Indagine conoscitiva sulla digitalizzazione ed interoperabilità delle banche dati fiscali

Audizione dell'Istituto nazionale di statistica
Dott. Massimo Fedeli

Direttore della Direzione Centrale per le tecnologie informatiche

Commissione parlamentare di vigilanza sull'anagrafe tributaria

20 ottobre 2021

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

Indice

Introduzione **5**

1. Il programma di modernizzazione dell'Istat e la trasformazione tecnologica e digitale dell'Istituto **6**

2. Le direttrici del cambiamento: tecnologia, metodologia, organizzazione **7**

2.1 Il Sistema Integrato dei Registri (SIR) *9*

2.2 Il Register-based Analytics Framework (RAF) *10*

2.3 Il sistema informativo EDAMIS e il protocollo SDMX *12*

2.4 Il progetto Hub della Statistica Pubblica e la piattaforma SDMX Istat Toolkit *13*

2.5 L'approccio strutturato alla standardizzazione delle piattaforme: il Generic Statistical Business Process Model *15*

2.6 La qualifica a Polo Strategico Nazionale: il percorso verso un Software Design Data Center e il private cloud *16*

3. L'esperienza maturata dall'Istat nell'interoperabilità delle banche dati fiscali per la produzione statistica **17**

4. L'Istat e l'interoperabilità nella PA: prospettive di sviluppo **19**

Appendice

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

Introduzione

Vorremmo innanzitutto ringraziare la Commissione per l'opportunità offerta di descrivere la qualità delle infrastrutture tecnologiche e dei servizi digitali dell'Istituto e, in particolare, le loro prospettive di sviluppo, anche in riferimento al progetto di Cloud nazionale.

In questa audizione richiameremo dapprima il programma di modernizzazione dell'Istat, avviato nel 2016, e l'impatto che esso ha avuto sulla trasformazione tecnologica e digitale dell'Istituto.

Verranno quindi descritte alcune delle principali iniziative intraprese nell'ambito degli interventi tecnologici e infrastrutturali di modernizzazione (che si intrecciano inevitabilmente ai cambiamenti metodologici e organizzativi succedutisi), che hanno caratterizzato l'attività dell'Istituto in questi anni. Particolare attenzione verrà dedicata alla creazione di una semantica comune delle informazioni volta a standardizzare il più possibile l'organizzazione dei dati, rendendo più efficiente la loro integrazione e interpretazione tra i vari domini statistici (si veda il paragrafo 2.1 sul Sistema Integrato dei Registri); altro tema rilevante è quello della valorizzazione delle informazioni mediante la creazione di soluzioni di *data science* e *business analytics* (si veda il paragrafo 2.2 sul *Register-based Analytics Framework*).

In generale, gli obiettivi del processo di rafforzamento tecnologico dell'Istituto perseguiti negli ultimi anni possono essere così sintetizzati:

- ✓ Accessibilità e Fruizione: facilitare l'accesso e l'analisi dei dati creando un ecosistema unico *multi-purpose* e disporre di strumenti avanzati per la fruizione del dato
- ✓ Efficienza e Rapidità: incrementare l'incisività dell'Istituto con infrastrutture informative che assicurino risposte "agili" e flessibili
- ✓ Valorizzazione e Consapevolezza: incrementare e facilitare l'utilizzo di basi dati sempre più complesse e granulari; introdurre competenze innovative, in linea con le esigenze di contesto; sviluppare percorsi formativi che rafforzino il know-how dell'Istituto
- ✓ Innovazione e Continuità: implementare le iniziative di cambiamento abilitate dalle tecnologie informatiche e garantire la continuità operativa
- ✓ Confidenzialità e Sicurezza: perseguire e realizzare iniziative tecnologiche sulla confidenzialità dell'informazione statistica e incrementare la consapevolezza sulle tematiche GDPR.

Nelle ultime pagine di questo documento ci soffermeremo sul tema dell'utilizzo delle banche dati fiscali nell'ambito dei processi di produzione statistica, per cui l'Istituto ha consolidato negli anni diverse soluzioni tecnologiche, metodologiche ed organizzative.

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

In chiusura, verranno fornite alcune considerazioni generali sulla partecipazione e il ruolo dell'Istat al tavolo trasversale sulla Data Strategy Nazionale.

1. Il programma di modernizzazione dell'Istat e la trasformazione tecnologica e digitale dell'Istituto

Negli ultimi anni l'Istat ha avviato un importante processo di modernizzazione, al fine di arricchire l'offerta e la qualità delle informazioni statistiche prodotte e migliorare l'efficacia e l'efficienza delle attività dell'Istituto.[1] Sul piano statistico, il Programma di modernizzazione prevede il passaggio da un modello basato prevalentemente sull'acquisizione diretta dei dati da cittadini e imprese attraverso le indagini, ad uno basato sull'integrazione di dati individuali provenienti da una pluralità di fonti disponibili (indagini, archivi amministrativi e nuove fonti, come i Big Data); uno degli obiettivi di tale Programma è senza dubbio quello di favorire l'armonizzazione e la standardizzazione dei processi e dei prodotti e ridurre la duplicazione di soluzioni non ottimali sviluppate all'interno di ogni singolo processo.

Coerentemente con il nuovo modello produttivo, è stato necessario rinnovare il modello organizzativo dell'Istituto per:

- ✓ accelerare l'evoluzione dei processi produttivi delle statistiche, i quali devono rapidamente adeguarsi ai bisogni della società e alla disponibilità di nuove tecnologie
- ✓ disegnare una struttura organizzativa più snella, che si sostituisca al modello a silos, basato su linee produttive condotte in parallelo e in modo indipendente
- ✓ affiancare a tali cambiamenti un meccanismo di governance, in grado di dare attuazione alla nuova strategia.

In tale contesto, in tema di infrastrutture tecnologiche e servizi digitali, l'Istat ha intrapreso un percorso volto a favorire la trasversalità dei servizi IT, con uno sviluppo omogeneo delle applicazioni e delle infrastrutture. Tale approccio garantisce un deciso supporto agli obiettivi della modernizzazione, recependo al contempo gli indirizzi dell'Agenzia per l'Italia Digitale (AgID) per quel che riguarda l'innovazione tecnologica e la razionalizzazione delle infrastrutture IT a livello Paese.

