dell'arte nei diversi istituti.

La domanda è aperta, quindi per ciascun settore il rispondente può inserire le informazioni che ritiene più opportune e, come espressamente richiesto, può allegare link e/o documentazione utile ad approfondimenti. Nello specifico la domanda è la seguente: *“Quali sono gli investimenti che il tuo istituto ha già sostenuto per avviare la produzione di TSS negli ambiti indicati?” (Puoi fornire, laddove disponibili, link e /o informazioni utili per ciascuno dei settori?)*

Alla domanda segue la specifica dei settori indagati, relativi a:

1. Metodologia;
2. Informatica;
3. Risorse umane;
4. Organizzazione,
5. Comunicazione e diffusione,
6. Settore legale.

Di seguito sono riportati i risultati raccolti per singola dimensione.

➤ **Settore metodologico**

La partecipazione degli Istituti ai progetti promossi da Eurostat diventa per molti dei rispondenti il momento in cui si sperimentano nuove soluzioni per l'implementazione di nuovi paradigmi metodologici in grado di sostenere i processi di integrazione delle nuove fonti, gli ESSnet Big Data I e II si caratterizzano per la forte collaborazione con Università, Accademie, altri enti di ricerca ed enti privati fornitori di dati che facilitano sperimentazioni di nuovi metodi e modelli.

Altri investimenti riguardano la formazione dei ricercatori su temi legati alla Data Science e all'analisi dei Big Data, anche attraverso la partecipazione a corsi ESTP¹²⁷. Più in generale i rispondenti indicano la partecipazione a progetti nazionali e/o europei come un contributo all'avanzamento del settore metodologico del proprio istituto. Alcuni fanno riferimento a specifici *pilot track* sviluppati all'interno dei WP degli ESSnet Big Data I e II legati a *scanner data*, all'indice dei prezzi, all'acquisizione di dati dalle imprese e così via.

In generale, ad esclusione di due istituti che affermano di non aver sostenuto nessun tipo di investimento in ambito metodologico, dalle risposte fornite emergono diverse iniziative messe in campo per sfruttare tecnologie, piattaforme, collaborazioni e per proseguire e valorizzare e consolidare quanto già fatto in ambito europeo. L'Istituto di Statistica del Regno Unito, *Office for national Statistics*, (d'ora in poi ONS nel testo) segnala una sezione del sito¹²⁸ chiamata *Data Science for Public Good*, costantemente in aggiornamento, che contiene informazioni su nuovi strumenti e tecnologie usate per sfruttare la disponibilità delle nuove sorgenti di dati e dei metodi innovativi implementati per trattarle al fine di fornire misurazioni e analisi sull'economia, l'ambiente e la società in generale.

L'esperienza dell'ONS è assolutamente un'esperienza virtuosa di gestione dei dati come bene pubblico. Nel presente capitolo alcuni aspetti segnalati dai ricercatori ONS saranno approfonditi e graficamente inseriti in riquadri per dare loro una maggiore visibilità. Questa scelta è dovuta al fatto che l'organizzazione e le pratiche all'interno dell'ONS rappresentano la realizzazione del disegno teorico che abbiamo delineato nei capitoli precedenti attraverso una serie di *best practices* che hanno garantito il raggiungimento di risultati rilevanti in termini di qualità delle informazioni e di innovazione di processo e di prodotto.

¹²⁷ I corsi ESTP fanno parte del programma europeo di formazione statistica (*European Statistical Training Program ESTP*) che offre agli statistici europei l'opportunità di partecipare a corsi di formazione internazionali, *workshop* e seminari a livello post-laurea. Il programma, coordinato da Eurostat nasce per soddisfare le esigenze specifiche del sistema statistico europeo e per cercare di allineare i diversi livelli di conoscenza statistica tra i ricercatori che afferiscono agli INS.

¹²⁸ <https://datasciencecampus.ons.gov.uk/>

Il Data Science Campus

All'interno dell'ONS già dal 2017 è stato istituito il *Campus Data Science* che coinvolge un nucleo di ricercatori qualificati appartenenti ad una solida rete di ricerca proveniente dal mondo istituzionale, governativo e accademico. Il Campus è un organismo intergovernativo che ha l'obiettivo di indagare sull'uso di nuove fonti di dati, compresi i dati amministrativi e i big data per il bene pubblico. Sviluppa progetti per l'ONS, il governo e le più importanti organizzazioni britanniche e internazionali per supportare i processi di *decision making* e fornire informazioni rilevanti ed aggiornate per l'ONS e altri *stakeholders* su vari ambiti tematici (per es. economia, il commercio, l'ambiente e la società...) per guidare l'innovazione per il bene pubblico.

Il caso dell'ONS e del suo *Hub* sulla Data Science rappresenta un esempio virtuoso in tema di gestione dei dati come bene pubblico. L'*hub* attraverso l'integrazione tra *data science* e Intelligenza Artificiale, costruisce competenze, applicando strumenti, metodi e pratiche, per creare nuove conoscenze e migliorare il processo decisionale per il bene pubblico. Il ruolo della *data science* all'interno dell'ONS è perfettamente in linea con quanto abbiamo segnalato in più parti nel presente lavoro, ossia rappresenta l'applicazione di strumenti, metodi e le pratiche dell'era digitale e dei dati per creare una nuova comprensione e migliorare il processo decisionale per diverse tipologie di *stakeholders*.

➤ Settore informatico

Gli investimenti informatici maggiormente segnalati sono tutti finalizzati al rafforzamento delle infrastrutture già presenti negli Istituti, in alcuni casi, all'acquisizione di nuove piattaforme (tipo Cloudera). In linea generale, l'analisi delle risposte evidenzia che gli investimenti in ambito informatico sono finalizzati a dotare gli istituti di infrastrutture IT maggiormente performanti e in grado di supportare l'analisi statistica di grandi set di dati.

Altri investimenti riguardano la progettazione di un server RStudio per utilizzare il linguaggio di programmazione R¹²⁹ in modo più efficace e ottenere una maggiore potenza di calcolo, la sostituzione di piccoli server di *web scraping* con un'infrastruttura di *web scraping* dedicata con un *server* di sviluppo e uno di produzione. In un caso sono stati fatti investimenti per data *warehouse* statistico finalizzato a migliorare la documentazione e l'accessibilità di tutte le fonti di dati (compresi i big data) all'interno dell'istituto.

L'ONS anche in questo caso riporta la sua esperienza virtuosa e presenta i suoi investimenti in ambito informatico, inserendoli nel contesto di una *vision* di più ampio respiro, facendo riferimento all'*Integrated Data Service* (IDS), un'infrastruttura sicura e affidabile, implementata per aumentare la condivisione dei dati con il governo, consentire l'analisi a sostegno della crescita economica e garantire servizi pubblici efficienti per il miglioramento della vita dei cittadini.

➤ **Settore comunicazione e diffusione**

Alla domanda su quali siano gli investimenti sostenuti dall'Istituto in ambito comunicazione e diffusione, la metà degli istituti dichiara di non aver avviato alcuna attività in questo campo. Gli investimenti indicati riguardano nella maggior parte dei casi la pubblicazione sul sito web istituzionale delle statistiche sperimentali, segnale questo che le statistiche sperimentali rappresentano non solo una nuova fonte di informazione ma anche uno strumento per comunicare risultati ottenuti con le nuove fonti dati. Anche i seminari vengono segnalati, insieme a strumenti di *data visualization*, come mezzi di diffusione di risultati. Dalle risposte date però, e soprattutto dalle mancate risposte, emerge con chiarezza che rispetto agli altri settori, quello comunicazione e diffusione è l'ambito in cui sono stati fatti meno investimenti, questo elemento è un chiaro segnale che manca ancora la giusta consapevolezza della cruciale importanza che riveste la comunicazione nella valorizzazione di questi prodotti e nel coinvolgimento del cittadino. I big data appartengono ai cittadini e in un'ottica di

¹²⁹ R un linguaggio di programmazione specifico per l'analisi statistica dei dati e la grafica computazionali

Citizen science ai cittadini dovrebbero essere “restituiti” sotto forma di informazione statistica.

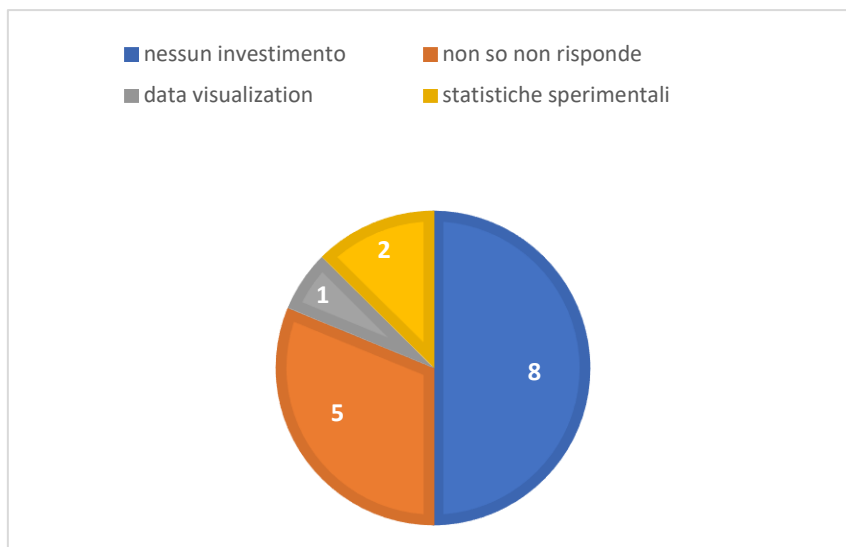


Grafico 2 - Investimenti nel settore comunicazione e diffusione

La sfida della comunicazione è enorme. Occorre ripensare ad un nuovo paradigma, a nuovi strumenti: non possiamo comunicare i nuovi prodotti con gli strumenti “vecchi”. In primo luogo, gli istituti di statistica hanno bisogno di nuovi approcci di visualizzazione per la rappresentazione interattiva e multidimensionale di dati estremamente complessi con minori perdite di informazioni utili e con maggiori possibilità dell'esplorazione dei dati. Le TSS hanno bisogno da un lato di complesse analisi di dati che si basano su una conoscenza estremamente avanzata in matematica e statistica, e dall'altro della capacità di progettare e realizzare nuovi strumenti per la comunicazione visiva dei risultati. Esperti provenienti da mondi completamente diversi devono lavorare fianco a fianco per cogliere la sfida della visualizzazione dei big data¹³⁰ e arrivare a quello che John Tukey ha scritto già nel 1977: *il più grande valore di un'immagine è quando ci costringe a notare ciò che non ci saremmo mai aspettati di*

¹³⁰ Secchi P., *On the role of statistics in the era of big data: a call for a debate* – novembre 2017, Elsevier

vedere” Si possono costruire delle macchine molto efficienti nel processare questi dati, ma è molto importante occuparsi del momento della comunicazione per chiunque produca¹³¹.

➤ **Settore risorse umane**

Alla domanda “Quali sono gli investimenti che il tuo Istituto ha già sostenuto nel settore delle Risorse Umane?”, quasi la metà dei rispondenti non risponde o afferma di non conoscere gli investimenti fatti dal proprio istituto nel settore indagato. Le rimanenti risposte fornite, confermano quanto già emerso sia nella prima domanda relativa alle funzioni svolte – che per la metà dei casi riguardava appunto funzioni legate alla *data science* - sia nella trattazione teorica del presente lavoro: gli Istituti più sensibili agli investimenti su nuove competenze, assumono e/o formano profili da *data scientist*. L’assunzione di *data scientist*, di analisti di dati favorisce e accelera spesso il rilascio di *output* statistici sperimentali. Per alcuni Istituti la partecipazione ai progetti europei, così come le prime statistiche sperimentali prodotte, hanno rappresentato una sorta di formazione sul campo per i ricercatori coinvolti.

Questo risultato conferma quanto già anticipato nel capitolo 1 e che qui brevemente riportiamo, relativamente al ruolo della *data science* come nuova disciplina in grado di aiutare e supportare l’estrazione di informazioni e conoscenza dai dati. L’appartenenza dei rispondenti al mondo della *data science* indica che l’approccio alla scienza si sta spostando dall’essere guidato da un paradigma *model-driven* (dove la modellazione e le astrazioni sono fondamentali e i dati sono solo di supporto a date ipotesi iniziali) a un paradigma *data-driven* (dove i modelli possono essere estratti direttamente dai dati). È sempre più richiesta la capacità di trarre informazioni da grandi insiemi di dati, utilizzando un insieme più ampio di competenze. I modelli di conoscenza stanno evolvendo da un modello a forma di T a modelli di conoscenza a forma di Pi (cfr figura sotto). Quando la capacità del ricercatore

¹³¹ Von Oppeln-Bronikowski S, Hagenkort-Rieger S., João Santos M., *New trends in communication: Branding and content marketing*, in *Power from Statistics: data, information and knowledge*, Eurostat 2018 pag. 117

supporta un approccio a forma di T, il ricercatore padroneggia sia la specializzazione del dominio (sull'asse verticale) che le competenze interdisciplinari (sull'asse orizzontale).

