
Python

07 nov 2018

1	Scopo del documento	1
2	Indice dei contenuti	3
2.1	Cos'è il DAF	3
2.2	A chi si rivolge	4
2.3	Benefici e Funzionalità per la PA	4
2.4	Roadmap di evoluzione	5
2.5	Modello di governance	6
2.6	Architettura di alto livello	6
2.7	Casi d'uso della fase sperimentale	8

CAPITOLO 1

Scopo del documento

Questo documento rappresenta il risultato dell'azione "Definizione e realizzazione del piano di sviluppo della fase di sperimentazione del Data & Analytics Framework", previsto nel [capitolo 9](#) del [Piano Triennale per l'Informatica nella PA \(2017-2019\)](#). Nel documento sono contenute informazioni relative al piano di sviluppo del DAF nella sua fase di sperimentazione (entro dicembre 2017).

Per richieste di chiarimento e domande di approfondimento si rimanda alla [sezione dedicata al DAF](#) presente su [Forum Italia](#).

2.1 Cos'è il DAF

Il **Data & Analytics Framework** (DAF) è una delle attività atte a valorizzare il patrimonio informativo pubblico nazionale approvata dal Governo italiano nell'ambito del **Piano Triennale per l'Informatica nella PA 2017-2019**. L'obiettivo principale del DAF è di abbattere le barriere esistenti nell'interscambio dei dati pubblici tra PA e promuoverne l'utilizzo a supporto del decision making pubblico, ottimizzare i processi di analisi dati e generazione di sapere, standardizzare e promuovere la diffusione degli open data, promuovere e supportare iniziative di ricerca scientifica favorendo la collaborazione con Università ed enti di ricerca.

Il DAF si compone di:

- Una **Piattaforma Big Data** costituita 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: le basi di dati che le PA generano per svolgere il proprio mandato istituzionale; i dati generati dai sistemi informatici delle Pubbliche Amministrazioni come log e dati di utilizzo che non rientrano nella definizione precedente; i dati autorizzati provenienti dal web e dai social network di potenziale interesse della Pubblica Amministrazione;
 - *big data engine* utile 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;
 - *strumenti per l'interscambio dei dati*, utili a favorire la fruizione dei dati elaborati da parte dei soggetti interessati, anche attraverso API che espongono dati e funzionalità ad applicazioni terze;
 - *strumenti di analisi e visualizzazione dei dati* offerti in modalità self-service agli utilizzatori del DAF.
- Un **Dataportal**, che rappresenta l'interfaccia utente per l'utilizzo delle funzionalità implementate nel DAF. In particolare, il dataportal si compone di:
 - un catalogo dei dataset basato su CKAN, che gestisce i metadati relativi sia ai dati contenuti nel DAF che agli open data harvestati dai siti delle PA;
 - interfacce utente per accedere ai tool di analisi e data visualization menzionati sopra;

- un modulo riservato alle PA per gestire il processo di *ingestion* e gestione dei dati e metadati nel DAF;
- un modulo per *data stories*, attraverso il quale gli utenti possono pubblicare le proprie analisi e collaborare con altri utenti.
- Da un **team di esperti di dati**, composto da data scientist, data engineer e big data architect 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'incentivazione della ricerca scientifica su tematiche di interesse per la PA.

Si rimanda al [capitolo 9 del Piano Triennale](#) per maggiori informazioni.

2.2 A chi si rivolge

I principali utenti del DAF sono le **Pubbliche Amministrazioni**, che a seconda dei casi possono utilizzare il DAF per:

- accedere in lettura ai dati presenti nel DAF ed effettuare analisi a supporto del *decision making* (eccezion fatta per i dati sensibili il cui accesso deve essere autorizzato dagli owner del dato);
- sviluppare nuovi servizi non critici rivolti ai cittadini e alle imprese utilizzando i dati e le funzionalità esposte dal DAF;
- sviluppare data application utili per migliorare i servizi erogati e i propri processi interni;
- utilizzare i servizi offerti dal DAF per automatizzare il processo di generazione e pubblicazione di Open Data, in modalità SaaS.

Alle PA si aggiungono le seguenti categorie di utenti:

- **data journalist**, che sono alla ricerca di informazioni analitiche per supportare le proprie tesi giornalistiche;
- **cittadini** che ricercano informazioni desumibili dagli strumenti di visualizzazione e dalle data stories pubblicate dal team di datascientist e dalla *community* sulla base dei dati e strumenti di analisi collegati al DAF;
- **community** di civic hacker, sviluppatori e aziende che fanno uso degli opendata per creare applicazioni e servizi a valore aggiunto;
- il **mondo della ricerca** e dell'innovazione. Grazie al DAF sarà, infatti, possibile promuovere iniziative atte a coinvolgere il mondo della ricerca e dell'innovazione su tematiche di interesse pubblico.