Il Programma di modernizzazione ha anche l'obiettivo di migliorare l'uso dei dati nei vari ambiti dell'Istituto, rendendo più efficiente ed agile l'infrastruttura informatica a supporto della sua operatività. Particolare attenzione, oltre alle tematiche "by design" di privacy, GDPR (*General Data Protection Regulation*) e confidenzialità, è stata posta alle modalità di diffusione dell'informazione statistica agli stakeholder,

[1] Si veda l'audizione dell'Istat alla I Commissione "Affari costituzionali della Presidenza del Consiglio e Interni" della Camera dei deputati, il 4 maggio 2016 (https://www.istat.it/it/archivio/185582).

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

guardando alle diverse tipologie di interlocutore e alla definizione delle modalità di suo uso e/o fruizione.

In particolare, la trasformazione tecnologica e digitale dell'Istat ha operato lungo tre direttrici principali:

- ✓ Tecnologia, come fattore abilitante in termini di *data & user centricity*
- ✓ Metodologia, come elemento distintivo per la gestione del dato e del metadato
- ✓ Organizzazione, come chiave per una corretta governance dell'intero ciclo di vita del dato;

e si è indirizzata, in special modo, verso le seguenti aree tematiche:

- ✓ Centralizzazione e Integrazione dei Dati
- ✓ Standardizzazione del Ciclo di Vita dei Dati
- ✓ Centralità dei Metadati
- ✓ Superamento dell'obsolescenza tecnologica e razionalizzazione delle spese infrastrutturali
- ✓ Razionalizzazione e consolidamento degli applicativi e delle tecnologie.

2. Le direttrici del cambiamento: tecnologia, metodologia, organizzazione

L'*Information Technology* sta rappresentando una leva di continuo miglioramento per la Pubblica Amministrazione, garantendo processi produttivi più efficienti e servizi più adeguati per gli utenti.

Il Programma di Modernizzazione dell'Istat ha sancito il passaggio da una gestione "ad hoc" dei sistemi e dei processi informatici, cuciti sulla base delle esigenze dei singoli dipartimenti, verso un livello di industrializzazione maggiore e trasversale, definendo *best practice*, favorendo la standardizzazione della gestione dell'intero ciclo di vita del dato e contribuendo all'evoluzione repentina e radicale – nonché alla governance completa – dei sistemi informatici, oggi presenti in modo pervasivo in ogni processo produttivo e di ricerca dell'Istituto, vista la mole di dati e di informazioni elaborate.

L'Istat si è dotata, in particolare, di un nuovo framework architetturale per la raccolta, analisi, validazione e diffusione dei dati (Figura 1 in Appendice), che ha permesso di gestire in maniera sinergica e specialistica tutti gli aspetti del ciclo di vita del dato, fornendo contemporaneamente un quadro di riferimento per le progettualità in essere e per quelle future. Tale framework è basato sulla selezione degli approcci e dei prodotti software che permettono di affrontare, indirizzare e risolvere le tematiche interne e gestire, esternamente, le modalità di rappresentazione e fruizione dell'informazione delle varie tipologie di stakeholder.

L'Istituto ha inoltre incrementato i propri metodi di gestione e produzione dei dati, migliorando l'offerta di contenuti di prodotti statistici. Insieme alla riorganizzazione

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

degli archivi dei dati, si è perseguito un costante raffinamento nella gestione del metadato, ovvero delle informazioni a corredo del dato per poterne meglio comprendere il valore, la generalità e particolarità, l'uso e la riusabilità e consentire l'individuazione delle relazioni puntuali tra i diversi set di dati, garantendo una maggiore integrazione e interoperabilità. L'introduzione di questo livello semantico ("ontologie", ovvero la rappresentazione esplicita del dominio di interesse) ha permesso all'Istituto di:

- ✓ incrementare l'interoperabilità tra i dati provenienti dalle molteplici fonti (amministrative e di indagine)
- ✓ validare i contenuti ed effettuare eventuali correzioni/integrazioni del dato
- ✓ comprendere più profondamente e sfruttare ancor di più le potenzialità dei dati raccolti.

Le ontologie sviluppate dall'Istat sono state riconosciute a livello europeo e adottate da AgID come *best practice* nell'ambito della PA Italiana.

L'evoluzione dell'Istituto ha seguito un approccio olistico, che ha visto cambiamenti anche dal punto di vista organizzativo, dei processi e delle competenze del personale. In particolare rispetto a:

- ✓ Organizzazione: l'Istituto ha intrapreso una serie di interventi per aggiornare l'assetto organizzativo e le relative attribuzioni di ruoli e responsabilità, in funzione delle nuove esigenze e per supportare la trasformazione digitale
- ✓ Processi: la maggiore complessità dei sistemi informativi e la loro crescente importanza per il raggiungimento degli obiettivi dell'Istituto ha comportato la necessità di definire dei modelli per il loro governo e di standardizzare ed "industrializzare", per quanto possibile, le attività svolte
- ✓ Persone: per affrontare le sfide della trasformazione digitale, l'Istituto sta promuovendo l'evoluzione delle competenze e delle capacità del proprio personale, in modo da poter supportare il nuovo modello di funzionamento e raggiungere gli obiettivi strategici.

La struttura organizzativa dell'Istat si è così trasformata nel tempo in risposta alle esigenze di digitalizzazione e ai servizi offerti, ai cambiamenti del contesto (come l'introduzione di normative specifiche, le richieste di AgID o a fronte di situazioni contingenti come la pandemia), nonché alle evoluzioni tecnologiche in corso.

In particolare, gli elementi chiave dell'evoluzione organizzativa della Direzione Centrale per le Tecnologie Informatiche sono stati:

- ✓ il passaggio da un modello distribuito al modello centralizzato di gestione dei sistemi informativi, con l'assegnazione della responsabilità di progettare, sviluppare e manutenere i servizi IT
- ✓ l'introduzione del Servizio ITC di Monitoraggio e Governance, che ha introdotto il concetto di Governance IT per standardizzare le attività della direzione e di monitoraggio delle performance

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

- ✓ la modifica dei profili organizzativi della Direzione, con la creazione dei servizi attualmente operativi in cui è messa in evidenza la centralità del ruolo dell'IT nel trattamento dei dati e si aumenta il focus sulla gestione della sicurezza informatica.

L'Istituto ha inoltre identificato i ruoli chiavi preposti alla gestione *data driven* del percorso di valorizzazione del proprio patrimonio informativo, avvalendosi, oltre che all'interno delle aree IT, anche del contributo specialistico e tematico delle altre Direzioni dell'Istituto.[2]

Nel seguito del paragrafo vengono descritte alcune delle principali iniziative intraprese dall'Istat nell'ambito degli interventi tecnologici, infrastrutturali, metodologici e organizzativi sopra richiamati.