In nuovo modello emergente *Pi-shaped*, aggiunge un'altra competenza verticale, relativa alle capacità statistiche e computazionali necessarie per affrontare l'analisi dei (big) data.

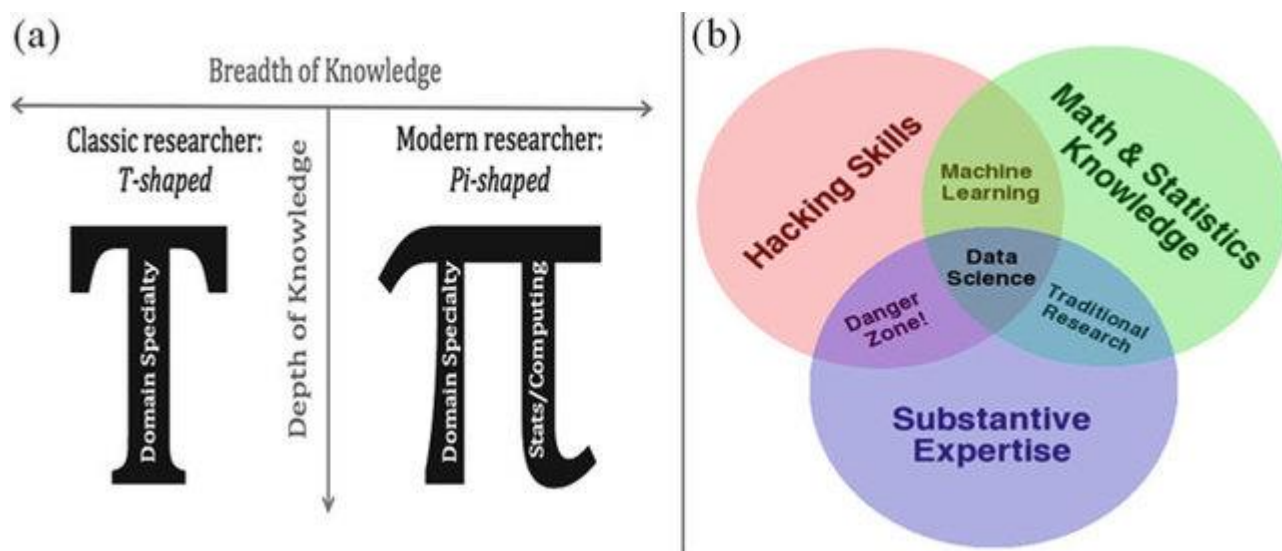


Fig. Rappresentazione delle competenze del data scientist: passaggio da T model a Pi model (from: DrewConway.com, retrieved 25/6/ 2015)

➤ Settore legale

Alla domanda “Quali sono gli investimenti che il tuo Istituto ha sostenuto nel settore legale?”, 3 degli istituti rispondenti affermano di non aver avviato alcuna attività nel settore, tutti gli altri mettono in evidenza la necessità di dotarsi un *framework* legale, di specifiche *policy* che consentano il pieno uso dei big data. La metà degli istituti cita linee guida europee sul *web scraping*¹³² come unica cornice di riferimento. Il *web scraping* (dall'inglese *to scrape*, grattare/raschiare), è una tecnica che permette di estrarre e memorizzare dati dalle pagine web, ad esempio, per scopi di analisi o per altri tipi di valutazione. Durante lo *scraping* vengono raccolti molti tipi diversi di informazioni. Può trattarsi, ad esempio, di informazioni di contatto, come indirizzi di posta elettronica o numeri di telefono, ma anche di singoli termini di ricerca o URL. Questi vengono quindi raccolti in database o tabelle locali. Tutti i paesi membri del Sistema statistico europeo, usano nuove fonti di dati per produrre

¹³² https://ec.europa.eu/eurostat/cros/content/item-04-web-scraping-policy_en.

statistiche e analisi in tutta Europa e il *web scraping* consente una raccolta più tempestiva di nuove forme di dati per produrre informazioni statistiche, per questo si è reso necessario produrre linee guida sulla politica di *web scraping* in grado di guidare il SSE e i suoi paesi membri nelle loro attività di estrazione, raccolta, elaborazione e uso dei dati.

L'adozione di queste linee guida da parte di un gran numero di INS rende evidente la necessità di regolamentazione: non appena a livello centralizzato vengono definite specifiche *policy*, queste diventano patrimonio comune dell'intera comunità statistica di riferimento.

La mancanza di investimenti nel settore legale per alcuni dei rispondenti è dettata da fattori ambientali esterni agli istituti e riconducibili a un contesto normativo nazionale poco favorevole all'uso delle nuove fonti. In altri Istituti sono invece in atto interlocuzioni interne e/o esterne al proprio istituto finalizzate a trovare soluzioni in grado di superare problemi di accesso ai dati e di tutela della *privacy* in generale. È da segnalare che è il processo di revisione in corso della legge statistica europea Regolamento CE N. 223/2009¹³³ del Parlamento Europeo e del Consiglio è attualmente arrivata allo stadio della consultazione pubblica di tutti gli interessati (produttori di dati pubblici e privati, utilizzatori istituzionali cittadini, ecc.), la è cominciata il 19 luglio 2022 e scade il 25 ottobre 2022.

Sebbene parzialmente riviste nel 2015 con il Regolamento UE n. 759¹³⁴, il Regolamento del 2009 riflette il modo in cui le statistiche ufficiali venivano prodotte all'inizio degli anni 2000. In questo lavoro abbiamo visto quanto e come si sia trasformato il mondo dei dati e l'ambiente questi dati si producono e i sistemi di produzione statistica in cui vengono sviluppate, prodotte e diffuse le statistiche europee. Le regole che disciplinano le statistiche ufficiali devono essere aggiornate per riflettere questi cambiamenti.

¹³³ <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:02009R0223-20150608>

¹³⁴ <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=celex:32015R0759>

➤ **Settore organizzativo**

La metà dei rispondenti, così come accaduto per il settore relativo a comunicazione e diffusione, alla domanda *“Quali sono gli investimenti che il tuo Istituto ha fatto nel settore organizzativo?”*, dichiara che l’Istituto di appartenenza non ha realizzato nessuna iniziativa. Le altre risposte offrono risultati interessanti perché i diversi rispondenti danno un’interpretazione diversa all’espressione *“settore organizzativo”*. Di seguito si riportano le differenti risposte ricevute perché offrono interessanti spunti di riflessione:

- *“all’interno del mio Istituto è stata creata una struttura organizzativa dedicata esclusivamente all’analisi dati”*,
- *“sono stati acquisiti server dedicati al web scraping”*.

In questi casi specifici è evidente che anche se si parla di settore organizzativo la visione predominante è ancora legata ad una vista prettamente tecnologica. Di contro un altro rispondente fa riferimento ad una struttura dedicata alla Ricerca *in new digital data* finalizzata al miglioramento della qualità dell’informazione statistica con specifico riferimento ad accuratezza, tempestività e rilevanza. I dati a cui si fa riferimento sono dati di telefonia mobile, dati da sensori e dati provenienti dall’IoT (Internet of Things). Questa risposta invece segnala chiaramente maggiore consapevolezza sul percorso da seguire derivante da esperienze e risultati già consolidati in questo ambito.

La mancanza di investimenti nel settore organizzativo e il fatto che un terzo degli istituti non risponda alla domanda, segnala che sebbene l’opportunità offerta dalle nuove fonti sia stata colta dalla totalità degli Istituti – seppure con livelli di approfondimento diversi -, questa non sempre è stata accompagnata da riorganizzazione dei processi e dalla implementazione di ambienti in grado di facilitare i nuovi sistemi produttivi.

4.3.4 Terza sezione – Investimenti programmati

La sezione precedente ci ha permesso una prima analisi sull'*as-is*, cioè sul posizionamento attuale dei INS nella costruzione di un sistema idoneo a favorire e valorizzare le TSS nei vari ambiti presi in riferimento, la terza sezione, che analizzeremo nel presente paragrafo, si focalizza sugli investimenti pianificati e ancora da realizzare nei medesimi settori e quindi tende a restituire uno scenario del *to-be*.

Anche in questa sezione la domanda “quali sono gli investimenti che il tuo Istituto ha programmato di sostenere per potenziare la produzione di TSS nel settore...?” viene ripetuta per ciascuno dei settori analizzati, così come nella sezione precedente. Quasi tutte le risposte fornite, come vedremo nelle singole sezioni, sono relative ad azioni di consolidamento di quanto già fatto. Di seguito si riportano i risultati relativi ai singoli settori:

➤ **Settore metodologico**

Alla domanda su quali siano gli investimenti programmati nel settore metodologico, tre rispondenti affermano di non aver pianificato nessun altro investimento, per un paio degli intervistati gli investimenti metodologici programmati coincidono con la prosecuzione degli investimenti e dei progetti metodologici già in essere e che riguardano, come già emerso, la partecipazione a progetti europei. Nelle risposte si fa riferimento ad alcune *best practice* fra cui emerge il progetto WIN *Web Intelligence Network* nato in ambito europeo per usare i servizi *Web Intelligence Hub* (WIH) e integrarli nei sistemi di produzione statistica nazionali. Quasi la metà dei partecipanti all'indagine afferma che i prossimi investimenti riguarderanno infrastrutture e nuovi paradigmi metodologici in grado di favorire l'integrazione di nuovi dati provenienti da diverse fonti con particolare riferimento all'acquisizione di dati di telefonia mobile da *providers* privati.

L'ONS per rispondere alla domanda segnala due progetti specifici:

- *Stima dell'attività veicolare e pedonale da telecamere del traffico cittadino e cittadino: realizzato attraverso l'estrazione di informazioni dalle telecamere del traffico con*

tecniche di riconoscimento automatizzato delle immagini e *deep learning*. I dati delle telecamere del traffico possono offrire un enorme valore alle autorità pubbliche fornendo statistiche in tempo reale per monitorare l'attività delle popolazioni locali e supportare le decisioni così come accaduto durante la pandemia. Le telecamere sul traffico forniscono dati tempestivi e ad alta frequenza che possono aiutare a rilevare specifici trends nel comportamento sociale.

- *Faster Indicators of UK Economic Activity* per produrre informazioni economiche più rapide e per creare, attraverso i big data, una serie di indicatori che consentano l'identificazione precoce di grandi cambiamenti economici e fornire informazioni sull'attività economica, a un livello di tempestività e granularità attualmente non possibili con le statistiche economiche ufficiali.

➤ **Settore informatico**

Tutte le risposte relative al settore informatico si riferiscono ad investimenti finalizzati al rafforzamento dell'infrastruttura e all'acquisizione di piattaforme specifiche per l'analisi di grandi moli di dati (*Kubernetes, Spark Cluster, Cloudera Data platform*) e il rafforzamento delle azioni già intraprese in questa direzione. Nelle risposte si fa riferimento all'implementazione di *cloud* e piattaforme per analisi di grandi dati, all'implementazione di una *sandbox*, all'acquisizione di *software* per il *web scraping*, alla migrazione di alcuni dati nella piattaforma Cloudera per analisi avanzate, all'acquisizione di strumenti per processare *scanner data*.

➤ **Settore comunicazione e diffusione**

Il settore comunicazione è il settore che meno di tutti ha pianificato investimenti di medio lungo termine. Le risposte in questo caso sono sovrapponibili a quelle registrate nella sezione relativa agli investimenti realizzati. Tra i rispondenti solo due Istituti dichiarano di aver pianificato investimenti in questo settore, fondamentalmente quasi tutti i rispondenti, gli investimenti nel settore comunicazione e diffusione consistono nel continuare a pubblicare statistiche sperimentali popolando e aggiornando lo spazio a loro dedicato sui vari siti

istituzionali. Uno dei rispondenti cita i seminari come possibile ulteriore strumento di comunicazione e diffusione, un altro le ulteriori collaborazioni con *providers* esterni per incrementare la pubblicazione di *small data* con granularità territoriale più fine. Quello che emerge nuovamente è il ruolo delle statistiche sperimentali che, oltre a rappresentare un *output* statistico, sono uno strumento su cui gli istituti hanno investito e vogliono continuare ad investire per valorizzare e comunicare all'esterno l'uso delle nuove fonti dati nella produzione statistica ufficiale.

