

POLITECNICO DI TORINO

Corso di Laurea Magistrale in
Ingegneria della Produzione Industriale e dell'Innovazione Tecnologica

Tesi di Laurea Magistrale

**Data Governance e l'industria dei Microprocessori: Un Caso di Studio
su Processi, Competenze e Strumenti**



Relatori
Prof. Guido Perboli
Dott.ssa Chiara Vandoni

Gianluca Rosa-Brusin
s309176

A.A. 2023-2024

Ringraziamenti

È difficile credere che il mio percorso universitario sia giunto al termine. Questo elaborato segna l'inizio di un nuovo capitolo e sono entusiasta di poter applicare le conoscenze e le capacità acquisite per contribuire ai progetti che mi stanno a cuore e lasciare un impatto positivo sulle persone che incontrerò.

In queste poche righe desidero esprimere la mia sincera gratitudine a tutte le persone che hanno sostenuto il mio cammino accademico e arricchito profondamente questi anni.

Innanzitutto, vorrei ringraziare il Professor Guido Perboli e la Dottoressa Chiara Vandoni per il supporto ricevuto durante la stesura dell'elaborato.

A tutta la mia famiglia, che crede in me, sempre. La considerazione che avete di me mi rende orgoglioso. Sono fortunato ad avere così tanti buoni esempi a cui posso ispirarmi.

A Nicolò, Stefano e Denis, i migliori compagni di viaggio che avrei potuto incontrare.

Ad Albi, ovunque sarò, so che avrò sempre un fratello acquisito che mi aspetta a casa.

A tutta la GVNG, compagni di una vita. E soprattutto ad Alessandro, che quest'anno ci ha insegnato che finché abbiamo fiato possiamo combattere.

Ad Alice, sempre al mio fianco. La mia supporter numero uno. La tua ambizione e determinazione mi spinge a dare sempre il massimo.

A tutte le persone che mi sono state vicine, anche da lontano.

Grazie.

Indice

1.	INTRODUZIONE ALL'ELABORATO	7
2.	IL DATA MANAGEMENT	8
3.	I DATI.....	10
3.1.	LA RELAZIONE TRA DATI E SAPERE	10
3.2.	IL VOLUME DI DATI GLOBALE	10
3.3.	IL CICLO DI VITA DEI DATI	12
3.4.	TIPI DI DATI.....	13
3.4.1.	Dati strutturati e non strutturati	14
3.4.2.	Dati operativi e analitici	16
3.4.3.	I dati e le funzioni nei sistemi informativi.....	17
3.4.4.	Big Data	17
3.5.	I DATI ANALITICI: STRUTTURE E FIGURE PROFESSIONALI.....	18
3.5.1.	Le strutture dei dati analitici.....	18
4.	I PRINCIPI DEL DATA MANAGEMENT	21
5.	IL VALORE DEI DATI	22
5.1.1.	Condivisione dei dati e valore.....	23
6.	DATA GOVERNANCE	24
6.1.	LA DATA GOVERNANCE COME FUNZIONE STRATEGICA	24
6.1.1.	Riduzione dei rischi	24
6.1.2.	Miglioramento dei processi.....	27
6.2.	I METADATI E LE SUE STRUTTURE	28
6.2.1.	Business Glossary.....	29
6.2.2.	Data Dictionary.....	30
6.2.3.	Data Lineage.....	30
6.3.	IL FRAMEWORK DELLA DATA GOVERNANCE	30
6.4.	I MODELLI DELLA DATA GOVERNANCE	31
7.	IL DATA MESH.....	32
7.1.	I PRINCIPI DEL DATA MESH	32
7.1.1.	Domain Ownership.....	32
7.1.2.	Data as a product.....	33
7.1.3.	Self-serve data platform	33
7.1.4.	Federated computational governance	34
8.	L'INDUSTRIA DEI MICROPROCESSORI.....	35
8.1.	PREMESSA	35
8.2.	OVERVIEW DELL'INDUSTRIA	35
8.3.	I TRANSISTORS.....	36
8.4.	SUPPLY CHAIN E PROCESSO DI PRODUZIONE DEI CHIP	36

8.4.1. Ricerca e contesto normativo	36
8.4.2. Progettazione e sviluppo	37
8.4.3. Fabbricazione.....	37
8.4.4. Applicazione e distribuzione	38
8.5. IL MERCATO DEI MICROPROCESSORI	38
8.6. PESTEL ANALYSIS	39
8.6.1. Politico	39
8.6.2. Economico	40
8.6.3. Sociale	40
8.6.4. Tecnologico.....	40
8.6.5. Ambientale	41
8.7. L'IMPORTANZA DELLA DATA GOVERNANCE NELL'INDUSTRIA DEI MICROPROCESSORI	41
9. ESPERIENZA DI TIROCINIO – LA DATA GOVERNANCE E L'INDUSTRIA DEI MICROPROCESSORI	43
9.1. IL CONTESTO PRIMA DEL PROGETTO E LE NECESSITÀ DELL'AZIENDA	43
9.2. L'APPROCCIO DI ACCENTURE	44
9.2.1. Discovery	44
9.2.2. Design.....	45
9.2.3. Pilot.....	45
9.2.4. Plan.....	46
9.2.5. Scale.....	46
9.3. ROADMAP ATTUALE	47
9.4. IL FRAMEWORK DI DATA GOVERNANCE ADOTTATO	47
9.4.1. Stabilire la visione e la strategia	47
9.4.2. Organizzazione e Ruoli	48
9.4.3. Tecnologie	51
9.4.4. Processi	52
9.5. IL PROCESSO DI METADATA MANAGEMENT	54
9.5.1. Business glossary – creazione e manutenzione	55
9.5.2. Creazione del data dictionary, data lineage e loro manutenzione	58
9.6. CAPACITÀ DEL TOOL DI DATA CATALOG	60
9.7. LE METRICHE DEI PROCESSI	61
9.1. IL FUNZIONAMENTO DEL FRAMEWORK DI DATA GOVERNANCE	62
9.2. SALES & MARKETING USE CASE – METADATA MANAGEMENT	63
9.2.1. Definizione dei ruoli e avvio del test	64
9.2.2. Popolamento del Business Glossary	65
9.2.3. Popolamento del data dictionary	66
9.2.4. Risultati del test.....	67
9.2.5. Conclusioni del test	67
9.3. DATA MATURITY FRAMEWORK	69
9.3.1. I livelli di maturità.....	69
9.3.2. I principi guida	70
9.3.3. Criteri di valutazione.....	72

10. CONCLUSIONI	73
11. ALLEGATI.....	75
12. BIBLIOGRAFIA.....	113

Indice delle figure

Figura 1: Il data management e la data governance come funzioni strategiche verticali nelle aziende	75
Figura 2: La Piramide DIKW.....	75
Figura 3:La relazione ciclica tra dati, informazioni, conoscenza e saggezza	76
Figura 4: Statista, volume di dati creati e consumati globalmente all'anno in zettabytes dal 2010 al 2025.....	76
Figura 5: La relazione tra il ciclo di vita dei dati e il processo di data management	77
Figura 6: Rappresentazione semplificata di un database relazionale	78
Figura 7: Framework del NIST per la Cybersecurity	79
Figura 8:I principi del data mesh	80
Figura 9: Tipi di semiconduttori	80
Figura 10: Vendite globali per tipologia di semiconduttori nel 2021	80
Figura 11: Numero di transistors per microprocessore	81
Figura 12: Mercato dei semiconduttori per applicazione.....	81
Figura 13: Roadmap di alto livello per il seguimiento del progetto di data governance.....	82
Figura 14: Rappresentazione dell'andamento del progetto attraverso la roadmap di alto livello	83
Figura 15: Il framework di data governance adottato.....	84
Figura 16: rappresentazione grafica del modello Hub&Spoke	84
Figura 17: Processo di metadata management completo.....	88
Figura 18: Processo di metadata management relativo al business glossary	91
Figura 19: Processo di metadata management e data dictionary	92
Figura 20: Codice VBA per la creazione della macro	94
Figura 21: Processo di harvesting dei metadati.....	96
Figura 22: Rappresentazione grafica del metamodello.....	102
Figura 23: Richiesta di dati extra-dominio prima del framework di data governance	103

Figura 24: Rappresentazione della mancanza di ownership dei dati, catalogazione delle entità e dei dati dei domini.....	103
Figura 25: Matrice complessità valore per la scelta del test di caso d'uso	104
Figura 26: Struttura del business glossary per il dominio di Sales & Marketing	104
Figura 27: Numero di termini di business per entità del glossario S&M	105
Figura 28: Tipi e numero di entità con errori riscontrati.....	106
Figura 29: Numero e tipi di errori riscontrati	107
Figura 30: Entità e attributi obbligatori.....	107
Figura 31: Struttura del data dictionary	108
Figura 32: Metriche del test sul dominio di Sales& Marketing.....	108
Figura 33: Miglioramento del numero di termini di business	109
Figura 34: Numero di errori rilevati, prima e dopo	109
Figura 35: Termini di business con i campi obbligatori compilati	110
Figura 36: Risultati del test di Sales & Marketing	111
Figura 37: Valutazione della maturità aziendale rispetto al processo di metadata management .	112
Figura 38: Esempio di valutazione globale della data governance	113

Indice delle tabelle

Tabella 1: Le differenze tra dati strutturati e non strutturati.....	78
Tabella 2: Gli artefatti della data governance e un esempio	79
Tabella 3: Classificazione della sensibilità delle informazioni nell'organizzazione	85
Tabella 4: Attributi del business glossary	86
Tabella 5: Attributi del data dictionary	87
Tabella 6: Attributi del data lineage	87
Tabella 7: Spiegazione della matrice RACI	89
Tabella 8: Matrice RACI del processo di metadata management	89
Tabella 9: Template per la definizione dei requisiti del data consumer	93
Tabella 10: Capacità del tool di data catalog.....	97
Tabella 11: Capacità di management e matrice delle responsabilità	98

Tabella 12: Capacità di modifica e matrice delle responsabilità.....	98
Tabella 13: Capacità di visualizzazione e matrice delle responsabilità.....	99
Tabella 14: Capacità di visualizzazione e matrice delle responsabilità.....	99
Tabella 15: Metriche dei processi di data governance.....	100
Tabella 16: Definizione e spiegazione degli errori identificati tra i termini di busines	105

1. INTRODUZIONE ALL'ELABORATO

Questo elaborato si propone di analizzare l'importanza strategica dell'applicazione di processi della data governance nelle aziende di oggi. Attraverso l'esperienza di tirocinio maturata tra il marzo e il giugno del 2024, si vuole contestualizzare la definizione del framework della data governance in un'azienda produttrice di microprocessori, approfondendo vari aspetti della gestione e del governo dei dati. L'obiettivo è integrare le conoscenze teoriche derivate dalla letteratura con l'esperienza pratica di tirocinio svolta presso Accenture spa, società globale di consulenza, tecnologia e operazioni, nell'ambito di un progetto pluriennale atto a supportare la definizione della strategia di governo dei dati in un'azienda leader del settore di circuiti integrati.

La necessità di digitalizzare i processi aziendali per mantenere una posizione competitiva sul mercato ha portato le aziende a dover gestire una mole sempre più grande di dati. Inoltre, i progressi delle tecnologie digitali stesse, stanno cambiando il modo in cui le aziende competono, come definiscono e forniscono valore. Questi progressi non sono mai stati così critici, né così necessari.

L'elaborato si propone quindi di approfondire i temi relativi al Data Management, esplorando l'importanza cruciale di una corretta gestione dei dati nell'era moderna e considerando questi ultimi come asset strategici per le organizzazioni. Al fine di comprendere le soluzioni e i processi definiti nell'ambito della data governance, è necessario fornire una dettagliata classificazione dei dati, distinguendo tra dati strutturati, non strutturati, operativi e analitici, e descrivendo il loro ciclo di vita. Questa sezione mette in luce i principi fondamentali della gestione dei dati, necessari per la loro valorizzazione e sicurezza.

Di seguito, si cerca di portare chiarezza sul valore dei dati e le difficoltà nel valutarlo correttamente, evidenziando come la loro condivisione e il loro utilizzo mirato possano incrementare significativamente il valore aziendale. Inoltre, un'ampia sezione è dedicata alla Data Governance, con un particolare focus sull'adattamento delle teorie alla pratica aziendale.

Al fine di contestualizzare il settore dei microprocessori, le sue sfide, e come la data governance può migliorare i processi aziendali, nonché migliorare la privacy e la sicurezza delle informazioni trattate, è fornita una panoramica dettagliata dell'industria dei microprocessori, analizzando in particolare come viene gestita la catena del valore.

Un'importante parte dell'elaborato è dedicata all'analisi dell'esperienza di tirocinio. Durante questa esperienza, i framework e i processi definiti dalla letteratura sono stati adattati alle esigenze specifiche dell'azienda, basandosi su modelli consolidati di data governance. I framework definiti da un punto di vista strategico non sono tutt'ora implementati all'interno dell'organizzazione, ma è stato identificato un primo caso d'uso per testare i processi, i ruoli, e le nuove responsabilità definite.

L'obiettivo principale di questo elaborato è dimostrare come l'implementazione dei processi della data governance possa migliorare significativamente le operazioni aziendali, riducendo i rischi, ottimizzando i processi e incrementando il valore strategico dei dati. Attraverso un'analisi dettagliata dei framework teorici e della loro applicazione pratica durante il tirocinio, si intende fornire un modello replicabile per altre aziende del settore.

In sintesi, l'elaborato offre una panoramica completa e integrata della teoria e della pratica della data governance. È importante menzionare che i temi sviluppati nell'attività di tirocinio hanno una connotazione prettamente strategica, in quanto la fase di implementazione delle strategie definite avverrà in un secondo momento.

2. IL DATA MANAGEMENT

Nell'epoca dell'Industria 4.0 (ormai 5.0¹) e dell'innovazione tecnologica, la produzione di dati ha raggiunto livelli senza precedenti, offrendo alle organizzazioni opportunità per l'ottimizzazione delle operazioni aziendali e la creazione di nuovi modelli di business nel passato inesistenti. Queste però sono accompagnate da un'urgente necessità: gestire propriamente i dati raccolti per estrarre ricchezza informativa in modo efficace ed efficiente.

I dati non sono più semplici aggregati di informazioni, ma si configurano come preziosi asset aziendali, equiparabili agli asset fisici. Tuttavia, quantificare ed estrarre valore da questa vasta riserva informativa richiede molto più che un'osservazione superficiale. A differenza degli asset fisici, infatti, il processo di estrazione del valore dei dati è un processo complesso con poca visibilità nel breve periodo, che richiede un impegno prolungato e una visione strategica da parte della leadership aziendale. L'accurata pianificazione, la corretta coordinazione, e il controllo diventano quindi elementi imprescindibili per convertire i dati in un vantaggio competitivo tangibile.

Questo processo appena definito è il Data Management.

Il DAMA² (DAMA International, 2017) descrive il Data Management come "l'insieme delle fasi di sviluppo, esecuzione e supervisione di progetti, politiche, programmi e pratiche che gestiscono, controllano, proteggono e migliorano il valore dei dati e degli asset informativi durante tutto il loro ciclo di vita." Le fasi di gestione e controllo, che comprendono tutte le politiche, i progetti e i programmi, rappresentano un approccio olistico verso il trattamento dei dati, focalizzato sulla loro direzione strategica per supportare gli obiettivi aziendali e migliorare l'efficacia della loro amministrazione. I dati rappresentano infatti un asset estremamente delicato per le aziende, in quanto spesso trattano informazioni sensibili, segreti industriali e dati confidenziali. La perdita, il furto o l'utilizzo improprio dei dati da parte delle organizzazioni possono avere conseguenze devastanti per le stesse, inclusi danni alla reputazione, perdite finanziarie e problemi legali. Inoltre, la legislazione sulla protezione dei dati, come il GDPR³, impone agli organismi di trattare i dati personali in modo trasparente e sicuro.

Pertanto, è fondamentale assicurare un efficace governo dei dati, per garantire che il loro trattamento avvenga seguendo regole di sicurezza, integrità e conformità, proteggendo così l'azienda da potenziali rischi e minacce. Allo stesso momento è necessario stabilire e comunicare chiaramente le aspettative dell'organizzazione nei confronti dei comportamenti delle persone e dei processi in relazione alla gestione dei dati.

Questo processo continuo è la Data Governance.

Il Data Management, così come la Data Governance, deve quindi essere considerata una funzione verticale (allegato 1) nelle organizzazioni, che implica la partecipazione di tutte le funzioni aziendali, coinvolgendo professionisti sia con competenze tecniche che strategiche di business, con il comune obiettivo di assicurare un'elevata qualità dei dati dell'organizzazione per raggiungere gli obiettivi strategici.

¹ L'industria 5.0 complementa l'attuale approccio dell'Industria 4.0 mettendo specificamente la ricerca e l'innovazione al servizio della transizione verso un'industria sostenibile, incentrata sull'uomo e resiliente (Commissione Europea, 2023)

² Data Management Body of Knowledge

³ The General Data Protection Regulation

I dati sono infatti la chiave di volta per le organizzazioni che intraprendono un percorso di trasformazione digitale. Senza una solida base di dati, le organizzazioni operano essenzialmente al buio, incapaci di valutare l'efficacia delle loro iniziative. Ancora più importante, i dati giocano un ruolo fondamentale nell'assicurare il consenso dei principali stakeholder e dei responsabili delle decisioni (Forbes, 2019). Le organizzazioni che dispongono di dati affidabili e di alta qualità sui loro clienti, prodotti, servizi e operazioni possono prendere decisioni migliori rispetto a quelle senza dati o con dati inaffidabili.

Il principale driver del data management è infatti consentire alle organizzazioni di ottenere valore dai loro beni informativi, proprio come la gestione efficace dei beni finanziari e fisici consente alle organizzazioni di ottenere valore da tali beni. I dati costituiscono il patrimonio strategico dell'organizzazione poiché attraverso la loro elaborazione è possibile rappresentare il passato dell'organizzazione, fornendo insight sulle performance passate e presenti, e offrire indicazioni preziose sui futuri trend e sviluppi, guidando così le decisioni e le strategie aziendali.

In un mondo sempre più digitalizzato, i dati sono ora onnipresenti in tutti i settori e le industrie, riconosciuti come una fonte e chiave per il consolidamento di vantaggi competitivi. Questo significa che i processi di Data Management e di Data Governance, purché in misure diverse, sono applicabili in tutti i settori economici.

Nel settore secondario questo fenomeno ha un'elevata visibilità: in qualsiasi processo di produzione industriale viene generato continuamente una grande quantità di dati, ad esempio, dalle linee di produzione. Una corretta gestione di questi dati permette di ottimizzare i processi identificando le aree problematiche, i colli di bottiglia e i passaggi inutili. Il settore terziario, invece, è totalmente dipendente da un intenso utilizzo dei dati: per qualsiasi azienda che fornisce servizi, tra cui banche, assicurazioni o società di investimento, l'analisi dei dati è un'attività essenziale. Grazie ai clienti, alle transazioni, ai conti, agli investimenti, ai prestiti e ad altre fonti, queste aziende producono volumi di dati enormi. In particolari industrie appartenenti al settore terziario, inoltre, normative come il GDPR sono particolarmente severe e stringenti.

Prima di identificare quindi i principi su cui si basa il framework del Data Management, è necessario approfondire la natura dei dati, la loro struttura, come questi vengono creati, immagazzinati, elaborati e di seguito utilizzati per estrarre del valore utile a prendere delle decisioni informate e consapevoli.

3. I DATI

3.1. LA RELAZIONE TRA DATI E SAPERE

Nella scienza dell'informazione i concetti di dati, informazione e conoscenza sono strettamente collegati. Un metodo grafico per rappresentare questa connessione è la piramide DIKW⁴ (Allegato 2, ResearchGate, 2016).

I dati sono una forma di rappresentare e interpretare eventi e fatti riguardanti il mondo, fornendo conoscenza e sapere, se utilizzati nel modo opportuno. Infatti, preso singolarmente, un dato non ha nessun tipo di impatto o utilizzo, ma è necessario prima fornirgli un contesto, dandogli significato. Solo quando i dati vengono organizzati e interpretati in un contesto specifico diventano informativi e utili per coloro che li utilizzano (Data Versity, 2024).

L'informazione è un insieme di dati contestualizzati che hanno una rilevanza per una o più persone in un determinato periodo di tempo. I dati grezzi sono trasformati in una forma comprensibile per il destinatario. Quando le informazioni sono conservate, per approfondire la loro comprensione attraverso l'esperienza e processi di comprensione o di associazione, è possibile generare conoscenza.

Per ultimo, la saggezza si manifesta nell'abilità di applicare la conoscenza, considerando il contesto e le conseguenze a lungo termine delle azioni intraprese.

È necessario però notare un punto importante. I dati non esistono di per sé. I dati devono essere creati per poterli elaborare ed estrarne valore. Questo processo di creazione dei dati, che avviene attraverso l'analisi o l'osservazione di processi ed eventi, richiede una profonda conoscenza.

Pertanto, per comprendere e raccogliere efficacemente i dati, è indispensabile possedere una solida comprensione dei processi osservati. Di conseguenza, quella che viene spesso rappresentata come una piramide, con i dati alla base e la conoscenza al vertice, dovrebbe in realtà essere concepita come un cerchio, in cui conoscenza e saggezza alimentano continuamente la produzione di dati (Allegato 3).

Prima di approfondire la natura dei dati, le sue caratteristiche e come viene gestito il ciclo della vita di un dato, è necessario comprendere il volume dei dati creati oggi, per rimarcare l'importanza e la magnitudine della presenza dei dati in tutte le organizzazioni del mondo e nelle vite dei singoli individui.

3.2. IL VOLUME DI DATI GLOBALE

Le aziende e organizzazioni hanno sempre gestito i propri dati, ma gli sviluppi tecnologici hanno portato alla loro continua e massiccia produzione, ampliando il loro significato e la loro gestione. Nel 1975, gli asset tangibili costituivano l'83% della valutazione di un'azienda; oggi, invece, il 90% è costituito da asset intangibili tra cui dati, proprietà intellettuale, marchio e reputazione (Anmut, 2019).

Dalla scoperta del transistor nel 1947 e del microchip integrato nel 1956 (World Economic Forum, 2021), la nostra società ha subito un rapido avanzamento tecnologico. In poco più di 50 anni,

⁴ Data, Information, Knowledge, Wisdom

abbiamo raggiunto una potenza di calcolo senza precedenti, che ha permesso di aumentare esponenzialmente la velocità con cui sviluppiamo nuove tecnologie, fino al raggiungimento di pietre miliari come l'intelligenza artificiale, la realtà aumentata e i veicoli autonomi.

Oggi giorno l'ubiquità degli strumenti digitali e la loro inclusione in ogni aspetto della vita umana sono i principali fattori che contribuiscono alla creazione continua e costante di dati di un volume senza precedenti, poiché ogni interazione, transizione o esperienza si traduce in un'impronta digitale, contribuendo così a un'enorme mole di dati in costante crescita.

L'allegato 4 (Statista, 2023) mostra la crescita dei dati globali creati, consumati e archiviati nel periodo compreso tra il 2010 ad oggi, comprendendo inoltre la previsione di crescita per l'anno 2025 in zettabytes. Nel 2024 sono stati consumati circa 150 zettabytes.

Per rendere più chiara la magnitudine di grandezza di un singolo zettabyte, si può effettuare un esempio: un video in alta definizione solitamente consuma circa 20 gigabyte (GB) di spazio di archiviazione per ogni ora di riproduzione. Uno zettabyte può quindi contenere all'incirca 50 miliardi di ore di video in HD, ovvero quasi 6000 anni di video in alta definizione. (allegato 5).

Nonostante la mole gigantesca di dati prodotti annualmente, solo circa il 2% dei dati prodotti e consumati nel 2020 è stato salvato e conservato nel 2021 (Statista 2023). Questo è un primo indicatore dell'importanza del Data Management. Ogni anno, infatti, il 98% circa dei dati che vengono prodotti globalmente sono persi per diversi motivi:

1. **Mancanza di archiviazione adeguata:** molte organizzazioni non dispongono di sistemi di archiviazione sufficientemente robusti per conservare grandi quantità di dati a lungo termine;
2. **Scarsa gestione dei backup:** in assenza di backup⁵ regolari dei dati per garantire la loro protezione, questi sono vulnerabili a eventi come malfunzionamenti hardware, errori umani o attacchi informatici, che possono causare la perdita permanente dei dati;
3. **Obsolescenza tecnologica:** i formati dei file e i sistemi di archiviazione possono diventare rapidamente obsoleti, rendendo difficile o impossibile l'accesso ai dati nel tempo;
4. **Cancellazione accidentale o deliberata:** errori umani o decisioni aziendali possono portare alla cancellazione accidentale o deliberata di dati, soprattutto se non vi è una corretta politica di gestione dei dati in atto;
5. **Dati non strutturati o non classificati:** la mancanza di una struttura organizzativa per i dati può rendere difficile recuperarli e conservarli in modo efficace nel lungo periodo.

Inoltre, è importante considerare che non tutti i dati meritano di essere archiviati e, di conseguenza, vengono eliminati. Allo stesso modo, è necessario garantire che i dati archiviati siano certificati, ovvero che rispettino certi requisiti di qualità e sicurezza imposti dall'azienda stessa, per validare la loro coerenza e accuratezza.

In linea con la forte crescita del volume di dati, si prevede che la base installata di capacità di archiviazione aumenti, con un tasso di crescita annuale composto del 19,2% nel periodo previsto dal 2020 al 2025 (Statista, 2023). Nel 2020, la base installata di capacità di archiviazione ha raggiunto i 6,7 zettabyte e si prevede che cresca fino a 16 zettabytes entro il 2025. Ad ogni modo è evidente come la capacità di archiviazione aumenti ma ad un tasso di crescita non sufficiente per stare al passo con la crescita dei dati.

⁵ Duplicazione di un file o di un insieme di dati su un supporto esterno al computer, per avere una copia di riserva.

3.3. IL CICLO DI VITA DEI DATI

Come tutti gli altri asset aziendali, anche i dati hanno un ciclo di vita che le organizzazioni devono pianificare e organizzare strategicamente per il raggiungimento degli obiettivi relativi al Data Management (IBM, 2024).

Le fasi del ciclo di vita dei dati tradizionalmente identificate sono le seguenti (Allegato 6):

- **Generazione/Raccolta:** perché il ciclo di vita dei dati abbia inizio, prima di tutto i dati devono essere generati. Alcuni dati sono generati dall'organizzazione stessa, altri dai clienti o fornitori e alcuni da terze parti ancora. Ogni vendita, acquisto, assunzione, comunicazione, interazione, tutto genera dati. Sebbene i dati possano essere generati in vari modi, la raccolta di tutti i dati disponibili non è necessaria per il successo dell'azienda, ma spetta ai team di Data Management identificare quali informazioni debbano essere catturate e i migliori mezzi per farlo, sulla base della qualità e rilevanza dei dati;
- **Storage/Archiviazione:** dopo che i dati sono stati raccolti ed elaborati, devono essere archiviati per un uso futuro. Questo viene comunemente realizzato attraverso la creazione di database o set di dati. Questi set di dati possono quindi essere memorizzati nel cloud, su server o utilizzando un'altra forma di archiviazione fisica come un hard disk, CD, cassetta o floppy disk⁶. Il metodo di archiviazione apre diversi temi che verranno identificati in seguito come la differenza tra dati strutturati e non strutturati, l'infrastruttura utilizzata per la loro archiviazione sulla base del grado di sicurezza necessario eventualmente anche per soddisfare i requisiti di privacy quelli delle politiche governative. Inoltre, è importante evidenziare l'importanza della ridondanza dei dati e come gestirla;
- **Elaborazione:** una volta raccolti, i dati spesso devono essere elaborati. L'elaborazione dei dati può fare riferimento a varie attività, tra cui uno studio approfondito mirato all'identificazione e correzione di errori quali dati mancanti, duplicati o inaccurati, oppure la loro trasformazione attraverso l'aggregazione di grandi volumi di informazioni. Un insieme di dati può essere anche arricchito, aggregando informazioni da fonti esterne. In alcuni casi alcuni processi di elaborazione possono avvenire in concomitanza con la fase di archiviazione. Ad esempio, attraverso la crittografia dei dati, un processo mediante il quale i dati vengono trasformati in una forma illeggibile e incomprensibile, chiamata testo cifrato, si utilizzano algoritmi matematici complessi per proteggere i dati da accessi non autorizzati;
- **Utilizzo e analisi:** durante la fase di utilizzo dei dati, i dati vengono sfruttati per svolgere attività specifiche o per supportare processi decisionali all'interno dell'organizzazione. Questo utilizzo può comprendere l'applicazione di modelli di machine learning per fare previsioni, l'esecuzione di query su database per ottenere informazioni rilevanti, la generazione di report per monitorare le prestazioni aziendali o l'utilizzo dei dati per alimentare applicazioni e servizi. L'integrazione con la fase di analisi dei dati avviene quando i dati utilizzati vengono esaminati e valutati criticamente per identificare pattern, tendenze o relazioni significative. L'analisi dei dati può fornire un quadro più approfondito delle informazioni contenute nei dati utilizzati durante la fase di utilizzo. Ad esempio, l'analisi dei dati può rivelare correlazioni tra variabili o fornire insight su comportamenti dei clienti che possono guidare decisioni aziendali;
- **Visualizzazione e Condivisione:** la visualizzazione dei dati si riferisce al processo di creazione di rappresentazioni grafiche delle tue informazioni, tipicamente attraverso l'uso di

⁶ È paradossale, ma per alcuni tipi di dati estremamente sensibili, i sistemi di archiviazione non connessi a Internet, come la carta, risultano essere i più sicuri.

uno o più strumenti di visualizzazione. La visualizzazione dei dati rende più facile comunicare rapidamente la tua analisi a un pubblico più ampio sia all'interno che all'esterno della organizzazione. La condivisione dei dati all'interno e all'esterno dell'organizzazione permette il trasferimento di valore dei dati;

- **Eliminazione dei dati:** in questa fase finale del ciclo di vita, i dati vengono eliminati dai registri e distrutti in modo sicuro. Le aziende eliminano i dati non più necessari per creare più spazio di archiviazione per i dati attivi. In questa fase, i dati vengono rimossi dagli archivi quando superano il periodo di conservazione richiesto o non hanno più uno scopo significativo per l'organizzazione;

Oltre al ciclo di vita dei dati, esistono altre due fasi cruciali che comprendono la gestione dei dati in senso più ampio:

- **Pianificazione:** prima di intraprendere qualsiasi azione nel ciclo di vita dei dati, è essenziale pianificare attentamente. Durante questa fase, vengono definiti obiettivi chiari e specifici per la gestione dei dati, identificando quali dati saranno raccolti, come verranno trattati e gestiti, e quali procedure saranno implementate per garantire la sicurezza e la conformità normativa. La pianificazione include anche la definizione delle politiche e delle procedure aziendali per la gestione dei dati, nonché la valutazione delle risorse necessarie per implementare con successo il piano;
- **Disegno e Implementazione:** una volta pianificato il processo di gestione dei dati, è necessario tradurre il piano in azione attraverso il disegno e l'implementazione di una struttura aziendale adeguata. Durante questa fase, vengono definite e implementate le infrastrutture tecnologiche e organizzative necessarie per la raccolta, l'elaborazione, l'archiviazione e l'analisi dei dati. Ciò include la progettazione di database, la configurazione dei sistemi di archiviazione, l'implementazione di procedure di sicurezza dei dati e la formazione del personale coinvolto nel processo di gestione dei dati. Inoltre, viene definita la struttura dei metadati per descrivere e categorizzare i dati, garantendo una gestione efficace e una facile ricerca e recupero dei dati quando necessario.

Al fine di comprendere meglio il ciclo di vita di un dato e il modo in cui è possibile estrarre valore da esso, è necessario analizzare i tipi di dati.

3.4. TIPI DI DATI

Nella letteratura accademica e professionale, i dati aziendali sono spesso suddivisi in diverse categorie, assumendo connotazioni differenti per rispondere a specifiche esigenze.

Alla base di questa suddivisione c'è la differenza tra i dati strutturati e i dati non strutturati. Queste due macrocategorie permettono di comprendere diverse caratteristiche dei dati analizzati, tra cui il formato, la loro forma di archiviazione e il miglior metodo per la loro analisi. Inoltre, è possibile identificare i dati operativi e analitici, che distinguono tra dati utilizzati per le operazioni quotidiane e dati usati per l'analisi strategica e decisionale.

Inoltre, è necessario approfondire il concetto di master data, reference data e il fondamentale ruolo dei metadati, ovvero quella categoria di dati che permette la classificazione di altri dati, in funzione del contesto d'uso all'interno dei sistemi informativi.

Per ultimo è importante menzionare i Big Data, che non identificano nello specifico una categoria di dati, bensì un agglomerato di dati, spesso di così elevate dimensioni e complessità, tale che diventa difficile gestirli con i tradizionali metodi di gestione dei dati.

3.4.1. DATI STRUTTURATI E NON STRUTTURATI

I dati strutturati e i dati non strutturati rappresentano due ampie categorie di dati che possono essere raccolti. La tabella 1 nell'allegato 7 riassume le principali differenze tra queste due categorie.

DATI STRUTTURATI

I dati strutturati, comprendono tipi di dati discreti quali numeri, testo breve e date. Sono dati che si adattano perfettamente alle strutture tabellari. Ad esempio, i dati che possono essere presentati in un file Excel o nei Fogli Google sono un esempio di dati strutturati.

La forma tabellare permette di organizzare i dati in righe e colonne. Ogni riga rappresenta il contenuto della singola istanza, mentre ogni colonna rappresenta un attributo diverso riferito alle diverse istanze (IBM, 2021).

Dato che ogni tabella generalmente contiene un solo tipo di istanze, può essere necessario alle volte collegare diverse tabelle, ovvero si possono considerare correlate in base ad una o più colonne comuni presenti in entrambe le tabelle. Nel caso in cui più tabelle vengano correlate, è possibile definire un database relazionale.

Nell'immagine nell'Allegato 8 la struttura semplificata di un database relazionale. I dati organizzati in tabelle sono formati dalle righe (le tuple) e colonne (attributi). Ogni tabella è caratterizzata da una chiave primaria che identifica univocamente il record, composta da una o più colonne. Per ultimo, le chiavi esterne sono utilizzate per creare relazioni fra le tabelle: la chiave esterna di una tabella, infatti, fa riferimento alla chiave primaria di un'altra tabella.

I database relazionali possono essere scritti, letti e manipolati utilizzando SQL⁷, un linguaggio sviluppato da IBM negli anni '70 (IBM, 2021) per supportare i suoi database mainframe. SQL è un linguaggio semplice, progettato per essere simile all'inglese scritto, rendendolo più intuitivo per gli utenti.

I database relazionali sono il tipo di database più comunemente utilizzato oggi. Sono progettati per gestire e archiviare un elevato volume di dati e recuperarli in modo logico e sistematico. La struttura logica che permette la rappresentazione visiva delle strutture di dati e per modellare è organizzare i dati in un database relazionale è il diagramma entità relazioni (esempio in allegato 8). Grazie alla loro struttura e facilità di interpretazione, la maggior parte dei processi di data governance che saranno analizzati in seguito, è applicata proprio a questo tipo di dati.

DATI NON STRUTTURATI

I dati non strutturati, che non seguono un formato predefinito o una struttura organizzata, non possono essere elaborati e analizzati con strumenti e metodi convenzionali. I dati non strutturati sono i dati moderni, come i testi, i video, gli immagini e gli audio, i dati dei sensori⁸ o ancora i dati

⁷ Structured Query Language

⁸ Qualsiasi dispositivo che rileva e risponde a qualche tipo di input dall'ambiente fisico, come la luce, il calore, il movimento, l'umidità, la pressione.

web. Questi sono caratterizzati dalla loro origine digitale e interoperabilità, dalla rapida creazione e continuo movimento, dalla loro multimodalità e interoperabilità e distribuzione spaziale.

A causa della loro natura, è infatti impossibile organizzare questo tipo di dati nelle righe e colonne di una tabella. In particolare, la caratteristica della multimodalità, ovvero che questi dati possono essere presentati in diversi formati e combinare diverse modalità di comunicazione, rende la loro organizzazione in strutture fisse pressoché impossibile.

Per questi motivi i dati non strutturati sono organizzati nei database non relazionali, molto flessibili e scalabili, che permettono la gestione di grandi volumi di dati eterogenei.

Nonostante i database non relazionali non organizzino i dati sottostanti in righe e colonne, hanno comunque delle modalità di archiviazione per permettere l'organizzazione dei dati che contengono. Tra queste è possibile menzionare i database NoSQL, progettati per gestire dati non strutturati, come documenti JSON (Ecma International, 2021), BSON (Mongo DB, n.d.), XML (w3Schols, n.d.). Ogni documento può contenere attributi e valori, e i documenti possono essere organizzati in collezioni o raggruppamenti logici. Inoltre, strutture come i Data Lakes o sistemi di archiviazione su Cloud, consentono di archiviare dati non strutturati in modo scalabile e affidabile, offrendo funzionalità per il backup, il versionamento e l'accesso sicuro ai dati.

Il ponte tra i dati strutturati e non strutturati sono i dati semi-strutturati. Hanno alcune caratteristiche strutturate, ma non seguono uno schema rigido come i dati strutturati. Utilizzano un formato che permette di avere una struttura flessibile.

DIFFERENZE, VANTAGGI E SVANTAGGI

È importante per le organizzazioni comprendere la differenza tra i dati strutturati e non strutturati e il modo di organizzarli e utilizzarli per ricavare valore. L'architettura tabellare dei dati strutturati permette la loro intuitiva manipolazione. Inoltre, ci sono diversi strumenti (Excel), e applicazioni (PowerBI) che permettono la gestione intuitiva dei dati strutturati, rendendo l'accesso disponibile a diversi utenti di business che non hanno necessità di avere una conoscenza approfondita dei tipi di dati e del loro funzionamento. D'altra parte, però, i dati che vengono definiti attraverso una struttura fissa spesso tendono ad avere una flessibilità e utilizzo limitata, svolgendo una sola funzione. Inoltre, una volta definita, cambiare la struttura dei dati sottostante ad essi può essere molto costoso e dispendioso in termini di tempo.

I dati non strutturati invece hanno una natura molto più dinamica, nel senso che siccome sono archiviati nel loro formato nativo, la loro struttura rimane indefinita finché non sorge la necessità da parte di un data consumer⁹, che raccoglie, cataloga e analizza solo i dati necessari. Questo rende il processo di collazionamento molto facile e rapido.

D'altra parte, i dati non strutturati sono molto più difficili da analizzare rispetto ai dati strutturati, in quanto richiedono una profonda comprensione dei tipi di dati, e richiedono tool specializzati.

Un'altra differenza tra i dati strutturati e non strutturati è l'architettura fisica in cui vengono archiviati: i dati strutturati sono raccolti in sistemi di archiviazione con schemi rigidi e libero accesso, per esempio le Data Warehouses, mentre i dati non strutturati sono archiviati in Data Lakes, che permettono un tipo di utilizzo pay-per-use, limitando i costi e permettendo un'elevata scalabilità. Entrambi questi concetti saranno approfonditi nel capitolo 3.5.1 (Le strutture dei dati analitici).

⁹ Qualsiasi utente che consuma dei dati, utilizzandoli per diversi scopi.

I dati non strutturati rappresentano oggi tra l'80% e il 90% di tutti i dati aziendali (IDC¹⁰, n.d.). I progressi tecnologici e i contenuti generati dall'uomo contribuiscono ad una crescita continua di proliferazione di dati dovuta al loro volume, varietà e velocità di produzione. Si stima infatti che questi ultimi crescano ad un ritmo tre volte maggiore rispetto ai dati strutturati.

Con l'avvento dei modelli di Large Language Models e la capacità di automatizzare la creazione, l'elaborazione e l'analisi dei contenuti, questa tendenza è solo destinata a continuare a crescere.

Ma la natura dei dati non strutturati come già menzionato pone delle difficoltà alle organizzazioni per poterli memorizzare in un database, nonché le loro caratteristiche li rendono difficili per la ricerca, la modifica e l'attribuzione. Questo è anche il motivo per cui la maggior parte delle organizzazioni concentra i propri sforzi nel Data Management di dati non strutturati. Fortunatamente nuovi strumenti come i LLM, permetteranno di aumentare la comprensione stessa dei dati non strutturati.

3.4.2. DATI OPERATIVI E ANALITICI

DATI OPERATIVI

I dati operativi sono le informazioni essenziali che supportano le attività quotidiane di un'azienda. Questi dati riflettono lo stato attuale dell'azienda e sono utilizzati per mantenere l'integrità delle transazioni in tempo reale. Sono raccolti, memorizzati e processati tramite sistemi di elaborazione delle transazioni (OLTP¹¹), che gestiscono operazioni come creazione, aggiornamento e cancellazione dei dati.

Queste transazioni possono essere considerate come "oggetti ad evento", perché ogni transazione o interazione è vista come un evento che genera dati e può attivare ulteriori processi all'interno del sistema. I sistemi OLTP (Oracle, n,d) sono infatti progettati per gestire un elevato numero di transazioni simultanee, richiedendo oltretutto un tempo di risposta molto rapido, dovendo soddisfare l'accesso simultaneo da più utenti mantenendo l'integrità dei dati. Gli utenti, infatti, interagiscono con applicazioni front-end modificando i database relazionali sottostanti. Esempi di applicazioni OLTP sono gli e-commerce, le banche, o qualsiasi sistema ERP¹², CRM¹³, MRP¹⁴.

DATI ANALITICI

I dati analitici forniscono all'azienda una visione storica, integrata e aggregata di tutte le attività aziendali. Sono il risultato dall'analisi di aggregati di dati operativi derivanti dell'esecuzione dei processi aziendali.

Questi dati, gestiti tramite dei sistemi OLAP¹⁵, sono i dati utilizzati e ottimizzati per l'analisi storica e per le previsioni, attraverso la creazione di report e visualizzazioni o di modelli di machine learning. Le tecnologie OLAP funzionano aggregando dati da varie fonti in strutture multidimensionali note come cubi OLAP.

Molti dati che gestiscono le aziende hanno infatti dimensioni multiple. Per esempio, i dati di vendita hanno diverse connotazioni rispetto ad una posizione (stato, regione, provincia, negozio). Nei data lakes, i data sets sono immagazzinati in tabelle che gestiscono due dimensioni alla volta e grazie ad

¹⁰ International Data Corporation

¹¹ Online Transaction Processing

¹² Enterprise Resource Planning: Oracle, Sap

¹³ Customer Relationship Manager: Salesforce, Hubspot

¹⁴ Material Resource Planning: Sap, Infor, Epicor

¹⁵ Online Analytical Processing

OLAP, i dati sono estratti e organizzati in tabelle multidimensionali che permettono un rapido accesso ai dati stessi. Inoltre, un altro punto importante è l'assegnazione delle gerarchie tra le dimensioni, come ad esempio la dimensione "tempo" con una gerarchia formata dagli anni, trimestri, mesi e giorni.