Il DAF si poggia su un sistema di sicurezza, gestione di accessi e data separation che permette di gestire i permessi di accesso ai dati. In questo modo, tutte le funzionalità del DAF saranno esposte a tutte le tipologie di utenti sopra indicate, che potranno accedere ai dati per cui il relativo profilo utente ha diritto di accesso.

2.3 Benefici e Funzionalità per la PA

Il DAF fornisce un sistema di *big data management* e strumenti di analisi e utilizzo dei dati in modalità SaaS/PaaS, con l'obiettivo di sgravare il più possibile le PA da attività di gestione operativa e tecnica. Inoltre, esse avranno accesso al team di data scientist e data engineer che supporterà le PA nell'utilizzo delle funzionalità del DAF, e potrà prendere in carico richieste di analisi e sviluppo specifiche, da valutare caso per caso.

Il DAF mette a disposizione di ciascuna PA aderente:

- un insieme di **tool per l'analisi e l'elaborazione dei dati** presenti nel DAF che sfruttano le capacità di gestione ed elaborazione di big data, tra cui:
- Un tool per generare dashboard e report.

- Un notebook per effettuare analisi sui dati presenti nel DAF a cui l'utente ha accesso.
- L'accesso a un insieme di **dati utili per l'elaborazione di analytics**. Di seguito, un elenco non esaustivo dei dati che a tendere saranno presenti nel DAF:
 - **Dati delle basi di dati d'interesse nazionale**: basi di dati autoritative rispetto alle "entità" che gestiscono (ad es. ANPR è autoritativa per l'"entità" residente). Nel DAF è possibile trovare una copia sempre aggiornata dei dati in esse contenute, fatte salve le eccezioni di norme e regolamenti, e in accordo con il Garante per la Privacy.
 - **Dati delle PA**: le PA sincronizzeranno (cfr. Piano triennale) una copia dei dati utili a svolgere il proprio mandato istituzionale nonché i dati generati dai propri sistemi informatici (es. log). Tali dati sono accessibili da parte di tutte le PA, ad eccezione di quei dati sui cui vigono norme in materia di protezione dei dati personali.
 - **Open data standard**: il DAF promuove la creazione di standard per la diffusione di open-data su temi di diffuso interesse pubblico (es. Mobilità, trasporti, turismo, eventi, ecc.). Grazie a tali standard un dataset può essere popolato in modo collaborativo da più PA.
 - **Dati di interesse pubblico di terze parti**: il DAF raccoglie e mette a disposizione di tutte le PA dati di terze parti di potenziale interesse pubblico (es. dati provenienti dai social networks, dati forniti da aziende, ecc.).
- **un servizio di pubblicazione di open-data di qualità**: poiché il DAF ospita copie aggiornate dei dati presenti nelle basi di dati delle PA, ciascuna PA può decidere di abilitare un servizio per la pubblicazione dei propri open-data direttamente tramite apposite API esposte dal DAF.
- **un insieme di data application** che implementano casi d'uso di interesse per interi cluster di PA (es. Monitoraggio incidenti stradali e individuazione zone a rischio, Servizio per la verifica della qualità delle informazioni contenuti nelle basi di dati della PA, previsionali per i comuni, sentiment analysis, ecc.).
- **servizi di accesso a best effort di eventi real-time** (*egestion*) su flussi, eventualmente arricchiti e/o normalizzati, di dati veicolati verso il DAF e potenzialmente utili per la realizzazione di servizi non critici.

Oltre a quanto messo a disposizione dalla piattaforma, attraverso il team di data scientist, il DAF offre un **supporto alle PA** per la costruzione di modelli di interconnessione delle diverse sorgenti dati, l'analisi dei dati, lo sviluppo di modelli di machine learning, il coordinamento dello sviluppo di data application e l'organizzazione di "competizioni" scientifiche su tematiche di interesse per la PA.

2.4 Roadmap di evoluzione

Lo sviluppo del DAF prevede due fasi principali:

- *Fase 1*: Realizzazione e sperimentazione
- *Fase 2*: Messa in produzione

La prima fase, le cui attività sono concentrate nel secondo semestre 2017, è finalizzata alla realizzazione della piattaforma tecnologica e alla sua sperimentazione sulla base di casi d'uso individuati in collaborazione con alcune PA selezionate. Sarà inoltre avviata una collaborazione con il Garante della Privacy per definire le modalità attraverso le quali le PA potranno formalizzare il rapporto con l'ente al quale, terminata la fase di sperimentazione, sarà affidata la gestione del DAF. In tal senso saranno anche stabilite le regole che, nel rispetto delle norme sulla privacy, definiranno le modalità di caricamento e di analisi dei dati sul DAF, nonché della diffusione dei dati e dei risultati delle analisi stesse.