➤ **Settore risorse umane**

Anche il settore risorse umane risente di una scarsa pianificazione, sono solo tre gli istituti che prevedono l'assunzione di nuovo personale proveniente dal mondo accademico a cui affidare specifiche analisi di fattibilità, in un altro caso l'istituto cerca un unico esperto di *web scraping* e uno statistico per la gestione di un *datawarehouse*. Il carattere marcatamente tecnico dei profili ricercati, fa emergere una programmazione di breve periodo finalizzata a risolvere problemi specifici. Sembra mancare una visione di lungo periodo che preveda formazione del personale interno e o l'assunzione di nuovo personale con competenze idonee a sostenere nuovi processi di produzione statistica. In altre sezioni del questionario, oltre che nella trattazione teorica sul tema dei big data, è già emersa la necessità di specifiche competenze per riuscire ad estrarre valore informativo da questi nuovi dati, eppure, a fronte di questa riconosciuta esigenza, si manifesta una scarsa capacità di programmare misure utili a colmare i gap di competenza emersi. Dall'analisi delle risposte risulta che gli istituti cercano soluzioni tampone per risolvere problemi di breve periodo, è emblematica la risposta relativa alla ricerca di un unico esperto di *web scraping* che segnala da un parte sicuramente l'urgenza di trovare una soluzione ad un problema imminente e dall'altra una incompleta visione sulla necessità di dotarsi in maniera strutturata di un sistema di competenze in grado di sostenere il passaggio ad un nuovo modello produttivo.

➤ Settore legale

Alla domanda “*quali investimenti il tuo istituto ha programmato nel settore legale*”, 5 rispondenti dichiarano che il proprio istituto non ha programmato nessuna iniziativa per adeguare il settore legale alle nuove esigenze produttive. Gli altri Istituti che invece sono più attivi su questo fronte, hanno avviato attività di approfondimento della normativa legata alla *privacy* a livello sia nazionale che europeo per poi adeguare la propria normativa interna al contesto normativo esterno. Gli Istituti che hanno partecipato alla stesura delle linee guida sull’uso del *web scraping* (a cui abbiamo fatto riferimento anche nella sezione precedente) citano queste linee guida e i loro possibili aggiornamenti come investimenti nel settore legale. Quattro istituti fanno invece riferimento alla creazione di *framework* normativi, attraverso anche l’interlocuzione con *stakeholders* esterni, che consentano o facilitino l’acquisizione di dati da providers esterni (dati di telefonia, sul traffico, sul consumo di elettricità ecc...).

Dall’analisi delle risposte emerge sicuramente che l’obiettivo di qualsiasi iniziativa avviata dai singoli Istituti nel settore legale è indissolubilmente legata alla necessità di garantire o comunque facilitare l’accesso ai dati detenuti da *providers* esterni. L’obiettivo è quello di creare una cornice normativa idonea a favorire l’accesso ai dati, a tutelare gli utenti proprietari dei dati e a rafforzare le collaborazioni con enti privati e pubblici in tema di scambio dei dati.

L’ONS ha programmato attività finalizzate a inserire nella cornice legale di Istituto, norme per garantire forme più avanzate di tutela della *privacy* tra cui emerge, per il suo carattere innovativo, la crittografia omomorfa, una tipologia di crittografia a chiave pubblica, caratterizzata dalla possibilità di intervenire sul contenuto crittografato senza un processo di decodifica prima di elaborare i dati. Il vantaggio è proprio sul versante della sicurezza, perché il contenuto codificato non apre mai il suo guscio crittografico e consente di eseguire i calcoli direttamente sui dati cifrati senza la possibilità di accedere ai dati in chiaro.

4.3.5 Quarta sezione – Investimenti auspicati

L'ultima sezione dedicata agli investimenti, indaga le aspettative degli intervistati. Come già detto, i ricercatori coinvolti nella *survey* hanno lunga esperienza di lavoro nei settori legati all'analisi e alla produzione di *output* statistici con l'uso delle nuove fonti sia a livello nazionale sia a livello internazionale. I loro suggerimenti su auspicabili investimenti da parte degli istituti di appartenenza offrono spunti di riflessione e possibili indicazioni strategiche per indirizzare le scelte future. Le sezioni precedenti hanno indagato una dimensione oggettiva (investimenti fatti e programmati), questa indaga una dimensione meramente soggettiva.

Anche in questa sezione, così come in quelle precedenti, la domanda *“in base alla tua esperienza, quali sono gli investimenti che secondo te il tuo Istituto dovrebbe fare per potenziare la produzione di TSS nel settore...?”* viene ripetuta per ciascuno dei settori analizzati. Dalle risposte ricaviamo un'analisi delle percezioni, viene valorizzato il ruolo del rispondente a cui viene infatti chiesto di fornire un'opinione personale che si basa sull'esperienza professionale maturata sui temi indagati. Ai ricercatori che hanno maturato esperienze specifiche sul tema viene quindi chiesto di valutare esigenze da colmare nei vari settori degli Istituti per andare verso una produzione strutturata di informazione statistica con l'uso delle nuove fonti.

In linea generale questa è la sezione in cui i ricercatori hanno dato maggiori indicazioni su possibili orientamenti, è la parte del questionario in cui i rispondenti hanno potuto indicare i loro suggerimenti personali e non solo fornire informazioni in loro possesso. Nel complesso le indicazioni date restituiscono un quadro completo su interventi ritenuti necessari per valorizzare quanto fatto e dare maggiore impulso al processo di messa in produzione delle TSS. In questa sezione sono meno frequenti le mancate risposte, questo segnala la predisposizione dei ricercatori a dare un loro contributo e il loro desiderio di sentirsi parte di un processo al quale hanno collaborato con il loro lavoro e le loro competenze. Le sezioni in cui emerge un numero più alto di mancate risposte, così come pure nelle sezioni

precedenti, sono quelle relative ai settori legale, organizzativo e di comunicazione e diffusione perché sono settori tradizionalmente più “lontani” dalle competenze dei ricercatori coinvolti e ritenuti, purtroppo, settori meno impattanti sull’obiettivo finale. Il contesto dei singoli istituti forse ancora non è maturo per far risaltare la rilevanza di questi settori nella costruzione complessiva del sistema di produzione TSS. Non è chiara evidentemente la necessità di un approccio integrato tra le diverse dimensioni né la complessità del sistema di produzione TSS, ovvero il concetto relativo alla sostenibilità organizzativa, illustrato nel capitolo precedente.

➤ **Settore metodologico**

Le indicazioni ricevute sugli investimenti da fare per migliorare il settore metodologico sono di grande interesse, abbracciano varie fasi del processo produttivo ed emergono per il loro specifico carattere trasversale. I suggerimenti spaziano dalla richiesta di un numero maggiore di risorse umane con competenze metodologiche, fino alla capacità di “*thinking out-of-the box*”, passando dall’esigenza di avere un maggiore supporto dal top management e dall’opportunità di sviluppare una strategia e una *roadmap* per le TSS a livello nazionale. Si chiede quindi una *governance* centralizzata in grado di assumersi l’onere di guidare i processi verso la costruzione del nuovo sistema di produzione attraverso, per esempio, l’adozione di linee guida e l’implementazione di iniziative comuni in tutti i paesi del SSE.

Emerge la necessità di rafforzare la ricerca, tramite il superamento del problema più volte emerso della carenza di risorse, si suggerisce di assumere nuovi *data scientist*, di formare un gruppo dedicato all’analisi e alla produzione di TSS, di acquisire dati da *smart systems*, di rafforzare i settori dedicati alla ricerca e sviluppare di nuovi paradigmi metodologici in grado di rafforzare l’integrazione di nuove fonti nella produzione di statistiche ufficiali.

In alcuni casi sono indicati specifici suggerimenti, probabilmente derivati da particolari problemi affrontati dai singoli istituti, è questo il caso dei ricercatori che suggeriscono di creare linee guida per indicatori di qualità da usare per studi di fattibilità sull’uso di nuove

fonti dati per la produzione statistica corrente, o una *dashboard* di tutti i metodi e modelli di produzione e dei relativi aggiornamenti.

Di grande interesse è la risposta *thinking of the box*, che si ripete anche nelle domande relative agli altri settori e che mette in chiara evidenza la necessità di assumere un nuovo approccio e di creare nuovi paradigmi per creare il sistema delle TSS. Le TSS rappresentano, come già specificato, un cambiamento culturale, un'opportunità, per coglierla occorre però usare nuovi approcci, nuovi strumenti, nuovi metodi che siano meglio adattabili ai nuovi dati senza derogare ai principi propri della statistica ufficiale. La statistica, gli Istituti di statistica devono evolvere insieme ai cambiamenti del mondo esterno se vogliono rispondere alla sfida della trasformazione digitale e per farlo devono pensare fuori dagli schemi della statistica "classica".

➤ **Settore informatico**

Anche nella parte relativa agli investimenti auspicati nel settore informatico, emerge la necessità di rafforzare le infrastrutture di cui sono già dotati gli istituti. La maggior parte dei ricercatori fa riferimento al *cloud computing*, a piattaforme tipo Sandbox che consentano la condivisione dei dati sia all'interno del sistema statistico sia con imprese private.

Seppure la domanda si riferisca ad un settore prettamente tecnologico, il cinque rispondenti richiedono investimenti su formazione del personale e la costituzione di strumenti di *governance* in grado di orientare il processo e, nella migliore delle ipotesi, di un gruppo dedicato alle TSS. Si richiedono investimenti finalizzati ad un uso più efficace ed efficiente delle infrastrutture tecnologiche di cui gli INS sono dotati, quindi maggiore formazione e nuove risorse.

Il tema delle competenze e il tema di una *governance* dedicata sono elementi che vengono spesso riportati nelle diverse sezioni, questo testimonia la rilevanza strategica che viene loro attribuita dai rispondenti e la necessità però di seguire un percorso strutturato e soprattutto comune a tutti gli INS del Sistema Statistico Europeo.

➤ **Settore legale**

Anche in questo caso, il numero di mancate risposte è elevato, emerge però tra i rispondenti la necessità di un allineamento innanzitutto con la normativa europea ovvero con il Regolamento Ue 2016/679 che dal 25 maggio 2018 è divenuto pienamente applicabile in tutti gli Stati membri. Il Regolamento, noto come GDPR (*General Data Protection Regulation*), norma la protezione delle persone fisiche con riguardo al trattamento e alla libera circolazione dei dati personali.

Emerge ancora una volta con forza la necessità di accordi con i fornitori di dati, e di un quadro normativo maggiormente flessibile che consenta in maniera più snella, l'uso delle nuove fonti dati. In mancanza di questo, viene segnalata la difficoltà di rendere continuativo e strutturale l'uso delle nuove fonti di dati nei processi di produzione statistica. A fronte della maggior parte di risposte che richiede soluzioni più ampie e di lungo respiro, uno dei rispondenti suggerisce investimenti in campo legale solo quando si presentano i problemi specifici.

➤ **Settore comunicazione e diffusione**

La domanda *“in base alla tua esperienza, quali sono gli investimenti che il tuo Istituto dovrebbe fare nel settore comunicazione e diffusione per rafforzare la produzione TSS?”* registra il minor numero di risposte. Pochissimi rispondenti forniscono indicazioni per proporre un nuovo paradigma di comunicazione e segnalano la necessità di avere una comunicazione chiara e accessibile, migliori strumenti di visualizzazione e una pubblicazione automatizzata di dati in tempo reale con una supervisione più leggera. La mancanza di suggerimenti più strutturati è dovuta al fatto che nessuno dei rispondenti ha dichiarato di possedere competenze e/o esperienze professionali in questo ambito disciplinare e questo comporta chiaramente una minore capacità di dare suggerimenti mirati. Il riferimento ad una migliore visualizzazione,

attraverso l'uso di *dashboard*, è la risposta comunque più ricorrente, gli strumenti di *data visualization* risultano quelli maggiormente capaci di garantire una comunicazione più efficace e maggiormente adattabile alla tipologia di *output* prodotti.

