



**Scouting tecnologico e primi test
del nuovo servizio di Business Analytics
della Direzione Sistemi Informativi INFN**

Claudio Galli¹

¹⁾ *INFN, Direzione Sistemi Informativi, 00044 Frascati (Roma) Italy*

Abstract

Dalla fine del 2021 in poi sempre più utenti hanno espresso la necessità di analizzare i dati provenienti dai sistemi della Direzione Sistemi Informativi (DSI) con maggior autonomia rispetto a quanto non fosse possibile fino a quel momento. Per dare una risposta adeguata a tali richieste è stata avviata una fase di *scouting* tecnologico coinvolgendo diversi *competitor* internazionali nel settore *Business Analytics*. L'obiettivo era identificare quale soluzione potesse meglio rispondere alle nuove esigenze, nel rispetto delle linee guida dell'Agenzia per L'Italia Digitale (AGID)[1] per la transizione digitale delle pubbliche amministrazioni (PA). Un requisito fondamentale è stato il grado di integrazione che tali prodotti garantivano. Si è valutato quindi la capacità di connettersi con altre tecnologie impiegate nella direzione, la sostenibilità infrastrutturale (risorse, attività di *DevOps* ...) e quanto si integrassero con gli strumenti usati abitualmente dalle altre direzioni ed utenti INFN. Dopo una breve introduzione, questo elaborato si concentra sul prodotto scelto, Microsoft Power BI Premium, le funzionalità più importanti che offre in associazione con il *Tenant 365* dell'INFN ed il *setup* del nuovo servizio di *Self Service BI*.

DOI: 10.15161/oar.it/76983

1 Scenari e Casi d'Uso

Nel corso degli ultimi anni nel servizio *data warehouse e Supporto all'Utenza* della *Direzione Sistemi Informativi (DSI)* è maturata la percezione che per rispondere meglio alle esigenze di alcune tipologie di utenti fosse necessario ampliare il servizio di *Business Intelligence INFN* con strumenti di *Self-BI ed Analytics*. Nonostante il *tool* su cui si basava il servizio fosse ottimo nel produrre report statici, caratterizzati da una struttura consolidata nel tempo, mancava della flessibilità necessaria a rispondere ai frequenti cambi di requisiti che alcuni scenari d'uso imponevano. Anche dal punto di vista delle più moderne tecniche di *Data Visualization* presentava diverse limitazioni. La maggior inadeguatezza però si palesava in quelle situazioni in cui l'utente esprimeva la necessità di condurre in autonomia delle analisi cambiando di volta in volta le logiche di aggregazione dei dati e/o il formato di presentazione dei risultati.

Tra il 2021 e la fine del 2022 il servizio *data warehouse e Supporto all'Utenza* della DSI si era dato l'obiettivo di identificare nuove strategie e *tools* da affiancare gli strumenti in uso, al fine di restituire agli utenti il giusto grado di autonomia nell'analisi dei dati e migliorare il servizio offerto.

1.1 La Strategia identificata

La strategia individuata lavorava contemporaneamente su diversi fronti tra loro complementari :

- 1 identificare uno strumento di *Self-BI*¹, o più in generale di *Business Analytics*, che consentisse agli utenti di essere autonomi in tutte le fasi di esplorazione dei dati e successivamente nella presentazione dei risultati ottenuti;
- 2 la progressiva ristrutturazione dei dati della DSI integrati nel *data warehouse* andando a costruire *ex-novo* una serie di modelli dati multidimensionali[2], consistenti, certificati, esplorabili e ben documentati, da fornire agli utenti come sorgenti autoritative per le loro analisi;
- 3 identificare quei gruppi di utenti che hanno esigenze particolari e delegare ad alcune persone al loro interno le responsabilità della creazione dei report e delle dashboard necessarie per l'analisi dei dati. Ovviamente, questo implica l'accesso controllato ai modelli dati ed un'adeguata formazione sull'uso degli strumenti;

La scelta di questo approccio presenta diversi vantaggi. In passato era capitato spesso infatti che il management INFN, o interlocutori esterni all'istituto, richiedessero una serie di informazioni alle varie direzioni, commissioni, o gruppi a vario titolo. Nel caso in cui i tool gestionali non consentissero di reperire i dati necessari, questi gruppi si

¹Questa attività si è conclusa a Dicembre 2022 con l'acquisto delle licenze Microsoft Power BI Premium per Capacity liv.P1 e il primo setup dell'ambiente che è argomento di questo elaborato.

trovavano costretti a girare le richieste al personale della DSI per ottenere le informazioni necessarie da elaborare e spesso anche le elaborazioni finali risultavano a carico del personale della DSI. Riportare lo sviluppo della reportistica in mano agli utenti vuol dire renderli autonomi nel gestire le richieste del management, evitando complesse interazioni con la DSI per raggiungere la struttura finale dei report. Questo implica anche valorizzare le competenze di quei gruppi di utenti dando maggior risalto al loro lavoro sui dati. D'altra parte significava anche ridurre il carico di lavoro del personale della DSI, in particolare del gruppo data warehouse che si sarebbe potuto concentrare maggiormente sulla costruzione di un *modello di conoscenza* integrato che favorisca la "*Data Governance*" all'interno della Direzione Sistemi Informativi.

D'altra parte era importante che lo strumento di *Business Analytics* scelto si potesse integrare nell'infrastruttura di *back-end* del servizio di *Business Intelligence*². Oltre a questo era fondamentale che fosse possibile condividere i risultati delle analisi e che si interfacciasse in maniera facile con i tool che le varie direzioni ed amministrazioni dell'ente erano solite utilizzare quotidianamente.

Per ottenere questo risultato sono state prese in considerazione diverse tecnologie e servizi prendendo a riferimento i produttori leader di mercato nel settore della *Business Analytics*.

²Il servizio si compone di un database che implementa il *data warehouse* e di un prodotto di TIBCO[3] per la manipolazione e preparazione dei dati che prende il nome di *Jasper ETL (Extract Transform Load)*

2 Posizionamento delle soluzioni nel mercato della Business Analytics

Il percorso di esplorazione tecnologica si è sviluppato in due principali fasi. La prima si è concentrata sulla ricerca di strumenti software e soluzioni *on premises* (OP)³ da poter installare sulla piattaforma di *Business Continuity* a cui si affida la DSI per i propri sistemi. La seconda fase invece ha allargato gli orizzonti verso soluzioni *on cloud* (OC)⁴ compatibili con le direttive AGID[1] per la PA (Pubblica Amministrazione). Un elenco aggiornato di tali servizi è disponibile oggi sul sito *Cloud Marketplace*[4] dell'ACN (Agenzia per la Cybersicurezza Nazionale)[5].

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms



Source: Gartner (March 2022)

Figura 1: Magic Quadrant for Analytics and Business Intelligence Platforms Marzo 2022.

Ugualmente possiamo identificare due differenti set di requisiti, o meglio di *desiderata*, tipici delle soluzioni *OP* piuttosto che delle soluzioni *OC*. Da esse conseguono

³per brevità d'ora in poi si indicheranno con OP le soluzioni basate su installazioni On Premises

⁴per brevità d'ora in poi si indicheranno con OC le soluzioni basate su servizi in Cloud

due distinte categorie di criteri di valutazione che porteranno poi alla scelta finale di una soluzione rispetto all' altra.

In figura [fig.1] è illustrato il quadrante di Gartner di Marzo 2022[6] che riporta il posizionamento dei vari tool e piattaforme di *Analytics* e *Business Intelligence* presenti sul mercato.