2.1 Il Sistema Integrato dei Registri (SIR)

L'utilizzo degli archivi amministrativi per la produzione dei registri statistici rappresenta uno dei pilastri della modernizzazione dell'Istituto. I due principali "canali informativi", da archivio amministrativo e da indagine, si completano infatti a vicenda e costituiscono una modalità incrociata di verifica e analisi statistica.

Uno degli obiettivi strategici individuati dall'Istat negli ultimi anni è stata la costruzione del Sistema Integrato dei Registri statistici (SIR), che, a partire dall'acquisizione di archivi amministrativi e risultati di indagine, fornisce un'architettura dati per analisi annuali e longitudinali nel tempo, intra e *inter-domain*. L'obiettivo del SIR è quello di creare dei dataset di alta qualità per garantire un'offerta informativa ampia e ricca e che consenta a tutti gli stakeholder, amministrazioni, cittadini ed imprese, di effettuare analisi su fenomeni complessi in maniera trasversale, coerente, flessibile, multidimensionale e al livello di dettaglio desiderato, garantendo l'affidabilità dei dati e la tutela della riservatezza. Il SIR rappresenta infatti ad oggi la più autorevole fonte di statistiche ufficiali che integra e certifica dati basati su diverse fonti amministrative, sulle indagini, nonché sulle fonti alternative, come i Big Data.[3]

Le linee guida emanate a livello nazionale ed internazionale vanno nella direzione di fornire le informazioni ai cittadini in maniera semplice, trasparente ed economica e spingono verso l'interoperabilità dei servizi, l'integrazione delle fonti disponibili (*interoperabile by design*) e il riuso dei dati (*once only*). Il SIR risponde appieno a questi requisiti e rappresenta un'innovazione, non solo dal punto di vista tecnologico

[2] Sono state istanziate, in particolare, due nuove Direzioni dedicate alla valorizzazione degli studi e delle ricerche in ambito economico e sociale: DVSE, Direzione centrale per gli studi e la valorizzazione tematica nell'area delle statistiche economiche; DVSS, Direzione centrale per gli studi e la valorizzazione tematica nell'area delle statistiche sociali e demografiche.

[3] La produzione dei registri statistici è in continuo aumento, nel rispetto della normativa GDPR, e i sistemi per l'acquisizione e il trattamento degli archivi amministrativi e per la produzione dei registri sono in costante evoluzione e perfezionamento.

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

ma anche dei processi di produzione statistica, garantendo una gestione organica ed unitaria delle diverse tematiche grazie al superamento dell'approccio a silos ed una maggiore efficienza, efficacia ed economicità dei processi di produzione.

Il tema è molto complesso e sfidante, sia dal punto di vista statistico-metodologico che dal punto di vista informatico e di governance. Gli aspetti da affrontare sono infatti molteplici:

- ✓ la governance degli accessi ai dati e, più in generale, l'aderenza alle indicazioni del GDPR e alle policy di sicurezza dell'Istituto
- ✓ la gestione della semantica dei dati attraverso l'introduzione di ontologie e opportuna metadatazione
- ✓ l'automatizzazione e il monitoraggio dei processi e la valutazione di aspetti di performance e di qualità.

Il punto di forza del SIR è nella capacità di integrare a livello concettuale, logico/fisco e statistico dati provenienti da diverse fonti anche afferenti a diverse aree statistiche (Figura 2 in Appendice); è quest'integrazione che consente di garantire un'offerta informativa più profonda, ampia e ricca.

2.2 Il Register-based Analytics Framework (RAF)

Valorizzare l'enorme mole di dati elaborati dall'Istituto è un passaggio fondamentale per la finalizzazione degli sforzi compiuti nella costruzione del Sistema Integrato dei Registri. Il programma denominato RAF (*Register-based Analytics Framework*) prosegue l'evoluzione dell'Istituto in ottica *data-driven* e si pone l'obiettivo di fornire – sia agli utenti meno esperti che a quelli con esigenze più complesse – una piattaforma per poter svolgere analisi statistiche in modo più o meno guidato. Esso prende le mosse, oltre che dalla costruzione del Sistema Integrato dei Registri, dai lavori svolti in ambito IT – in collaborazione con le altre Direzioni dell'Istituto – di *assessment* e disegno del framework futuro abilitante il consolidamento dei dati, della pre-diffusione e diffusione delle informazioni.

Alla base di tutta la progettualità sviluppata, c'è il rigoroso rispetto della normativa sulla privacy secondo l'approccio privacy *by design*.

Di rilevanza strategica per l'intero progetto è il porre l'utente al centro (*user centered approach*) di questo nuovo strumento, facendo propria la necessità di sviluppare una soluzione facile da usare, flessibile nell'accesso ai dati e intuitiva nella valutazione dei risultati richiesti.

Il RAF rappresenta di fatto una infrastruttura abilitante un'analisi visuale basata sulla multidimensionalità, multidisciplinarietà e interrelazione tra differenti indicatori e aree tematiche. La vastità e complessità dell'ecosistema informativo dell'Istituto ha portato a definire, nell'ambito delle finalità del programma RAF, molteplici *stream* progettuali: ogni *stream* ha il compito di affrontare la porzione specifica di sua competenza e collaborare e integrarsi con gli altri. Gli *stream* identificati coinvolgono

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

la maggior parte delle Direzioni dell'Istat e spaziano dalla definizione delle aree di ricerca tematiche, all'identificazione della metodologia e metadatazione, alla definizione delle classi di utenza, passando per la creazione di una nuova infrastruttura tecnologica che sia abilitante e di supporto all'obiettivo del progetto e che faccia leva su strumenti e architetture allo stato dell'arte.

Come detto, il sistema RAF si basa sul Sistema dei Registri – oltre che sui Censimenti Permanenti – e mette l'accento sulle modalità di rappresentazione e fruizione dei dati validati. Esso si pone quindi come risposta avanzata alla crescente domanda di informazione statistica per differenti tipologie di utenza, interna ed esterna all'Istituto, in modo da poter misurare la complessità e la multidimensionalità dei fenomeni economici e sociali tramite un paradigma che pone al centro la fruibilità, la semplicità e l'agilità nel trovare le informazioni di interesse. L'elevata granularità del nuovo sistema offre opportunità rilevanti che non devono essere confinate ai soli guadagni di efficienza ed efficacia (accuratezza delle stime) conseguiti nelle linee della produzione statistica ufficiale, ma rispondere a nuove strategie di produzione di informazione statistica.

