Share-Ing S.r.l.
Ingegneria dei Processi e del Software

Corso di Alta Formazione «Web Security and Privacy Officer»

Convegno «Cybersecurity e Data Protection»

Big Data, privacy by design and by default

Nicola Pagliarulo

Share-Ing S.r.l.

Università di Bologna - Palazzo Hercolani, Sala Poeti - Strada Maggiore n. 45, Bologna

Indice

Big Data

- Cosa sono i Big Data
- Qual è la reale novità dei Big Data
- Fasi di un progetto Big Data

«Privacy by design» e «Privacy by default»

- Cosa di intende per «Privacy by design» e per «Privacy by default»
- Fasi di un processo di “Privacy by design”
  - Analisi: definizione contesto e obiettivi del sistema per la riservatezza
  - Progettazione: utilizzo di Design Pattern
  - Progettazione: adozione di Design Strategies
  - Sviluppo: utilizzo di Privacy-Enhancing Technologies (PETs)

Big Data O “Privacy by design”, oggi

- Alcuni esempi
- Alcuni problemi
- Alcune false soluzioni

Big Data E “Privacy by design”, domani?

Cosa sono i Big Data?

Sono la "versione attuale" di ciò per cui l'informatica ("trattamento automatico di informazioni") è nata: la gestione di grandi quantità di dati. E questo almeno dalla gestione dei dati del censimento della popolazione americana del 1880, da cui nacque IBM.

L'altro ieri di parlava di Sistemi di Supporto alla Decisione (*Decision Support Systems*, DSS), ieri si diceva *Business Intelligence* (BI), oggi si parla di ***Big Data***, di *Analytics*, di *Big Data Analytics*.

Al di là della moltitudine di nomi e di sigle, di prodotti e di slogan, si tratta sempre di ***Analisi dei Dati***, cioè di *quell'insieme di tecniche e di tecnologie che consente di rendere disponibile il patrimonio di conoscenza contenuto nei dati e che permette conseguentemente di prendere delle decisioni consapevoli*.

Qual è la reale novità dei Big Data?

La novità dei Big Data è caratterizzata comunemente con l'acronimo delle ***3V: Volume, Velocity, Variety***. In italiano: quantità, velocità [di produzione], varietà [di tipi e formati][dei dati].

- *Volume*. Gli strumenti di Internet, ed in particolare le piattaforme "social", la disponibilità di sensori a basso costo, generano una quantità di dati inaudita, prima inconcepibile.
- *Velocità*. Alla quantità dei dati si aggiunge la rapidità con la quale questi dati sono continuamente (o ineluttabilmente) prodotti.
- *Varietà*. È dato tutto ciò che è informazione: anagrafiche strutturate, output di sensori, testi, immagini, tutto ciò che può essere digitalizzato.

Qual è la reale novità dei Big Data?

Chiamiamo Big Data Analytics *l'insieme di tecniche e di tecnologie (per lo più in divenire) che permettono per gestire l'estrema numerosità e/o variabilità e/o varietà delle informazioni a disposizione*.

La principale tra le tecnologie abilitanti dei Big Data è quella del ***Cloud***, cioè della *distribuzione geografica su base planetaria delle risorse di immagazzinamento e di trattamento* (con tutta le complessità che ne deriva proprio rispetto al tema della Privacy).

Con una formula un po' semplicistica potremmo scrivere che:

Internet + Cloud = Big Data

Quello dei Big Data è un *campo ancora totalmente aperto*, sia per quanto riguarda le soluzioni tecniche da adottare, sia per quanto riguarda i problemi teorici da affrontare.

Fasi di un progetto "Big Data"

(e, più in generale, di un progetto di Analisi dei Dati)

Elenco delle *fasi successive di un progetto Big Data* (elenco puramente intuitivo e non rigoroso):

- *Raccolta* dei dati (Collection e, contestualmente, Pulizia dei dati, Cleaning)
- *Caricamento* dei dati nelle strutture (di qualsivoglia natura) predisposte per la loro archiviazione
- *Analisi* dei dati
- *Visualizzazione* dei risultati

Cosa si intende per "Privacy by design"?

REGOLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (***Regolamento generale sulla protezione dei dati***)

(***GDPR: General Data Protection Regulation***)

Articolo 25: Protezione dei dati fin dalla progettazione e protezione per impostazione predefinita

(Article 25: Data protection by design and by default)

1. Tenendo conto dello stato dell'arte e dei costi di attuazione, nonché della natura, dell'ambito di applicazione, del contesto e delle finalità del trattamento, come anche dei rischi aventi probabilità e gravità diverse per i diritti e le libertà delle persone fisiche costituiti dal trattamento, sia al momento di determinare i mezzi del trattamento sia all'atto del trattamento stesso[,] ***il titolare del trattamento mette in atto misure tecniche e organizzative adeguate,*** quali la pseudonimizzazione, ***volte ad attuare in modo efficace i principi di protezione dei dati***, quali la minimizzazione, e a integrare nel trattamento le necessarie garanzie al fine di soddisfare i requisiti del presente regolamento e tutelare i diritti degli interessati.