Diversi aziende forniscono servizi di database OLAP, spesso in Cloud, come IBM, AWS, Microsoft e altre. I dati analitici complementano quelli operativi e sono fondamentali per il funzionamento efficiente di un'azienda, fornendo le basi per decisioni strategiche e supportando il processo di miglioramento continuo. L'applicazione dei framework strategici ed operativi che fanno riferimento al Data Management si riferiscono principalmente ai dati analitici.

RELAZIONE TRA I DATI OPERATIVI E ANALITICI

I processi operativi sono influenzati dalla strategia aziendale. Lo svolgimento dell'insieme delle attività e compiti produce un grande volume di dati, che viene contenuto all'interno dei database OLTP. Il processo di ETL¹⁶ estrae i dati dal database OLTP e li trasforma nell'area di staging, includendo la pulizia dei dati e l'ottimizzazione per l'analisi. I dati trasformati vengono quindi caricati nel database di elaborazione analitica online (OLAP).

3.4.3. I DATI E LE FUNZIONI NEI SISTEMI INFORMATIVI

Un'ultima categoria di dati fa riferimento alla funzione all'interno dei sistemi informativi aziendali. Sulla base della loro connotazione e utilizzo possiamo distinguere le seguenti categorie.

I Master data rappresentano le entità chiave di business su cui si basano le transazioni. Sono fondamentali per il funzionamento dell'azienda perché aiutano a descrivere e organizzare i processi aziendali, siccome sono utilizzati trasversalmente in tutte le funzioni o domini. Esempi di entità descritti dai Master Data riguardano i clienti, i prodotti, i fornitori, e i Master Data l'insieme di dati che li descrivono. Sono oggetti concettuali che hanno un'esistenza indipendente per scopi di business, caratterizzati da un insieme di campi e attributi e una chiave per essere identificabili univocamente.

I reference data invece sono insiemi di dati utilizzati per classificare o categorizzare altre informazioni all'interno dell'azienda. Servono come standard di riferimento che rimane relativamente stabile nel tempo. Possono essere assimilati ai dati analitici, ma hanno un significato aggiunto. I reference data sono stabili e standardizzati e permettono di convalidare la coerenza e l'integrità dei dati transazionali, assicurando coerenza nei sistemi e nei processi aziendali. Per esempio, i codici di valuta, le classificazioni industriali o le tabelle dei paesi sono esempi di reference data.

Per ultimo, i metadati sono dati che descrivono altri dati e permettono di migliorare la comprensione e organizzazione. Sono utilizzati per gestire e governare la qualità, la conformità dei dati, la loro accessibilità e tracciabilità. Il ruolo e funzionamento dei metadati verrà largamente approfondito nel capitolo 6.2 (I metadati e le sue strutture). Questi sono estremamente importanti in vari aspetti della gestione dei dati e svolgono un ruolo cruciale nell'ambito di progetti di Data Governance.

3.4.4. BIG DATA

I Big Data si riferiscono ad agglomerati di dati estremamente grandi e complessi che non possono essere gestiti, elaborati o analizzati con i tradizionali strumenti di gestione dei dati. I Big Data non

¹⁶ Extract, Transform, Load

fanno riferimento a tipi specifici di dati, bensì è un termine che descrive grandi volumi di dati, strutturati e non strutturati, difficili da gestire, che inondano le aziende ogni giorno (SAS, n.d.)

I Big Data sono fonte di molte opportunità per le organizzazioni dato il loro volume, velocità e varietà di creazione. I Big Data rendono possibile il processo inverso di estrarre conoscenza a partire dai dati. Di solito la statistica applicata ai processi produttivi permette alle aziende di analizzare dati strutturati con il fine di descrivere tendenze passate, ad esempio valutare l'andamento della domanda di un particolare modello di microchip in un anno. Ci sono anche modelli che sono utilizzati per prevedere comportamenti futuri, ma, purtroppo, questi non hanno un'elevata affidabilità.

La Data Science, ovvero l'integrazione di metodi provenienti dalla matematica, statistica, informatica, elaborazione dei segnali, modellazione della probabilità, riconoscimento dei modelli, apprendimento automatico, modellazione dell'incertezza e visualizzazione dei dati hanno permesso di analizzare e trarre valore dai Big Data.

In particolare, gli algoritmi di machine learning, ovvero metodi computazionali che permettono ai computer di imparare da dati senza essere esplicitamente programmati, vengono utilizzati per analizzare grandi quantità di dati e identificare pattern o relazioni complesse al loro interno, spesso non identificabili attraverso l'aggregazione di dati, al fine di fare previsioni o prendere decisioni.

Con l'introduzione dei Big Data negli ambienti di Data Warehousing e Business Intelligence, le tecniche di Data Science vengono utilizzate per fornire una visione futura dell'organizzazione. Le capacità predittive, in tempo reale e basate su modelli, utilizzando diversi tipi di fonti di dati, offrono alle organizzazioni una migliore visione della direzione in cui si trovano. I data scientist sono gli esperti che si occupano di sviluppare e applicare questi algoritmi, analizzando i dati e interpretando i risultati per estrarre valore commerciale.

3.5. I DATI ANALITICI: STRUTTURE E FIGURE PROFESSIONALI

3.5.1. LE STRUTTURE DEI DATI ANALITICI

La forma in cui i dati analitici vengono gestiti nelle aziende ha subito dei cambi nel tempo, dovuti all'aumentare delle complessità organizzative, la proliferazione di fonti di dati e la crescita delle aspettative del valore estraibile dai dati.

LA DATA WAREHOUSE

Uno dei primi sistemi di archiviazione e gestione dei dati progettato per consentire l'analisi e l'elaborazione dei dati aziendali è la Data Warehouse. Questo tipo di architettura dei dati consente di integrare e conservare dati provenienti da diverse fonti aziendali in un unico repository centralizzato, ottimizzato per l'analisi.

I dati sono infatti:

- Estratti da numerosi database e fonti operative;
- Trasformati in uno schema universale, rappresentato in un formato tabellare;
- multidimensionale e variabile nel tempo;
- Caricati nelle tabelle della Data Warehouse;
- L'accesso è gestito attraverso query di tipo SQL;
- Servono principalmente per i casi d'uso di reporting e visualizzazione analitica.

Il processo di utilizzo dei dati in una Data Warehouse è gestito principalmente dai Data Analyst, ovvero i responsabili di interpretare i dati per estrarre insights preziosi che guidano le decisioni aziendali. Oggigiorno queste figure, facendo parte principalmente dei team di IT, hanno conoscenze approfondite degli strumenti e tecnologie analitiche, come i linguaggi SQL, Python o R e strumenti di visualizzazione come Tableau o Power BI. A seconda del caso d'uso, gli analisti identificano, raccolgono ed elaborano dati al fine di creare Dashboard e reportistica per soddisfare le richieste strategiche o per la risoluzione dei problemi.

La struttura monolitica della Data Warehouse (Data Mesh, 2022) presenta vantaggi e svantaggi. Se da un lato la centralizzazione dei dati permette di organizzare diverse fonti in un'unica posizione, risparmiando sia tempo e risorse per l'integrazione, pulizia e analisi dei dati, dall'altra man mano che le complessità organizzative aumentano, nuovi modelli di business vengono implementati e le fonti dei dati crescono a dismisura, la gestione diventa complicata.

Inoltre, la progettazione, implementazione e gestione di una Data Warehouse può essere complessa e costosa e, una volta implementata, effettuare dei cambi per allineare la sua struttura alle esigenze strategiche aziendali può essere dispendioso.

Per ultimo, la crescente complessità nel creare una tecnologia unificata per gestire i dati su larga scala all'interno di un'organizzazione comporta anche una maggiore difficoltà di gestione. Questo si traduce in un'ulteriore specializzazione del personale coinvolto, rendendolo costoso e difficile da sostituire.

Un'alternativa (spesso utilizzata complementariamente ai Data Warehouse) sono i Data Mart, ovvero subset di dati più piccoli e specializzati, orientati verso specifici dipartimenti o funzioni aziendali. Essi semplificano l'accesso e l'analisi dei dati fornendo un'ottica focalizzata su determinati processi o aree di business, riducendo così la complessità del data warehouse e migliorando le prestazioni delle query per gli utenti finali.

IL DATA LAKE

Come anticipato, le tecnologie che facilitano le analisi dei dati hanno subito uno sviluppo tecnologico. L'introduzione degli algoritmi di machine learning per l'analisi dei dati, le strutture di Data Warehouse non erano adatte a soddisfare i nuovi modelli di analisi dei dati.

Inoltre, la crescente raccolta e utilizzo dei Big Data ha causato la progressiva perdita di necessità di avere strutture dati con uno schema universale, rappresentato in un formato tabellare multidimensionale e variabile nel tempo. I dati possono essere infatti raccolti in modo strutturato e non strutturato in formato nativo.

In questo contesto i Data Lake si configurano come strutture molto più scalabili e flessibili rispetto ai Data Warehouse (Microsoft Azure, n.d.), caratteristiche importanti considerando l'elevato volume di dati che caratterizza i Big Data.

Questo approccio flessibile consente ai data scientist di accedere a una vasta gamma di dati in un unico luogo e di esplorarli liberamente per identificare le relazioni e i pattern che possono essere utilizzati per addestrare modelli di machine learning.

Al contrario delle Data Warehouses, all'interno dei Data Lakes, i dati:

- appresentano il più possibile il contenuto e la struttura originali
- vengono trasformati minimamente per adattarsi ai formati di archiviazione più diffusi

- vengono caricati in uno storage di oggetti scalabile
- I data scientist accedono principalmente allo storage del data lake per l'analisi e l'addestramento dei modelli di machine learning

ARCHITETTURA CLOUD

L'architettura cloud si riferisce alla struttura e ai componenti tecnologici che formano un ambiente di elaborazione basato sul cloud. Comprende l'infrastruttura, le piattaforme, i servizi e le applicazioni che operano e interagiscono nel cloud, facilitando l'archiviazione, l'elaborazione e la gestione dei dati su una rete di server remoti piuttosto che su server locali o computer personali.

Le organizzazioni che adottano l'architettura cloud spesso spostano le risorse IT sul cloud, eliminando la necessità di server e storage on-premises¹⁷, riducendo la necessità di immobili, raffreddamento ed energia per i data center IT e sostituendoli con una spesa IT mensile.

Queste architetture possono gestire sia lo streaming di dati in tempo reale che il batch processing¹⁸. Inoltre, i servizi di cloud riducono la necessità di gestire fisicamente l'infrastruttura, in quanto permettono grande scalabilità e flessibilità.

Per ultimo le tecnologie cloud convergono le strutture del data warehouse e del data lake in un'unica tecnologia, estendendo il data warehouse per includere l'addestramento ML integrato, o alternativamente costruendo l'integrità del data warehouse, la transazionalità e i sistemi di interrogazione nelle soluzioni di data lake.

Questa soluzione permette l'analisi dei dati in tempo reale e allo stesso tempo riduce il costo derivante dall'utilizzo di infrastrutture per elaborare grandi volumi di dati.

¹⁷ Server locale, è un server fisico situato in loco presso il centro dati o la struttura dell'azienda

¹⁸ Le transazioni sono elaborate attraverso il loro raccoglimento in gruppi

4. I PRINCIPI DEL DATA MANAGEMENT

I principi del Data Management (Dama International, 2017) forniscono un quadro per garantire l'accuratezza, la sicurezza e l'usabilità dei dati, consentendo decisioni informate, migliorando l'efficienza operativa e arricchendo l'esperienza del cliente.

Tuttavia, raggiungere una gestione efficace dei dati richiede la ricerca di equilibrio tra esigenze strategiche e operative, assicurandosi che le pratiche di gestione dei dati siano allineate sia con gli obiettivi aziendali che con i requisiti operativi quotidiani. Le esigenze strategiche richiedono che la gestione dei dati supporti gli obiettivi a lungo termine, l'innovazione e la competitività, mentre le esigenze operative richiedono una gestione dei dati efficiente, affidabile e sicura per sostenere le operazioni quotidiane.

Senza un approccio equilibrato, le organizzazioni rischiano di avere sistemi di gestione dei dati troppo rigidi e inflessibili o troppo caotici e imprevedibili.

I principi di gestione dei dati sono essenziali per raggiungere questo equilibrio. Il Dama cita che:

- **I dati sono un asset con proprietà uniche:** oltre alla presenza dei metadati che ne descrivono le proprietà, i dati hanno una caratteristica importante che li differenzia dagli altri tipi di asset, ovvero che non sono consumati quando sono utilizzati;
- **Il valore dei dati può e dovrebbe essere espresso in termini economici:** questo principio pone una grande sfida in capo alle organizzazioni, in quanto non esiste uno standard definito per attribuire un valore economico ai dati;
- **Gestire i dati significa gestire la qualità dei dati:** garantire che i dati siano adatti allo scopo è un obiettivo principale della gestione dei dati;

Come ogni altro processo e progetto aziendale con una valenza strategica, i progetti di Data Management richiedono un'accurata pianificazione, esecuzione e controllo, ma, soprattutto, per la sua corretta esecuzione richiedono figure professionali con diverse skills.

Toccando tutte le aree funzionali delle aziende, la gestione dei dati richiede competenze tanto tecniche quanto di business, nonché un controllo continuo per garantire il raggiungimento degli obiettivi.

Al fine di pianificare correttamente, è necessario comprendere che l'applicazione del Data Management ha una natura locale, sulle singole aree funzionali, ma deve essere applicato all'impresa intera per poter avere un'efficacia misurabile. Questo è anche il motivo per cui il Data Management e la Data Governance sono interconnessi.

Inoltre, senza un chiaro coinvolgimento della leadership in questi tipi di progetti, è molto verosimile che la pianificazione non si trasformi in best practices da seguire a livello operativo. Vedremo in seguito come questo punto è un tema tanto fondamentale quanto delicato e critico per le organizzazioni.

Per riassumere, l'applicazione di un progetto di Data Management e di Data Governance richiede un vero e proprio processo di cambiamento, anche definito Change Management, che coinvolge tutti i domini aziendali, e che è guidato da una visione strategica della leadership.

5. IL VALORE DEI DATI

Come anticipato, attribuire un valore economico ai dati è un compito estremamente difficile. La natura complessa dei dati stessi rappresenta un ostacolo significativo.

Sebbene ci sia un ampio consenso sul fatto che i dati siano fondamentali per la creazione di valore all'interno delle organizzazioni, c'è poco consenso sul come quantificare il valore dei dati. Ci sono tre tecniche principali utilizzate dalle organizzazioni per stimare il valore dei loro dati (World Economic Forum, 2023), che si concentrano su diversi aspetti, come il loro costo, il rischio della loro gestione e i potenziali ritorni derivanti dal loro utilizzo.

Il metodo più diretto, ma anche meno preciso, è il metodo dei costi. Questo metodo consiste nel calcolare i costi associati alla gestione dei dati, come ad esempio i costi di infrastruttura (es. il costo di acquisto e manutenzione degli hardware e software utilizzati per archiviare e gestire i dati), i costi del personale (es. le spese relative ai dipendenti che si occupano della gestione e manutenzione dei dati) e i costi energetici (es. il costo dell'energia elettrica necessaria per alimentare i dispositivi di archiviazione e gestione dei dati). Il problema di questo metodo è che tendenzialmente non riesce a catturare i futuri rendimenti economici che potrebbero essere ottenuti dai dati tramite la loro elaborazione, utilizzo e condivisione. Questo metodo, infatti, si basa sul presupposto che le organizzazioni sono interessate a considerare i processi di gestione dei dati come delle operazioni a valore aggiunto. Di conseguenza, l'organizzazione può stabilire che l'insieme del patrimonio informativo raccolto, elaborato e conservato dovrebbe valere almeno quanto il suo costo di gestione.

Il secondo approccio è quello del valore aggiunto, basato sul potenziale ritorno aggiuntivo dall'utilizzo dei dati, chiamato metodo di valutazione dei dati. Questo approccio si basa sul reddito aggiuntivo prodotto dalle attività di un'impresa, misurando i flussi di cassa. La valutazione deve essere svolta confrontando questi ultimi prima e dopo l'utilizzo dei dati, quantificando il loro valore sotto la forma di ricavi aggiuntivi, costi ridotti o entrambi. Ad esempio, un'azienda manifatturiera che produce e commercializza una macchina industriale può utilizzare i dati sulla manutenzione preventiva per ottenere un vantaggio competitivo. Analizzando anni di dati sui guasti e le frequenze di manutenzione delle loro macchine in uso ai clienti, l'azienda potrebbe notare una manutenzione irregolare, causando spesso guasti imprevisti. Offrendo servizi di manutenzione preventiva personalizzati, come promemoria periodici e sconti, l'azienda potrebbe ridurre i tempi di inattività delle macchine dei clienti e aumentare la loro fedeltà. Questo può tradursi in un aumento del reddito grazie alla vendita di pezzi di ricambio e servizi di manutenzione.

Il terzo metodo è quello del valore di mercato. I dati come ogni altro asset possono essere scambiati e venduti. Questo metodo consiste nel ricercare il valore di mercato di set di dati simili o database per determinare un prezzo equo per i dati. La legge della domanda e dell'offerta regolano i meccanismi di determinazione del valore di un set di dati. Si può prendere come esempio il settore farmaceutico, in cui i database sono estremamente importanti perché forniscono informazioni cruciali sui pazienti, sulle malattie e sulla risposta ai trattamenti. Le aziende farmaceutiche spesso investono considerevoli risorse finanziarie all'acquisizione di dati accurati e dettagliati per informare la loro ricerca e lo sviluppo di nuovi farmaci (Forbes, 2022).

Un ultimo tipo di approccio è quello basato sui rischi: consiste nel valutare i rischi associati alla perdita, furto o corruzione dei dati, e assegnare un valore ai dati in base alla loro sensibilità e importanza.

Sebbene non ci sia un metodo più corretto degli altri, la letteratura enfatizza l'importanza di associare il valore dei dati agli specifici casi d'uso, facendo un collegamento con le fasi del ciclo di vita dei dati.

5.1.1. CONDIVISIONE DEI DATI E VALORE

Il valore dei dati accresce man mano che i dati avanzano negli step del loro ciclo di vita. Un insieme di dati grezzi può avere un certo valore nella fase iniziale di generazione o raccolta, ma questo si moltiplica man mano che l'organizzazione lo elabora, lo analizza, e lo condivide con altri partner. Il valore maggiore è raggiunto nella fase di analisi e condivisione. Al di fuori dei metodi per la valutazione del valore dei dati che sono stati presentati, bisogna ricordare quanto le informazioni ricavabili attraverso l'analisi, possono creare valore migliorando i processi di decisione e di efficienza operativa, ultimamente migliorando i prodotti e servizi dell'organizzazione.

A maggior ragione, molte organizzazioni che migliorano i propri sforzi per garantire un ottimo Data Management eventualmente si rendono conto che partner, fornitori o clienti dispongono di dati e approfondimenti adiacenti che, se combinati in modo ponderato attraverso sistemi e piattaforme, possono creare proposte di valore più ricche per tutti. Dal 2017 al 2019, il numero di aziende con partnership legate ai dati è passato dal 21% al 40% e questo dato è in costante crescita (McKinsey, 2019).

La cosiddetta pratica di "Data Sharing" non deve essere intesa come condivisione gratuita. Ci sono diversi modelli o forme di condivisione che stanno emergendo man mano che l'economia dei dati evolve (European Union, 2018). Tra diversi modelli è possibile distinguere:

- **Monetizzazione dei dati:** approccio unilaterale attraverso il quale le aziende generano ricavi aggiuntivi dai dati che condividono con altre aziende. I dati possono anche essere monetizzati tramite la fornitura di servizi;
- **Data marketplaces:** intermediari che mettono in contatto fornitori di dati e utenti di dati per scambiare dati su una piattaforma online sicura. Queste imprese generano ricavi dalle transazioni di dati che avvengono sulla piattaforma.
- **Piattaforme di dati industriali:** questo approccio collaborativo e strategico consiste nello scambio di dati tra un gruppo ristretto di aziende ambienti chiusi, sicuri ed esclusivi per promuovere lo sviluppo di nuovi prodotti/servizi e/o migliorare la loro efficienza interna.
- **Abilitatori tecnici:** aziende specializzate e specificamente dedicate a facilitare la condivisione dei dati tramite una soluzione tecnica. I ricavi provengono dall'installazione, dall'uso e/o dalla manutenzione della soluzione (non dai dati scambiati).
- **Politica dei dati aperti:** aziende che scelgono di condividere dati gratuitamente per promuovere lo sviluppo di nuovi prodotti e/o servizi.

Indipendentemente dal modello scelto, uno studio di Gartner riporta che le aziende che condividono i dati all'esterno della propria organizzazione generano un valore economico misurabile tre volte superiore rispetto a quelle che non lo fanno (Gartner, 2021).

Alla base di queste dinamiche c'è la necessità di costruire relazioni profonde e sostenibili con clienti, stakeholders e consumatori. Inoltre, per essere in grado di condividere i propri dati all'esterno, le organizzazioni devono stabilire processi di data management e data governance ben definiti, per classificare e organizzare i dati in modo appropriato, evitando così eventuali perdite o problematiche.

6. DATA GOVERNANCE

Il DAMA definisce la Data Governance come “l'esercizio dell'autorità e del controllo (pianificazione, monitoraggio e applicazione) sulla gestione degli asset di dati”. I processi di Data Management e quelli di Data Governance sono estremamente connessi ed in alcuni casi coincidenti, ma l'obiettivo dei due è differente. Infatti, se l'obiettivo del Data Management è istituire processi e meccanismi che rendano possibile per l'organizzazione estrarre valore dai propri dati, la Data Governance si concentra su come vengono prese le decisioni riguardanti i dati e su come si prevede che le persone e i processi si comportino in relazione ai dati.

6.1. LA DATA GOVERNANCE COME FUNZIONE STRATEGICA

La governance dei dati dovrebbe essere considerata una funzione strategica per le aziende perché affronta sfide fondamentali intrinseche nella gestione degli asset di dati, sempre più vitali per il successo organizzativo.

Nel panorama attuale guidato dai dati, dove questi ultimi sono considerati un asset aziendale critico, una governance dei dati efficace garantisce che i dati siano gestiti in modo completo lungo tutto il loro ciclo di vita, dall'acquisizione alla dismissione. Attraverso l'istituzione di politiche, standard e processi per la gestione dei dati, la governance dei dati aiuta le organizzazioni a strutturare i processi di Data Management per l'estrazione del valore dei loro dati, riducendo allo stesso momento i rischi legati alla loro scarsa qualità, alle violazioni della sicurezza e alla non conformità normativa. La Data Governance non deve essere considerata quindi come una attività progettuale puntuale bensì come una funzione strategica essenziale e svolta continuamente, a cui sono legati due principali obiettivi, o driver di business, che sono la riduzione dei rischi e il miglioramento dei processi (DAMA International, 2017).

6.1.1. RIDUZIONE DEI RISCHI

I rischi in ambito della Data Governance riguardano la sensibilità dei dati utilizzati in azienda. In un contesto in cui i processi aziendali sono per lo più digitalizzati e le organizzazioni fanno affidamento sui dati giornalmente per la presa di decisioni, l'esposizione ai rischi legati alla loro gestione non fa che aumentare. La Data Governance supporta la gestione e riduzione dei rischi nelle organizzazioni attraverso dei processi che permettano la loro identificazione, la valutazione dell'impatto e la loro gestione.

È possibile definire due aree principali riguardanti la riduzione dei rischi, ovvero quella relativa alla sicurezza dei dati, e quella relativa alla privacy dei dati. Queste due aree sono interconnesse perché hanno un obiettivo specifico: sulla base dell'importanza strategica che i dati rappresentano, o al fine di essere conformi alle politiche e regolamenti in merito alla gestione dei dati, è necessario determinare la segretezza e confidenzialità dei dati, gestire chi ha accesso a questi ultimi e le modalità di condivisione.

SICUREZZA DEI DATI

La sicurezza dei dati riguarda la pratica di proteggere le informazioni digitali da accessi non autorizzati, danneggiamenti o furti durante tutto il ciclo di vita. La sicurezza dei dati comprende sia l'aspetto fisico della gestione dei dati, ovvero degli hardware e dispositivi di archiviazione, sia l'aspetto della sicurezza logica delle applicazioni software. La sicurezza dei dati non mira solo a

stabilire delle misure contro attacchi informatici esterni, ma anche dalle minacce interne causate dagli errori umani.

Infatti, secondo un report della Stanford University (Stanford University, Tessian, 2023), nel 2023 circa l'88% delle situazioni in cui si sono verificate delle violazioni di dati sono state causate da errori umani, attraverso pratiche di phishing¹⁹, l'invio di e-mail al destinatario sbagliato e il riutilizzo di password deboli. Secondo un report di IBM (IBM, 2023), il costo globale medio di un evento che ha causato la violazione dei dati di un'organizzazione è di 4.45 milioni di dollari, cifra aumentata del 15% nei soli ultimi tre anni.

È chiara la ragione del motivo per cui lo stesso report presenta come il 51% delle organizzazioni stiano pianificando di aumentare gli investimenti per aumentare la sicurezza dei propri dati aziendali.

È possibile identificare 3 tipi di misure di protezione dei dati:

- **Crittografia (ISO, n.d.):** è un metodo di protezione dei dati che trasforma le informazioni leggibili in un formato illeggibile utilizzando un algoritmo e una chiave di crittografia. Solo chi possiede la chiave corretta può decifrare i dati e renderli nuovamente leggibili;
- **Controlli di Accesso:** sono misure di sicurezza che regolano chi può visualizzare o utilizzare le risorse di un sistema informativo. Questo include l'autenticazione degli utenti, ovvero le modalità attraverso le quali il sistema verifica l'identità di un utente che cerca di accedere ad una risorsa, e l'autorizzazione dei loro privilegi, il processo che determina quali risorse o dati un utente autenticato può visualizzare o modificare;
- **Autenticazione Multifattoriale:** è un tipo di sistema di autenticazione che richiede agli utenti di fornire due o più verifiche indipendenti per accedere a una risorsa. Questo aumenta significativamente la sicurezza rispetto all'uso della sola password.

Queste misure di per sé non sono sufficienti a garantire la protezione completa dei dati di un'organizzazione. È necessario, infatti, che quest'ultima stabilisca un vero e proprio Framework e degli Standard di sicurezza che aiutino a rafforzare le pratiche di sicurezza dei dati. A livello internazionale sono stati stabiliti degli standard per impostare e manovrare un sistema di gestione della sicurezza delle informazioni come l'ISO/IEC 27001 (ISO/IEC, n.d.). L'ISO/IEC è uno standard internazionale che fornisce alle aziende di qualsiasi dimensione e di tutti i settori di attività una guida per stabilire, implementare, mantenere e migliorare continuamente un sistema di gestione della sicurezza delle informazioni. Implementare lo standard ISO/IEC 27001 offre un approccio olistico alla sicurezza delle informazioni, comprendendo vari aspetti dell'organizzazione, non solo i sistemi fisici e digitali, ma fornisce un quadro gestionale centralizzato, per preparare persone, processi e tecnologie a fronteggiare i rischi tecnologici e altre minacce, e consente di risparmiare aumentando l'efficienza e riducendo le spese per tecnologie di difesa inefficaci.

Infatti, se le 3 misure di protezione dei dati identificati precedentemente siano un metodo attuato principalmente per prevenire e controllare l'accesso a dati sensibili aziendali, in realtà queste spesso non sono sufficienti. Il National Institute of Standards and Technology (NIST) stabilisce un framework, chiamato il NIST Cybersecurity Framework (NIST, 2024) che identifica cinque macroaree per supportare le organizzazioni nella gestione e riduzione dei rischi relativi alla cybersecurity:

- **Identificazione:** sviluppare una comprensione organizzativa per gestire il rischio di cybersecurity per sistemi, persone, asset, dati e capacità;

¹⁹ Il phishing è un tipo di truffa effettuata su Internet attraverso la quale un malintenzionato cerca di ingannare la vittima convincendola a fornire informazioni personali, dati finanziari o codici di accesso, fingendosi un ente affidabile in una comunicazione digitale.

- **Protezione:** sviluppare e implementare adeguate salvaguardie per garantire la fornitura di servizi critici dell'infrastruttura, tra cui rientrano i sistemi di controllo degli accessi e crittografia;
- **Rilevazione** (Kron, n.d.): sviluppare e implementare adeguate attività per identificare il verificarsi di un evento di cybersecurity. Un attacco informatico è possibile che non sia immediatamente rilevabile. Nel corso dei decenni dall'inizio di Internet, gli attacchi di phishing si sono evoluti in imprese altamente sofisticate che prevedono l'uso dell'ingegneria sociale, convincendo i bersagli che gli aggressori sono contatti legittimi che fanno richieste legittime di informazioni.
- **Risposta:** sviluppare e implementare adeguate attività per agire in relazione a un evento di cybersecurity rilevato. Infatti, è necessario che le organizzazioni, coscienti dei rischi presenti, elaborino un piano di contingenza per ciascun rischio, per minimizzare il tempo di risposta
- **Recupero:** Sviluppare e implementare adeguate attività per mantenere piani di resilienza e ripristinare eventuali capacità o servizi compromessi a causa di un evento di cybersecurity.

L'allegato 9 mostra il framework elaborato dal NIST. È importante notare come alla base di tutti gli step del Framework sia presente l'attività di Governance, fondamentali per contestualizzare le attività di cybersecurity nella più ampia strategia di gestione del rischio aziendale (ERM²⁰) di un'organizzazione.

Inoltre, è necessario menzionare che senza un sistema sottostante che classifichi i dati e li organizzi in base alla loro sensibilità e importanza aziendale, i processi di mitigazione dei rischi di violazione dei dati mancherebbero di una base solida su cui fondarsi. È essenziale che l'azienda conosca, identifichi e gestisca correttamente i dati considerati strategici, in modo da poterli proteggere accuratamente. Senza una chiara comprensione di quali dati debbano essere trattati con particolare riguardo, non sarà possibile implementare efficacemente i processi di gestione del rischio. Gli artefatti della data governance risolvono questa problematica, fornendo una struttura e degli strumenti per una gestione efficace e sicura dei dati strategici dell'azienda.

PRIVACY DEI DATI

La privacy dei dati è diventata un tema sempre più rilevante e critico a causa della crescente raccolta, elaborazione e condivisione di dati personali da parte di individui, organizzazioni e governi. Diverse normative definiscono e impongono obblighi alle organizzazioni per gestire il modo in cui i dati raccolti sono elaborati, trattati ed eventualmente condivisi.

In particolare, si fa riferimento alle PII²¹, ovvero le informazioni che possono essere utilizzate per identificare un individuo in modo univoco. Queste informazioni comprendono nomi, indirizzi, numeri di telefono, indirizzi e-mail e altre.

In Europa è possibile citare il General Data Protection Regulation (GDPR, n.d.) che, riconoscendo la privacy come un elemento fondamentale per proteggere la dignità, l'autonomia e i diritti degli individui, pone chiarezza su questi temi, applicando misure destinate alla tutela dei consumatori dei loro dati personali, stabilendo anche i casi e le multe per la violazione o il mancato corretto trattamento di questi ultimi. Queste sanzioni raggiungono un massimo di 20 milioni di euro o del 4% del fatturato globale, oltre al risarcimento eventualmente dovuto agli interessati.

²⁰ Enterprise Risk Management: un approccio integrato e congiunto alla gestione del rischio in un'organizzazione.

²¹ Personally Identifiable Information

La normativa indica che, nel caso in cui un'azienda o un'organizzazione processi dei dati, lo debba fare seguendo 7 principi (GDPR, 2023), stabiliti dall'articolo 5 del GDPR:

1. **Liceità, correttezza e trasparenza:** il trattamento deve essere lecito, equo e trasparente nei confronti dell'interessato o proprietario ai dati;
2. **Limitazione della finalità:** è necessario trattare i dati per le finalità legittime specificate esplicitamente all'interessato quando questi vengono raccolti;
3. **Minimizzazione dei dati:** è necessario raccogliere e trattare solo quanto è assolutamente necessario per le finalità specificate;
4. **Precisione:** è necessario mantenere i dati personali accurati e aggiornati;
5. **Limitazione della conservazione:** è possibile conservare solo i dati personali identificativi per il tempo necessario alla finalità specificata;
6. **Integrità e riservatezza:** il trattamento dei dati deve essere effettuato in modo tale da garantire la sicurezza, l'integrità e la riservatezza adeguate;
7. **Responsabilizzazione:** il titolare del trattamento è responsabile di poter dimostrare la conformità al GDPR con tutti questi principi;

Come per la sicurezza dei dati, la normativa prevede anche un insieme di strumenti che devono essere utilizzati dalle organizzazioni per valutare i potenziali rischi per la privacy associati a nuovi processi, progetti e tecnologie. I processi di Data Governance permettono di caratterizzare e classificare i dati in modo opportuno al fine di poter rispettare i 7 principi del GDPR.

6.1.2. MIGLIORAMENTO DEI PROCESSI

Come analizzato, la data governance permette la riduzione dei rischi legato alle potenziali ripercussioni derivanti da un inaccurato trattamento dei dati aziendali, soprattutto per le aziende operanti in settori fortemente regolamentati, come i servizi finanziari o la sanità, in cui le organizzazioni che vi operano a raccogliere un elevato volume di PII. Nonostante ciò, se da un lato rispondere all'evoluzione della legislazione richiede processi di governance dei dati rigorosi, dall'altro l'esplosione dell'analitica avanzata e della data science ha creato la necessità di catalogare accuratamente i dati e curare la loro qualità. In questo modo, la Data Governance permette il miglioramento dei processi aziendali in diversi modi.

DATA QUALITY IMPROVEMENT

La qualità dei dati è un tema estremamente sensibile che si riferisce al livello in cui i dati sono adatti ad essere utilizzati per scopi specifici. Quando la qualità dei dati soddisfa gli standard per l'uso previsto, gli utenti possono fidarsi dei dati e sfruttarli per migliorare il processo decisionale, portando allo sviluppo di nuove strategie aziendali o all'ottimizzazione di quelle esistenti. Tuttavia, quando uno standard non è rispettato, gli strumenti di qualità dei dati forniscono un valore aggiunto aiutando le aziende a diagnosticare i problemi di fondo dei dati. La qualità dei dati è rappresentata attraverso diverse caratteristiche che riguardano i dati:

- **Accuratezza:** i dati devono essere corretti e privi di errori. Questo significa che le informazioni devono riflettere la realtà e essere precise
- **Completezza:** I dati devono essere completi e contenere tutte le informazioni necessarie. La mancanza di dati può compromettere infatti l'analisi e le decisioni basate su di essi
- **Coerenza:** i dati devono essere coerenti tra diversi sistemi e contesti. Le informazioni dovrebbero essere uniformi e non contraddittorie. Molto spesso i dati sono copiati in specifiche strutture per soddisfare richieste di analisi puntuali. È imperativo che attraverso queste procedure i dati mantengano la propria coerenza. Ad esempio, le modifiche effettuate su un dato devono avvenire in ogni luogo in cui il dato è conservato

- **Aggiornamento:** i dati devono essere aggiornati regolarmente per riflettere le informazioni più recenti. Dati obsoleti possono portare a decisioni errate;
- **Accessibilità:** i dati devono essere facilmente accessibili agli utenti che ne hanno bisogno, con adeguati controlli di sicurezza per proteggere le informazioni sensibili;
- **Rilevanza:** i dati devono essere pertinenti e utili per gli scopi per cui sono stati raccolti e vengono utilizzati;
- **Affidabilità:** i dati devono essere affidabili e degni di fiducia, in modo che le decisioni basate su di essi siano accurate e valide.

SCAMBIO DI DATI

Come affrontato nel capitolo 5.1.1 (Condivisione dei dati e valore), le pratiche di condivisione e acquisto di dati tra le aziende è un fenomeno in costante crescita. Come menzionato, alla base di queste pratiche c'è la necessità di garantire la fiducia e l'integrità dei dati.

Per ovviare a questi problemi tendenzialmente vengono redatti dei contratti per garantire che i fornitori rispettino gli accordi relativi ai dati. Questo include la definizione di requisiti chiari per il trattamento dei dati nei contratti e la stabilità di metriche di performance per valutare l'aderenza dei fornitori agli standard concordati.

Inoltre, quando si svolgono pratiche di questo tipo, bisogna ricordare i requisiti di sicurezza e privacy dei dati stabiliti dalla normativa. Si suppone infatti che l'organizzazione sia responsabile dei dati che gestisce durante tutta la sua catena del valore.

Attraverso processi di governance dei dati stabiliti, è possibile valutare l'entità del rischio dei terzi, ovvero la sicurezza e la conformità dei fornitori rispetto agli standard e alle normative applicabili.

Allo stesso modo vale per operazioni di outsourcing: le aziende devono assicurarsi che i fornitori rispettino le leggi e i regolamenti locali relativi alla privacy dei dati e alla protezione dei dati personali. Ciò può includere la verifica che i dati siano archiviati e gestiti in conformità con le leggi sulla privacy del paese in cui si trovano, e l'adozione di misure di sicurezza adeguate a proteggere i dati durante il trasferimento transfrontaliero.

6.2. I METADATI E LE SUE STRUTTURE

Nel capitolo 3.4.3 (I dati e le funzioni nei sistemi informativi) sono stati introdotti i metadati, come dati che descrivono altri dati. I metadati sono infatti informazioni che aiutano a comprendere, classificare, gestire e utilizzare i dati. Sono essenziali per la gestione efficace dei dati all'interno di un'organizzazione, poiché forniscono un contesto e una struttura ai dati. I metadati hanno diverse utilità. Per fare un esempio che migliori la comprensione dell'utilità dei metadati, prendiamo in considerazione un'azienda che produce microprocessori:

- **Descrizione dei Dati:** i metadati forniscono informazioni dettagliate sui dati, come il formato, la data di creazione, l'autore, e così via. Questo rende più facile capire di cosa trattano i dati e come possono essere utilizzati. Ad esempio, un file di progetto di un nuovo microprocessore potrebbe avere metadati che indicano il formato del file (ad esempio, CAD), la data di creazione (15 marzo 2024), l'autore (Ingegnere Marco Rossi) e una breve descrizione del progetto (Progetto del nuovo microprocessore X123);
- **Organizzazione e Classificazione:** i metadati aiutano a organizzare e classificare i dati, facilitando la loro ricerca e recupero. Per esempio, all'interno del database aziendale, i progetti dei microprocessori possono essere classificati per generazione (Gen1, Gen2, ecc.),

tipologia (ARM, x86, ecc.), o stato del progetto (in sviluppo, completato, in revisione). Un ingegnere che cerca specifiche tecniche per un microprocessore particolare può facilmente trovare il documento rilevante grazie alla classificazione basata sui metadati.

- **Tracciabilità e Provenienza:** i metadati tracciano l'origine dei dati e le modifiche apportate nel tempo. Questo è fondamentale per garantire l'integrità e l'affidabilità dei dati e permettere di risalire facilmente alle modifiche apportate, garantendo che le informazioni siano accurate e aggiornate. Ad esempio, un documento di specifiche tecniche potrebbe includere metadati che indicano tutte le revisioni effettuate, chi ha apportato le modifiche (Ingegnere Sara Bianchi, revisione il 1 aprile 2024) e la ragione delle modifiche (miglioramento dell'efficienza energetica).
- **Sicurezza e Conformità:** i metadati possono includere informazioni sulle politiche di accesso e sulle normative applicabili, aiutando a garantire che i dati sui microprocessori siano utilizzati in conformità con le leggi e i regolamenti. Ad esempio, un file contenente informazioni sensibili sulla produzione dei microprocessori potrebbe avere metadati che specificano chi ha il permesso di accedere ai dati (solo il team di sviluppo) e che la gestione dei dati deve rispettare le normative GDPR per la protezione dei dati personali. Questo aiuta l'azienda a garantire la sicurezza delle informazioni e a rispettare le normative vigenti.

Per approfondire la comprensione dell'utilizzo dei metadati, è necessario evidenziare la differenza tra i metadati di business e i metadati tecnici, che si organizzano principalmente in due strutture²², il Business Glossary e il Data Dictionary (Fonti interne Accenture). Entrambi questi strumenti raccolgono Metadati per contestualizzare dati, ma la funzione e gli utenti a cui queste strutture sono rivolte sono differenti. Inoltre, è importante menzionare il Data Lineage, per tracciare il percorso del ciclo di vita dei dati.

6.2.1. BUSINESS GLOSSARY

Il business glossary si concentra principalmente sui concetti e sui termini aziendali²³, fornendo definizioni chiare e comprensibili per gli utenti non tecnici dell'organizzazione. Include termini utilizzati comunemente nei processi aziendali e nelle comunicazioni interne, garantendo una comprensione uniforme dei concetti aziendali tra i dipartimenti.

Spesso questi termini di business si riferiscono alle caratteristiche delle entità definite nell'organizzazione. Che cosa si intende per "cliente"? E per "fornitore"? Oppure per "magazzino"? Il business glossary è stato creato per rispondere a questo tipo di domande, favorendo l'interoperabilità dei dati tra le diverse funzioni interne aziendali (per il dipartimento Vendite e quello di Supply Chain il "cliente" dovrebbe avere lo stesso significato) e anche esternamente tra le aziende che operano nello stesso settore.

Il business glossary è destinato principalmente a figure del business, come dirigenti, manager, analisti di business e altri stakeholder non tecnici, che utilizzano i dati per prendere decisioni aziendali informate e comunicare efficacemente all'interno dell'organizzazione. Queste persone possono consultare il glossario per comprendere i termini aziendali e garantire che la comunicazione sia chiara e uniforme.

²² Anche definiti "Artefatti"

²³ Anche definiti termini di business

6.2.2. DATA DICTIONARY

Il data dictionary è la controparte tecnica del Business Glossary, e fornisce una visione approfondita degli elementi di dati utilizzati nei sistemi informativi dell'azienda. Ogni flusso di dati (file o tabella di database) utilizzato all'interno dell'organizzazione deve essere registrato nel Data Dictionary. Questo dovrebbe includere i dettagli dei dati, come il tipo, la lunghezza consentita, il nome tecnico, le trasformazioni (il lineage) e qualsiasi altro dettaglio tecnico rilevante. Questi sono appunto i Metadati tecnici.

Il data dictionary è principalmente destinato a figure dell'IT, come sviluppatori, analisti di dati, amministratori di database e architetti, che sono responsabili della progettazione, dello sviluppo e della gestione dei sistemi informativi aziendali. Queste persone utilizzano il data dictionary per comprendere la struttura e le caratteristiche tecniche dei dati, guidando lo sviluppo e la manutenzione dei sistemi informativi aziendali.

6.2.3. DATA LINEAGE

I dati raramente hanno una natura statica. Le specifiche del ciclo di vita dei dati all'interno di una data organizzazione possono essere piuttosto complicate, perché i dati non attraversano solo il ciclo di vita descritto nel capitolo 3.3, ma sono caratterizzati dalla cosiddetta data lineage, cioè un percorso lungo il quale si spostano dal punto di origine al punto di utilizzo. Il data lineage è la tracciabilità completa e dettagliata dell'intero percorso dei dati dall'origine fino alla destinazione finale. In altre parole, si riferisce alla capacità di seguire e comprendere il viaggio dei dati lungo tutte le fasi del loro ciclo di vita, compresa la loro origine, trasformazione, movimento e utilizzo.

