



9 Data & Analytics Framework

Il *Data & Analytics Framework* (DAF) fa parte delle attività atte a valorizzare il patrimonio informativo pubblico nazionale. Il DAF ha l'obiettivo di sviluppare e semplificare l'interoperabilità dei dati pubblici tra PA, standardizzare e promuovere la diffusione degli *open data*, ottimizzare i processi di analisi dati e generazione di conoscenza. L'idea è quella di aprire il mondo della Pubblica amministrazione ai benefici offerti dalle moderne piattaforme per la gestione e l'analisi dei *big data*, agendo lungo quattro direttrici principali:

- amplificare sensibilmente il valore del patrimonio informativo della PA mediante l'utilizzo delle tecnologie big data che consentono di creare conoscenza per i *decision maker* e ridurre drasticamente i tempi di analisi. La scalabilità orizzontale di queste tecnologie permette, infatti, di estrarre informazioni dall'incrocio di molteplici basi di dati e di processare dati *real-time* consentendo di avere più prospettive di analisi su un dato fenomeno, in maniera tempestiva;
- favorire e ottimizzare lo scambio dei dati tra PA, minimizzandone i costi transattivi per l'accesso e l'utilizzo. Sarà infatti possibile superare lo schema delle convenzioni uno a uno, che portano a molteplici copie degli stessi dati e consentire un accesso standardizzato ad un dato sempre aggiornato;
- incentivare la diffusione degli *open data* e renderne più efficace l'utilizzo. Il DAF, infatti, consente di centralizzare e redistribuire i dati pubblici attraverso API, garantendo standardizzazione di formati e modalità di riutilizzo su dati sempre aggiornati;
- favorire l'analisi esplorativa dei dati da parte di team di *data scientist*, sia in seno alle singole PA che a livello centrale, al fine di migliorare la conoscenza dei fenomeni sociali. Le tecniche di analisi utilizzate consentiranno anche lo sviluppo di applicazioni "intelligenti" che sfruttino le regolarità nei dati per offrire servizi a cittadini, imprese e Pubbliche amministrazioni;
- infine, il *framework* consentirà la promozione di iniziative di ricerca scientifica su tematiche di interesse specifico per la PA, favorendo la collaborazione con università ed enti di ricerca.

Il DAF sarà strutturato in conformità con quanto definito nel CAD e nel Modello di interoperabilità, nelle Infrastrutture immateriali e in coerenza con le esigenze relative alle funzioni di monitoraggio del Piano.

Il DAF si basa su una **Piattaforma big data**, composta da: un *data lake*, un insieme di *data engine* e strumenti per la comunicazione dei dati.

Nel *data lake* vengono memorizzati, nel rispetto delle normative in materia di protezione dei dati personali, dati di potenziale interesse quali, ad esempio: (i) le basi di dati che le PA generano per svolgere il proprio mandato istituzionale; (ii) i dati generati dai sistemi informatici delle Pubbliche amministrazioni come log e dati di utilizzo che non rientrano nella definizione precedente; (iii) i dati autorizzati provenienti dal web e dai social network di potenziale interesse della Pubblica amministrazione.

I *Big Data Engines* sono utili ad armonizzare ed elaborare, sia in modalità *batch* che *real-time*, i dati grezzi memorizzati nel *data lake*, e a implementare modelli di *machine learning*.

Infine gli strumenti per la comunicazione dei dati sono utili a favorire la fruizione dei dati elaborati da parte dei soggetti interessati, anche attraverso API che espongono dati e funzionalità ad applicazioni terze.



La realizzazione e la successiva gestione del DAF è affidata al **BDT-PA**, ovvero *Big Data Team della PA*, un team composto da *data scientist*, *big data architect* ed esperti di dominio che provvedono al disegno e all'evoluzione concettuale della Piattaforma *big data*, alla costruzione di modelli di interconnessione delle diverse sorgenti dati, all'analisi dei dati, allo sviluppo di modelli di *machine learning*, al coordinamento dello sviluppo di *data application* e all'organizzazione di "competizioni" scientifiche su tematiche di interesse per la PA .