Cosa si intende per "Privacy by default"?

REGOLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (***Regolamento generale sulla protezione dei dati***)

(***GDPR: General Data Protection Regulation***)

Articolo 25: Protezione dei dati fin dalla progettazione e protezione per impostazione predefinita

(***Article 25: Data protection by design and by default***)

2. ***Il titolare del trattamento mette in atto misure tecniche e organizzative adeguate 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. In particolare, dette misure garantiscono che, per impostazione predefinita, non siano resi accessibili dati personali a un numero indefinito di persone fisiche senza l'intervento della persona fisica.

In altre parole: "Privacy by design" e "Privacy by default"

*Il principio della "**Privacy by design**" prevede di tenere conto delle esigenze relative alla tutela privacy fin dall'inizio della fase di progettazione dei sistemi di trattamento e per tutto il ciclo di vita del dato.*

*Il principio della "**Privacy by default**" prevede che i dati degli utenti siano già tutelati per impostazione predefinita dei sistemi di trattamento.*

Attività preliminare di ***Privacy Impact Assessment (PIA)***:

- *Identificazione degli attori* del sistema e confronto con gli attori del sistema
- *Identificazione dei rischi* (tenendo conto della percezione degli attori)
- *Identificazione delle soluzioni e formulazione di raccomandazioni*
- *Implementazione delle raccomandazioni*
- *Revisioni (reviews), verifiche (audit) e misure (measures)*

Progettazione: utilizzo di Design Pattern

Per semplicità di descrizione si può dire che un "***Design Pattern***" è *un modello di architettura del software, che risolve un problema di progettazione di ordine generale. Il progettista aggiunge poi al modello quelle caratteristiche necessarie per risolvere lo specifico obiettivo di progettazione*.

Un Design Pattern ha il vantaggio di semplificare la progettazione e lo svantaggio di vincolarla.

Esistono Design Pattern orientati alla Privacy come problema di progettazione di ordine generale (cfr http://privacypatterns.org).

Progettazione: adozione di Design Strategies

I Design Pattern sono "oggetti" molto tecnici: si tratta già di codice software. Ad un maggior livello di astrazione, e quindi di libertà del progettista, ci sono le ***Design Strategies*** che *non impongono una specifica struttura al sistema, ma vincolano la struttura al raggiungimento di determinati obiettivi*.

Progettazione: adozione di Design Strategies

Dal GDPR sono derivabili ***otto Privacy Design Strategies***.

1) Strategie orientate ai dati

Minimizzazione
L'obiettivo è quello di restringere al minimo possibile l'insieme dei dati personali da trattare.

Nascondimento
L'obiettivo è quello di impedire ogni visione integrale dei dati personali.

Separazione
L'obiettivo è quello di gestire i dati personali separandoli, partizionandoli, distribuendoli ed impedendo che le diverse partizioni possano essere ricongiunte.

Aggregazione
L'obiettivo è quello di trattare i dati personali al massimo livello di aggregazione possibile cioè con il minore grado di dettaglio possibile.

Progettazione: adozione di Design Strategies

2) Strategie orientate ai processi

Informazione
L'obiettivo è quello della trasparenza, cioè quello della massima disponibilità di informazioni sul trattamento di cui sono oggetto i dati personali.

Controllo
L'obiettivo è quello di fornire agli utenti la possibilità di intervenire sui propri dati (in lettura, aggiornamento e cancellazione).

Rafforzamento
L'obiettivo è quello di non limitarsi al rispetto formale degli obblighi di legge, ma di attuare strategie a rinforzo di tale rispetto.

Dimostrazione
L'obiettivo è quello di permettere al gestore del trattamento dei dati di dimostrare la correttezza e la completezza della sua gestione.

Sviluppo: utilizzo di PETs

Nella fase di sviluppo le ***Privacy-Enhancing Technologies (PETs)*** sono *tecniche intrinsecamente in grado di favorire la riservatezza dei dati personali* (quali, a titolo di esempio: crittografia, protocolli per la comunicazione anonima, credenziali basate su attributi, ecc.)

Big Data O "Privacy by design", oggi

Nella realizzazione di un progetto di "Analisi dei Dati" (alias Analytics) alcuni obiettivi sono costanti.

- Massimizzare la quantità di dati disponibili.
  - Per massimizzare la quantità di informazioni estraibili
  - Per individuare tendenze (trends) o ripetizioni (patterns)
- Integrare e collegare informazioni che provengono da fonti differenti
- Ricostruire o inferire informazioni su soggetti specifici
- Utilizzare e riutilizzare lo stesso insieme di dati per le più diverse finalità di analisi.

Anche guardando al di fuori dell'ambito dell'Analisi Dati, quello della System Integration (e con esso della Data Integration) è una delle attività nobili della progettazione informatica.

