

# Data Governance

Baroum Mrad

**La Data Governance è un set di linee guida per come le persone si comportano e prendono decisioni sui dati.**

*"un sistema di diritti decisionali e di responsabilità per i processi legati alle informazioni, eseguiti secondo modelli concordati che descrivono chi può intraprendere quali azioni con quali informazioni, e quando, in quali circostanze, usando quali metodi". - Data Governance Institute (2014)*

## Principali componenti di una data governance

### *Come punto di partenza*

- *Regolamenti costituzionali dell'azienda*
- *Codice di condotta e codice etico*

Il primo importante passo nella costruzione di un framework efficace è la definizione degli indirizzi strategici in ambito di gestione del dato, che individuino quindi:

- le finalità,
- il perimetro,
- le tempistiche,
- gli investimenti da destinarvi.



## La Data Governance nel concreto

*La Data Governance è quindi la capacità di gestire i dati come un vero e proprio asset aziendale. Per riuscirci è necessario seguire delle “regole”, ovvero un framework di Data Governance, al fine di assicurarsi che le informazioni siano dirette adeguatamente in relazione agli obiettivi dell’impresa.*



## Differenza tra Data Governance e Data Management

La Data Governance consiste nella definizione, implementazione ed esecuzione di una policy che costituisce il **Data Management** da applicare, con tutte le **procedure** e gli **strumenti** necessari per la gestione e fruizione dei dati.

Definire un completo **processo di gestione e valorizzazione dei dati aziendali** e di declinarsi in modo proporzionale alle esigenze della specifica realtà:

- **Data Dictionary:** fornisce una visione dei metadati delle basi dati importate nel modulo di Data Governance, approfondendo ogni singolo campo/tabella sia dei metadati che delle statistiche ad essi legati.
- **Data Quality:** gestisce il core engine dei controlli di qualità registrati, effettuati e monitorati. È possibile effettuare controlli di qualità sia in modalità guidata che custom, verificando poi score e risultati.
- **Monitoring:** gestisce la visualizzazione e l'interpretazione dei risultati registrati durante i controlli schedulati, fornendo la possibilità di verificare l'andamento dello score di qualità per ogni dimensione.
- **Data Lineage:** identifica il ciclo di vita del dato tramite una rappresentazione a grafo, così da poter verificare sia l'origine dei dati e le loro trasformazioni durante i diversi processi.
- **Business Glossary:** fornisce un repository utile per comunicare e governare i termini aziendali, le definizioni associate e le loro relazioni. Il glossario è formato da asset con specifiche relazioni e obiettivi.
- **Loader:** definisce un resoconto di tutte le attività del modulo, in particolare la sintesi dei caricamenti effettuati con il dettaglio della relativa frequenza ed esito.

## Differenza tra Data Governance e Data Management

la Data Governance consiste nella definizione, implementazione ed esecuzione di una policy che costituisce il **Data Management** da applicare, con tutte le **procedure** e gli **strumenti** necessari per la gestione e fruizione dei dati.

Definire un completo **processo di gestione e valorizzazione dei dati aziendali** e di declinarsi in modo proporzionale alle esigenze della specifica realtà:

- **Data Dictionary:** fornisce una visione dei metadati delle basi dati importate nel modulo di Data Governance, approfondendo ogni singolo campo/tabella sia dei metadati che delle statistiche ad essi legati.
- **Data Quality:** gestisce il core engine dei controlli di qualità registrati, effettuati e monitorati. È possibile effettuare controlli di qualità sia in modalità guidata che custom, verificando poi score e risultati.
- **Monitoring:** gestisce la visualizzazione e l'interpretazione dei risultati registrati durante i controlli schedulati, fornendo la possibilità di verificare l'andamento dello score di qualità per ogni dimensione.
- **Data Lineage:** identifica il ciclo di vita del dato tramite una rappresentazione a grafo, così da poter verificare sia l'origine dei dati e le loro trasformazioni durante i diversi processi.
- **Business Glossary:** fornisce un repository utile per comunicare e governare i termini aziendali, le definizioni associate e le loro relazioni. Il glossario è formato da asset con specifiche relazioni e obiettivi.
- **Loader:** definisce un resoconto di tutte le attività del modulo, in particolare la sintesi dei caricamenti effettuati con il dettaglio della relativa frequenza ed esito.

## La Data Governance nel concreto

1. **Data Architecture Management:** corrispondente alla struttura complessiva dei dati e delle risorse che costituiscono l'architettura aziendale;
2. **Data Development** (modellazione e progettazione dei dati): relativa a tutti gli aspetti legati all'analisi, alla progettazione, alla costruzione, al testing e alla manutenzione delle fasi operative;
3. **Database Operations Management** (archiviazione dei dati e operazioni): in merito alla distribuzione degli storage dei data asset strutturati, ad esempio nei vari sistemi di database;
4. **Data Security Management** (sicurezza dei dati): nel rispetto della conformità prevista per garantire privacy e autenticazioni secondo le normative vigenti;
5. **Reference & Master Data Management:** comprende tutte le operazioni utili ad acquisire, estrarre, trasformare, spostare, consegnare, replicare, federare e virtualizzare i dati strutturati;
6. **Datawarehouse & Business Intelligence Management:** per garantire la gestione dell'analisi descrittiva dei dati per le attività di report utili a supportare i processi decisionali
7. **Document & Content Management:** comprende tutte le operazioni utili ad archiviare, proteggere, indicizzare e accedere ai dati non strutturati per garantire integrazione ed interoperabilità con i dati strutturati;
8. **Metadata Management:** utilizzo di standard per ridurre la ridondanza e garantire una migliore qualità dei dati e utilizzo di Data Catalogue per la raccolta, categorizzazione, manutenzione, integrazione, controllo, gestione e distribuzione dei metadati;
9. **Data Quality Management:** consiste nella definizione, monitoraggio e mantenimento dell'integrità dei dati, oltre che al progressivo miglioramento qualitativo dei dati stessi.