Nella fase successiva, che andrà a regime dopo che sarà ufficializzato l'ente a cui sarà affidata l'operatività e l'evoluzione del DAF e avviati gli opportuni interventi normativi, il Team Digitale ed AgID predisporranno le procedure atte a rendere operativo quanto prodotto durante la fase sperimentale.

Di seguito una roadmap di alto livello, redatta in base alle informazioni e disponibilità delle PA con cui si è attualmente in contatto, e che potrà essere soggetta a cambiamenti e integrazioni in base ai futuri sviluppi.

- **Entro Ottobre 2017:** sviluppo primo MVP del Dataportal e rilascio per user testing da parte di *community* e PA selezionate.
- **Entro Ottobre 2017:** definizione delle PA centrali e locali coinvolte nella fase di sperimentazione. Si pubblicherà un elenco delle PA aderenti, con aggiornamento continuo.
- **Entro Novembre 2017:** rilascio secondo MVP del Dataportal che recepisce indicazioni degli utenti sulla base dell'utilizzo del primo MVP.
- **Entro Dicembre 2017:** Completamento onboarding delle PA coinvolte nella fase di sperimentazione.
- **Entro Dicembre 2017:** Rilascio della prima release del Dataportal.
- **Entro Dicembre 2017:** Definizione della governance a tendere del DAF e individuazione della PA che avrà in gestione il progetto.
- **Da Gennaio 2018:** passaggio di consegne e training verso la PA che gestirà il DAF. Attivazione processo di onboarding per le PA non ancora aderenti al DAF, sia centrali che locali.
- **Entro Marzo 2018:** sviluppo e rilascio dei casi d'uso elencati nel documento, con modalità *best effort*.

2.5 Modello di governance

Durante la fase sperimentale la governance del DAF è in carico al Team Digitale che 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. Il Team Digitale si farà anche carico di ingaggiare i rapporti con le PA coinvolte nella fase di sperimentazione e lavorerà insieme a loro per lo sviluppo di casi d'uso indicati nella roadmap.

Parallelamente, il Team Digitale, in stretta collaborazione con le istituzioni competenti, lavorerà per l'individuazione di una PA che prenderà in carico il progetto. Il Team Digitale e la PA selezionata lavoreranno in sinergia durante la fase di messa in produzione del DAF, durante la quale si effettueranno i passaggi di consegna e training.

2.6 Architettura di alto livello

L'architettura logica del DAF è basata sui seguenti layers:

- *Micro-Service Layer:* composto da tutti i servizi necessari per implementare le funzionalità della piattaforma. Tutte le componenti sono implementate come microservizi.
- *Ingestion Layer:* dedicato alla realizzazione di tutte le attività utili al caricamento dei dati.
- *Hadoop Computational Layer:* contiene tutte le piattaforme computazionali che tipicamente fanno parte dell'ecosistema Hadoop, tra cui si evidenzia [Spark](#) sul quale si basano la maggior parte delle elaborazioni eseguite nel DAF. I microservizi presenti nel Micro-service layer utilizzano il livello computazionale per eseguire task utili all'accesso ai dati e alle operazioni di manipolazione e trasformazione. L'ingestion layer usa il computational layer per eseguire operazioni di conversione e trasformazione dei dati.
- *Hadoop Storage Layer:* ovvero la piattaforma di memorizzazione dei dati fornita da [Hadoop](#). Benché tutti i dati del DAF siano memorizzati su HDFS (il filesystem distribuito del sistema Hadoop), a seconda delle esigenze è possibile che i dati siano replicati su [Kudu](#) e [HBase](#) per favorire l'efficienza computazionale dei tool di analisi.

L'immagine seguente riassume la vista logica dell'architettura del DAF:

