

Governance dei dati e conformità
in materia di privacy:

16 passi pratici per prepararsi al GDPR con Talend

Sunil Soares e Jean-Michel Franco

8 luglio 2017



L'Unione europea (EU) ha pubblicato il regolamento generale per la protezione dei dati (in inglese General Data Protection Regulation, GDPR) nel maggio 2016. Dopo un periodo di transizione di due anni, le norme saranno applicabili a partire dal 25 maggio 2018. Il regolamento concerne il trattamento dei dati personali di tutti gli individui interessati, compresi i clienti, i collaboratori e i prospect di un'azienda. Per dato personale si intende qualsiasi informazione riguardante una persona fisica identificata o identificabile (l'«interessato»); si considera identificabile la persona fisica che può essere individuata, direttamente o indirettamente, con particolare riferimento ad una caratteristica come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale. Esempi tipici di tali dati sono i nomi dei clienti, i contatti di una base CRM, gli stipendi, i bonus e le valutazioni delle performance del personale, ma il regolamento si applica anche a una nuova tipologia di dati, ovvero quelli in provenienza dai sensori, che possono indicare la posizione di un veicolo e il comportamento del guidatore.

Il regolamento si applica ai residenti nell'Unione europea, anche qualora i dati siano elaborati da entità che operano di fuori del territorio europeo, nell'ambito di altre giurisdizioni quali gli Stati Uniti, la regione Asia-Pacifico, il Medio Oriente e l'Africa. La violazione delle disposizioni della GDPR è soggetta a sanzioni pecuniarie fino a 20 000 000 EUR, o per le imprese, fino al 4 % del fatturato mondiale totale annuo dell'esercizio precedente, se superiore.

Un solido programma di governance dei dati riveste un ruolo cruciale nel quadro generale della GDPR, ma per mettersi in regola con le normative occorre tenere presente anche gli ambiti di applicazione classici della governance dei dati, come la data ownership (proprietà dei dati), la gestione dei metadati, dei modelli e della qualità dei dati. Inoltre, un progetto GDPR implica l'integrazione di soluzioni self-service per la preparazione dei dati e la data stewardship, al fine di incentivare ogni utente aziendale ad assumere la responsabilità della protezione dei dati in modo concreto e verificabile e non come semplice applicazione di direttive legislative astratte.

Approccio globale

La figura 1 mostra i 16 passi da seguire per mettere a punto un programma di adeguamento al regolamento GDPR.

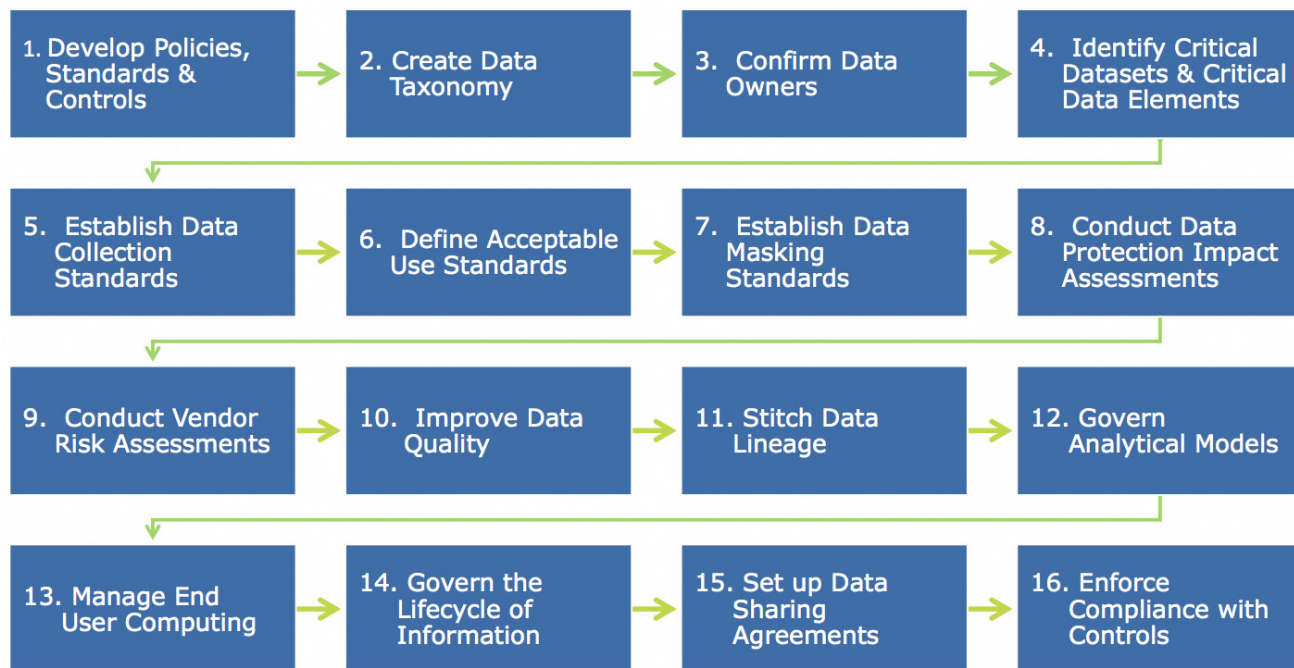


Figura 1: un approccio end-to-end per soddisfare ai requisiti del GDPR.

1. Sviluppare regole, standard e controlli di governance dei dati

Al fine di rispettare il regolamento, le aziende devono definire una strategia di governance dei dati basata su regole, standard e controlli. La tabella 1 offre una sintesi dei controlli richiesti e una lista dei tool di Talend che consentiranno di automatizzarli e tracciarli.

Tabella 1: esempi di controlli GDPR			
Articolo GDPR	Descrizione GDPR	Controlli GDPR	Soluzione Talend
Articolo 6	Liceità del trattamento	<ul style="list-style-type: none">• Ottenere l'approvazione dell'ufficio legale e del responsabile della conformità durante la fase di concezione di ogni nuovo progetto che richieda l'elaborazione di dati personali.	<ul style="list-style-type: none">• Talend Metadata Manager• Talend Master Data Management (MDM)
Articolo 7	Condizioni per il consenso	<ul style="list-style-type: none">• Ottenere il consenso informato dell'interessato e fornire la prova che quest'ultimo abbia dato tale consenso.	<ul style="list-style-type: none">• Talend Data Quality• Talend Big Data and/or Talend MDM
Articolo 9	Trattamento di categorie particolari di dati personali che rivelino l'origine razziale o etnica, o le opinioni politiche	<ul style="list-style-type: none">• Identificare categorie particolari di dati come CDE (Critical Data Elements)• Ottenere l'approvazione dell'ufficio legale e del responsabile della conformità per il trattamento di categorie particolari di dati durante la fase di concepimento del progetto.	<ul style="list-style-type: none">• Talend Metadata Manager• Talend Data Quality
Articolo 11	Trattamento che non richiede l'identificazione	<ul style="list-style-type: none">• Mascheramento dei dati	<ul style="list-style-type: none">• Talend Data Quality• Talend Data Preparation
Articolo 30	Registri delle attività di trattamento	<ul style="list-style-type: none">• Data lineage (Tracciabilità) per i dati sensibili all'interno dell'azienda e per le terze parti	<ul style="list-style-type: none">• Talend Metadata Manager

2. Creare la mappa dei dati

Il team addetto alla governance dei dati dovrà classificarli, in funzione dell'architettura dati dell'azienda, in categorie e sottocategorie. Ciò consentirà di ottenere una visione olistica di tutti i dati che l'azienda deve amministrare ai fini di adeguarsi alla GDPR. La mappa dei dati potrà ad esempio includere le informazioni relative ai dipendenti come categoria di livello 1 (figura 2). Le altre informazioni che li riguardano possono essere ulteriormente classificate in categorie multiple di livello 2 (stipendio e bonus, identità, contatti, informazioni mediche protette, social network, valutazione delle performance).