La disponibilità di una elevatissima mole di dati caratterizzata da un dettaglio molto puntuale a livello settoriale, territoriale e dimensionale – nonché potenzialmente riclassificabile secondo nuovi schemi classificatori più analitici –, sta già ponendo problemi rispetto alla piena e tempestiva fruibilità dei dati in considerazione della sostanziale inadeguatezza degli attuali sistemi informativi per la fruizione dell'informazione e la tutela della riservatezza – progettati ormai da tempo per gestire moli di dati notevolmente inferiori, con tempi di rilascio notevolmente più lunghi e secondo un approccio di tipo statico e rigido (sulla base di esigenze informative pianificate a priori e non modificabili ex post).

IL RAF rappresenterà comunque nel futuro la modalità con cui l'Istat fornirà "contenuti statistici" alla propria utenza, creando visualizzazioni di facile interpretazione e fruizione (*dashboard*) e mettendo a disposizione "nodi" informativi evoluti che integrino e correlino i dati raccolti, organizzati, interpretati e validati dall'Istituto.

Un primo caso di applicazione del RAF riguarderà il mondo delle imprese: sulla base di alcune metriche e indicatori statistici caratterizzanti, verrà offerta la possibilità di effettuare analisi e correlazioni a differenti livelli di granularità del dato, sempre mantenendo il corretto livello di privacy e confidenzialità. Le informazioni statistiche che il RAF, in prima battuta, indirizzerà per il mondo delle imprese saranno:

- ✓ la "Produttività apparente del lavoro", ovvero il rapporto tra il valore aggiunto e il numero di addetti delle imprese;
- ✓ la "Propensione all'esportazione", ovvero la percentuale del fatturato esportato sul fatturato totale delle imprese;
- ✓ il "Costo del lavoro", ovvero la media delle retribuzioni lorde dei dipendenti delle imprese;

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

- ✓ il “Tasso di profittabilità”, ovvero il rapporto tra il margine operativo lordo e il valore aggiunto delle imprese.

Queste metriche, opportunamente raggruppate in schede tematiche, permetteranno di analizzare agilmente alcune sfaccettature del mondo delle imprese. Ogni scheda riporterà infatti grafici e indicazioni con informazioni del particolare ambito di interesse. Gli oggetti presenti in ogni scheda potranno essere “navigati” in maniera indipendente, andando ad effettuare selezioni (ad es. geografia, ATECO) e *drill-up/-down*, anche in modo integrato. Il *marking* che l’utente effettua su di un grafico genererà l’aggiornamento automatico degli altri grafici presenti sulla medesima scheda.

Successivamente, nell'ottica del *continuous improvement*, verrà rilasciata anche la possibilità di avere dipendenza tra grafici/indicatori multi-scheda e la messa a disposizione di nodi informativi in cui utenti autorizzati possono fare analisi in autonomia sui dati, costruire visualizzazioni da usare in proprio e in collaborazione con altri soggetti, opportunamente autorizzati.

2.3 Il sistema informativo EDAMIS e il protocollo SDMX

In qualità di partner dell’ESS (European Statistical System), insieme agli altri istituti nazionali di statistica e alle varie autorità nazionali dei Paesi membri responsabili dell’elaborazione, della produzione e della diffusione delle statistiche europee, l’Istat collabora con Eurostat, l'ufficio statistico dell'Unione europea, nell’adempimento della sua missione di principale fornitore di statistiche di alta qualità sull’Europa. Lo scambio di dati, asse portante della collaborazione, ruota attorno al concetto di *Single Entry Point* (SEP): i dati raccolti da tutte le organizzazioni dell’ESS giungono in un’area di ricezione comune presso Eurostat, dove possono essere automaticamente monitorati, controllati e consegnati alla destinazione attraverso un insieme di strumenti informatici comuni.

Per la trasmissione di tutte le serie di dati regolari a Eurostat dal 1º luglio 2008 è obbligatorio utilizzare EDAMIS (*Electronic Data files Administration and Management Information System*), il sistema informativo che realizza il SEP. L’uso del SEP implica che i dataset in arrivo siano identificati quali istanze di un dataset contenuto nell’inventario dei dataset da trasmettere dai Paesi membri a Eurostat. Si tratta di un ambiente *mission-critical* per lo scambio di dati nell’ESS che viene utilizzato quotidianamente da una grande comunità di utilizzatori.

Il Regolamento europeo EBS (*European Business Statistics*), al termine del periodo di transizione gennaio 2021-dicembre 2023, prevede l’adeguamento, anche per la parte che riguarda la trasmissione dei dataset dei relativi domini statistici, al formato standard di scambio dati e metadati SDMX (*Statistical Data & Metadata Exchange*).

Nei prossimi mesi, diversi domini aggiuntivi (*Population statistics, Asylum and managed migration, Education, Agricultural statistics, Euro-Group Registers*)

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

inizieranno a passare a trasmissioni conformi a SDMX, adottandone la stessa architettura per la validazione basata su questo standard.

Per quanto concerne gli aspetti architetturali e tecnologici legati alla generazione dei file SDMX-ML da parte delle strutture di produzione dell'Istat ai fini di reporting verso Eurostat, si distinguono normalmente le seguenti casistiche:

- ✓ Nel caso in cui i dati siano già memorizzati all'interno di basi dati consolidate con una propria struttura, al fine di evitare ridondanze, tali dati sono "mappati" direttamente da database sulle strutture dati SDMX fornite da Eurostat: questo è l'approccio seguito per il reporting dei dati di Contabilità Nazionale o quello di dati relativi a diverse indagini congiunturali.
- ✓ Nel caso in cui per alcune indagini non si disponga di un database vero e proprio (ad esempio i dati sono memorizzati su file testuali oppure su file excel) oppure si voglia implementare un database specificamente per il reporting già strutturato sulla base del modello informativo SDMX industrializzando e standardizzando così il processo, è disponibile il tool Meta&Data Manager sviluppato da Istat utilizzando le librerie standard di SDMX prodotte da Eurostat: tale soluzione, a meno di casistiche particolari, è la soluzione più avanzata proposta dalla Direzione Centrale per le tecnologie informatiche, in particolare per le indagini che stanno ora iniziando le attività di passaggio al formato SDMX proposto da Eurostat (ad esempio l'Indagine sui Trasporti Marittimi).