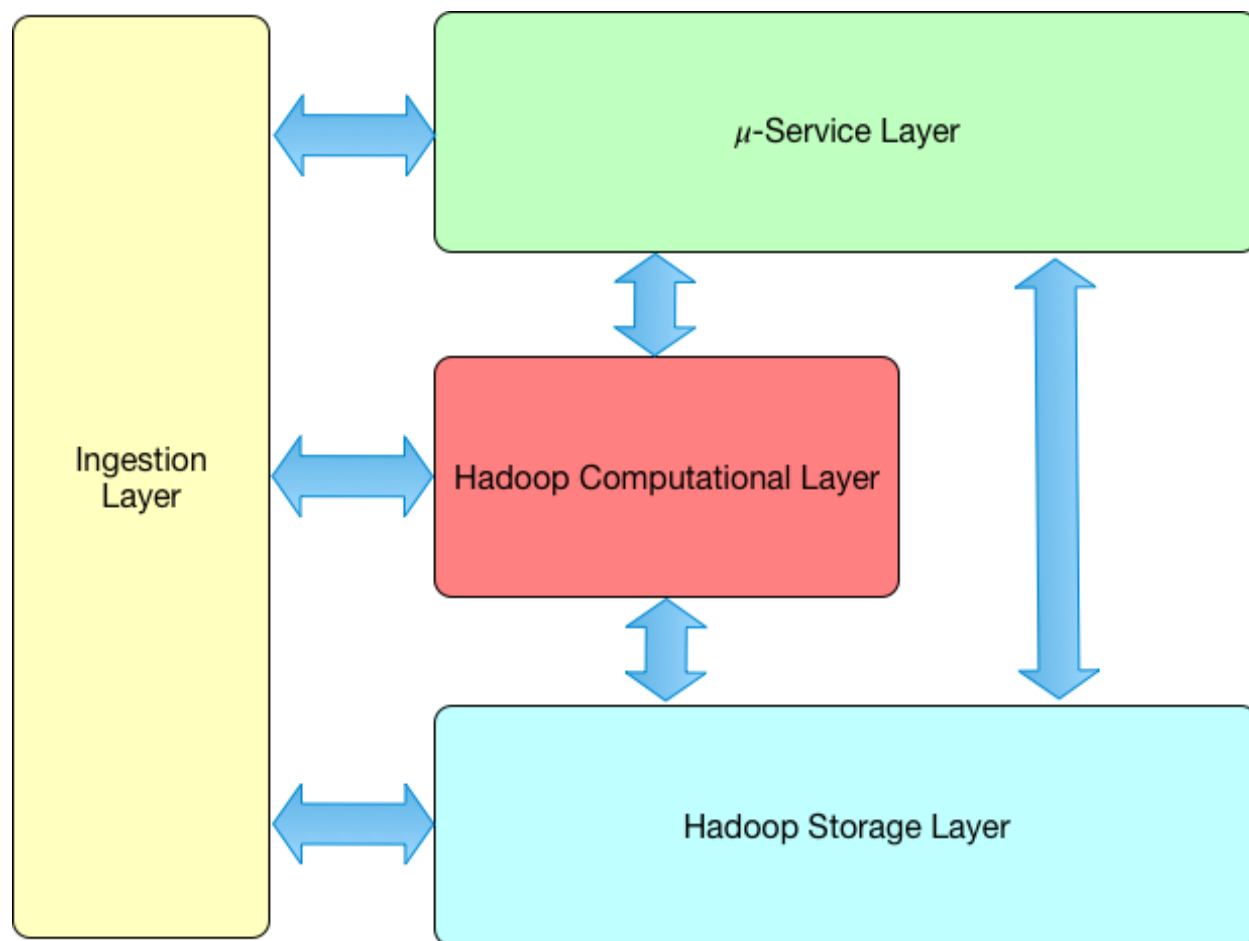


Fig. 1: Logical View

Dal punto di vista dell'architettura utile a ospitare il DAF, la piattaforma DAF è progettata per essere installata su due cluster disgiunti di macchine, così come mostrato nella figura successiva:

In dettaglio:

- *Kubernetes Cluster* - ovvero un cluster Kubernetes composto da nodi che svolgono il ruolo di edge per il cluster Hadoop: tali nodi ospitano i microservizi che fanno, in modalità client, uso delle risorse computazionali e di memorizzazione offerte dal cluster Hadoop.
- *Hadoop Cluster* - ovvero un cluster di macchine su cui è installata una distribuzione Hadoop out-of-the-box in modalità multi-node.

2.7 Casi d'uso della fase sperimentale

A seguire un elenco di casi d'uso già individuati dal team di data scientist del DAF e su cui sono in corso attività di sviluppo e sperimentazione a *best effort*. Questo elenco è da considerarsi dinamico e pertanto sarà arricchito/modificato nel tempo sulla base delle attività concordate con le PA che partecipano alla fase sperimentale. All'interno delle schede di dettaglio sono indicate delle date di scadenza suscettibili a variazioni perché calcolate secondo le informazioni attualmente in nostro possesso. Esse devono infatti essere verificate con le PA coinvolte nelle singole sperimentazioni sulla base dei tempi necessari a queste ultime per fornire i dati o implementare i meccanismi di ingestion.