Il data lineage fornisce una visualizzazione chiara delle relazioni e delle dipendenze tra i dati e consente agli utenti di comprendere come i dati sono stati creati, modificati e utilizzati nel contesto delle applicazioni e dei processi aziendali. Mentre il business glossary e il data dictionary si concentrano sulla definizione e la struttura dei dati, il data lineage traccia il percorso dei dati attraverso i vari sistemi e processi, dalla loro origine fino al punto di utilizzo finale. Può essere considerato come una forma di ciclo di vita del dato, ma che fornisce una visibilità, più che sullo stato dell'utilizzo del dato, sulle trasformazioni e le manipolazioni dei dati lungo il percorso.

Seguendo il flusso di dati, le figure tecniche possono identificare eventuali errori o incoerenze. L'esempio nell'allegato 10 chiarisce le differenze tra i tre "artefatti" introdotti che sono utilizzati per la gestione dei metadati nelle organizzazioni.

6.3. IL FRAMEWORK DELLA DATA GOVERNANCE

Il Framework della Data Governance è una struttura definita che dirige l'implementazione di linee guida, protocolli, processi e regole stabilite per i dati in un'azienda (Dama International, 2017). È il fondamento di un programma di governance dei dati.

Il Framework è composto da politiche, processi, ruoli, metriche e standard che garantiscono l'uso efficace e controllato delle informazioni all'interno di un'organizzazione, definendo chiaramente le responsabilità e le aspettative per la gestione dei dati e assicurando che tutte le parti interessate siano allineate sugli obiettivi comuni e sulle pratiche migliori per il trattamento dei dati.

I metodi includono le tecniche e gli strumenti per implementare e gestire il framework. Questi includono policies per regolare vari aspetti della gestione dei dati, e guide dettagliate su come implementare le politiche di data governance, tra cui best practices e procedure step by step.

I ruoli e le responsabilità devono essere chiaramente definiti per ciascun soggetto della Data Governance, per attribuire chiaramente le aspettative di ogni soggetto relative ai processi della data governance.

Per ultimo, il framework comprende l'insieme di strumenti e piattaforme utilizzate per permettere e implementare le attività di Data Governance.

Promuovere una cultura orientata ai dati, in cui tutti i dipendenti comprendono l'importanza della gestione dei dati e sono responsabili del trattamento corretto dei dati, è fondamentale per il successo del framework.

Tutti questi elementi verranno analizzati in profondità contestualizzando il framework implementato durante l'esperienza di tirocinio.

6.4. I MODELLI DELLA DATA GOVERNANCE

I modelli operativi della data governance definiscono il modo in cui le attività di gestione dei dati vengono organizzate, gestite e coordinate all'interno di un'organizzazione.

I tre principali modelli operativi sono il centralizzato, il replicato e il federato (Dama International, 2017). La scelta del modello operativo dipende da diversi fattori, tra cui la dimensione dell'organizzazione, la complessità del suo ambiente dati, la cultura aziendale e i requisiti normativi. Alcune organizzazioni possono preferire un approccio centralizzato per massimizzare il controllo sui dati e garantire l'aderenza alle politiche aziendali, mentre altre possono optare per un modello federato per favorire la flessibilità e l'innovazione.

- **Modello Centralizzato:** in questo modello, la responsabilità della governance dei dati è concentrata in un'unica entità o dipartimento all'interno dell'organizzazione. Questo approccio offre un maggiore controllo e coerenza nella gestione dei dati, ma può rallentare i processi decisionali e limitare la flessibilità.
- **Modello Replicato:** nel modello replicato, la responsabilità della governance dei dati è distribuita tra diverse entità all'interno dell'organizzazione. Ogni divisione o dipartimento gestisce i propri dati in modo indipendente, ma rispetta le linee guida e le politiche aziendali. Questo modello favorisce la decentralizzazione e la personalizzazione, ma può causare duplicazione dei dati e inconsistenza nelle pratiche di gestione.
- **Modello Decentralizzato:** approccio in cui il controllo e la gestione dei dati sono distribuiti tra più entità o partecipanti anziché essere centralizzati in un'unica autorità. Questo modello cerca di affrontare le sfide legate alla privacy, alla sicurezza e alla proprietà dei dati, consentendo a più stakeholder di contribuire alla governance dei dati.

Di seguito, analizzando i principi del Data Mesh, vedremo come per molte organizzazioni l'approccio vincente sia uno che combina elementi del modello centralizzato e del modello decentralizzato, in cui diverse entità all'interno dell'organizzazione mantengono un certo grado di autonomia nella gestione dei dati, ma collaborano e condividono risorse per garantire coerenza e allineamento con gli obiettivi aziendali.

7. IL DATA MESH

Il Data mesh è una nuova metodologia per l'acquisizione, la gestione e l'accesso ai dati analitici su larga scala, specialmente in ambienti complessi e di grande dimensione, sia all'interno che tra le organizzazioni. Questo approccio si concentra sull'utilizzo efficace dei dati analitici, fondamentali nei casi d'uso predittivi o diagnostici, per generare insight per le aziende e addestrare modelli di machine learning per migliorare le decisioni basate sui dati.

La necessità di un nuovo paradigma come il data mesh nasce da diversi fattori. In primo luogo, i modelli tradizionali di governance dei dati spesso faticano a scalare e a adattarsi all'aumento del volume, della varietà e della velocità dei dati generati all'interno delle organizzazioni (come, per esempio, per la gestione dei Big Data).

In secondo luogo, la diffusione delle architetture distribuite, del cloud computing e dei data lake ha decentralizzato l'archiviazione e l'elaborazione dei dati, sfidando il controllo centralizzato tipico dei modelli di governance tradizionali. Questa decentralizzazione richiede un approccio più flessibile e scalabile alla governance che possa accomodare diverse fonti e formati di dati.

Inoltre, la crescente domanda di analisi e insights in tempo reale richiede un approccio alla governance dei dati più agile e collaborativo. Il data mesh enfatizza l'importanza della collaborazione tra esperti di dominio, ingegneri dei dati e scienziati dei dati per garantire che i dati siano gestiti, governati e utilizzati in modo efficace per guidare i risultati aziendali.

La metodologia del data mesh prevede quattro principi chiave che le organizzazioni devono adottare per implementare al meglio i temi della Data Governance analizzati in precedenza. L'allegato 11 descrive semplicemente le relazioni tra i principi.

7.1. I PRINCIPI DEL DATA MESH

7.1.1. DOMAIN OWNERSHIP

Il principio di domain ownership rappresenta un approccio fondamentale che sottolinea la decentralizzazione e la distribuzione delle responsabilità dei dati alle persone più vicine ai dati stessi. Questo principio è stato concepito con l'obiettivo di supportare una struttura scalabile e consentire cicli di cambiamento e miglioramento continui e rapidi.

In questo contesto, la decomposizione dei dati e la loro integrazione sono guidate dai confini delle unità organizzative, anziché dalle soluzioni tecnologiche sottostanti. Ciò significa che i confini dei domini dei dati seguono le divisioni di responsabilità allineate con l'attività aziendale anziché con la struttura tecnologica. Questo principio si contrappone alle architetture tradizionali dei dati, che tendono a suddividere i dati in base alla tecnologia e a concedere la proprietà dei dati a team specifici, come il team del data warehouse o il team del data pipeline. Al contrario, nel data mesh, la responsabilità della condivisione dei dati viene affidata a ciascun dominio aziendale, poiché ogni dominio è considerato il primo utilizzatore o il punto di origine dei dati di cui è responsabile.

Questo approccio, si basa sull'idea di ownership del dato-prodotto (secondo principio) nel lungo termine, che implica la creazione, la modellazione, la manutenzione, l'evoluzione e la condivisione dei dati come prodotto per soddisfare le esigenze degli utenti dei dati. Inoltre, va sottolineato che il concetto di ownership dei dati in questo contesto si limita alla responsabilità delle organizzazioni nel mantenere la qualità, la longevità e l'accessibilità legale dei dati che generano.

7.1.2. DATA AS A PRODUCT

Fino ad ora abbiamo definito i dati come un asset aziendale strategico. Il principio di "data as a product" rappresenta uno step ulteriore nell'approccio alla gestione dei dati all'interno delle organizzazioni. L'obiettivo principale è quello di garantire che i dati siano trattati con la stessa attenzione e rigore riservati ai prodotti e ai servizi tradizionali, riconoscendo il loro valore strategico per l'organizzazione.

In primo luogo, il principio di "data as a product" sottolinea l'importanza di considerare i dati come un asset aziendale critico, che richiede una gestione proattiva e mirata. Questo approccio implica la definizione chiara dei requisiti di qualità, sicurezza e accessibilità dei dati, nonché la designazione di proprietari di prodotto responsabili della cura e dell'evoluzione dei dati all'interno dei rispettivi domini aziendali.

Inoltre, il concetto di "data as a product" enfatizza l'importanza di trattare i dati come un bene destinato all'uso e al consumo da parte di una vasta gamma di stakeholder aziendali. Ciò implica la necessità di sviluppare e mantenere un'infrastruttura tecnologica e operativa che supporti l'erogazione efficiente e affidabile dei dati a tutti coloro che ne hanno bisogno, sia internamente che esternamente all'organizzazione.

Un altro aspetto cruciale di questo principio è la promozione di una cultura organizzativa orientata ai dati, che incoraggi la collaborazione trasversale e l'innovazione nell'uso dei dati per supportare le decisioni aziendali e generare valore aggiunto.

Questo principio rimarca inoltre l'importanza di adottare un approccio iterativo e orientato al cliente²⁴ nello sviluppo e nell'evoluzione dei dati. Ciò significa che i dati devono essere costantemente valutati e migliorati in base al feedback degli utenti e alle esigenze aziendali emergenti, garantendo così che rimangano sempre rilevanti e utili nel contesto operativo dell'organizzazione.

7.1.3. SELF-SERVE DATA PLATFORM

Il principio della "self-serve data platform" mira a gestire come le organizzazioni affrontano l'accesso e l'utilizzo dei dati. Si basa sull'idea di fornire agli utenti un accesso diretto e semplificato ai dati di cui hanno bisogno, eliminando la dipendenza da team centralizzati e riducendo i tempi di attesa per l'ottenimento delle informazioni necessarie. Anche questo principio stressa l'importanza di una governance decentralizzata dei dati.

Questo principio è fondamentale per favorire l'agilità operativa e l'innovazione all'interno dell'organizzazione, consentendo agli utenti di esplorare, analizzare e utilizzare i dati in modo autonomo e tempestivo.

In primo luogo, il concetto di "self-serve data platform" si concentra sull'implementazione di infrastrutture tecnologiche e strumenti che consentano agli utenti di accedere e manipolare i dati in modo intuitivo e sicuro. Ciò include la creazione di portali self-service, dashboard interattive, e strumenti di interrogazione dei dati che semplifichino il processo di scoperta e utilizzo delle informazioni.

²⁴ Inteso come l'utilizzatore del dato

Ne consegue l'importanza dell'adozione di approcci basati su metadati e cataloghi dati che consentano agli utenti di scoprire e comprendere meglio la disponibilità e l'utilizzo dei dati all'interno dell'organizzazione.

7.1.4. FEDERATED COMPUTATIONAL GOVERNANCE

Il principio della "federated computational governance" implica una distribuzione delle responsabilità di governance dei dati e dei processi computazionali tra le diverse unità organizzative. Questo principio è nato per trovare un posizionamento ottimale tra i due estremi di un modello di governance totalmente centralizzato e quello decentralizzato.

In sostanza, anziché imporre una funzione di governance centralizzata con comando e controllo senza tenere conto delle sfumature di ciascun dominio, il paradigma Data Mesh sposa un modello di governance federata. Ciò significa bilanciare le esigenze e le conoscenze specifiche del dominio con i requisiti generali dell'azienda per produrre una divisione delle responsabilità chiaramente articolata e concordata.

In genere, tuttavia, l'IT centrale si occuperà della reportistica sul lignaggio dei dati, degli standard inter-domini, dell'autenticazione e dell'articolazione degli standard e delle politiche globali di rischio e conformità. I domini, quindi, saranno responsabili della qualità dei dati (sia per la definizione che per la misurazione), della provenienza dei dati, dell'autorizzazione, della classificazione dei dati, dell'adesione agli standard di conformità e terminologici e della definizione della standardizzazione delle entità dei dati tra i domini.

8. L'INDUSTRIA DEI MICROPROCESSORI

8.1. PREMESSA

L'obiettivo della prima parte dell'elaborato era fornire al lettore un contesto generale sui dati, sul data management e sulla data governance, al fine di facilitare la comprensione di questi concetti applicati durante l'implementazione di un progetto di data governance in un'azienda che produce e commercializza microprocessori.

Dopo aver esaminato in dettaglio i vari tipi di dati, le loro caratteristiche e il loro valore, nonché i vantaggi derivanti dall'implementazione di processi di gestione e governo dei dati, è necessario fornire un contesto adeguato sull'industria dei microprocessori, le tendenze del settore e i potenziali sviluppi futuri. Questo sarà utile per elaborare un'analisi approfondita sia del settore e dei principali fattori che lo influenzano, sia di come l'implementazione del progetto contribuirà a ridurre i rischi legati alla gestione dei dati e a migliorare l'efficienza e l'efficacia dei processi aziendali.

8.2. OVERVIEW DELL'INDUSTRIA

L'industria dei microprocessori²⁵ costituisce una componente essenziale dell'evoluzione tecnologica contemporanea. Un microprocessore è un circuito integrato che contiene la funzione di unità centrale di elaborazione (CPU) di un computer su un singolo chip. Questo componente svolge operazioni aritmetiche, logiche e di controllo, rendendo possibile l'esecuzione di programmi complessi, tra cui sistemi operativi, applicazioni software e multimediali.

Spesso i termini semiconduttore e chip vengono utilizzati in modo intercambiabile, ma non sono la stessa cosa. Secondo l'OCSE²⁶, i semiconduttori possono essere distinti in due grandi categorie (OCSE, 2019): circuiti integrati (IC²⁷) e OSD²⁸ (chip optoelettronica, sensori e semiconduttori discreti). I chip (o IC) possono essere considerati un sottoinsieme dei semiconduttori (Allegato 12). Gli IC rappresentano la maggior parte delle vendite globali di semiconduttori (Rabobank, n.d.), rappresentando nel 2021 più dell'80% del volume di vendita globale (Allegato 13).

Oggi, i microprocessori sono fondamentali per il funzionamento di una vasta gamma di dispositivi, dai computer ai telefoni cellulari, dalle automobili ai dispositivi IoT, nonché alla base di tecnologie emergenti come l'intelligenza artificiale. I semiconduttori sono tecnologicamente complessi e richiedono livelli così elevati di investimenti in R&S e spese di capitale che esistono solo pochi fornitori di chip al mondo. La creazione di nuove fonti di approvvigionamento è quindi spesso proibitiva.

I chip possono assumere la forma di circuiti logici, di memoria, micro o analogici. Tutti questi chip, pur essendo utilizzati nei prodotti di consumo finali, hanno funzionalità e caratteristiche diverse. I circuiti logici, ad esempio, sono il cosiddetto cervello dell'informatica e funzionano in base a codici binari. I circuiti di memoria immagazzinano le informazioni necessarie per i calcoli. Il micro comprende microprocessori e microcomputer, come le unità di elaborazione centrale (CPU), le unità di elaborazione grafica (GPU) e le unità di elaborazione accelerata (APU) che combinano CPU e

²⁵ Anche definiti all'interno di questo elaborato come chip, o semiconduttori

²⁶ Organizzazione per la cooperazione e lo sviluppo economico

²⁷ Integrated Circuits

²⁸ On-screen Display

GPU. Infine, i circuiti integrati analogici trasformano i dati analogici, come le registrazioni vocali, in dati digitali. L'architettura dei microprocessori, ovvero il modo in cui sono progettati e organizzati internamente per eseguire le istruzioni e gestire i dati, influenza la sua potenza computazionale, l'efficienza energetica, la capacità di gestire carichi di lavoro specifici e la compatibilità con alcuni tipi di software.

Alla base del funzionamento dei microprocessori ci sono i transistor, un semiconduttore che amplifica o commuta i segnali elettronici.

8.3. I TRANSISTORS

I transistor sono costituiti da materiali come il silicio o il germanio che sono in grado di far scorrere la corrente elettrica in modo controllato.

In questo modo, un transistor può agire come un interruttore o un gate per i segnali elettronici. In pratica, questo significa che utilizziamo i transistor come interruttori elettronici che accendono o spengono i circuiti elettronici. Si tratta di una funzione di base che utilizziamo nei circuiti logici digitali, come quelli presenti nei computer, dove utilizziamo i transistor per rappresentare gli uni e gli zeri del codice binario.

Gli sviluppi tecnologici hanno permesso di ridurre con un ritmo costante la dimensione dei componenti dei transistor, conseguendo la progressiva miniaturizzazione degli stessi. Seguendo la prima Legge di Moore (Allegato 14), elaborata nel 1965 dal cofondatore dell'azienda Intel, il numero di transistor su un circuito integrato è quasi raddoppiato ogni due anni. Oggigiorno alcuni processori superano i 140 miliardi di transistor (AndandTech, 2023).

Bisogna considerare però che il rapido avanzamento tecnologico in questo settore ha fatto sì che negli ultimi anni, un numero sempre minore di aziende è stato in grado di tenere il passo nella corsa alla costruzione dei chip. I processi di progettazione e produzione sono diventati estremamente lunghi, complessi e costosi, richiedendo attrezzature e conoscenze sempre più specialistiche. Minore è la dimensione, maggiore è il costo del chip e il costo del processo produttivo. Prima di analizzare l'industria dei microprocessori e i maggiori players presenti è necessario approfondire il processo di produzione dei microchip, contestualizzato all'interno della supply chain dei semiconduttori.

8.4. SUPPLY CHAIN E PROCESSO DI PRODUZIONE DEI CHIP

La catena di fornitura dei semiconduttori comprende una complessa rete di aziende, organizzazioni e individui coinvolti nella progettazione, produzione, collaudo, imballaggio e distribuzione dei semiconduttori.

Infatti, la produzione di un singolo chip per computer richiede spesso più di 1.000 passaggi (CSET, 2021), passando attraverso i confini internazionali 70 o più volte prima di raggiungere il cliente finale.

8.4.1. RICERCA E CONTESTO NORMATIVO

L'intensa attività di ricerca e sviluppo è necessaria per ideare nuovi design e migliorare le tecnologie esistenti. Da un lato, questo è necessario per soddisfare le necessità di mercato, dall'altro le aziende coinvolte in questo settore devono mantenere un'attenta vigilanza sul contesto normativo, poiché la

produzione e la distribuzione di semiconduttori sono soggette a regolamentazioni rigide e in continua evoluzione a livello globale. Queste regolamentazioni possono riguardare la proprietà intellettuale, la sicurezza dei dati, l'uso di sostanze chimiche pericolose, e le restrizioni commerciali tra paesi. Le aziende devono assicurarsi di rispettare tutte le normative locali e internazionali per evitare sanzioni e garantire una produzione etica e sicura.

8.4.2. PROGETTAZIONE E SVILUPPO

Le aziende produttrici di semiconduttori, dette anche “fabless”, progettano e sviluppano nuovi chip, in proprio o attraverso collaborazioni con altre aziende. Sono chiamate fabless perché non producono direttamente i chip che progettano e commercializzano. La progettazione comprende determinare le specifiche necessarie, effettuare la progettazione logica, la progettazione fisica e le fasi di validazione e verifica. Le specifiche dei chip determinano il loro funzionamento nel sistema che lo utilizza. La progettazione logica crea un modello schematico dei componenti elettrici interconnessi, mentre la progettazione fisica traduce questo modello in un layout fisico dei componenti elettrici e delle interconnessioni. La validazione e la verifica assicurano che i chip basati sul progetto funzionino come previsto.

8.4.3. FABBRICAZIONE

La produzione vera e propria dei chip per semiconduttori è affidata a fabbriche specializzate, chiamate fonderie. Le fonderie ricevono i progetti dei chip dalle aziende fabless e utilizzano apparecchiature personalizzate per creare e incidere i chip su un wafer di silicio. All'interno della fase di produzione, possiamo distinguere tra produzione front-end e back-end. La fase front-end include l'estrazione e purificazione del silicio, la crescita del cristallo e la creazione dei wafer. Il silicio viene estratto dalla sabbia e purificato, quindi fuso e raffreddato per formare lingotti cilindrici tramite un processo chiamato crescita del cristallo. I lingotti vengono poi tagliati in dischi sottili chiamati wafer e lucidati per ottenere una superficie uniforme e pulita. La fotolitografia è una parte essenziale della fase front-end, dove la luce ultravioletta e maschere fotografiche trasferiscono modelli di circuiti sui wafer di silicio. Strati sottili di vari materiali vengono depositati sui wafer e incisi secondo i modelli definiti dalla fotolitografia, formando i componenti e le interconnessioni del microprocessore. L'impiantazione ionica modifica le proprietà elettriche delle regioni specifiche del wafer.

La fase back-end comprende la stratificazione e interconnessione, il test e il confezionamento. I wafer passano attraverso numerosi cicli di stratificazione e modellazione per costruire i componenti finali, seguiti dal cablaggio metallico per completare le connessioni. Dopo la fabbricazione, i microprocessori vengono testati per assicurare la funzionalità e prestazioni, confezionati in alloggiamenti protettivi e spediti ai clienti per l'integrazione nei dispositivi elettronici.

Dopo la produzione, i chip devono essere testati per garantire il corretto funzionamento. Questo compito viene svolto da aziende specializzate nel collaudo e nell'assemblaggio. Una volta che i chip hanno superato i test richiesti, vengono assemblati in prodotti che possono essere utilizzati nell'elettronica e in altri dispositivi.

L'estrema precisione della fabbricazione di questi prodotti richiede processi eseguiti in ambienti puliti e protetti dalle vibrazioni. Qualsiasi particella, anche più piccola di un batterio, potrebbe compromettere seriamente la produzione di interi lotti di chip. In questo senso l'intervento umano è sempre meno richiesto, e l'automazione dei processi produttivi svolge un ruolo molto importante.

Attualmente gran parte dell'industria statunitense dei semiconduttori produce attrezzature per la fabbricazione piuttosto che fabbricare i semiconduttori stessi (Digikey, 2023). L'automazione è associata anche alle operazioni di manipolazione di prodotti chimici, chip e wafer e all'uso di robot per camere bianche prodotti da produttori come KUKA Robotics. Questi ultimi svolgono un ruolo importante nella riduzione delle perdite causate da errori umani.

Nella produzione di semiconduttori, l'automazione è spesso più legata all'elaborazione dei dati e all'automazione delle decisioni che ne derivano. Le fabbriche utilizzano algoritmi automatizzati per il controllo avanzato dei processi (APC) e per il controllo statistico dei processi (SPC). Questi sistemi tengono traccia delle variazioni di processo e dei difetti di produzione che ne derivano, da ridurre attraverso il controllo in tempo reale dei processi produttivi. Tali sistemi possono impiegare l'intelligenza artificiale e l'apprendimento automatico per identificare modelli all'interno di serie di dati molto ampie che tracciano molti parametri di processo e metriche di qualità.

8.4.4. APPLICAZIONE E DISTRIBUZIONE

i chip confezionati vengono poi distribuiti alle aziende che li vendono agli utenti finali o li incorporano nei loro prodotti, la vendita avviene direttamente o tramite distributori.

8.5. IL MERCATO DEI MICROPROCESSORI

Il mercato dei microprocessori è in crescita grazie all'aumento della domanda di calcolo ad alte prestazioni e all'incremento dell'uso di applicazioni IoT, 5G e AI, che favoriscono la crescita del mercato. Grazie alla loro velocità, alle dimensioni compatte e alla facilità di manutenzione, i microprocessori sono componenti essenziali dell'elettronica di consumo, come desktop, PC, smartphone, tablet e server (Grandview Research, 2023). L'allegato 15 mostra le principali applicazioni dei microprocessori.

I continui progressi nell'efficienza energetica e nella tecnologia dei semiconduttori giocano un ruolo significativo, mentre la concorrenza di mercato e la continua innovazione spingono il mercato a crescere. Di conseguenza, il mercato dei microprocessori è in costante crescita e fornisce la potenza di elaborazione essenziale per l'informatica moderna e le tecnologie emergenti.

Il mercato globale dei microprocessori è stato valutato a intorno ai 110 miliardi di dollari nel 2023 e si prevede che raggiungerà i 185 miliardi di dollari nel 2031, con un CAGR del 6,44% per il periodo previsto tra il 2024 e il 2031 (Markes&Data, 2023).

Il settore dei microprocessori è cruciale non solo per le aziende coinvolte, ma anche per i paesi che competono per il suo controllo. I microprocessori sono infatti fondamentali per la sicurezza nazionale, l'economia e l'innovazione tecnologica, rendendo il dominio in questo campo una priorità strategica per molte nazioni. La capacità di progettare e produrre microprocessori avanzati determina la competitività globale e l'autosufficienza tecnologica di un paese.

L'Asia (Taiwan, Cina, Giappone e Corea del Sud principalmente) detiene la quota maggiore del mercato grazie alla crescente penetrazione di smartphone e altri dispositivi elettronici (PIIE, 2022). Il Nord America detiene una quota considerevole del mercato globale dei microprocessori grazie alla presenza di leader di mercato negli Stati Uniti e in Canada (Precedence Research, 2022).

A causa dell'elevato grado di specializzazione e dei processi produttivi ad alta intensità di capitale, non sono molte le aziende coinvolte nella progettazione e nella produzione di chip. È interessante

notare come nel 2020 (Statista, 2023), circa il 70% delle quote globali fossero in mano a 8 aziende produttrici (in ordine, Intel, TSMC, Qualcomm, SK Hynix, Broadcom, Samsung, Nvidia, Sony, Micron). A maggio del 2024, cinque fra queste otto sono ancora le più grandi aziende di semiconduttori per capitalizzazione di mercato (Companies Market Cap, 2024).

8.6. PESTEL ANALYSIS

Al fine di riassumere i punti precedenti e fornire un'analisi comprensiva dell'industria dei microprocessori, è stata elaborata un'analisi PESTEL²⁹ del settore, ovvero un framework strategico che descrive i fattori macro-ambientali che influenzano e caratterizzano un particolare settore.

8.6.1. POLITICO

Le politiche governative relative a sovvenzioni, tariffe e regolamentazioni possono influenzare significativamente l'industria dei microprocessori.

Da un lato, le leggi sulla privacy dei dati e le normative sulla sicurezza informatica richiedono alle aziende di implementare sistemi robusti di protezione dei dati. Dall'altro, incentivi fiscali e sussidi per la ricerca e sviluppo (R&D) possono stimolare l'innovazione e la produzione di microprocessori.

Ne è un esempio l'UE Chips Act (European Parliamentary Research Service, 2022. European Commission, 2019), entrato in vigore nel settembre del 2023, che ha destinato oltre 43 miliardi di euro di investimenti pubblici e privati per sostenere lo sviluppo di capacità tecnologiche e innovazione su larga scala che permettano di raggiungere una capacità produttiva pari al 20% della capacità mondiale di semiconduttori entro il 2030. Inoltre, il piano prevede di affrontare la carenza di competenze, attrarre nuovi talenti e sostenere l'emergere di una forza lavoro qualificata.

Allo stesso modo, il Dipartimento del Commercio americano sta supervisionando 50 miliardi di dollari per rivitalizzare l'industria dei semiconduttori statunitense, di cui 39 miliardi di dollari in incentivi per i semiconduttori (The White House, 2022).

Un altro fattore importante è l'impatto delle guerre commerciali. Le tensioni commerciali tra paesi, come la guerra commerciale tra USA e Cina, possono influenzare le catene di approvvigionamento e creare incertezza nel mercato. A partire dal 2022 gli Stati Uniti hanno iniziato a imporre controlli sulle esportazioni di aziende che progettano, producono e creano strumenti per chip di silicio avanzati, limitando le loro vendite alla Cina (BBC, 2022).

Queste tensioni possono portare a tariffe più alte, restrizioni sulle esportazioni e difficoltà nell'acquisizione di materie prime e componenti essenziali. Queste situazioni possono porre particolari problemi ad aziende come Nvidia, Broadcom, AMD, Intel, Taiwan Semiconductor, Samsung, Lam Research, KLA e Tokyo Electronics – che hanno ricevuto entrate sostanziali dalla Cina nel 2023 (New York Times, 2024).

²⁹ Political, Economic, Social, Technological, Environmental, Legal

8.6.2. ECONOMICO

Come menzionato la produzione di chip è dipendente da materiali come semiconduttori e metalli. I prezzi delle materie prime come silicio, rame e altri metalli influenzano significativamente i costi di produzione dei microprocessori. La volatilità dei prezzi delle materie prime può portare a margini di profitto variabili e influenzare la stabilità finanziaria delle aziende.

Il crescente utilizzo del silicio da parte di varie industrie, tra cui quella automobilistica e solare, contribuiranno a far crescere il mercato globale di quest'ultimo del 5,4% annuo nei prossimi 5 anni, passando da una produzione globale di 3,27 milioni di tonnellate a 4,25 milioni entro il 2029 (Mordor Intelligence, n.d.).

Gli alti investimenti in R&D sono cruciali per rimanere competitivi nel settore. Le aziende che investono di più in R&D possono sviluppare tecnologie avanzate e innovare più rapidamente, mantenendo una posizione di leadership sul mercato. Ne è un esempio Nvidia, che con la produzione del chip H100 ha conquistato una posizione di leader nel mercato.

Il costo del lavoro e la disponibilità di personale qualificato possono influenzare le operazioni e i costi delle aziende. Settori con una forte domanda di ingegneri e scienziati possono incontrare difficoltà nel reperire talenti, influenzando la capacità di innovare e produrre. Questo fattore è particolarmente importante nell'industria dei semiconduttori in cui "ognuna delle mille del processo di produzione di chip richiede "un livello di conoscenza da tesi di dottorato" come citato da a Jeff Koch (Financial Times, 2024), analista di SemiAnalysis una società di ricerca indipendente specializzata in semiconduttori e IA.

8.6.3. SOCIALE

Le tendenze nei comportamenti dei consumatori influenzano le strategie di produzione e sviluppo prodotto delle aziende: ad esempio la crescita della domanda di dispositivi elettronici avanzati, come smartphone, laptop e dispositivi IoT, stimola la domanda di microprocessori avanzati.

Questa domanda è alimentata dalla crescente digitalizzazione della società e l'aumento dell'uso di tecnologie avanzate stanno cambiando le dinamiche sociali e creando nuove opportunità di mercato. L'adozione di tecnologie come l'AI e l'IoT stimola la domanda di semiconduttori avanzati.

La disponibilità di programmi educativi e di formazione in ingegneria elettronica e informatica può influenzare la capacità di un paese di sostenere un'industria dei microprocessori. I paesi con forti programmi educativi possono attrarre investimenti e diventare hub tecnologici.

8.6.4. TECNOLOGICO

Le tendenze nei comportamenti dei consumatori influenzano le strategie di produzione e sviluppo prodotto delle aziende: ad esempio la crescita della domanda di dispositivi elettronici avanzati, come smartphone, laptop e dispositivi IoT, stimola la domanda di microprocessori avanzati.

Questa domanda è alimentata dalla crescente digitalizzazione della società e l'aumento dell'uso di tecnologie avanzate stanno cambiando le dinamiche sociali e creando nuove opportunità di mercato. L'adozione di tecnologie come l'AI e l'IoT stimola la domanda di semiconduttori avanzati.

La disponibilità di programmi educativi e di formazione in ingegneria elettronica e informatica può influenzare la capacità di un paese di sostenere un'industria dei microprocessori. I paesi con forti programmi educativi possono attrarre investimenti e diventare hub tecnologici.

8.6.5. AMBIENTALE

La produzione di microprocessori richiede processi ad alta intensità energetica e l'uso di sostanze chimiche pericolose, che possono avere impatti ambientali significativi. La fase di fabbricazione dei wafer è una delle fasi più redditizie dell'intera catena del valore dei semiconduttori (Ruberti, 2023), ma è anche la fase di produzione con il più alto impatto ambientale, poiché comporta l'utilizzo di enormi quantità di materie prime, acqua, energia e la generazione di rifiuti ed emissioni.

Le aziende produttrici di semiconduttori saranno sottoposte a crescenti pressioni per la decarbonizzazione, in particolare per sostenere la limitazione dell'aumento globale delle temperature a 1,5°C, come dichiarato nell'Accordo di Parigi (BCG, 2023). Inoltre, il 30% dei contratti con i fornitori degli OEM del settore automobilistico prevede già limiti alle emissioni di CO₂ e si prevede che tale percentuale raggiungerà l'80 per cento entro il 2030.

Nonostante ciò, la crescita della domanda di chip e le difficoltà del processo di produzione rendono però la decarbonizzazione del settore un obiettivo ancora molto lontano.

Infine, le barriere normative, come la direttiva dell'Unione Europea sulla rendicontazione della sostenibilità aziendale e le norme proposte dalla Securities and Exchange Commission degli Stati Uniti, sono state concepite per costringere le aziende a essere più trasparenti e precise sui dati relativi alle emissioni lungo la loro catena di approvvigionamento, come primo passo per lo sviluppo di piani di mitigazione.

8.7. L'IMPORTANZA DELLA DATA GOVERNANCE NELL'INDUSTRIA DEI MICROPROCESSORI

La data governance riveste un ruolo importante nell'industria dei microprocessori, un settore caratterizzato da una catena del valore complessa. Di seguito vengono analizzate le principali aree in cui la data governance può migliorare i processi aziendali e migliorare la sicurezza dei dati trattati.

L'esempio più immediato è la gestione accurata dei dati relativi ai clienti, agli ordini e alla fatturazione. Le aziende di microprocessori tendenzialmente si rivolgono ad un numero limitato di clienti, ma che spesso sono a loro volta aziende multinazionali.

Le aziende di microprocessori tendono a rivolgersi a un numero limitato di clienti, che spesso sono aziende multinazionali. La gestione dei dati di tali clienti è complessa poiché le diverse sedi fanno capo alla stessa azienda madre e vi possono essere differenze tra l'entità a cui viene emessa la fattura e quella a cui viene effettuata la spedizione di prodotti. Questo richiede un'accurata sincronizzazione dei dati per garantire che tutte le informazioni siano corrette e aggiornate, permettendo una gestione efficiente dei processi di fatturazione e spedizione.

Lo stesso vale per il tracciamento e la gestione delle forniture. Come analizzato, la catena di fornitura dei semiconduttori è complessa e coinvolge numerosi fornitori e passaggi logistici. Una gestione efficace dei dati può migliorare la visibilità lungo la catena di fornitura, permettendo di migliorare il

tracciamento dei componenti e dei materiali in tempo reale. Questo aiuta a prevenire interruzioni nella produzione, ottimizzare l'inventario e ridurre i costi operativi.

Un ulteriore aspetto di rilievo riguarda la gestione dei dati delle macchine industriali automatiche, le quali generano ingenti quantità di dati operativi utili per l'ottimizzazione dei processi produttivi e la previsione di eventuali guasti. Il processo produttivo si articola in una moltitudine di passaggi, ciascuno dei quali richiede la misurazione e il monitoraggio di specifiche variabili, come ad esempio la temperatura. Per trarre valore da questa imponente mole di dati, è imprescindibile che essi siano accuratamente gestiti e affidabili.

La gestione delle informazioni personali è un altro tema di rilevante importanza. Sebbene le aziende di microprocessori non debbano sottostare a criteri stringenti come quelli imposti ad esempio nel settore assicurativo, in cui il volume e la natura dei dati trattati sono particolarmente critici, è comunque necessario che esse siano conformi al GDPR e alle altre normative internazionali e locali, comprese quelle relative alla sicurezza dei dati e alla protezione ambientale. La data governance facilita la tracciabilità e la trasparenza, rendendo più semplice dimostrare la conformità alle normative durante gli audit. Inoltre, aiuta anche a mantenere la documentazione in ordine e accessibile, riducendo il rischio di sanzioni legali.

Per ultimo, come visto la protezione della proprietà intellettuale è uno dei temi più importanti. La ricerca e lo sviluppo nel settore dei microprocessori sono attività estremamente competitive e richiedono la protezione delle innovazioni tecnologiche.

Proteggere la proprietà intellettuale attraverso una gestione rigorosa dei dati garantisce che le informazioni sensibili siano adeguatamente protette e che i brevetti e i segreti commerciali siano salvaguardati da eventuali minacce interne ed esterne.

9. ESPERIENZA DI TIROCINIO – LA DATA GOVERNANCE E L'INDUSTRIA DEI MICROPROCESSORI

I seguenti paragrafi hanno l'obiettivo di contestualizzare l'esperienza di tirocinio svolta presso Accenture spa, in qualità di Technology, Strategy and Advisory Intern. In particolare, l'esperienza di tirocinio è stata svolta all'interno della business unit di Strategy & Consulting, nello specifico all'interno dell'industria High Tech.

L'intera esperienza di tirocinio è stata concentrata in un progetto di implementazione delle pratiche di data governance in un'azienda operante nel settore dell'industria dei microprocessori. La suddetta, per motivi di privacy e sicurezza delle informazioni, verrà denominata "l'azienda", la "compagnia" oppure "l'organizzazione". Inoltre, alcune informazioni specifiche e tecniche riguardo all'azienda saranno mascherate per non compromettere l'integrità della stessa e delle informazioni che la riguardano.

Gli obiettivi del progetto e dell'attività di tirocinio hanno una connotazione prettamente strategica, e si concretizzano in consegne al cliente, in diversi formati (Microsoft Word, Excel, Powerpoint), che hanno lo scopo di definire il framework di data governance nella sua interezza, nonché nell'implementazione e utilizzo del tool di data governance per supportare l'organizzazione a stabilire i metodi, i processi e le responsabilità necessarie per implementare un corretto governo dei dati.

9.1. IL CONTESTO PRIMA DEL PROGETTO E LE NECESSITÀ DELL'AZIENDA

Nel maggio 2023 l'azienda ha rivisto la propria organizzazione per dare maggiore attenzione alla Data Governance a livello aziendale per consentire il raggiungimento di obiettivi strategici (tra cui un obiettivo di fatturazione per l'anno 2027), tramite un miglioramento della gestione e governo dei dati.

La visione ultima del progetto è avere a disposizione dati affidabili che apportano valore alle decisioni strategiche, diventando così un'azienda più orientata ai dati. Per raggiungere questo obiettivo, l'organizzazione ha richiesto il supporto ad Accenture per i seguenti punti:

- **Stabilire la visione e la strategia:** l'azienda ha definito la visione e la strategia della Data Governance, scegliendo il modello operativo "Hub & Spoke" con l'obiettivo di passare in 2 anni da un modello operativo centralizzato a uno federato;
- **Definire e centralizzare le regole e le politiche comuni:** l'obiettivo dell'azienda è rendere la data governance un dominio aziendale trasversale alle altre funzioni, tramite la scelta di un modello operativo, un quadro di riferimento comune, regole centrali, politiche, standard e KPI;
- **Concretizzare la strategia definita:** attraverso un approccio pragmatico e incrementale per aree di business, l'obiettivo è stabilire best practices e processi che si possano integrare nella struttura aziendale esistente. L'obiettivo è identificare, prioritizzando sulle base delle necessità aziendali, dei casi d'uso su specifici domini aziendali, che diventeranno il modello da replicare per diffondere i processi della data governance a tutta l'organizzazione;
- **Diffondere ed evangelizzare la governance dei dati:** questa categoria comprende tutte le attività di formazione e familiarizzazione dei dipendenti dell'azienda riguardo alla data governance. Comprende la diffusione di una nuova mentalità relativa al trattamento dei dati, lo stabilimento di un piano di comunicazione efficace per diffondere i pilastri della data governance, piani di apprendimento e di condivisione delle conoscenze.

Per riassumere, l'obiettivo del progetto consiste nel supportare l'azienda a strutturare da zero i processi di data governance, assicurando la definizione dei ruoli e delle responsabilità, nonché i processi e le attività da svolgere per garantire un corretto governo dei dati. È importante sottolineare che l'azienda aveva già un sistema di gestione dei dati, principalmente basato su Master e Reference data, ma mancava di regole e processi comuni a tutta l'azienda che portava alle seguenti problematiche:

- Non erano presenti delle strutture interne all'azienda dedicate esclusivamente al corretto governo dei dati;
- Allo stesso modo, le figure e le persone con responsabilità nei confronti dei dati non erano definite seguendo un framework stabilito;
- Non era presente un'unica piattaforma che permetteva l'accesso e la comprensione di tutti i dati aziendali;
- Non era presente un sistema che permettesse l'accesso controllato ai dati rispetto a delle capacità assegnate a ogni utente;
- Non esistevano dei processi definiti relativi alla gestione dei metadati e ai temi più importanti riguardanti i dati, tra cui la loro privacy, sicurezza e qualità;
- I team di sviluppo IT possedevano tutta la conoscenza relativa ai dati, alla loro archiviazione e struttura. Di conseguenza l'accesso ai dati era possibile solo attraverso queste figure altamente specializzate;
- Non erano stabilite delle metriche riguardo alla gestione dei dati e dei metadati. In questa forma l'efficacia delle iniziative relative che necessitavano l'utilizzo di questi ultimi era difficile da misurare;
- Non era presente un framework per misurare la maturità dell'organizzazione rispetto alla data governance;
- Non era presente un piano strutturato per formare i dipendenti dell'organizzazione rispetto al corretto governo dei dati e all'importanza della data governance.

9.2. L'APPROCCIO DI ACCENTURE

L'approccio standard di Accenture per l'implementazione ed esecuzione di progetti di questo tipo, compreso delle fasi di Discovery, Design, Pilot, Plan, Scale, è stato dettagliato e cucito sulle necessità dell'azienda. Alcune delle fasi menzionate sono state sviluppate in parallelo, in quanto le singole attività appartenenti ad ogni fase, in alcuni casi, non richiedevano il completamento di fasi precedenti per essere svolte. Pertanto, di seguito sono menzionate le fasi, ma nel dettaglio saranno analizzati i risultati ottenuti dallo svolgimento delle attività.

Nell'allegato 16 è fornita una roadmap di alto livello dei documenti e delle attività da svolgere riguardo la definizione e centralizzazione delle regole e politiche comuni, la concretizzazione delle strategie definite e la diffusione ed evangelizzazione della governance dei dati. È possibile notare come non è presente la fase di identificazione della visione e della strategia in quanto è stato definito e approvato in fase di definizione del ruolo di Accenture nell'implementazione del progetto di data governance, ed è pertanto ulteriori azioni non sono necessarie.