➤ **Settore risorse umane**

I suggerimenti che vengono dati per migliorare il settore delle risorse umane sono tutti focalizzati su due aspetti:

- incremento del numero di risorse e quindi nuove assunzioni con specifiche competenze legate alla data science,
- formazione continua del personale organizzata e gestita in maniera strutturata nel lungo periodo, particolarmente focalizzata sul rafforzamento delle competenze metodologiche e informatiche.

In alcuni casi viene suggerito di creare gruppi dedicati alle nuove fonti all'interno dei singoli Istituti in modo da poter seguire e supportare il processo di costruzione del sistema TSS.

Le indicazioni fornite sono perfettamente sovrapponibili alle esigenze che i ricercatori hanno messo in luce nelle varie sezioni della *survey*, i loro suggerimenti sono orientati a colmare i *gap* emersi. Il tema della carenza delle risorse viene spesso citato, il personale è impegnato a garantire la produzione corrente, legata al rispetto di Regolamenti comunitari, è difficile quindi liberare risorse da impegnare esclusivamente sul fronte ricerca e innovazione. Pertanto si fa pressante il richiamo all'incremento del numero di risorse e al miglioramento delle loro competenze.

4.3.6 Sezione dati e strumenti usati

La sezione del questionario "Dati e strumenti usati" raccoglie informazioni sulla tipologia dei dati usati e sugli strumenti e sulle necessarie forme di collaborazione utili a produrre i primi *output* innovativi. Fondamentalmente viene richiesto ai rispondenti di esplicitare con quali mezzi hanno potuto garantire i primi rilasci statistici con l'uso delle nuove fonti. Uno

degli aspetti che si è voluto approfondire, riguarda nello specifico le collaborazioni con enti pubblici (Enti di ricerca, università...) e privati (fornitori di dati) perché le collaborazioni con il mondo esterno sono imprescindibili per la ricerca e l'innovazione, e per creare e alimentare un vero e proprio ecosistema della ricerca. Una delle sfide principali per gli istituti produttori di statistiche ufficiali è quella di posizionare questi prodotti nel mercato dell'informazione e farli diventare parte dell'ecosistema informativo. Per fare questo gli Istituti di statistica sono necessariamente chiamati ad instaurare una proficua collaborazione con nuovi *partner*. La disponibilità di nuovi dati rende necessaria una ridefinizione dei modelli di produzione che prevede l'attivazione e il rafforzamento delle collaborazioni con altri soggetti pubblici e privati. Come si evince dal Grafico n. 3 la maggior parte dei dati usati sono *Web Data*, cioè dati provenienti da Internet (IaD *Internet as Data Source*) seguiti da *scanner data* e da immagini satellitari. I dati di telefonia mobile, sebbene rappresentino una fonte in grado di restituire informazioni importanti su settori quali turismo e mobilità, sono solo al quarto posto per la difficoltà di accesso a questi dati.

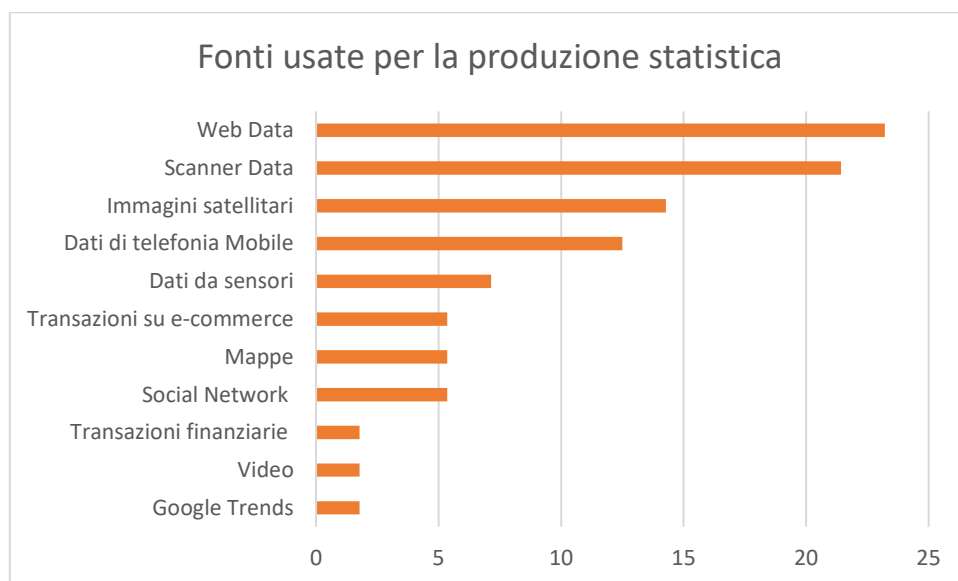


Grafico 3 - Fonti usate per la produzione statistica

Tredici Istituti hanno avuto accesso diretto ai dati, attraverso tecniche di *web scraping* nella maggioranza dei casi o di tecniche miste a seconda del tipo di fonte trattata (Grafico 4). Tra

le tecniche miste maggiormente citate rientra l'accesso tramite API (*application programming interface*).

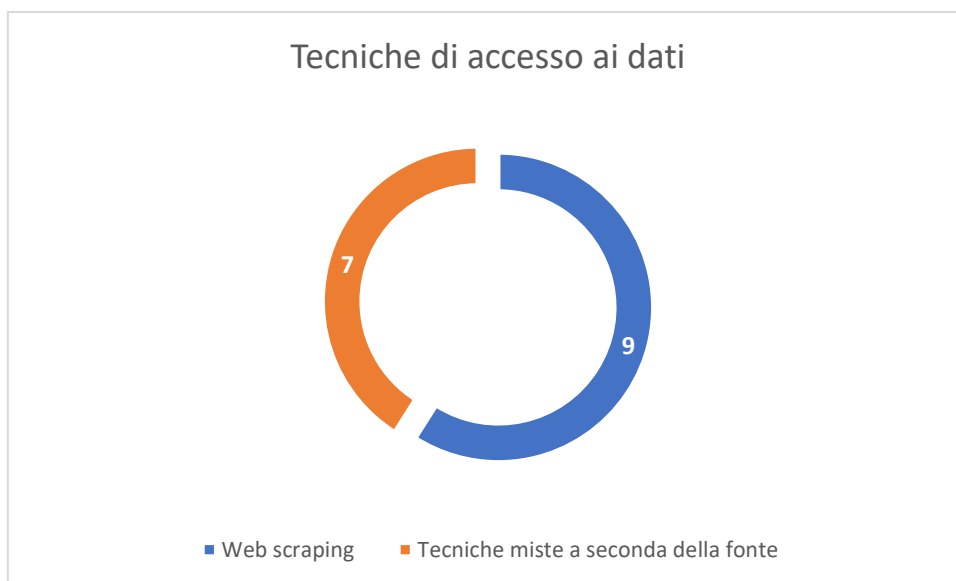


Grafico 4 - Tecniche di accesso ai dati

La domanda n. 12 chiede informazioni sulla natura delle collaborazioni esterne e in particolare sulla tipologia di ente maggiormente coinvolto nella fornitura di dati. Gli operatori telefonici sono gli enti con cui si stringono maggiori forme di collaborazione, senza delle quali l'accesso ai dati MNO non sarebbe chiaramente possibile. Seguono le Istituzioni governative e le imprese private.

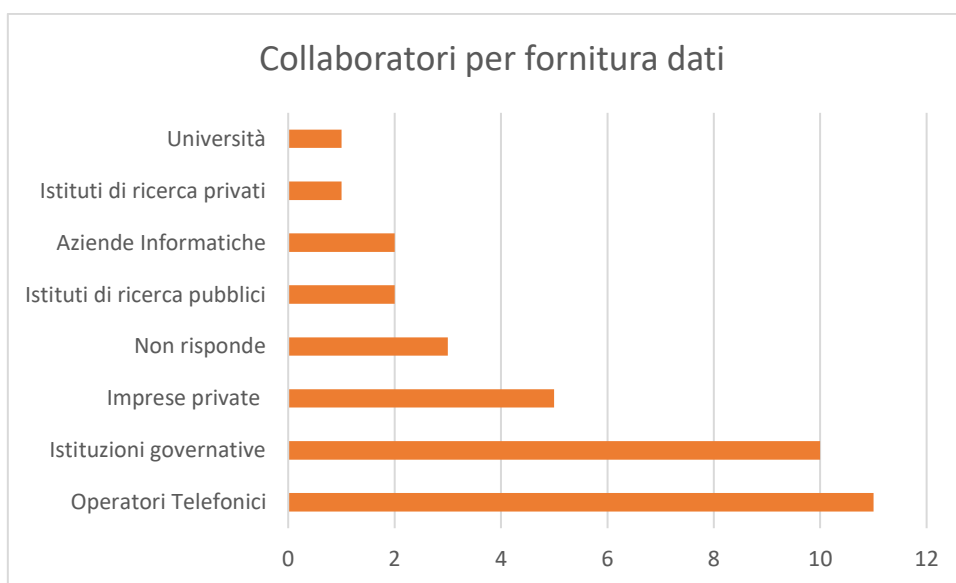


Grafico 5 - Tipologia di soggetti con cui gli stipulano accordi di collaborazione

La domanda successiva chiede “*Il partner esterno che ha fornito dati, ha partecipato ad altre fasi del processo produttivo?*”. In tre casi i *providers* hanno collaborato anche ad altre fasi del processo statistico e nello specifico in un caso il partner esterno ha partecipato a tutte le fasi del processo produttivo, negli altri casi segnalati la partecipazione ha riguardato la fase di validazione dei risultati.

Alla domanda relativa alle tecniche di elaborazione usate per la fase di inferenza dei big data, 10 intervistati fanno riferimento a tecniche di *Machine Learning* e gli altri 6 dichiarano di usare tecniche miste a seconda della tipologia dei dati e dei singoli progetti. Altre tecniche citate sono il *Deep Learning* e altre procedure statistiche di tipo più tradizionale (analisi di regressione, analisi di correlazione es.).

Estremamente interessanti sono le risposte ricevute alla domanda relativa ai problemi riscontrati nella fase di elaborazione dati. Le varie segnalazioni contengono una serie di indicazioni per gli istituti di statistica che potrebbero orientare investimenti futuri verso la risoluzione dei problemi segnalati.

Nella gran parte dei casi i ricercatori mettono in luce difficoltà in fase di accesso ai dati (i dati spesso sono archiviati e accessibili da file CSV), segnalano *software* con scarsa potenza di calcolo, spazi di archiviazione ridotti, problemi di *privacy*, mancate competenze specifiche, di riservatezza e di sicurezza soprattutto per i dati di telefonia mobile. Si registra anche una mancanza di competenze tecniche e difficoltà amministrative nei settori preposti ad avviare processi di acquisizione di piattaforme *cloud* presso fornitori esterni. In generale quello che emerge è una difficoltà di carattere organizzativo, che paradossalmente non emerge nella sezione dedicata al settore organizzativo. La produzione di queste statistiche appare non inserita appieno nella macchina organizzativa quindi i settori impegnati nella produzione statistica con le nuove fonti dati trovano spesso difficoltà nei settori a supporto (legale, acquisizione di beni e servizi...).

A fronte dei problemi segnalati, a ciascuno dei rispondenti viene anche chiesto di specificare le soluzioni messe in campo per superare le difficoltà segnalate. Spesso la soluzione trovata è individuale e non rispecchia una politica generale di Istituto. Vengono segnalate collaborazioni con partner esterni, aggiornamenti di *software* o accesso ai dati a pagamento, altri stanno lavorando a possibili soluzioni senza specificare quali, altri hanno fatto ricorso a formazione interna. L'ONS segnala che la soluzione consiste in gruppi di lavoro orizzontali tra diverse realtà di ricerca e istituzionali quindi nella cooperazione tra i soggetti coinvolti.

Tra le risposte ne segnaliamo due in particolare che descrivono due approcci agli antipodi: uno dimostra una visione strategica e la esprime ancora una volta segnalando la necessità di *thinking out of the box* e l'altro suggerisce soluzioni *ad hoc* ai problemi che di volta in volta si presentano, escludendo in questo modo la necessità di una riorganizzazione complessiva del sistema di funzionamento degli Istituti e proponendo quindi un modello di lavoro *stovepipe*.

4.3.7 Sezione risultati raggiunti

La domanda n. 17 raccoglie informazioni sugli *output* statistici già prodotti e rilasciati con le nuove fonti dati, e quindi fondamentalmente sulle statistiche sperimentali. In particolare chiede di indicare in quali ambiti tematici ricadono i primi risultati prodotti. Come si evince dal Grafico 6, le aree tematiche di riferimento riguardano maggiormente l'economia, i trasporti e a seguire turismo popolazione e prezzi. I settori più coinvolti sono quelli su cui gli INS hanno investito di più già in fase sperimentale durante gli EssNet: scanner data, immagini satellitari, AIS...