Come si può notare il grafico divide i tool in quattro categorie, ovvero *Challengers*, *Leaders*, *Visionaries*, *Niche Players*. Si riporta una breve descrizione generale delle quattro aree :

- **Niche Players** : sono quelle realtà fortemente specializzate nella risoluzione di un determinato problema. Non mostrano una ampia visione di mercato, ma dimostrano comunque una buona efficienza nell'assolvere il compito che si prefiggono.
- **Visionaries** : queste sono le compagnie con una chiara visione del mercato e delle sue esigenze o di come rimodellare la percezione di una determinata tecnologia; mancano però di efficacia nel dimostrare o affermare la loro visione sul mercato.
- **Challengers** : Sono realtà ben strutturate che hanno una visione chiara della tecnologia e tendono a definire degli standard di mercato, ma non sempre riescono a cogliere appieno le esigenze che gli utenti dimostrano o dimostreranno in futuro con il maturare delle tecnologie.
- **Leaders** : vengono considerate Leader quelle realtà che hanno dimostrato anche in passato la capacità di utilizzare al meglio gli standard, guidare il mercato e interpretare le esigenze presenti e future degli utenti con una *roadmap* concreta e chiara per il futuro.

I quattro prodotti che sono stati esaminati provengono principalmente dal quadrante dei *Leader* di mercato con l'unica eccezione di Tibco Spotfire che viene posto tra i prodotti *Visionaries*. Per una prima indagine superficiale delle caratteristiche dei servizi e prodotti ci si è affidati ai risultati pubblicati e condivisi sulla piattaforma *SelectHub.com* [7] che espone informazioni utili per un primo approccio agli strumenti e una indicazione di massima anche dei *feedback* degli utenti. In una successiva fase sono stati presi contatti con tutti i partner selezionati e si è proceduto con delle call esplorative e in alcuni casi con dei test in locale e su servizi online.

2.1 Dove e cosa cercare

Come già detto le prime indagini sono state condotte sotto l'ipotesi di una gestione *OP* del nuovo servizio di *Analytics*. L'installazione sulle risorse hardware messe a disposizione della DSI dall'infrastruttura di Business Continuity INFN impone alcuni vincoli sulla soluzione da adottare. La scelta del sistema operativo per le VM (*Virtual machines*), la famiglia di processori, la quantità e qualità dello storage o della memoria erano sicuramente alcuni degli aspetti che potevano condizionare la scelta del tool. A queste si affiancava l'esigenza di integrare la configurazione del sistema con le attuali procedure di gestione da parte del servizio Infrastruttura della DSI.

Questo primo vincolo, insieme con la necessità di integrare la gestione IT del servizio nell'attuale flusso di lavoro da parte del servizio Infrastruttura della DSI, implicava al tempo delle prime indagini l'esclusione delle soluzioni *Windows based*, la cui gestione non era integrata nel normale flusso di *DevOps* del servizio Infrastruttura.

Questo di fatto portava all'esclusione automatica di due *competitor* su quattro; Microsoft e Qlick che prevedevano installazioni *OP* solo su macchine *Windows Server*.

Per quanto riguarda gli altri due *competitor*, Tibco Spotfire e Salesforce Tableau, sono stati fatti test aggiuntivi, anche con installazioni desktop di prova e si è guardato più nel dettaglio alle risorse *hardware* necessarie. Proprio il computo dei requisiti *hardware* è risultato un secondo punto critico. Infatti ci si è accorti subito che, per offrire all'utente un servizio *responsive*, sarebbe stato necessario allocare una quantità di risorse considerevole.

Per fare un esempio, in tabella [tbl:1] sono riportate le risorse minime dichiarate rispettivamente da Tibco in [8] e Tableau in [9] per una installazione *Single Instance* del servizio. Nel caso di Tableau viene riportata anche l'indicazione della variazione richiesta per una configurazione minimale di una installazione *multi node*.

Spotfire	Tableau
<i>CPU</i> : 64bit 2Core (4 vCPUs)	64bit 4Core (8vCPUs)
<i>RAM</i> : 16GB	64GB nodo iniziale + 16GB per ogni nodo aggiuntivo
<i>DISK</i> : 10GB	15GB

Tabella 1: Risorse *hardware* minime per installazione *OP Single Instance*.

A prima vista sembrerebbero risorse equiparabili, ma andando a cercare nella documentazione o investigando con i produttori quali siano i valori raccomandati troviamo cifre ben diverse. Innanzi tutto in entrambi i casi, anche se in modi diversi, il servizio può essere scomposto in sotto servizi dedicati a specifiche funzioni che in molti casi risultano onerosi dal punto di vista computazionale o per l'impatto sulla memoria delle VM che li ospitano. Quindi spesso si incontrano raccomandazioni che promuovono a vario titolo la configurazione del servizio in modalità *multi nodes*. In secondo luogo per avere un

servizio di medio carico che risulti *responsive* nei confronti degli utenti tipo e taglia delle CPU e quantitativo di RAM varia notevolmente.

In tabella [tbl:2] sono riportate alcune stime fatte raccogliendo informazioni in rete, leggendo la documentazione o parlando con consulenti delle varie realtà. Essendo una stima, che ha valore solo a titolo esemplificativo e nel periodo in cui è stata fatta, non ha la pretesa di essere una indicazione puntuale, bensì dare semplicemente l'idea di come e quanto possa discostarsi una eventuale installazione di sviluppo da uno scenario più vicino ad un servizio *production ready*

Spotfire	Tableau
<i>CPU</i> : 64bit 2Core (4 vCPUs)	64bit 4Core (8vCPUs)
<i>RAM</i> : 16GB	64GB nodo iniziale + 16GB per ogni nodo aggiuntivo
<i>DISK</i> : 10GB	15GB

Tabella 2: Risorse *hardware* raccomandate per una installazione *OP Multi Nodes*.

Mentre venivano prese in considerazione le varie opzioni *OP*, si è anche iniziato a considerare l'eventualità che il nuovo servizio di *Analytics* potesse essere basato su un servizio in *cloud*. Le premesse per questa nuova opportunità derivavano in parte dal fatto che i *competitor* stessi si facevano promotori di soluzioni *OC*, in parte dal fatto che le stesse soluzioni *cloud* iniziavano ad essere certificate sul portale dell'AGID. Quindi si è presa in considerazione anche l'ipotesi di soluzioni *OC*, dato che l'ente già si stava avvalendo di alcuni servizi *cloud* certificati per la PA come Microsoft Office365.

3 Setup del servizio Microsoft Power BI Premium per Capacity

A Giugno 2022 Microsoft ha attivato gratuitamente una sottoscrizione di test per il servizio *Power BI Premium per Capacity* sul *tenant* Office 365 dell'INFN. Questo ha permesso di effettuare le prime prove sul funzionamento del servizio e di iniziare a immaginare come configurare i vari componenti e gestire le funzionalità a disposizione.

Il primo elemento preso in esame è stato il modello di sviluppo. Il servizio infatti permette di implementare diversi scenari d'uso in base al tipo di licenze (per utenti o per *Capacity*), sottoscrizioni, utenze (in termini di amministratori, collaboratori, visualizzatori ecc...) e livelli di servizio che si vuole garantire.

La figura [fig:2], e nella successiva [fig:3], sono estrapolate dal documento "*Planning Power BI Enterprise Deployment*" [10] v3.1 di Maggio 2020 scaricabile e liberamente fruibile in formato pdf dal sito ufficiale della documentazione Microsoft. Qui vengono utilizzate per gentile concessione di Microsoft al fine di mettere in evidenza alcune caratteristiche degli scenari d'uso considerati.

Nella prima delle due viene schematizzato un ipotetico ciclo di vita dei report suddiviso in tre fasi, (*Development*, *Testing/QA*⁵, *Production*). Le tre fasi vengono implementate attraverso tre diverse aree di lavoro, *workspace*⁶.