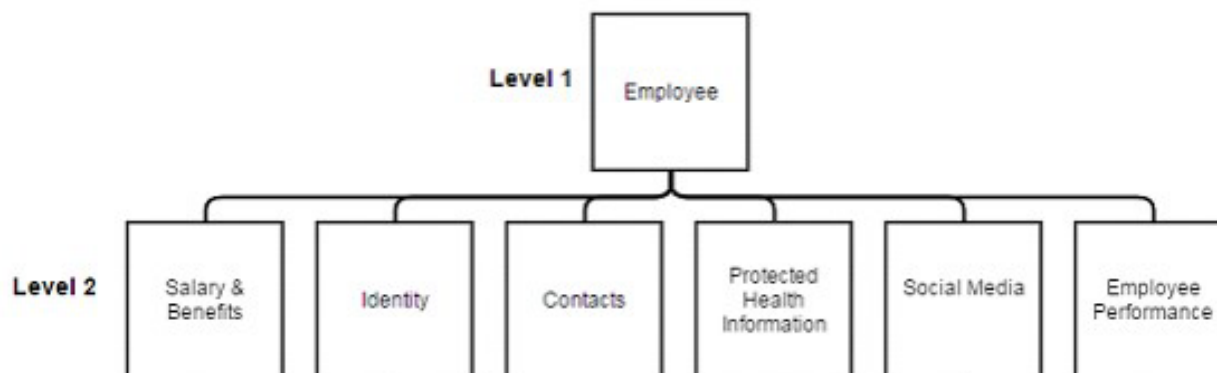


Figura 2: mappa dei dati per le informazioni relative ai dipendenti.

Talend Metadata Manager fornisce un glossario aziendale che può essere utilizzato per definire degli insiemi di termini e per collegarli a categorie e sottocategorie. La compilazione di un simile glossario è semplice: basta integrare (dragging) in un modello di dati esistente e ben consolidato, importare termini e definizioni provenienti da altre fonti (ad es. file CSV, Excel) oppure creare contenuto interattivo a partire dall'interfaccia utente durante la fase di classificazione degli elementi. Una volta pubblicato, il glossario è accessibile da parte di qualsiasi utente autorizzato tramite un'interfaccia di ricerca (figura 3).



Figura 3: il glossario aziendale di Talend Metadata Manager.

3. Confermare i proprietari dei dati (data owners)

Il passo successivo consiste nel definire i proprietari dei dati che in ultima analisi saranno responsabili di una o più categorie e sottocategorie di dati. Questi designeranno i data steward che sovrintenderanno alle operazioni ricorrenti sui dati. A loro volta, i data steward dovranno identificare set di dati e CDE (Critical Data Elements) nonché impostare standard per la raccolta, l'uso e il mascheramento dei dati. Talend Metadata Manager può anche catalogare i proprietari dei dati e i data steward per ogni categoria e sottocategoria di dati ed assegnare loro i ruoli e i workflow associati.

Talend Data Manager può inoltre elencare i proprietari dei dati per “cliente”, ma anche per “identità cliente”, “fatturazione”, “contatto cliente” e “indirizzo di spedizione”. Con Talend Data Stewardship si possono coinvolgere i proprietari dei dati e i data steward in workflow con controlli d'accesso sicuri e verificabili basati sui ruoli per abilitare processi self-service quali data arbitration, risoluzione degli errori e validazione degli standard (figura 4).

Source name	Trust score
1 Master Data Management	80% %
2 Marketo (from our Partner LeadMaster360)	35% %
3 Twitter Opt-ins with Social Sign-on	50% % ✕

Figura 4: assegnazione di ruoli collaborativi per il processo di data arbitration, la risoluzione degli errori e le campagne di merging con Talend Data Stewardship.

4. Identificare i set di dati e i CDE sensibili

Questo approccio completo in 16 passi dimostra che la governance dei dati è un esercizio molto rigoroso. I data steward devono concentrare i loro sforzi sull'identificazione dei set di dati e dei CDE sensibili nelle rispettive categorie. Nell'esempio precedente, l'identità di un collaboratore è composta da un certo numero di elementi, tra i quali nome, sesso, data di nascita, codice fiscale. Le informazioni sui social network utilizzati sono composti da diversi set di dati sensibili provenienti da Facebook®, Twitter®, LinkedIn® (figura 2). Il team che ha in carico la governance dei dati deve valutare se le regole per la raccolta e l'utilizzo dei dati sono meglio definite a livello dei set di dati sensibili piuttosto che a livello dei singoli elementi. Per esempio, standard accettabili di utilizzo e di sicurezza possono essere gestiti più efficacemente per l'insieme dei dati Facebook (set di dati sensibili) piuttosto che per il solo ID Facebook (elemento sensibile).

Talend Metadata Manager supporta un glossario ISO 11179 che contiene termini aziendali associati a dati personali. Inoltre, riprendendo l'esempio precedente, Talend Metadata Manager può contenere un inventario di termini aziendali relativi all'identità dei clienti, quali "nome", "indirizzo email", "numero di telefono". Talend Metadata Manager supporta anche la mappatura di tali elementi di dati personali verso campi fisici nei sistemi sorgente dell'azienda. Questi campi fisici saranno basati su metadati tecnici raccolti dai sistemi sorgente, che riposano su una grande varietà di connettori di Talend Metadata Manager (figura 5). Questi connettori forniscono una visione sempre aggiornata dell'insieme dei dati, un po' come un navigatore GPS che allerta il conducente quando le condizioni del traffico cambiano.

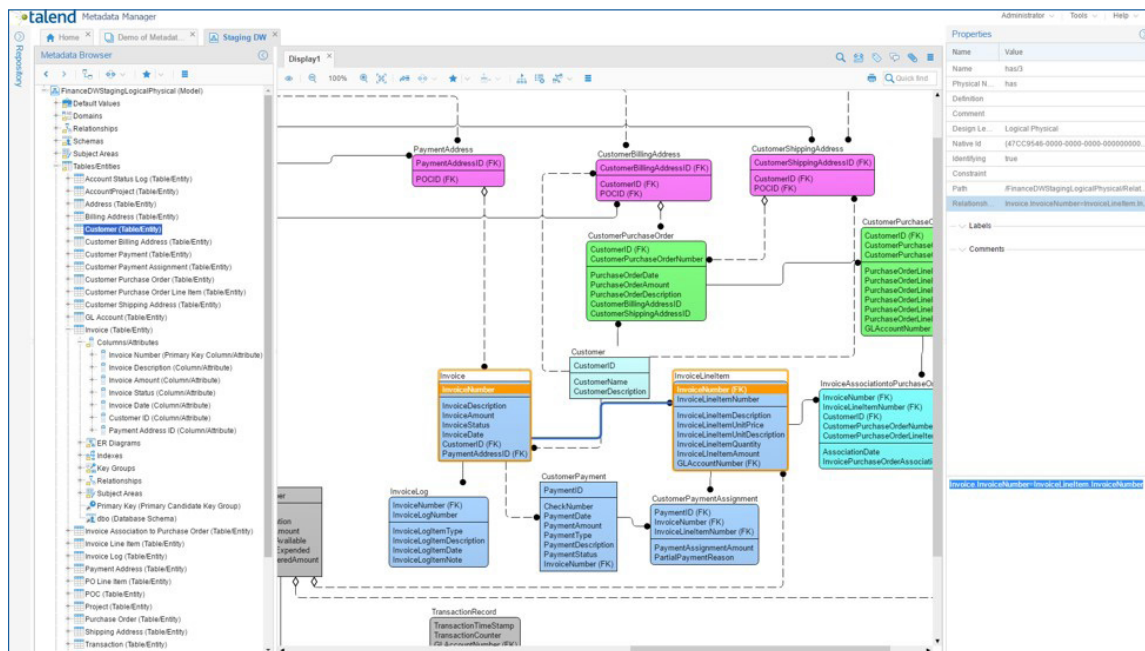


Figura 5: definizione – o retroingegneria – di modelli di dati ed elementi di dati con Talend Metadata Manager.

5. Stabilire degli standard per la raccolta dei dati

L'articolo 25 della GDPR concerne la protezione dei dati fin dalla progettazione e la protezione per "impostazione predefinita". In base a tale articolo, il titolare del trattamento (ovvero l'entità che determina le finalità e i metodi di trattamento dei dati personali) mette in atto misure tecniche e organizzative adeguate, quali ad esempio la minimizzazione dei dati, per garantire che siano trattati, per impostazione predefinita, solo i dati personali necessari per ogni specifica finalità del trattamento. Tale obbligo vale per la quantità dei dati personali raccolti, la portata del trattamento, il periodo di conservazione e l'accessibilità. Il team addetto alla governance dei dati deve stabilire controlli che ottengano l'approvazione del servizio giuridico e del responsabile della conformità durante la fase di concepimento di ogni nuovo progetto. Ad esempio, un'azienda di soluzioni software può utilizzare un modulo online per consentire agli utenti di scaricare una versione di prova. In questo scenario, se è ragionevole che l'azienda richieda il nome e l'indirizzo email dell'interessato, esigere anche la data di nascita e il codice fiscale sarebbe inappropriato.