9.2.1. DISCOVERY

La fase di Discovery comprende un insieme di attività destinate ad analizzare nel dettaglio lo stato attuale e il livello di maturità della governance dei dati, per identificare i casi d'uso critici che hanno

un impatto sul valore aziendale e sulle metriche di processo, partendo dai processi già esistenti nell'azienda. Questo comprende:

- **Allineamento con la strategia dei dati e analisi delle soluzioni IT:** l'allineamento è condotto attraverso workshop e riunioni dedicate. Lo scopo dell'attività è quello di analizzare principi, politiche, processi, metriche, strumenti tecnologici, ruoli e responsabilità inerenti alla gestione dei dati definiti e adottati nell'organizzazione;
- **Interviste con i domini interessati per identificare i principali bisogni, esigenze e mancanze:** utilizzando check list, documenti operativi e condivisi l'obiettivo è avere una chiara visione dei processi aziendali, quali sono consumatori e quali produttori di dati e il livello di conoscenza e alfabetizzazione aziendale riguardo ai dati e alle pratiche di data governance;
- **Long list di use cases e prioritizzazione:** attraverso le riunioni e workshop sopra menzionate è possibile definire la long list di use case in cui testare i nuovi framework e processi stabiliti.

9.2.2. DESIGN

La fase di Design comprende la definizione delle modalità operative e del processo di gestione dei dati. Tra questi è necessario definire i controlli di qualità dei dati, le regole e procedure comuni, nonché la definizione dei KPI aziendali da misurare.

Questo comprende:

- **Definizione del modello organizzativo:** il punto di partenza per stabilire l'approccio strategico inizia con la definizione dei principi, politiche e standard di utilizzo dei dati da applicare per garantire l'accessibilità, la sicurezza e l'affidabilità dei dati. L'insieme di questi elementi sarà contenuto all'interno del documento di Data Governance Framework;
- **Definizione del modello operativo:** In concomitanza, è necessario disegnare il modello operativo scelto, ovvero l'Hub&Spoke, attraverso l'identificazione dei principali ruoli coinvolti e delle loro responsabilità fondamentali. Queste fasi sono estremamente interconnesse in quanto il modello operativo influenza quello organizzativo e viceversa;
- **Metadata Management Process:** questo comprende l'identificazione dei processi di gestione dei dati e del ciclo di vita dei dati, nonché la mappatura dei ruoli del modello operativo target sui processi di gestione dei dati;
- **Controllo qualità dei dati tecnici:** questo comprende la definizione dei principali controlli di qualità dei dati tecnici e delle linee guida che devono essere eseguiti per i campi chiave, nonché la definizione degli altri processi fondamentali della data governance, che si basano su quello del metadata management.

9.2.3. PILOT

La fase di Pilot comprende il lancio di progetti pilota per testare nuove regole, processi e linee guida, implementazione di azioni specifiche sui casi d'uso identificati e la misurazione dei risultati ottenuti, per validare l'efficacia dei suddetti progetti e in seguito implementare il framework a livello aziendale.

Questo comprende:

- **Prioritizzazione degli use cases:** è necessario effettuare una selezione delle iniziative più rilevanti in base alla matrice di priorità e all'aderenza alle esigenze dell'organizzazione. La

short list è composta al massimo da 2 casi d'uso per ciascun dominio. A partire dalla long list, l'esecuzione di workshop con i domini interessati sarà fondamentale per analizzare in dettaglio i loro processi relativi alla creazione e consumo dei dati, eventuali punti critici, e determinare la loro complessità per scegliere dei casi d'uso che permettano uno sviluppo fattibile dei casi d'uso;

- **Lancio degli use cases:** lancio di 2 casi d'uso per ogni dominio in perimetro guidati dalla matrice di priorità. Sono proposte diverse attività da avviare:
 - Gestione dei dati: ad esempio, è necessario definire una serie di controlli funzionali e tecnici sulla qualità dei dati, progettazione di un PoC³⁰ di uno specifico strumento di gestione dei dati, definizione di nuove procedure di gestione dei dati e dei metadati su misura;
 - Revisione e miglioramento: è necessario stabilire i processi, strumenti, approccio, modalità di lavoro, documentazione, persone/organizzazioni, gestione del cambiamento;
 - Pulizia dei dati per stabilire i dati di base da cui partire: a partire dai dati esistenti, è necessario effettuare una pulizia di questi, eventualmente eliminando duplicati, dati non utilizzati o di poco valore;
 - Analitica: ad esempio, attraverso la realizzazione di un MVP³¹ di un cruscotto o dashboard per automatizzare un report esistente, e l'esecuzione di un'analisi dei dati per risolvere problemi specifici.

9.2.4. PLAN

La fase di Plan comprende le attività di definizione di una roadmap per adottare il framework comune in tutti i domini, per lanciare altre azioni, per affrontare altri casi d'uso/aree, e per raggiungere gli obiettivi aziendali.

Questo comprende:

- La definizione di un'ipotetica roadmap di trasformazione per l'adozione dei nuovi strumenti di Data Governance individuati nella fase di "Design" e per l'adozione dei processi del nuovo modello operativo target, definito anche Maturity Framework della data governance;
- La definizione di una roadmap di implementazione per i casi d'uso selezionati, compresa la valutazione delle dipendenze tra le attività e degli impatti sull'architettura IT

9.2.5. SCALE

La fase di Scale comprende il rilascio e industrializzazione delle nuove "best practice ST" a livello aziendale. In questo momento del progetto questa fase non sarà analizzata in quanto, al fine di poter stabilire dei modelli di processi e best practices da replicare per diffondere i processi della data governance a tutta l'organizzazione, è necessario che questi vengano testati e approvati.

³⁰ Proof of Concept

³¹ Minimum Viable Product

9.3. ROADMAP ATTUALE

L'allegato 17 rappresenta lo stato attuale del progetto. Come menzionato nell'introduzione di questo capitolo, in questa fase del progetto il supporto di Accenture si traduce principalmente nella redazione di documenti (di diverso formato) che aiutino gli utenti di business dell'organizzazione a stabilire formalmente i processi della data governance. Questi documenti hanno un duplice scopo: da un lato diventare il punto di riferimento per gli utenti che si interfacciano con la data governance per comprendere il framework e i processi definiti, e dall'altro, senza dei documenti presentabili e consultabili, il coinvolgimento delle figure dirigenziali sarebbe impossibile.

In verde sono rappresentate le attività che sono state completate e le milestone raggiunte. In particolare, saranno analizzato di seguito il contenuto dei documenti elaborati per definire:

1. Il framework di data governance adottato
2. Il modello operativo identificato
3. I ruoli definiti e le loro responsabilità
4. I processi della data governance definiti
5. Dettaglio del processo di metadata management
6. Le metriche dei processi stabiliti
7. Il framework di maturità della data governance stabilito

In giallo sono indicate le attività ancora in corso o in fase di definizione. Il progetto attualmente si sta sviluppando attraverso le fasi di Discovery e Design.

9.4. IL FRAMEWORK DI DATA GOVERNANCE ADOTTATO

9.4.1. STABILIRE LA VISIONE E LA STRATEGIA

Durante la fase di definizione della visione e del ruolo strategico della data governance è stato evidenziato che per l'azienda, che possiede una ricchezza di informazioni in costante aumento, sia fondamentale sfruttare i dati per trarre vantaggio dal loro utilizzo, aumentando l'efficienza dei processi aziendali, supportando le strategie aziendali e abilitando nuovi servizi in diverse aree dell'azienda. Di conseguenza, la Data Governance viene riconosciuta come un elemento fondamentale per consentire una gestione appropriata e affidabile, la protezione e lo sfruttamento di uno degli asset più preziosi dell'azienda: i dati.

Il documento di Data Governance Framework è stato progettato per fornire un approccio olistico strutturato alla gestione e organizzazione degli asset di dati, mantenere la qualità dei dati, garantire sicurezza, privacy e conformità, nonché assicurare l'accesso, l'uso, l'ubicazione, la conservazione e l'archiviazione dei dati.

Per garantire i vantaggi e i benefici di una gestione dei dati adeguata e strutturata, il Data Governance Framework (allegato 18) è articolato in Organizzazione e Ruoli, Processi e Strumenti & Tecnologie.

Inoltre, senza regole precise e concordate, la realizzazione della supervisione della data governance è impossibile; pertanto, è necessario un insieme di politiche di data governance. Infine, per monitorare il funzionamento del framework di data governance e la sua adozione, è essenziale definire metriche di valutazione, sia qualitative che quantitative, che permetteranno di correggere la direzione intrapresa, se necessario. L'intero funzionamento dei processi di data governance deve

basarsi su una serie di principi, già menzionati all'interno della tesi, che servono da linea guida per la definizione del Framework stesso:

- **Considerare i dati come Asset Strategico:** i dati non sono solo un sottoprodotto dei processi aziendali, ma un asset vantaggioso che può guidare l'innovazione, l'efficienza e la competitività;
- **Innovazione basata sui dati:** è fondamentale promuovere una cultura dell'innovazione in cui nuove idee e lo sviluppo di prodotti sono basati su approfondimenti derivati dai dati e sulle tendenze di mercato;
- **Allineamento al Business:** il Framework di Data Governance deve essere allineato con gli obiettivi strategici dell'azienda, in particolare in termini di innovazione, sviluppo di prodotti e competitività sul mercato.
- **Proprietà dei dati e responsabilità aziendale:** delineare i ruoli e le responsabilità dei principali stakeholder nella gestione e protezione degli asset di dati critici e di alto valore permette anche di definire chiaramente la proprietà dei dati e le strutture di responsabilità all'interno di ST
- **Cultura e consapevolezza dei dati:** la strategia aziendale deve sposare la Cultura dei Dati attraverso corsi di formazione, workshop e iniziative di sensibilizzazione per educare dipendenti e stakeholder sull'importanza degli asset di dati e della loro governance, sui loro ruoli e sulle migliori pratiche da adottare.

9.4.2. ORGANIZZAZIONE E RUOLI

IL MODELLO OPERATIVO – HUB & SPOKE

Prima dell'implementazione del progetto di data governance, l'azienda basava la gestione dei propri dati attraverso una governance centralizzata.

Con l'obiettivo di iniziare la transizione verso un modello federato, l'azienda ha identificato l'Hub & Spoke come modello organizzativo per gestire i processi di data governance (allegato 19). Si tratta di una struttura organizzativa federata in cui un'autorità centrale, il Data Governance Office (il "Hub"), stabilisce e applica regole e politiche di governance dei dati, standard e best practice. Gli spokes rappresentano invece i singoli domini aziendali che gestiscono e utilizzano i dati.

L'Hub funge da punto centrale di controllo e supervisione delle attività di governance dei dati assicurando che tutte le pratiche relative alla governance dei dati siano allineate con gli obiettivi e i requisiti di conformità dell'organizzazione. Gli spokes hanno autonomia nella gestione delle loro specifiche esigenze e processi di dati, pur dipendendo dall'hub centrale in termini di politiche e standard di data governance.

Il modello si basa su una stretta collaborazione tra l'Hub centrale e gli spokes. Il primo fornisce indicazioni, supporto e competenze, mentre gli ultimi possiedono e conoscono il dominio aziendale specifico e implementano le pratiche di data governance nelle loro operazioni quotidiane. Questo consente di operare con elevata flessibilità e personalizzazione, poiché le diverse unità aziendali possono avere requisiti specifici sui dati e un diverso livello di maturità della governance dei dati.

Per ultimo, l'hub comunica le politiche e gli standard agli spokes e fornisce formazione e supporto per garantire la comprensione e la conformità.

FRAMEWORK ORGANIZZATIVO

Sono stati istituiti due organi di presidenza distinti, creati strategicamente per fornire una chiara leadership, autorità decisionale e supervisione delle politiche e delle procedure di Data Governance, allineando i loro sforzi con gli obiettivi e i requisiti più ampi dell'azienda:

- **Data Governance Office:** fornisce una funzione centralizzata per soddisfare le esigenze aziendali di governance dei dati e le capacità di gestione in modo disciplinato e coerente, facilitare la convalida delle informazioni e la proprietà dei dati e facilitare l'identificazione, l'implementazione e l'allineamento con l'azienda di strategie, politiche, procedure e standard sui dati;
- **Steering Committee:** questo team di leadership con funzioni eterogenee (composto da rappresentanti delle unità di business, dell'ufficio di Data Governance, delle funzioni aziendali e delle funzioni IT) è stato creato per fornire direzione e supervisione all'implementazione di processi, politiche, procedure e standard comuni della data Governance. Questo gruppo avrà una supervisione a livello aziendale e avrà la responsabilità e l'autorità di prendere decisioni e risolvere potenziali conflitti per l'azienda.

Di seguito, vengono i ruoli principali per la governance dei dati nell'organizzazione, categorizzanti in ruoli all'interno dell'ufficio di Data Governance, come parte dell'Hub, e ruoli all'interno dei domini di business, indicati come spokes nel quadro organizzativo precedentemente delineato. Il Data Governance Office è composto da due tipi di ruoli principali:

- **Head of Data Governance:** è un dirigente aziendale responsabile della supervisione della governance dei dati, con particolare attenzione alle politiche, ai processi e alle procedure per garantire la qualità, la sicurezza e l'efficacia della gestione dei dati aziendali. Collaborando con altri dirigenti, lavora per promuovere una governance dei dati coesa in tutta l'organizzazione.
- **Data Governance Expert:** implementa le personalizzazioni nello strumento di governance dei dati e fornisce aggiornamenti regolari sullo stato di salute dei processi di governance dei dati. Questo ruolo garantisce la coerenza e l'efficienza delle pratiche di governance dei dati in tutta l'organizzazione.

Al momento il DGO è composto da un Head e da quattro DGE, con cui il team di Accenture si interfaccia regolarmente per monitorare lo stato dell'avanzamento lavori. La nomina del DGO è stato il primo passo per l'implementazione dei processi di data governance, in quanto, prima di questo progetto, non esisteva un'autorità centrale che definisse l'approccio strategico riferito ai dati. I dati aziendali erano gestiti principalmente dal dipartimento IT che aveva il compito di identificare le principali soluzioni per l'archiviazione, per gestire pipeline di dati, e per garantire, attraverso figure come data analysts e data scientists, che questi dati fossero disponibili e utilizzabili nel momento del bisogno. L'intervento di dipartimenti esterni era, come ad esempio quello legale, era necessario per identificare quelle situazioni in cui venivano gestiti dati sensibili o PII, riservando particolare attenzione al loro trattamento. Nonostante ciò, non era previsto un framework e un piano strategico per fare in modo che sistematicamente nell'azienda venissero applicate delle best practices per il trattamento dei dati.

RUOLI IT

All'interno dell'Hub, come punto di riferimento per questioni tecniche relative al trattamento dei dati, sono state definite dei nuovi ruoli.

- **Solution Architect:** è un professionista che progetta e supervisiona l'implementazione di soluzioni strategiche, traducendo i requisiti in strutture architettoniche efficienti, colmando il divario tra dettagli tecnici e obiettivi aziendali;
- **Data Technical Architect:** è un professionista specializzato nell'architettura dei dati, che sfrutta le competenze tecniche per supportare il Solution Architect nell'implementazione delle attività di gestione del repository. Questo ruolo si concentra sulla progettazione di sistemi efficienti di archiviazione e recupero dei dati all'interno di quadri di soluzione più ampi, garantendo l'integrità dei dati e l'ottimizzazione dei processi;
- **Platform Admin:** è un ruolo essenziale responsabile dell'assegnazione e del monitoraggio dei privilegi di accesso ai dati, della garanzia di stabilità e della disponibilità delle risorse all'interno del catalogo di dati identificato dall'azienda. Questo amministratore supervisiona le autorizzazioni degli utenti, mantenendo la piattaforma di catalogazione di dati sicura e ben organizzata per facilitare la sua gestione, sostenendo l'integrità dei dati e l'affidabilità del sistema.

RUOLI DI BUSINESS – DATA OWNERSHIP E DATA STEWARDSHIP

All'interno dei domini, i principali ruoli che hanno la responsabilità per il funzionamento dei dati e per implementare le politiche della data governance sono il data owner e il data steward.

La proprietà dei dati (Data Ownership) si riferisce all'obbligo e alla responsabilità per i dati all'interno di un'organizzazione. Assegnare la proprietà dei dati assicura che ci sia un individuo o un team designato responsabile dell'accuratezza, della privacy, della sicurezza e della disponibilità dei dati.

A valle dell'identificazione della proprietà dei dati è necessario identificare i domini di appartenenza di questi ultimi, ovvero categorizzare i dati dell'organizzazione in domini logici o aree tematiche. Di solito i domini sono definiti a partire dalle aree funzionali aziendali (es. HR, Supply Chain, Sales). Tra le aree funzionali è importante individuare e scegliere come proprietari dei dati degli individui che abbiano le conoscenze e le competenze necessarie nei rispettivi domini dei dati. I proprietari dei dati sono infatti spesso manager o dirigenti senior che comprendono il valore strategico dei dati.

Il data owner è la più alta carica all'interno del dominio a cui fare riferimento, è il soggetto che sviluppa e fa rispettare le politiche e gli standard di governance dei dati.

Infatti, sebbene la Steering Committee definisca le strategie e politiche aziendali che devono essere utilizzate come linea guida per l'implementazione di processi relativi alla gestione dei dati dai vari domini, nella pratica ogni dominio ha le sue particolarità e specifiche. Non è sempre detto quindi che tutte le linee guida applicate dall'autorità di governance centrale possano riflettersi in best practices in tutti i domini. Il compito del data owner è proprio quello di gestire i dati specifici del suo dominio, adottando le politiche definite dal CDO e dalla Steering Committee e adattandola per quanto possibile.

L'identificazione della ownership dei dati è fondamentale per mantenere l'autonomia degli spokes, ovvero dei domini. Il data owner ha sufficiente autonomia per allocare le risorse disponibili per il dominio, nonché decidere come e dove investire per migliorare i propri processi di gestione dei dati. Questo permette di velocizzare i processi di provisioning dei dati, modellazione dei dati e distribuzione dei dati. Inoltre, avendo grande conoscenza del proprio dominio, ciascun data owner è responsabile della comunicazione con gli altri domini, al fine di avvicinare i produttori e consumatori dei dati.

Il data owner può anche far parte della Steering Committee, come membro votante o non votante. In questo modo il comitato può essere rappresentato in modo diversificato all'interno dell'organizzazione, per garantire che gli standard stabiliti siano adatti all'ampia gamma di esigenze dell'organizzazione e che modifiche specifiche ai singoli domini siano applicati il meno possibile per garantire consistenza e interoperabilità dei dati.

La Data Stewardship, invece, è un concetto essenziale nella gestione dei dati, poiché consiste nella supervisione e della cura dei dati all'interno di un'organizzazione. I data steward sono responsabili di garantire che i dati siano gestiti in modo accurato, sicuro e conforme alle normative.

Le direttive e linee guida imposti dal CDO, dalla Steering Committee e dai data owner è messa in pratica dai data steward, che hanno il compito di supervisionare l'intero ciclo di vita dei dati e dei loro metadati, garantendo la qualità e la conformità normativa.

I data steward sono utenti di business, anche responsabili di facilitare la collaborazione e la comunicazione tra i differenti domini aziendali. Questi professionisti tendenzialmente hanno ottime competenze analitiche e tecnologiche, e sono esperti del proprio dominio, ma non appartengono al dipartimento IT delle organizzazioni. Giocano un ruolo chiave soprattutto nella gestione dei metadati, supervisionando la realizzazione degli strumenti della Data Governance, tra cui soprattutto il business glossary.

Per ultimo, sono state identificate due figure tecniche che faranno parte, a fine del processo di transizione dal modello centralizzato a quello federato, degli spokes. Queste sono il data custodian e il data modeler. Il primo gestisce le strutture di dati per verificare la completezza dei requisiti tecnici, operando direttamente sul data dictionary. Il secondo invece ha il compito di allineare i requisiti di business stabilendo la definizione del modello semantico, nonché del diagramma entità-relazioni. Le attività di entrambe le figure saranno analizzate soprattutto all'interno del processo di Metadata Management.

9.4.3. TECNOLOGIE

Gli strumenti e le tecnologie si riferiscono ad applicazioni software, piattaforme e sistemi specificamente progettati e utilizzati per facilitare, automatizzare e migliorare vari aspetti delle pratiche e dei processi di governance dei dati. Questi strumenti e tecnologie vengono utilizzati per semplificare la gestione dei dati, garantire la qualità dei dati, applicare misure di sicurezza, monitorare la conformità e supportare le iniziative generali di governance dei dati all'interno dell'organizzazione.

La piattaforma principale che sarà utilizzata dall'azienda per fare ciò è Talend Data Catalog³², ovvero una piattaforma di catalogazione dei dati che offre un unico punto di controllo per i dati aziendali. All'interno del catalogo di dati troviamo infatti il business glossary, il data dictionary e il data lineage. Grazie a strumenti per la ricerca e l'individuazione e ai connettori per estrarre metadati da praticamente qualsiasi fonte di dati, il catalogo dei dati semplifica la protezione dei dati, il governo delle analisi, la gestione delle pipeline di dati e l'accelerazione dei processi ETL.

³² <https://www.talend.com/products/data-catalog/>

9.4.4. PROCESSI

Nel contesto della Data Governance, i processi si riferiscono alla serie strutturata e sistematica di attività, compiti e flussi di lavoro stabiliti per gestire, controllare e ottimizzare il ciclo di vita delle risorse di dati all'interno dell'organizzazione. I processi di governance dei dati sono progettati per garantire che i dati siano gestiti in modo coerente, sicuro e in linea con gli obiettivi organizzativi e i requisiti normativi.

Sono stati identificati otto processi principali per il raggiungimento degli obiettivi strategici dell'organizzazione. In particolare, il processo di Metadata Management pone le basi per tutti gli altri processi, che dipendono dal corretto utilizzo dei metadati.

Questo sarà analizzato con profondità in quanto è l'unico degli otto processi che è stato definito con precisione.

PROCESSO 1 – CLASSIFICAZIONE DEI DATI

La classificazione dei dati è l'attività di categorizzazione o etichettatura dei dati in base al loro contenuto, all'importanza, alla sensibilità e ai requisiti legali per gestirli e proteggerli meglio. L'azienda necessitava un sistema di classificazione per organizzare e proteggere i dati in base alla loro importanza e ai requisiti di accesso, consentendo una gestione più efficace dei dati, il controllo degli accessi e le misure di sicurezza. In questo contesto specifico, le risorse di dati possono essere classificate in quattro categorie: segrete, riservate, ristrette e pubbliche, in ordine dalla più alla meno protetta (allegato 20).

Le difficoltà di questa attività sono molteplici in quanto la classificazione dei dati deve avvenire non solo sulla base delle necessità aziendali interne, garantendo che le risorse di dati siano gestite, protette, utilizzate e condivise in modo appropriato all'interno dell'organizzazione. Infatti, tutti gli asset di dati devono essere classificati in base ai vincoli normativi/linee guida del Paese/Regione in cui il dato è utilizzato.

La complessità di essere un'azienda multinazionale in termini di governance dei dati risiede oltretutto nel fatto che i dati sono soggetti a diverse normative che possono variare significativamente tra loro. Di conseguenza, in base al luogo in cui i dati sono raccolti e utilizzati, possono sorgere discrepanze informatiche riguardo a come i dati vengono archiviati, se sono mascherati o meno, e altre modalità di gestione.

È compito dei data owners dei rispettivi domini, definire la classificazione dei dati per i dati rilevanti, revisionare e approvare i cambi che riflettono un'evoluzione delle necessità di business o per cambiamenti normativi. Questi avranno il supporto del corrispondente (o dei corrispondenti) data steward in quanto esperto delle procedure aziendali, che dovrà effettuare processi di audit per verificare la conformità con le Politiche e gli Standard.

PROCESSO 2 – QUALITÀ DEI DATI

Le iniziative di Qualità dei Dati non sono delle attività una tantum, ma un processo iterativo che coinvolge tutti gli aspetti del framework: persone, processi e tecnologie. Le dimensioni della qualità dei dati sono quelle definite dalla letteratura e menzionate precedentemente nel capitolo 6.1.2. In particolare, sono state identificate la completezza, l'accuratezza, l'aggiornamento, l'unicità, la consistenza e la validità.

Il Data owner definisce le regole di qualità dei dati del proprio dominio, mentre il data steward verifica la correttezza di queste, sia da un punto di vista tecnico che funzionale.

Inoltre, è necessario definire le metriche di Data Quality per l'analisi specifica della qualità dei domini di dati, KPI che saranno identificati in seguito. Oltre ad una visione strategica sulla qualità dei dati, l'azienda necessitava di strutture utili al monitoraggio della qualità dei dati.

PROCESSO 3 – SICUREZZA E PRIVACY DEI DATI

Nel contesto della governance dei dati, questi due concetti già precedentemente definiti lavorano in tandem. Le misure di sicurezza dei dati vengono implementate per proteggere i dati da minacce esterne e interne, mentre le pratiche di privacy dei dati garantiscono che i dati vengano raccolti ed elaborati in modo da rispettare i diritti e le aspettative di privacy delle persone.

Per quanto riguarda i principi di accesso dei dati, i data owner dei domini hanno l'obiettivo di identificare e classificare i dati, determinano i diritti di accesso in conformità con le politiche e i requisiti di sicurezza e privacy dei dati già in vigore presso l'azienda. Dal punto di vista operativo i data Stewards implementano i controlli di accesso, assicurando che agli utenti siano concessi solo i privilegi di accesso minimi necessari per i loro ruoli.

PROCESSO 4 – RESIDENZA DEI DATI

La residenza dei dati è la pratica strategica di specificare e designare con precisione la posizione geografica per l'archiviazione, l'elaborazione e il transito dei dati, in linea con i mandati legali e normativi. Ciò comporta un'attenta valutazione del luogo in cui si trovano i dati, del modo in cui vengono elaborati e della possibilità di trasferirli oltre i confini internazionali. La conformità a questi requisiti di residenza dei dati è essenziale per sostenere la governance, la privacy e la sicurezza dei dati, in quanto garantisce l'allineamento con i diversi quadri giuridici delle rispettive regioni o paesi. Lo stesso discorso affrontato per la completezza dei dati vale anche per la residenza di questi ultimi.

PROCESSO 5 – CONSERVAZIONE E ARCHIVIAZIONE

Queste politiche stabiliscono per quanto tempo i diversi tipi di dati devono essere conservati, quando devono essere smaltiti in modo sicuro perché non più utilizzati attivamente e infine eliminati. Queste politiche preservano i record storici, soddisfacendo i requisiti di conformità e liberando risorse di archiviazione primaria. Seguendo il ciclo di vita dei dati, devono essere definite delle procedure per stabilire:

- **Il periodo di conservazione dei dati** sulla base della classificazione dei dati e dei requisiti di business, garantendo un approccio equilibrato alla conformità legale, alla sicurezza dei dati, all'efficienza dei costi e ai requisiti aziendali in tutti gli stati dei dati (attivi, archiviati o programmati per l'eliminazione);
- **Le procedure di conservazione e cancellazione dei dati:** questo comprende sia i metodi che gli strumenti di archiviazione dei dati e dei metadati, sia i sistemi utilizzati per identificare gli asset di dati che sono attualmente in uso, archiviati o pronti per lo smaltimento;
- **I requisiti e tempi di ripristino dei dati**, con l'obiettivo di garantire un recupero dei dati efficiente ed efficace, ottimizzando l'utilizzo delle risorse e riducendo al minimo le potenziali interruzioni;
- **Metodi di smaltimento e distruzione dei dati:** quando necessario, bisogna garantire uno smaltimento sicuro e irreversibile dei dati, in linea con la loro sensibilità, per ridurre i rischi di violazione dei dati e di accesso non autorizzato;

- **Procedure di notifica per lo smaltimento e l'archiviazione dei dati:** garantire la conformità ai requisiti di gestione dei dati e la comunicazione tempestiva delle azioni relative ai dati alle parti interessate;

PROCESSO 6 – CONDIVISIONE DEI DATI E COLLABORAZIONE

In base alla classificazione dei dati, l'archiviazione e il backup non devono essere utilizzati con servizi di terze parti come Google Drive, Dropbox, Microsoft Teams. I dati devono essere condivisi solo da canali approvati con un adeguato controllo degli accessi. Le politiche di condivisione e collaborazione dei dati sono componenti essenziali della governance dei dati, soprattutto quando si tratta di gestire e proteggere le informazioni sensibili e di consentire una collaborazione efficace all'interno e all'esterno dell'organizzazione.

PROCESSO 7 – MONITORAGGIO DEI DATI

Il monitoraggio dei dati si riferisce all'osservazione, al tracciamento e all'analisi continua e sistematica delle attività, dei processi e degli asset legati ai dati all'interno dell'azienda. Il suo scopo principale è garantire l'integrità, la sicurezza, la qualità e la conformità dei dati.

Il monitoraggio dei dati comporta la raccolta e l'analisi continua dei dati per rilevare le anomalie, identificare le tendenze e valutare le prestazioni rispetto ai parametri e agli standard stabiliti. Questo approccio proattivo consente all'azienda di prendere decisioni informate, ridurre i rischi, affrontare i problemi e mantenere le risorse di dati in linea con gli obiettivi e le politiche.

PROCESSO 8 – METADATA MANAGEMENT

La gestione dei metadati è uno dei processi fondamentali della governance dei dati e rende possibile tutti gli altri processi sopra menzionati. Per il momento, inoltre, è l'unico che è stato completamente definito e testato.

Questo processo comporta la creazione, l'aggiornamento e la gestione dei metadati, favorendo una migliore comprensione e consumo dei dati, come base per approfondimenti, decisioni e azioni guidate dai dati.

9.5. IL PROCESSO DI METADATA MANAGEMENT

I metadati forniscono informazioni strutturate sui dati, aiutando a comprendere e gestire efficacemente gli asset di dati, garantendo qualità, conformità, affidabilità e accessibilità.

Uno strumento di gestione dei metadati (definito strumento di governance dei dati, strumento di catalogazione dei dati o semplicemente catalogo dei dati) viene utilizzato per catalogare, gestire il versionamento e favorire la scoperta e la comprensione degli asset di dati da parte degli utenti dell'azienda. L'obiettivo è istituire un'organizzazione ordinata e strutturata di tutte le informazioni aziendali, in modo che ogni utente all'interno dell'azienda possa facilmente individuare, scoprire, comprendere e utilizzare gli asset di dati in casi d'uso e progetti che generano valore.

Il catalogo dei dati può supportare l'utente nel processo di scoperta e utilizzo dei metadati, offrendo la possibilità di sfogliare i metadati (sia aziendali che tecnici), creare raccolte, dashboard, fogli di lavoro.

Il catalogo di dati è composto principalmente dalle 3 strutture, o artefatti, già analizzati in precedenza, ovvero il business glossary, il data dictionary e il data lineage. All'interno di questo strumento, ciascuno dei termini inseriti, che termini di business o asset tecnici, sono identificati da attributi che ne migliorano e approfondiscono la comprensione. Attraverso gli allegati 21, 22 e 23 è possibile identificare i principali attributi che caratterizzano gli elementi contenuti all'interno degli artefatti definiti. Tutti gli attributi (ovvero i metadati) elencati, rendono possibile la strutturazione degli altri processi di data governance.

È possibile notare come alcuni attributi del business glossary sono indicati con una (M)³³, ad indicare quei campi che devono essere sicuramente presenti quando un termine di business è inserito all'interno del glossario aziendale.

Il diagramma di flusso nell'allegato 24 dettaglia il processo di metadata management. Come è possibile notare, ogni attività deve essere svolta attraverso uno strumento differente, tra i principali troviamo sistemi di documentazione come Microsoft Word ed Excel e la piattaforma di catalogazione dei dati Talend Data Catalog.

A questo processo è stata associata una matrice RACI³⁴, o matrice delle responsabilità (allegato 26). Questa serve per identificare chiaramente quali sono i ruoli e responsabilità relativi ad ogni attività.

Questo processo in realtà è costituito da attività che coinvolgono sia figure di business, che si interfacciano principalmente con il business glossary per l'identificazione dei termini di business, sia figure tecniche, che associano i termini di business ai termini tecnici. Per questo motivo, il diagramma di flusso è stato diviso in due, il primo rappresentante il processo di creazione e manutenzione del business glossary (allegato 27), il secondo il processo di creazione e manutenzione del data dictionary (allegato 28).

Questi due processi sono strettamente interconnessi: ogni fase dei processi è riconducibile ad una attività del processo di metadata management.

9.5.1. BUSINESS GLOSSARY – CREAZIONE E MANUTENZIONE

DEFINIZIONE DEI REQUISITI

Il processo di creazione del business glossary inizia quando un consumatore di dati, sia esso un data analyst, un data scientist o qualsiasi altro stakeholder, incontra la necessità di eseguire inferenze o analisi per generare report o dashboard, o per altre attività relative ai processi della data governance identificati in precedenza.

Ad esempio, il data consumer può richiedere dati per verificare che siano conformi al GDPR, assicurando che tutte le informazioni personali siano gestite in modo adeguato e legale oppure per effettuare un'analisi di rischio, valutando potenziali minacce e vulnerabilità all'interno dei processi aziendali.

Da questa esigenza nasce la necessità di dati strutturati, approvati e certificati per garantire l'accuratezza e l'affidabilità dei processi decisionali. Durante la fase di definizione dei requisiti, il data consumer spiega le ragioni della sua richiesta attraverso un rapporto dettagliato (Microsoft Word o

³³ Mandatory

³⁴ Responsible, Accountable, Consulted, Informed. Vedi Allegato 25 per maggiore dettaglio

Excel, esempio dei campi necessari per la richiesta in allegato 29) contenente le esigenze specifiche, i casi d'uso e gli scenari in cui i dati saranno utilizzati.

ANALISI DEI REQUISITI E DEL PERIMETRO DEI DATI

Il Data Steward del dominio corrispondente analizza la richiesta del data consumer e raccoglie le informazioni necessarie continuando la compilazione del suddetto modello in Microsoft Excel o Microsoft Word, in particolare, identificando la lista di termini di business necessari per soddisfare la richiesta del data consumer. Il data steward, consultando i data custodian e i data modeler, è responsabile dell'analisi dei requisiti specificati e del perimetro dei dati, garantendo la conformità agli standard, facilitando la comunicazione tra termini tecnici e di business e valutando gli impatti sul data catalog. In questa fase il data owner deve essere informato quando l'attività viene svolta.

IDENTIFICAZIONE DEI TERMINI DI BUSINESS

Il Data Steward completa il suddetto modello predefinito con l'elenco dei termini aziendali che potrebbero soddisfare la richiesta del data consumer. Il data owner è responsabile della supervisione di questo modello completato.

Attraverso l'analisi dell'elenco dei termini commerciali originati (realizzato dal data steward, con il supporto dei data custodians e dei modelers), sarà chiaro se i termini di business necessari sono già presenti e disponibili nel Data Catalog Tool o se è necessario integrarli all'interno del data catalog.

COMPILAZIONE MANUALE O BULK LOAD

Nel caso in cui il termine di business non sia già presente e disponibile, il data steward sarà responsabile della compilazione del business glossary. In base alla portata dell'iniziativa del business glossary, questo può essere compilato attraverso la compilazione manuale (direttamente sullo strumento Data Catalog) o il Bulk Load (attraverso la compilazione del BG Import Sheet su Excel, seguito dal caricamento sullo strumento Data Catalog).

Il BG Import Sheet è un documento Excel in cui è stata implementata una macro (codice in allegato 30) per gestire il suo funzionamento. Questo file è necessario per assicurare che una volta che i termini di business sono caricati all'interno del business glossary, ogni termine abbia presenti i suoi corrispondenti attributi obbligatori (identificati nell'allegato 21). Inoltre, deve funzionare come riferimento per tutti i domini per facilitare l'operazione di bulk load.

L'operazione di bulk load avviene attraverso il caricamento di un file ".csv" che include i termini di business e i metadati. Il file excel, con la macro, svolge principalmente due funzioni:

- Nel momento del salvataggio, evidenzia in rosso per ogni termine di business inserito le celle obbligatorie che non sono state compilate, bloccando l'operazione di salvataggio. In questo modo l'utente che compila il business glossary è obbligato a compilare tutte le celle obbligatorie;
- Nel momento del salvataggio, il file excel esporta il foglio in cui è contenuto il business glossary oggetto del bulk load in un file ".csv" e lo salva direttamente nello stesso repository in cui il file "BG Import Sheet" è salvato. In questo modo, il file è pronto per essere importato sul catalogo di dati.

Il data owner è responsabile del corretto adempimento del business glossary. Una volta che il Business Term è presente nel Data Catalog Tool, è indispensabile verificare la completezza e l'accuratezza delle informazioni fornite e incluse nel BG Import Sheet. Questo è particolarmente

importante se l'adempimento del business glossary avviene tramite Bulk Load, dove è più facile incorrere in errori. Il data steward, insieme al data custodian, supervisiona la verifica delle informazioni caricate sul Data Catalog. In questa fase, il data steward si concentra sul business glossary, mentre il data custodian si occupa del data dictionary e del data lineage.

Il data owner è responsabile di verificare manualmente che le informazioni caricate soddisfino i requisiti per la successiva approvazione e pubblicazione delle modifiche sul data catalog. Nel caso in cui vengano rilevate informazioni mancanti o imprecise, è necessario tornare alla fase di adempimento del BG, per verificare come sono state svolte le attività sullo strumento data catalog e risolvere eventuali problemi.

CONNESSIONE TRA I TERMINI DI BUSINESS E I METADATI TECNICI

Dopo l'approvazione del data owner, il data steward, consultando il data custodian, gestisce la mappatura e le connessioni tra i metadati di business (termini, definizioni ed entità di dati) e gli asset tecnici (tabelle, colonne e fonti di dati) sullo strumento data catalog. Il data owner è responsabile della corretta creazione di tali connessioni. Effettuare la connessione tra termini di business e asset tecnici significa creare un'associazione semantica tra i due termini.

All'interno del catalogo di dati, è possibile infatti creare degli attributi personalizzabili, che permettono di creare delle relazioni tra uno o più termini di business. Questo ha a che fare con la definizione del metamodello. Il metamodello è una rappresentazione dei concetti, relazioni e regole che governano l'organizzazione e la gestione dei dati all'interno di un sistema di data governance. Esso fornisce una struttura per definire e collegare i vari elementi dei dati, sia a livello di business che tecnico. L'associazione utilizzata nel catalogo dei dati per gestire la connessione tra i termini di business e quelli tecnici è la "Define/Defined By", perché un termine tecnico deve per forza almeno definire un termine di business.

CURA DEI METADATI

Per avere un catalogo dei dati affidabile, aggiornato e significativo, è necessario mantenere i metadati. Il data steward è responsabile della cura dei metadati aziendali sul Glossario aziendale all'interno del catalogo dati.

FASE DI CERTIFICAZIONE

Infine, il data owner è responsabile della certificazione dei termini aziendali nel business glossary. Questo processo prevede la convalida e l'approvazione dell'accuratezza, della rilevanza e dell'affidabilità delle voci di metadati all'interno del catalogo, indicando che esse soddisfano determinati criteri o standard stabiliti dall'organizzazione.

Con questa certificazione, il data owner:

- Assicura che per ogni termine aziendale è definita almeno un'associazione con un elemento del dizionario dei dati;
- Chiarisce la proprietà e la responsabilità delle risorse di dati;
- Convalida che tutte le informazioni rilevanti siano state acquisite e documentate nel catalogo dei dati;
- Supporta la conformità alle politiche, alle normative e agli standard di governance dei dati, garantendo il rispetto delle linee guida stabilite;
- Costruisce la fiducia nel catalogo dei dati come fonte affidabile di informazioni per le attività relative ai dati in azienda.

MANUTENZIONE CONTINUA

La manutenzione del catalogo dati e del glossario aziendale garantisce l'accuratezza dei dati, la conformità alle normative e l'efficienza operativa. Aggiornando e verificando regolarmente le informazioni, le organizzazioni mantengono la fiducia, consentono di prendere decisioni informate e riducono i rischi associati a dati obsoleti o imprecisi.

Per questo motivo, l'attività di manutenzione del catalogo dati e del glossario aziendale è un'attività continua che non deve essere trascurata. Si tratta di una verifica continua che coinvolge tutti gli attori del processo, per assicurare che le informazioni aggiunte siano accurate, aggiornate e certificate, anche in caso di modifiche alle politiche o alle normative sulla governance dei dati

9.5.2. CREAZIONE DEL DATA DICTIONARY, DATA LINEAGE E LORO MANUTENZIONE

IDENTIFICAZIONE DELLE ENTITÀ DI DATI E FONTI DEI DATI

Il data custodian utilizza le informazioni raccolte dal data steward per identificare gli asset di dati tecnici specifici necessari a soddisfare la richiesta. Le informazioni raccolte saranno utilizzate per completare il modello in Microsoft Excel, indicando i termini tecnici necessari per descrivere l'asset di business. Questa attività è considerata la controparte tecnica dell'attività di identificazione dei termini aziendali dettagliata nel processo di creazione del glossario aziendale.

Una volta che tutte le informazioni sono disponibili in un unico template, se necessario, consultando il data steward, il data custodian verificherà se le fonti di dati richieste sono già state raccolte e se quindi sono già disponibili nel data catalog. In tal caso, il data custodian, consultandosi con il data steward e il data modeler, sarà responsabile della verifica della completezza e dell'accuratezza delle informazioni (questa fase sarà trattata in modo approfondito più avanti).

Se le fonti di dati non sono già disponibili nel data catalog, il data custodian dovrà rivolgersi al platform admin per verificare se il connettore di harvesting per le specifiche fonti di dati è già registrato. In tal caso, l'amministratore della piattaforma sarà responsabile della configurazione del connettore di harvesting (questa fase sarà trattata di seguito).

Il processo di harvesting di metadati comprende l'insieme di attività che permettono raccogliere metadati da una fonte di dati. Questa raccolta viene effettuata utilizzando connettori nativi o personalizzati del catalogo di dati che sostanzialmente sono ponti che utilizzano un driver specifico per connettersi a un sistema di origine dei dati e raccoglierne i metadati.

Prima di raccogliere i metadati, è necessario conoscere alcune informazioni (ad esempio dove risiedono i metadati, quali tecnologie sono necessarie per estrarli e quale processo deve essere seguito per garantire una corretta estrazione). Bisogna anche assicurarsi di avere una connettività adeguata alla fonte di metadati in formato esterno, nonché l'accesso a qualsiasi risorsa ausiliaria.

Di seguito è riportato un elenco di alto livello delle azioni che devono essere eseguite quando si effettua il processo di harvesting dei metadati (allegato 31) :

- Identificare i sistemi delle fonti di dati che agiscono sia come produttori di dati (come archivi operativi di dati) sia come consumatori di dati (come piattaforme di dati, strumenti di business intelligence, strumenti di modellazione AI, ecc.);

- Identificare il flusso di dati concettuale o logico per comprendere il ciclo di vita dei dati e dove risiedono;
- Selezionare e configurare un bridge e raccogliere i metadati per ogni sistema;
- Verificare la coerenza delle informazioni.

SUPERVISIONE DELLA CREAZIONE DI UN NUOVO CONNETTORE

Nel caso in cui il connettore di harvesting non sia registrato, il data custodian verificherà se un connettore nativo è presente nell'elenco di quelli disponibili. In tal caso, il data custodian sarà responsabile della configurazione del connettore di harvesting (questa fase sarà trattata di seguito).