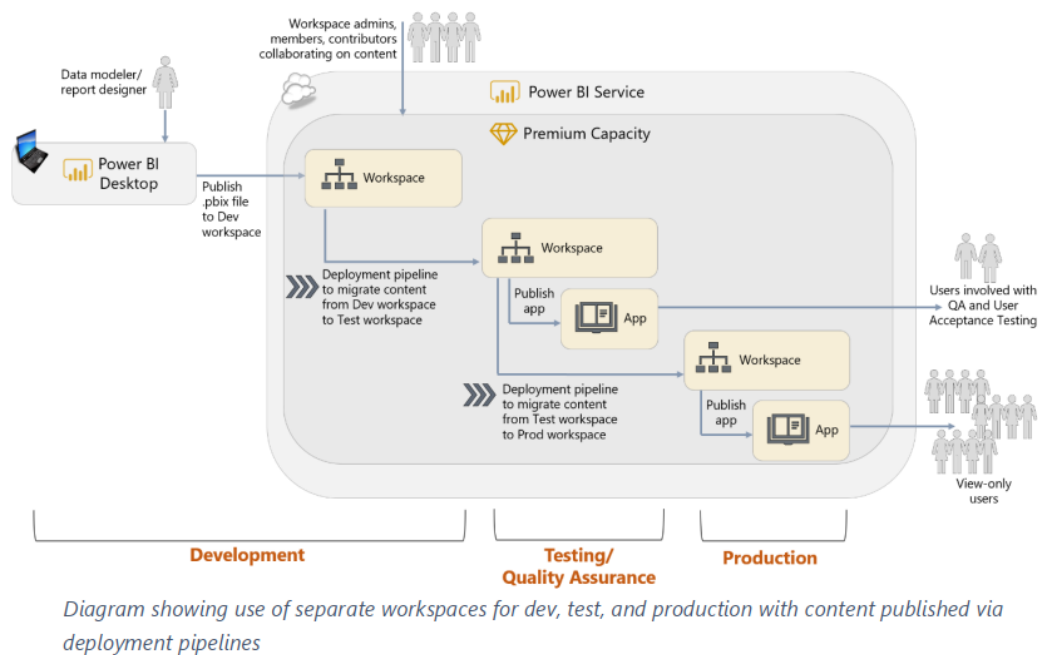


Figura 2: Ciclo di distribuzione dei contenuti tra "aree di lavoro" e produzione di "App" per la distribuzione dei reports.

⁵QA in genere indica *Quality Assurance* e si riferisce a tutte quelle attività volte a certificare la qualità di un prodotto

⁶Il fatto che i *workspace* siano di tipo *Premium* è rappresentato dall'icona del diamante.

Una caratteristica importante che viene messa in luce è che il set di utenti che operano sulle diverse aree di lavoro può essere configurato a seconda delle proprie esigenze. Sono stati identificati almeno tre gruppi di utenti. Il primo, composto da personale della DSI, si occuperà dell'amministrazione del servizio, di sviluppare e documentare i modelli dati (*dataset*). A partire da questi si occuperà di creare la reportistica minima necessaria per consentirne la validazione.

Successivamente i *dataset* promossi in ambiente di *Test/QA* verranno certificati da un secondo gruppo di utenti che chiamiamo *Esperti di Dominio* i quali collaborano con la DSI e hanno competenze specifiche sui processi e gli applicativi a cui il *dataset* fa riferimento.

Infine ci sono uno o più gruppi di utenti, esterni alla DSI, che saranno i veri "clienti" finali dell'infrastruttura di *Analyticse* che accederanno ai report e *dataset* condivisi tramite il *workspace* di produzione.

Questa condivisione può avvenire in diversi modi a seconda delle caratteristiche degli utenti che potremmo riassumere in due distinti profili, base o avanzato.

La principale discriminante che distingue i due profili è il grado di autonomia di cui hanno bisogno gli utenti nel trattamento dei dati.

Profilo Base: identifica utenti, come i "*Resp. Nazionali*" ed i "*Financial Officer*", che non hanno necessità di un'area dedicata in cui sviluppare autonomamente report e dashboard, ma che hanno bisogno di tenere sotto controllo degli indicatori e fare analisi dei dati secondo modalità consolidate e percorsi guidati. Tipicamente sono gruppi numerosi in cui il modello di analisi dei dati è noto e ben consolidato.

Profilo Avanzato: questa seconda tipologia di utenti invece manifesta l'esigenza di una maggiore autonomia nella creazione e distribuzione dei report costruiti a partire dai *dataset* certificati dalla DSI. Tipicamente in questi gruppi di utenti è necessario che sia presente qualcuno che conosca bene lo strumento di *Power BI Desktop* per la creazione di report e dashboard, ma che allo stesso tempo possa lavorare con i modelli dati e conosca le esigenze del gruppo di riferimento. In questa categoria potrebbero ad esempio rientrare le varie direzioni dell'Amministrazione Centrale (AC).

Per ottenere questo risultato è necessario poter ricondurre diversi scenari d'uso del servizio alla stessa "*struttura organizzativa*" di riferimento. Dove molti strumenti e servizi di *Business Analytics* impongono una scelta netta sull'implementazione del servizio, *Power BI Premium* offre flessibilità e aree di lavoro configurabili di fatto consentendo di *sovrapporre* diversi schemi architetturali in maniera semplice ed efficace.

La configurazione finale del servizio dovrà tenere conto di un processo di sviluppo *ibrido* che consenta di sviluppare *dataset* per gli utenti con profilo *avanzato* ed al contempo mettere a disposizione degli utenti con profilo *base* dei percorsi guidati nell'analisi dei dati. Questa seconda modalità di condivisione dei report viene identificata nella figura [fig:2] con il termine "*App*".

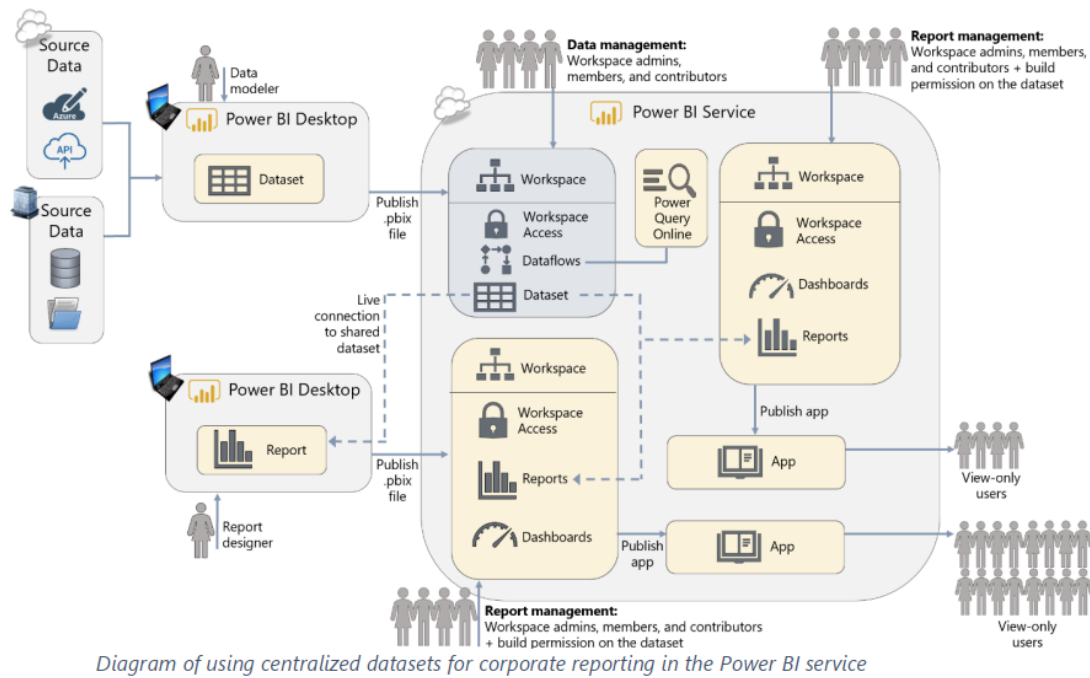


Figura 3: Scenari e *workspace* scelti tra i possibili presentati nel *Whitepaper* di riferimento [10]

Tramite la creazione di App, secondo la definizione di Microsoft, è possibile creare un pacchetto che includa il *dataset* con i riferimenti alle sorgenti e le politiche di aggiornamento, le regole di accesso ai dati ⁷ e tutto quello che serve per la visualizzazione ed esplorazione dei dati; il tutto fruibile attraverso un "portale" che facilita la raccolta di più report diversi sotto lo stesso contesto. In 3.4 ne viene descritto un esempio di come rendere disponibili i percorsi guidati per gli utenti con profilo "base" che nella [fig:2 e 3] sono indicati come *View-only User*.