Talend MDM, Talend Big Data e Talend Data Quality supportano la creazione di un data lake conforme alle disposizioni del GDPR, in cui possono essere riconciliate tutte le informazioni relative all'interessato, compresi i dati personali e i consensi. Queste soluzioni permettono di verificare che "James Smith" e "Jim Smith" siano la stessa persona, anche se solo James Smith ha espresso il suo consenso (opt-in). Il data lake GDPR può anche gestire un singolo archivio che raccoglie tutti i consensi (ricevere offerte via email, accettazione dei cookie, essere contattato telefonicamente ...). Questo repository conserva, per ogni interessato, un elenco di tutti i consensi ottenuti a partire da ogni applicazione aziendale. Fornisce inoltre una traccia di audit (audit trail), una tracciabilità a livello di record dei dati relativi al consenso di un certo individuo per una determinata applicazione. Quando tutte le informazioni saranno state riconciliate nel data lake, il team della governance dei dati sarà in grado di fornire i servizi necessari per soddisfare le richieste emananti dal diritto alla portabilità e dal diritto all'oblio.

Un approccio basato sul data lake, sebbene non sia l'unico possibile, costituisce un obiettivo pragmatico nell'era dei Big Data, nella quale i dati provengono da ovunque, non solo da sistemi centralizzati e ben gestiti. In questi ambienti, i dati in entrata possono non essere completamente strutturati e documentati. Inoltre è necessario catturare le impronte dei dati personali provenienti da questi set di dati grezzi e stabilire le regole lavorando accanto a specialisti quali i data scientist, che devono elaborare questi dati per creare dei modelli di machine learning.

6. Definire standard di utilizzo accettabili

L'articolo 6 del GDPR tratta della liceità del trattamento dei dati personali. Il trattamento è lecito solo in determinate situazioni, nella misura in cui l'interessato ha espresso il consenso al trattamento dei dati, quando il trattamento è necessario all'esecuzione di un contratto, per adempiere un obbligo legale, quando il trattamento è necessario per la salvaguardia degli interessi vitali dell'interessato o di un'altra persona fisica. I responsabili della data governance devono stabilire controlli per garantire che ogni nuovo progetto che richieda l'uso di dati personali sia approvato dall'ufficio legale e dal responsabile della conformità durante la fase di pianificazione.

In base all'articolo 7 comma 1 della GDPR, il titolare del trattamento deve essere in grado di dimostrare che l'interessato ha prestato il proprio consenso al trattamento dei propri dati personali. La governance dei dati deve collaborare con i servizi IT, l'ufficio legale e il responsabile della conformità per creare un repository che raccolga tutti i dati relativi ai consensi dell'azienda. Tale archivio deve contenere un registro immutabile del consenso dato dalle persone interessate, ad esempio quello, da parte dei visitatori del sito web, all'uso di cookie. Come accennato in precedenza, è possibile utilizzare un data lake conforme alla GDPR con Talend MDM, Talend Big Data e Talend Data Quality per catalogare tutti i consensi espressi dai soggetti interessati. La Figura 6 illustra un job Talend che, mediante Talend MDM, viene usato per recuperare dati di opt-in da un sistema di terze parti e pubblicare tali informazioni in tutte le applicazioni che lo richiedono. Questo job sfrutta anche Talend Data Quality per riconciliare i dati tra il sistema esternalizzato e il catalogo centralizzato e conforme alla GDPR, e per riconoscere e associare duplicati in funzione delle soglie di probabilità. Questa pipeline di dati è completamente automatizzata e può essere gestita in modo collaborativo per consentire agli utenti aziendali di definire i controlli sui dati con Talend Data Preparation e / o di elaborare i processi di certificazione dei dati e di data curation con Talend Data Stewardship.

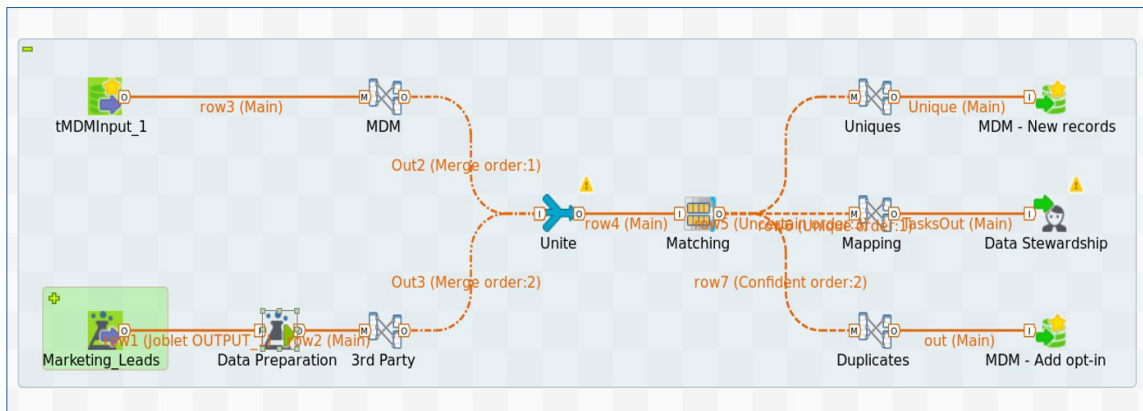


Figura 6: un job Talend riunisce qualità dei dati, data stewardship e integrazione dei Big Data in un ambiente grafico unificato per raccogliere, standardizzare, riconciliare, certificare, proteggere e propagare i dati personali.

Talend Metadata Manager può anche fungere da catalogo di standard di utilizzo accettabili per gli elementi di dati personali. Ad esempio, se nell'ambiente Big Data viene introdotto un nuovo valore, "Halal", per l'attributo "Preferenze alimentari", i workflow di Talend possono servire per ottenere le necessarie autorizzazioni legali, poiché questo campo potrebbe potenzialmente essere sfruttato per determinare le convinzioni religiose dell'interessato, caso disciplinato espressamente dall'articolo 9 del GDPR. Talend MDM conserva un log degli aggiornamenti dei dati master, inclusi i consensi. Nella figura 7, il log mostra che le informazioni relative al consenso sono state aggiunte al profilo del cliente "Pierre Flores" il 14 giugno 2017. Un'ulteriore analisi di questo record fornirà uno storico completo (lineage) che dimostrerà che il sito web della società è l'applicazione che ha raccolto il consenso.

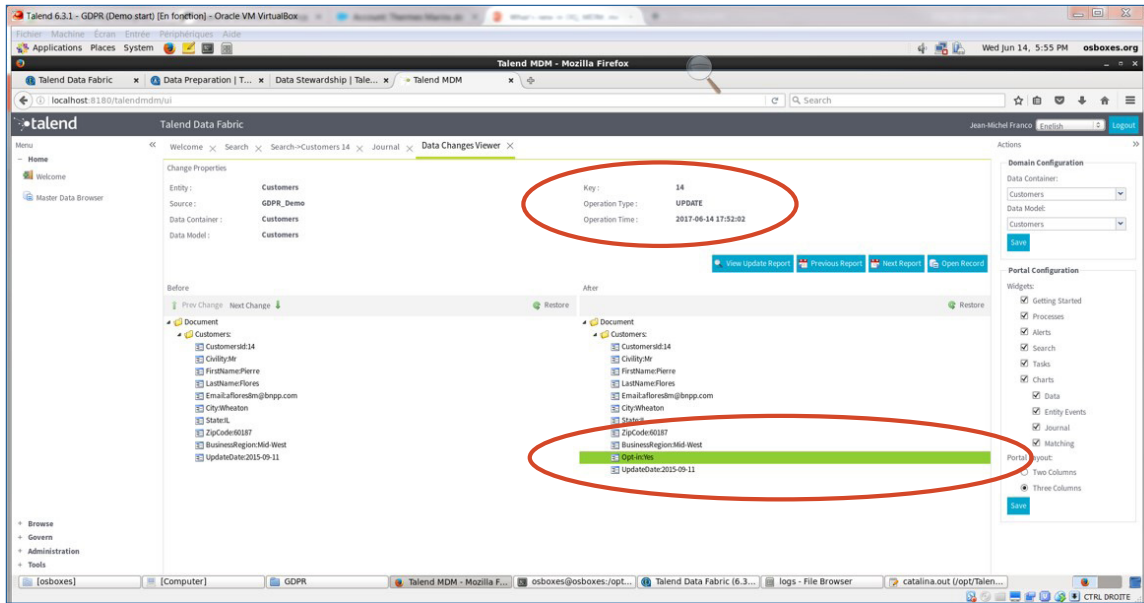


Figura 7: Talend MDM fornisce lo storico (data lineage) dei dati a livello di record, con funzionalità per annullare/ripetere le operazioni, mettendo così a disposizione una traccia di audit (audit trail) per i consensi e altri dati riguardanti un particolare individuo.