Se non è presente un connettore nativo, sarà necessario creare un nuovo connettore. Il processo di creazione di un nuovo connettore può includere diverse attività, tra cui la scelta se contattare il fornitore di servizi data catalog o creare il connettore internamente. Questa scelta, così come le altre possibili fasi di questo processo, devono essere ulteriormente definite in modo più strutturato, possibilmente in una fase di maggiore maturità dell'iniziativa. Tuttavia, la supervisione di questa attività è affidata al platform admin.

CONFIGURAZIONE DEL CONNETTORE

Se un connettore di harvesting è presente e/o già registrato, il platform admin sarà responsabile della configurazione di tale connettore, ad esempio specificando quali campi di dati devono essere importati, impostando filtri, definendo regole di estrazione dei dati, specificando la frequenza di recupero dei dati e altri parametri rilevanti.

Questa fase comprende generalmente il collaudo della connessione configurata, per garantire che funzioni come previsto. Il platform admin può avvalersi dell'aiuto del data custodian, che ha una conoscenza approfondita dei requisiti per l'harvesting.

HARVESTING DEI METADATI TECNICI

Il data custodian esegue il connettore e procede all'estrazione dei metadati tecnici. Questa attività può includere la fornitura di credenziali di autenticazione o altri meccanismi di autenticazione per garantire che il processo di raccolta abbia tutte le autorizzazioni necessarie per recuperare dalle fonti di dati.

APPROVAZIONE E CURA

Seguendo il processo di metadata management, il data owner darà la sua approvazione formale dopo la verifica che i termini tecnici rispondano effettivamente alle richieste dei consumatori di dati.

Di seguito, avverrà la connessione di questi con i termini di business, nonché la cura dei metadati tecnici.

CREAZIONE DEL DATA LINEAGE

Parallelamente, il data custodian è anche responsabile della cura della linea orizzontale dei dati (idealmente creata automaticamente) all'interno del data catalog, fornendo integrazione e arricchimento manuale. Il data custodian creerà anche il lineage semantico per tracciare il flusso semantico dei dati.

CREAZIONE DEL DIAGRAMMA ENTITA' RELAZIONI

Dopo aver creato le connessioni tra i metadati aziendali e tecnici, il data modeler è incaricato di definire il diagramma Entità-Relazioni. Si tratta di una rappresentazione visiva utilizzata per descrivere le relazioni tra le diverse entità all'interno dei set di dati catalogati, migliorando la comprensione della struttura dei dati e delle relazioni, facilitando l'esplorazione dei dati e garantendone l'integrità. Inoltre, deve incaricarsi di gestire il metamodello, ovvero l'insieme di associazioni tra i dati che permettono di gestire quelle situazioni in cui diversi dati di diversi domini possono avere lo stesso nome o significato ed è necessario stabilire un significato o connotazione comune ai due.

CERTIFICAZIONE E MANTENIMENTO

Per queste sezioni, valgono gli stessi commenti effettuati per il business glossary.

9.6. CAPACITÀ DEL TOOL DI DATA CATALOG

La creazione delle strutture del business glossary e del data dictionary è svolta principalmente all'interno dello strumento di catalogazione dei dati. All'interno del tool di data catalog sono previste le seguenti capacità, che permettono ad ogni attore e utente di sviluppare le proprie attività sullo strumento stesso, assicurando la sicurezza e privacy dei dati:

- **Capacità di management:** gli utenti con capacità di gestione hanno il permesso di creare e gestire oggetti nel catalogo. Ciò include la possibilità di aggiungere, modificare o eliminare metadati, nonché di gestire altri aspetti rilevanti relativi alla sicurezza, all'accessibilità e alla personalizzazione di Talend Data Catalog;
- **Capacità di modifica:** gli utenti con capacità di modifica possono apportare modifiche dirette agli oggetti del catalogo. Questo si traduce nella possibilità di modificare gli attributi dei metadati, aggiungendo informazioni rilevanti per favorire la scoperta e la comprensione dei metadati;
- **Capacità di visualizzazione:** gli utenti con capacità di visualizzazione possono vedere i metadati e le risorse di dati nel catalogo, ma non hanno il permesso di apportare modifiche. Questo livello di accesso è spesso assegnato agli utenti che devono solo visualizzare le informazioni senza apportare modifiche. Per massimizzare i vantaggi offerti da un'efficace data governance, tutti gli utenti del dominio almeno avranno capacità di visualizzazione sui dati del proprio dominio;
- **Capacità di Workflow:** queste funzionalità possono essere attivate per gestire i processi pertinenti definendo una sequenza specifica di attività - ciascuna associata all'attore pertinente - con l'obiettivo di rivedere, approvare e pubblicare le modifiche al Data Catalog assicurando una stretta aderenza ai processi definiti. Al momento la capacità di workflow non è attiva all'interno del catalogo di dati.

L'allegato 32 permette di identificare i gruppi di capacità individuate e la descrizione delle azioni che li riguardano.

All'interno del data catalog ognuna di queste capacità è stata assegnata ad una o più ruoli definiti in precedenza, per forzare automaticamente all'interno del catalogo di dati le attività che ciascun ruolo può svolgere (Allegato 33). In questo modo, non si incorre nel rischio di modifiche o effettuare aggiornamenti non permessi da parte di ruoli che non dispongono delle necessarie capacità.

9.7. LE METRICHE DEI PROCESSI

Le metriche chiave della data governance sono definite in base alle best practice, seguendo i processi precedentemente menzionati. Sono fondamentali per comprendere l'applicazione delle politiche e delle pratiche di data governance in tutta l'organizzazione. Oltre alle metriche operative elencate di seguito, sono previste anche metriche specifiche per il business e orientate ai vari domini.

Le metriche principali identificate per il momento sono contenute all'interno dell'allegato 34, insieme alla loro descrizione e regola di calcolo.

Per la definizione di queste metriche, è stato scelto un approccio di tipo incrementale. Questa strategia ha l'obiettivo di enfatizzare il progresso graduale e iterativo nella definizione e nel perfezionamento delle metriche più importanti per guidare il miglioramento continuo. Inoltre, si vuole consentire la definizione di metriche ampliando gradualmente l'ambito di interesse (ovvero le aree e i processi di interesse) e il grado di precisione e dettaglio delle metriche stesse.

L'approccio graduale e iterativo significa che, dopo aver identificato una serie di KPI iniziali che possono essere realmente misurati, la loro implementazione sarà sottoposta a una fase pilota che ne consentirà la verifica. Questo dà agli utenti la possibilità di identificare lacune e inefficienze e di garantire che i KPI si evolvano per soddisfare lo sviluppo dell'implementazione del progetto di data governance. Nonostante, infatti, i KPI contenuti all'interno dell'allegato 34 siano stati definiti sulla base di esperienze pregresse, alcuni KPI (es. gestione della sicurezza e privacy, accuratezza dei dati, monitoraggio dei dati...) non sono immediatamente misurabili e di conseguenza implementati.

Altri invece, come la completezza dei metadati, non possono essere misurati nella loro completezza, ma la stessa metrica può essere divisa in sotto metriche che insieme comporranno il KPI per intero.

Per il momento, sono state definite le metriche relative principalmente al business glossary e alla completezza dei suoi attributi.

Nella prima fase la priorità sarà data alle metriche relative al business glossary e alla sua completezza. Come è già stato definito all'interno del processo di metadata management, il business glossary pone le basi per la comprensione dei termini di business in tutta l'azienda. Per questo motivo, il popolamento del business glossary per ogni dominio è il primo passo verso il corretto governo dei dati. Per la rappresentazione dei dati analizzati di seguito, viene utilizzata la dashboard fornita dal tool di data catalog. Questa permette di effettuare inferenze e filtrare tra i termini di business contenuti nel catalogo, permettendo di ricavare informazioni sul suo stato. Con l'obiettivo di controllare i termini aziendali e i loro attributi, sono state identificate le seguenti metriche:

- Facendo riferimento al KPI M1 "Percentuale di dati classificati", è stata stabilita la metrica "Dati classificati". Il primo livello di implementazione di questa metrica si riferisce al numero di termini con la corretta applicazione delle etichette di sensibilità;
- In riferimento al KPI M5 "Gestione dei metadati", è stata stabilita la metrica "Campi obbligatori dei metadati". I termini di business sono stati classificati in 3 gruppi, per identificare il loro grado di completezza in relazione alla presenza di campi obbligatori definiti all'interno dell'allegato 21. Infatti, per motivi tecnici, non è stato subito possibile implementare il caricamento massivo attraverso il template BG Import Sheet definito nel processo di creazione e cura del business glossary. Per questo motivo, era necessario tenere traccia di quali termini di business fossero catalogati con i loro attributi obbligatori;
- In riferimento al KPI M4 è stata stabilita la metrica "Area di Business e Processo aziendale". Poiché ogni Business Term appartiene a una sola area di business e a un solo processo,

questa metrica consente di identificare il volume di Business Term assegnati per ogni dominio.

- In riferimento al KPI M5 “Gestione dei metadati” è stata stabilita la metrica “Termini di business con associazione”. Questo KPI riguarda solo i Business Terms che definiscono almeno un asset tecnico, attraverso l’associazione³⁵ “Define/Defined by”. Come accennato, ogni termine di business deve essere identificato e associato ad almeno un asset tecnico.

Con la seguente definizione dei processi della data governance, sarà possibile implementare e definire le metriche per gli stessi. Inoltre, tramite il graduale approfondimento delle capacità del catalogo di dati, attivando i processi di workflow sarà possibile misurare temi come il numero di termini di business all’interno del workflow di approvazione e certificazione o quello per l’assegnazione dell’etichetta di sensibilità.

9.1. IL FUNZIONAMENTO DEL FRAMEWORK DI DATA GOVERNANCE

L’allegato 35 rappresenta i concetti affrontati sino ad ora e le relazioni che intercorrono tra gli elementi menzionati. Questo diagramma è fondamentale in quanto mostra il funzionamento delle strutture e dei concetti definiti fino ad ora e relativi al framework della data governance. Le relazioni sono rappresentate di colore diverso sulla base degli elementi che connettono.

Iniziando l’analisi dalle strutture organizzative, si mostrano i nuovi ruoli definiti e adibiti alla data governance. L’allegato 36 mostra come, prima di questo progetto, al netto degli uffici di supporto e dei domini di business, non esistevano gli apparati della Steering Committee, il DGO, nonché tutte le figure responsabili delle attività della data governance.

Uno dei problemi principali riscontrati prima della definizione del framework, è che spesso, nel caso in cui un dominio avesse avuto bisogno di dati appartenenti ad un altro dominio per effettuare delle analisi, avrebbe dovuto comunicare la propria necessità al dipartimento IT, che, accedendo ai repository dei domini, avrebbe dovuto reperire i dati necessari e comunicarli al primo dominio. Questo è causato principalmente dalla creazione di silos organizzativi, dalla mancanza di ownership dei dati, e di interoperabilità degli stessi. Inoltre, anche nel caso in cui questi dati fossero reperiti, la loro accuratezza e qualità non poteva essere garantita.

Infatti, come è possibile notare nell’allegato 37, nonostante le entità di dati fossero già esistenti, queste non erano propriamente descritte attraverso i loro metadati in un catalogo centralizzato, accessibile ai differenti domini e certificate attraverso un processo definito formalmente.

Inoltre, era impossibile identificare la proprietà univoca dei singoli dati. Questo significa che, nel momento del bisogno, era difficile localizzare e accedere ai dati stessi, ed era complicato anche comprendere a quali figure del dominio rivolgersi per poter ottenere i permessi per accedere ai dati. Questo problema causava spesso perdite di tempo, poiché gli utenti incaricati delle analisi erano costretti a impegnarsi in una serie di scambi di e-mail e messaggi per individuare la figura corretta a cui rivolgersi.

Allo stesso modo, significa che i singoli domini non avevano un vero e proprio controllo sulla completezza e qualità dei propri dati. Questo fenomeno è presente soprattutto in organizzazioni di

³⁵ L’associazione è il modo per attribuire un valore semantico ai dati. Ci sono delle associazioni predefinite, ma è possibile crearne di personalizzabili dandogli il significato ricercato

grandi dimensioni. Questi problemi erano riconducibili alla mancanza dei processi della data governance e, in particolare, quello del metadata management, abilitatore di tutti gli altri.

Con l'implementazione del framework della data governance e l'utilizzo del catalogo di dati, si intende rimuovere le barriere create da una mancanza di struttura e processi per la gestione dei dati.

Innanzitutto, le strutture, ruoli e responsabilità definiti assicurano l'ownership diretta dei dati, obbligando i vari responsabili dei domini a mantenere certi standard di gestione degli stessi, facilitando la condivisione e l'interoperabilità tra i domini. Le entità, propriamente descritte all'interno del catalogo di dati saranno sottostanti ai processi della data governance, imponendo regole di qualità, condivisione, privacy e condivisione attraverso tutto il ciclo dei dati.

Come menzionato in precedenza, attualmente il framework non è implementato del tutto, pertanto, molti di questi problemi ancora non sono stati risolti. Nonostante ciò, è stato sviluppato un test per implementare i processi e le strutture definite, analizzato nel seguente paragrafo.

9.2. SALES & MARKETING USE CASE – METADATA MANAGEMENT

Nonostante, come riportato nell'allegato 17, non siano stati ancora implementati dei veri e propri casi d'uso su cui testare l'implementazione completa del framework di data governance, durante l'esperienza di tirocinio è stato sviluppato un primo test sul processo di metadata management sul dominio di Sales & Marketing.

Questo dominio è stato scelto preliminarmente attraverso la realizzazione di una matrice Complessità-Valore (allegato 38). Innanzitutto, sono stati identificati gli otto domini su cui si sarebbe potuto effettuare il test sul processo di metadata management. Sulla base di fattori come il volume di dati creati, la complessità del dominio, la reperibilità e supporto delle figure interne al dominio, è stata valutata la complessità (divisa in 5 livelli, da Minima a Molto Alta) del test. Allo stesso modo, è stata valutata la capacità di creare valore per il dominio stesso, tramite le metriche definite in precedenza. Come conseguenza, il dominio di Sales & Marketing è stato scelto come miglior opzione per questo specifico test, in quanto mappare le entità dei clienti e le loro informazioni è un'attività relativamente gestibile, ma il valore aggiunto è elevato poiché consente di ottimizzare le strategie di marketing, migliorare la personalizzazione delle offerte e, di conseguenza, incrementare significativamente le vendite e la soddisfazione del cliente.

Inoltre, è importante notare come questo test era circoscritto ad un numero di entità ristretto del dominio di Sales & Marketing per poter mantenere contenuto il volume di dati e metadati da gestire.

Questo test aveva l'obiettivo la creazione delle strutture del business glossary e del data dictionary all'interno del Catalogo di Dati e la loro conseguente popolazione con dei termini di business e tecnici, che soddisfacessero un certo grado di qualità definito in fase di analisi degli stessi. Inoltre, grazie agli artefatti creati e a continue riunioni con l'organizzazione, si intendeva comprendere l'attuale situazione dell'azienda e del dominio in rispetto alla gestione dei loro dati e metadati e migliorare eventualmente la qualità di questi ultimi.

Il motivo di questo test era di testare il processo di metadata management appena definito, definendo i ruoli e le responsabilità all'interno del dominio di Sales & Marketing permettendo di porre le basi per il test degli altri processi della data governance, quando questi saranno definiti. Questo test, iniziato la prima settimana di aprile, è ancora in fase di sviluppo. Nonostante ciò, l'obiettivo dei seguenti paragrafi è mostrare le attività svolte e gli obiettivi raggiunti fin ora.

9.2.1. DEFINIZIONE DEI RUOLI E AVVIO DEL TEST

Per prima cosa, al fine di assicurare il successo del test, all'interno del team di Sales & Marketing con cui Accenture si è interfacciato sono state identificate le figure di business definite nel data governance framework.

Per fare ciò, è stato necessario per prima cosa effettuare diverse riunioni congiunte con il DGO e le figure responsabili del dominio al fine di identificare correttamente uno o più data owner. Avendo validato che il numero di business term avrebbe dovuto essere di un volume contenuto, al fine di mantenere il perimetro del test limitato, in prima istanza è stato identificato un singolo data owner, tra le figure manageriali del dominio di Sales & Marketing.

La nomina del data owner è stato il primo passaggio critico del test. Le attività da sviluppare per portare a termine questo test si aggiungono a quelle già in capo al data owner identificato. Per questo motivo è stato necessario identificare una figura manageriale che fosse sufficientemente vicino ai dati del dominio, ma che avesse sufficiente disponibilità di tempo per poter portare avanti questo progetto, che avrebbe richiesto almeno 4 ore alla settimana (10% del monte ore settimanale), tra le riunioni di aggiornamento, le attività da sviluppare autonomamente, e la formazione.

Il data owner identificato, per poter familiarizzare con i temi chiave della data governance (a lui estranei prima di questo test), nonché sull'utilizzo del catalogo di dati, ha infatti seguito una formazione generale (svolta dal team di Accenture e dal DGO) di 1h sulla data governance e le strutture definite (business glossary e data dictionary), e di 1h sull'utilizzo e sulle capacità del tool di catalogazione dei dati.

La nomina del primo data owner è stata fondamentale iniziare a raccogliere i principali termini di business all'interno di un file excel (issue log, analizzato nel prossimo paragrafo), per poter stabilire le gerarchie tra le principali entità e, a seconda di queste, identificare altri owner dei dati, per garantire la massima comprensione di questi ultimi, nonché non sovraccaricare il primo data owner identificato. Al momento, è stato nominato un data owner per ognuno delle 3 entità identificate all'interno del dominio di Sales & Marketing (Customer, Backlog e Demand, analizzate nel seguente paragrafo).

Allo stesso modo, è stata identificato e nominato un data steward del dominio. La nomina è avvenuta tramite il supporto diretto del DGO e del data owner. Il data steward ha seguito lo stesso percorso di formazione del data owner, con un focus maggiore (1h in più) sulle capacità del tool di catalogazione dei dati, in quanto, seguendo il processo di metadata management, il suo intervento è richiesto in vari momenti ed è critico per il successo dello stesso.

Per quanto riguarda le figure tecniche, era già stato identificato il platform admin, che ha supportato Accenture nelle attività sul catalogo di dati. Questa figura era già stata assegnata a tempo quasi completo al progetto di data governance in fase di definizione del framework, in quanto la gestione del catalogo di dati ha richiesto un intenso periodo di formazione per esplorare le capacità e le possibilità dello stesso.

Inoltre, per il trattamento dei termini tecnici, sono stati identificati due data custodian. Queste figure sono state identificate all'interno del dipartimento IT, che avevano profonda conoscenza dei connettori di harvesting, avendo accesso ai database in cui i termini tecnici sono contenuti. Inoltre, sono state coinvolte altre figure di supporto, quali 1 data scientist e un data analyst interno dell'azienda.

Per ultimo, agli utenti del team di Accenture, all'interno del tool di catalogazione dei dati, hanno assunto ciascuno le capacità di un ruolo (data owner, steward, modeler, custodian, admin) per poter testare le attività e le capacità di accesso ai dati e alle informazioni presenti.

9.2.2. POPOLAMENTO DEL BUSINESS GLOSSARY

Al fine di migliorare la comprensione delle entità fondamentali del dominio di Sales & Marketing, nonché i termini di business che lo definiscono, sono state condotte diverse riunioni il cui output è mostrato nell'allegato 39, che mostra la struttura del business glossary implementata sul catalogo di dati. Come è possibile notare, la cartella del dominio è formata dai due artefatti della data governance, il Glossario e il Data Dictionary.

Il Glossario contiene 3 entità principali, il Cliente, la Domanda, e il Backlog, identificate grazie al supporto del data owner principale. Le 3 entità sono state identificate a partire dai processi del dominio di Sales & Marketing. In breve, il Customer rappresenta tutte le informazioni relative ai clienti dell'azienda. Questo include dati anagrafici come nome, indirizzo, contatti, e dettagli specifici come preferenze di acquisto, storico degli ordini, e comportamenti di acquisto. Queste informazioni sono cruciali per la gestione delle relazioni con i clienti e per personalizzare i servizi e le offerte.

Il Demand raccoglie le informazioni relative agli ordini dei clienti. Comprende dettagli quali il numero dell'ordine, i prodotti richiesti, le quantità, le date di ordine e di consegna previste, e qualsiasi altra specifica richiesta del cliente. Questi dati aiutano a gestire e monitorare il flusso degli ordini, garantendo che le esigenze dei clienti siano soddisfatte in modo efficiente e tempestivo.

Per ultimo, il "Backlog" rappresenta l'insieme degli ordini dei clienti che non sono ancora stati evasi. Include informazioni sugli ordini in sospeso, come la quantità di prodotti ancora da consegnare, le date previste di consegna, lo stato attuale di ciascun ordine e le cause di eventuali ritardi. La gestione del backlog è essenziale per garantire che gli ordini siano completati nel rispetto dei tempi, migliorando la soddisfazione del cliente e l'efficienza operativa dell'azienda.

L'entità Customer è formata a sua volta da 4 sotto-entità che descrivono il cliente e i suoi processi. Infatti, l'entità Company definisce le informazioni relative all'azienda, la Billing racchiude le informazioni riguardante la fatturazione degli ordini, e le Shipping e Location permettono di incrociare l'azienda cliente con le sue succursali, nonché le informazioni relative alla consegna dei prodotti finiti. Al di sotto di queste entità, all'ultimo livello, infatti, sono contenuti i termini di business che le descrivono.

I termini di business sono stati raccolti e catalogati con gli opportuni attributi menzionati nell'allegato 21. Sono stati raccolti e catalogati 118 termini di business (87 relativi al Customer, 11 relativi al Demand, 20 relativi al Backlog, mostrati in allegato 40). Per prima cosa è stato evidente che pochi dei termini di business raccolti fossero in possesso degli attributi obbligatori.

Prima di affrontare questo punto, è stato però necessario valutare la qualità e la coerenza dei termini di business raccolti. Infatti, è stato notato che, principalmente sugli attributi "Nome" e "Descrizione" era possibile identificare numerose discrepanze, presentando incongruenze tra diversi termini di business con nome simile ma significato diverso, nome diverso ma significato simile, o la loro descrizione del tutto non chiara.

Al fine di affrontare questo problema, sono stati definiti cinque tipi di errori, riscontrati nell'analisi dei termini di business. L'allegato 41 descrive il tipo di problema, il loro impatto, e le azioni necessarie da portare a termine per la loro risoluzione.

L'allegato 42 indica per ogni entità il numero di termini di business con e senza errori. Il totale di errori identificato è 145, sottolineando come alcuni termini di business siano caratterizzati da più di un errore (l'allegato 43 mostra il numero totale di errori riscontrati).

Il team di Accenture ha lavorato con i data owner del dominio al fine di identificare e rimuovere per quanto possibile gli errori identificati. Per importanza, l'errore di tipo 1 è stato il primo ad essere affrontato, seguito dall'errore di tipo 2. Di seguito, è stato analizzato l'errore di tipo 2, che si pensava avrebbe indirettamente migliorato la situazione degli errori 3 e 4. Allo stesso modo, l'errore di tipo 5, che veniva riscontrato almeno in 81 termini di business, richiede un'analisi dettagliata e pertanto è un'attività che richiede molta dedizione e impegno in termini di tempo, pertanto è ancora in processo.

Le riunioni con i data owner sono state utili per affrontare un altro tema, ovvero la proprietà dei dati del dominio. Nell'analizzare i termini di business ci si è posti l'obiettivo di identificare quali di questi fossero creati dai processi del dominio di S&M e quali fossero derivati dai processi di altri domini e consumati all'interno del primo. È stato identificato, ad esempio, che alcuni dei termini appartenenti alle entità Shipping e Billing fossero riconducibili al dominio di Supply Chain e Finance. Pertanto, alcuni di questi termini di business sono stati eliminati, senza perdere l'informazione che questi ultimi sono utilizzati all'interno del dominio, attraverso le opportune associazioni all'interno del catalogo di dati.

Come secondo step, il caricamento dei termini di business sul Catalogo di dati è avvenuto tramite la creazione di un file in formato csv. I primi tentativi di import sono stati eseguiti dal team di Accenture e sono avvenuti senza l'utilizzo del BG_ImportSheet in quanto la maggior parte degli attributi di business non soddisfacevano nemmeno il primo livello del KPI "Campi obbligatori dei metadati", non includendo principalmente i campi "Description" e "Master/Reference". Nell'allegato 44 è presente la rappresentazione del KPI al momento del primo upload sul Catalogo di dati.

Nelle settimane successive, sono state svolte diverse riunioni con i rappresentanti del dominio e del DGO per poter affrontare questi due temi, ovvero la risoluzione degli errori dei termini di business, e migliorare la completezza dei metadati di questi ultimi. Inoltre, con il popolamento dei termini tecnici all'interno del data dictionary, è stato possibile effettuare il collegamento di questi ultimi con quelli di business.

9.2.3. POPOLAMENTO DEL DATA DICTIONARY

Seguendo il processo di metadata management, in concomitanza della armonizzazione dei termini di business, è stato possibile procedere con l'harvesting dei termini tecnici.

Questo processo è avvenuto seguendo il flow chart definito nell'allegato 31. Per effettuare l'harvesting dei termini tecnici sono stati utilizzati 3 connettori differenti, accedendo quindi a 3 database distinti.

I termini tecnici sono stati correttamente ordinati e organizzati secondo i loro schemi, tabelle e colonne all'interno del catalogo di dati, secondo la gerarchia rappresentata all'interno dell'allegato 45. Inoltre, seguendo il processo di metadata management, con la presenza dei termini tecnici sul catalogo di dati, è stato possibile effettuare le connessioni ai termini di business ai termini tecnici tramite l'associazione Define/Defined By. Questo è un'altra operazione che è stata svolta manualmente.

9.2.4. RISULTATI DEL TEST

Sono state identificate delle metriche per la valutazione dell'andamento del test. La maggior parte sono state definite all'interno del framework, altre invece fanno riferimento a parametri del test stesso (allegato 46).

Come è possibile notare, oltre alle metriche generali definite in precedenza, sono presenti KPI relativi al caso d'uso specifico, includendo fattori come lo sforzo del team di Accenture e del DGO per coinvolgere e formare le persone, nonché del dominio stesso per portare avanti le proprie attività.

L'implementazione di un registro degli errori ha permesso di avviare discussioni con i data owner per correggere gli errori e introdurre il concetto di cura dei metadati e di associazione tra le varie entità aziendali.

Innanzitutto, le discussioni con il dominio hanno permesso di ridurre il numero di errori di tipo 1 a 0 unità. Questo significa che al momento all'interno del business glossary di S&M non ci sono termini con lo stesso nome. Allo stesso modo, sono stati ridotti di molto gli errori di tipo 2, 3, 4.

Questo ha permesso ridurre il numero dei termini raccolti inizialmente. Dei 118 originali, ne sono stati mantenuti 76 (allegato 47), una riduzione del 35%.

L'errore di tipo 5 è tutt'ora in fase di analisi. Modificare le descrizioni dei termini aggiungendo parole chiave per migliorarne la comprensione è un'attività che richiede un elevato consumo di tempo. L'errore di tipo 5 è infatti quello che ha subito meno cambiamenti. Nell'allegato 48 è possibile notare come cambia il numero di errori rilevati dall'inizio del test. Da un totale iniziale di 157 errori, al momento ne sono rilevati solamente 65, una riduzione di quasi il 60%, di cui i rimanenti sono principalmente errori di tipo 5.

Questo significa che al momento, almeno 10 (la differenza tra 76 e 65) termini di business non è caratterizzato da alcun errore.

Per quanto riguarda gli attributi dei termini caricati, al momento tutti soddisfano i primi 2 livelli e il terzo livello è in fase di compilazione (allegato 49). Nonostante non sia stato seguito un workflow stabilito, è stata assegnata manualmente ad ogni termine l'etichetta di classificazione, permettendo quindi di restringere l'accesso ai termini di business particolarmente sensibili.

Allo stesso modo, il processo di associazione dei termini di business a quelli tecnici è stato svolto manualmente sulla piattaforma, permettendo l'associazione del 75% dei termini (57/75).

L'allegato 50 riassume i risultati raggiunti rispetto alle metriche del test definite in principio.

9.2.5. CONCLUSIONI DEL TEST

Nonostante i risultati raggiunti, è evidente come il test non sia ancora terminato. In prima istanza sarà necessario completare la rimozione degli errori tra i termini di business. Di seguito, saranno necessarie ulteriori analisi per completare questi ultimi almeno con tutti i metadati obbligatori e procedere con la compilazione dei termini di business non obbligatori, fondamentali per avere una visione completa sulle entità e le loro caratteristiche.

Un tema fondamentale che non è stato ancora trattato è la creazione del data lineage, per poter chiaramente identificare il percorso dei dati e dei termini di business.

Quando questo sarà fatto, il workflow di approvazione e certificazione dei termini inseriti nel catalogo di dati sarà possibile, permettendo ai data owner di marcare i termini di business che soddisfano tutte le caratteristiche di completezza e accuratezza. Un termine sarà certificato solo e soltanto se:

- Tutti gli attributi, obbligatori e non obbligatori, sono presenti all'interno del catalogo di business
- Il termine è associato almeno ad un termine tecnico
- Il termine è caratterizzato dal data lineage

Nel pratico questo test, una volta concluso, permetterà di semplificare il modo in cui gli utenti possono accedere ai dati, nonché fornire maggiore sicurezza sull'integrità e sulla completezza degli stessi. A partire da questo caso d'uso, sarà possibile espandere il numero di entità analizzate all'interno del dominio di Sales & Marketing, permettendo eventualmente di mappare completamente il significato e l'utilizzo dei termini che caratterizzano il dominio stesso.

Allo stesso modo, seguendo la matrice complessità-valore analizzata nell'allegato 38 sarà possibile implementare casi d'uso simile all'interno di altri domini. Questo permetterà nel lungo periodo di creare connessioni tra i termini di business di domini differenti, definendo con chiarezza la proprietà dei dati sottostanti.

In concomitanza, la definizione e sviluppo degli altri 7 processi della data governance, permetterà di espandere lo scopo dei test, come quello appena analizzato. Questo significa che i termini di business saranno arricchiti con nuovi attributi, i quali permetteranno di creare sinergie e estrarre sempre più valore dal loro utilizzo. Questo permetterà la definizione di casi d'uso completi che, sebbene contestualizzati ad ogni dominio, consentiranno di testare tutti i processi della data governance. L'espansione di questi casi d'uso, eventualmente, toccherà ogni dominio dell'azienda, rendendo la data governance un grande processo verticale a tutte le funzioni aziendali. Nel seguente capitolo si analizza il framework della maturità della data governance, che permette di delineare una roadmap per lo sviluppo del progetto della data governance e la sua diffusione, fino a raggiungere l'organizzazione intera.

9.3. DATA MATURITY FRAMEWORK

Un framework della maturità della è un modello strutturato di alto livello per monitorare il grado di maturità dell'azienda rispetto a uno specifico argomento o all'implementazione di un progetto. In questo caso, il Data Governance Maturity Framework ha l'obiettivo di disegnare una chiara roadmap per l'azienda per sviluppare, implementare e migliorare le pratiche di data governance, includendo anche gli strumenti per valutare in modo qualitativo e quantitativo lo stato di maturità effettivo e quello desiderato, identificando i passi necessari e le azioni da intraprendere per raggiungere il risultato desiderato.

Nell'ambito di definizione delle politiche e principi del framework della data governance, è stato disegnato il framework di maturità aziendale relativo alla data governance, che verrà implementato e utilizzato a seguito della fase pilota e del test degli use case. Questo è il motivo per cui nell'allegato 16 il data maturity framework è rappresentato dopo la fase di pianificazione. Nonostante questo sia stato definito, attualmente l'azienda non ha sufficiente visibilità sui processi in atto per poterlo adottare e misurare.

Il Framework è composto da:

- Livelli di maturità, ognuno dei quali rappresenta un diverso stadio di sviluppo nel percorso di governance dei dati della ST, dalle fasi iniziali a quelle ottimizzate e pienamente integrate;
- Principi guida, che rappresentano le linee guida, la roadmap e le raccomandazioni per stabilire i requisiti che la ST deve soddisfare per passare al livello di maturità successivo. Sono incluse fasi specifiche, tappe fondamentali e tempistiche;
- Criteri di valutazione, utilizzati per determinare lo stato della ST a ciascun livello del quadro di maturità. Questi sono rappresentati dalle metriche definite

9.3.1. I LIVELLI DI MATURITÀ

I livelli di maturità identificati permettono di articolare il progetto di adozione delle pratiche di data governance in vari stadi, garantendo così alle figure dirigenziali e agli esecutivi dell'azienda una visione chiara e costante dello stato di avanzamento. I livelli identificati sono 6:

- **Livello di maturità 0 (No Capability):** Le politiche di data governance, i ruoli e i processi non sono definiti a livello aziendale. Le pratiche relative ai dati sono incoerenti e ci sono potenziali lacune nella qualità e nell'integrità dei dati.
- **Livello di maturità 1 (Initial):** viene riconosciuta l'importanza della data governance, ed è stata stabilita una struttura formale; tuttavia, l'adozione è limitata a pochi esperti. Ruoli e responsabilità sono definiti, ma non assegnati a persone specifiche all'interno dell'organizzazione. Si stanno definendo casi d'uso in diversi domini per affrontare problemi e testare soluzioni. I problemi relativi ai dati sono identificati e categorizzati, ma non affrontati.
- **Livello di maturità 2 (Repeatable):** La consapevolezza aziendale sulla data governance aumenta. I casi d'uso nei vari domini sono per lo più stabiliti, con l'uso di uno strumento centralizzato. Ruoli e responsabilità sono assegnati alle persone corrette, seguendo le politiche ed i processi stabiliti. In questa fase, l'implementazione del progetto non si basa solo su un piccolo gruppo di esperti, ma coinvolge anche gli utenti regolari, che partecipano attivamente alle pratiche di data governance nei loro ruoli. Man mano che i casi d'uso vengono implementati, i problemi relativi ai dati vengono affrontati puntualmente. C'è una consapevolezza organizzativa sulla data governance e sui problemi relativi ai dati, ma le

migliori pratiche e i processi formali sono definiti ma ancora da implementare in tutta l'organizzazione;

- **Livello di maturità 3 (Defined):** I ruoli, le responsabilità, i processi e gli standard definiti sono parzialmente compresi e applicati da vari team, domini o casi d'uso. Anche se non ancora pienamente pervasivi, queste pratiche assumono sempre di più la forma di routine, integrate nelle attività aziendali quotidiane e, infine, è necessario sempre meno intervento manuale poiché viene raggiunta la definizione formale dei processi. L'iniziale standardizzazione dei processi potrebbe consentire di scalare la data governance in tutta l'organizzazione;
- **Livello di maturità 4 (Governed):** processi, politiche e standard non sono solo definiti, ma vengono anche applicati in modo coerente in tutta l'organizzazione ST grazie a una pianificazione centralizzata e a funzioni di governance stabilite. Queste pratiche sono gestite autonomamente dai vari dipartimenti, indicando un forte cambiamento culturale verso la valorizzazione della data governance. Vengono utilizzate metriche sia qualitative che quantitative per monitorare e valutare l'adozione e l'efficacia degli sforzi di data governance;
- **Livello di maturità 5 (Optimization):** L'organizzazione ha integrato e standardizzato completamente i processi, le politiche e gli standard di data governance. Questi sono monitorati rigorosamente mediante metriche ben comprese per garantire la conformità e l'efficacia continuative. Le metodologie stabilite per rivedere e migliorare le pratiche di data governance sono focalizzate sul miglioramento continuo per rispondere alle esigenze in evoluzione.

9.3.2. I PRINCIPI GUIDA

La seguente serie di principi guida, attività e milestone, dovrebbe definire gli obiettivi a lungo e a breve termine per consentire all'organizzazione di realizzare progressi misurabili nell'implementazione del progetto di Data Governance, permettendo di passare da un livello di maturità a quello successivo.

PASSAGGIO DAL LIVELLO 0 AL LIVELLO 1

Di seguito, l'insieme di azioni richieste affinché l'organizzazione possa passare dal livello di “No Capability” al livello “Initial”:

- Identificazione e nomina di un Responsabile della Data Governance incaricato di guidare l'iniziativa di data governance;
- Creazione di una bozza iniziale dello Statuto di Data Governance che delinea visione, missione, obiettivi e ambito;
- Identificazione di alcuni progetti pilota per affrontare immediatamente i problemi relativi ai dati e testare soluzioni potenziali;
- Identificazione di ruoli e responsabilità per i processi di data governance.

PASSAGGIO DAL LIVELLO 1 AL LIVELLO 2

Di seguito, l'insieme di azioni richieste affinché l'organizzazione possa passare dal livello “Initial” al livello “Repeatable”:

- Garanzia del sostegno da parte degli esecutivi aziendali;
- Definizione chiara e documentazione dei processi, delle politiche e delle metriche di data governance;

- Assegnazione di ruoli e responsabilità specifici per i processi di data governance come la gestione dei metadati, la gestione della qualità dei dati, ecc.;
- Selezione e implementazione di strumenti e tecnologie centralizzati per supportare le attività di data governance;
- Implementazione di casi d'uso in vari domini per affrontare problemi specifici di data governance e dimostrare il valore di una data governance strutturata;
- Sviluppo di una formazione completa per gli utenti nei domini in cui viene implementato il progetto di DG, spiegando il nuovo modello di data governance, sviluppando materiali di formazione e guide per gli utenti.

PASSAGGIO DAL LIVELLO 2 AL LIVELLO 3

Di seguito, l'insieme di azioni richieste affinché ST possa passare dal livello "Repeatable" al livello "Defined":

- Espansione dei casi d'uso della Data Governance a livello aziendale;
- Sviluppo di workshop per le figure esecutive dell'organizzazione con il fine di illustrare i benefici della data governance con esempi reali e casi di studio, nonché per gli utenti funzionali e tecnici, concentrandosi sui loro ruoli e sul loro coinvolgimento nei processi di data governance in tutta l'organizzazione. A questo punto, la Data Governance dovrebbe essere conosciuta a livello di tutta l'organizzazione;
- Creazione di una struttura di supporto o di una community per rispondere alle domande e fornire orientamento sulle politiche e sugli standard di data governance.

PASSAGGIO DAL LIVELLO 3 AL LIVELLO 4

Di seguito, l'insieme di azioni richieste affinché ST possa passare dal livello "Defined" al livello "Governed":

- È provato che i processi e gli standard di data governance siano applicati in modo coerente in tutta l'organizzazione;
- Sviluppo di una documentazione dettagliata dei processi e di procedure operative standard;
- Definizione e adozione di metriche qualitative e quantitative per monitorare l'efficacia della data governance (qualità dei dati, sicurezza e privacy dei dati, ecc.);
- Sviluppo e applicazione di un modello di valutazione della maturità per valutare la maturità della data governance dell'organizzazione;
- Promozione dello sviluppo e dell'esecuzione di casi d'uso che supportino i processi aziendali e dimostrino il valore della data governance.

PASSAGGIO DAL LIVELLO 4 AL LIVELLO 5

Di seguito, l'insieme di azioni richieste affinché ST possa passare dal livello "Governed" al livello di "Optimization":

- Adozione di metodologie strutturate per il monitoraggio e la valutazione continui dei processi di data governance e implementazione di processi di miglioramento continuo per adattare e migliorare le pratiche di data governance;
- Organizzazione di revisioni e aggiornamenti regolari delle politiche e degli standard di data governance per riflettere i nuovi requisiti aziendali;
- Promozione della cultura basata sui dati attraverso campagne di educazione e sensibilizzazione continue;

- Creazione di meccanismi per monitorare i benefici e la creazione di valore derivanti dall'adozione delle politiche e degli standard di data governance.

9.3.3. CRITERI DI VALUTAZIONE

Dopo aver definito il livello di maturità che serve da roadmap per l'implementazione del processo di data governance all'interno dell'organizzazione e le linee guida per passare da un livello al successivo, è fondamentale stabilire un modello quantitativo per misurare il livello di adozione della data governance rispetto ai processi definiti e associarlo a tutti i livelli sopra citati.

Per effettuare una valutazione della maturità dell'organizzazione rispetto alla data governance è necessario prendere in considerazione i seguenti temi:

- Gli ambiti della valutazione (ovvero gli 8 processi della data governance)
- Le dimensioni della valutazione
 - Coinvolgimento e Adozione: ovvero la scala di comprensione e adozione del modello di Data Governance e dei tool a supporto a livello aziendale;
 - Tool e Tecnologie: la disponibilità degli strumenti di Data Governance (es. Business Glossary) e delle tecnologie abilitanti (es. Catalogo dei dati);
 - Politiche e Standard: livello di definizione e adozione delle politiche e degli standard definiti in ambito Data Governance;
 - Realizzazione di valore: comprensione e adozione di metriche quantitative per monitorare i benefici della Data Governance e il valore generato dai dati aziendali.
- Scala di valori (ovvero i livelli di maturità selezionati)

Attraverso questi elementi, per ognuno dei processi menzionati, è possibile definire quantitativamente il livello di maturità del processo stesso. L'allegato 51 mostra la tabella per la valutazione della maturità del processo di metadata management.

Sarà compito della Steering Committee, congiuntamente con il DGO, completare il modello di valutazione per ognuno degli altri sette processi menzionati. Per ultimo, effettuando una media della valutazione dei processi, è possibile ottenere il livello di maturità dell'organizzazione rispetto alla data governance. L'allegato 52 mostra un esempio di valutazione globale dei processi rispetto alla loro maturità nell'ambito del governo dei dati. È importante notare che è possibile ricavare una metrica quantitativa sia globale, sia sui singoli processi, sia sulle dimensioni identificate.

Per quanto riguarda la maturità dell'organizzazione supportata da Accenture, è troppo presto per definirne un livello di maturità quantitativamente. Siccome solo uno degli otto processi è stato definito ma ancora non implementato (metadata management), seguendo la descrizione del livello di maturità, è possibile dire che l'azienda si trova in una fase di transizione tra il livello 0 e il livello 1.

10. CONCLUSIONI

Come anticipato precedentemente, le azioni intraprese nel Q2 del 2024 nell'ambito del progetto di data governance, fanno parte di un piano di almeno due anni per supportare l'organizzazione a effettuare la transizione da un modello di gestione dei dati centralizzato, ad uno federato.

È evidente come le attività svolte sino ad ora hanno il fine di porre delle solide basi per definire e di seguito implementare i processi della data governance. Facendo riferimento all'allegato 17, i prossimi passi prevederanno la graduale definizione e introduzione degli 8 processi della data governance.

In concomitanza, verranno definiti i primi casi d'uso per implementare i processi definiti.