In [fig:3] è invece descritto come si possa usare il Servizio di *Power BI*, nel nostro caso con licenza *Premium per Capacity*, per pubblicare *dataset* e renderli disponibili ad altri utenti tramite *workspace* dedicati. Questo meccanismo di delega è possibile grazie alla configurazione di *access policy* all'interno dei *workspace* e sui *dataset* prodotti.

Questo permetterebbe alla DSI di indirizzare le esigenze degli utenti con profilo avanzato attraverso la configurazione di *workspace* secondari⁸.

Sarebbe poi responsabilità degli **amministratori** e **collaboratori** gestire il *workspace* e creare la reportistica per i **membri** del *workspace* secondario, mentre resterebbe in capo al personale della DSI fornire la documentazione necessaria al corretto utilizzo dei dati contenuti nei *dataset*.

⁷Tipicamente si fa riferimento alla tecnica denominata RLS, *Row Level Security*, che garantisce l'accesso alle sole righe del *dataset* che riportano uno o più valori specifici in genere associati a dei profili dell'utente o del gruppo a cui appartiene.

⁸In questo caso vengono identificati come *workspace* secondari tutti i *workspace* dedicati a gruppi di utenti e direzioni dell'INFN che necessitano di uno spazio autonomo in cui produrre i propri reports e dashboard

La linea grigia tratteggiata di figura [fig:3] mostra come lo stesso *dataset* possa alimentare diversi *workspace* secondari o report distribuiti tramite App.

In questo complesso meccanismo di delega ci sono alcuni punti particolarmente delicati da tenere in considerazione, tra cui citiamo :

- 1) le modalità di accesso alle risorse *OP INFN*.
- 2) la promozione di un report dall'ambiente di Sviluppo, a Test ed infine a Produzione.
- 3) la visualizzazione delle dipendenze tra sorgenti dati, *dataset*, report, dashboard e App prodotte dalla DSI e le risorse prodotte dalle altre direzioni.
- 4) l'analisi dell'impatto che una modifica o aggiornamento dei *dataset* potrebbe avere sui report che vi attingono.
- 5) la distribuzione dei report tramite App verso tutti quegli utenti con profilo base.

Nelle prossime sezioni vengono presentate in maniera sintetica le soluzioni implementate e le funzionalità proposte da *Power BI Premium* per la gestione di questi aspetti del servizio.

3.1 Profili di Accesso alle Risorse

Il primo punto critico da affrontare è permettere al servizio *OC* di accedere in sicurezza ai dati INFN contenuti nel *data warehouse* della DSI. Questo implica un accesso controllato a risorse interne all'infrastruttura INFN. L'obiettivo primario restava comunque ottenere una configurazione quanto più possibile semplice e sicura, che privilegiasse il disaccoppiamento delle infrastrutture, e consentisse di rendere disponibili i dati tramite il Servizio di *Power BI*. La scelta di un livello di licenza *Premium P1 per Capacity* è stata fondamentale in quanto garantisce all'INFN che il servizio venga erogato tramite hardware dedicato su un *virtual datacenter*, anch'esso dedicato all'INFN. Fatta questa scelta, restava il dubbio su come implementare il collegamento tra servizio online e *data warehouse* della DSI.

Dopo aver investigato diverse ipotesi, si è scelto di installare su macchine virtuali della DSI il servizio ancillare, *Power BI Gateway*⁹. In particolare avendo tre ambienti *data warehouse* (dev, test e produzione) si è scelto, come da *best practice Microsoft*, di installarne 3 *Gateway* distinti. Questa scelta consente di isolare eventuali problemi ed investigare separatamente il carico di lavoro sui vari ambienti.

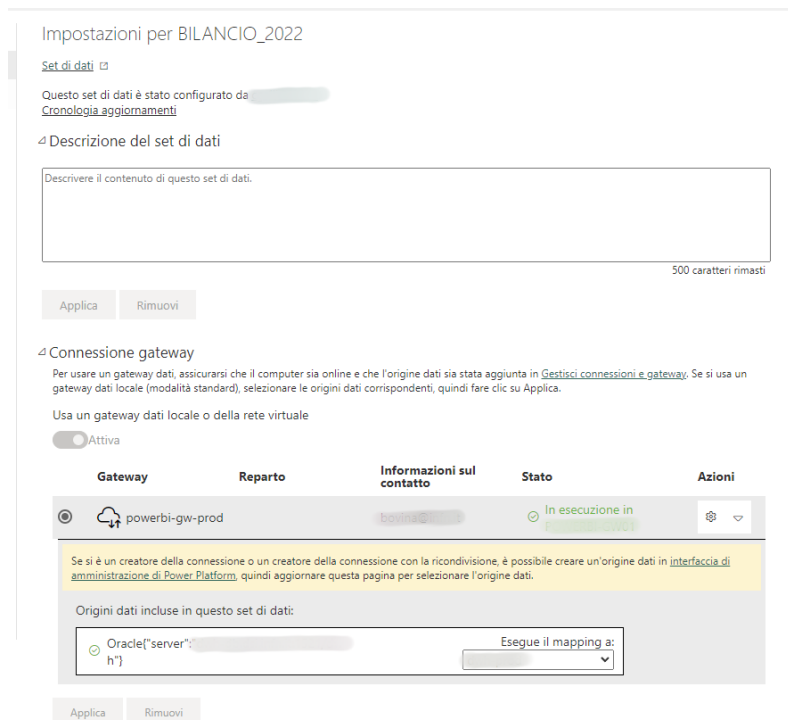


Figura 4: MS Power BI Analisi Dipendenze *dataset* e Report.

Il servizio *Gateway* si comporta fundamentalmente come un *proxy* per le richieste provenienti dal servizio *Power BI* e dirette verso il corrispondente *data warehouse* della

⁹L'installazione dei *Gateway*, così come la gestione e manutenzione dei servizi infrastrutturali è in capo all'Ufficio Infrastruttura della DSI che segue ogni attività di installazione, manutenzione ed aggiornamento dei servizi e delle macchine in uso

DSI. Ogni *Gateway* garantisce un accesso riservato verso un solo *data warehouse* tramite un'utente a bassi privilegi; ovvero un utente on accesso *read only* che espone le sole risorse necessarie.

Una volta che un *dataset* viene sviluppato tramite l'applicativo *Power BI Desktop* va poi pubblicato nell'area di sviluppo. A questo punto è necessario garantirgli l'accesso al *Gateway* appropriato per consentire l'aggiornamento delle informazioni contenute nel modello dati. In figura [fig:4] viene mostrato il *setup* da effettuare per garantire questo canale di comunicazione tra *dataset* online e *data warehouse*. Nell'esempio viene mostrata la configurazione del *dataset* dei dati di Bilancio 2022 (ambiente di sviluppo) che consente al *dataset* di aggiornare giornalmente le informazioni tramite il *Gateway* di sviluppo.

Affinché il *Gateway* funzioni una volta pubblicato online, è necessario che contenga le informazioni di connessione appropriate alla tecnologia "dietro" al *Gateway*. Questo vuole dire che se il *Gateway* è configurato per esporre i dati di un database *Oracle*, ma il *dataset* è stato configurato a partire da dati presenti su un database *Postgres*, allora non sarà possibile configurare correttamente gli aggiornamenti automatici dei dati. Esiste quindi una sorta di mappatura tra la configurazione del *Gateway* da un lato e la tecnologia delle risorse *OP*, che sono le sorgenti autoritative dall'altro.

In figura [fig:5] e [fig:6] vengono mostrate le principali tecnologie che il servizio, e i *dataset* in generale, sono in grado di intercettare come sorgenti dati tramite il *Gateway*.