2.4 Il progetto Hub della Statistica Pubblica e la piattaforma SDMX Istat Toolkit

Nell'ambito del progetto "Hub della Statistica Pubblica", denominato anche "Sistan Hub", è stata realizzata una piattaforma di diffusione centralizzata e standardizzata dei dati degli enti Sistan territoriali e centrali, memorizzati in fonti dati distribuite e delocalizzate presso gli enti stessi o in cloud presso Istat (modalità SaaS).

L'obiettivo è quello di rendere tale piattaforma fortemente industrializzata, interoperabile e basata su standard statistici consolidati a livello internazionale (in particolare SDMX) per essere utilizzata sia nell'ambito del progetto Hub della Statistica Pubblica, sia in tutti i contesti di diffusione corporate di dati statistici in modalità interattiva e machine-to-machine con configurazioni architetturali scalabili.

La piattaforma denominata "SDMX Istat Toolkit" è una piattaforma open-source (sotto licenza EUPL) che permette di facilitare la standardizzazione dei seguenti processi in ambito statistico:

- ✓ diffusione (sia di tipo interattivo sia machine-to-machine) dei dati e metadati statistici
- ✓ reporting verso altre organizzazioni a livello nazionale ed internazionale
- ✓ Statistical Data warehousing

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

- ✓ raccolta di dati provenienti da altre organizzazioni secondo una struttura definita e standardizzata, provenienti da altre organizzazioni o da differenti strutture della medesima
- ✓ gestione dei metadati sia strutturali (descrittivi dei dati e della loro strutturazione) che referenziali (ovvero descrittivi del processo statistico che genera i dati, della qualità dell'indagine e dei dati medesimi e di altre informazioni necessarie o utili per l'utilizzo corretto dei dati)

Le applicazioni di cui è composta la piattaforma sono le seguenti:

Meta&Data manager

Applicazione web per la gestione e la pubblicazione di dati e metadati statistici sulla base dello standard internazionale SDMX (Figura 3 in Appendice), che nello specifico permette di:

- ✓ gestire e diffondere metadati strutturali (funzionalità di registry)
- ✓ creare database SDMX-compliant per la diffusione ed il reporting, nonché gestire e diffondere o rendere disponibili machine-to-machine i relativi dati
- ✓ creare e pubblicare metadati referenziali (SDMX metadataset)
- ✓ creare e pubblicare cataloghi di dataset open secondo lo standard DCAT-AP_IT (Data Catalogue-Application Profile per la pubblicazione e federazione di cataloghi open data, promossa da AgID) per la diffusione e la federazione di tali cataloghi
- ✓ creare, gestire e pubblicare glossari tematici.

Data Browser

Applicazione web che interagisce con i servizi web SDMX consentendo agli utenti di navigare, ricercare, interrogare, visualizzare set di dati ed effettuarne il download in vari formati (xml, csv, json) mantenendo comunque una struttura compatibile con il modello informativo SDMX. La rappresentazione dei dati relativi ad un dataset può avvenire sotto forma di tabella multidimensionale, grafico o mappa tematica.

Tale applicazione può essere utilizzata all'interno di una singola organizzazione al fine di diffondere dataset archiviati in uno o più database, oppure nell'ambito di un progetto "multi-source" (architettura Hub), dove più organizzazioni espongono i propri database tramite SDMX Web Services.

Le principali funzionalità rese disponibili da tale applicazione sono le seguenti:

- ✓ muoversi tra diversi database distribuiti
- ✓ sfogliare uno o più alberi dei temi e selezionare il dataset di interesse
- ✓ effettuare, per ogni sorgente dati, una ricerca testuale dei dataset
- ✓ impostare i filtri per interrogare i dataset
- ✓ creare, per ogni dataset, grafici e visualizzare mappe tematiche
- ✓ configurare il layout della tabella e del grafico

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

- ✓ calcolare le variazioni congiunturali e tendenziali
- ✓ consultare dashboard che permettono di visualizzare in tempo reale dati provenienti da sorgenti differenti
- ✓ memorizzare query utilizzabili in altre sessioni di lavoro
- ✓ configurare, come amministratore, le connessioni alle sorgenti, gestire le utenze e creare dashboard e visualizzazioni.

2.5 L'approccio strutturato alla standardizzazione delle piattaforme: il Generic Statistical Business Process Model

Con il proliferare di tecniche nuove e strumenti sempre più diversificati e moderni, idonei ad essere utilizzati in organizzazioni di produzione statistica e ricerca come l'Istat, il disegno e l'implementazione di una indagine statistica sono stati nel corso del tempo a rischio di forte personalizzazione.[4] Col tempo si è sentita quindi la necessità di provare a standardizzare le terminologie, gli strumenti e i processi attraverso un glossario comune e un catalogo di servizi statistici/metodologici ufficiali.

Da queste considerazioni sono nati standard che aiutano il lavoro dello statistico e del personale che realizza gli strumenti informatici a supporto delle indagini. Il principale standard di riferimento per la modellazione statistica dei processi di produzione del dato è il *Generic Statistical Business Process Model* (GSBPM), che descrive e definisce l'insieme dei processi di business necessari a produrre statistiche ufficiali. Il GSBPM è nato grazie al lavoro di esperti internazionali, riuniti sotto l'egida della Commissione economica per l'Europa (UNECE)[5], motivati a catalogare gli aspetti di processo comuni per l'implementazione di un processo di business statistico.

Il GSBPM fornisce un riferimento unico nel suo genere, offrendo una terminologia armonizzata che aiuta le organizzazioni che si occupano di statistica, nazionali o internazionali, a condividere metodi e strumenti, al loro interno e anche tra diverse indagini. Questo modello identifica i possibili passaggi nel processo statistico e le interdipendenze tra di essi. Sebbene l'implementazione effettiva del GSBPM segua spesso la stessa sequenza di passaggi nella maggior parte dei processi statistici, alcuni elementi del modello possono collocarsi in ordine diverso, a seconda delle esigenze specifiche.

In Istat si è cercato di seguire un approccio integrato, metodologico e tecnologico, per standardizzare e strutturare le indagini statistiche attraverso questo modello. Il raggiungimento di questi obiettivi è comunque condizionato ad una profonda

[4] Spesso, ad esempio, alcuni processi sono legati a specifici ricercatori, alla loro conoscenza, presenza e competenze; la forte connessione tra persone e processi è una barriera alla condivisione e standardizzazione.

[5] https://unece.org/.

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

revisione del processo di produzione statistica, in ottica di armonizzazione, che sia dunque fondato su una serie di piattaforme più generalizzate possibile, che coprano una o più fasi del processo statistico.