Man mano che gli 8 processi della data governance saranno definiti, lo scopo dei casi d'uso verrà espanso, migliorando la profondità e la cura dei dati dell'area funzionale. Di seguito, sono analizzate le difficoltà riscontrate nell'implementazione del progetto di data governance, e le attività che sono state portate avanti per risolvere:

- **Mancanza di supporto dai vertici:** l'assenza di sponsor esecutivi ha rallentato l'adozione delle pratiche di data governance. Senza il forte supporto e l'impegno della leadership aziendale, i progetti di data governance possono infatti mancare di risorse e priorità. Molto spesso succede perché l'implementazione di questi progetti avviene parallelamente allo svolgimento delle attività quotidiane dei domini. Questo significa che molto spesso i leader delle aree funzionali e le figure esecutive, se sottovalutano l'importanza di questi progetti, possono causare ritardi non voluti. È stato fondamentale articolare chiaramente i benefici del progetto di data governance, evidenziando come migliorerà i processi decisionali e operativi. Inoltre, l'utilizzo di use case passati e l'esperienza pregressa sono stati fondamentali per assicurare il coinvolgimento di figure che possano promuovere l'iniziativa all'interno dell'azienda;
- **Resistenza al cambiamento:** le organizzazioni spesso incontrano resistenza interna da parte dei dipendenti che sono abituati a lavorare con metodi e processi esistenti, soprattutto in multinazionali di grandi dimensioni. Inoltre, molto spesso le attività di data governance vengono svolte in aggiunta al carico di lavoro predefinito, soprattutto per gli utenti di business come i data owner e i data steward. In questo senso è fondamentale calcolare chiaramente la quantità di tempo dedicabile da ogni figura, per evitare di sovraccaricare il dipendente con nuove attività;
- **Complessità tecnica:** La data governance comprende una vasta gamma di processi e attività che richiedono competenze diversificate e specializzate. La carenza di conoscenze specifiche e di expertise in materia tra i dipendenti può costituire un significativo impedimento al successo del progetto. Questo è uno dei motivi per cui le aziende frequentemente ricorrono al supporto esterno per l'implementazione di tali progetti. Tuttavia, è essenziale che durante il corso del progetto l'azienda investa risorse adeguate nella formazione e nell'acquisizione di personale esperto nella gestione dei dati. Questo processo richiede tempo ed è normale che in una fase iniziale la mancanza di esperienza possa rallentare le operazioni e l'implementazione dei processi;
- **Complessità infrastrutturale:** come citato, per l'implementazione dei processi di data governance si trova supporto nelle strutture di catalogazione dei dati. Queste piattaforme quasi sempre sono caratterizzate da delle interfacce intuitive che permettono la loro semplice gestione. Ad ogni modo, l'implementazione di nuove piattaforme richiede tempo. A priori, è necessaria una dettagliata analisi dei requisiti tecnici per selezionare le piattaforme più appropriate che si integrino più facilmente con i sistemi esistenti;

- Misurazione degli obiettivi: definire metriche di successo e KPI per monitorare l'efficacia della data governance è essenziale ma può essere complicato. Tra tutti, il ritorno degli investimenti in progetti di data governance è molto difficile da valutare, come evidenziato nel capitolo 5 (Il valore dei dati). Questo è un altro motivo per cui assicurare il coinvolgimento delle figure dirigenziali spesso risulta complicato.

Nonostante le sfide e le complessità incontrate nell'implementazione del progetto di data governance, il trend di crescita esponenziale dei dati e le nuove applicazioni emergenti, come l'intelligenza artificiale, renderanno imprescindibile per le organizzazioni gestire correttamente i propri dati. In futuro, le aziende che sapranno valorizzare i propri dati attraverso pratiche di governance ben strutturate saranno in grado di estrarre un valore significativamente maggiore, rispetto a quelle che continueranno a trascurare questo aspetto cruciale. La capacità di sfruttare i dati in modo efficace sarà il fattore determinante che distinguerà le organizzazioni di successo in un panorama tecnologico e competitivo in continua evoluzione.

11. ALLEGATI

11.1. ALLEGATO 1: DATA GOVERNANCE E DATA MANAGEMENT

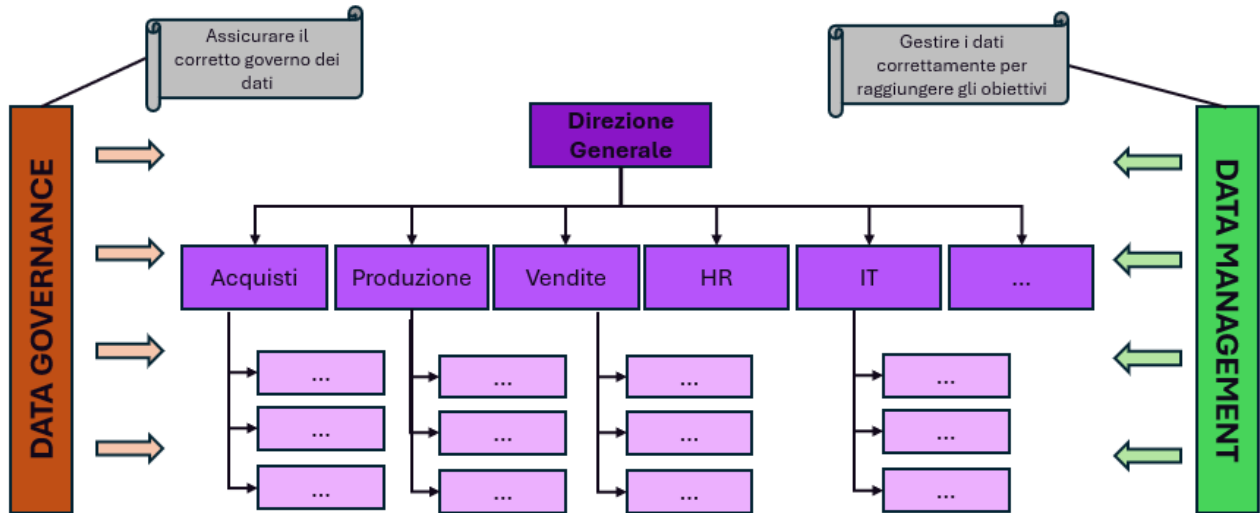


Figura 1: Il data management e la data governance come funzioni strategiche verticali nelle aziende

11.2. ALLEGATO 2: PIRAMIDE DIKW



Figura 2: La Piramide DIKW

11.3. ALLEGATO 3: IL CICLO DEI DATI E DEL SAPERE

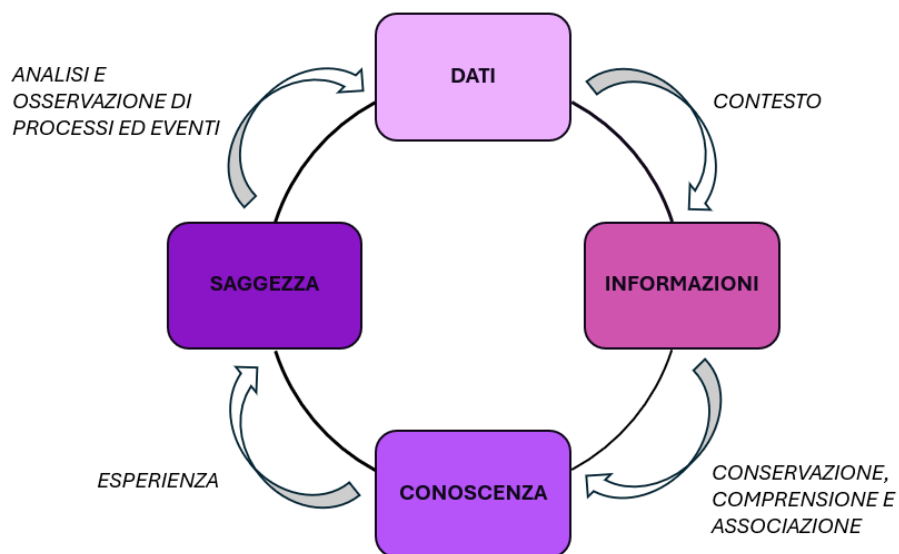


Figura 3: La relazione ciclica tra dati, informazioni, conoscenza e saggezza

11.4. ALLEGATO 4: VOLUME DI DATI CREATI

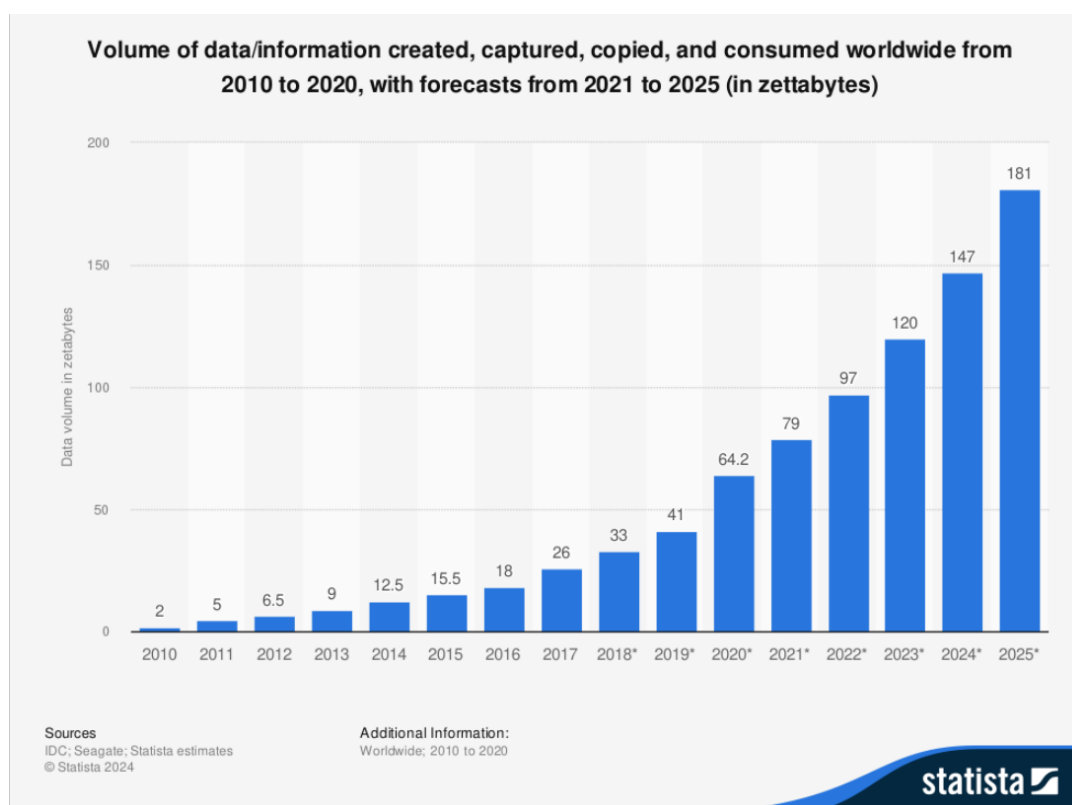


Figura 4: Statista, volume di dati creati e consumati globalmente all'anno in zettabytes dal 2010 al 2025

11.5. ALLEGATO 5: DIMENSIONE DI UNO ZETTABYTE

$$1 \text{ zettabyte} = 10^{21} \text{ byte}$$

Poiché un byte è composto da 8 bit, abbiamo:

$$1 \text{ zettabyte} = 8 * 10^{21} \text{ bit}$$

Ora, considerando che un video HD di alta qualità consuma circa 20 gigabyte (GB) di spazio di archiviazione per ogni ora di riproduzione, convertiamo 20 GB in bit:

$$20 \text{ GB} = 20 * 8 * 10^9$$

Quindi, dividiamo il numero totale di bit in uno zettabyte per il numero di bit in 20 GB:

$$\frac{8 * 10^{21}}{20 * 8 * 10^9} = 50.000.000.000 \text{ ore di video in HD}$$

$$\frac{50.000.000.000}{24 \text{ ore} * 365 \text{ giorni}} = 5.700 \text{ anni}$$

11.6. ALLEGATO 6: IL CICLO DI VITA DEI DATI E DATA MANAGEMENT

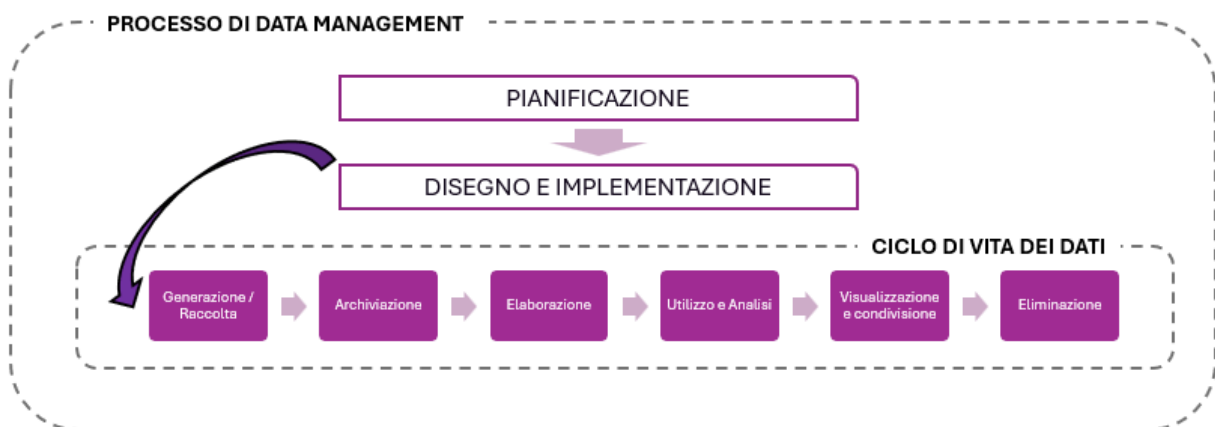


Figura 5: La relazione tra il ciclo di vita dei dati e il processo di data management

11.7. ALLEGATO 7: DATI STRUTTURATI E NON STRUTTURATI

Caratteristica	Dati Strutturati	Dati Non Strutturati
Struttura	Formato predefinito, organizzati in tabelle con righe e colonne.	Nessun formato predefinito, senza struttura organizzata.
Esempi	Numeri, testo breve, date, tabelle Excel, database relazionali...	Testi, video, immagini, audio, dati web, dati dei sensori...
Organizzazione	Facilmente organizzati e rappresentabili in database relazionali (SQL).	Organizzati in file system, database NoSQL, data lakes, cloud storage.
Volume di dati dei database gestibili	Generalmente gestiscono volumi moderati di dati, ma sono scalabili.	Progettati per gestire grandi volumi di dati eterogenei.

Interoperabilità	Alta interoperabilità con strumenti tradizionali di elaborazione dati.	Alta interoperabilità grazie alla loro origine digitale e formati flessibili.
Elaborazione	Analizzabili con strumenti convenzionali come SQL.	Richiedono strumenti avanzati e specifici per l'analisi, come NLP, strumenti di elaborazione video/audio.
Tempo di Creazione	Dati raccolti periodicamente e aggiornati regolarmente.	Creati rapidamente e continuamente, spesso in tempo reale.
Distribuzione	Solitamente centralizzata.	Distribuiti spazialmente, spesso su sistemi di storage distribuiti.
Scalabilità	Scalabilità limitata alla capacità del database relazionale.	Altamente scalabili grazie a database NoSQL, data lakes e cloud storage.
Accessibilità	Accesso controllato tramite DCL e SQL.	Accesso tramite varie API e servizi cloud, con funzionalità di backup e versioning.
Utilizzo	Adatti per applicazioni aziendali tradizionali, reportistica, e analisi strutturata.	Utilizzati per analisi di grandi dati, machine learning, e applicazioni multimediali.

Tabella 1: Le differenze tra dati strutturati e non strutturati

11.8. ALLEGATO 8: DATABASE RELAZIONALE

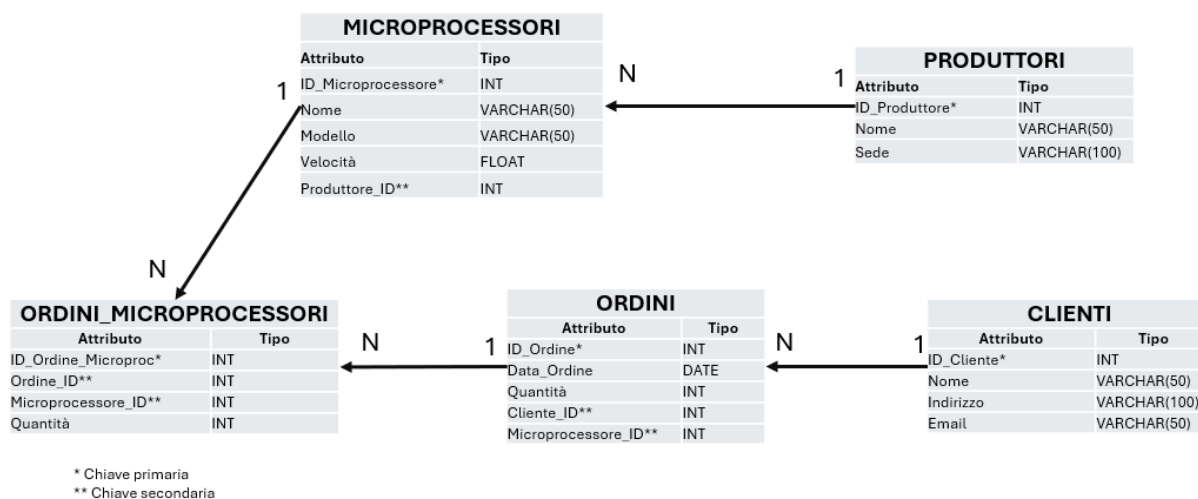


Figura 6: Rappresentazione semplificata di un database relazionale

11.9. ALLEGATO 9: FRAMEWORK NIST



Figura 7: Framework del NIST per la Cybersecurity

11.10. ALLEGATO 10: ARTEFATTI DELLA DATA GOVERNANCE

Elemento	Contenuto	Esempio
Business Glossary	Definizioni e descrizioni dei termini e delle espressioni utilizzati nel settore.	Termine: Wafer Definizione: Una sottile fetta di materiale semiconduttore, come il silicio, su cui sono costruiti i circuiti integrati.
Data Dictionary	Dettagli tecnici dei dati, incluse le descrizioni degli attributi, i tipi di dati e le restrizioni.	Tabella: Microprocessori Attributo: Frequenza Tipo di Dato: Float Descrizione: La velocità operativa del microprocessore in GHz.
Data Lineage	Tracciabilità dell'origine, del movimento e delle trasformazioni dei dati nel sistema.	Fonte Dati: Sensore di Temperatura Trasformazione: Dati aggregati ogni 5 minuti Destinazione: Dashboard di Monitoraggio Produzione.

Tabella 2: Gli artefatti della data governance e un esempio

11.11. ALLEGATO 11: I PRINCIPI DEL DATA MESH

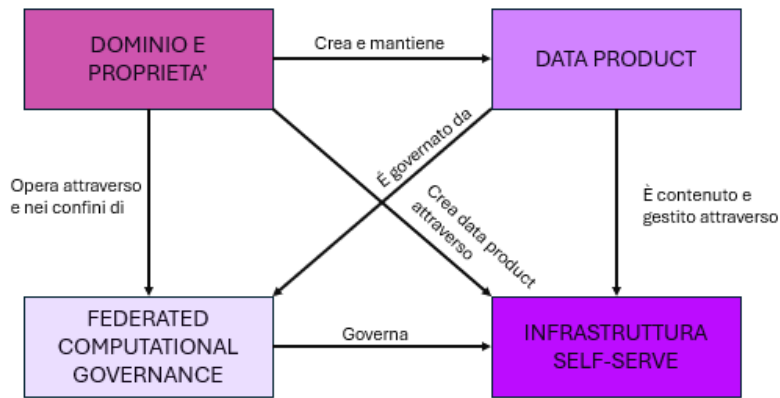


Figura 8: I principi del data mesh

11.12. ALLEGATO 12: TIPI DI SEMICONDUTTORI

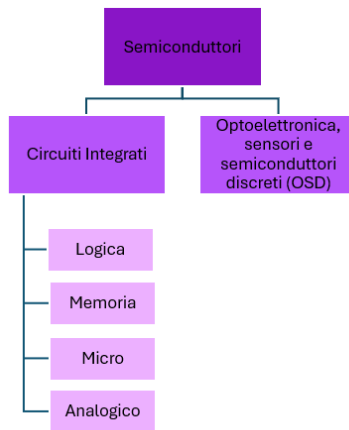


Figura 9: Tipi di semiconduttori

11.13. ALLEGATO 13: VENDITE GLOBALI DI SEMICONDUTTORI 2021 PER TIPO

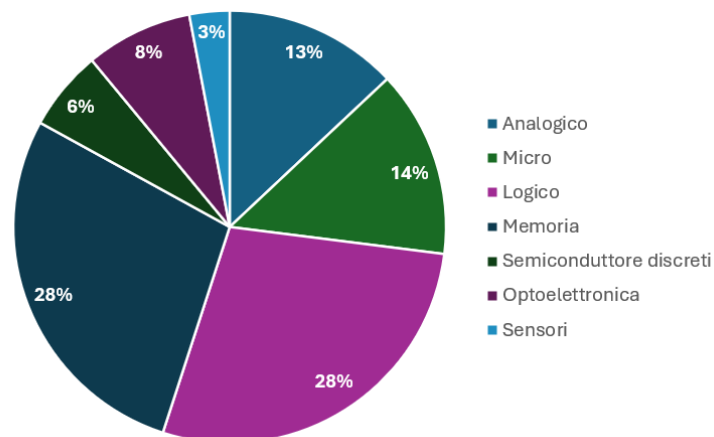


Figura 10: Vendite globali per tipologia di semiconduttori nel 2021

11.14. ALLEGATO 14: LEGGE DI MOORE

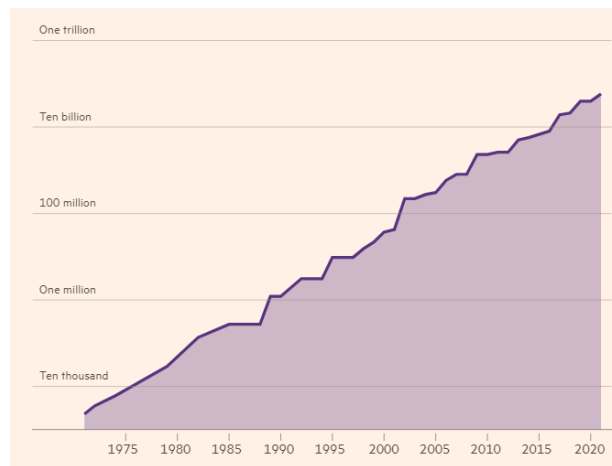


Figura 11: Numero di transistors per microprocessore

11.15. ALLEGATO 15: MERCATO DEI SEMICONDUCTORI PER APPLICAZIONE

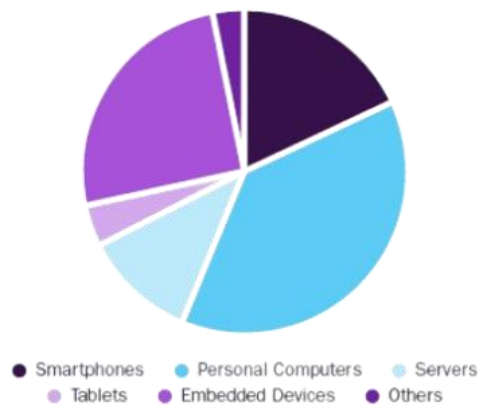


Figura 12: Mercato dei semiconduttori per applicazione

11.16. ALLEGATO 16: ROADMAP DI ALTO LIVELLO

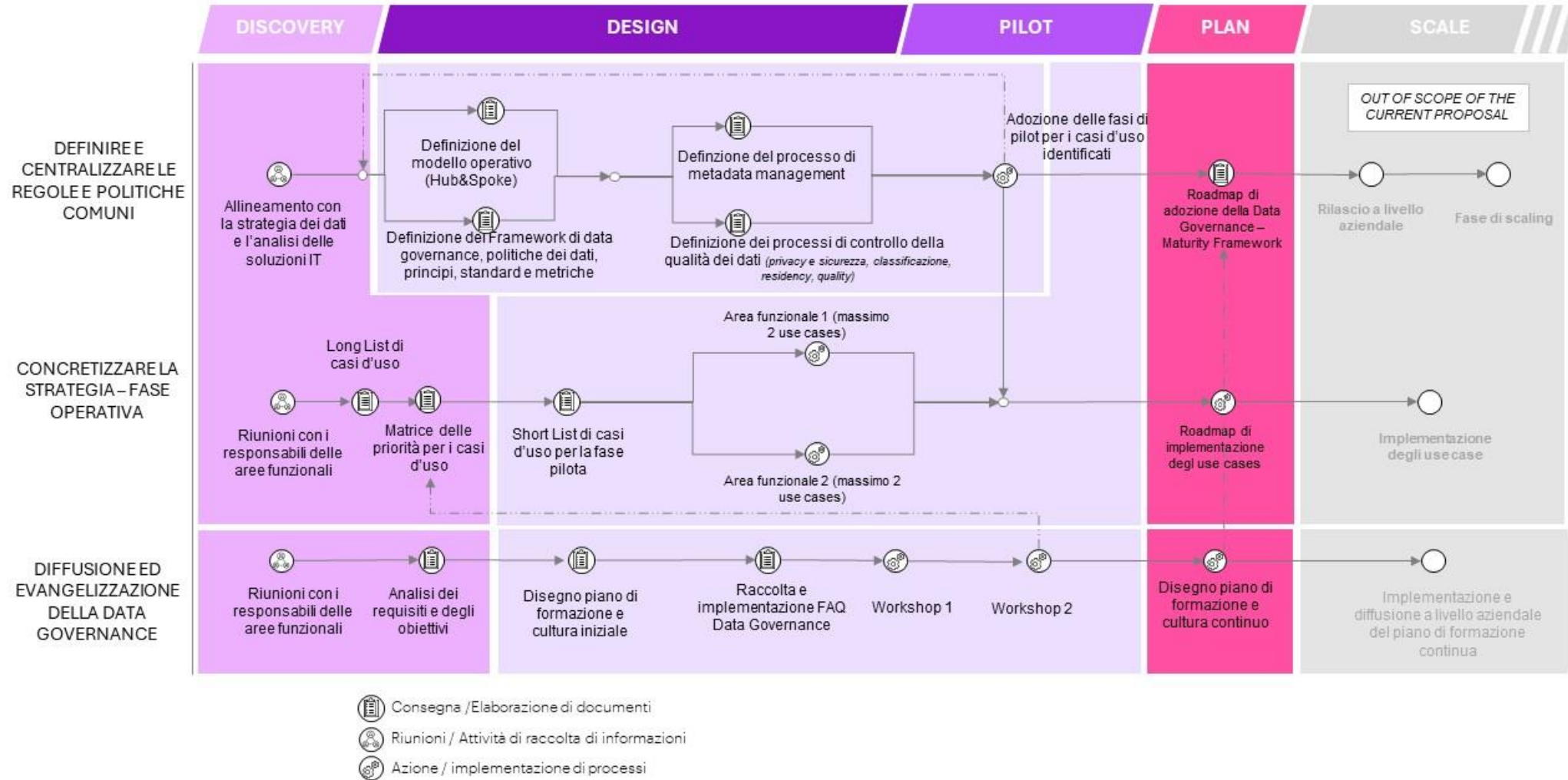


Figura 13: Roadmap di alto livello per il seguito del progetto di data governance

11.17. ALLEGATO 17: ANDAMENTO DEL PROGETTO

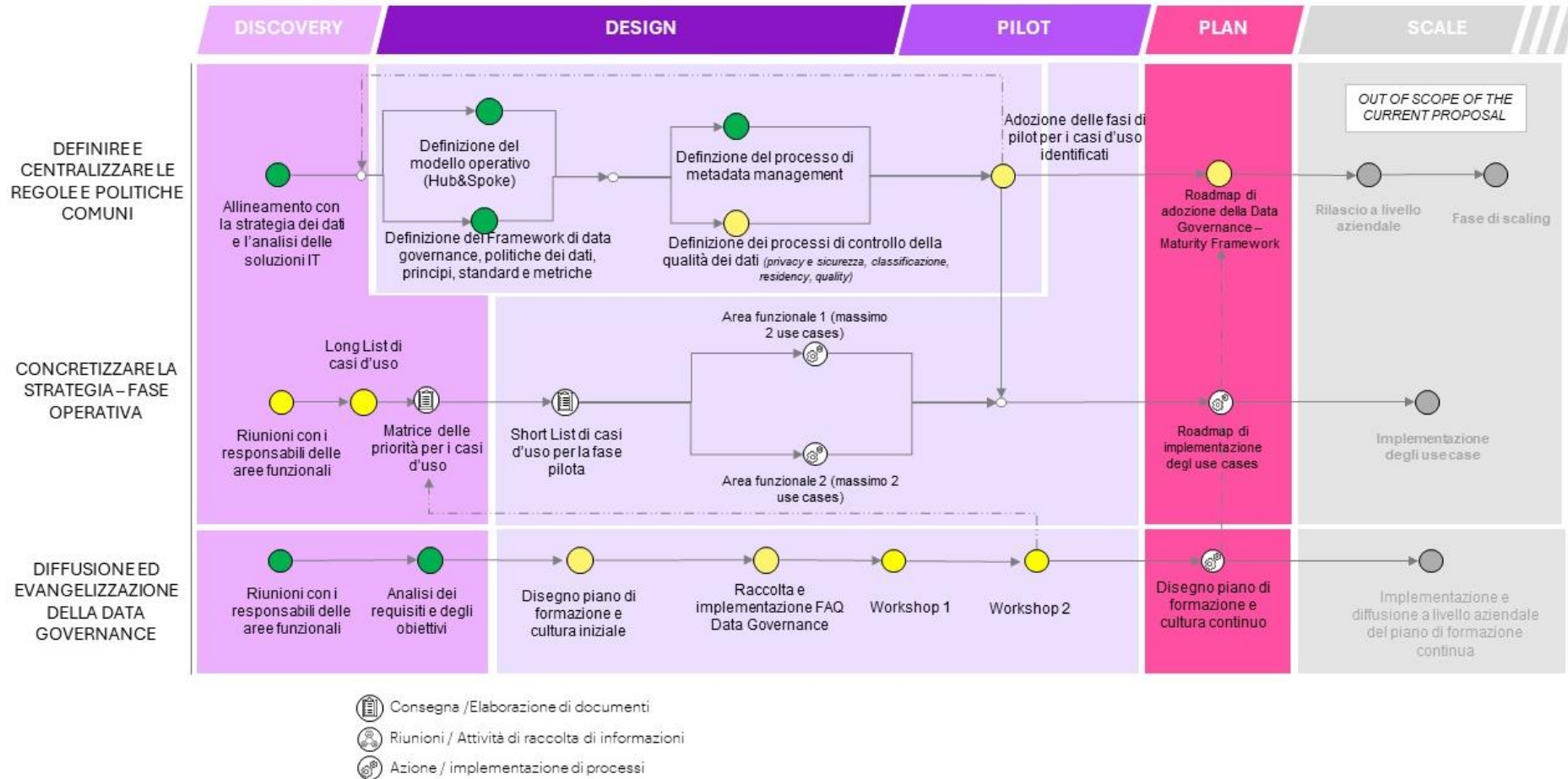


Figura 14: Rappresentazione dell'andamento del progetto attraverso la roadmap di alto livello

11.18. ALLEGATO 18: IL FRAMEWORK DI DATA GOVERNANCE

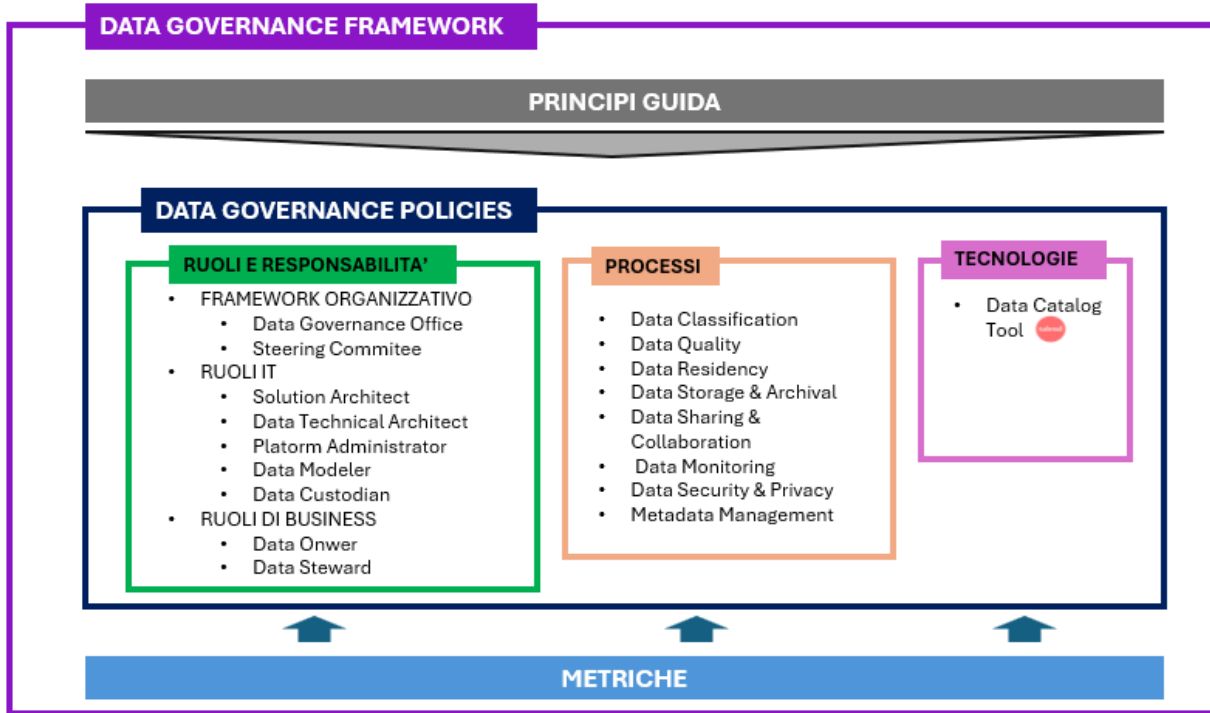


Figura 15: Il framework di data governance adottato

11.19. ALLEGATO 19: IL MODELLO HUB & SPOKE

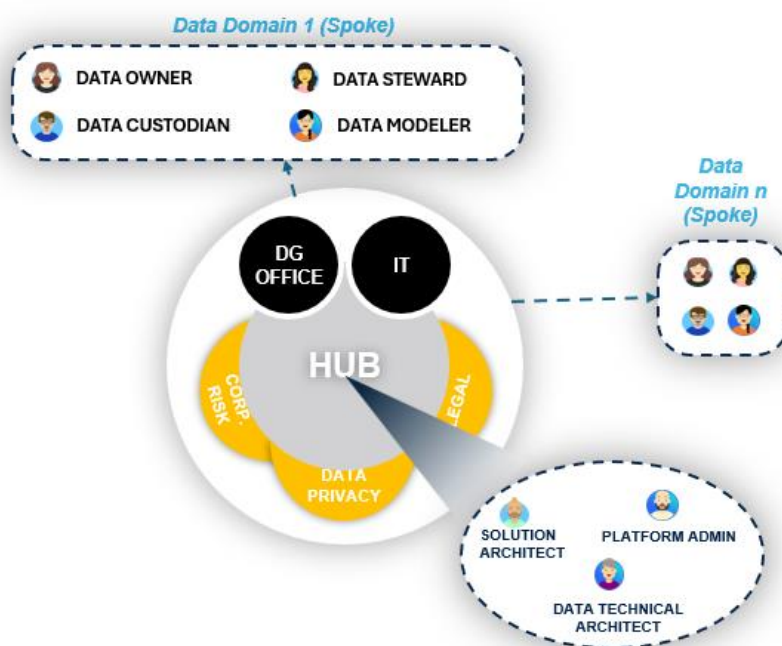


Figura 16: rappresentazione grafica del modello Hub&Spoke

11.20. ALLEGATO 20: SECRETEZZA DELLE INFORMAZIONI

LIVELLO	PUBBLICO	RISTRETTO	CONFIDENZIALE	SEGRETO
DESCRIZIONE	Informazioni in possesso dell'organizzazione che sono state rese disponibili al pubblico, attraverso la loro pubblicazione sul sito web aziendale, o altri fonti di comunicazione esterne all'azienda.	Informazioni processate a partire di attività giornaliere, la quale alterazione, condivisione o perdita, potrebbe causare un danno moderato all'azienda.	Informazioni la quale alterazione, condivisione o perdita potrebbe causare danni significativi all'organizzazione	Informazioni la quale alterazione, condivisione o perdita potrebbe causare danni catastrofici e/o irreversibili all'organizzazione
ESEMPIO	<ul style="list-style-type: none"> • Annunci pubblici • Informazioni sulle performance finanziarie • Materiale di marketing • Offerte di lavoro • Post sui social media 	<ul style="list-style-type: none"> • Annunci interni all'organizzazione • Informazioni relative a iniziative interne generali • Informazioni non sensibili riguardo allo sviluppo di progetti o prodotti 	<ul style="list-style-type: none"> • Informazioni sensibili riguardo allo sviluppo di progetti o prodotti • Informazioni sensibili riguardo a proprietà intellettuali • Documentazione finanziaria non pubblica • Strategia di mercato • PII di dipendenti, clienti, partner... 	<ul style="list-style-type: none"> • Dettagli di progetti di M&A • Informazioni e PII sensibili • Informazioni strategiche sensibili •

Tabella 3: Classificazione della sensibilità delle informazioni nell'organizzazione

11.21. ALLEGATO 21: ATTRIBUTI DEL BUSINESS GLOSSARY

Attributo	Descrizione
Nome (M)	Nome del Termine
Contesto (M)	Informazioni sulla struttura gerarchica del glossario aziendale a cui appartiene il termine aziendale
Tipo	Tipo di termine (termine aziendale o acronimo) da inserire nel glossario aziendale
Descrizione (M)	Descrizione dettagliata del termine
Regola Logica	Indicazione della regola logica (o regola di calcolo), se esiste, utilizzata per creare le informazioni a cui si riferisce il termine aziendale
Contiene	Si riferisce ai possibili termini che possono essere contenuti o strettamente correlati al termine aziendale in modo gerarchico; questo attributo generalmente si riferisce all'elenco di termini aziendali che sono a un livello gerarchico inferiore rispetto al termine aziendale in questione
Etichette	Etichette opzionali che offrono ulteriore contesto riguardo ai dati appartenenti a una categoria specifica o rilevanti per un particolare argomento (es. GDPR, ESG).
Collezioni	Collezione di metadati che ogni utente può creare per personalizzare e facilitare l'accesso e la condivisione di un sottoinsieme specifico di metadati senza dover navigare nel repository per trovarlo (es. collezione contenente l'elenco di termini aziendali da utilizzare per il Report X)
Associazioni	Si riferisce alle possibili associazioni che possono esistere tra termini aziendali a livelli diversi e tra termini aziendali e attributi tecnici; ogni termine aziendale dovrebbe avere almeno un'associazione con un asset tecnico
Area Aziendale (M)	Area aziendale che genera i dati o a cui i dati sono più logicamente associati (es. Market Intelligence, Global Logistic & Warehousing).
Processo Aziendale (M)	Processo aziendale (appartenente all'area aziendale) che genera i dati o a cui i dati sono più logicamente associati (es. Demand Planning).
Data Owner (M)	Utente responsabile del glossario aziendale a livello di dominio.
Data Steward (M)	Responsabile della definizione, esecuzione e cura del glossario aziendale del proprio dominio.
Tipo di Termine (M)	Tipo di campo registrato nel glossario aziendale (termine grezzo, termine calcolato, KPI).
Tipo di Dato (M)	Tipo di dato (elenco di valori: carattere, numerico, data...)
Classificazione o Etichetta di Sensibilità dei Dati	Classificazione dei dati basata sul livello di sensibilità dei dati secondo le politiche interne di ST e, se applicabile, i vincoli normativi.
Sistema di Origine (M)	Sistema che genera asset di dati sotto i termini aziendali definiti.
Master (M) / References (R)	Indica se il termine è un Master Data o un Reference Data

Tabella 4: Attributi del business glossary

11.22. ALLEGATO 22: ATTRIBUTI DEL DATA DICTIONARY

Nome	Descrizione
Nome	Nome dell'asset tecnico (schema, tabella, colonna)
Documentazione Aziendale	Nome aziendale locale o attributo di descrizione per fornire ulteriori informazioni sull'oggetto utilizzando un linguaggio aziendale
Etichette	Etichette opzionali che offrono ulteriore contesto riguardo ai dati appartenenti a una categoria specifica o rilevanti per un particolare argomento (es. GDPR, ESG)
Collezione	Collezioni di metadati che consentono la creazione di gruppi di informazioni per navigare meglio nel catalogo dei dati
Associazioni	Relazioni definite tra diversi oggetti per favorire la comprensione attraverso la creazione di collegamenti logici; ciascun elemento del glossario aziendale dovrebbe essere collegato ad almeno un elemento del dizionario dei dati
Tipo di Dato	Tipo di formato dei dati (es. varchar, numero)
Classificazione o Etichetta di Sensibilità dei Dati	Classificazione dei dati basata sul livello di sensibilità dei dati secondo le politiche interne di ST e, se applicabile, i vincoli normativi
Classi di Dati	Classe o categoria di dati a cui i dati appartengono logicamente (es. riga di indirizzo)
Data Owner	Utente responsabile del glossario aziendale a livello di dominio
Data Custodian	responsabile dell'adempimento e della cura del dizionario dei dati e della linea di provenienza dei dati del proprio dominio
Data di Aggiornamento	Data dell'ultimo aggiornamento

Tabella 5: Attributi del data dictionary

11.23. ALLEGATO 23: ATTRIBUTI DEL DATA LINEAGE

Nome	Descrizione
Applicazioni, sistemi, schemi, tabelle, colonne di origine	Nome dei sistemi, schemi, tabelle, colonne da cui i dati originano
Processi di trasformazione dei dati	Elenco dei processi e delle procedure (ETL) che trasformano i dati
Sistemi, schemi, tabelle, colonne temporanee	Nome dei sistemi, schemi, tabelle, colonne (area di staging) dove i dati sono temporaneamente archiviati per attività di gestione dei dati
Applicazioni, sistemi, schemi, tabelle, colonne di destinazione	Nome dei sistemi, schemi, tabelle, colonne dove i dati sono archiviati e consumati.