Big Data O "Privacy by design", oggi

Gli obiettivi di un progetto di "Analisi dei Dati" appaiono oggi come opposti agli obiettivi della "Privacy Design Strategy":

- Minimizzazione dei dati personali acquisiti
- Separare i dati personali
- Nascondere ed aggregare al massimo livello possibile i dati
- Informare circa l'uso fatto dei dati personali.

Alcuni esempi

A seguire alcuni esempi di applicazioni Big Data in cui il rispetto della Privacy «non è prioritario»:

- Qualsiasi forma di "Advertising" sul Web è il frutto del tracciamento delle operazioni dell'utente (in qualche caso tra miliardi di altri).
- La maggior parte delle aziende di assicurazioni utilizza diffusamente il tracciato delle "scatole nere" installate sulle automobili.
- Un sistema di analisi della spesa medica regionale avrebbe certamente come obiettivo l'identificazione di comportamenti (e soggetti) fraudolenti.
- En passant, è escluso dal perimetro del DGPR, ma anche senza Big Data, da almeno quindi anni sarebbe potenzialmente possibile condurre indagini molto analitiche sui comportamenti fiscali.

Alcuni problemi

GDPR non si applica solo ad aziende stabilite nell'Unione Europea, ma si applica anche ad Controllori, Processori, Terze Parti situati al di fuori dell'Unione Europea, ma che trattano dati personali di soggetti situati nell'Unione Europea.

In Europa, estremizzando, tutto ciò che non è esplicitamente permesso è vietato. Negli Stati Uniti , altresì estremizzando, tutto ciò che non è esplicitamente vietato è permesso.

In molte parti del mando il controllo puntuale e personale di tutto ciò avviene sul Web è parte integrante della politica interna ed estera.

L'armonizzazione della normativa europea con prassi molto diverse sarà certamente molto difficile.

Alcuni problemi

GDPR non si applica solo ad aziende stabilite nell'Unione Europea, ma si applica anche ad Controllori, Processori, Terze Parti situati al di fuori dell'Unione Europea, ma che trattano dati personali di soggetti situati nell'Unione Europea.

Alcune tra le aziende private di maggior valore a livello mondiale traggono il loro valore proprio dalla disponibilità e sfruttamento di informazioni personali, spesso "liberamente" rese disponibili dagli utenti.

Sembra piuttosto improbabile una volontaria rinuncia alla fonte del proprio valore, né l'apparato sanzionatorio (ammesso che sia opportuno) ha molte possibilità d'essere efficace.

Alcuni problemi

Oggi i sistemi informatici sono considerati alla stregua di "utility" aziendali ed i fornitori dei sistemi informatici sono conseguentemente retribuiti. Appare molto improbabile che la "compliance" dei sistemi informativi alle norme del GDPR porti con sé il riconoscimento di un maggior valore economico, e questo a fronte di un maggior lavoro richiesto, con la conseguenza di una redditività ulteriormente ridotta.

Alcune false soluzioni

- Proliferazione di moduli e documenti di "consenso informato" di improbabile lettura, che, se pure risolvessero un problema formale, non avrebbero relazione con l'obiettivo sostanziale.
- Certificazioni e certificazioni di certificazioni che siano solo fini a se stesse e che suggeriscano agli utenti una sicurezza più o meno veritiera circa il trattamento dei loro dati personali.

Big Data E "Privacy by design", domani?

- Il tema della Riservatezza del dato (che potrebbe interessare poco le aziende) va di pari passo con il tema della Sicurezza del dato (che interessa sempre molto le aziende); la grande attenzione al secondo tema non può che avere positive ricadute sul primo.
- Alcuni grandi fornitori di tecnologia hanno già iniziato a proporre "soluzioni" di gestione dei dati compliant con le esigenze del GDPR. Questo implica che nel medio termine i progettisti informatici potranno disporre di Privacy-Enhancing Technologies "a scaffale", predefinite, senza necessità di onerosa iper-progettazione.
- Esistono già embrionali soluzione di criptaggio per il Cloud.

Big Data E "Privacy by design", domani?

- Le tecniche di anonimazione del dato più semplici sono già utilizzate nel caso "Big Data Analysis" di dati medici o genetici.
- Il processo reso necessario dalla "Privacy by design":
  - valutazione
  - attuazione (anche attraverso PETs)
  - monitoraggio

  è conforme e sovrapponibile a qualsiasi processo di Ingegneria del Software e quindi è facilmente integrabile nelle attività di Gestione di un Progetto informatico.
- Un obbligo normativo, se attuato con sensatezza operativa (più che con minacce di sanzioni da "grida manzoniana"), non può che essere un forte stimolo.

SHARE-ING
Ingegneria dei processi e del software

Grazie della vostra pazienza.

Share-Ing S.r.l.
Via Giulio Caccini, 1 – 00198 Roma
www.share-ing.eu

Nicola Pagliarulo
Fondatore e Responsabile operativo
320 0406975
nicola.pagliarulo@share-ing.eu