7. Stabilire standard per il mascheramento dei dati

L'articolo 26 della GDPR stipula che è auspicabile applicare i principi di protezione dei dati a tutte le informazioni relative a una persona fisica identificata o identificabile. I principi di protezione dei dati non dovrebbero pertanto applicarsi a informazioni anonime, vale a dire informazioni che non si riferiscano a una persona fisica identificata o identificabile o a dati personali resi sufficientemente anonimi da impedire o da non consentire più l'identificazione dell'interessato.

L'articolo 11 del GDPR affronta il trattamento che non richiede l'identificazione. Se le finalità per cui un titolare del trattamento tratta i dati personali non richiedano o non richiedano più l'identificazione dell'interessato, gli obblighi incombenti al titolare del trattamento sono notevolmente minimizzati.

L'articolo 32 della GDPR concerne la sicurezza del trattamento. Tenendo conto dello stato dell'arte e dei costi di attuazione, nonché della natura, dell'oggetto, del contesto e delle finalità del trattamento, come anche del rischio di varia probabilità e gravità per i diritti e le libertà delle persone fisiche, il titolare del trattamento e il responsabile del trattamento (l'entità che effettua il trattamento per conto del titolare del trattamento) dovranno mettere in atto opportune misure tecniche e organizzative per garantire un livello di sicurezza adeguato al rischio.

Gli addetti alla governance dei dati devono prevedere delle procedure per mascherare o codificare convenientemente i dati personali sensibili. Le regole relative al mascheramento dovranno garantire che i dati non possano essere ricostruiti quando vengono combinati più campi. Ad esempio, i data scientist possono richiedere che il campo contenente il nome e il cognome dei dipendenti sia mascherato prima di effettuare un'operazione di analisi. Tuttavia, un data scientist esperto potrà probabilmente risalire all'identità di un certo collaboratore collegando i dati relativi alla denominazione del posto, allo stipendio e al sesso dell'interessato (ad es. "Responsabile risorse umane, donna, stipendio di base di 200 000 \$"). In tale situazione, sarebbe più opportuno occultare la denominazione del posto e fornire solo un'indicazione sulla fascia di stipendio ("più di 100 000 \$").

Grazie a Talend Data Quality, il mascheramento e il mescolamento dei dati (data shuffling) possono essere effettuati in qualsiasi fase di una pipeline (figura 8). Il mescolamento dei dati (data shuffling) è un metodo di mascheramento mediante il quale una colonna (o un set di dati più complessi quali un gruppo di colonne o una partizione) è mescolata a caso in modo tale da nascondere l'identità, mentre i valori pertinenti restano al loro posto. La privacy è quindi rispettata, ma l'analisi e il test dei dati possono essere portati a termine utilizzando i valori iniziali dei dati.

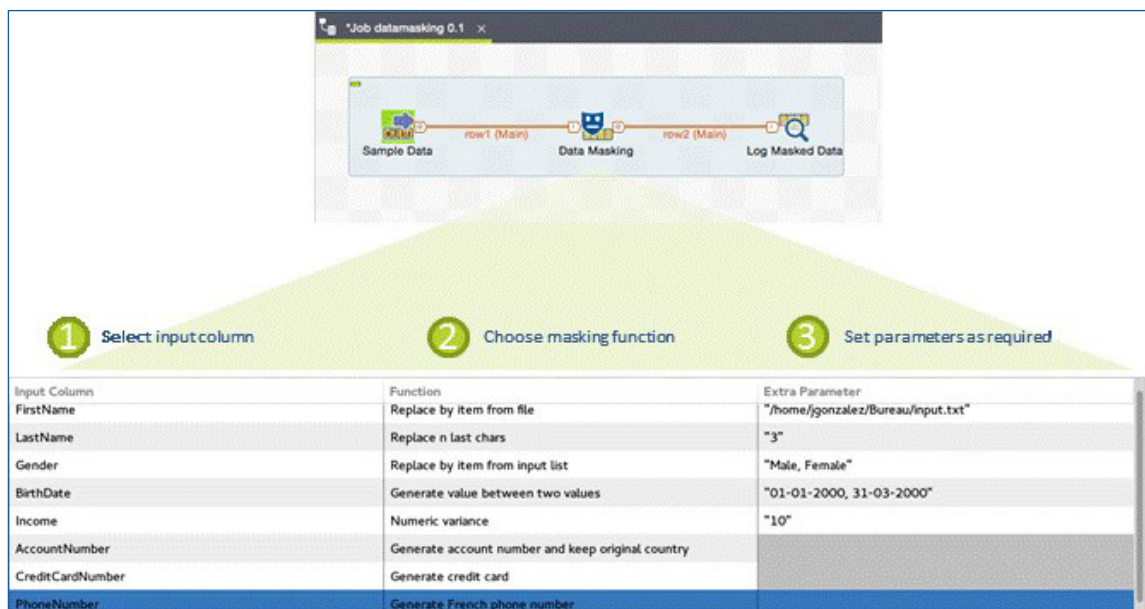


Figura 8: il mascheramento e il mescolamento (data shuffling) dei dati possono essere effettuati per flussi in modo batch o in tempo reale, con funzioni preconfigurate o personalizzate che si adattano ai dati personali più ricorrenti.

Con Talend Data Preparation, il mascheramento dei dati può inoltre essere realizzato ad hoc, permettendo agli utenti aziendali di proteggere i dati sensibili prima di condividerli con altri collaboratori. Si prenda l'esempio di un responsabile marketing che voglia effettuare un report sul successo di una campagna per poi comunicarlo a un partner. In un caso come questo, il responsabile marketing può condividere il set di dati per realizzare l'analisi dopo aver proceduto all'anonimizzazione di quei dati che potrebbero rivelare in modo inappropriato informazioni relative alla vita privata degli interessati. (figura 9).

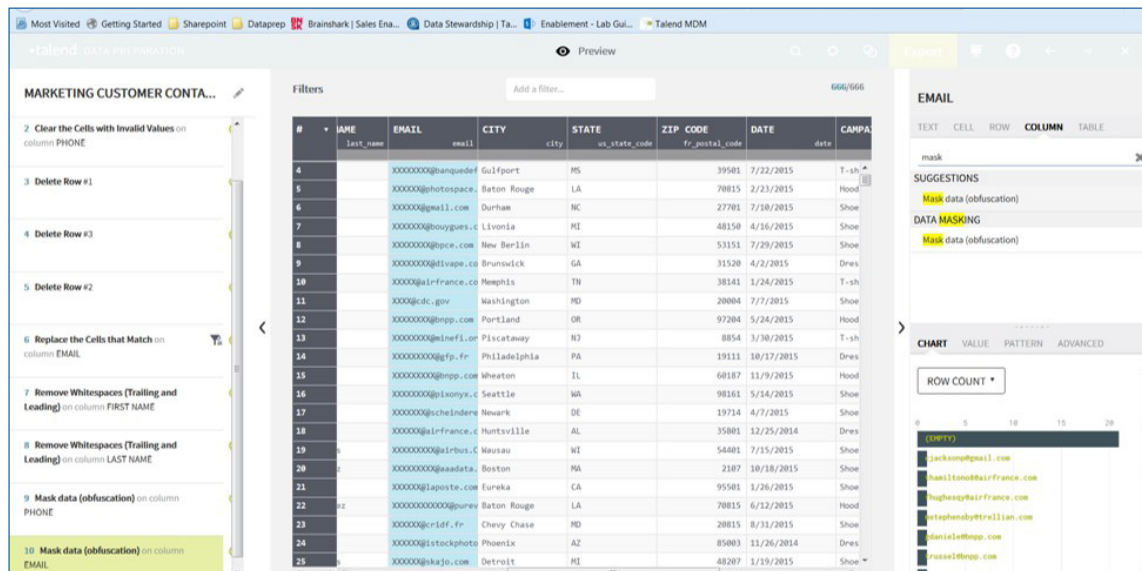


Figura 9: mascheramento dei dati self-service per gli utenti aziendali con Talend Data Preparation.