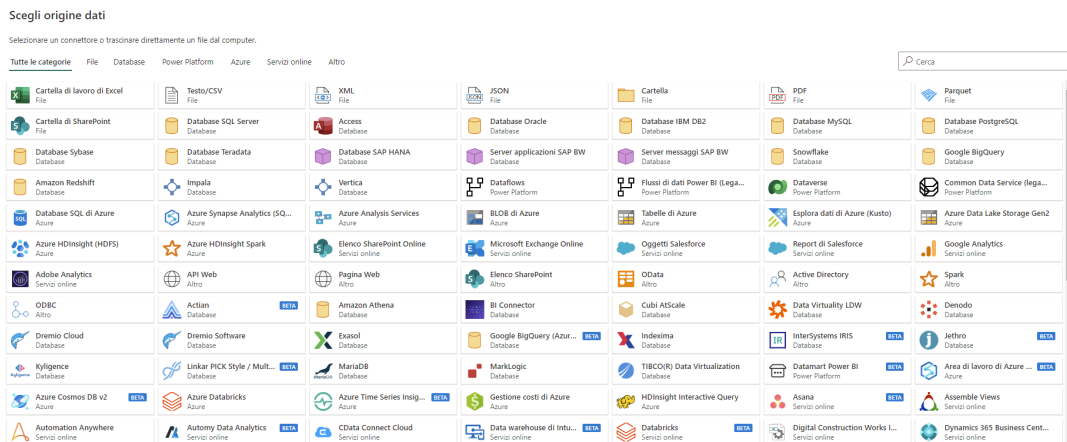


Figura 5: *Power BI Service*: tipologie di sorgenti dati accessibili.

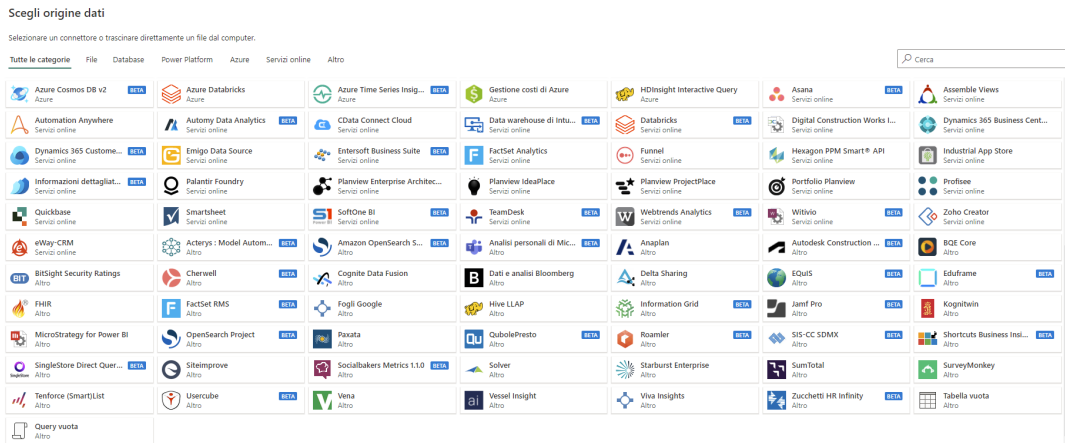


Figura 6: *Power BI Service*: tipologie di sorgenti dati accessibili.

3.2 Data Pipeline

Il secondo punto che viene presentato dipende direttamente della scelta di segregare verticalmente gli ambienti di lavoro e le rispettive connessioni alla *data warehouse* corrispondente tramite i *Gateway*. Questa misura di mitigazione è basata su una convenzione volta a impedire la presenza all'interno dello stesso *workspace* di connessioni tramite *Gateway* diversi che renderebbe quasi impossibile gestire una attività di *troubleshooting*. Si è deciso quindi che ogni *workspace* si avvalga di un unico *Gateway* per l'accesso ai dati presenti nella *data warehouse* di riferimento. Se da un lato questa decisione riduce la complessità di alcune attività, dall'altro pone il problema di come replicare dati, configurazioni e report, da un ambiente all'altro fino ad arrivare al *workspace* di produzione.

Fortunatamente il servizio di *Power BI Premium* propone una soluzione elegante anche a questo problema tramite la gestione delle *pipeline* di distribuzione [fig:7 e Fig:8]

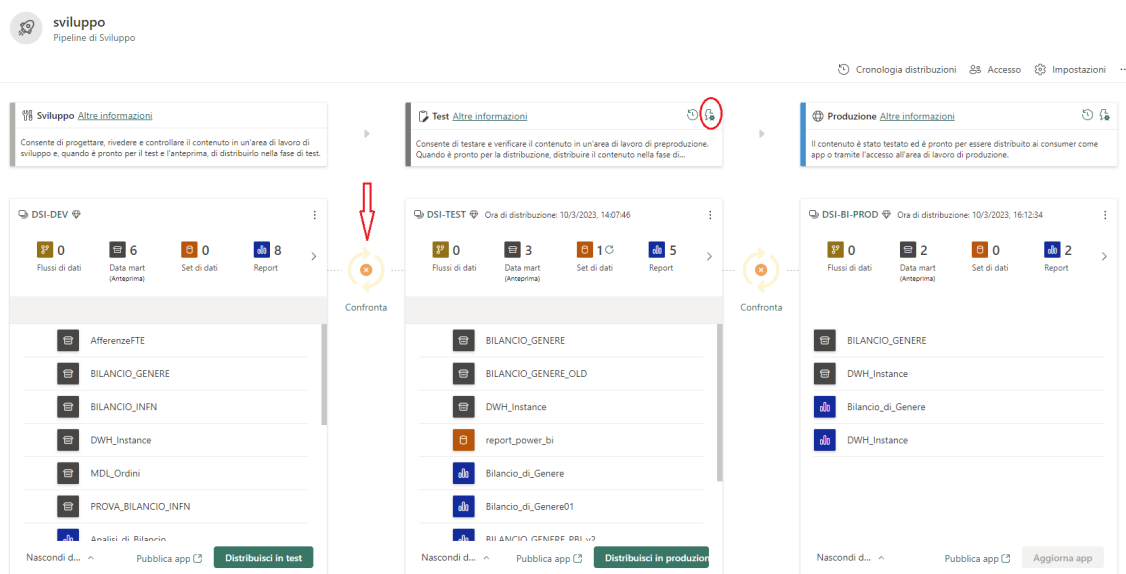


Figura 7: Esempio di gestione di una *pipeline* di distribuzione.

Nella figura [fig:7] viene messo in evidenza come la pipeline sia preimpostata su tre stadi ai quali sono associabili diversi *workspace premium*¹⁰. Al secondo e terzo stadio, che rappresentano il passaggio in test e quello in produzione, sono associabili delle "Automation" ovvero operazioni automatiche che sono indicate dal simbolo in alto a destra con il fulmine. Un esempio è la conversione automatica del parametro che identifica la Gateway usato dal dataset all'atto della promozione da un *workspace* al successivo nella pipeline. Un'altra importante funzionalità offerta da questo strumento è indicata dal pulsante "Confronta" tra uno stadio ed il successivo. Questa funzione, simile ad una sorta di *diff* tra *workspace*, mette in evidenza le differenze tra due *workspace*¹¹. Il risultato del confronto viene mostrato nella figura [fig:8].

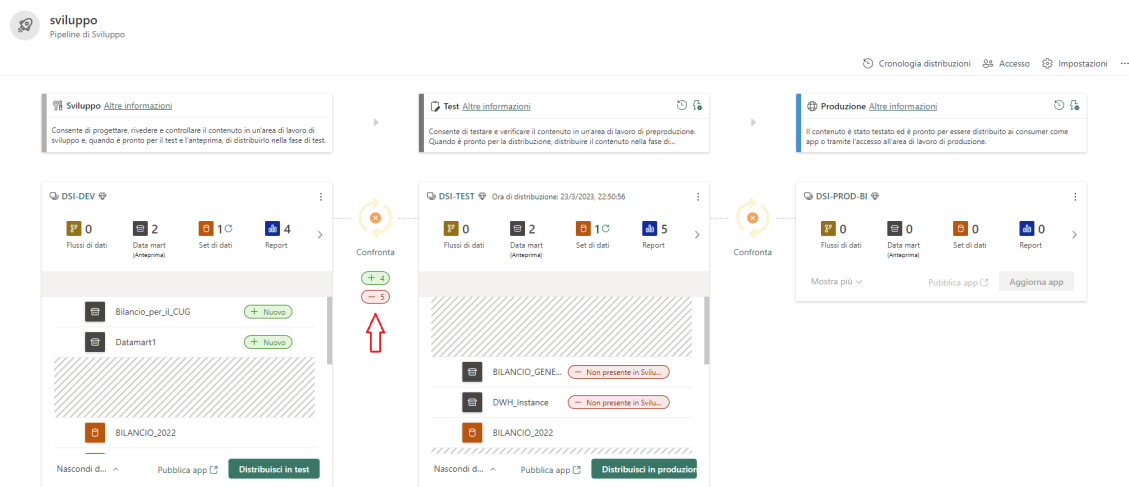


Figura 8: pipeline.