L'elenco dei casi d'uso è il seguente:

2.7.1 Implementazione e fruizione Dataset Chiave

Descrizione

Il DAF ausilia l'implementazione di **Dataset Chiave**, fornendo un framework per ingestion e normalizzazione di dati potenzialmente provenienti da molteplici fonti, esponendo una modalità di interrogazione unificata e con regole comuni.

Motivazione

Rendere facilmente disponibili e interoperabili una serie di dataset Open Data ritenuti strategici sotto il profilo dei fenomeni in essi descritti. Per questi dataset, si vogliono minimizzare i costi transattivi di utilizzo, prevedendo standard e regole di esposizione più stringenti rispetto a quelle normalmente previste per gli Open Data.

Casi applicativi

- Implementazione Dataset Chiave Ambito Mobilità e Trasporti
- Implementazione Dataset Chiave Ambito Economia e Finanze
- Implementazione Dataset Chiave Ambito Cultura e Turismo

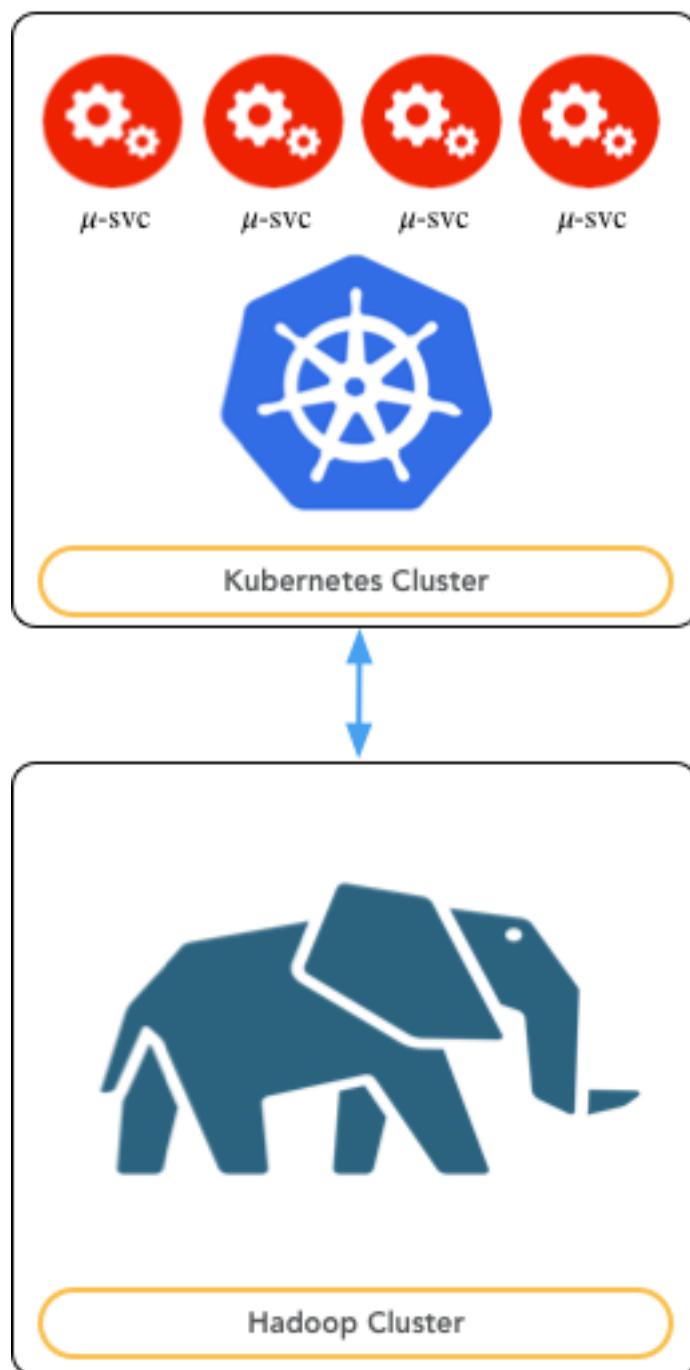


Fig. 2: Deployment View

Schede di dettaglio casi applicativi

Caso applicativo	Implementazione Dataset Chiave Ambito Mobilità e Trasporti
PA coinvolte	TaskForce OpenData, Motorizzazione Civile, ACI
Dataset utilizzati	<i>Statici</i> : Lista Parcheggi, Parco Macchine, Perimetro, varchi e orari ZTL, Patenti Attive, Incidenti Stradali. <i>Real Time</i> : Traffico e viabilità in tempo reale, stato dei parcheggi in tempo reale.
Road-map	Entro dicembre 2017
Stato	In attesa dei dataset
Note	In lavorazione la documentazione relativa alla struttura dei dati per ogni dataset chiave, le pipeline di trasformazione dal dato grezzo in quello Open Data, le procedure di ingestion automatiche dei dati necessari per produrre i dataset chiave menzionati.