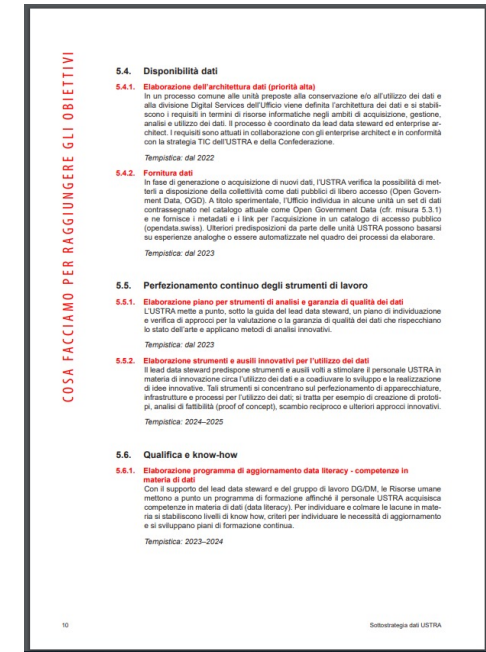
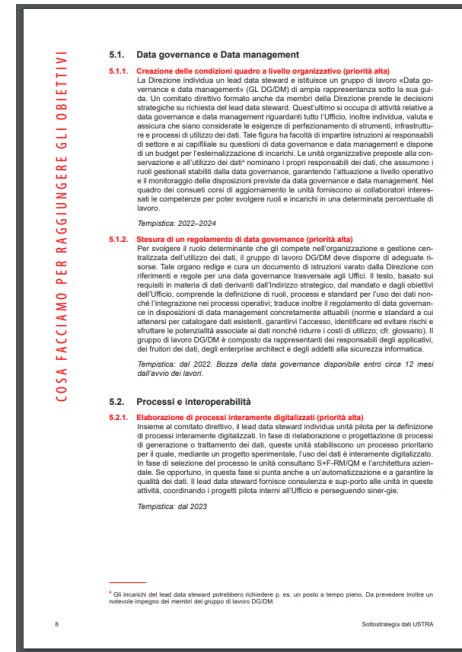
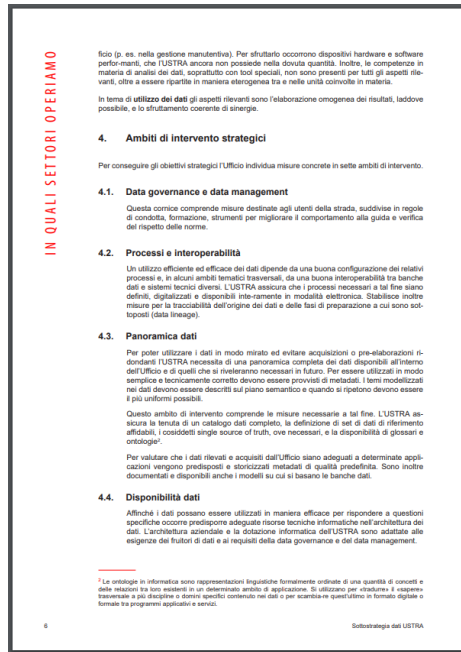






# La Data Governance nel concreto

## Esempio: Documento «Sottostrategia Dati - ASTRA»

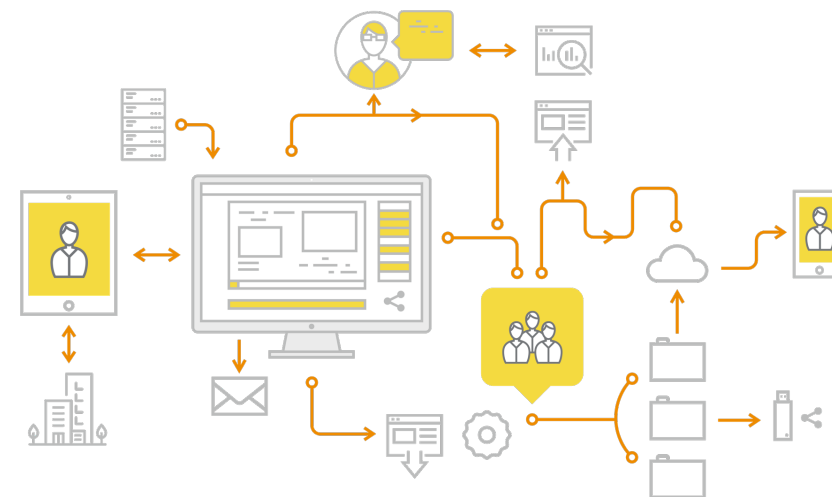


<https://www.astra.admin.ch/dam/astra/it/dokumente/direktion/teilstrategie-daten.pdf.download.pdf/Sottostrategia%20dati.pdf>

[www.progetti.interreg-italiasvizzera.eu](http://www.progetti.interreg-italiasvizzera.eu)

## La Data Governance nel concreto: Descrizione dei dati

<b>DATASET DESCRIPTION</b>	Descrivere i dati esistenti o che si intendono creare, indicandone provenienza, natura e ordine di grandezza. Motivare la creazione di nuovi dataset e relativo valore aggiunto.
<b>Provenance of data and metadata (content)</b>	Descrivere se i dati provengono da interviste, indagini, se sono estratti da archivi disciplinari, banche dati e/o da altri progetti
<b>Type of data</b>	Descrivere se sono qualitativi, quantitativi
<b>Nature and formats</b>	Descrivere natura e formato dei dati (meglio se di formato non proprietario), ad esempio: documenti testo (DOC, ODF, PDF, TXT, etc); immagini (JPG, GIF, SVG, PNG, TIFF); video/film (MPEG, AVI, WMV, MP4); registrazioni audio (MP3, WAV, AIFF, OGG, etc); dati strutturati (HTML, JSON, TEX, XML, RDF); tabelle (CSV, ODS, TSV, XLS, SAS, Stata, SPSS portable); codici sorgente (C, CSS, JavaScript, Java, etc); configuration data (INI, CONF, etc) database (MS Access, MySQL, Oracle)
<b>Amount of data</b>	Stimare l'ordine di grandezza dell'intero dataset (GB, MB, TB, PB)
<b>Requirements for software e hardware</b>	Specificare l'eventuale utilizzo di software o hardware specifici
<b>Data re-use and integration</b>	Nel caso di riutilizzo di dati, dichiarare la loro integrazione e riuso nel progetto, definire il contesto di riutilizzo, definire i destinatari. Pianificare se sono previste traduzioni
<b>New Dataset value</b>	Motivare la creazione del nuovo dataset, definendone il valore aggiunto per la comunità scientifica o altro
<b>Audio-visual materials</b>	In caso di video, segnalare la durata (il dato è utile per pianificare i costi di progettazione e archiviazione)
<b>Software application</b>	Descrivere se verranno prodotti software application per facilitare ricombinazioni di dataset di diversa origine



