La gestione dei contenuti e la nuova normativa europea sulla data privacy

General Data Protection Regulation (GDPR)

Giorgio Anselmi, IBM Client Technical Manager - Cloud Content Services
Giulia Caliari, IBM - IT Security Architect

Roma, 22 Giugno 2017
Clients are responsible for ensuring their own compliance with various laws and regulations, including the European Union General Data Protection Regulation.

Clients are solely responsibility for obtaining advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulations that may affect the clients’ business and any actions the clients may need to take to comply with such laws and regulations.

The products, services, and other capabilities described herein are not suitable for all client situations and may have restricted availability.

IBM does not provide legal, accounting or auditing advice or represent or warrant that its services or products will ensure that clients are in compliance with any law or regulation.
The General Data Protection Regulation (GDPR) was published on 4 May 2016, and will be immediately applicable after a 2 year transition period on 25 May 2018 to any organisation which operates in the EU market.

- Introduces cross-industry 72H breach reporting to regulators and without undue delay to individuals with associated risk of severe reputational harm.

- Non-compliance has the potential to lead to huge fines of up to €20m or 4% of total annual worldwide turnover, so now is the time to build on the foundations you already have to ensure you Protect, Govern and Know Your Data.
Alcuni esempi di dati personali

- **Dati Demografici**
  - Nome
  - Gender
  - Data di Nascita
  - Età
  - Nazionalità

- **Canali**
  - Numero di telefono
  - Indirizzo Postale
    - Città
    - Provincia
    - Codice postale
    - Nome strada e civico
  - Indirizzo Email

- **ID Nazionali**
  - Carta Identità
  - Codice Fiscale
  - Passaporto
  - Targa Automobilistica
  - Patente

- **Organizzazione**
  - Nome
  - Forma legale

- **Conti finanziari**
  - IBAN
  - BIC

- **Numeri delle carte**
  - Bancomat
  - American Express
  - VISA
  - MasterCard
  - Dinners Club
  - Discover
  - JCB
  - ...

- **Social Media**
  - URL FaceBook
  - URL Linkedin
  - URL Pinterest
  - URL Instagram

- **Dati Sensibili**
  - Salute
  - Politico
  - Religioso
  - Filosofico
  - Genetico
  - Biometrico
  - Razza
  - Etnia

- **Identificativi Digitali**
  - IP Address (V4, V6)
  - MAC Address
  - X/Y Coordinate Geografiche
GDPR: 5 main obligations categories

**Data Protection By Design and By Default** Art 25
- «... the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures ...»
- A) for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed; personal data are not made accessible without the individual’s intervention to an indefinite number of natural persons ..»
- such measures shall ensure that by default personal data are not made accessible without the individual’s intervention to an indefinite number of natural persons

**Accountability** Art. 5, 24 and others
- Data controller is responsible ... and be able to demonstrate compliance with... «(accountability)»
- Controllers and Processors have the duty to prove the compliance with the GDPR principles, and therefore the obligation to trace processing activities and lawfulness, the gathering of notice and consent, the activities to manage, the security measures implemented, the access, etc..
- Obligation for “Records of processing activities” (Art. 30) Registri dei Trattamenti
- Privacy Impact assessment (Art. 35)

**Lawfulness and Consent** (Art 5-8)
- Personal data are: processed legitimately, correctly and transparently... ; gathered for purposes well-defined, evident and legitimate...; appropriate, relevant and limited to the need ...; stored for a period to guarantee the appropriate security;
- Lawfulness (Art 6) Liceità del Trattamento
- Consent (Art. 7 e 8)

**EU citizen rights** Art. 12 a 20 and others
- Transparent information
- Rights to access, rectification and erasure (“right to be forgotten”)
- Rights to restriction of processing
- Right to data portability
- Right to object and automated individual decision-making (“No Profiling”)

**Security of personal data**
Art. 5, 24, da 32 a 34
- ... appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality')
- ... appropriate technical and organisational measures... to ensure a level of security appropriate to the risk ...
- the pseudonymisation and encryption of personal data;
- the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;
- the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident..
- ... in particular of the risks...from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed
- Notification of a personal data breach to the supervisory authority (Art 33)
- Communication of a personal data breach to the data subject (Art 34)
GDPR Solution Framework – IBM Technology