Oltre al dettaglio sintetico è possibile espandere le informazioni e vedere un dettaglio esteso sulla base del quale decidere quali risorse "promuovere" al *workspace* successivo della pipeline. L'associazione di un *workspace* ad un determinato stadio della pipeline non è immutabile, ma può essere riconfigurato all'occorrenza, tipicamente è possibile invertire un pezzo di pipeline per "ripartire" con gli sviluppi dall'ultima versione presente sull'ambiente di test o produzione implementando un meccanismo analogo a quanto attualmente fatto per l'aggiornamento periodico degli ambienti di test e sviluppo degli applicativi all'interno della DSI. La promozione da un ambiente al successivo avviene selezionando la risorsa e avviando il processo di distribuzione (pulsante "distribuisce"). Nel caso della promozione da sviluppo a test questo equivale a sottoporre agli esperti di dominio il dataset per la sua certificazione.

¹⁰In realtà è possibile configurare pipeline anche solo con due stadi legati al *workspace* sorgente e quello di destinazione.

¹¹Il confronto tra *workspace* sorgente e quello di destinazione è bidirezionale ed evidenzia quali risorse manchino in uno dei due e quali risultino più aggiornate.

3.3 Dependency Visualization & Impact Analysis

Dato che, come si è già detto, l'organizzazione del lavoro risponde a due diversi scenari, è necessario prestare particolare attenzione alle dipendenze che si instaurano tra report e dashboard ed i *dataset* a partire dai quali queste vengono sviluppate.

Anche in questo il servizio di *Power BI* facilita il lavoro tramite la visualizzazione delle dipendenze interne al *workspace*. Nella figura [fig:9] vengono mostrati alcuni esempi in cui si può vedere come diversi report possano in realtà attingere ai dati degli stessi *dataset*. Non è possibile creare report che attingano a più *dataset*, nonostante sia possibile creare dashboard che includano oggetti visivi mutuati da diversi report, che di conseguenza attingano dati da diversi *dataset*. In alternativa è possibile agire tramite l'applicazione *Power BI Desktop* per attingere da diversi "Hub Dati", *dataset* pubblicati online sui *workspace* della DSI, nella creazione di nuovi report; una volta che questi report vengono pubblicati un un'area personale o *Premium* costituiscono la creazione di un nuovo *dataset* che integra tutte le risorse dei *dataset* originali.

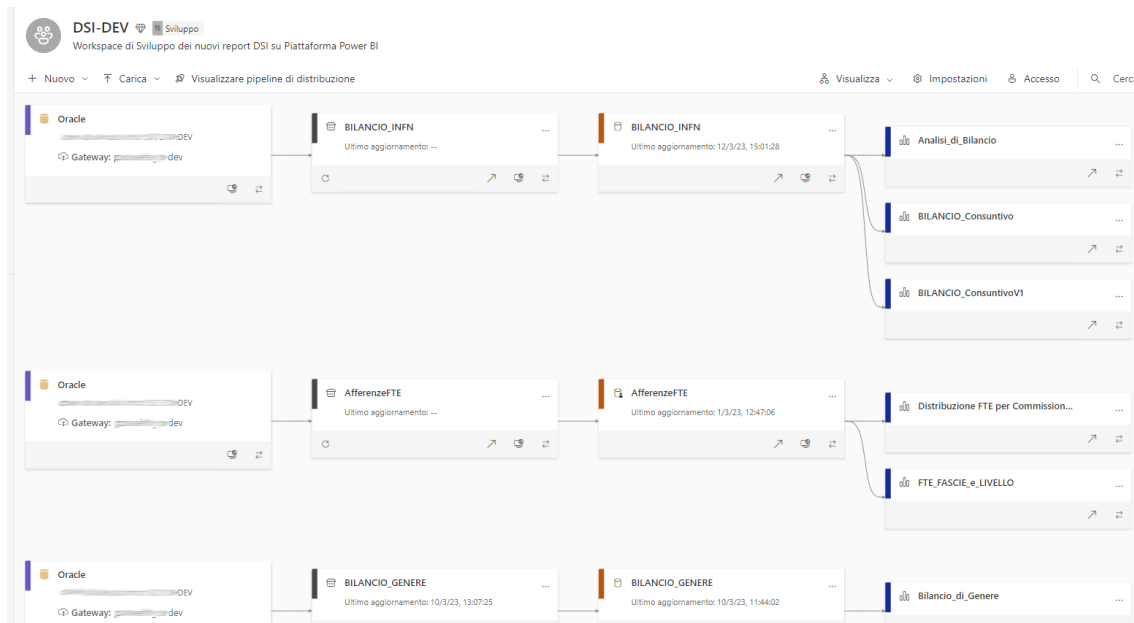


Figura 9: Visualizzazione per Derivazioni (dipendenza dei report dai *dataset* e dai *Gateway*).

Nella "Visualizzazione delle Derivazioni" vengono presentati diversi dati tra cui: la cronologia di aggiornamento dei *dataset*, i *Gateway* da cui dipendono per l'aggiornamento dei dati, le dipendenze, i link per l'apertura delle risorse direttamente con gli strumenti online e il link che avvia l'analisi di impatto che vediamo poi in [fig:10]. Nella figura [fig:9] invece si vede un altro componente che abbiamo testato. Questa tecnologia, ancora in beta, prende il nome di *Datamart*. La cosa interessante di questa soluzione è che offre la possibilità di svolgere tutti i passi di configurazione direttamente online senza passare dall'applicativo desktop e integra al suo interno una istanza di database SQL Server "privata" al *datamart*. Purtroppo al momento questa soluzione presenta ancora alcuni

problemi nella gestione degli aggiornamenti dei dati, quindi non verrà inclusa nel servizio INFN anche se terremo monitorato il suo sviluppo in attesa di un consolidamento.

Nella figura [fig:10], viene proposta la schermata di "analisi di impatto" che eventuali modifiche nei *dataset* avrebbero sui tutti gli elementi che ne fanno uso.

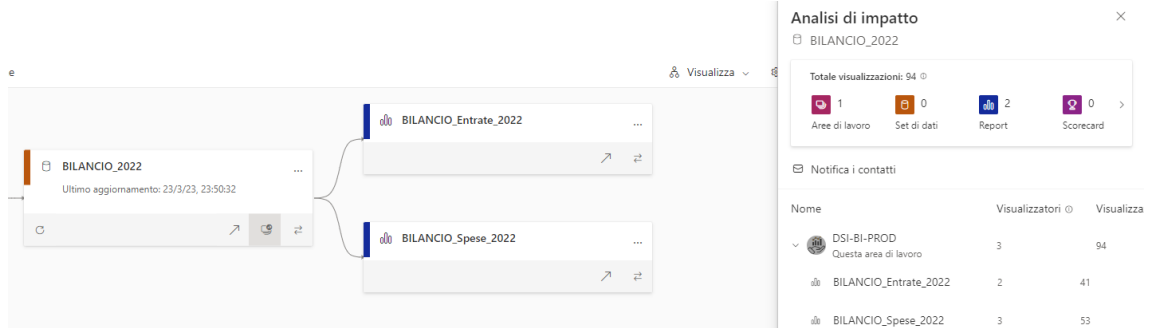


Figura 10: Esempio di resoconto di Impact Analysis relativo a un *dataset*.