8. Effettuare valutazioni d'impatto sulla protezione dei dati

L'articolo 35 della GDPR concerne le valutazioni d'impatto sulla protezione dei dati. Tali valutazioni possono essere necessarie qualora il trattamento implichi nuove tipologie o categorie particolari di dati, quali le informazioni sull'origine etnica o razziale. I responsabili della governance dei dati devono stabilire dei controlli per fare in modo che le valutazioni d'impatto relative alle operazioni di trattamento che si prevede di effettuare siano condotte conformemente alle disposizioni della GDPR. Si prenda il caso di un dettagliante che desideri utilizzare un sistema di riconoscimento facciale nei punti vendita dei centri commerciali. Tale software si baserà sul sesso e sull'età approssimativa dei visitatori per consentire la personalizzazione delle offerte pubblicitarie. Una valutazione d'impatto sulla protezione dei dati potrà obbligare il dettagliante a eliminare qualsiasi informazione in suo possesso entro il minuto seguente l'apertura, da parte del cliente, del messaggio pubblicitario, ma anche ad astenersi dal collegare tali dati ai profili sui social network.

Talend Data Quality e Talend Metadata Manager possono catturare, scoprire ed effettuare la profilazione di nuovi set di dati e della relativa semantica in modo altamente automatizzato, quindi applicare tali regole di controllo su larga scala. Si tratta quindi di due strumenti che possono svolgere un ruolo attivo nella valutazione dell'impatto sulla protezione dei dati per qualsiasi tipo di sistema. Supponiamo che una società voglia condurre una valutazione di questo tipo sul data lake aziendale, che assorbe ingenti volumi di informazioni provenienti da dispositivi connessi. Talend Data Quality potrà aiutare ad effettuare tale valutazione identificando nel set di dati le informazioni a carattere personale, che dovranno quindi essere mascherate prima di essere immesse nel data lake.

9. Condurre valutazioni del rischio fornitori

L'articolo 28 comma 1 della GDPR tratta degli obblighi incombenti a un responsabile del trattamento terzo. Qualora un trattamento debba essere effettuato per conto del titolare del trattamento, quest'ultimo ricorre unicamente a responsabili del trattamento terzi che presentino garanzie sufficienti per mettere in atto misure tecniche e organizzative adeguate in modo tale che il trattamento soddisfi i requisiti del regolamento. L'articolo 28(2) stabilisce inoltre che il responsabile del trattamento non possa ricorrere a un altro responsabile senza previa autorizzazione scritta, specifica o generale, del titolare del trattamento. Gli addetti alla governance dei dati devono assicurarsi che l'ufficio legale e il responsabile della conformità approvino la valutazione del rischio fornitori prima di condividere dati a carattere personali con questi ultimi. Se un fornitore condivide tali dati con dei subappaltatori a valle, valgono le stesse regole, quindi l'ufficio legale e la direzione della conformità dovranno convalidare anche le valutazioni dei rischi dei subappaltatori.

Come abbiamo visto al passo 8, le tecnologie Talend possono supportare la valutazione dei rischi in situazioni che comportano la raccolta o la condivisione dei dati con i fornitori. Ma le soluzioni Talend vanno oltre la fase della valutazione e applicano automaticamente e quotidianamente i controlli GDPR riguardanti lo scambio di dati. Talend Data Quality può incorporare controlli all'interno di un flusso di integrazione dati, mentre Talend Data Stewardship può delegare la responsabilità della data stewardship ai fornitori.

10. Migliorare la qualità dei dati

Il regolamento GDPR prevede un caso d'uso specifico riguardante la qualità dei dati. L'articolo 16 del GDPR esige che le aziende rettifichino i dati personali inesatti e integrino le informazioni mancanti senza ingiustificato ritardo. Il team della data governance deve quindi stabilire meccanismi per permettere ai clienti, ai dipendenti e agli altri interessati di richiedere la risoluzione tempestiva di qualsiasi problema di qualità dei dati riguardante le loro informazioni personali. La riconciliazione dei dati è un altro fattore chiave, poiché le informazioni relative ai clienti e ai collaboratori aziendali possono essere frammentate in varie applicazioni aziendali. La GDPR stabilisce che queste informazioni siano riconciliate in una vista completa e coerente tale da poter essere comunicata agli interessati che ne fanno richiesta.

Talend Data Quality genera un codice nativo per l'esecuzione dei controlli sulla qualità dei dati e l'anonimizzazione nel momento opportuno e al posto giusto, in locale in un cluster Hadoop o nel Cloud, su dati statici o su dati in streaming (figura 10). Fornisce inoltre sofisticate funzionalità per la deduplicazione dei dati e l'identificazione di corrispondenze (data matching), consentendo così di riconciliare o collegare dei set di dati attraverso i sistemi.

The Talend logo, consisting of the word "talend" in a lowercase, sans-serif font, is positioned on the right side of a dark blue horizontal bar. To the left of the text, there are several overlapping circles and lines in a lighter blue color, creating a network-like pattern.

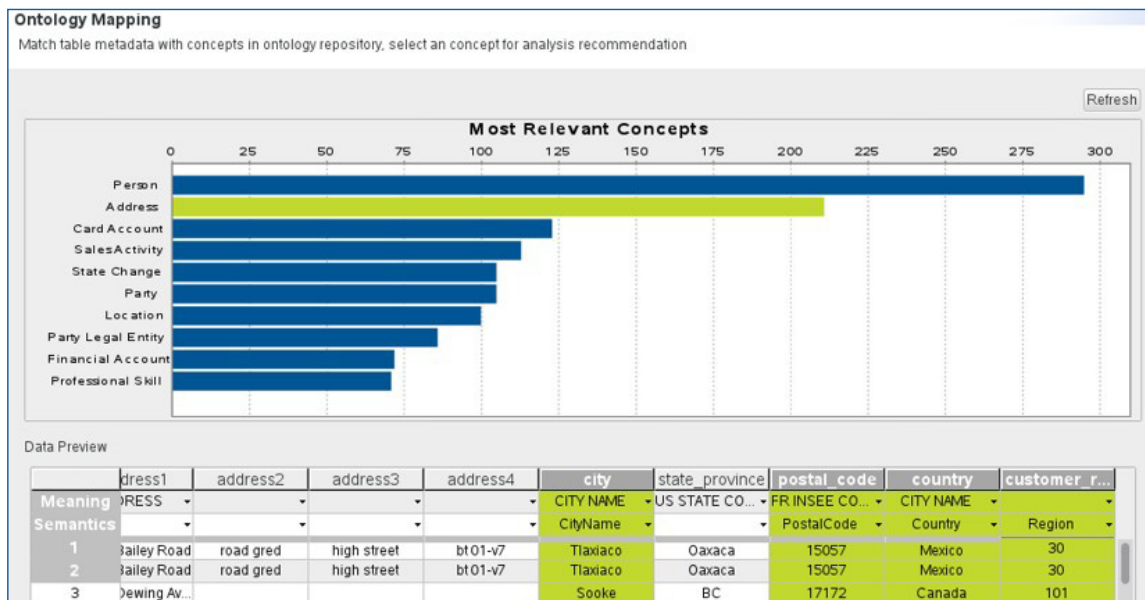


Figura 10: Talend Data Quality può abbinare automaticamente i dati personali alle nuove fonti di dati basandosi su modelli, dizionari e ontologie, quindi etichettare o applicare regole ai dati evidenziati.

Le organizzazioni dovrebbero inoltre spostare la responsabilità della protezione dei dati dai Data Protection Officer ai data steward o agli utenti aziendali. Infatti, un commerciale potrebbe essere la persona più indicata per verificare se i dati di contatto dei suoi clienti sono aggiornati. A un responsabile marketing, invece, potrebbe competere il compito di assicurarsi che un sistema per la gestione dei consensi sia stato predisposto nella base dati. Al fine di garantire che tutti i collaboratori in azienda possano gestire i dati con i quali lavorano nel rispetto delle norme della GDPR, le organizzazioni dovranno dotare i diversi servizi di applicazioni basate sul workflow, come Talend Data Preparation e Talend Data Stewardship, e dar loro più autonomia, senza tuttavia mettere a rischio i dati (figura 11).