2.6 La qualifica a Polo Strategico Nazionale: il percorso verso un Software Design Data Center e il private cloud

In linea con i principi e le linee guida emanate da AgID, anche l'Istat ha avviato un percorso verso un "Software Design Data Center" (SDDC) e la costituzione di un "private Cloud" per facilitare un percorso verso un modello di "Cloud della Statistica".

In quest'ottica le evoluzioni infrastrutturali sono state realizzate per fornire servizi di housing, hosting, e cloud verso altre PA che necessitano di un soggetto affidabile e garantito dallo Stato per la gestione e il trattamento di dati strategici.

Ad oggi, i servizi informatici dell'Istat sono erogati da due data center (DC) ubicati, il primo, nella sede di Roma di via Cesare Balbo 16 (sito primario) e, il secondo, in housing nel Data Center Inail di Roma[6] (sito secondario). Nel loro insieme, Rete e Data Center, sono le fondamenta su cui si basano la fornitura dei servizi IT e le politiche di Business Continuity (BC)/Disaster Recovery (DR). La connettività dei data center con Internet è realizzata tramite i servizi di connettività del consorzio GARR (Gestione Ampliamento Rete Ricerca) e tutte le sedi accedono a Internet tramite i servizi erogati dai due data center. I collegamenti tra le sedi Istat avvengono tramite il Sistema Pubblico di Connettività (SPC). SPC fornisce inoltre un collegamento con la rete Infranet delle Pubbliche Amministrazioni presso il data center di via Cesare Balbo. Le reti sono realizzate tramite reti virtuali VLAN, realizzando perciò un'astrazione rispetto alle effettive reti fisiche e semplificandone quindi la gestione. La sicurezza dell'accesso alla rete è garantita anche da un sistema di Network Access Control (NAC). Grazie a tale servizio è possibile avere una visione in tempo reale di ciò che è effettivamente connesso alla rete dell'Istituto e di gestire gli accessi degli utenti. Questo costituisce un primo importante livello di sicurezza che viene ulteriormente rafforzato dalle policy definite sugli apparati firewall.

I sistemi di storage presenti nelle due sedi sono in configurazione di completa ridondanza, sono connessi alle Storage Area Network (SAN) di sede e sono dotati di meccanismi di sincronizzazione remota per il supporto del piano di Disaster Recovery.

L'infrastruttura di virtualizzazione di Istat si articola in tre ambienti, distinti per tecnologia di virtualizzazione e architettura:

- ✓ Primo ambiente: tecnologia Microsoft Hyper-V su architettura classica a due livelli Storage/Computing

[6] Via del santuario Regina degli apostoli, 33.

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

- ✓ Secondo ambiente: tecnologia VMware su architettura classica a due livelli Storage/Computing
- ✓ Terzo ambiente: tecnologia VMware su architettura iperconvergente, in cui le risorse Storage/Computing risultano aggregate in un'unica infrastruttura hardware integrata.

Tali infrastrutture consentono di fruire dei benefici tipici dell'utilizzo della virtualizzazione:

- ✓ erogazione di servizi in maniera più agile con una distribuzione ottimale del carico sulle risorse fisiche, e con alta scalabilità
- ✓ incremento della produttività, riduzione dei costi e flessibilità nella gestione IT.

Queste evoluzioni infrastrutturali consentono alla Direzione Centrale per le Tecnologie Informatiche di accompagnare la *Digital Transformation* dell'Istituto e di continuare nel percorso di *Cloud Enablement*, integrando le soluzioni di virtualizzazione con una piattaforma di gestione definita secondo i nuovi modelli di erogazione as-a-Service (*Cloud Management Platform*).

L'integrazione di una *Cloud Management Platform* abilita l'esposizione di un Catalogo Servizi, l'automazione dei processi di *fulfillment* delle richieste e una gestione "elastica" dell'utilizzo delle risorse, consentendo un costante allineamento alle esigenze di business e introducendo la possibilità di effettuare procedure di Disaster Recovery verso cloud esterni.

Il processo di *cloud transformation* prevede la fornitura di tutti i tre principali modelli di servizio (IaaS, PaaS, SaaS) e Istat, in qualità di Polo Strategico Nazionale, intende erogare questi servizi attraverso la propria *Cloud Management Platform*, in modo da poterli offrire sia all'utenza interna che a terze parti con cui l'Istituto collabora per svolgere le proprie funzioni istituzionali.

3. L'esperienza maturata dall'Istat nell'interoperabilità delle banche dati fiscali per la produzione statistica

Sul tema dell'interoperabilità delle banche dati fiscali, con particolare riguardo all'utilizzo integrato delle informazioni presenti nell'anagrafe tributaria – temi di specifica competenza della Commissione –, l'Istat ha maturato nel tempo un elevato know-how tecnico-organizzativo nella progettazione, integrazione e utilizzo di questi archivi di fonte fiscale per la produzione di statistiche ufficiali.

Ad oggi, l'Istat utilizza oltre 30 archivi di dati fiscali provenienti dall'Agenzia delle Entrate, acquisiti tramite protocolli sicuri di trasmissione sulla base di una pianificazione annuale, che trova nel Piano Statistico Nazionale (PSN) e nei regolamenti statistici comunitari la principale fonte di liceità per il loro utilizzo a fini di produzione statistica ufficiale. Questi archivi vengono impiegati in modo intensivo ed integrato, e in particolare forniscono un input per la produzione statistica di ben

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

222 progetti statistici definiti nel PSN a titolarità Istat. Con riguardo all'Anagrafe tributaria delle persone fisiche e giuridiche con partita Iva, questa specifica banca dati fiscale viene utilizzata da ben 10 progetti Istat inclusi nel PSN, che riguardano sia i registri statistici di base che le rilevazioni dirette. L'utilizzo ampio ed integrato delle banche dati fiscali nell'ambito della produzione statistica è frutto di un percorso che ha portato l'Istat a sperimentare, progettare e poi consolidare opportune soluzioni tecnologiche, metodologiche ed organizzative per l'utilizzo di questi dati in modo intensivo nell'ambito dei processi di produzione statistica.

Nel seguito del paragrafo si riportano sinteticamente le tappe principali di questo percorso.