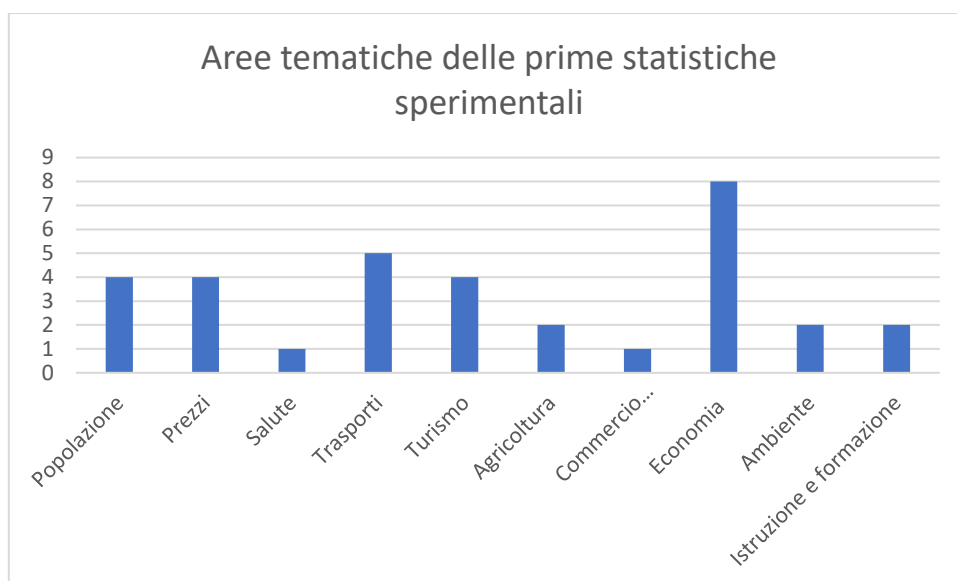


Grafico 6- Ambiti tematici delle statistiche sperimentali pubblicate fino al settembre 2020

Come già indicato nelle pagine precedenti, tutti gli Istituti che hanno prodotto statistiche sperimentali le hanno pubblicate sui siti istituzionali in un'area specificatamente dedicata, contrassegnata molto spesso da un logo dedicato che la distingue dagli spazi contenenti prodotti di statistica ufficiale. Ciascuna statistica sperimentale ha, a corredo dei risultati, una nota metodologica esplicitiva dei processi che hanno prodotto l'*output* pubblicato. In ogni area di ciascun istituto, vengono raccolti anche i *feedback* degli utenti per capire come vengono recepiti questi nuovi *output*, in cosa sono migliorabili e quali sono gli usi che gli utenti ne fanno, in sostanza il *feedback* diventa uno strumento di coinvolgimento degli utenti e anche un criterio su cui basare l'eventuale passaggio in produzione delle statistiche sperimentali stesse.

La domanda n. 18 chiede appunto agli intervistati se hanno avuto modo di consultare i *feedback* degli utenti, ma solo due dei ricercatori coinvolti rispondono affermativamente e segnalano in un caso che gli utenti dichiarano di voler accedere direttamente ai dati grezzi e quindi non e nell'altro caso segnalano che i *feedback* sempre positivi perché provengono quasi esclusivamente dal mondo accademico.

Alla domanda n. 20 “secondo te qual è la piattaforma più funzionale alla diffusione delle TSS” le risposte sono pressoché uniformi e danno i risultati riportati nel Grafico 7. Viene richiesto al rispondente di attribuire a ciascuna delle piattaforme indicare un punteggio da 1 a 4 relativo a diversi gradi di funzionalità in una scala in cui 1 equivale a per nulla funzionale e 4 a molto funzionale.

In your opinion, what is the most functional platform for the dissemination of Trusted Smart Statistics?

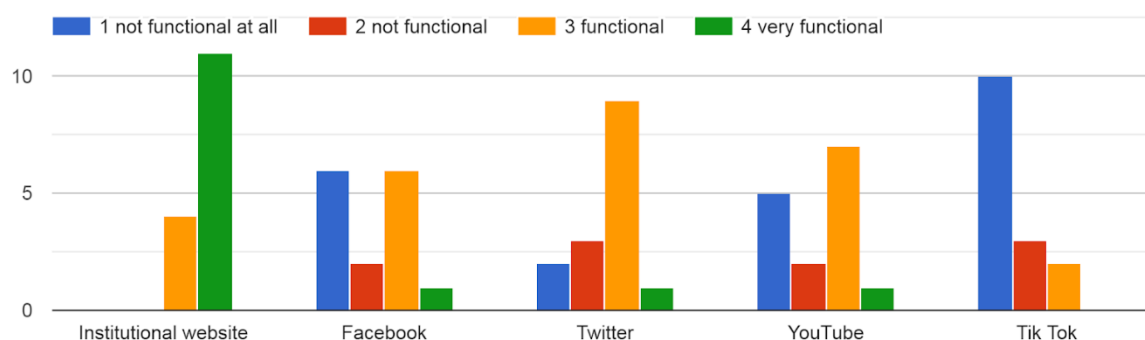


Grafico 7 - Piattaforme più funzionali alla diffusione di TSS

Tutti i ricercatori coinvolti sono concordi nel sostenere che la piattaforma più idonea a diffondere statistiche con i big data sia il sito istituzionale in quanto è l’interfaccia dei singoli istituti con l’utenza, rappresenta lo strumento in grado di suscitare maggiore fiducia. Al contrario *TikTok* è la piattaforma percepita come meno adatta alla diffusione di questi prodotti. *Twitter* è invece ritenuto, immediatamente dopo il sito istituzionale, lo strumento più funzionale per diffondere i dati e statistiche, seguito da *YouTube* e *Facebook*. *Twitter* è anche segnalato come lo strumento per entrare in contatto con la comunità scientifica.

Le risposte date mettono da più parte in luce la necessità di segnalare con attenzione agli utenti che le statistiche sperimentali e/o quelle prodotte con l’uso delle nuove fonti non sono statistiche ufficiali, pertanto occorre estrema cura nell’esplicitare l’iter di produzione di

questi prodotti in modo che non sia intaccata la fiducia che gli utenti hanno nei confronti degli istituti di statistica che, in ciascuno dei paesi coinvolti, rappresentano il maggiore produttore di statistica ufficiale.

Emerge qui, più che altrove, l'importanza non solo di metter in atto tutti i metodi e gli strumenti più idonei a garantire gli stessi standard di qualità, rilevanza ed affidabilità tipici delle statistiche tradizionali al fine di non mettere in pericolo la *trust*, ma emerge soprattutto la imprescindibile necessità di comunicare all'esterno l'esistenza e l'adozione di queste garanzie. Gli istituti di statistica da sempre in grado assicurare soluzioni a difesa della sicurezza dei dati tradizionali, ora devono dapprima attrezzarsi per assicurare le stesse garanzie anche per i prodotti rilasciati con dati digitali e poi comunicare a tutti gli *stakeholders* l'esistenza e l'adozione di queste garanzie a tutela degli *stakeholders* stessi.

Attiva anche la presenza degli Istituti su Instagram. Allo stato attuale emerge un carattere *crossmediale* dei messaggi diffusi, lo stesso messaggio, ovvero lo stesso prodotto statistico, viene diffuso su più piattaforme. L'evoluzione di questa comunicazione potrebbe assumere un carattere *transmediale* offrendo agli utenti esperienze di narrazione diversa a seconda della piattaforma usata e veicolando nuove e distinte informazioni. Il carattere transmediale aiuterebbe alla comprensione dei dati perché spingerebbe l'utente a ricostruire il significato complessivo dei prodotti statistici diffusi. Sicuramente si può affermare che la presenza degli istituti sulle varie piattaforme comunque garantisce sia una diffusione più capillare dell'informazione statistica sia un dialogo costante con l'utenza e rafforza il ruolo degli enti come attori importanti nella diffusione dell'informazione statistica e nella comunicazione scientifica in generale.

L'ultima parte del questionario ritorna sul fondamentale tema delle risorse umane. Ai rispondenti viene chiesto di indicare quale competenza risulta più importante per la produzione di statistiche con l'uso dei big data, e il grado di importanza che ciascuno di essi attribuisce alle singole famiglie di competenze (in una scala che permette di scegliere un'opzione fra 4 livelli di utilità: per nulla utili, poco utili, utili e molto utili).

Emerge che i profili più utili sono quelli informatici seguiti da *data scientist* e statistici. Questo risultato rafforza quindi il riferimento teorico del modello *pi shaped* come utile a estrarre valore conoscitivo dai dati, oltre all'informatica è necessaria una competenza di tipo statistico. Seguono gli esperti di dominio e analisti dei dati e metodologi, le ultime categorie in ordine di importanza sono gli esperti di *data visualization* e i matematici. In pochissimi casi vengono segnalati come poco utili le competenze matematiche, statistiche e di *data visualization*. In nessun caso i profili segnalati vengono considerati per nulla utili.

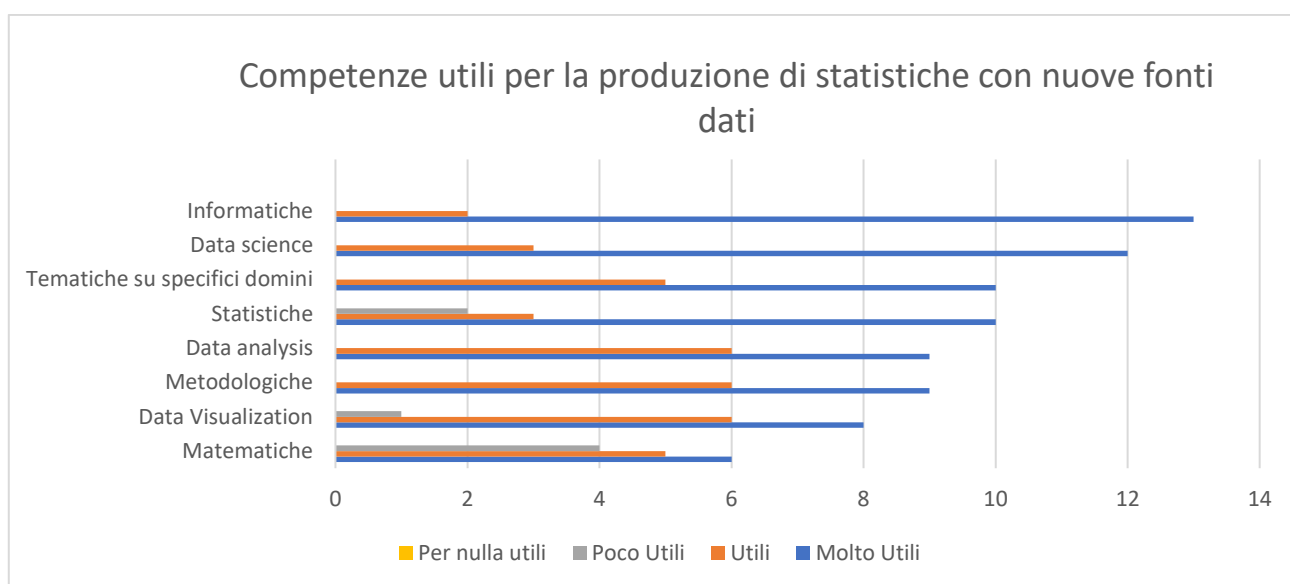


Grafico 8 - Competenze più utili alla produzione di TSS

La domanda successiva chiede invece su quali famiglie di competenze l'Istituto di appartenenza sta investendo per migliorare la produzione di *output* statistici con le nuove fonti di dati e il grado di importanza che ciascun rispondente attribuisce a ciascuna di esse.