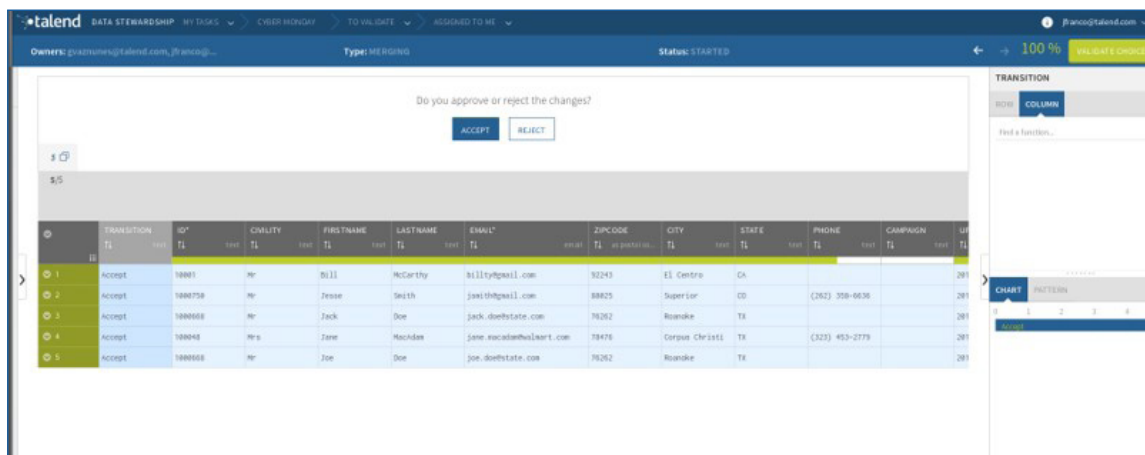


Figura 11: Talend Data Stewardship permette di dirigere workflow di data stewardship e di delegare le attività potenzialmente a tutti i collaboratori in azienda.

11. Tracciare un chiaro percorso dei dati (Data Lineage)

L'articolo 30 della GDPR esige che le aziende tengano un registro delle attività di trattamento svolte. Tale registro deve comportare una descrizione delle categorie di dati personali; una descrizione delle categorie di destinatari a cui i dati personali sono stati o saranno comunicati, compresi i destinatari di paesi terzi o di organizzazioni internazionali; i trasferimenti verso un paese terzo o un'organizzazione internazionale. Questi obblighi si estendono anche ai fornitori che effettuano il trattamento per conto di un'organizzazione. Per rispettare quest'articolo, i responsabili della governance dei dati devono rafforzare la gestione dei metadati e le funzioni di tracciabilità (data lineage).

Talend Metadata Manager prende in carico tali funzioni attraverso numerose piattaforme tra cui business intelligence, Hadoop, NoSQL e ETL (figura 12). Talend Big Data Platform si integra inoltre con Apache Atlas e Cloudera Navigator per fornire la tracciabilità dei flussi di dati all'interno di un data lake.

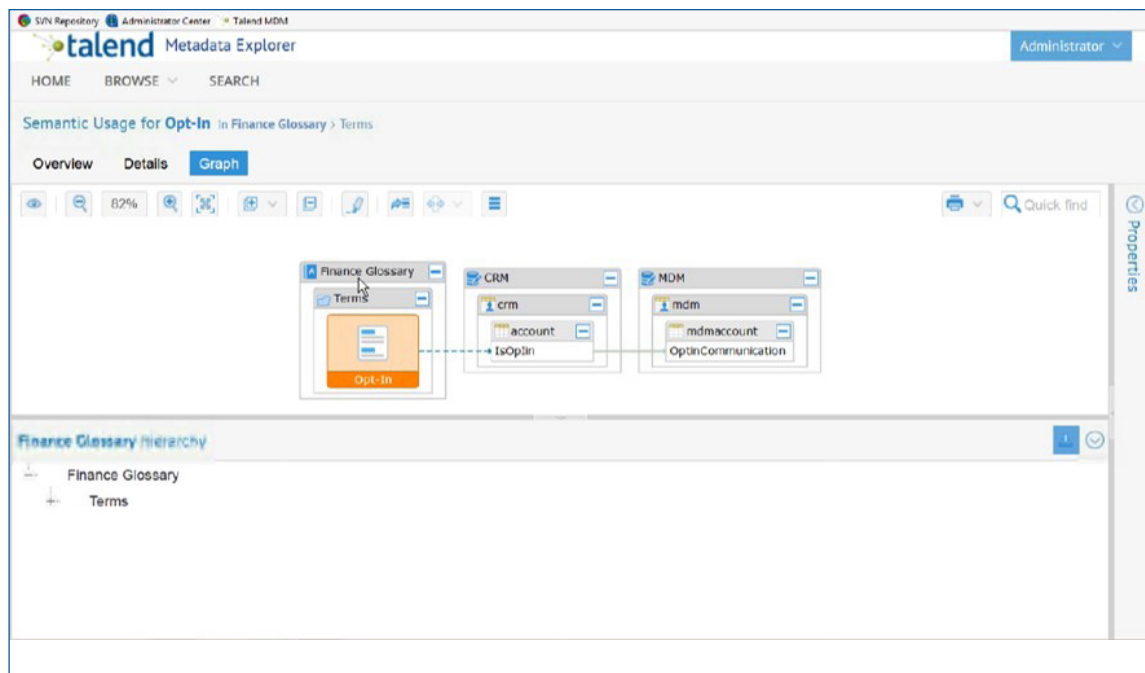


Figura 12: Talend Metadata Manager presenta una vista end to end dei dati critici quali gli opt-in, per capire da dove vengono e dove vanno i dati.

12. Controllare i modelli analitici

L'articolo 22 della GDPR tratta del processo decisionale automatizzato relativo alle persone fisiche, compresa la profilazione. In base a tale articolo, l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona. Le aziende devono stabilire una strategia di governance per quanto concerne i loro modelli analitici, che tengano conto dei livelli di propensione al rischio e al marketing.

Consideriamo per esempio una banca che utilizza dei codici postali nei suoi modelli analitici per proporre offerte di credito ai clienti. Servirsi di tali codici postali può dare adito a risultati disparati qualora la banca escluda certi codici postali corrispondenti a zone dove le minoranze sono preponderanti¹. Occorre che i responsabili della data governance stabiliscano dei controlli in modo tale che tutti i modelli ottengano l'autorizzazione dell'ufficio legale e del responsabile della conformità prima di essere messi in produzione.

¹ Federal Trade Commission. Big Data: A Tool for Inclusion or Exclusion? (Gennaio 2016).

<https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>.

Questi modelli analitici dovrebbero essere regolati in modo simile ai CDE. Come minimo, gli addetti alla governance dei dati possono costruire un inventario grazie a Talend Metadata Manager, comprendente il nome del modello, il proprietario, le variabili di input et di output, la metodologia del modello, la data di creazione e una prova che attesti l'approvazione dell'ufficio legale e del responsabile della conformità.

13. Gestire l'End User Computing (EUC)

L'articolo 32 della GDPR concerne la sicurezza del trattamento dei dati personali. Le organizzazioni generano un importante volume di applicazioni EUC, quali fogli di calcolo e data base che possono essere memorizzati sui PC o in archivi Microsoft® SharePoint®. Tipicamente, le applicazioni EUC sfuggono al controllo dell'IT e contengono spesso dati personali sensibili.

Esistono due modi per riprendere in mano la situazione. Il primo consiste nel fornire a coloro che devono lavorare con dati personali un'opzione migliore rispetto alle soluzioni di burocratica. Infatti, secondo un sondaggio realizzato da IDC², circa l'8% dei collaboratori aziendali effettua operazioni di preparazione dei dati in self-service senza alcuna supervisione tramite vari spreadsheet, l'integrazione dei dati essendo il caso d'uso più frequente (60%). Talend Data Preparation è un tool molto più efficace che aiuta gli utenti aziendali a realizzare la preparazione dei dati, anche quando comporta dati personali. Ma, cosa ancora più importante, è in grado di seguire e tracciare le attività associate catturando automaticamente ogni informazione individuale identificabile all'interno della sorgente dati, nonché le azioni portate a termine su tali dati dagli utenti aziendali (figura 13).