## La Data Governance nel concreto: Standard e Metadata

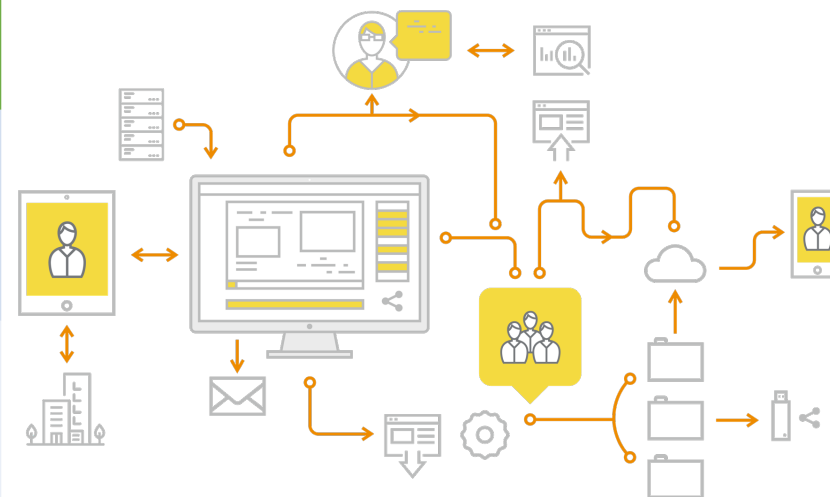
<b>STANDARDS AND METADATA</b>	Motivare la creazione di nuovi dati, in termini contenuto e qualità. Descrivere i metadati e la documentazione accompagnatoria dei dataset, con riferimenti a standard di disciplina, interoperabilità e normativa
<b>Metadata standards and data documentation</b>	Descrivere il tipo di metadati con riferimento a standard (es: <a href="#">Dublin Core</a> o Datacite Metadata Schema)
<b>Methodologies for data collection/generation</b>	Descrivere le metodologie di raccolta e produzione dei dati durante il processo di ricerca. <ul style="list-style-type: none"> <li>• Chi e come raccoglie i dati;</li> <li>• Chi e come li struttura e memorizza;</li> <li>• Chi e come li elabora;</li> <li>• Chi e come li distribuisce.</li> </ul>
<b>Data quality and standards</b>	<ul style="list-style-type: none"> <li>• Documentare come garantire coerenza e qualità ai dati raccolti o prodotti in termine di standard, calibrazione, validazione, review</li> <li>• riferirsi a idonei standard disciplinari se esistenti: nel caso in cui allo stadio attuale le tipologie di dati non rientrano in standard, prevenirne la conversione;</li> <li>• in caso di riuso di metadati, dichiarare se ci sono vincoli al loro riutilizzo;</li> </ul>
<b>Metadata quality</b>	La riusabilità dei dati è assicurata dalla qualità dei metadati a loro associati





## La Data Governance nel concreto: Data Management

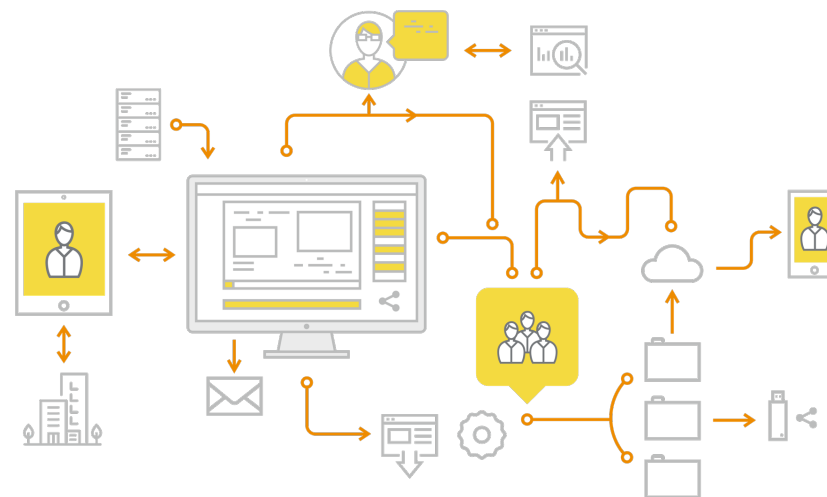
<b>DATA MANAGEMENT, DOCUMENTATION AND CURATION</b>	Descrivere in termini tecnici i processi adottati per la gestione, documentazione, cura e conservazione dei dati. È un Sommario. Stimare i costi associati alla gestione e mantenimento dei dati per rendere FAIR il progetto, facendo attenzione che i costi del primo anno saranno più alti.
<b>Managing, storing and curating data</b>	Predisporre un <b>Preservation Plan</b> che descriva le modalità di: <ol style="list-style-type: none"> <li>1. memorizzazione;</li> <li>2. backup;</li> <li>3. trasmissione;</li> <li>4. cura dei dati nel breve e medio periodo, con riferimenti a pratiche, standard e regolamenti ove applicabil</li> </ol>
<b>Preservation strategy and standards</b>	<ol style="list-style-type: none"> <li>1. Descrivere le modalità di conservazione a lungo termine dei dati in repositories certificati (Iris o altro repository di ateneo, di comunità, di dominio o specifici del progetto);</li> <li>2. se la conservazione riguarda solo una parte dei dati raccolti,</li> <li>3. dichiarare la durata di conservazione dei dati nel repository e l'eventuale cancellazione dopo un determinato periodo di tempo;</li> <li>4. indicare i costi annui (definire da quando) per:                         <ul style="list-style-type: none"> <li>• I repository, gli streaming services, altri servizi o infrastrutture</li> <li>• l'eventuale preparazione dei dati per l'archiviazione (costi da pareri legali sui diritti o soluzioni tecniche per la conversione)</li> <li>• il riuso/visualizzazione dei dati.</li> </ul> </li> </ol>