In particolare nella figura si possono apprezzare una serie di informazioni utili come l'elenco dei *workspace* coinvolti, nel caso in cui siano stati prodotti report su altri *workspace* a partire dal *dataset* della DSI, quanti utenti abbiano grant di accesso alla risorsa (report o *dataset*) e quanti stiano realmente accedendo/visualizzando i dati oppure quante siano le dashboard coinvolte dalla modifica. Infine, se opportunamente configurati gli avvisi e le caselle di posta, è possibile mandare a tutti gli utenti coinvolti una notifica con un semplice testo che descriva le nuove funzionalità introdotte.

3.4 Distribuzione dei report tramite "App"

Nella figura [fig: 11] viene mostrata la bozza di un report¹² sviluppato in un' area di lavoro di test. Lo scopo di questa immagine è semplicemente mostrare come verrebbe visto il report da un utente "membro" dello stesso *workspace* in cui è stato sviluppato il report. Al report possono essere associati anche dei filtri con cui l'utente può variare alcuni parametri di presentazione dei dati. In alto viene visualizzata una barra degli strumenti le cui voci dipendono dai privilegi concessi all'utente all'interno del *workspace*. Quindi un *administrator*, un *contributor* e un semplice *member* avranno a disposizione funzionalità differenti come ad esempio la possibilità di salvarsi offline una copia del report, se amministratori del *workspace*¹³. E' anche possibile sottoscrivere gli aggiornamenti periodici per un report e riceverne una copia con scadenze definite direttamente dall'utente se le configurazioni del tenant lo consentono.

Un'altro elemento interessante di questa soluzione è la possibilità di visualizzare i report anche direttamente tramite Microsoft Teams, ponendo l'attenzione sul tema del lavoro collaborativo e sulla condivisione dei risultati durante *call*, un *meeting* e presentazioni ufficiali. Inoltre le funzionalità di esportazione intelligente dei report consentono di riceverne una versione in pdf o *powerpoint* formattate in maniera adeguata per una presentazione formale dei risultati.

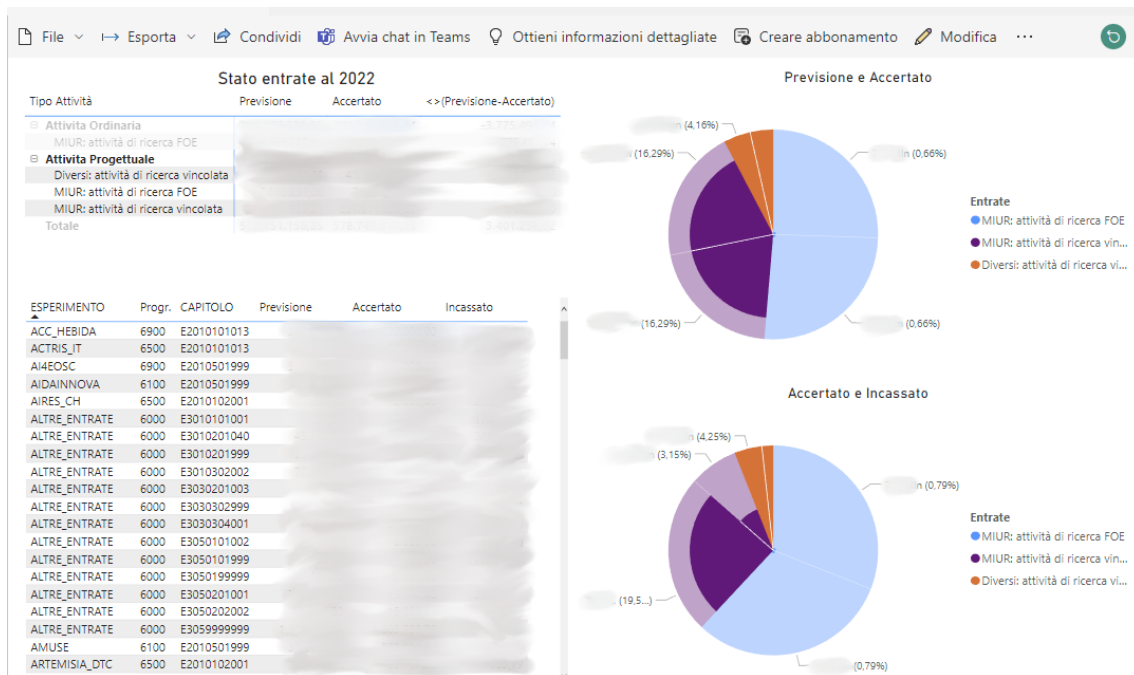


Figura 11: Esempio di Report visto da un Membro dello stesso *workspace*

Qualora invece fosse necessario distribuire il report ad un gruppo di utenti più vasto, sprovvisto di licenza PRO, e soprattutto non interessato a sviluppare la propria reportistica

¹²I dati di budget sono stati mascherati perché non significativi e non realistici al momento dello sviluppo del report in questione

¹³il formato di esportazione previsto è *.pbix usabile tramite Power BI Desktop

ma solo ad avere uno strumento di analisi dedicato, il metodo di distribuzione previsto è quello della creazione delle "App" (o Application).

Una App, come già accennato, è un modo per raccogliere sotto uno stesso portale (url), tutte le risorse necessarie alla visualizzazione di uno o più report. Questo può essere fatto a partire dalla pipeline di distribuzione del *workspace* di produzione selezionando tutti i *dataset*, i report e le dashboard che si vogliono includere nel portale che ne risulterà. Ogni *dataset* si porterà dietro le configurazioni per l'accesso tramite *Gateway*.

Per l'utente finale sarà come disporre di un portale dedicato alle sue esigenze di analisi, personalizzato in base al suo ruolo e profilo, tramite cui poter condurre le proprie analisi ed il monitoraggio delle metriche senza doversi preoccupare di altro.

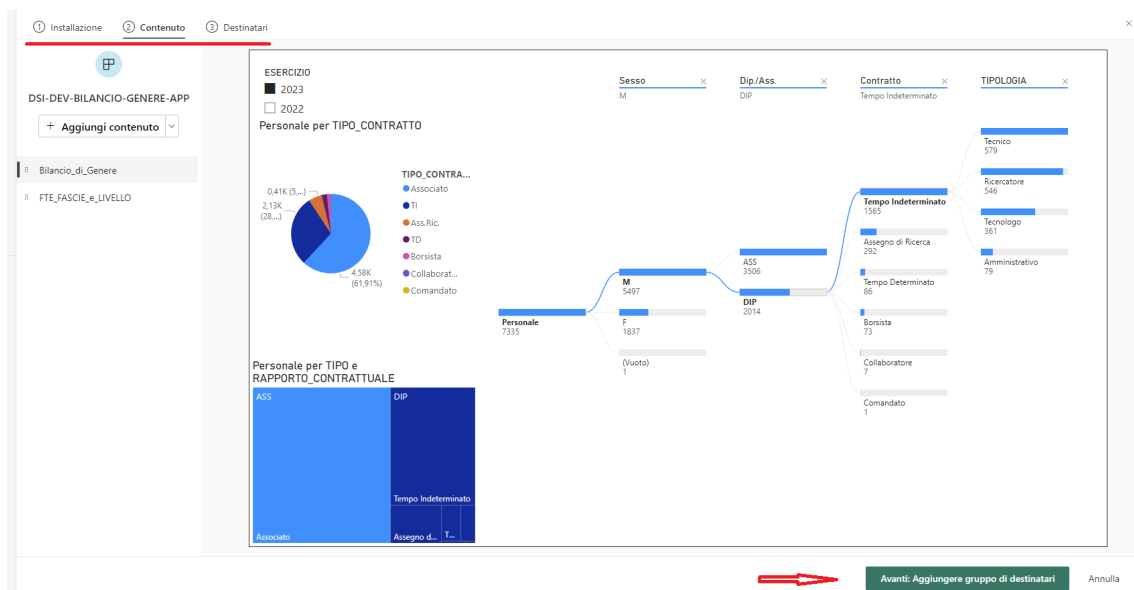


Figura 12: Creazione di un pacchetto App contenente dati e report da condividere con altri utenti

In [fig:12] viene mostrata la schermata principale del wizard di creazione di una *App Power BI*. In questa schermata viene definito il contenuto dell'App, in termini di *dataset* e visualizzazioni. E' anche possibile includere in maniera trasparente nella stessa App più report e dashboard legate a *dataset* diversi. Quindi risulta un modo efficace per proporre all'utente una visione integrata di tutti i report e i dati che possono interessargli e sono a sua disposizione. In una successiva pagina è possibile definire gli utenti o i gruppi che possono accedere alla risorsa, oltre che decidere se questi potranno modificare l'app a loro volta, condividerla, oppure salvare una copia dei dati di dettaglio per una esportazione. Al termine della configurazione e delle operazioni richieste, viene prospettato un link che sarà quello tramite cui l'utente finale andrà ad accedere all'App. Un link all'App sarà disponibile anche attraverso il *workspace* personale dell'utente sul tenant INFN.