Risulta che gli Istituti stanno investendo su competenze metodologiche e relative alla *data science*, quindi su professionalità in grado di migliorare le capacità di creazione di nuovi modelli e *framework* metodologici utili all'analisi dei dati, per scoprire relazioni nascoste tra i dati, per l'estrazione di informazione dai dati, per individuare strumenti e applicazioni per ottimizzare la raccolta, l'analisi e la gestione dei dati

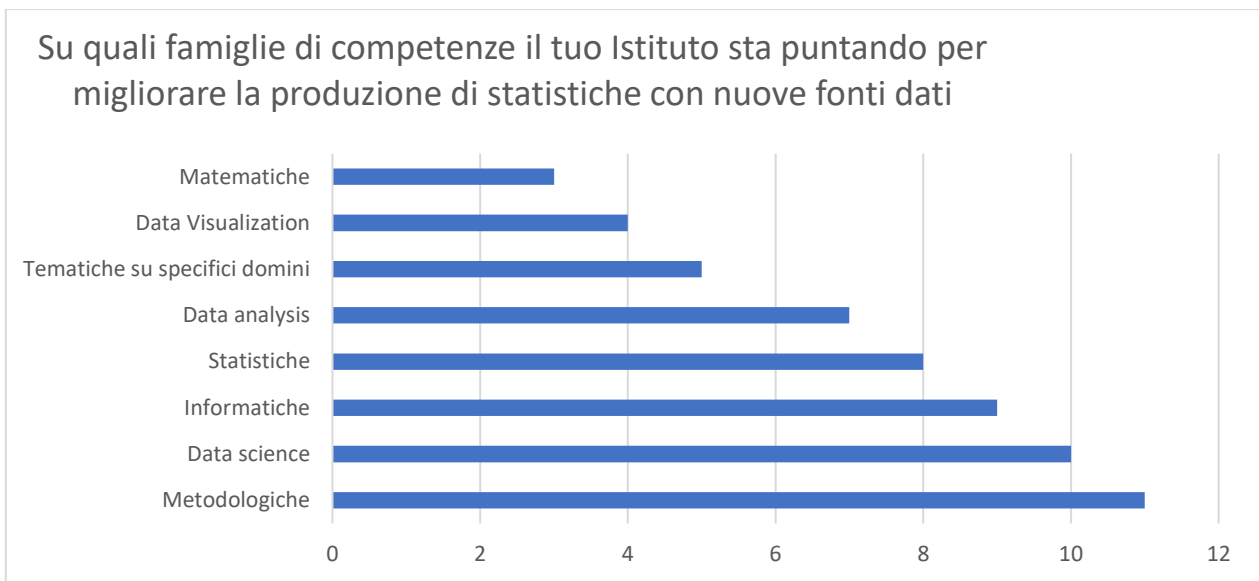


Grafico 9 - Competenze su cui gli Istituti stanno investendo

Il questionario si chiude con la richiesta di una valutazione sui principali vantaggi ottenibili dall'uso dei big data nella produzione di nuovi *output* statistici. Anche qui per ciascuno degli *item* scelti i rispondenti hanno attribuito un grado di importanza (per nulla importante, poco importante, importante, molto importante).

Il vantaggio a cui è attribuita maggiore importanza è la riduzione dell'onere sui rispondenti, seguito dalla produzione di nuovi indicatori e dalla possibilità di indagare fenomeni emergenti. Un alto grado di importanza viene anche attribuito alla tempestività e alla granularità territoriale. Le TSS ancora non sembrano poter contribuire al miglioramento della reputazione degli Enti e poca importanza viene anche data alla riduzione dei costi.

Vantaggi dell'uso dei big data nella produzione statistica

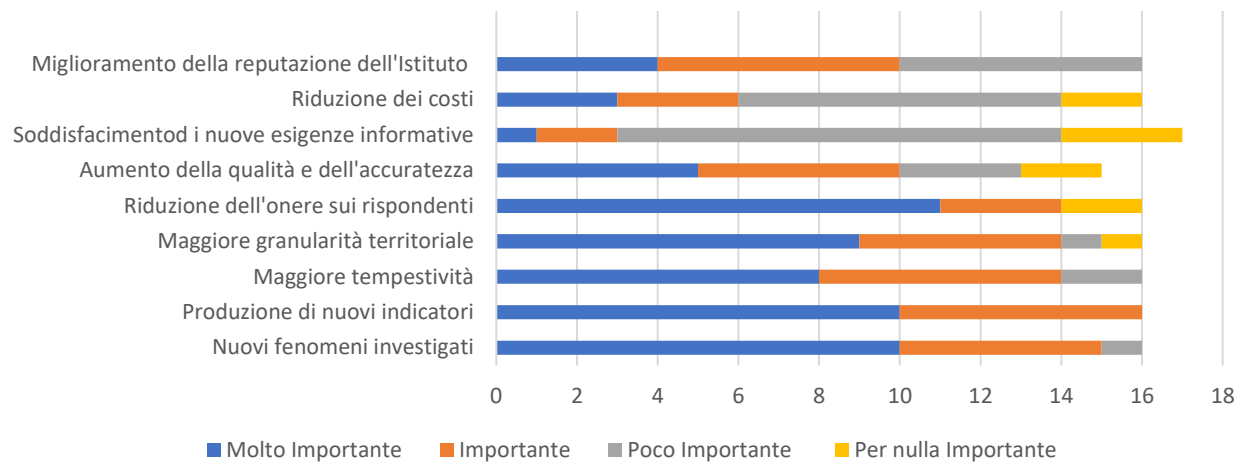


Grafico 10 - Vantaggi ottenuti dall'integrazione dei big data nei processi di produzione

Allegato n. 4 Il questionario

Dear colleagues,

This questionnaire is an integral part of a Doctoral Thesis at La Sapienza University of Rome, Department of Communication and Social Research (CoRiS), in collaboration with the Directorate for Methodology and Design of Statistical Processes of the National Institute of Statistics (ISTAT).

The cognitive objective of the research project is to detect the contribution provided by Big Data to the improvement of statistical production, the development of Trusted Smart Statistics in European Statistical Institutes, as well as to the investments already made, and those in the implementation phase who develop statistical products with the use of new data sources.

Therefore - considering your specific experience on the subject - your contribution, together with that of other representatives of the National Statistical Institutes in Europe, will offer useful information for the development of a comparative analysis between European countries in order to highlight best practices and possible suggestions on future developments.

We thank you for the precious help you will want to give to the research and we ask you to answer all the questions in order to provide us, in each of the investigated areas, with a more complete view on the results being achieved and the tools used to reach them, together with the strategic objectives of long term shared by the Board of your Institute.

Where possible, please report any links to useful materials for further information on the individual items investigated.

All data will be processed exclusively for scientific research purposes and in compliance with current legislation on *privacy*.

We appreciate and thank you for your collaboration.

1 Country

2 Organisation

3 Name of the person filling the questionnaire

4 Organizational structure of belonging

5 Function Performed

- Data Analyst
- IT
- Methodologist
- Domain Expert
- Expert in Communication and Dissemination of *Output*
- Data Scientist
- Data Visualization Expert
- Statistical
- Other (specify)_____
-

6 Is your institution making investments to benefit the opportunities offered by Big Data?

- Yes
- No

If so, for how long has it been?

7 Does your Institute have a *governance* that directs and monitors the production of statistics with Big Data?

- Yes
- No

If yes, can you provide further information or links to useful resources, if available?

8 What are the investments your Institute has already made to start the production of Trusted Smart statistics in the various sectors indicated below?
(Can you possibly provide links to useful information for each of the sectors, if available?)

Methodology

Information Technology

Communication and Dissemination

Organization

Human Resources

Legal sector (for everything concerning the need to mitigate *privacy* issues, in relation to the use of new data sources as well as to define forms of partnership and collaborations.)

9 What are the investments your Institute has already planned to enhance for the production of Trusted Smart Statistics in each of the sectors indicated?

(Can you possibly provide links to useful sources/facts/information for each of the sectors, if available?)

Methodology

Information Technology

Communication and Dissemination

Organization

Human Resources

Legal sector (for everything concerning the need to mitigate *privacy* issues, in relation to the use of new data sources as well as to define forms of partnership and collaborations.)

10 Based on your experience, what are the investments that you think your Institute should make to enhance the production of Trusted Smart Statistics in each of the sectors indicated?

Methodology

Information Technology

Communication and Dissemination

Organization

Human Resources

Legal sector (for everything concerning the need to mitigate *privacy* issues, in relation to the use of new data sources as well as to define forms of partnership and collaborations.)

11 What kind of data do you use and / or have you used to produce new statistical outputs from Big Data sources?

- Social Networks: Facebook, Twitter...
- Images: Instagram, Flickr, Picasa
- Satellite Images
- Video (Youtube ...)
- Maps
- Mobile Phone Data
- Financial Transactions
- Transactions on E-Commerce Sites
- Sensor Data
- Scanner Data
- Web Data
- Other (specify)

12 Did you have direct access to this data?

- Yes
- No

If yes, through which techniques? (web scraping, API access, NLP processing, etc.)

12 a On the other hand, if the data were obtained thanks to a collaboration with subjects outside the Institute, can you indicate the type of partner?

- Public Research Bodies Institutes
- Private Research Institutes
- Universities or Academic Institutions
- IT Companies
- Telephone Operators
- Government Institutions
- Other (specify)_____

13 Did the external partner providing the data, being mentioned previously, also participate in other phases of the statistical process?

- Yes
- No

If yes, can you specify which phases are you referring to?

14 What software do you use in the various phases to implement the techniques referred to in question 12? (E.g. : Scrapy, I-Macros, NLTK, PyTorch-NLP, TensorFlow, SciKit learn etc)

15 What processing methodology / technique do you use for the Big Data inference phase? (E.g. : machine learning, deep learning, social network analysis ...)

16 What problems have you encountered in data processing? (Eg: software with low computing power, low storage space, *privacy*, access to data, organizational problems, etc ...).

16 a What main solutions have you adopted to overcome them?

17 What are the thematic areas of the Trusted Smart Statistics already produced or in the production phase (also as experimental statistics)?

- population
- prices
- health
- transport
- tourism
- agriculture
- international trade
- economy
- environment
- education and training
- Other (specify)

Can you possibly provide links to useful materials, if available?

18 Have you had the opportunity to check user feedback on the Trusted Smart Statistics produced?

- Yes

If yes, what are the aspects detected by users that you consider most significant?

- No

20. In your opinion, what is the most functional platform for the dissemination of Trusted Smart Statistics?

- Institutional Website
- Facebook
- Twitter
- YouTube
- Other (specify)_____

Can you explain why?

21. Which skills have been most useful in producing Trusted Smart Statistics?

- Methodological
- Statistical
- Mathematical
- IT
- Data Analysis
- Data Science
- Data Visualization
- Domain Experts
- Other (specify)

22. What skills is your Institute focusing on to support the Trusted Smart Statistics production?

- Methodological
- Statistical
- Mathematical
- Statistical
- IT
- Data Analysis
- Data Science
- Data Visualization
- Domain Experts
- Other (specify)

23. Below are some of the advantages attributed to the use of Big Data in the production of statistical outputs. Assign a value of 1 (not important at all), 2 (not important), 3 (important), 4 (very important) to each of the items listed.

Indicators				
New Phenomena Investigated	1	2	3	4
Production of New Indicators	1	2	3	4
Better Timelines				
Greater Territorial Granularity	1	2	3	4
Reduction of Burden on Respondents	1	2	3	4
Increased Quality / Accuracy of Statistical <i>Output</i>				
Fulfilment of New Cognitive Questions				
Cost Reduction	1	2	3	4
Improvement of the Institution's Reputation	1	2	3	4
Other (specify)	1	2	3	4

Thanks for your valuable contribution. If you are interested in knowing the results of the survey, leave your email

Conclusioni

Tirare le somme di questo lavoro è un compito arduo. Il finale è completamente aperto a ulteriori sviluppi e approfondimenti. Il nuovo ambiente digitale in cui ci muoviamo tutti noi e anche la comunità statistica è in continua evoluzione. Le opportunità e gli scenari che si prospettano sono innumerevoli e promettenti. La statistica ufficiale continuerà a rimettersi in gioco per intrecciarsi a queste trasformazioni continue e continuare a raccontarle. L'Istat è da anni coinvolto in un processo di innovazione e di evoluzione permanente che gli ha permesso di mantenere indiscusso il suo ruolo di produttore di informazione statistica per varie tipologie di *stakeholders*.

Il nuovo sistema di produzione delle nuove statistiche, verso cui l'Istat si sta muovendo, è multidimensionale, la sua complessità risiede proprio nel fatto che abbraccia e coinvolge tantissimi aspetti, e su ciascuno di essi bisogna lavorare per garantire un avanzamento *sostenibile* e coerente verso l'implementazione del nuovo impianto produttivo. Tutte le dimensioni coinvolte offrono possibilità di approfondimento anche in vista delle possibili reali applicazioni nel contesto Istat.