A partire dagli anni '90 ed in anticipo rispetto ad altri paesi avanzati, l'Istat ha iniziato ad utilizzare le banche dati fiscali per la produzione e l'aggiornamento continuo dei registri statistici sulle imprese, riducendo così l'onere informativo sui rispondenti delle rilevazioni dirette e consentendo anche l'aggiornamento annuale del registro statistico delle imprese attive e residenti in Italia (ASIA).

In una seconda fase, anche grazie all'elevato grado di standardizzazione ed integrabilità delle fonti fiscali con altre informazioni di carattere statistico, l'utilizzo della fonte fiscale è stato esteso alla produzione di statistiche strutturali sui conti economici delle imprese nonché alla produzione di ulteriori registri di base o estesi. Più recentemente, è stata avviata un'attività di progettazione per l'utilizzo di dati fiscali provenienti da recenti innovazioni normative (per esempio i dati fiscali relativi alla fatturazione elettronica) nell'ambito dei conti economici nazionali e delle rilevazioni congiunturali sul fatturato dell'industria e dei servizi.

La presenza nelle banche dati fiscali di identificativi diretti, nonché di variabili caratterizzate da un elevato grado di standardizzazione e, nel complesso, un buon livello di qualità delle informazioni disponibili ha consentito un notevole grado di integrazione, a livello micro, di questi dati nell'ambito del Sistema integrato dei registri dell'Istituto. Le operazioni di trattamento statistico di questi dati al fine di verificarne la qualità per la produzione statistica hanno riguardato sia la classificazione statistica degli operatori economici sia l'analisi della coerenza e completezza delle variabili fiscali a fini statistici, secondo metodologie e procedure sperimentate e consolidate nel tempo da Istat, a partire dalle linee guida sulla qualità dei dati definite a livello nazionale ed internazionale.

Un altro elemento importante progettato e implementato al fine di rispettare la normativa sulla tutela dei dati personali e le disposizioni fornite dall'Autorità Garante per la Privacy ha riguardato l'adozione di tecniche di pseudonimizzazione e la progressiva implementazione dei domini specifici di integrazione ai dati personali inclusi negli archivi fiscali già nella fase di acquisizione di questi dati da parte dell'Istat.

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

4. L'Istat e l'interoperabilità nella PA: prospettive di sviluppo

Come descritto nei paragrafi precedenti, l'Istat ha intrapreso internamente, ed in rapporto al proprio ruolo istituzionale, molte iniziative che stanno rendendo l'Istituto sempre più in linea con gli standard internazionali e con le direttrici del Piano Triennale dell'Informatica e alle direttive e linee guida di AgID.

L'obiettivo dell'Istituto è porsi come punto di riferimento sia per gli aspetti riguardanti la gestione del dato (Figure 4 e 5 in Appendice), la governance, le metodologie di integrazione e validazione, sia per gli aspetti più organizzativi e culturali in ambito di un approccio *data-driven*. Le esperienze dell'Istat possono essere infatti messe a disposizione della Pubblica Amministrazione italiana e in tale ambito l'Istituto sta partecipando al tavolo trasversale sulla Data Strategy Nazionale. L'Istat è del resto un soggetto unico nel panorama nazionale, con le competenze tecniche e le esperienze necessarie per facilitare la trasformazione digitale nei seguenti ambiti:

- ✓ Supporto nella definizione del modello e del framework centralizzato di riferimento in ambito interoperabilità della PA:
    - Acquisizione, elaborazione e diffusione dei dati
    - Creazione di vocabolari ed ontologie (metadati)
    - Partecipazione a progetti di interoperabilità e di scambio dati a livello europeo e internazionale
    - Modelli informativi e standard per la gestione (incluso la condivisione) dei dati e metadati
    - Progettazione e sviluppo di soluzioni IT per l'interoperabilità e lo scambio dei dati
- ✓ Governance di progetti complessi tra Amministrazioni in ambito Sistan
- ✓ Supporto nella definizione del servizio di Governance centralizzato:
    - Modello organizzativo e definizione dei ruoli e responsabilità
    - Integrazione, Qualità e Validazione Semantica (ontologie e metadati)
    - Formazione e incremento della Cultura del Dato
- ✓ Supporto nella definizione della Diffusione, Condivisione e Analisi dei dati.

L'Istat è, del resto, già delegata a:

- ✓ definire nomenclature e metodologie di base per la classificazione e la rilevazione dei fenomeni (art. 15, d. lgs. n. 322/89)
- ✓ definire metodi e formati per le PP.AA. per lo scambio e l'utilizzo in via telematica dell'informazione statistica e finanziaria, nonché coordinamento modificazioni, integrazioni e nuove impostazioni della modulistica e dei sistemi informativi della P.A. (d.P.R. n. 166/2010).

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

Appendice

Figura 1. Infrastruttura di riferimento per la raccolta, analisi, validazione e diffusione dei dati