9.1 Scenario attuale

Ad oggi non esiste nella Pubblica amministrazione un *framework* di analisi, standardizzazione e interscambio dei dati pubblici che favorisca, al contempo, la definizione e il monitoraggio di politiche *data-driven*. Già a partire dal 2013, AgID ha verificato la possibilità di utilizzare tale tipologia di strumenti nello specifico dominio della Pubblica amministrazione per mezzo di numerose iniziative sperimentali condotte in collaborazione con enti di ricerca nazionali e con varie università italiane, nell'ambito del progetto *Italia.gov.it*, il motore dell'amministrazione digitale.

Negli ultimi anni, le tecnologie big data sono maturate a tal punto da essere utilizzate non solo negli ambienti di produzione delle maggiori aziende IT (ad es. Google, Facebook, Twitter, Linked-In), ma anche in quelli di banche, società di assicurazione, operatori di lotterie e scommesse, società di trading. Conseguentemente sono emersi nuovi profili professionali, come il *data scientist* e il *big data architect*, le cui competenze sono da considerarsi necessarie per il governo e l'uso delle piattaforme di *big data*.

Per quanto riguarda lo scambio di dati tra le PA, lo scenario attuale vede ancora diffusa la pratica di stipule di convenzioni o di accordi diretti tra PA per regolare lo scambio di dati necessari per lo svolgimento di attività istituzionali. Tale pratica non risulta scalabile e pone limiti alla condivisione dell'informazione del settore pubblico.

9.2 Obiettivi strategici

- Valorizzare il patrimonio di informazioni della Pubblica amministrazione facilitando l'accesso ai dati da parte delle PA e favorendo la costituzione di team agili di analisi sia centrali che federati.
- Puntare su qualità e standardizzazione dei dati. Il DAF, infatti, è lo strumento operativo che consente di implementare in maniera coordinata gli sforzi descritti nel paragrafo 4.1 "Dati della PA", ponendo attenzione sui processi di generazione, gestione, aggiornamento e diffusione dei dati.
- Agevolare lo sviluppo e la diffusione degli open data e della *API economy*, attraverso i quali la società civile può riutilizzare, nel rispetto della legge, il patrimonio di informazioni pubbliche e creare nuove opportunità di business. A tal fine, il DAF permetterà di sviluppare API standardizzate su basi di dati sempre aggiornate, per favorire la costruzione di applicazioni e servizi al cittadino.
- Incentivare le collaborazioni con università ed enti di ricerca. Si darà loro accesso a *sandbox* contenenti campioni significativi dei dati opportunamente anonimizzati, per dare stimolo alla ricerca e creare conoscenza utile per la collettività.
- Favorire lo scambio di dati tra Pubbliche amministrazioni superando le limitazioni dell'attuale pratica di modalità di accesso ai dati basata su convenzioni tra singole amministrazioni.
- Razionalizzare le risorse impegnate nello scambio di dati e in iniziative di *analytics*, inclusi *data*



warehouse e *business intelligence*. Tali iniziative, spesso non coordinate, sono di frequente caratterizzate da alti costi di licenze e hardware dedicato e hanno un'elevata tendenza a rispondere molteplici volte allo stesso bisogno.

- Offrire strumenti che misurino in maniera puntuale e tempestiva gli stati di avanzamento dell'attuazione del Piano e che permettano di individuare azioni integrative o correttive sulla base di logiche *data-driven*.

9.3 Linee di azione

Il DAF, come detto, si basa sullo sviluppo di una Piattaforma *big data* e sulla costituzione di un team di *data scientist*, *big data architect* e *data engineer*. Il *Big Data Team* della PA, istituito all'interno del Team digitale, ha il compito di gestire attivamente la fase di sviluppo concettuale e implementativo dell'infrastruttura, insieme a tutte le fasi del ciclo di vita del dato, dall'ingestione all'analisi e sviluppo di applicazioni. Inoltre, BDT-PA svilupperà partnership tecnologiche e di progetto tra le PA coinvolte.