Nella figura [fig:13] viene presentato l'aspetto finale di un portale così come viene visto da un utilizzatore finale dell'App.

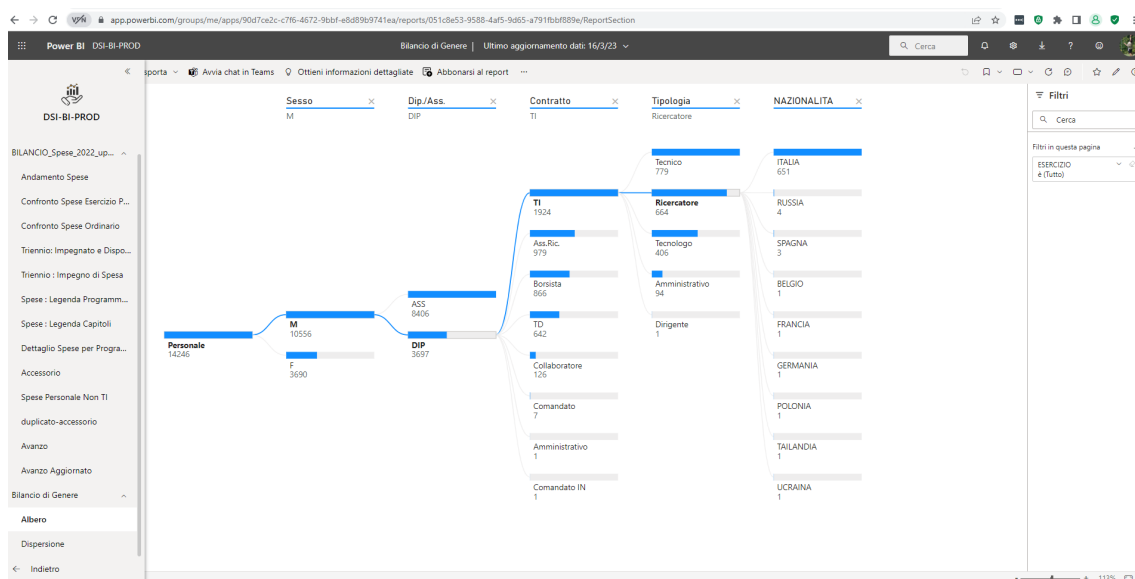


Figura 13: Visualizzazione finale del report tramite una App distribuita agli utenti *view-Only*

4 Conclusioni

Il lavoro di *scouting* tecnologico svolto tra la fine del 2021 e quella del 2022 ha evidenziato diversi punti di forza di una soluzione *On Cloud* rispetto alla gestione di una installazione *On Premises* su *hardware* proprietario. Tra questi vanno citate la conformità di questi servizi *cloud* con le normative GDPR e le certificazioni AGID per la Pubblica Amministrazione, garantita in gran parte dalla scelta della licenza *Premium per Capacity P1*, e l'aderenza di questa scelta tecnologica con la "*Strategia Cloud della PA*" [11] promossa dal 2019 in poi da AGID .

Come detto nell'introduzione questo elaborato si concentra esclusivamente su aspetti tecnici legati alle funzionalità offerte. Proprio da questo punto di vista, *Microsoft Power BI* è un servizio in grado di accogliere una gran varietà di esigenze ed indirizzare soluzioni per una gran varietà di scenari d'uso. Questo rappresenta un punto di forza tanto per gli utilizzatori finali, quanto per chi deve gestire il servizio o creare contenuti. Non stupisce quindi che Gartner ponga la suite di strumenti al più alto livello dei *leader* di mercato nel settore della *Business Analytics*.

Dal punto di vista INFN il primo elemento di preferenza rispetto ai *competitor* è l'elevata integrazione con tutti gli strumenti della suite Office365 che sono già largamente in uso all'INFN. E' infatti possibile visualizzare i report e le dashboard direttamente da Microsoft Teams durante una *call*, oppure è possibile esportare le informazioni in formato power point, già pronte e formattate per una presentazione. Inoltre nella versione *Power BI Desktop* è possibile elaborare direttamente da file excel i dati; anche se questo significa lavorare con dati *offline*, non aggiornati né certificati, ma resta comunque possibile condurre una veloce analisi parziale all'occorrenza.

Un secondo vantaggio, non meno importante, è la grande versatilità delle funzionalità offerte per la gestione delle aree di lavoro. La Visualizzazione dei flussi di derivazione¹⁴, l'analisi di impatto delle modifiche apportate ai *dataset*¹⁵ e soprattutto la gestione delle *pipeline* di distribuzione sono perfettamente in linea con le politiche di gestione degli sviluppi che la direzione si è data nel tempo.

In particolare la possibilità di diversificare i gruppi di utenti per i vari *workspace*, consente di implementare quel processo di certificazione dei *dataset* che passa per l'interazione con gli esperti di dominio.

Restano alcuni punti aperti nella sperimentazione delle funzionalità aggiuntive che riguardano l'uso delle *policy RLS* basate sul profilo degli utenti, l'uso delle *Sensitivity Labels* nei *dataset*¹⁶ e l'accesso a funzionalità di AI integrate nella piattaforma oppure raggiungibili tramite *endpoint online*.

¹⁴Rappresentano la possibilità di visualizzare in maniera veloce ed intuitiva le dipendenze tra *dataset* e report sviluppati

¹⁵ad esempio modificando un *dataset* su cosa si impatta o quali dashboard sono influenzate dalla modifica di un report

¹⁶Le *Sensitivity Labels* sono parte di un meccanismo di tagging dei dati a livello di colonna che impedisce che il dato a cui viene applicato il *tag* possa essere stampato, esportato o condiviso al di fuori dello scope del report.

Riferimenti bibliografici

- [1] ©Agenzia per l'Italia Digitale. Agenzia per l'italia digitale.
- [2] Stefano Rizzi Matteo Golfarelli. *Data™Warehouse: Teoria e pratica della progettazione*. In McGraw-Hill Education, editor, *book*, pages 1–448. McGraw-Hill Education (Italy) S.r.l, January 2006. Un manuale completo e aggiornato per la progettazione di un Data Warehouse.
- [3] ©2023 TIBCO. Tibco jaspersoft etl.
- [4] ©Agenzia per la Cybersicurezza Nazionale. Catalogo dei servizi cloud per la pa qualificati.
- [5] ©Agenzia per la Cybersicurezza Nazionale. Agenzia per la cybersicurezza nazionale.
- [6] Gartner Peerinsights. Gartner Magic Quadrant 2021 per analytics e business intelligence.
- [7] selecthub.com. Best business intelligence software tools.
- [8] ©2022 TIBCO Spotfire ®. System requirements for spotfire server version 12.3.
- [9] ©2022 Tableau Software®. Minimum installation hardware requirements.
- [10] ©Microsoft Power BI. White papers for power bi.
- [11] ©Agenzia per la Cybersicurezza Nazionale. La strategia cloud della pa.