## La Data Governance nel concreto: Confidenzialità e sicurezza

<b>DATA SECURITY AND CONFIDENTIALITY OF POTENTIALLY DISCLOSED INFORMATION</b>	E' la sezione più importante per i progetti di ricerca che raccolgono dati sensibili, anche se tutti i progetti devono prevedere informazioni sulla sicurezza delle procedure adottate per conservare, trasmettere e scambiare dati. Quanto descritto in questa sezione è soggetto alla validazione da parte del Comitato Etico
<b>Confidential and Sensitive Data</b>	Identificare se ci sono i dati di natura confidenziale / sensibili
<b>Data security standards</b>	Identificare formalmente gli standard (Es: ISO 27001) l'istituto aderisce
<b>Main risks to data security</b>	<p>Descrivere i rischi per i partecipanti alla raccolta dei dati, in termini di confidenzialità e sicurezza.</p> <p>Ovvero, per ogni rischio, identificare:</p> <ul style="list-style-type: none"> <li>• Livello del rischio (basso, medio, alto);</li> <li>• Probabilità di occorrenza del rischio (bassa, media, alta);</li> <li>• L'impatto rispetto ai diversi attori coinvolti;</li> <li>• Le modalità di monitoraggio;</li> <li>• Le contromisure adottate in caso di occorrenza.</li> </ul>
<b>Ethics and Legal Compliance</b>	Con riferimento all'identificazione dei rischi legati alla gestione di dati sensibili e di tutti i metadati a loro associati, specificare se la ricerca coinvolge cellule staminali embrionali umane, persone, animali, ambiente, salute, sicurezza nazionale eccetera







## La Data Governance nel concreto: Responsabilità e risorse

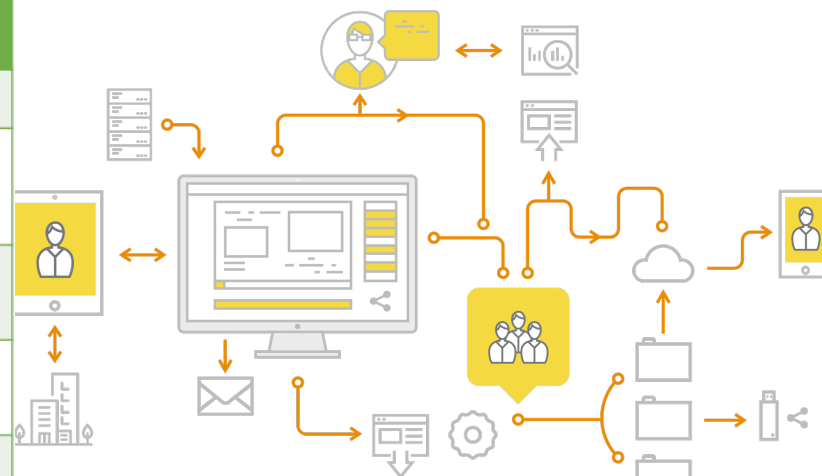
RESPONSIBILITIES AND RESOURCES	Specificare le responsabilità a livello di progetto/Istituto. Se possibile individuare anche il nominativo del referente . Segnalare e giustificare inoltre eventuali risorse aggiuntive
<b>Responsibility</b>	specificare chi ha la responsabilità: <ul style="list-style-type: none"> <li>• nella raccolta dei dati e creazione di metadati</li> <li>• nel garantire la qualità dei dati</li> <li>• nel garantire la sicurezza dei dati</li> <li>• nel garantire l'archiviazione e condivisione dei dati</li> <li>• nel garantire la conformità con la policy dei dati dell'istituto</li> </ul>
<b>Additional resources required to deliver the plan</b>	specificare le richieste, possibilmente corredate da una stima dei costi per: - Specialist/Technical expertise, consulenza legale, consulenza tecnica (o costi per formazione dello staff) per la gestione dei dati e l'archiviazione a lungo termine, Hardware o Software, Costi relativi ai Data Repository
<b>Compliance with ethical and legal requirements</b>	Dichiarare se il progetto è conforme con i requisiti legali e etici di istituto





## La Data Governance nel concreto: Policy di riferimento

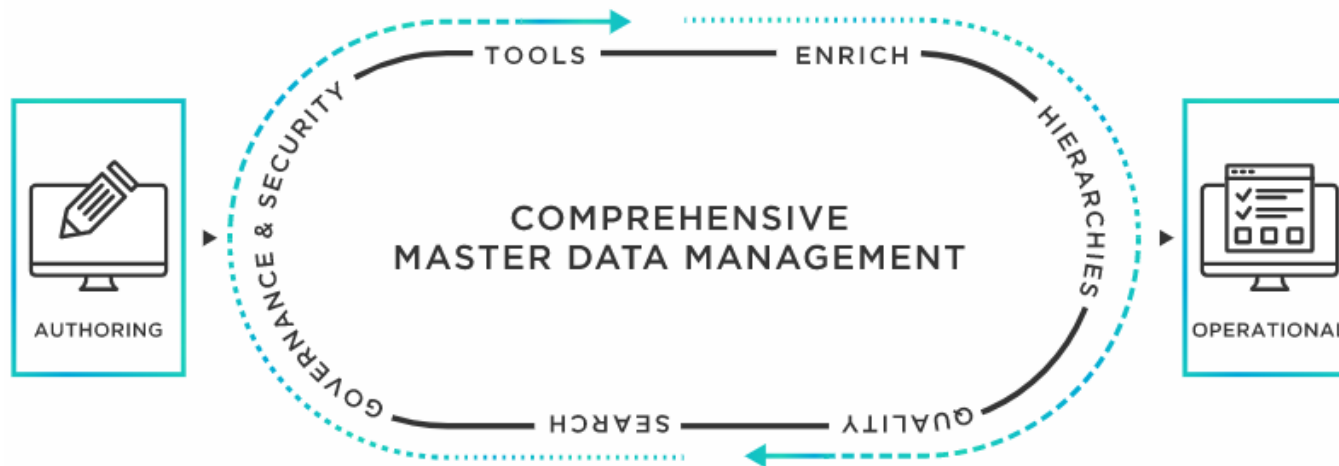
RELEVANT INSTITUTIONAL POLICIES ON DATA SHARING AND DATA SECURITY	Riferimenti alle politiche di Istituto
<b>Research Data Policy</b>	URL alla Policy
<b>Open Access Repository and Scientific Publications</b>	URL alla Policy sul repository di Ateneo
<b>Where the DMP will be deposited</b>	Indicare che il DMP verrà caricato in Iris o altro repository
<b>How DMP will be accessible</b>	Indicare come viene assicurato l'accesso al DMP
<b>DMP preservation Responsibility</b>	Indicare chi è responsabile per la conservazione del DMP