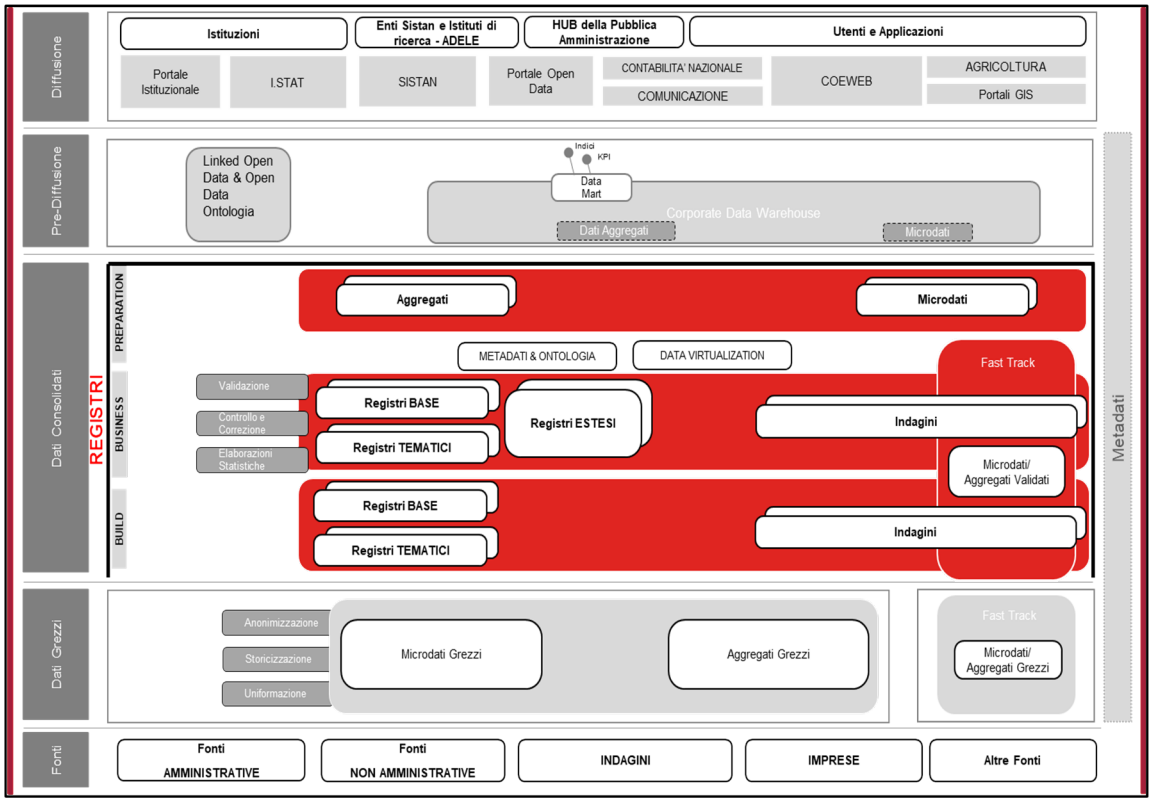
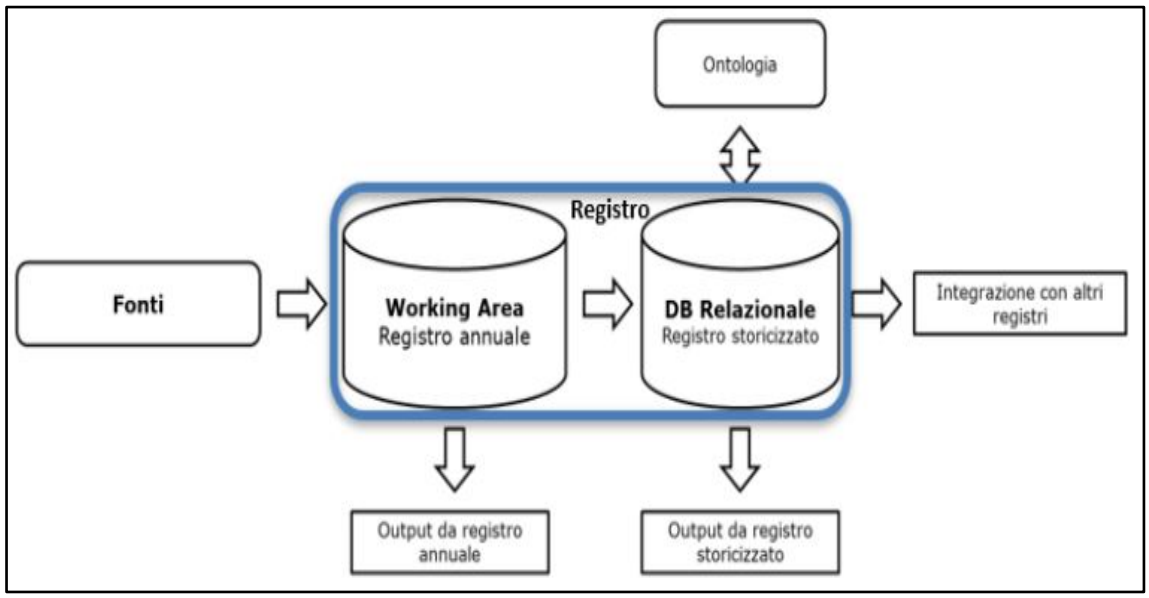


Figura 2. Architettura logica del Sistema Integrato dei Registri (SIR)



Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB

Figura 3. Schematizzazione Meta e Data Manager

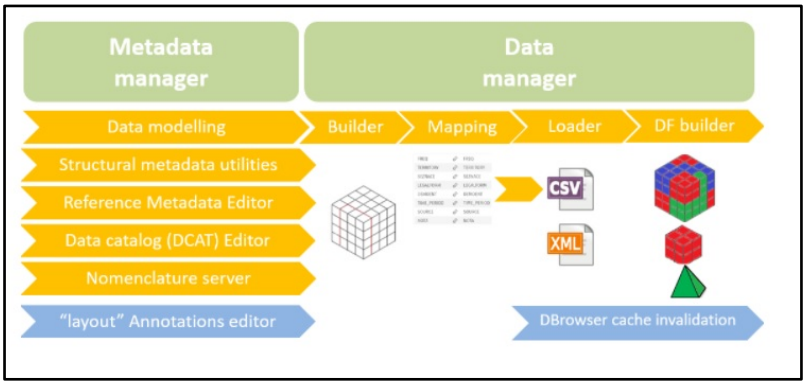


Figura 4. Proposta di architettura logica di alto livello per l'interoperabilità e governo della Data Strategy Nazionale

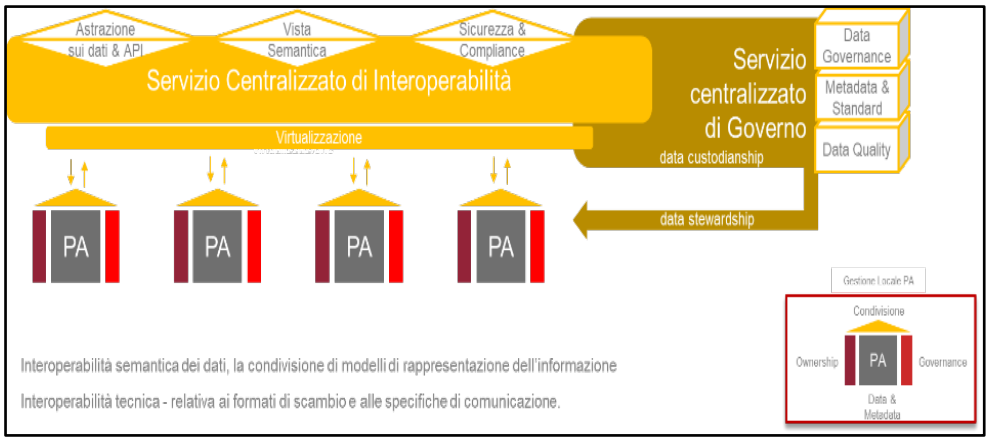


Figura 5. Proposta di identificazione dei ruoli in ambito di governance della Data Strategy Nazionale

Com. anagrafe tributaria ARRIVO 19 ottobre 2021 Prot: 2021/0000082/ANATRIB