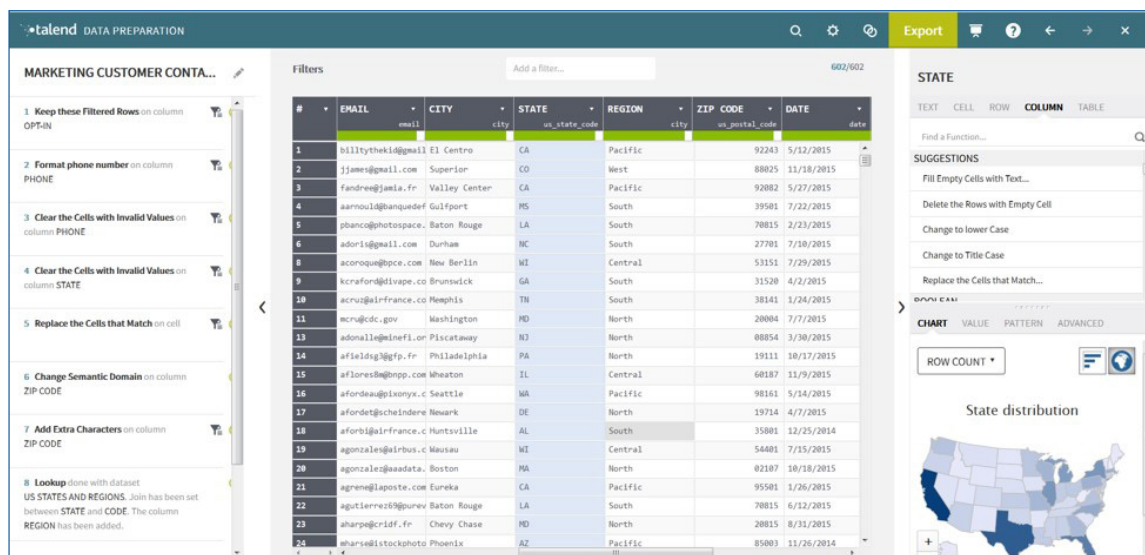


Figura 13: Talend Data Preparation fornisce migliori opzioni self-service rispetto a Microsoft Excel e altre applicazioni di burocratica per la preparazione dei dati e per il monitoraggio delle attività svolte dagli utenti, garantendo così un controllo e una riutilizzabilità più sicuri.

Il secondo approccio consiste nell'acquisire dei dati personali a partire da un testo in formato libero dopo averlo raccolto, quindi ad adottare le misure appropriate quali la categorizzazione e il mascheramento. Talend Data Quality offre funzionalità per il trattamento del linguaggio naturale per estrarre e etichettare dati potenzialmente personali a partire dal testo libero di un documento (email, file Microsoft Word o PDF), da un'applicazione (come Salesforce, Zendesk o Service Now), o ancora dal contenuto generato dagli utenti di un'applicazione Web o mobile (social network, fori di discussione). Grazie a funzionalità di machine learning sofisticate, eseguibili su vasta scala grazie ad Apache Spark, i riferimenti potenziali a dati personali in formato di testo libero possono essere automaticamente etichettati, quindi elaborati ulteriormente.

² The State of Self-Service Data Preparation and Analysis Using Spreadsheets, dicembre 2016.

14. Gestire il ciclo di vita dell'informazione

L'articolo 17 del GDPR tratta del diritto alla cancellazione, altrimenti detto "diritto all'oblio". L'interessato ha il diritto di ottenere dal titolare del trattamento la cancellazione dei dati personali che lo riguardano senza ingiustificato ritardo, e ciò in numerose situazioni, compreso quando i dati personali non sono più necessari rispetto alle finalità per le quali sono stati raccolti.

I responsabili della governance dei dati devono stabilire procedure che permettano agli interessati di richiedere la cancellazione dei loro dati. Devono inoltre mettere in atto dei controlli operativi in modo tale che tali domande vengano trattate in modo tempestivo. Per esempio, un'azienda potrebbe proporre sul suo sito web un servizio di assistenza al trattamento dei dati personali, attivato grazie a Talend Data Services. Più generalmente, Talend Data Services indica i punti di accesso ai dati, permettendo così di rispettare i diritti degli interessati, ivi compresi il diritto di accesso (articolo 15), il diritto di rettifica (articolo 16), il diritto all'oblio (articolo 17). Grazie a Talend Data Integration (figura 14), i dati possono essere automaticamente estratti in modo sicuro in un formato leggibile da dispositivo automatico, conformemente al diritto alla portabilità dei dati (articolo 20).

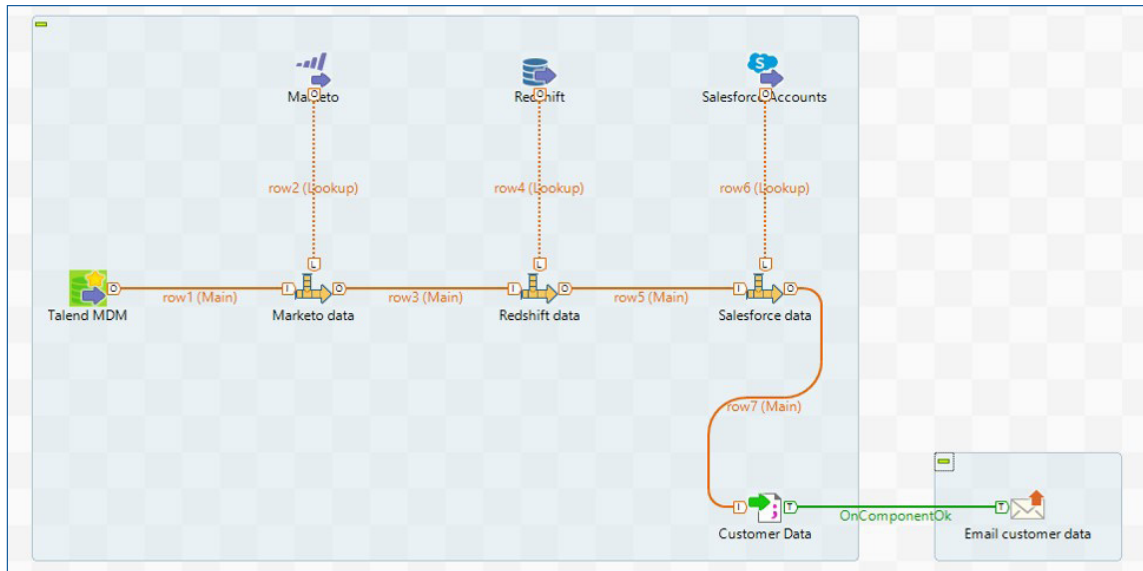


Figura 14: rispetto del diritto alla portabilità dei dati con Talend Data Integration.

Talend Metadata Manager fornisce un archivio che svolge un ruolo chiave per identificare rapidamente tutti i luoghi dove si trovano le informazioni riguardanti una data persona nelle applicazioni dell'organizzazione. Mappare la posizione di ogni elemento sensibile nei sistemi sorgente è cruciale. Precisamente, la sfida risiede nel fatto che le aziende conoscano generalmente i loro clienti o collaboratori in più contesti. Una compagnia aerea può aver identificato un cliente come passeggero, come frequent flyer nell'ambito del suo programma di fidelizzazione o ancora tramite il suo conto Twitter. Talend Big Data e Talend MDM integrano un componente nativo per la gestione della qualità dei dati che permette di far corrispondere dati eterogenei. Ciò consentirà all'azienda di capire che John Smith è la stessa persona che si cela dietro l'indirizzo email jsmith@widgets.com e lo pseudonimo @JohnSmith. Dopo aver riconciliato i dati nel flusso unificato, Talend Metadata Manager visualizza una vista end to end della catena di approvvigionamento informatica (figura 15) e l'impiega come base per assegnare i relativi ruoli e responsabilità per la governance e la data stewardship.