Tabella 6: Attributi del data lineage

11.24. ALLEGATO 24: PROCESSO DI METADATA MANAGEMENT

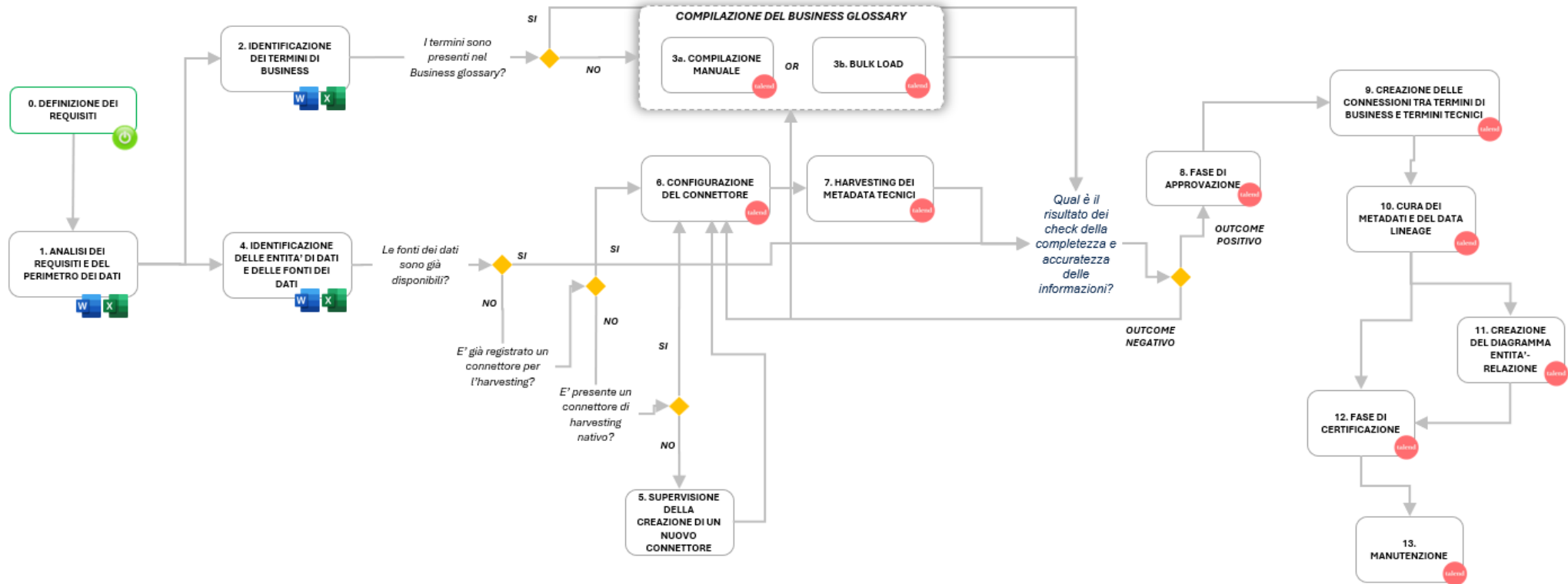


Figura 17: Processo di metadata management completo

11.25. ALLEGATO 25: MATRICE RACI

Ruolo	Definizione	Esempio
Responsible	Questo ruolo rappresenta la persona o le persone che eseguono il lavoro per completare l'attività. Sono coloro che hanno la responsabilità di portare a termine il compito.	Ad esempio, se si tratta di sviluppare un software, gli sviluppatori sono i "Responsible".
Accountable	La persona Accountable è quella che ha la responsabilità ultima del compito. È la persona che deve garantire che l'attività sia completata correttamente e che tutte le responsabilità siano state rispettate. Solo una persona può essere designata come Accountable per ogni attività, per garantire chiarezza e responsabilità univoca.	Ad esempio, in un progetto di sviluppo software, il project manager potrebbe essere Accountable.
Consulted	Questi sono i soggetti che devono essere consultati prima di prendere una decisione finale o completare l'attività. Sono solitamente esperti o stakeholder che offrono input preziosi e feedback. La consultazione implica una comunicazione bidirezionale.	Ad esempio, i designer UX potrebbero essere consultati durante lo sviluppo di un'interfaccia utente.
Informed	Queste persone devono essere informate dei progressi e delle decisioni, ma non sono direttamente coinvolte nell'esecuzione del compito. Ricevono comunicazioni unidirezionali, come aggiornamenti di stato o rapporti di avanzamento.	Ad esempio, la direzione aziendale potrebbe essere informata sugli sviluppi di un progetto.

Tabella 7: Spiegazione della matrice RACI

11.26. MATRICE RACI DEL PROCESSO DI METADATA MANAGEMENT

ID	Nome dell'attività	Ruoli Accountable	Ruoli Responsible	Ruoli Consulted	Ruoli Informed
0	Definizione dei requisiti	Data owner	Data consumer	Data steward	
1	Analisi dei requisiti e del perimetro dei dati	Data owner	Data steward Data custodian	Data consumer Data Modeler	
2	Identificazione dei termini di business	Data owner	Data steward	Data custodian Data modeler Data consumer	DG expert
3a	Compilazione manual del business glossary	Data owner	Data steward	Data consumer	
3b	Bulk load del business glossary	Data owner	Data steward	Data consumer	

4	Identificazione delle entità dei dati e delle fonti dei dati	Data owner	Data custodian	Data steward Data modeler	Data consumer
5	Supervisione della creazione di un nuovo connettore	Platform admin	Platform admin	Data custodian Data modeler	Data owner DG expert
6	Configurazione del connettore	Data owner	Data custodian	Data modeler Platform admin	Data steward
7	Harvesting dei metadata tecnici	Data owner	Data custodian	Data modeler Platform admin	Data steward
8	Fase di approvazione	Data owner	Data owner	Data steward Data custodian Data modeler	DG expert
9	Creazione delle connessioni tra I termini di business e tecnici	Data owner	Data steward Data custodian	Data modeler	Data consumer DG expert
10	Cura dei dati metadata (business e tecnici) e del data lineage	Data owner	Data steward	Data custodian Data modeler	Data consumer
11	Creazione del diagramma entità-relazioni	Data owner	Data modeler	Data custodian Data Steward	Data consumer DG expert
12	Certificazione	Data owner	Data owner	Data steward Data custodian Data modeler	Data consumer DG expert
13	Manutenzione	Data owner	Data steward Data custodian	Data modeler	

Tabella 8: Matrice RACI del processo di metadata management

11.27. ALLEGATO 27: PROCESSO DI METADATA MANAGEMENT PER IL BUSINESS GLOSSARY

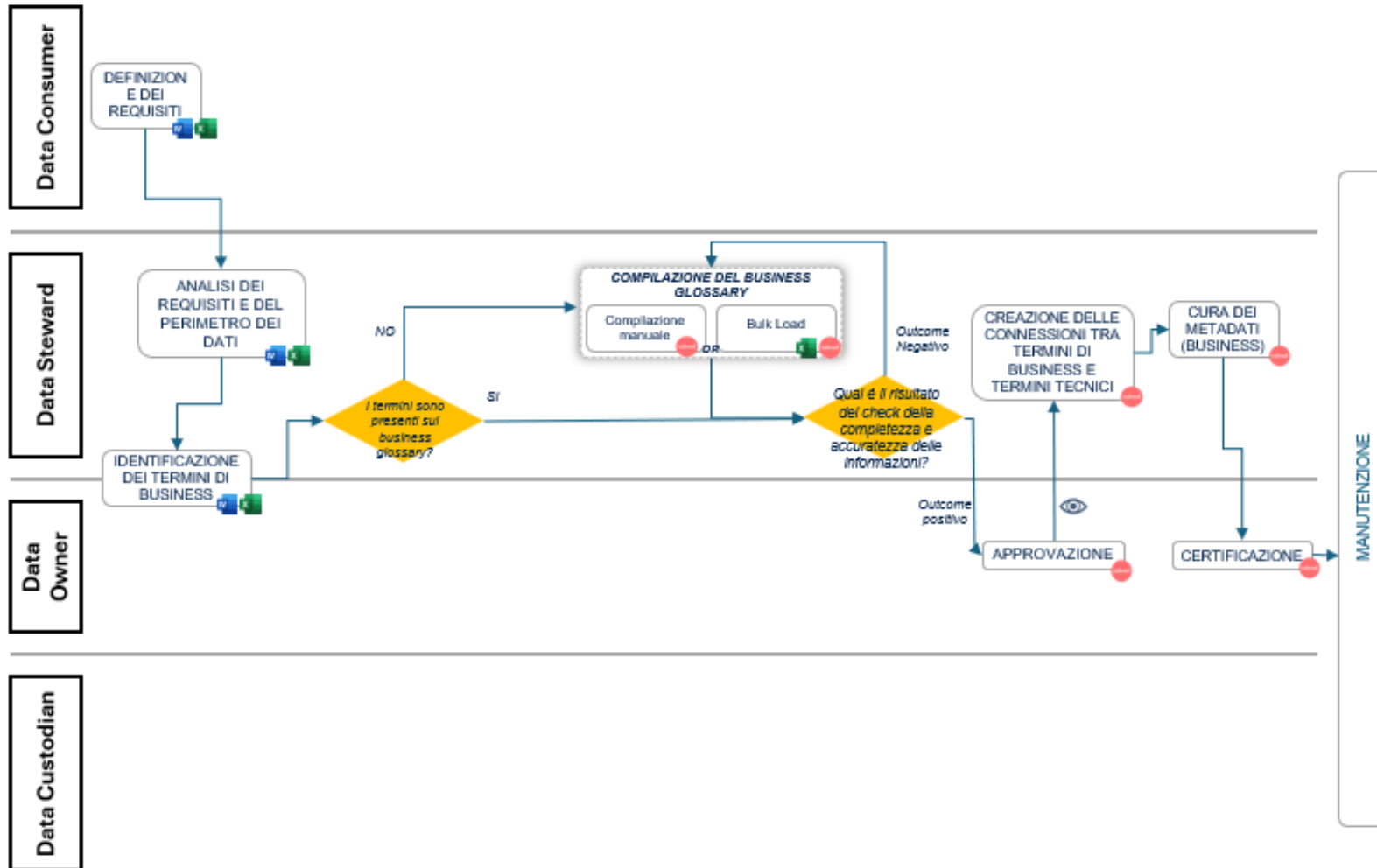


Figura 18: Processo di metadata management relativo al business glossary

11.28. ALLEGATO 28: PROCESSO DI METADATA MANAGEMENT E DATA DICTIONARY

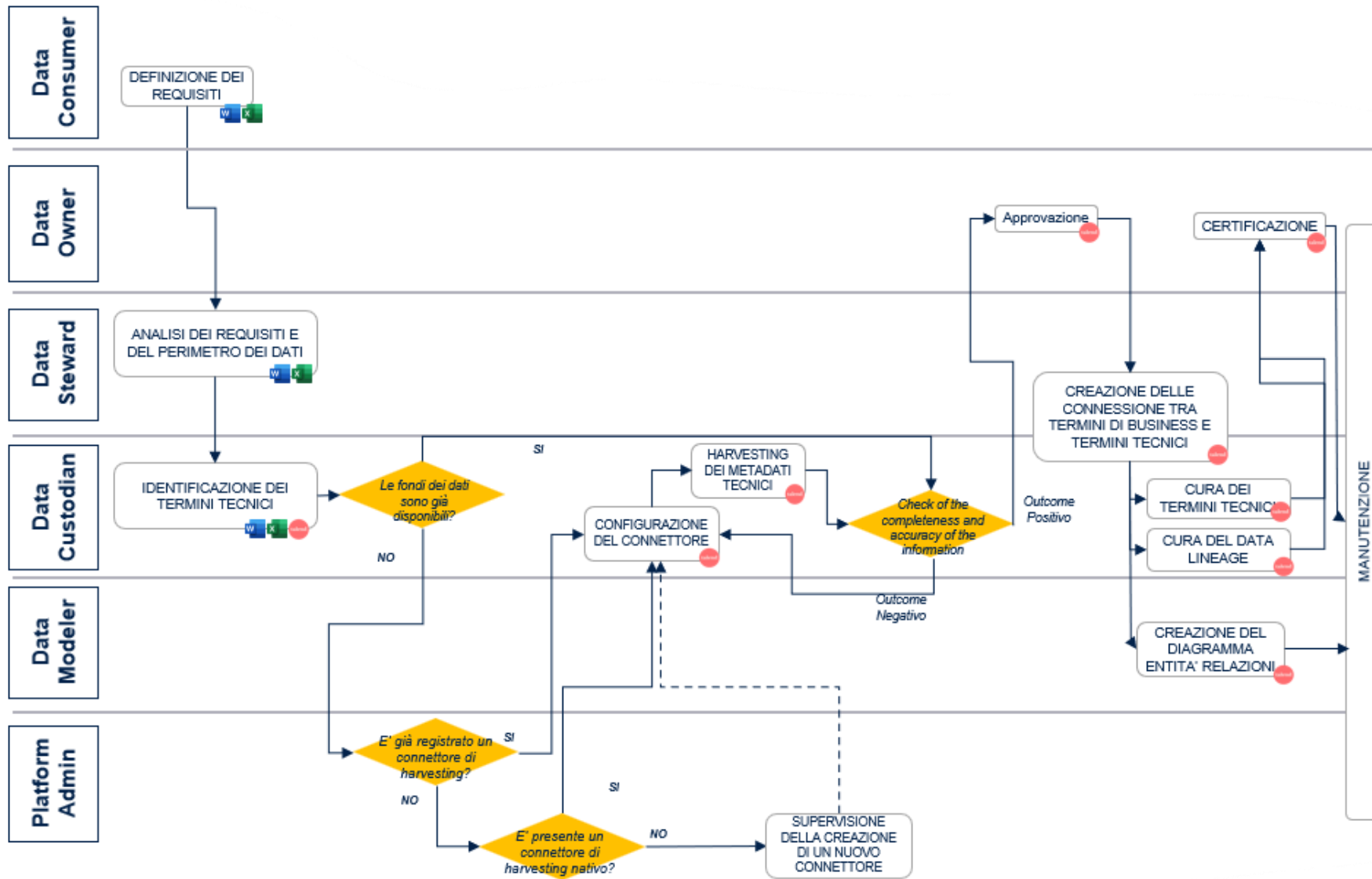


Figura 19: Processo di metadata management e data dictionary

11.29. ALLEGATO 29: TEMPLATE PER LA RICHIESTA DEL DATA CONSUMER

Campo	Descrizione	Esempio
Richiedente	Nome del data consumer che richiede i dati	Maria Rossi
Data Richiesta	Data in cui è stata fatta la richiesta	01/06/2024
Obiettivo	Scopo della richiesta	Analisi di rischio
Necessità dell'analisi	Motivo dettagliato dell'analisi richiesta	Verifica conformità al GDPR
Casi d'Uso	Descrizione dei casi d'uso specifici	Reportistica di conformità
Scenari	Descrizione degli scenari in cui verranno utilizzati i dati	Monitoraggio delle minacce e vulnerabilità aziendali
Tipologia dei Dati	Tipo di dati richiesti (strutturati, non strutturati, ecc.)	Dati strutturati
Fonte dei Dati	Origine dei dati richiesti	Database clienti, Log server
Frequenza	Frequenza con cui i dati devono essere aggiornati	Mensile

Tabella 9: Template per la definizione dei requisiti del data consumer

11.30. ALLEGATO 30: CODICE MACRO DEL BG IMPORT SHEET

MODULO CENTRALE

```
// Definizione delle variabili //
Public Sub CheckAndSave()
    Dim ws As Worksheet
    Dim LastRow As Long
    Dim LastColumn As Long
    Dim errorMessages As String
    Dim hasError As Boolean
    Dim colNamesToCheck As Variant
    Dim colsToCheck As Variant
    Dim cell As Range
    Dim NewFilePath As String
    Dim selectedSheetIndex As Long
    Dim nameCol As Long

    // Chiedere all'utente di selezionare il foglio contenente i campi obbligatori //
    selectedSheetIndex = SelectSheetWithMandatoryFields()
    If selectedSheetIndex = 0 Then
        MsgBox "No sheet has been selected. Operation cancelled. Please, try again.",
        vbExclamation, "Error"
        Exit Sub
    End If

    // Impostare il foglio di lavoro sul foglio selezionato //
    Set ws = ThisWorkbook.Sheets(selectedSheetIndex)

    // Definire il nome delle colonne, ovvero gli attributi obbligatori //
    colNamesToCheck = Array("$Name", "$Type", "$Context", "Description", "Associations", "Term
    Type", "Business Area", "Data Type", "Source System", "Business Process", "Data Steward", "Data
    Owner", "Data Classification / Sensitivity Label")

    // Trova i numeri di colonna corrispondenti per i nomi delle colonna definiti //
    ReDim colsToCheck(1 To UBound(colNamesToCheck) + 1)
    For i = 0 To UBound(colNamesToCheck)
        // Cerca il numero della colonna //
        On Error Resume Next
        colsToCheck(i + 1) = ws.Rows(1).Find(colNamesToCheck(i), LookIn:=xlValues,
        LookAt:=xlWhole).Column
        On Error GoTo 0

    // Verifica se il nome della colonna è stato trovato //
    If colsToCheck(i + 1) = 0 Then
        MsgBox "Column " & colNamesToCheck(i) & " not found in the selected sheet.",
        vbExclamation, "Error"
        Exit Sub
    End If
    Next i

    // Trova il numero di colonna di "$Nome", ovvero la colonna contenente il nome del business term
    //
    nameCol = colsToCheck(1)
```

```
// Trova l'ultima riga della colonna "$Nome" //
LastRow = ws.Cells(ws.Rows.Count, nameCol).End(xlUp).Row

// Trova l'ultima colonna popolata nella riga 1 //
LastColumn = ws.Cells(1, ws.Columns.Count).End(xlToLeft).Column

' Initialize error flag and messages
hasError = False
errorMessages = ""

// Eseguire un loop tra la matrice formata dalle righe e le colonne appena trovate per verificare la
presenza di celle vuote //
For i = 2 To LastRow
    For Each col In colsToCheck
        Set cell = ws.Cells(i, col)

        // Guarda se la cella è vuota //
        If cell.Value = "" Then
            // colora la cella //
            cell.Interior.Color = RGB(255, 0, 0) ' Red color

            // Se è vuota, aggiorna il messaggio di errore con le coordinate della cella trovata //
            hasError = True
            errorMessages = errorMessages & "Cell in row " & i & " and column " & col & " is empty."
        }
    }
    & vbCrLf
    Else
        // Se non è vuota rimuovi la formattazione che cambia il colore della cella in rosso //
        cell.Interior.ColorIndex = xlNone
    End If
    Next col
Next i

// Se ci sono errori, mostra il messaggio di errore contenente le celle evidenziate //
If hasError Then
    MsgBox "The following mandatory fields are missing or empty:" & vbCrLf & errorMessages,
    vbExclamation, "Incomplete Data"
    Exit Sub
End If

// Salva una copia del foglio in csv //
NewFilePath = ThisWorkbook.Path & "\ " & ws.Name & "_compiled.csv"
If ThisWorkbook.Path = "" Then
    // Se la cartella di lavoro non è stata salvata, chiedere all'utente di salvarla prima //
    If Not Application.Dialogs(xlDialogSaveAs).Show Then Exit Sub ' User cancelled saving
    NewFilePath = Application.GetSaveAsFilename(FileFilter:="CSV Files (*.csv), *.csv")
End If

SaveSheetAsCSV ws, NewFilePath, LastRow, LastColumn

MsgBox "The file was correctly saved.", vbInformation, "Success"
End Sub
```

SOTTOFUNZIONE PER DETERMINARE IL FOGLIO DI INTERESSE

```
Public Function SelectSheetWithMandatoryFields() As Long
    Dim ws As Worksheet
    Dim sheetNames() As String
    Dim selectedSheetIndex As Long

    // Popola un vettore con i nomi dei fogli //
    ReDim sheetNames(1 To ThisWorkbook.Sheets.Count)
    For Each ws In ThisWorkbook.Sheets
        sheetNames(ws.Index) = ws.Name
    Next ws

    // Chiedi all'utente di scrivere il nome del foglio interessato //
    selectedSheetIndex = Application.WorksheetFunction.Match(Application.InputBox("Please, write
the full name of the sheet containing the mandatory fields (be careful to write the exact sheet name
otherwise it won't be identified): " & vbCrLf & Join(sheetNames, vbCrLf), "Select Sheet"),
sheetNames, 0)

    SelectSheetWithMandatoryFields = selectedSheetIndex
End Function
```

SOTTOFUNZIONE PER SALVARE IL FILE IN CSV

```
Public Sub SaveSheetAsCSV(ws As Worksheet, FilePath As String, LastRow As Long,
LastColumn As Long)
    Dim rng As Range

    // Definisci il range per il salvataggio del csv //
    Set rng = ws.Range(ws.Cells(1, 1), ws.Cells(LastRow, LastColumn))

    ' Export the range to CSV
    rng.Copy
    With CreateObject("Excel.Application")
        .Visible = False
        With .Workbooks.Add(1)
            .Sheets(1).Paste
            .SaveAs FilePath, 6 ' xICSV
            .Close False
        End With
        .Quit
    End With
    Application.CutCopyMode = False
End Sub
```

Figura 20: Codice VBA per la creazione della macro

11.31. ALLEGATO 31: PROCESSO DI HARVESTING DEI METADATI

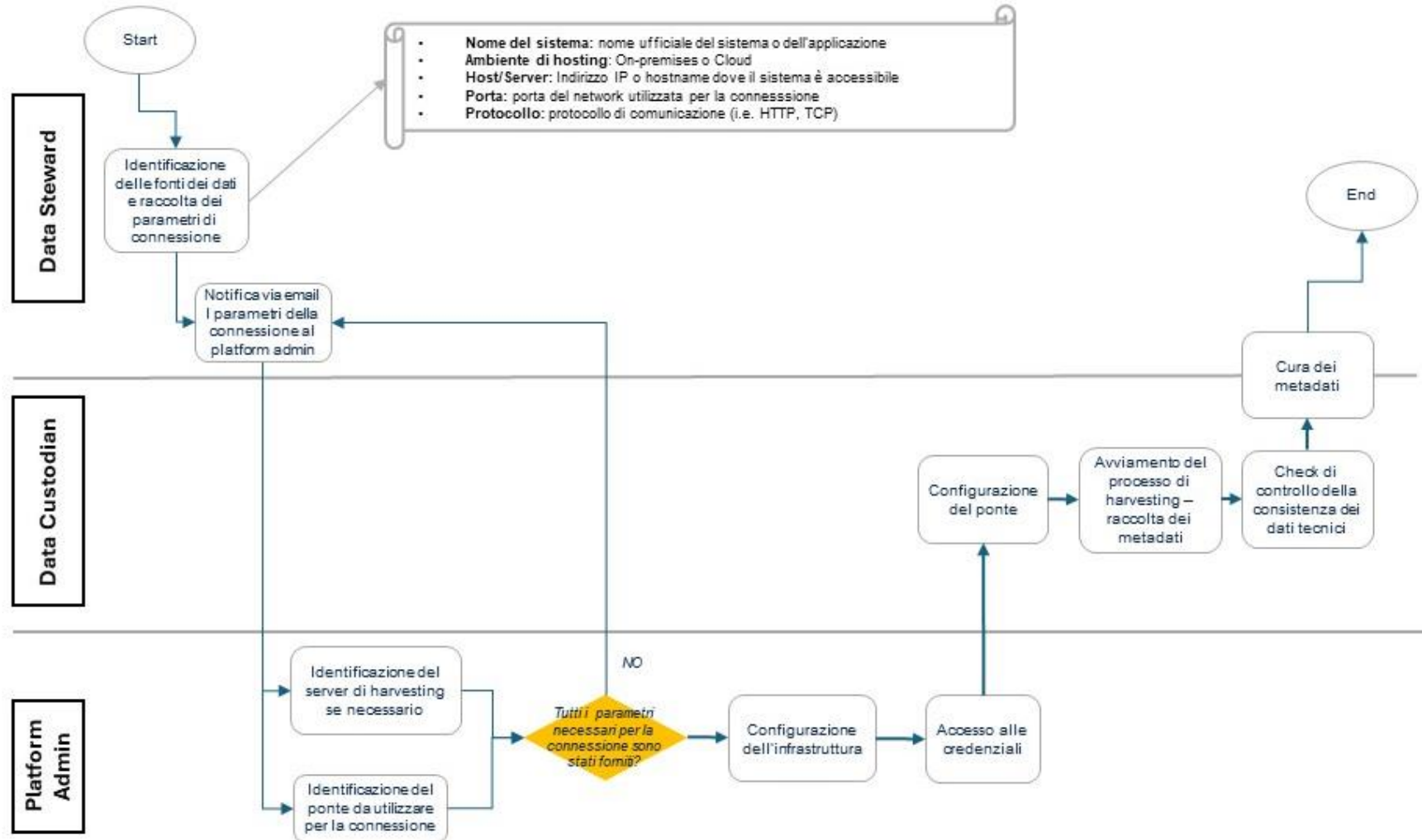


Figura 21: Processo di harvesting dei metadati

11.32. ALLEGATO 32: CAPACITÀ DEL TOOL DI DATA CATALOG

Capacità	Gruppo di Attività
Gestione del Repository	Creare, aggiornare e eliminare oggetti del repository
Gestione della Sicurezza	Visualizzare e assegnare ruoli sugli oggetti del repository
Gestione dei Metadati	Importare ed esportare metadati utilizzando bridge o file .csv
Gestione dei Dati	Importare campioni di dati e informazioni di profilazione
Gestione del Workflow	Abilitare e configurare il workflow sugli oggetti del contenuto
Gestione delle Certificazioni	Modificare ed eliminare certificazioni
Gestione degli Avvisi	Modificare ed eliminare avvisi sugli oggetti
Gestione dei Commenti	Aggiornare ed eliminare commenti sugli oggetti
Gestione degli Osservatori	Aggiungere e rimuovere assegnazioni di osservatori
Modifica dei Metadati	Modificare i metadati
Modifica della Documentazione Aziendale	Impostare il nome aziendale e la descrizione di un oggetto
Modifica dei Diagrammi	Creare e modificare diagrammi nei modelli di database
Modifica della Documentazione dei Termini	Documentare gli oggetti utilizzando i termini del glossario
Modifica della Classificazione dei Dati	Approvare, suggerire, rifiutare una classe di dati o un'etichetta di sensibilità su un oggetto
Modifica delle Etichette	Aggiornare le etichette su un oggetto
Modifica degli Allegati	Aggiungere e rimuovere allegati su un oggetto
Modifica delle Certificazioni	Creare, aggiornare ed eliminare le proprie certificazioni su oggetti
Modifica degli Avvisi	Creare, aggiornare ed eliminare i propri avvisi su oggetti
Modifica dei Commenti	Creare, aggiornare ed eliminare i propri commenti su oggetti
Modifica degli Osservatori	Iniziare o interrompere l'osservazione di un oggetto
Modifica del Workflow	Assegnare e regolare i processi di workflow
Revisione del Workflow	Revisionare i contenuti coinvolti in un processo di workflow
Approvazione del Workflow	Approvare i contenuti coinvolti in un processo di workflow
Pubblicazione del Workflow	Pubblicare i contenuti coinvolti in un processo di workflow
Visualizzazione del Repository	Visualizzare l'albero del repository e interagire con le versioni
Visualizzazione dei Metadati	Visualizzare i metadati memorizzati in un contenuto
Visualizzazione dei Dati	Visualizzare i dati importati inclusi campioni di dati e tutte le informazioni di profilazione dei dati

Tabella 10: Capacità del tool di data catalog

11.33. ALLEGATO 33: CAPACITÀ DEL DATA CATALOG E MATRICE DELLE RESPONSABILITÀ

Ruolo	Management							
	Repository	Metadata	Data	Sicurezza	Workflow	Certificazione	Avvisi	Commenti
Data Governance Expert					✓			
Data Owner (domain-based)		✓	✓			✓	✓	✓
Data Steward (domain-based)		✓	✓		✓	✓	✓	✓
Data Custodian (domain-based)	✓	✓	✓				✓	✓
Data Modeler (domain-based)								
Data Analyst								
Data Scientist								
Data Engineer								
Solution Architect	✓	✓					✓	✓
Data Technical Architect	✓							
Platform Admin				✓				
Data Consumer								

Tabella 11: Capacità di management e matrice delle responsabilità

Ruolo	Editing				
	Metadata	Document	Diagram	Documentazione	Classificazione dei dati
Data Governance Expert					
Data Owner (domain-based)	✓	✓	✓	✓	✓
Data Steward (domain-based)	✓	✓		✓	✓
Data Custodian (domain-based)	✓			✓	
Data Modeler (domain-based)	✓	✓	✓	✓	
Data Analyst					
Data Scientist					
Data Engineer			✓		
Solution Architect					
Data Technical Architect					
Platform Admin					
Data Consumer					

Ruolo	Editing					
	Etichette	Allegati	Certificazione	Avvisi	Commenti	Visualizzazione
Data Governance Expert				✓	✓	
Data Owner (domain-based)	✓	✓	✓	✓	✓	✓
Data Steward (domain-based)	✓	✓	✓	✓	✓	✓
Data Custodian (domain-based)	✓	✓		✓	✓	✓
Data Modeler (domain-based)	✓	✓		✓	✓	✓
Data Analyst					✓	✓
Data Scientist					✓	✓
Data Engineer					✓	✓
Solution Architect				✓	✓	✓
Data Technical Architect					✓	
Platform Admin					✓	✓
Data Consumer						✓

Tabella 12: Capacità di modifica e matrice delle responsabilità

Ruolo	Visualizzazione		
	Repository	Metadata	Data
Data Governance Expert	✓	✓	
Data Owner (domain-based)	✓	✓	
Data Steward (domain-based)	✓	✓	
Data Custodian (domain-based)	✓	✓	
Data Modeler (domain-based)	✓	✓	
Data Analyst	✓	✓	
Data Scientist	✓	✓	
Data Engineer	✓	✓	
Solution Architect	✓	✓	
Data Technical Architect	✓	✓	
Platform Admin	✓	✓	
Data Consumer	✓	✓	✓

Tabella 13: Capacità di visualizzazione e matrice delle responsabilità

Ruolo	Workflow			
	Modifica	Revisione	Approvazione	Pubblicazione
Data Governance Expert	✓	✓	✓	✓
Data Owner (domain-based)				
Data Steward (domain-based)	✓	✓	✓	✓
Data Custodian (domain-based)		✓		
Data Modeler (domain-based)		✓		
Data Analyst		✓		
Data Scientist		✓		
Data Engineer		✓		
Solution Architect		✓		
Data Technical Architect				
Platform Admin				
Data Consumer				

Tabella 14: Capacità di visualizzazione e matrice delle responsabilità

11.34. ALLEGATO 34: LE METRICHE DEI PROCESSI

Processo	ID	Nome	Regola di Calcolo	Unità di Misura	Regola Computazionale
Classificazione dei Dati	M1	Percentuale di Dati Classificati	La proporzione di risorse di dati che sono state classificate in base alla loro sensibilità e importanza	%	$\frac{\# \text{Totale di dati classificati}}{\# \text{totale di dati}}$
Gestione della Sicurezza e Privacy	M2	Conformità alla Sicurezza e Privacy dei Dati	Il livello di conformità alle normative sulla privacy dei dati, come il GDPR o HIPAA	%	Per ogni normativa sulla privacy, $\frac{\# \text{Totale di dati certificati associati alla normativa}}{\# \text{Totale di dati con una normativa associata}}$
Gestione della Conservazione e Archiviazione dei Dati	M3	Conformità alla Politica di Conservazione	L'entità con cui le risorse di dati sono conservate e archiviate in conformità con le politiche stabilite	%	$\frac{\# \text{Totale di dati conformi al periodo di conservazione dei dati}}{\# \text{Totale di dati con un periodo di conservazione stabilito}}$
Gestione dei Metadati	M5	Completezza dei Metadati	La percentuale di risorse di dati con metadati completi e accurati (es. tutti gli attributi obbligatori, associazioni, genealogia dei dati)	%	$\frac{\# \text{Totale di business terms con metadati completi}}{\# \text{Totale di business terms}}$
Qualità dei Dati	M7	Accuratezza dei Dati	La percentuale di record di dati che sono accurati (es. record senza errori che possono essere usati come fonte affidabile di informazione)	%	$\frac{\# \text{Totale di dati accurati}}{\# \text{Totale di dati}}$
Qualità dei Dati	M8	Completezza dei Dati	La percentuale di record di dati che sono completi (es. se tutti i dati richiesti sono presenti, può essere misurato con diverse granularità)	%	$\frac{\# \text{Totale di dati certificati associati alla normativa}}{\# \text{Totale di dati con una normativa associata}}$

Qualità dei Dati	M9	Coerenza dei Dati	L'entità con cui i dati sono coerenti (es. qualità dei dati di essere uniformi, accurati e coerenti tra vari sistemi, database e applicazioni) tra sistemi e fonti	%	$\frac{\# \text{ Totale di dati dove non si osservano incoerenze}}{\# \text{ Totale di dati}}$
Monitoraggio dei Dati	M10	Copertura del Monitoraggio dei Dati	La proporzione di risorse di dati monitorate attivamente per anomalie o problemi	%	Numero di risorse di dati monitorate / Numero totale di dati $\frac{\# \text{ Totale di dati monitorati}}{\# \text{ Totale di dati}}$

Tabella 15: Metriche dei processi di data governance

11.35. ALLEGATO 35: RELAZIONI TRA GLI ELEMENTI DEL FRAMEWORK

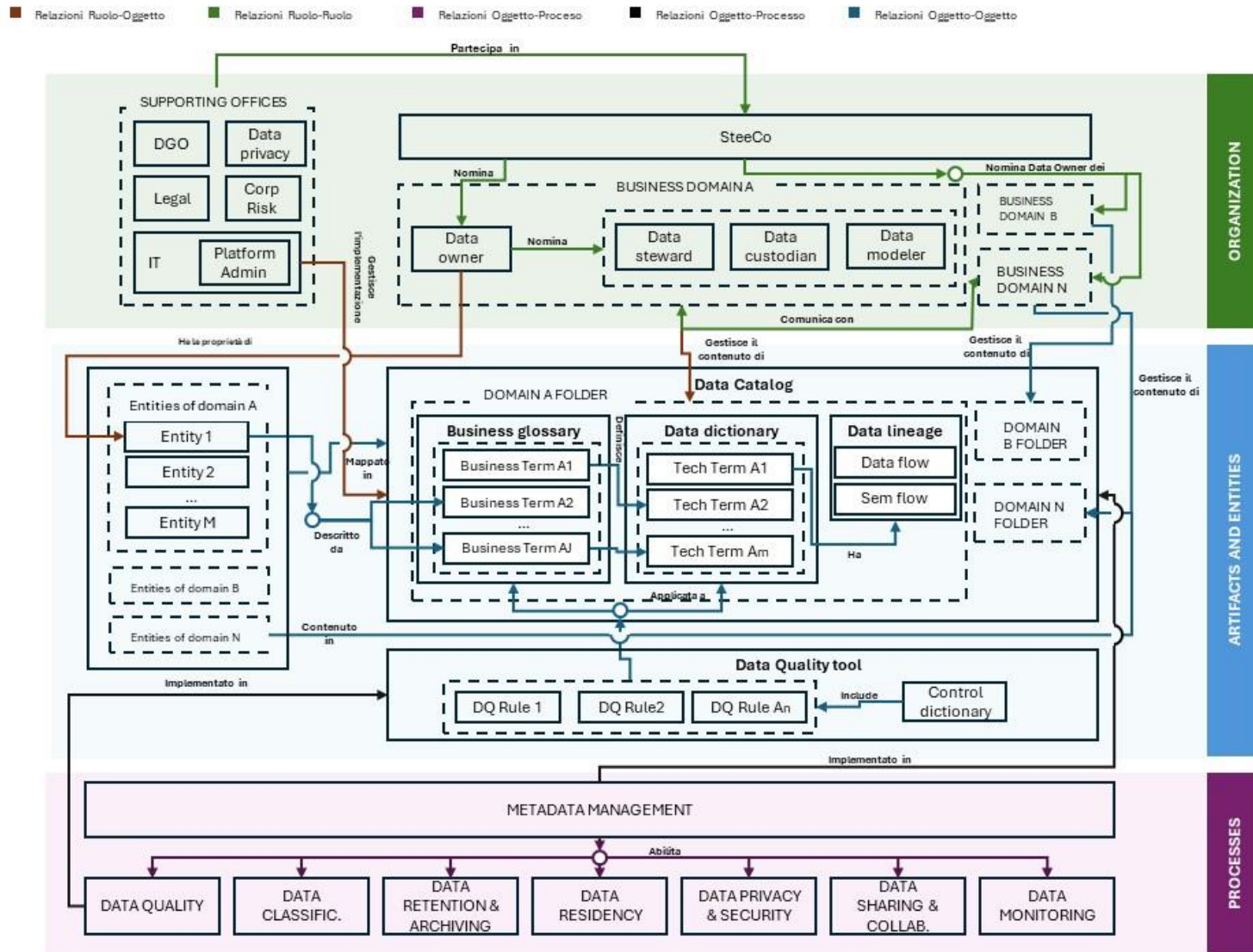


Figura 22: rappresentazione grafica del metamodello

11.36. ALLEGATO 36: PRIMA DEL FRAMEWORK

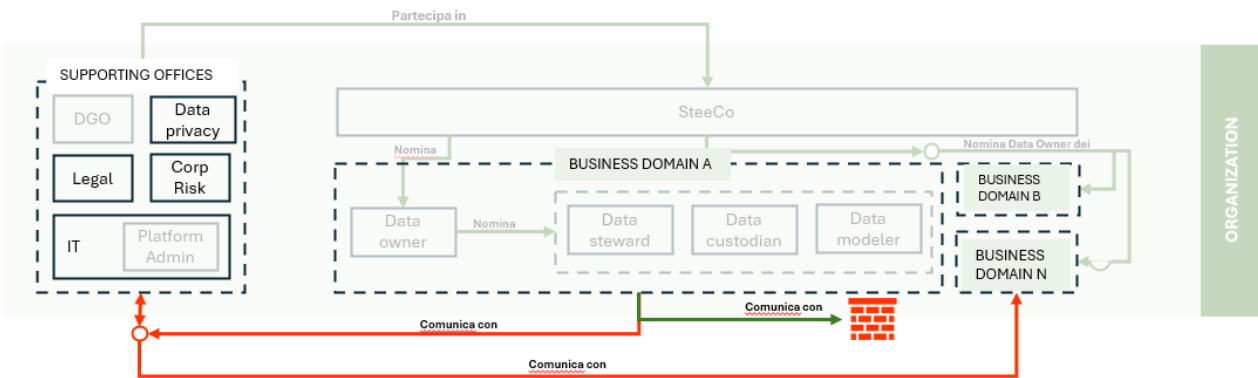


Figura 23: Richiesta di dati extra-dominio prima del framework di data governance

11.37. ALLEGATO 37: MANCANZA DI OWNERSHIP, CONDIVISIONE E DEFINI

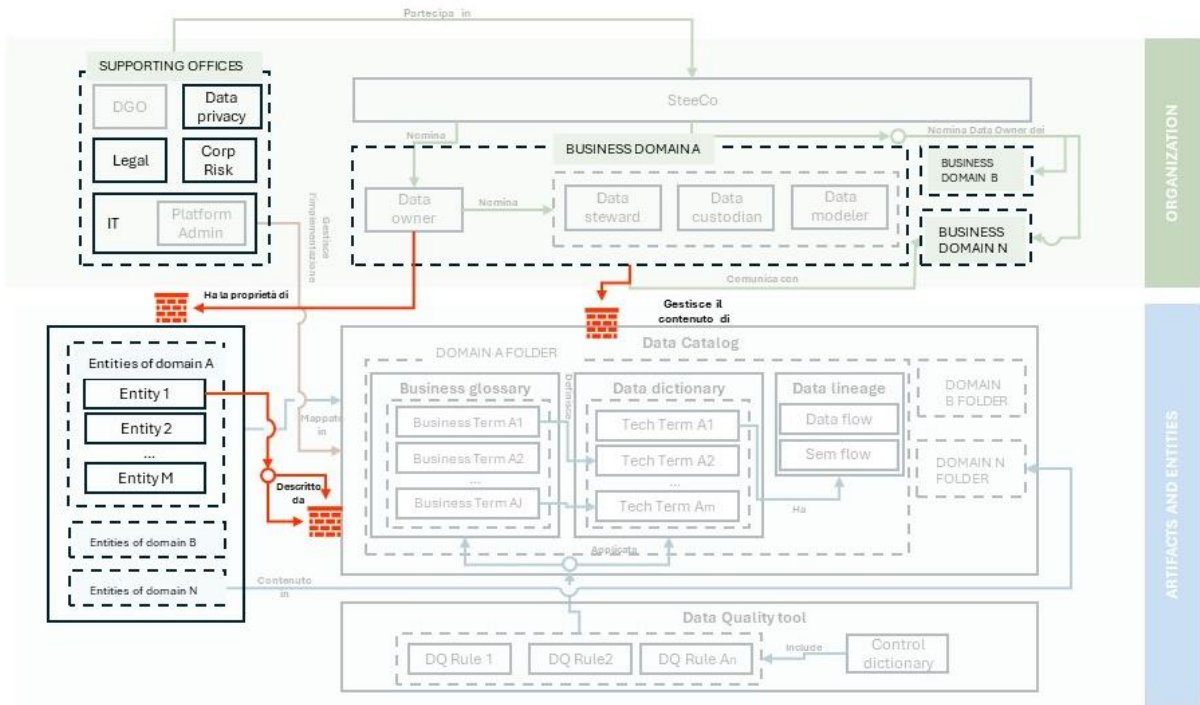


Figura 24: Rappresentazione della mancanza di ownership dei dati, catalogazione delle entità e dei dati dei domini