Caso applicativo	Implementazione Dataset Chiave Ambito Economia e Finanze
PA coinvolte	UnionCamere, Agenzia delle Entrate
Dataset utilizzati	<i>Statici</i> : Registro delle Imprese (dati anagrafici e bilanci), punti fiduciari, comuni d'Italia e stati esteri, dati statistici su immobili per comune, statistiche su transazioni immobiliari.
Road-map	Entro dicembre 2017
Stato	In attesa di via libera di UnionCamere per contenuto che sarà possibile rendere open. In attesa di apertura dati Agenzia Entrate.
Note	In lavorazione la documentazione relativa alla struttura dei dati per ogni dataset chiave, le pipeline di trasformazione dal dato grezzo in quello Open Data, le procedure di ingestion automatiche dei dati necessari per produrre i dataset chiave menzionati.

Caso applicativo	Implementazione Dataset Chiave Ambito Cultura e Turismo
PA coinvolte	MIBACT, TaskForce OpenData
Dataset utilizzati	<i>Statici</i> : Istituti e luoghi della cultura, eventi culturali, altri eventi assimilabili, catalogo dei beni culturali, cammini e percorsi, altri percorsi assimilabili, strutture ricettive, guide turistiche.
Road-map	Entro dicembre 2017
Stato	In lavorazione
Note	In lavorazione la documentazione relativa alla struttura dei dati per ogni dataset chiave, le pipeline di trasformazione dal dato grezzo in quello Open Data, le procedure di ingestion automatiche dei dati necessari per produrre i dataset chiave menzionati.

2.7.2 Pubblicazione Open Data “certificati” in SaaS (Certified ODaaS)

Descrizione

Alle Pubbliche amministrazioni che conferiscono i dati al DAF, verrà offerta di *default* la possibilità di creare delle pipeline di generazione e pubblicazione di Open Data sul Dataportal (evoluzione di dati.gov.it). In particolare, per i dataset conferiti che andranno a far parte di Dataset Standard del DAF, il sistema prevederà pipeline di trasformazione e pubblicazione tipizzate e automatiche.

Motivazione

- Sgravare le PA dai processi di generazione, aggiornamento e pubblicazione degli Open Data, e di sviluppo e manutenzione dei cosiddetti portali/cataloghi. Il servizio verrà svolto nativamente dal Dataportal che ospiterà delle sezioni dedicate alle singole PA.
- Pubblicare Open Data che soddisfano by design i requisiti previsti dalle linee guida sul tema, con particolare attenzione alla metadatazione, utilizzo di vocabolari controllati e interoperabilità, con modalità di fruizione che incentivano comunicazione machine-to-machine.

Casi applicativi

- Sperimentazione con Comune di Milano

Schede di dettaglio casi applicativi

Caso applicativo	Sperimentazione con Comune di Milano
PA coinvolte	Comune di Milano
Dataset utilizzati	In fase di definizione da parte del Comune
Road-map	In fase di avvio lavori
Stato	Entro dicembre 2017
Note	Output del progetto è di utilizzare le pipeline di ingestion per creazione di dataset standard nel DAF, da cui far partire pipeline di trasformazione degli stessi in dataset Open Data con relativa pubblicazione nella homepage di Milano all'interno del Dataportal.

2.7.3 Strumenti di Analisi & Dashboard Self Service

Descrizione

Attraverso il Dataportal, le PA aderenti al DAF avranno la possibilità di accedere ed analizzare i dati a cui sono abilitati. Di base, saranno messi a disposizione un tool per data exploration, querying, data visualization & costruzione dashboard, e un notebook per analisi più complesse.

Motivazione

Rendere facilmente disponibili e interoperabili una serie di dataset Open Data ritenuti strategici sotto il profilo dei fenomeni in essi descritti. Per questi dataset, si vogliono minimizzare i costi transattivi di utilizzo, prevedendo standard e regole di esposizione più stringenti rispetto a quelle normalmente previste per gli Open Data.

Casi applicativi

- Sperimentazione di Analytics & Dashboarding

Schede di dettaglio casi applicativi

Caso applicativo	Sperimentazione di Analytics & Dashboarding
PA coinvolte	Elenco aperto, correntemente rapporti avviati con Milano, Torino
Dataset utilizzati	In fase di definizione
Roadmap	Dicembre 2017 - Marzo 2018
Stato	In fase di avvio lavori
Note	Si lavorerà insieme alle PA selezionate per guidarle nell'utilizzo degli strumenti messi a disposizione, capirne sul campo i relativi bisogni e realizzare insieme a loro dei PoC esemplificativi.