Il BDT-PA progetta e definisce le modalità realizzative e di uso della Piattaforma *big data* della PA attraverso:

- l'individuazione del modello di governance che preveda un ruolo di guida e controllo da parte del Team digitale, in collaborazione con AgID e sentito il Garante della privacy;
- la pianificazione degli eventuali adeguamenti normativi che favoriscano la realizzazione del progetto;
- la definizione delle sorgenti di dati del *data lake* e le relative modalità di popolamento. Queste saranno inserite nelle linee guida prodotte nell'ambito del Modello di interoperabilità;
- la definizione dell'architettura logica della piattaforma e l'individuazione delle tecnologie implementative;
- l'identificazione delle esigenze informative utili alla definizione di *data driven policy* e alla realizzazione dei relativi strumenti analitici;
- l'uso di *cloud* pubblico e privato per *storage* e *computing*;
- il coinvolgimento della comunità scientifica per la promozione di iniziative atte a favorire la conduzione di attività ricerca su tematiche di interesse per la PA;
- le direttive di utilizzo e consultazione.

Nei prossimi mesi, il BDT-PA rilascerà il Piano di sviluppo del DAF. Esso prevedrà un *roll-out* incrementale in base agli accordi con le PA che il Team digitale sta implementando. In particolare sarà prevista una fase di sperimentazione che interesserà un gruppo di PA centrali e locali selezionate, e avrà l'obiettivo di realizzare modelli di scambio dati e casi di utilizzo con servizi per le PA, cittadini e imprese.

I dati presenti nel DAF saranno utilizzati anche per sintetizzare conoscenza utile alle attività di monitoraggio descritte nel capitolo 10 "Gestione del Cambiamento": a tal riguardo AgID e il Team digitale predisporranno strumenti che andranno a complementare il kit di strumenti descritti nella linea di azione "Strumenti per il monitoraggio dell'attuazione del Piano" del citato capitolo 10.



Oggetto	Definizione e realizzazione del piano di sviluppo della fase di sperimentazione del <i>Data & Analytics Framework</i>
Tempi	Entro dicembre 2017
Attori	AgID, Team digitale
Descrizione	Individuazione del modello di governance del DAF e delle PA che faranno parte della fase di sperimentazione. Definizione dell'architettura della piattaforma e della sua roadmap di evoluzione. Definizione dei casi d'uso per lo sviluppo di servizi per Pubbliche amministrazioni, cittadini e imprese. Questa attività è coordinata con il Garante della privacy. Realizzazione dell'infrastruttura tecnologica, coerentemente con il piano di sviluppo della fase di sperimentazione del DAF, che implementa tutte le componenti necessarie per il funzionamento della Piattaforma.
Risultato	Piano di sviluppo del DAF (<i>data di rilascio: giugno 2017</i>). Cluster big data e Test delle componenti e dei casi d'uso (<i>data di rilascio: dicembre 2017</i>).

Oggetto	Ingestione dei dati nel DAF - Fase sperimentale
Tempi	Da giugno 2017 a dicembre 2017
Attori	PA, AgID, Team digitale
Descrizione	Definizione delle basi di dati da inserire nel progetto in fase sperimentale e messa in esercizio delle procedure di estrazione e ingestione. Definizione dei rapporti tra il gestore del DAF e le PA aderenti all'iniziativa.
Risultato	Regolazione dei rapporti con le PA. Definizione standard operativi nel rispetto delle norme sulla privacy. Definizione delle procedure di ingestione dei dati nella piattaforma. Alimentazione del DAF (<i>data di rilascio: dicembre 2017</i>).

Oggetto	Messa in Produzione del DAF
Tempi	Da gennaio 2018
Attori	Team digitale, AgID, PA
Descrizione	Il Team digitale e AgID predisporranno le procedure atte al subentro del futuro <i>owner</i> del DAF che gestirà l'operatività e l'evoluzione del progetto. L' <i>owner</i> del DAF curerà le interazioni con le PA per definire piani di inclusione delle relative basi di dati e casi d'uso. Le PA di volta in volta coinvolte definiranno le modalità di ingestione dei dati e utilizzo del DAF



	consone alle proprie attività.
Risultato	Piano di subentro. Operatività corrente (data di rilascio: da definire).

Oggetto	Realizzazione strumenti di supporto al monitoraggio del Piano
Tempi	Da aprile 2017
Attori	AgID e Team digitale
Descrizione	Il Team digitale e AgID realizzano strumenti che, sulla base delle informazioni contenute nel DAF, forniscono informazioni utili alle attività di monitoraggio del Piano descritte nel capitolo 10 "Gestione del Cambiamento".
Risultato	Strumenti di supporto al monitoraggio del Piano <i>(da gennaio 2018)</i>