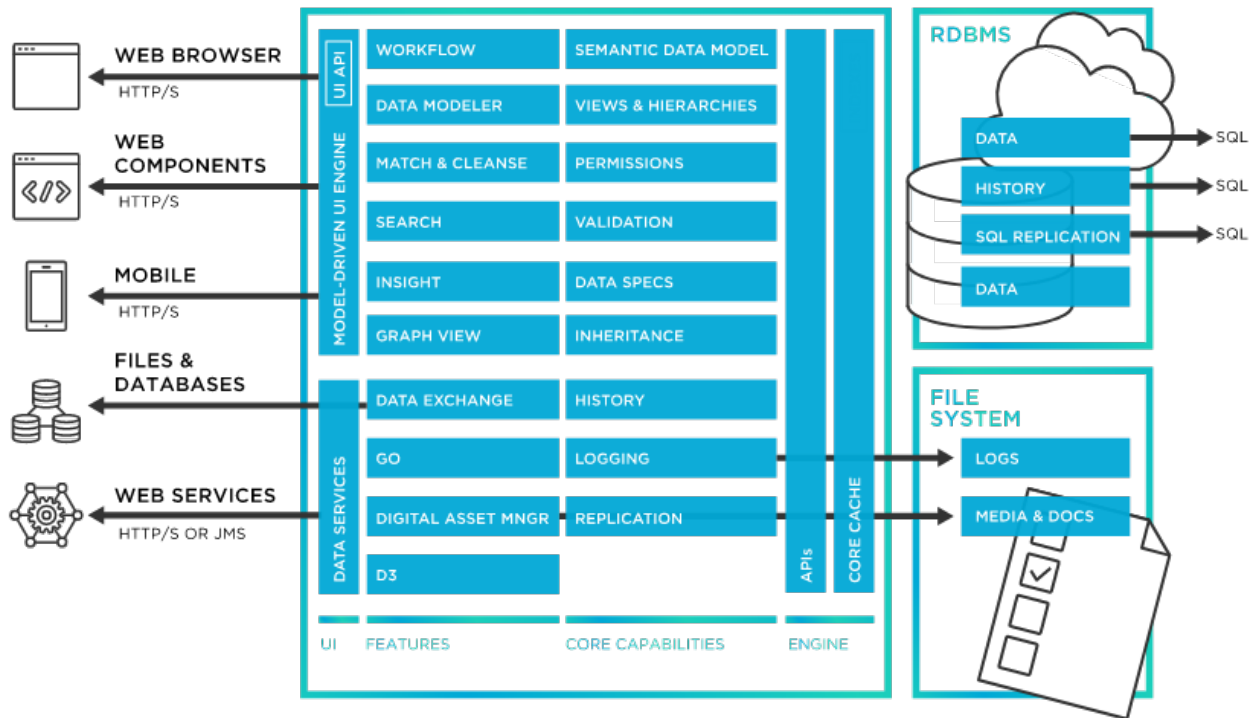




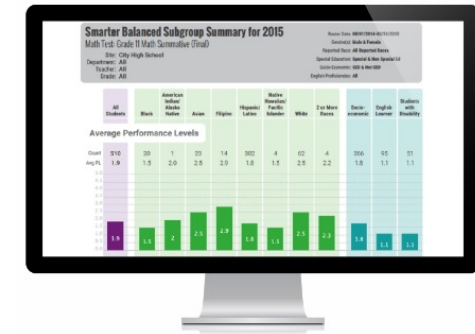
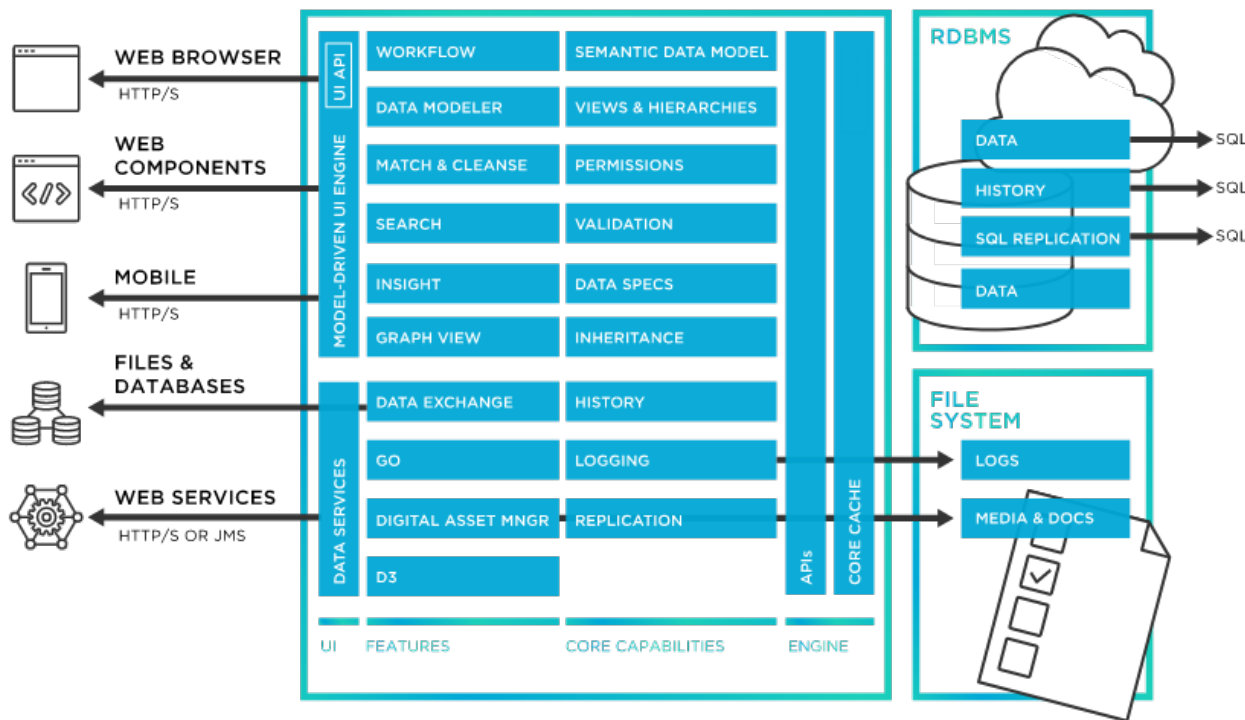
# Master Data Management