Sicuramente come ho illustrato nelle pagine precedenti le sfide che hanno posto le nuove sorgenti di dati agli Istituti di statistica sono state colte e hanno dato risultati sempre più vantaggiosi in termini di arricchimento dell'offerta statistica, ma hanno lasciato spazio a sfide nuove, hanno lasciato aperti molti ambiti di ricerca e approfondimento che richiedono interazioni e collaborazioni sempre più forti con altri attori pubblici e privati. Si tratta di strade intraprese già, ma che devono essere necessariamente percorse fino in fondo per far sì il patrimonio di dati prodotti che tutti noi produciamo quotidianamente possa essere trasformato in informazione statistica di cui fidarsi e diventare bene comune. Fra gli ambiti da continuare ad esplorare con rigore riguardano per esempio le policy sull'uso etico delle nuove fonti di dati, gli strumenti di *explainable Artificial Intelligence* e anche qui l'etica di questi strumenti, *l'input privacy*, le Linee Guida per l'uso delle nuove sorgenti di dati per la statistica ufficiale, l'implementazione di nuovi paradigmi di comunicazione per veicolare nel modo

più adeguato i nuovi *output* statistici. Ciascuna di queste dimensioni garantisce possibilità di riflessioni multidisciplinari per coglierne impatti e potenzialità e proporre possibili applicazioni. I risultati del questionario hanno confermato queste conclusioni aperte. C'è consapevolezza della bontà dei risultati già consolidati (tra cui ricordiamo ancora una volta le statistiche sperimentali) e delle ulteriori potenzialità di nuovi dati, metodi e strumenti in via di ulteriore sperimentazione, ma manca la piena consapevolezza che risultati migliori siano possibili grazie a grossi cambiamenti organizzativi che non coinvolgano solo aspetti metodologici e informatici. La contezza di ciò l'ho acquisita anche grazie alla mia esperienza fuori dall'Istat, la contaminazione del mio percorso professionale in Istat con l'esperienza accademica mi ha permesso di comprendere ancora di più gli impatti e l'importanza dell'interdisciplinarietà come valore fondante di ricerca e innovazione. Le TSS non sono solo un nuovo prodotto statistico, sono un nuovo modo di pensare, di proiettare i mutamenti esterni all'interno dei sistemi di produzione statistica, di usare i big data come nuovi misuratori sociali e come strumenti messi a disposizione dall'ambiente digitale.

Questo lavoro racconta lo stato dell'arte delle innovazioni di prodotto e di processo che si sono susseguite negli ultimi anni e che auspicabilmente continueranno al fine di rafforzare la ricaduta in produzione degli investimenti in ricerca e innovazione su cui l'Istituto è coinvolto, ma racconta anche le innumerevoli opportunità che si aprono per i sistemi statistici in generale. Per dirla con Beck, è necessaria per la comunità statistica una "*forma riflessiva di scientizzazione*": le statistiche ufficiali richiedono molto più dei metodi statistici, metodologici e delle tecnologie dell'informazione che rappresentano una parte della soluzione, strumenti fondamentali ma non esclusivi.

Serve una forte consapevolezza, serve la capacità, come ha più volte suggerito uno dei rispondenti alla *survey*, di *thinking of the box*, serve la conoscenza dei contesti esterni e una capacità di contaminarsi, di coinvolgere, di rafforzare relazioni con tutte le tipologie di utenti, con i cittadini, con il mondo accademico, con le Istituzioni pubbliche e private.

È anche grazie a questo che si rafforza la *social trust* che, in sociologia, è potenza, è collante dei gruppi e motore invisibile del progresso sociale.

Bibliografia

AGCOM Autorità per le garanzie nelle comunicazioni “Indagine conoscitiva sui big data”, 2018A.

AGCOM Autorità per le garanzie nelle comunicazioni, Servizio economico statistico, **Rapporto sul consumo di informazione, febbraio 2018.**

AGCOM, Autorità per le garanzie nelle comunicazioni, News vs fake news nel sistema dell'informazione, novembre 2018.

AGID Agenzia Italiana per lo sviluppo digitale, L'Intelligenza artificiale al servizio del cittadino, <https://ia.italia.it/assets/librobianco.pdf>, 2018

Alleva G., Focus: Pubblica Amministrazione: Digitalizzazione, Informazione e Dati, In Nuova Etica Pubblica Rivista dell'Associazione Etica PA, Anno 6, n. 11 – luglio 2018

Alleva G., Relazione di apertura, Dall'incertezza alla decisione consapevole: un percorso da fare insieme Atti della 13a Conferenza nazionale di statistica Roma, 4-6 luglio 2018

Amaturo E., Aragona B., Per un'epistemologia del digitale: note sull'uso di big data e computazione nella ricerca sociale, *Quaderni di Sociologia*, 81- LXIII | 2019.

Anderson C., The end of Theory: the data deluge makes the scientific method obsolete, Wired 2008.

Arbia, G., Statistica, nuovo empirismo e società nell'era dei big data, Edizione Nuova Cultura, 2018.

Barcaroli G., De Francisci S., Righi P., Scannapieco M., Big data: si apre un'enorme possibilità per la statistica ufficiale, pubblicato il 28 gennaio 2016 su ForumPa. <https://www.forumpa.it/pa-digitale/big-data-si-apre-unenorme-opportunita-per-la-statistica-ufficiale/>

Batini, C., Datacy, perché una scienza per studiare l'impatto dei dati sulla società. <https://www.agendadigitale.eu/cittadinanza-digitale/datacy-perche-una-scienza-per-studiare-limpatto-dei-dati-sulla-societa/> Articolo pubblicato su [agendadigitale.eu](https://www.agendadigitale.eu) il 19/03/2018.

- Bauman Z.**, *Modernità liquida*, Laterza 2013.
- Biggeri L.**, I principi e caratteristiche della statistica ufficiale tra passato e futuro, Atti della VII Conferenza nazionale di statistica, Istat, 2004
- Boccia Altieri G., Marinelli A.**, Piattaforme, algoritmi e formati. Come sta evolvendo l'informazione *online* in *Problemi dell'informazione* – Anno XLIII, N. 3, Dicembre 2018.
- Boccia Artieri G.**, Fare sociologia attraverso l'algoritmo: potere, cultura e *agency*, *Sociologia Italiana - AIS Journal of Sociology* n. 15, 2020.
- Boccia Artieri G.**, Introduzione al libro *Scienza connessa Rete Media e Social* a cura di Sveva Avveduto, Gangemi Editori 2012.
- Boyd D. and Crawford K.**, Critical Question for Big Data. Provocations for a cultural, technological, and scholarly phenomenon. "Journal of Information Communication & Society", Vol. 15, 2012. <http://dx.doi.org/10.1080/1369118X.2012.678878>
- Calude, C.S., Longo, G.** The Deluge of Spurious Correlations in Big Data. *Found Sci* 22, 2017.
- Carciotto A., Signore M.**, Improving relevance: Istat experience on experimental statistics , *Statistical Journal of the IAOS* 37 (2021) 593–601 593 DOI 10.3233/SJI-200764 IOS Press
- Cassese, S.** *L'arena pubblica. Nuovi paradigmi per lo Stato*. Rivista trimestrale di diritto pubblico Numero 3, 2001
- Castells, M.**, *Comunicazione e potere*, Università Bocconi Editore, Milano, 2009.
- Castells, M.**, *La nascita della società in rete*, Università Bocconi, 2014.
- Cesarano, A.I.**, *Social Network Society*, Rivista di scienze sociali, 2020.
- Ceri S.**, On the role of statistics in the era of big data: A computer science perspective, *Statistics & Probability Letters*, Volume 136, 2018, Pages 68-72, ISSN 0167-7152.
- Chandler D.**, *A World without Causation: Big Data and the Coming of Age of Posthumanism*. Millennium. 2015.
- Commissione Europea**, libro bianco sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_it.pdf, 2020.

Commissione Europea Regolamento (CE) n . 223/2009 del Parlamento europeo e del Consiglio dell' 11 marzo 2009 relativo alle statistiche europee (Testo rilevante ai fini del SEE e per la Svizzera).

Commissione Europea Regolamento (UE) 2015/759 del Parlamento europeo e del Consiglio, del 29 aprile 2015, che modifica il regolamento (CE) n. 223/2009, relativo alle statistiche europee (Testo rilevante ai fini del SEE e per la Svizzera).

Creteur J. Big data are here to stay! *Anaesth Crit Care Pain Med.* 2019 Aug;38(4):339-340.

doi: 10.1016/j.accpm.2019.05.001. Epub 2019 May 9. PMID: 31078797

De Mauro, A. Impariamo a conoscere i big data. Articolo apparso su apogeeonline <https://www.apogeeonline.com/articoli/impariamo-a-conoscere-i-big-data-andrea-de-mauro/>, 15 novembre 2019.

De Ruvo V., Il decreto delegato sui sistemi informativi automatizzati delle amministrazioni pubbliche, *Informatica e diritto*, XX annata, Vol. III, 1994, n. 1, pp. 121-131

Du Gay, P., Pryke, M., *Cultural Economy: Cultural Analysis and Commercial Life*, London: Sage, 2002.

Epifani S., *Sostenibilità digitale*, Digitale Transformation Institute, 2020.

European Statistical System - ESS, Directors General of the National Statistical Institutes - DGINS. 2013. Scheveningen Memorandum. Big Data and Official Statistics. Scheveningen – The Hague, The Netherlands, 25th – 27th September 2013: DGINS Conference.

European Statistical System - ESS, Directors General of the National Statistical Institutes - DGINS. 2018. Bucharest Memorandum on Official Statistics in a Datafied Society (Trusted Smart Statistics), Bucharest, 10th and 11th October 2018. 104thDGINS Conference

European Statistical System, Big Data Action Plan and Roadmap, Task Force composed of representatives of national statistical offices, the OECD, the UNECE, DG CONNECT, the JRC and academic expert, European Statistical System Committee, 22nd meeting on 26 Sep 2014 in Riga.

European Statistical System: ESSnet Big Data II Grant agreement Workpagage Coordination and communication, Derivable A4 Final version, 29 June 2021

European statistical system Power from statistics, Final Report, Eurostat 2018

European statistical system Item 1b. Beyond ESS Vision 2020: Innovation actions implementing the multiannual statistical programme 2021-2027, 39th Meeting of the European Statistical System Committee, 7 February 2019

European Statistical System: Strategic Plan 2020-2024

European Statistica System European statistical programme 2020 – 2017

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37. <https://doi.org/10.1609/aimag.v17i3.1230>, 1996

Firmani, D., Mecella, M., Scannapieco, M. et al. On the Meaningfulness of “Big Data Quality” (Invited Paper). *Data Sci. Eng.* 1, 6–20 (2016). <https://doi.org/10.1007/s41019-015-0004-7>

Floridi L., La quarta rivoluzione, Raffaello Cortina Editore, 2017

Futur of life Institute, I 23 principi di Asilomar, <https://futureoflife.org/ai-principles/>, 2017.

Garante per la protezione dei dati personali, doc. web n. 9802796] Parere sui lavori statistici IST 02834 Studio dei Mobile Network Data a fini statistici e IST 02829 La violenza raccontata dai social - Registro dei provvedimenti n. 235 del 9 giugno 20202, disponibile al link [www.garanteprivacy.it/home/docweb/-/docweb display/docweb/9802796#note](http://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9802796#note)

Giannotti F., Nanni M., Thanos C., Rauber A., Big Data Analytics: towards a European research agenda, White paper ERCIM, 2 gennaio 2015.

Giannotti, F., D. Pedreschi, Il bias dell’agoritmo e la polarizzazione delle opinioni, Gnosys 2/2019.

Greco P., Il quarto paradigma, *Micron*, n. 23 ottobre 2013, consultabile al link: https://www.arpa.umbria.it/resources/docs/micron%2023/06-MICRON_23.pdf

Gruppo di esperti nominati dal Ministero per lo Sviluppo Economico (MISE), Proposte per una Strategia italiana per l’intelligenza artificiale, 2020.

Gruppo indipendente di esperti di alto livello sull’intelligenza artificiale istituito dalla Commissione Europea nel giugno 2018, Orientamenti etici per un’IA affidabile, <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>, 2019

<https://www.agcom.it/documents/10179/3744102/Allegato+22-11-2018/3aff8790-8039-4456-8f9a-dae2497289a4?version=1.0>, novembre 2018C

Hassain, J. *Disinformation in Democracies: Improving Societal Resilience to Disinformation*. Riga: NATO Strategic Communications Centre of Excellence 2022

Hey, T., Tansley, S. Tolle, K. Jim Grey: *A transformed scientific method*, in Hey, T., Tansley, S. and Tolle, K. (eds) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research; Redmond, Washington. xvii-xxi, 2009.