Dynamic Policy Management:
- Define what, why, how long

Implementation Services:
- Distribute policies to data sources

Data Infrastructure:
- Control use, align cost to value

Data Management

Policies, Rules, Audit
Processes, Analyses

Security & Compliance Monitoring

Rights of EU Data Subjects
Security of Personal Data
Lawfulness and Consent
Accountability of Compliance
Design and Default

- Databases & Data Warehouse
- ECM & Collaboration
- Archive Platform
- Hadoop Platform
- Master Data
- Email Servers
- User Devices & File Shares
- Cloud & Social
GDPR – Approccio IBM Analytics

IBM GDPR approach:
- Assess
- Design
- Transform
- Operate
- Conform

Quick Win Analytics for the IT department:
- **Information Analysis**
  - Discover Structured and Non Structured citizens data in your IT environment
  - Discover current role and responsibilities
- **Short term actions**
  - Identify the quick actions that allow to foster GDPR compliance
  - Implement the actions
- **Information Governance**
  - Support the GDPR overall policies based on knowledge acquired through the quick win outcome

Support to GDPR workstream (example):
- **GDPR Work-stream**
  - 360° view of citizen (e.g. right to erase)
- **Data quality**
- **Privacy document management**

IBM Analytics Technology:
- MDM
- StoredIQ
- MDM, Infosphere Quality Stage
- Case Manager

IBM supporting technology:
- StoredIQ
- Information Analyzer
- IBM Optim Suite
- Information governance catalogue
An initial approach to data management for GDPR

1) Understand and Analyze Data & Documents containing Personal Information

2) Add additional insight into data sources through automated analysis

3) Define Policies and Rules required by GDPR to correctly manage Personal Information (what, why, how long)

4) ACTS: move, delete, archive, classify, masking….
Information Analyzer
Govern Data Wherever It Lives - Discovery on Structured Data

Content-based database analysis
- Data profiling
  - Domain, frequency and format analysis
  - Identification Keys
  - Identification of table relations
  - Data Classification (i.e. PII)
- Shared Metadata generation
- Automatic Documentation

Business Rules: compliance monitoring
- Led development of the control rules
- Definition of control thresholds
- Trend Management
- Output control from the command line
- Automatic documentation

Role in the GDPR
- Discovery & classification
- Metadata generation and management
- Data quality
Information Governance Catalog
Policies & Rules definition, Audit and Personal Data Catalogue

Defining the terms and rules

Category: Costs
Term: Tax Expense
Description: Tax to be paid on Gross Income
Status: Current
Data Steward: Roger West
Related Term: Tax Liability
Related Asset: Income_tbl

Collaboration between business & technical teams

Source: Income_tbl (Oracle Table)
Derivation: SUM(TAXAMT)
Data Rule: TAXEXP > $1
ETL Job: TXExp
Business Term: Tax Expense

Role in the GDPR
• Personal Data Catalogue
• Data Ownership
• Data classification

• Audit & Compliance
• By design & by default
Master Data Management: the golden source of citizen data

Provide a single source of truth for citizen data and linkage to operational systems

Point of access and control for stewardship and to support specific GDPR requirements (i.e. Citizen Data Updating, Movement, Deletion, Privacy Preference management…)

Role in the GDPR
- Single source of truth
- Data quality
- Update, transfer, delete,…
InfoSphere Optim Solution for Data Privacy, Test Data Management and Archive

Identify Relevant & Sensitive Data
Find what data must be retained, protected or removed

Optimize Test Data
Automate and optimize the application test processes that rely on data to enable continuous testing & DevOps

Dispose of Unnecessary Data
Remove unnecessary data from critical transactional or analytics applications

Retain Essential Data
Historical inactive data is safely retained while easily accessible for reports and compliance

Protect Sensitive Data
Private Data: Customer IDs, credit cards and financial data are masked or redacted
Gestione Processi: IBM Case Manager

IBM Case Manager for Investigations supporta:
- **Processo decisionale informato** – utilizzando dati disponibili su diverse sorgenti
- **Indagini complete** - workflows e dati strutturati e non
- **Vista dettagliata a 360 gradi** – del’intero caso
- **Misure investigative avanzate**
- **Collaborazione ottimizzata e trasparente**

Coordina e gestisce i processi previsti dal GDPR, che richiedano la supervisione umana al fine di garantire la conformità alle policy, quali ad esempio:

- Accesso ai propri dati da parte del Cittadino
- Diritto all’oblio
- Trasferimento Dati
Come identificare i Dati Critici, quelli Sensibili o Regolati, presenti nei contenuti archiviati in azienda?