# Master Data Management



# Master Data Management



# Master Data Management

MDM	Master ID	First Name	Middle Name	Last Name	Address	City	State	ZIP
	M-10001	Robert	John	Downey	10880 Malibu Point	Malibu	CA	90265

Source System 1	SS1_ID	First Name	Middle Name	Last Name	Address	City	State	ZIP
	12345	Robert		Downey	10880 Malibu Point		CA	90265

Source System 2	SS2_ID	First Name	Middle Name	Last Name	Address	City	State	ZIP
	40035398	Bob	John	Downey	10880 Malibu Pt.		CA	90265

Source System 3	SS3_ID	First Name	Middle Name	Last Name	Address	City	State	ZIP
	C987654	Robert	J	Downey	10880 M Point	Malibu		



**Golden Records**  
 Customer Golden Records 1.1M  
 Publish | Close rules

Filter	Tax Numb...	Customer ...	Customer ...	Addr_City	Addr_Co...
STR_1 MOSKOV...	5043025182	4151104031		SERPUKHOV	Russia
JURJUA GAGARIN...	174382563	631029687		BELGRADE	Serbia
"A.S. GLOBAL F...	AGF100810802	3548894254		CIUDAD DE MEXICO	Mexico
"AIR FLOW PVT...	HARDHIAN SING...	9654105493		NEW DELHI	India
"AL GHATH FOR...	AL WAKRAH MA...	44649139		DOHA	Qatar
"AVS-LAY" LTD,	UL KLIMOVA, 50...	9723953233		NOGINSK	Russia
"B" LINES	NGO DAC KE, 74...	031184366		HO CHI MINH	Vietnam
"BAZIS-IMPORT...	SEGEDSKAYA ST...	39645701037		ODESA	Ukraine
"BEIJING SODIY...	LIANG MA QIAO R...	1064652928		BEIJING	China

1 of 21,303 | Records per page: 50 | Configure table

**Golden Records rules**  
 Edit rules | Expand all

- Customer Name**  
 Most common value | 0 conditions | 0 overrides
  - RULE: Most common value
  - Input attribute: Customer Name
  - CONDITIONS: No Conditions
- Addr\_City**  
 Most common value | 0 conditions | 0 overrides
- Addr\_Country**  
 Most common value | 0 conditions | 0 overrides
- Addr\_Ln\_2**  
 Most common value | 0 conditions | 0 overrides
- Addr\_Ln\_3**  
 Most common value | 0 conditions | 0 overrides
- Addr\_PO\_Box**  
 Most common value | 0 conditions | 0 overrides
- Addr\_Postcode**