2.7.4 Verifica della qualità delle informazioni delle base di dati

Descrizione

Il DAF mette a disposizione la sua capacità di calcolo, la presenza (al suo interno) di basi di dati autoritative, nonché l'integrazione con il Registro dei vocabolari controllati, al fine di fornire un servizio per la verifica batch della qualità delle informazioni contenute all'interno delle basi di dati della PA.

Motivazione

Le PA che caricano nel DAF le proprie basi di dati possono rendere più semplice e sistematica la verifica della qualità dei dati.

Casi applicativi

- Indice delle Pubbliche Amministrazioni (IPA)

Schede di dettaglio casi applicativi

Ca- so ap- pli- ca- ti- vo	Indice delle Pubbliche Amministrazione (IPA)
PA coin- vol- te	AgID, Team digitale
Da- ta- set uti- liz- zati	IPA
Road- map	Entro dicembre 2017
Sta- to	In fase di pianificazione
No- te	In IPA sono memorizzate informazioni che si prestano a una serie di verifiche/controlli automatici come, ad esempio le URL, gli indirizzi email, gli indirizzi PEC. Una volta importato IPA nel DAF e avviato un flusso utile al suo aggiornamento, si abilita un flusso di dati per l'aggiornamento periodico e si sviluppa un programma batch che periodicamente verifica la qualità e la completezza dei dati. Il batch produce in modo automatico un report su tutto l'IPA e può inviare via email a ciascun referente IPA la sezione del report che lo riguarda.

2.7.5 Strumenti di monitoraggio automatico

Descrizione

Il DAF può utilizzare la sua capacità computazionale e di storage per effettuare attività di monitoraggio automatico su tematiche di interesse pubblico o di singole PA.

Motivazione

Per sua natura il DAF può offrire strumenti utili al monitoraggio automatico di fenomeni che possono essere misurati intercettando eventi e/o recuperando informazioni sul Web e/o su sistemi informatici accessibili alla PA.

Casi applicativi

- Strumenti di supporto al monitoraggio del Piano Triennale
- Monitoraggio canali social delle PA

Schede di dettaglio casi applicativi

Caso applicativo	Strumenti di supporto al monitoraggio del Piano Triennale
PA coinvolte	AgID, Team Digitale
Dataset utilizzati	IPA, altre datasource da identificare insieme ad AgID
Roadmap	Entro dicembre 2017
Stato	In attesa della definizione degli indicatori del Piano Triennale
Note	Il team di data scientist ed AgID verificheranno quali indicatori del piano triennale si prestano ad essere misurati in modo automatico per mezzo del DAF.

Caso applicativo	Monitoraggio canali social delle PA
PA coinvolte	Team Digitale
Dataset utilizzati	Dati provenienti da uno o più social network
Roadmap	Marzo 2018
Stato	In fase di studio
Note	L'output minimo è un dashboard che offre misurazioni sull'uso dei social network da parte delle PA (es. Classifiche delle PA più attive, PA più seguite, ecc). Sarà inoltre valutata la possibilità di realizzare cruscotti utili a monitorare in tempo reale gli argomenti più trattati sui canali della PA, in modo da offrire uno strumento utile a velocizzare l'individuazione delle esigenze dei cittadini.

2.7.6 Servizi di accesso ad eventi di interesse pubblico

Descrizione

Il DAF offre un'infrastruttura Big data in grado di svolgere il ruolo di aggregatore ed erogatore flussi di dati relativi ad eventi di interesse pubblico. In funzione della quantità e della frequenza, il DAF può implementare Streaming API, fornire servizi RSS, inviare notifiche mediante piattaforme di social network, ecc.

Motivazione

Grazie al DAF è possibile aggregare e uniformare flussi di eventi che ad oggi sono spesso erogati da molteplici sorgenti in modo disomogeneo. La capacità del DAF di intercettare ed eventualmente memorizzare, tali eventi abilita lo sviluppo: di applicazioni real-time per cittadini e imprese di servizi di notifica best effort utili a favorire la circolarità dell'informazione tra PA.

Casi applicativi

- Servizio notifica dei trasporti pubblici

Schede di dettaglio casi applicativi

Caso applicativo	Servizio notifica dei trasporti pubblici
PA coinvolte	MIT, Task Force DatiPubblici
Dataset utilizzati	Dataset direttamente forniti da PA che collaborano attivamente alla sperimentazione e dataset già disponibili sotto forma di open-data.
Road-map	In fase di studio
Stato	Marzo 2017
Note	Il dominio dei trasporti pubblici è particolarmente favorevole alla conduzione di attività sperimentali perché ci sono un numero significativo di dati già pubblicati come opendata utilizzando formati che rappresentano standard de-facto di dominio (es. gtfs, gtfs-realtime).