Istat. Big Data Committee – Annual Report ISTAT, Chapter 1 –Alleva Giorgio.

<https://www.istat.it/it/files//2018/09/Big-data-committee.pdf>

Istat, Rapporto Annuale 2022. La situazione del Paese, 2022

Leti G., Cerbara L. *Elementi di statistica descrittiva*, Il Mulino Manuali, 2009.

Kennedy H., How People Feel About What Companies do with Their Data is Just as Important as What They Know About It. *Impact of Social Sciences Blog*. 2018

Kitchin R., (2017) Thinking critically about and researching algorithms. *Information, Communication & Society* 20(1) 14–29 [http:// dx.doi.org/10.1080/1369118X.2016.1154087](http://dx.doi.org/10.1080/1369118X.2016.1154087).

Kitchin R., McArdle G., What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets in Big data and Society, 2016, <https://doi.org/10.1177/2053951716631130>

Kitchin, R., Big Data and Official Statistics: Opportunities, Challenges and Risks. *SSRN Electronic Journal*. 10.2139/ssrn.2595075, 2015

Laney D., *Application Delivery Strategies*, Meta Group, 2001

Latour, B., Tarde's idea of quantification in *The Social after Gabriel Tarde: Debates and Assessments*, Edited by: Candea, M. 145–162. London: Routledge, 2009.

Lazarsfeld P. F., *Interpretation of statistical relations as a research operation*, in Lazarsfeld P.F., Rosenberg M. (a cura di), *The language of social research: A reader in the methodology of social research*, Glencoe, Free Press, 1955.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. e Van Alstyne,

M. Social Science: Computational Social Science'. *Science* 323 (5915), pp. 721–723. DOI: 10.1126/science.1167742, 2009

Licklider, J. C. R., Man–computer symbiosis. *IRE Transactions on Human Factors in Electronics*, 1, 1960

Longo A., Scorza G., *Intelligenza Artificiale, L'impatto sulle nostre vite, diritti e libertà*, Mondadori Università, 2020.

Lorusso, A., *Postverità*, Laterza, Roma-Bari 2019.

MacFeely S., *The Continuing Evolution of Official Statistics: Some Challenges and Opportunities*, *Journal of Official Statistics*, pp. 789–810 Vol. 32, No. 4, 2016.

Marinelli, A., Andò R., *Television(s). Come cambia l'esperienza televisiva tra tecnologie convergenti e pratiche social*, Guerini e Associati, 2018

Marres N., *Digital sociology: The reinvention of social research*, New York, John Wiley & Sons, 2017.

Marr, B., Big data: The 5 vs everyone must know. *LinkedIn Pulse* 6, 2014.

Marz, N., Warren J., *Big Data: Principles and best practices of scalable realtime data systems*. Hanning 2015

'O Neil C., *Armi di distruzione matematica. Come i big data aumentano la disuguaglianza e minacciano la democrazia*, 2017.

OECD, *Artificial Intelligence in Society*, OECD Publishing, Paris, <https://doi.org/10.1787/eedfee77-en>, 2019.

OECD, *Data driven innovation: Big Data for Growth and Well Being*, OECD Publications, Paris, consultabile al link https://read.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en#page302, 2015

Pariser E., *The Filter Bubble*, New York, Penguin 2011.

Pedreschi D., *Dalla macchina all'uomo. Tracce digitali per un nuovo umanesimo*, partecipazione alla rubrica Digital Word, RAI Scuola, <http://www.sobigdata.eu/blog/digital-world-dino-pedreschi-rai-television> 8 luglio 2019.

Pedreschi D., *Explanaible AI. Aprire le scatole nere per un'intelligenza artificiale umana*. *Gnosis* 2/2019.

Pedreschi D., <http://www.datamanager.it/2018/09/dino-pedreschi-la-societa-riflessa-nei-big-data/> intervista a Data Manager pubblicato 14 settembre 2018a.

Pedreschi D., <https://www.ilsole24ore.com/art/the-black-box-society-come-fanno-imprese-fidarsi-loro-prodotti-AEc7FIJF> articolo pubblicato il 23 luglio 2018b.

Pedreschi D., Miliou J., Artificial Intelligence (AI): new developments and innovations applied to e-commerce, Policy Department for Economic, Scientific and Quality of Life Policies Directorate-General for Internal Policies, European Parliament, May 2020

Pitrelli, N. (2017). 'Big data e metodi digitali per la ricerca in comunicazione della scienza: opportunità, sfide e limiti'. JCOM 16 (02), C01_it.

Porter, T. M.. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*, Princeton: Princeton University Press, 2020. <https://doi.org/10.1515/9780691210544>

Porter T., *Statistics in the digital era*, in *Power from Statistics: data, information and knowledge*, pag.12, 2018

Pugliesi, R., Big data e scienze sociali. Una nuova sfida alla teoria. Dossier Economia Digitale I Copernicani, 10/2018, 52-59. Consultabile al link: <http://www.fondazionecomunica.org/wp-content/uploads/2018/10/CopernicaniDossier-1-Economia-digitale.pdf>

Quattrociocchi W.; Vicini A., *Liberi di crederci*, codice Edizione, Le Scienze, 2018

Quattrociocchi W.; Vicini A., *Misinformation*, FrancoAngeli Editori, 2016

Quattrociocchi W., Scienza, dati e decisione. Intervista a Walter Quattrociocchi – pubblicato sul numero 3 del 2020 di Pandora Rivista dedicato alle piattaforme.

Radermacher W. *Governing-by-the-numbers/Statistical governance: Reflections on the future of official statistics in a digital and globalised society*, Statistical Journal of the IAOS 35 (2019) 519–537, DOI 10.3233/SJI-190562 IOS Press

Radermacher W., *Literacy in statistics for the public discourse*, Statistical Journal of IAOS, Settembre 2021 747-752

Radermacher W., *Official Statistics 4.0 Verified Facts for people in the 21th Century*, Springer 2020

Ricciato, F., Wirthmann, A., & Hahn, M. (2020). Trusted Smart Statistics: How new data will change official statistics. *Data & Policy*, 2, E7. doi:10.1017/dap.2020.7

Ricciato, F., Skaliotis M., Wirthmann A., Giannakouris K., and Reis F.,. *Towards a Reference Architecture for Trusted Smart Statistics*. In DGINS 2018. Bucharest: Statistics Romania.

Rossi, F., *Il confine del futuro*, Feltrinelli 2020.

Rossi, F., *Intelligenza Artificiale benefica e sicura: iniziative accademiche governative e industriali*, Sistemi Intelligenti, a. XXIX, n. 3, dicembre 2017.

Rossa S., *Contributo allo studio delle funzioni amministrative digitali*, CEDAM 2021

Ruppert, E. *Sociotechnical Imaginaries of Different Data Futures: An Experiment in Citizen Data*. 3e Van Doornlezing. Rotterdam, NL: Erasmus School of Behavioural and Social Sciences., 2018

Russel S., Norvig P., *L'Intelligenza Artificiale. Un approccio moderno*. Terza Edizione Pearson 2010.

Scannapieco M, Righi A, *Come cambia la produzione statistica con i Big Data: il Centro e la Roadmap per le TSS*, Presentazione alla XIV Conferenza Nazionale di Statistica, all'interno della sessione Trusted Smart Statistics: sfide e opportunità, 30 novembre 2021

Sistan Codice italiano per la qualità delle statistiche, Direttiva n. 12/Comstat Pubblicata nella G.U. n. 23 (Serie Generale) del 29 gennaio 2022.

Spiegelhalter D., *L'arte della statistica. Cosa ci insegnano i dati*. Einaudi I Maverick, ,2020

UNECE United Nations Economic Commission for Europe Conference of European Statisticians Sixty-second plenary session Paris, 9-11 April 2014 In-depth review of big data Prepared by the temporary Task Team on Big Data and the Secretariat

https://unece.org/DAM/stats/documents/ece/ces/2014/7-In-depth_review_of_big_data.pdf

UNECE United Nations Outcomes of the UNECE Project on Using Big Data for Official Statistics – 2015

<https://statswiki.unece.org/display/bigdata/Big+Data+Projects?preview=/77170975/12245402/2/Outcomes%20of%20the%20UNECE%20Project%20on%20Using%20Big%20Data%20for%20Official%20Statistics.docx>

UNECE – HLG MOS How Big is Data? Exploring the role of Big Data in Official Statistics 2014 <https://statswiki.unece.org/pages/viewpage.action?pageId=99484307&preview=/99484307/99451129/Virtual%20Sprint%20Big%20Data%20paper.docx>

UNECE United Nations – Conference of European Statisticians “What does big data means for official UNECE statistics?” – Marzo 2013
<https://statswiki.unece.org/pages/viewpage.action?pageId=77170622>

UNECE United Nations HLG MOS “What does Big data means for official statistics”
Paper di conclusione dell’HLG MOS al Seminario Seminar on Modernization of Statistical Production and Services (San Pietroburgo Russia, 3-5 October 2012)
<https://statswiki.unece.org/pages/viewpage.action?pageId=77170622>

UNECE United Nations. Fundamental principles of official statistics. Official Resolution adopted by the UN General Assembly on 29/1/2014. <https://unstats.un.org/unsd/dnss/gp/fp-new-e.pdf>

Vespignani A., L’algoritmo e l’oracolo, Il Saggiatore, 2019.

Weinberger D., La stanza intelligente: La conoscenza come proprietà della rete, Torino, Codice Edizioni, 2012

Zook M., Barocas S., Boyd D., Crawford K., Keller E., Gangadharan SP, et Al. Ten simple rules for responsible big data research. PLoS Comput Biol 13(3): e1005399. <https://doi.org/10.1371/journal.pcbi.1005399> , 2017

Zuliani A., L’evoluzione dell’informazione statistica, le indagini di opinione e le sfide della statistica ufficiale, Istat 2015 disponibile su <https://www.istat.it/it/files/2011/05/Zuliani.pdf>

Ringraziamenti

Ringrazio innanzitutto Piero Falorsi, il *Director*, mio Direttore in Istat, per i suoi continui riferimenti alla rilevanza, alla qualità, all'importanza di volare alto, di guardare oltre perché è lì che si innescano i processi nuovi, quelli che portano innovazione veramente. E ringrazio Monica Scannapieco, che anni fa ci ha visto lungo e mi ha introdotto in un covo di nerd in cui mai avrei pensato di trovare spazio, invece ho trovato uno spazio pure confortevole, e la ringrazio perché è da dentro al covo che ho colto il valore dell'interdisciplinarietà e che ho capito che dei nerd non bisogna avere paura. Tutto questo è stato fondamentale anche per la stesura di questa tesi.

E ringrazio i miei genitori - mio padre ne sarebbe veramente molto fiero - perché mi hanno mostrato l'etica del lavoro, e hanno alimentato il mio forte senso istituzionale e la mia voglia di fare.

Ringrazio il Coris e la Direzione per la Metodologia e il disegno dei processi statistici dell'Istat per avermi dato l'opportunità di lavorare e studiare, di mettere a confronto due mondi, di intrecciare un'esperienza di formazione così significativa nel mio percorso professionale, ho riportato nella tesi le mie esperienze professionali in Istat e in Istat le nuove competenze acquisite in questi tre anni, uno scambio per me estremamente proficuo.

Ringrazio il Prof. Alberto Marinelli e il Prof. Stefano Epifani, miei tutor di questo percorso, che mi hanno indicato la direzione da prendere e offerto una chiave di lettura nuova. E ringrazio il Prof. Ruggiero per le risposte che ha sempre dato.

E ovviamente un ringraziamento speciale va ai maschi Jona, è solo grazie al meraviglioso casino che mi creano intorno ogni giorno, che ho trovato la forza di mettere ordine, trovare uno spazio e fermarmi, continuare a lavorare ma anche frequentare lezioni, seminari, conferenze e convegni e leggere e studiare e scrivere queste pagine.

E quindi grazie a Pietro per l'amore e soprattutto per le ottime cene che ha preparato, grazie a Davide ed Emanuele che hanno il coraggio di essere esattamente come vogliono essere e che fieri, mi rivolgono il loro sguardo più attonito quando racconto loro che ho scelto io, di mia spontanea volontà, senza nessuna costrizione, di "tornare a scuola".

Infine un pensiero va ai colleghi del XXXV, e un pensiero veramente speciale va a due di loro, a Patrizio e Michele che difficilmente mi scrollerò di dosso!! Per fortuna...