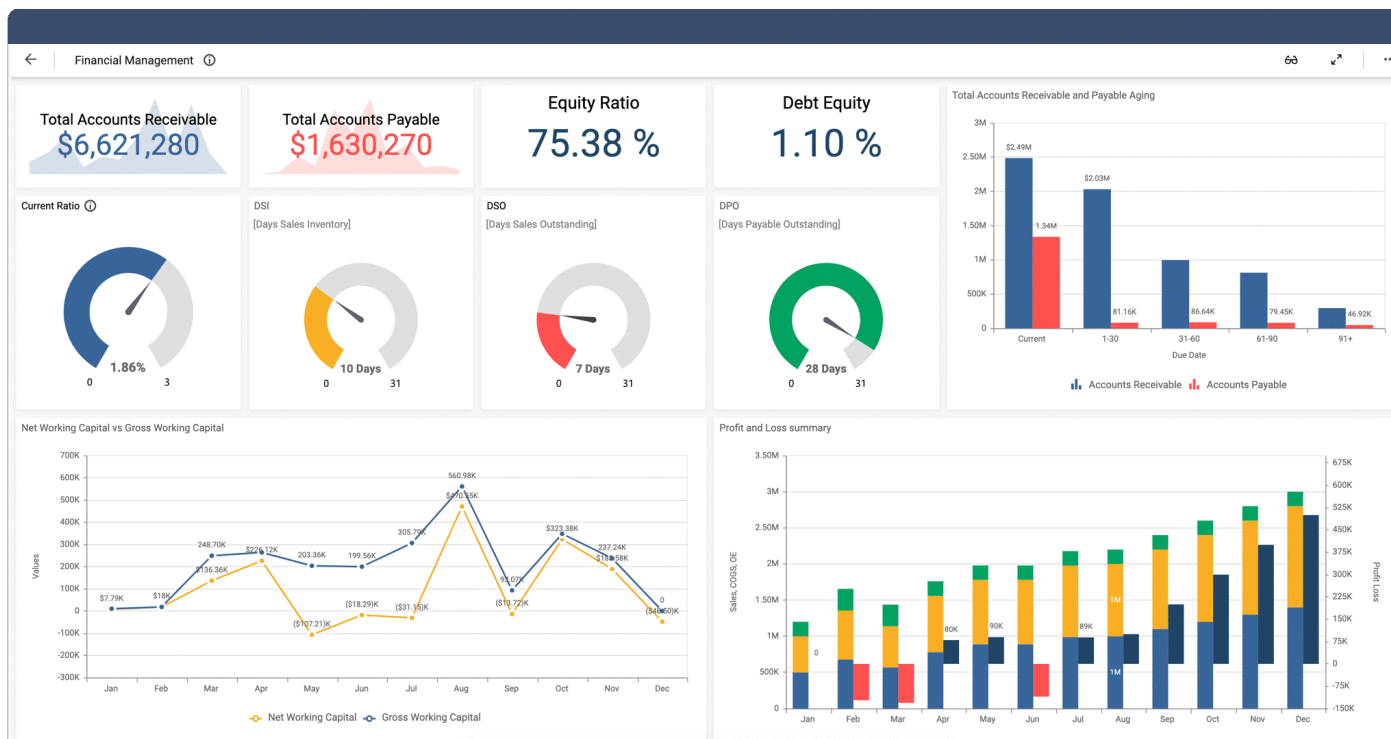
# Master Data Management

## GLOSSARIO DEI TERMINI

TERMINE	DESCRIZIONE	SINONIMI	COLLEGAMENTI LOGICI
<i>Ditta</i>	Organizzazione che esegue le prestazioni attraverso la manodopera del suo personale, emette fatture e proposte di preventivo ai clienti.	Società	
<i>Indirizzo</i>	Insieme delle proprietà: Via, Numero civico, CAP, Città, Provincia	Sito	
<i>Prestazione</i>	Attività lavorativa, quantificata in €/mq (prezzo unitario).		Dettaglio fattura, Dettaglio preventivo
<i>Dettaglio fattura</i>	Componente della fattura contenente gli elementi: prestazione, quantità, iva, importo.	Dettaglio	Fattura, Prestazione
<i>Fattura</i>	Documento emesso dalla ditta verso il cliente, recante un importo per le prestazioni fornite in un certo cantiere.		Pagamento, Dettaglio fattura
<i>Cantiere</i>	Luogo, di proprietà di un cliente, dove il personale della ditta esegue le prestazioni. Si precisa che 2 cantieri differenti possono trovarsi nello stesso luogo.	Opera	Bolla, Cliente, Personale
<i>Personale</i>	Persone fisiche che compongono la ditta e svolgono per essa le prestazioni nei cantieri. Il personale è composto dai soci, dai dipendenti e dai prestatori di manodopera.		Cantiere
<i>Cliente</i>	Soggetto che richiede alla ditta una o più prestazioni in uno o più cantieri di sua proprietà. Può richiedere proposte di preventivo ed effettuare pagamenti di fatture. Può essere privato o un'impresa.		Pagamento, Proposta, Cantiere
<i>Preventivo</i>	Documento emesso dalla ditta verso il cliente, recante una proposta riguardante l'importo da pagare a seguito di prestazioni che potranno essere eseguite in un futuro cantiere. Può essere accettato o meno dal cliente.		Dettaglio Preventivo, Proposta
<i>Dettaglio preventivo</i>	Componente del preventivo contenente gli elementi: prestazione, quantità, importo.		Preventivo, Prestazione
<i>Pagamento</i>	Azione svolta dal cliente verso la ditta a seguito dell'esecuzione di una o più prestazioni. Il pagamento può avvenire in varie modalità (ad es. in contanti, assegno, bonifico), e può assumere 2 stati (pagamento avvenuto e non).		Fattura, Cantiere, Cliente

## GLOSSARIO DEI TERMINI

# Data Visualization – la visualizzazione dei dati





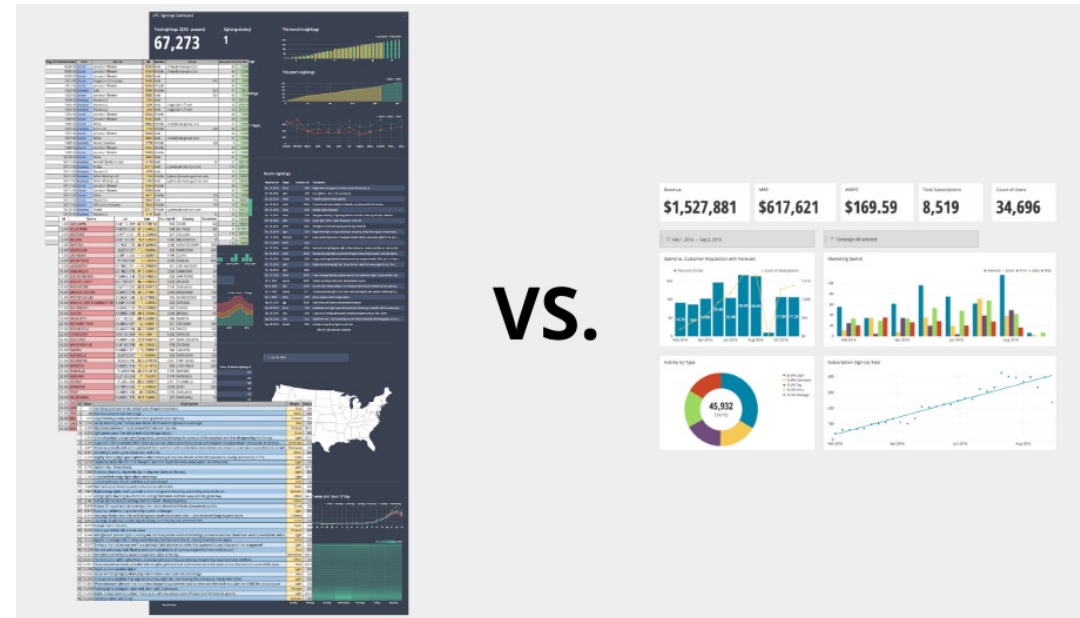
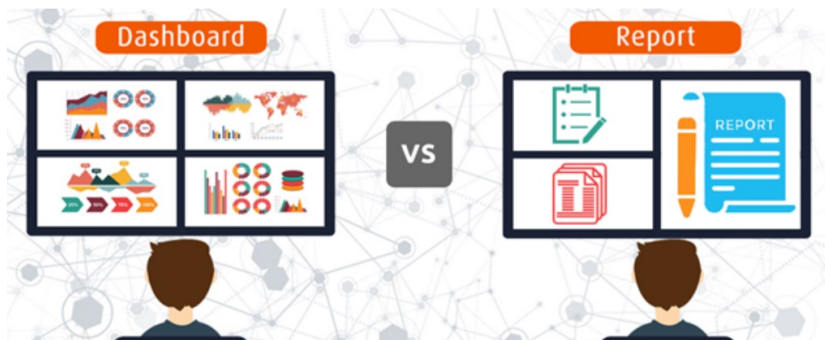
# Data Visualization – la visualizzazione dei dati