2.7.7 Data Applications

Descrizione

Una data application è un prodotto software che svolge delle attività ed espone funzionalità sulla base di un contenuto informativo estratto dai dati. Il DAF offre capacità per utilizzare i dati in esso contenuti per sviluppare ed esporre le funzionalità.

Motivazione

Le funzionalità esposte dalle data application costituiscono una modalità di valorizzazione del patrimonio informativo pubblico, perché offrono a cittadini e PA servizi basati sull'utilizzo delle informazioni estratte dalle banche dati pubbliche.

Casi applicativi

- Sviluppo algoritmo di network analysis & collocation mining applicato al registro delle imprese e arricchito con altri dataset contenuti nel DAF
- Mappa Interattiva Sicurezza Urbana
- Analisi e monitoraggio sinistri stradali
- Analisi per ottimizzazione politiche su pianificazione asili

Schede di dettaglio casi applicativi

Caso applicativo	Algoritmo di network analysis e collocation mining applicato al registro delle imprese
PA coinvolte	Unioncamere - InfoCamere
Dataset utilizzati	Registro delle imprese, altri dataset da identificare
Roadmap	Entro Dicembre 2017
Stato	In fase di studio
Note	Al momento lo studio sarà effettuato sul campione di imprese identificato da InfoCamere. L'obiettivo è di fornire uno strumento utile a capire i collegamenti tra imprese (network analysis) e zone con concentrazione di attività vs indicatori di profittabilità relativi.

Caso applicativo	Mappa Interattiva Sicurezza Urbana
PA coinvolte	Comune di Torino
Dataset utilizzati	Dataset su segnalazioni da parte dei cittadini al call center della Polizia Municipale
Roadmap	Entro Dicembre 2017
Stato	In fase di studio
Note	Il deliverable è una mappa in cui visualizzare le segnalazioni, con indicazioni di quali sono le zone più a rischio della città; una dashboard per la visualizzazione dei dati relativi.

Caso applicativo	Analisi e monitoraggio sinistri stradali
PA coinvolte	Comune di Torino
Dataset utilizzati	Dataset sui sinistri stradali
Roadmap	Entro Dicembre 2017
Stato	In fase di studio
Note	Il deliverable è una mappa in cui visualizzare i sinistri, con indicazioni di quali sono le zone più a rischio della città; una dashboard per la visualizzazione dei dati relativi.

Caso applicativo	Analisi per ottimizzazione politiche su pianificazione asili
PA coinvolte	Comune di Milano
Dataset utilizzati	Dataset su popolazione residente, elenco asili, elenco servizi, e altro da identificare
Roadmap	Entro Dicembre 2017
Stato	In fase di studio
Note	Il deliverable è un modello che aiuta ad identificare le aree dove è ottimale l'apertura di nuovi asili.

2.7.8 Valorizzazione del patrimonio informativo pubblico

Descrizione

La valorizzazione del patrimonio informativo pubblico passa anche per iniziative volte a sensibilizzare le PA sulle potenzialità offerte da una migliore gestione delle proprie basi di dati: ciò vale sia in termini di conoscenza utile all'ottimizzazione dei servizi esistenti, sia alla creazione di servizi innovativi. Inoltre poiché l'individuazione di algoritmi e tecniche di analisi innovative richiede spesso competenze scientifiche e tecniche specifiche, si reputa utile avviare un processo virtuoso di coinvolgimento della comunità scientifica su tematiche di interesse pubblico.

Motivazione

Il team di data scientist del DAF promuove iniziative volte alla valorizzazione del patrimonio informativo pubblico.

Casi applicativi

- Coinvolgimento della comunità dei data scientist e del mondo della ricerca

Schede di dettaglio casi applicativi

Ca-so ap-pli-ca-tivo	Coinvolgimento di data scientist e della comunità della ricerca scientifica
PA coin-vol-te	Team Digitale
Da-ta-set uti-liz-zati	I dataset saranno selezionati tra gli open-data contenuti nel DAF
Road-map	Marzo 2018
Sta-to	In fase di studio
No-te	I data scientist del DAF definiranno, a partire dai dati pubblici presenti nel DAF e dalle esigenze applicative comunicate dalle PA, task specifici da sottoporre alle comunità in oggetto. I task saranno composti da una chiara descrizione del problema applicativo da risolvere, da un dataset su cui condurre sperimentazioni e da un insieme di metriche per mezzo delle quali misurare e confrontare la bontà delle soluzioni individuate.