11.38. ALLEGATO 38: MATRICE COMPLESSITÀ VALORE

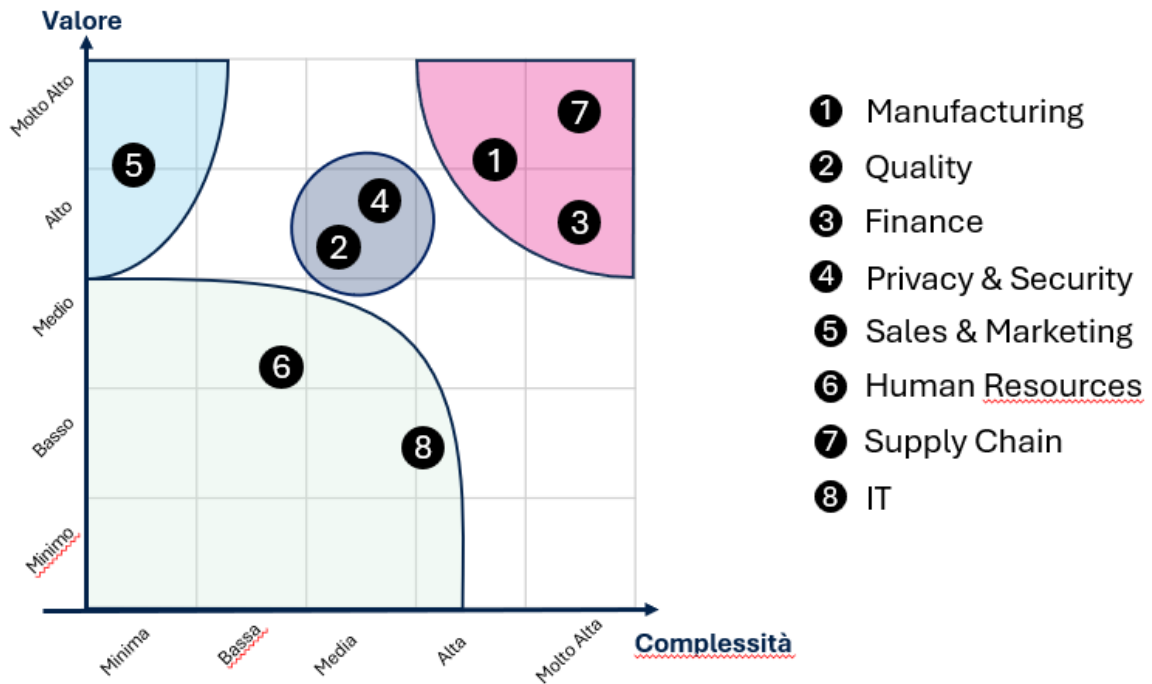


Figura 25: Matrice complessità valore per la scelta del test di caso d'uso

11.39. ALLEGATO 39: STRUTTURA DEL BG DEL DOMINIO S&M

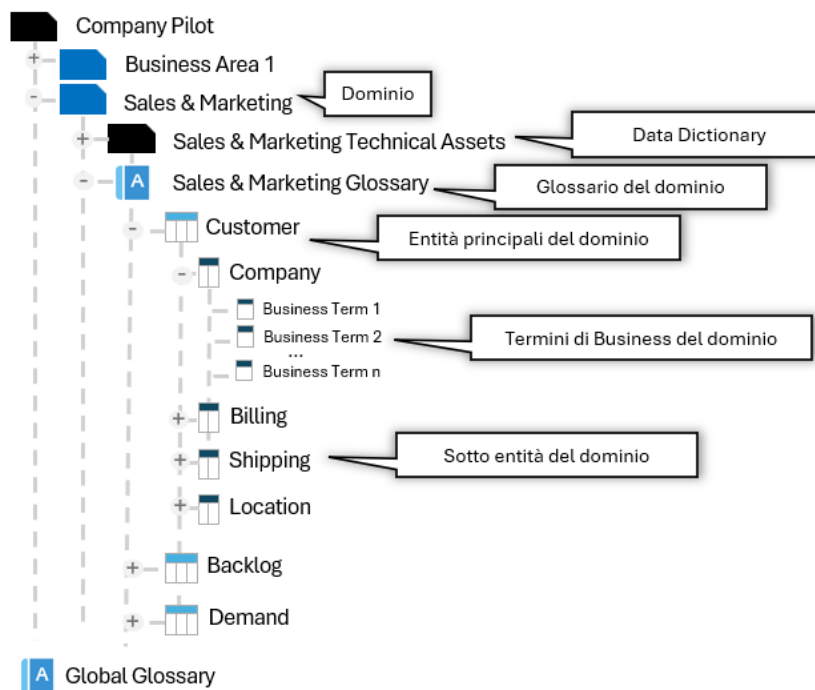


Figura 26: Struttura del business glossary per il dominio di Sales & Marketing

11.40. ALLEGATO 40: TERMINI DI BUSINESS PER ENTITÀ

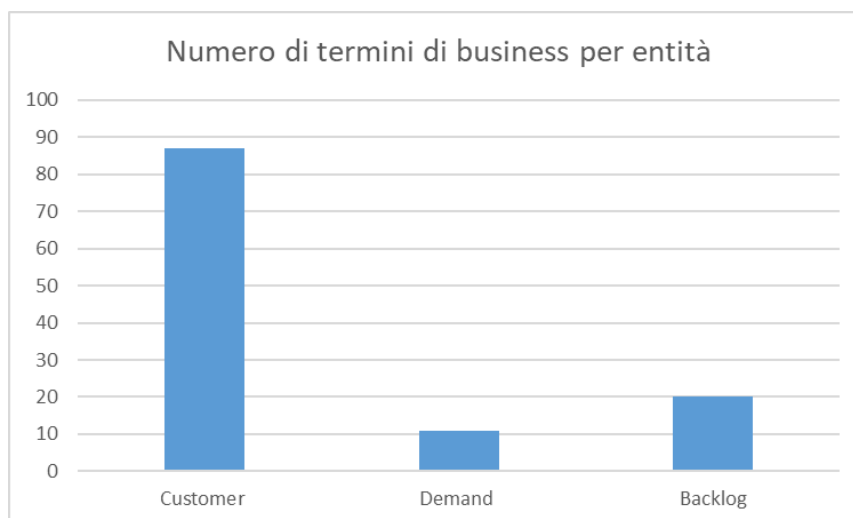


Figura 27: Numero di termini di business per entità del glossario S&M

11.41. ALLEGATO 41: DEFINIZIONE DEGLI ERRORI

ID	Descrizione del tipo di errore	Impatto dell'errore	Azione di rimedio
1	Molteplici termini di business con lo stesso nome	-Bassa qualità e accuratezza dei metadati -Alta probabilità di errori per analisi e report, influenzando decisioni chiave -Ricerca di informazioni dispendiosa in termini di tempo	1) Con un'analisi accurata del significato, se i termini sono assimilabili ad un singolo, eliminare i riferimenti ai termini duplicati e creare relazioni logiche tra i termini 2) Se il significato non è lo stesso, modificare i nomi dei termini aggiungendo più parole chiave per enfatizzare la differenza e creare tali termini all'interno del glossario di business del dominio
2	Nomi di termini di business non chiari o troppo generici	-Bassa trasparenza e sinergie -Ricerca di informazioni dispendiosa in termini di tempo	Modificare i nomi dei termini aggiungendo più parole chiave e creare tali termini all'interno del glossario di business del dominio (glossari specifici del dominio)
3	Termini di business con nomi simili a cui potrebbe essere attribuito lo stesso significato	-Aumentata probabilità di errori per analisi / report, influenzando decisioni chiave -Ricerca di informazioni dispendiosa in termini di tempo	1) Con un'analisi accurata del significato, se è lo stesso, eliminare i riferimenti ai termini duplicati e creare relazioni logiche tra i termini 2) Se il significato non è lo stesso, modificare i nomi dei termini aggiungendo più parole chiave e creare tali termini all'interno del glossario di business del dominio
4	Termini di business con nomi diversi ma a cui potrebbe essere attribuito lo stesso significato di business	-Aumentata probabilità di errori per analisi / report, influenzando decisioni chiave -Ricerca di informazioni dispendiosa in termini di tempo	1) Con un'analisi accurata del significato, se è lo stesso, eliminare i riferimenti ai termini duplicati e creare relazioni logiche tra i termini 2) Se il significato non è lo stesso, modificare i nomi dei termini aggiungendo più parole chiave e creare tali termini all'interno del glossario di business del dominio

5	<p>Gli attributi dei termini (es. descrizioni) devono essere revisionati perché incompleti, poco chiari o fuorvianti</p>	<p>-Bassa qualità e accuratezza dei metadati -Disallineamento tra l'organizzazione / i team</p>	<p>Modificare le descrizioni dei termini aggiungendo più parole chiave e creare tali termini all'interno del glossario di business del dominio (glossari specifici del dominio)</p>
----------	--	---	---

Tabella 16: Definizione e spiegazione degli errori identificati tra i termini di business

11.42. ALLEGATO 42: ENTITÀ CON ERRORI RILEVATI

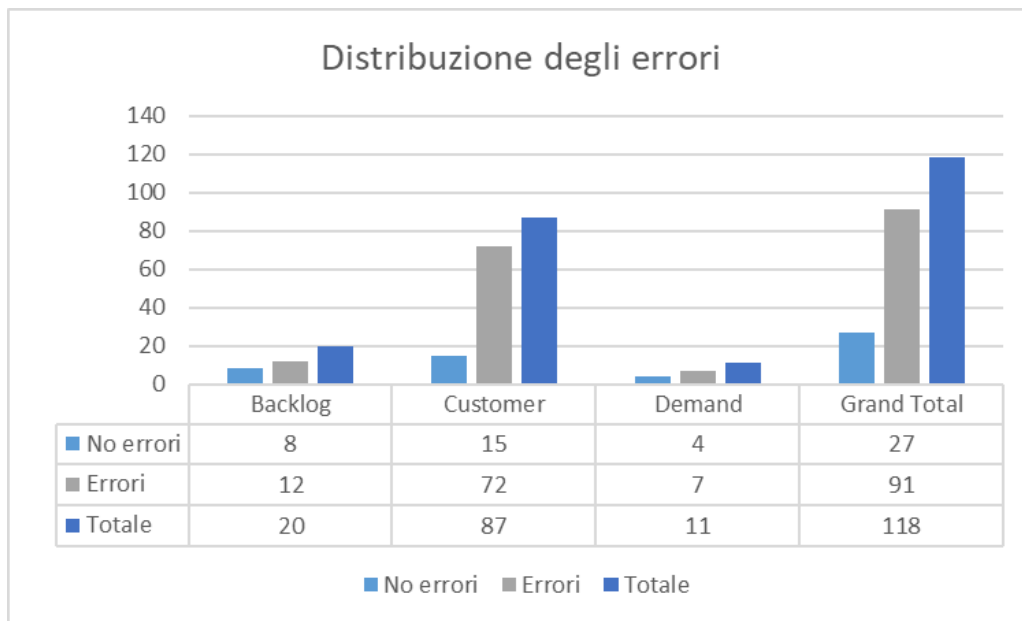


Figura 28: Tipi e numero di entità con errori riscontrati

11.43. ALLEGATO 43: NUMERO DI ERRORI RISCONTRATI PER TIPOLOGIA

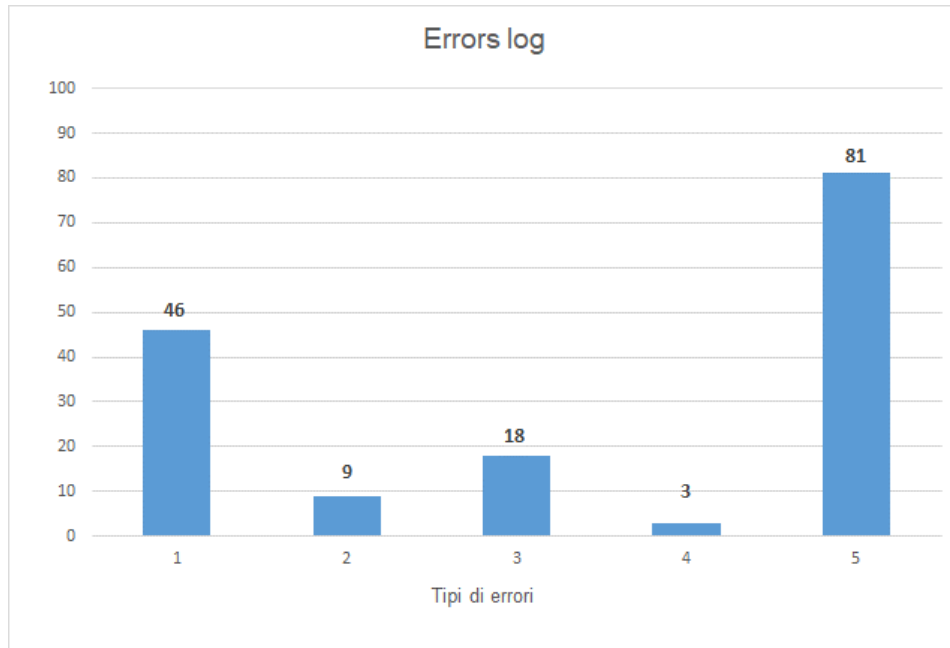


Figura 29: Numero e tipi di errori riscontrati

11.44. ALLEGATO 44: METADATA COMPLETENESS, PRIMO UPLOAD

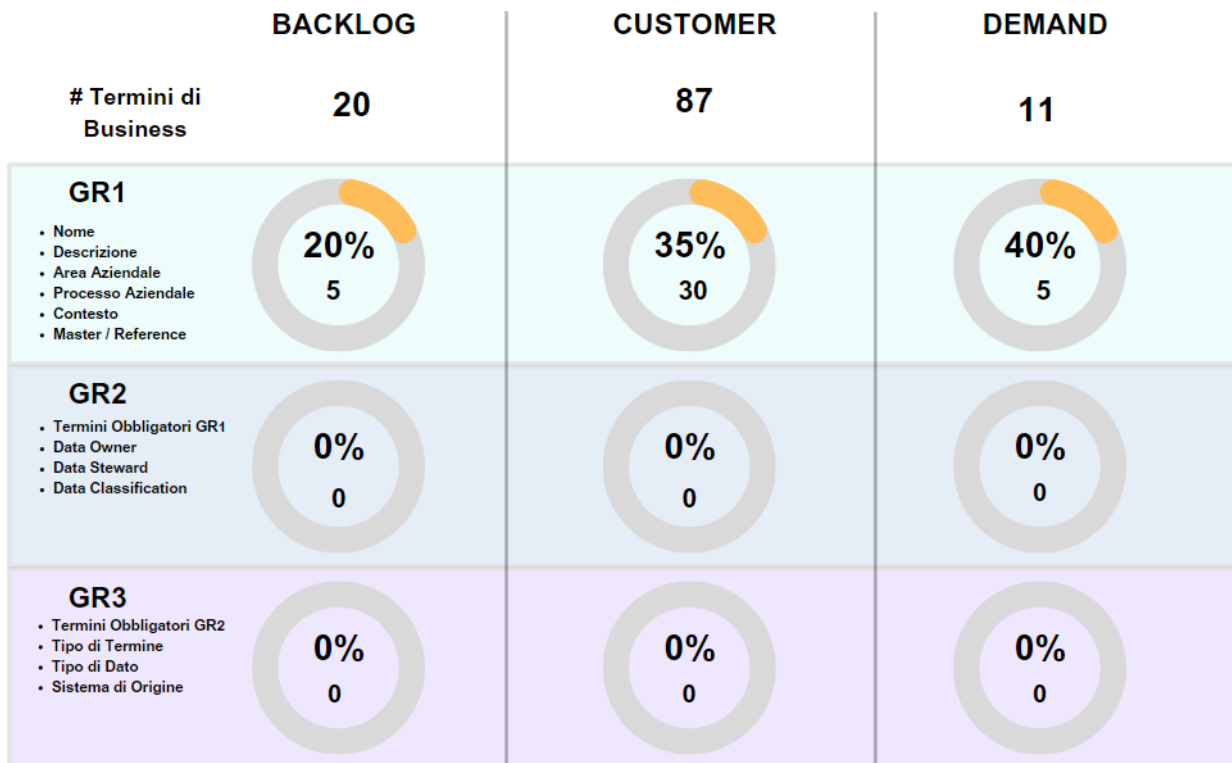


Figura 30: Entità e attributi obbligatori

11.45. ALLEGATO 45: STRUTTURA DEL DATA DICTIONARY

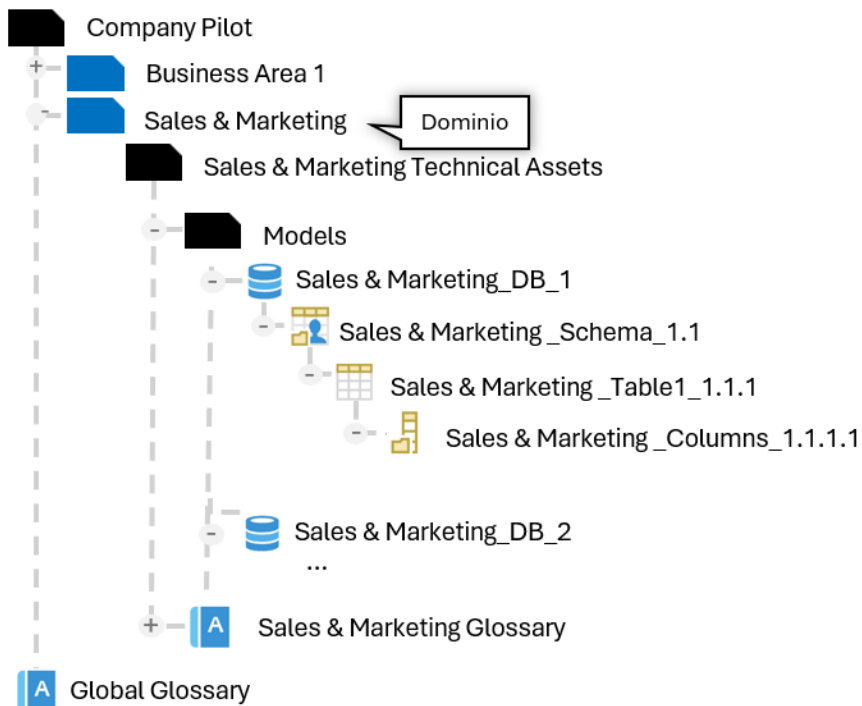


Figura 31: Struttura del data dictionary

11.46. ALLEGATO 46, METRICHE DEL TEST

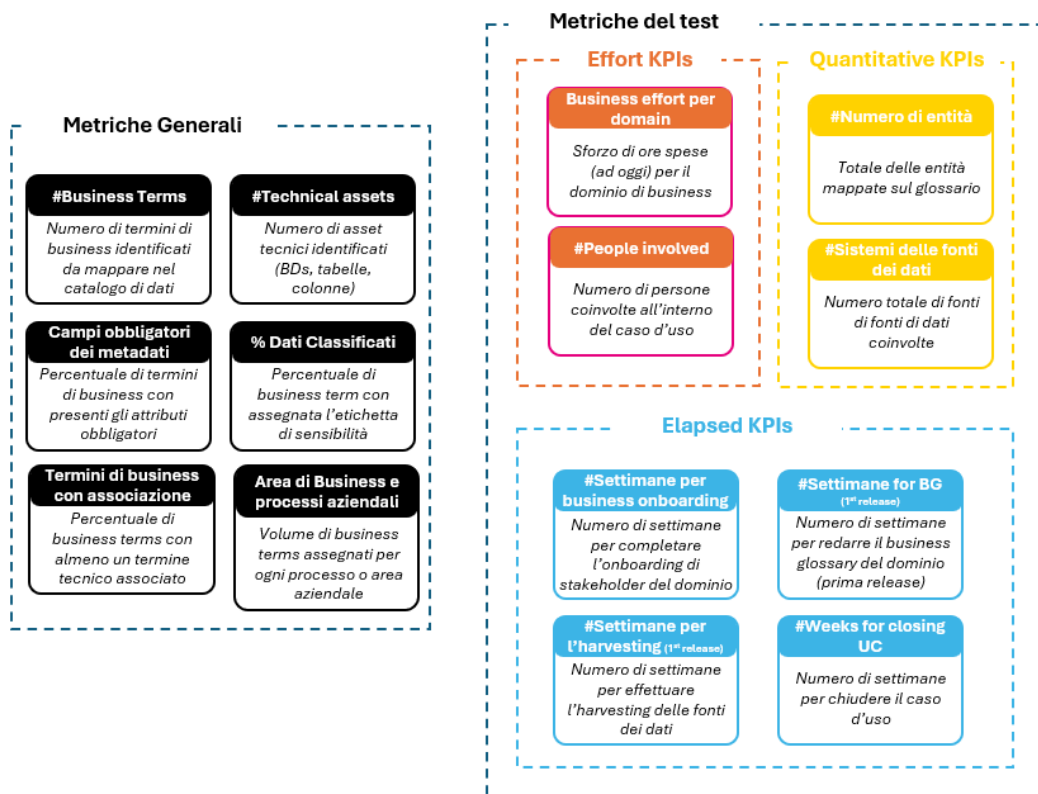


Figura 32: Metriche del test sul dominio di Sales& Marketing

11.47. ALLEGATO 47, TERMINI DI BUSINESS, PRIMA E DOPO

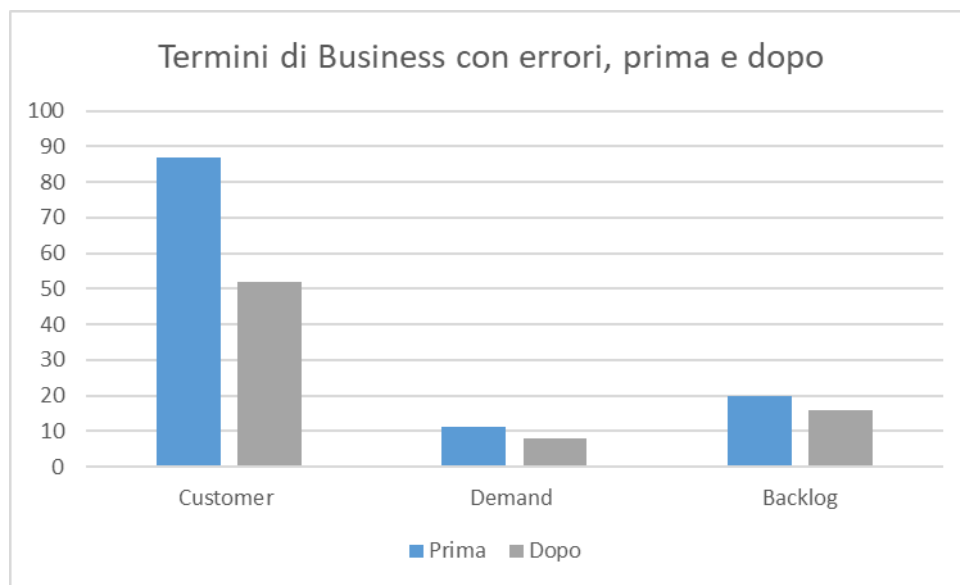


Figura 33: Miglioramento del numero di termini di business

11.48. ALLEGATO 48, NUMERO DI ERRORI RILEVATI

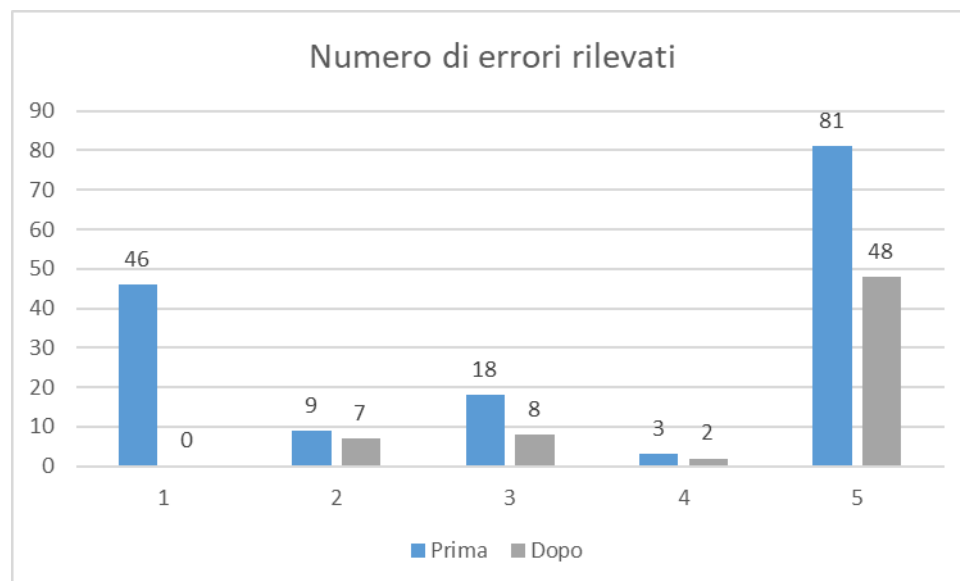


Figura 34: Numero di errori rilevati, prima e dopo

11.49. ALLEGATO 49, DASHBOARD CAMPI OBBLIGATORI DEI METADATI

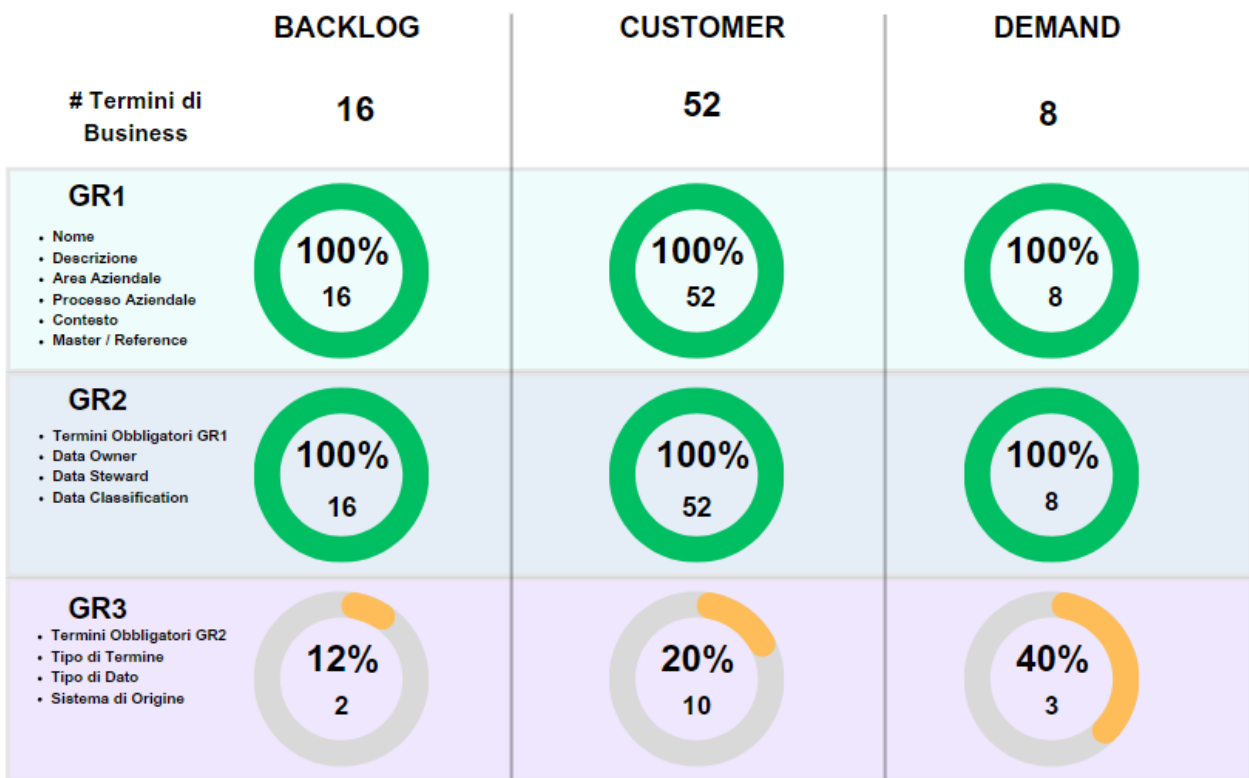


Figura 35: Termini di business con i campi obbligatori compilati

11.50. ALLEGATO 50, METRICHE DEL TEST

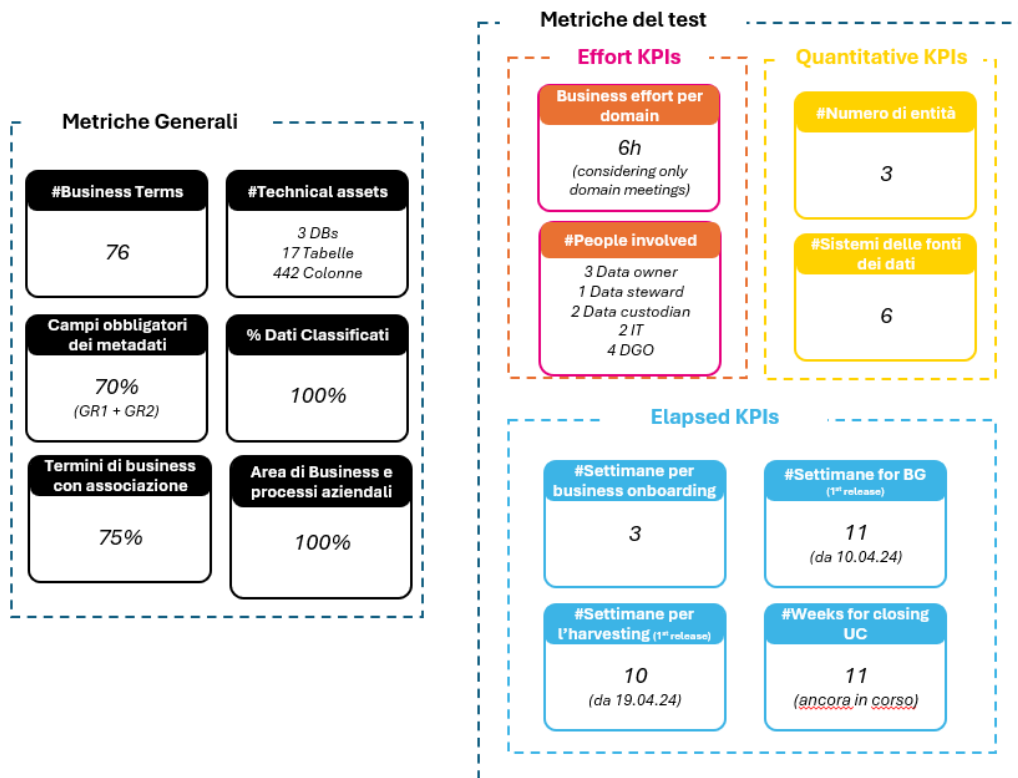


Figura 36: Risultati del test di Sales & Marketing

11.51. ALLEGATO 51: MODULO DI VALUTAZIONE DELLA MATURITÀ

VALUTAZIONE DELLA MATURITA'					
Ambito:		Metadata Management			
Dimensioni:		Engagement & Adoption	Tool e Tecnologie	Policy & Standard	Value Realization
MATURITY LEVEL	0 - No capability	Nessuna comprensione delle tipologie di metadati e dei potenziali benefici a livello aziendale	Non sono presenti strumenti per la gestione dei metadati	Assenza di Policy & Standard legate a Metadata Management	Nessuna comprensione del valore generato dalla gestione dei dati
	1 - Initial	Limitata comprensione delle tipologie di metadati e dei potenziali benefici a livello aziendale	Sono ricercati e definiti i tool a supporto per la raccolta dei metadati	Sono definite le Policy & Standard, ma in fase di test in un gruppo di esperti	Limitata comprensione del valore generato dalla gestione dei metadati
	2 - Repeatable	Comprensione delle tipologie di metadati e dei potenziali benefici solo in alcuni domini	I tool a supporto per la raccolta dei metadati sono stati implementati in alcuni domini (es. Business Glossary, Data Dictionary)	Sono applicate le Policy & Standard di alto livello legate a attività di Metadata Management solo in alcuni domini	Comprensione del valore generato dalla gestione dei metadati
	3 - Defined	Comprensione e prima adozione dei processi e per la gestione dei metadati nella maggior parte dei domini	I tool disponibili per la gestione dei metadati offrono solo parzialmente le funzionalità richieste dagli utenti	Sono definite e adottate le principali Policy & Standard di dettaglio legate a attività di Metadata Management	Comprensione del valore generato dalla gestione dei metadati e definizione di metriche quantitative per monitorare i relativi benefici e i costi
	4 - Governed	Adozione a livello aziendale dei processi e degli strumenti per la gestione dei metadati	I tool disponibili per la gestione dei metadati offrono tutte le principali funzionalità richieste dagli utenti	Standardizzazione di Policy & Standard per l'esecuzione di attività di Metadata Management per favorire l'adozione in diverse aree di business	Adozione e monitoraggio di metriche quantitative per monitorare i benefici e i costi legati al Metadata Management
	5 - Optimization	Suggerimenti su miglie e standardizzazione dei processi e degli strumenti per la gestione dei metadati	I tool disponibili per la gestione dei metadati offrono tutte le funzionalità richieste dagli utenti e sono attivi processi di miglioramento continuo	Fine-tuning continuo di Policy & Standard per ottimizzare l'esecuzione di attività di Metadata Management	Ottimizzazione e ricerca continua dei benefici legati al Metadata Management tramite iniziative abilitanti nuove fonti di creazione del valore

Figura 37: Valutazione della maturità aziendale rispetto al processo di metadata management

11.52. ALLEGATO 52: VALUTAZIONE GLOBALE DELLA MATURITÀ DELLA DATA GOVERNANCE

	RISULTATI MATURITÀ DELLA DATA GOVERNANCE				MEDIA
	Engagement & Adoption	Tool e Tecnologie	Policy & Standard	Value Realization	
Classificazione dei dati	2	3	2	4	2,75
Qualità dei dati	1	1	2	1	1,25
Sicurezza e Privacy dei dati	1	3	4	2	2,5
Residenza dei dati	0	0	0	0	0
Conservazione e archiviazione	1	1	2	1	1,25
Condivisione dei dati e collaborazione	0	1	4	2	1,75
Monitoraggio dei dati	1	2	1	1	1,25
Metadata Management	3	3	2	2	2,5
TOTALE					1,66

Figura 38: Esempio di valutazione globale della data governance

12. BIBLIOGRAFIA

- European Commission. (2023). Industry 5.0. European Commission. https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/industry-50_en
- DAMA International. (2017). DAMA-DMBOK: Data Management Body of Knowledge (2nd ed.). Technics Publications
- European Commission. (n.d.). Data protection in the EU. 2016, from https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_en
- Ahlstrom, R. (2019, January 17). The role of data in the age of digital transformation. Forbes. <https://www.forbes.com/sites/forbestechcouncil/2019/01/17/the-role-of-data-in-the-age-of-digital-transformation/#:~:text=By%20making%20data%20quality%20management%20and%20governance%20the,using%20artificial%20intelligence%20to%20generate%20important%20new%20insights>
- Ali Intezari, Nazim Taskin, David J. Pauleen (2016), The DIKW Hierarchy and Management Decision-Making. Conference Paper
- Brackett, M. (2024b, January 20). The Data-Information-Knowledge Cycle - DATAVERSITY. <https://www.dataversity.net/the-data-information-knowledge-cycle/>
- Neely, A. (2024, May 7). Why you should be treating your data as an asset. Anmut. <https://www.anmut.co.uk/why-you-should-be-treating-your-data-as-an-asset/>
- The world's data explained: how much we're producing and where it's all stored. (2022, May 20). World Economic Forum. <https://www.weforum.org/agenda/2021/05/world-data-produced-stored-global-gb-tb-zb/>
- Data growth worldwide 2010-2025 | Statista. (2023, November 16). Statista. <https://www.statista.com/statistics/871513/worldwide-data-created/>
- What is Data Lifecycle Management (DLM)? | IBM. (n.d.). <https://www.ibm.com/topics/data-lifecycle-management>
- Structured vs. unstructured data: What's the difference? | IBM. (n.d.). <https://www.ibm.com/think/topics/structured-vs-unstructured-data>
- Edgar F. Codd | IBM. (n.d.). <https://www.ibm.com/history/edgar-codd>
- ECMA-404 - ECMA International. (2021, February 4). Ecma International. <https://ecma-international.org/publications-and-standards/standards/ecma-404/>
- MongoDB. (n.d.). Explaining BSON with examples. <https://www.mongodb.com/resources/languages/bson>
- XML Introduction. (n.d.). https://www.w3schools.com/XML/xml_what_is.asp
- Worldwide Global DataSphere and Global StorageSphere Structured and Unstructured Data Forecast, 2023–2027. (n.d.). IDC: The Premier Global Market Intelligence Company. <https://www.idc.com/getdoc.jsp?containerId=US50397723&pageType=PRINTFRIENDLY>
- Che cos'è OLTP? (n.d.). <https://www.oracle.com/it/database/what-is-oltp/>
- Sas. (n.d.). Unlocking the Potential of Big Data: An Insightful Guide. https://www.sas.com/en_us/insights/big-data/what-is-big-data.html#:~:text=Big%20data%20refers%20to%20data,to%20process%20using%20traditional%20methods
- Dehghani, Z. (2022). Data Mesh: Delivering Data-Driven Value at Scale. O'Reilly Media
- What is a Data Lake? Data Lake vs. Warehouse | Microsoft Azure. (n.d.). <https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-a-data-lake/>
- World Economic Forum. (2023, November 7). World Economic Forum. <https://www.weforum.org/publications/articulating-value-from-data/>
- Nandwana, A. (2022, May 30). Data is Pharma's next blockbuster. Forbes. <https://www.forbes.com/sites/forbesbusinesscouncil/2022/05/27/data-is-pharmas-next-blockbuster/>
- Hoffman, W., Bick, R., Boral, A., Henke, N., Olukoya, D., Rifai, K., Roth, M., & Youldon, T. (2019, May 30). Collaborating for the common good: Navigating public-private data partnerships. McKinsey & Company. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/collaborating-for-the-common-good>
- Publications Office of the European Union. (2018). Study on data sharing between companies in Europe: case studies. Publications Office of the EU. <https://op.europa.eu/en/publication-detail/-/publication/004fcf02-49c7-11e8-be1d-01aa75ed71a1/language-en>
- Goasduff, L., "Data Sharing Is a Business Necessity to Accelerate Digital Business", Gartner, 20 May 2021,

<https://www.gartner.com/smarterwithgartner/data-sharing-is-a-business-necessity-to-accelerate-digital-business>

- Tessian Limited. (2023, October 25). Psychology of Human Error 2022 | Research Report | Tessian. Tessian. <https://www.tessian.com/resources/psychology-of-human-error-2022/>
- Cost of a data breach 2023 | IBM. (n.d.). https://www.ibm.com/reports/data-breach?utm_content=SRCWW&p1=Search&p4=43700076824884778&p5=b&p9=58700008442980784&gad_source=1&gclid=Cj0KCQjwmMayBhDuARIsAM9HM8dglMOrJiDsVCZWY18f80SCNvvDwMuYgUHSZx8p6Hv9_mIIcVp8PD8aAo5dEALw_wcB&gclsrc=aw.ds
- What is cryptography? (n.d.). ISO. <https://www.iso.org/information-security/what-is-cryptography>
- ISO/IEC 27001:2022. (n.d.). ISO. <https://www.iso.org/standard/27001>
- National Institute of Standards and Technology. (2024, February 26). The NIST Cybersecurity Framework. <https://doi.org/10.6028/NIST.CSWP.29>
- 7 basic steps to identify a data breach | Kron. (n.d.). [krontech.com. https://krontech.com/7-basic-steps-to-identify-a-data-breach](https://krontech.com/7-basic-steps-to-identify-a-data-breach)
- GDPR.eu. (n.d.). GDPR Archives - GDPR.eu. <https://gdpr.eu/tag/gdpr/>
- Wolford, B. (2023, September 14). What is GDPR, the EU's new data protection law? GDPR.eu. <https://gdpr.eu/what-is-gdpr/>
- The path to modern data governance. (n.d.). Eckerson Group. <https://www.eckerson.com/articles/modern-data-governance-problems>
- Let the chips fall where they may - OECD. (2019). <https://www.oecd.org/trade/let-the-chips-fall-where-they-may/>
- Ciani, A. and Nardo, M. (2022). The position of the EU in the semiconductor value chain: evidence on trade, foreign acquisitions, and ownership. JRC Working Papers in Economics and Finance, 2022/3.
- OECD, 2019. Measuring distortions in international markets: the semiconductor value chain.
- Powell, K. J. L. N. J. (2020). Mapping global supply chains – the case of semiconductors - Rabobank. Rabobank. <https://www.rabobank.com/knowledge/d011371771-mapping-global-supply-chains-the-case-of-semiconductors>
- Oppermann, A. (2023, February 13). What is a transistor? Built In. <https://builtin.com/hardware/transistor>
- Inside the miracle of modern chip manufacturing. (2024, February 28). Financial Times. <https://ig.ft.com/microchips/>
- Metal-Oxide-Semiconductor Field-Effect Transistors (MOSFETs) | Toshiba Electronic Devices & Storage Corporation | Asia-English. (n.d.). <https://toshiba.semicon-storage.com/ap-en/semiconductor/knowledge/e-learning/discrete/chap3/chap3-5.html>
- Smith, R. (2023, January 5). CES 2023: AMD Instinct MI300 Data Center APU Silicon in hand - 146B transistors, Shipping H2'23. AnandTech. <https://www.anandtech.com/show/18721/ces-2023-amd-instinct-mi300-data-center-apu-silicon-in-hand-146b-transistors-shipping-h223>
- Khan, S. M., Mann, A., & Peterso, D. (2021, January). The semiconductor supply chain: Assessing national competitiveness (CSET Issue Brief). Center for Security and Emerging Technology.
- Muelaner, J. (2023, May 25). How Automation is Helping U.S. Manufacturers Scale Semiconductor Manufacturing. DigiKey. <https://www.digikey.com/en/articles/how-automation-is-helping-scale-semiconductor-manufacturing>
- Microprocessor Market Size, Share & Trend Analysis Report By Architecture Type (ARM MPU, x64, x86, MIPS), By Application, By Region, and Segment Forecasts, 2023 - 2030. (2020, May 26). <https://www.grandviewresearch.com/industry-analysis/microprocessor-market>
- Global Microprocessor Market Share, Size & Growth Report 2031F. (n.d.). Markets & Data. <https://www.marketsanddata.com/industry-reports/microprocessor-market>
- Major semiconductor producing countries rely on each other for different types of chips. (2022, December 5). PIIE. <https://www.piie.com/research/piie-charts/major-semiconductor-producing-countries-rely-each-other-different-types-chips>
- Microprocessor market size to hit around USD 185.39 BN by 2032. (2023, November 3). <https://www.precedenceresearch.com/microprocessor-market>
- Statista. (2023, August 3). Microprocessor market share worldwide 2018-2020, by vendor. <https://www.statista.com/statistics/883715/microprocessor-market-share-worldwide-by-vendor/>
- Largest semiconductor companies by market cap. (2024). <https://companiesmarketcap.com/semiconductors/largest-semiconductor-companies-by-market-cap/>
- Clark, D. (2024, May 22). Nvidia, powered by A.I. Boom, reports soaring revenue and profits. The New York Times. <https://www.nytimes.com/2024/05/22/technology/nvidia-quarterly-earnings-results.html>
- Ragonnaud, G. (2022, November). The EU chips act: Securing Europe's supply of semiconductors (PE 733.596). European Parliamentary Research Service. <https://doi.org/10.xxxx/yyyy>

- European Chips Act. (2019). European Commission. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-chips-act_en
- House, W. (2023, February 3). FACT SHEET: CHIPS and Science Act will lower costs, create jobs, strengthen supply chains, and counter China. The White House. <https://www.whitehouse.gov/briefing-room/statements-releases/2022/08/09/fact-sheet-chips-and-science-act-will-lower-costs-create-jobs-strengthen-supply-chains-and-counter-china/>
- Josephs, B. S. T. a. J. (2022, December 16). US-China chip war: How the technology dispute is playing out. <https://www.bbc.com/news/business-63995570>
- Sommer, J. (2024, May 24). The business ties that bind the U.S. and China are strong but fraying. The New York Times. <https://www.nytimes.com/2024/05/24/business/united-states-china-chips-profits.html>
- Silicon Metal market Insights. (n.d.). <https://www.mordorintelligence.com/industry-reports/silicon-metal-market>
- Ruberti, M. (2023). The chip manufacturing industry: Environmental impacts and eco-efficiency analysis. *Science of the Total Environment*, 858, 159873. <https://doi.org/10.1016/j.scitotenv.2022.159873>
- Mohr, J., Tembey, G., Breidenbach, K., Sah, N., Jeschke, J., & Harder, T. (2023, June 13). For chip makers, the decarbonization challenge lies upstream. BCG Global. <https://www.bcg.com/publications/2023/why-chip-makers-need-to-focus-on-the-upcoming-decarbonization-challenges>