Dashboards vs. Reports



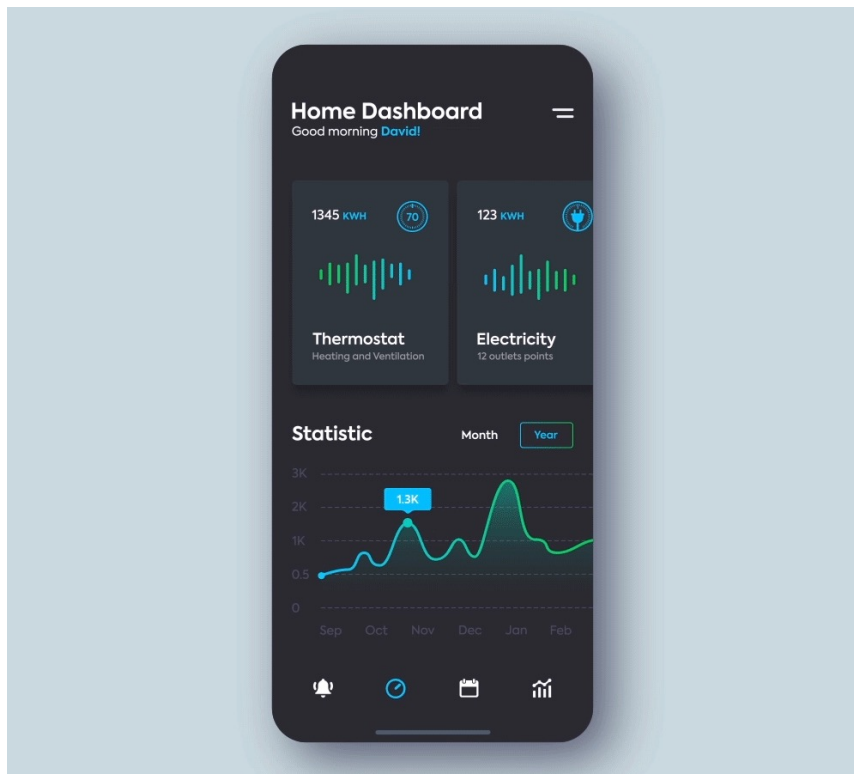
# Data Visualization – la visualizzazione dei dati

Dashboards vs. Reports





# Data Visualization – la visualizzazione dei dati








# Registro trattamento dei dati



I campi presi in considerazione sono:

- Finalità;
- Basi giuridiche;
- Categorie di dati;
- Categorie di interessati;
- Titolare;
- RPD;
- Misure di sicurezza;
- Periodo di conservazione.

RIFERIMENTI ORGANIZZATIVI	INFORMAZIONI DI PROCESSO	TIPOLOGIE	TIPOLOGIA DI DATI TRATTATI	FINALITÀ						CONSENSO	CONSERVAZIONE DEI DATI
01 - Intercomunicazione in Regione Lombardia/RSAS 02 - Supporto amministrativo e Clientele al Designato	01 - Gestione sicurezza attività informative in privacy	01 - Misure Clienti (Gestione Automabile)	01 - Nome, Cognome, Numero di Passaporto	01 - Prof/Incarico	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati
03 - Amministrazione	01 - Gestione sicurezza attività informative in privacy	01 - Misure Clienti (Gestione Automabile)	01 - Nome, Cognome, Numero di Passaporto	01 - Prof/Incarico	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati
04 - Ricerca e Sviluppo Innovazione ICT	01 - Gestione sicurezza attività informative in privacy	01 - Misure Clienti (Gestione Automabile)	01 - Nome, Cognome, Numero di Passaporto	01 - Prof/Incarico	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati
05 - Marketing	01 - Gestione sicurezza attività informative in privacy	01 - Misure Clienti (Gestione Automabile)	01 - Nome, Cognome, Numero di Passaporto	01 - Prof/Incarico	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati
06 - Supporto Database	01 - Gestione sicurezza attività informative in privacy	01 - Misure Clienti (Gestione Automabile)	01 - Nome, Cognome, Numero di Passaporto	01 - Prof/Incarico	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati	01 - Finalità per cui sono trattati i dati



Mostra 10 dati
 Cerca:

Divisione	Scopo	Motivi giustificativi per il trattamento	Soggetti interessati	Destinatari	Misure di sicurezza	Dati personali trattati	Risk Assessment	Richiesta DPIA
Business Intelligence	Analisi statistiche e monitoraggio delle prestazioni aziendali	Interesse preponderante del Titolare del trattamento	Collaboratori, Dipendenti, Pazienti	Autorità pubbliche cantonali, Ufficio federale statistico	[REDACTED]	Dati particolari	●	No
Fisica medica	Radioprotezione		Pazienti	[REDACTED]	[REDACTED]	Dati particolari	●	No
HR	Rilevazione delle presenze	Contratto	Dipendenti	[REDACTED]	[REDACTED]	Dati personali	●	No
HR	Formazione del personale	Contratto	Dipendenti	Consulenti / liberi professionisti	[REDACTED]	Dati personali	●	No
HR	Conservazione della documentazione del	Contratto	Collaboratori, Dipendenti		[REDACTED]	Dati	●	No