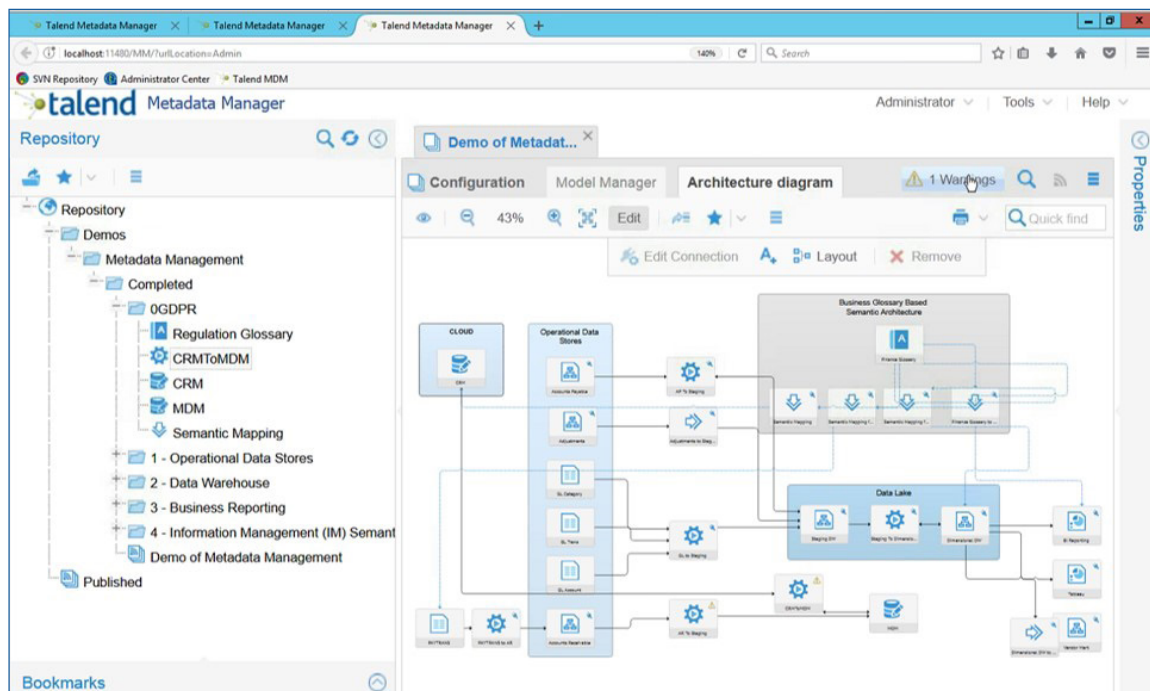


Figura 15: creazione di un archivio di dati per soddisfare ai requisiti del regolamento con Talend Metadata Manager.

15. Concludere un accordo di condivisione dei dati

L'articolo 28 (3) della GDPR asserisce che i trattamenti da parte di un responsabile del trattamento sono disciplinati da un contratto o da altro atto giuridico a norma del diritto dell'Unione o degli Stati membri, che vincoli il responsabile del trattamento al titolare del trattamento e che stipuli la materia disciplinata e la durata del trattamento, la natura e la finalità del trattamento, il tipo di dati personali e le categorie di interessati, gli obblighi e i diritti del titolare del trattamento. L'articolo 28 (4) esige che, quando un responsabile del trattamento ricorre a un altro responsabile del trattamento per l'esecuzione di specifiche attività di trattamento per conto del titolare del trattamento, su tale altro responsabile del trattamento sono imposti, mediante un contratto o un altro atto giuridico a norma del diritto dell'Unione o degli Stati membri, gli stessi obblighi. Tali obblighi possono essere riuniti in un accordo di condivisione dei dati.

I responsabili della governance dei dati devono assicurarsi che l'ufficio legale e il responsabile della conformità convalidino tali accordi prima che venga effettuato qualsiasi trasferimento di dati personali dall'UE verso un paese non UE, da un'organizzazione all'interno dell'UE verso un subappaltatore, e da un subappaltatore verso un altro subappaltatore a valle. Gli accordi di condivisione dei dati possono essere integrati semanticamente in Talend Metadata Manager.

16. Imporre l'applicazione delle norme con i controlli GDPR

I controlli GDPR sono un elemento primordiale delle misure volte a garantire la conformità regolamentare. Al passo 1 abbiamo visto come i responsabili della governance dei dati devono stabilire un quadro per tali controlli, che possono riguardare i consensi e un registro delle attività di trattamento dei dati. In seguito, abbiamo illustrato più dettagliatamente i diversi aspetti da tenere in considerazione a livello di governance dei dati e come implementare i necessari controlli e le piste di audit.

L'ultimo passo consiste nel rendere operazionali tali controlli e a mantenerli in modo durevole via via che l'insieme dei dati e/o le norme evolvono. Imporre l'applicazione delle disposizioni regolamentari e dei controlli richiede un impegno congiunto che i responsabili della governance dei dati devono dirigere in collaborazione con l'ufficio legale e il responsabile della conformità e del rispetto della privacy.

Riepilogo

La GDPR entrerà in vigore il 25 maggio 2018 e si applicherà al trattamento dei dati personali di tutti i residenti nell'Unione europea, compresi i clienti, i collaboratori e i prospect delle aziende. Questo libro bianco descrive 16 passi pratici per l'attuazione di una strategia di governance dei dati conforme al GDPR con la piattaforma Talend.

Spetta alle aziende pilotare azioni congiunte in collaborazione con il loro ufficio legale, il responsabile della conformità e della privacy e gli addetti alla gestione dei dati per riutilizzare i modelli esistenti di governance dei dati in modo tale da sostenere il rispetto delle disposizioni. In particolare, occorre che le organizzazioni definiscano quali sono gli elementi relativi ai "dati personali" con l'obiettivo di conformarsi al GDPR e mappare tali attributi alle applicazioni nell'archivio dei metadati.

La GDPR è l'esempio a noi più vicino di regolamento sul controllo dei dati che si estende ad altre giurisdizioni. Altre normative simili comprendono il Privacy Act in Australia, l'Anti-Spam Law in Canada, la Cybersecurity Law in Cina e la Personal Data Localization Law in Russia. Le migliori pratiche di governance dei dati descritte in questo libro bianco rispettano anche l'osservanza di tali regolamenti.

A proposito degli autori

Sunil Soares è il fondatore e il direttore associato di Information Asset. Oltre alle sue funzioni di consulente presso numerosi Chief Data Officer, Sunil è anche l'autore di numerosi libri, tra i quali *Selling Information Governance to the Business*, *Big Data Governance*, *Data Governance Tools*, *The Chief Data Officer Handbook for Data Governance*, *Data Governance Compliance for BCBS 239 and DFAST* e *Data Sovereignty and Enterprise Data Management: Extending Beyond the European Union General Data Protection Regulation*. Prima di ricoprire queste cariche, Sunil ha diretto il dipartimento della governance delle informazioni di IBM.

Jean-Michel Franco è Senior Director of Product Marketing di Talend. Ha consacrato la sua carriera allo sviluppo e all'ampiamiento delle pratiche di gestione dei dati nelle aziende. Ha iniziato la sua esperienza professionale presso HP, dove ha messo a punto la procedura di business intelligence, per poi proseguire con SAP, di cui è stato direttore delle soluzioni marketing, e con Business & Decision, in qualità di Innovation Director. È autore di 4 libri, pubblica regolarmente articoli specializzati e presenzia a conferenze nell'ambito di eventi e saloni professionali. Potete seguirlo su Twitter: [@jmichel_franco](https://twitter.com/jmichel_franco).

A proposito di Information Asset e Talend

Information Asset fornisce servizi esperti di consulenza in materia di governance e gestione dei dati aziendali. Per maggiori informazioni: www.information-asset.com.

Talend è un leader di nuova generazione nel settore dei Big Data e delle soluzioni Cloud che aiuta le aziende a trasformare i loro dati in capitale strategico. La piattaforma di integrazione unificata di Talend, aperta ed evolutiva, permette di accedere, in tempo reale e a livello aziendale, a informazioni preziose sui clienti e sui partner nonché sul loro settore di attività. Talend ha per missione quella di offrire l'agilità dei dati necessaria per affrontare le esigenze in evoluzione delle imprese di oggi.

Per maggiori informazioni: www.talend.com.

Copyright © 2017 Information Asset, LLC, and Talend, Inc. Tutti i diritti riservati. QUESTO MATERIALE NON PUÒ ESSERE RIPRODOTTO, VISUALIZZATO, MODIFICATO O DISTRIBUITO SENZA PREVIA AUTORIZZAZIONE SCRITTA ESPRESSA DI INFORMATION ASSET, LLC.

I nomi di altri prodotti e società qui citati possono essere marchi di fabbrica o marchi registrati dei rispettivi proprietari.

Questo rapporto ha scopo unicamente informativo ed è fornito "tale e quale" senza garanzia alcuna, ivi compreso senza garanzia di commerciabilità né di attitudine per usi particolari e senza garanzia derivante da qualsiasi proposta, specifica o campione.

The Talend logo is displayed in white lowercase letters on a dark blue background. To the left of the logo, there is a decorative graphic consisting of several overlapping circles and lines in shades of blue and white, resembling a network or data flow diagram.