- Un esperto può trattare 120-200 documenti o ripulire circa 40 cartelle al giorno

- L'accuratezza “Umana” nella Categorizzazione è di circa il 60%

- Lo sapevate che: 100TB = 500.000.000 di Documenti oppure = 1 Miliardo di eMail

- In media i contenuti raddoppiarono ogni 2–3 anni
Dove e come li posso trovare?
Da dove iniziare?  Un po’ di pulizia!

- 20% - 40% ROT* Data
- 25% Business value: Archive records for value
- 5% Regulatory record keeping: Hold and dispose of data
- 1% Legal hold: Collect and retain evidence
- 69% Information without duty and value = unnecessary data

%’s based on CGOC Summit 2012 Survey
CGOC = Compliance, Governance and Oversight Council

(*ROT = Redundant Obsolete Trivial)
Data cleanup business case – un caso reale

Information Lifecycle Governance / Legacy Data Cleanup

Business Case – Legacy Data Cleanup

Estimated achievable cumulative storage savings of €9.7m over 3 years (2015-2017) or €20.2m over 5 years (2015-2019) based on information received by [blank], past projects and industry averages.

Calculation assumptions used:
1. Unstructured Data Volume by end of 2014 = 2.4Pb
2. Unstructured data is growing 15% year over year
3. Storage Cost per TB: 2'500€ per year, including HW & Services
4. Storage cost decreasing year over year 10%
5. 27% ROT Potential, with 5% follow on year reduction

Saving Potentials - Details:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2015 savings</td>
<td>2.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016 savings</td>
<td></td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017 savings</td>
<td></td>
<td></td>
<td>3.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018 savings</td>
<td></td>
<td></td>
<td></td>
<td>3.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019 savings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>2020 savings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.1</td>
</tr>
<tr>
<td>Total Annual Storage Savings in M</td>
<td>2.5</td>
<td>2.5</td>
<td>3.2</td>
<td>3.9</td>
<td>4.5</td>
<td>1.1</td>
</tr>
<tr>
<td>Cumulative Storage Saving in M</td>
<td>2.1</td>
<td>4.5</td>
<td>7.8</td>
<td>11.7</td>
<td>16.2</td>
<td>21.3</td>
</tr>
</tbody>
</table>

Volume Savings over 3.2Pb in 3 Years

Note: A=Actual figures; P=Projected figures
E poi cerchiamo quello che ci interessa

- **20% - 40%** ROT* Data
  - (*) ROT = Redundant Obsolete Trivial

- **25%** Business value:
  - Archive records for value

- **5%** Regulatory record keeping:
  - Hold and dispose of data

- **1%** Legal hold:
  - Collect and retain evidence

- **69%** Information without duty and value = unnecessary data

%’s based on CGOC Summit 2012 Survey

CGOC = Compliance, Governance and Oversight Council
Rilevanti per l’azienda

StoredIQ – Come Funziona

Capire
- Data Assessment
- Indicizzazione “In-place” per comprendere il contenuto laddove si trova:
  - Data Visualization Maps
  - Content Mining

Ridurre
- Isolare i Non-Business Data

Identificare
- Classificare i Records
  - Metadata & Content ID
  - Contextual Analysis
  - Semantic Analysis
  - Bayesian Classifier
  - Based on Records Taxonomy

Agire
- Move, Copy,
- Protect
- Dispose

Rilevanti per l’azienda
StoredIQ
Come identifica i dati - gli Infoset
StoredIQ - La Tecnología

Application Server

Gateway Server

Distributed, Scalable Indexing Layer

Data Servers

- Distributed, Scalable Indexing Layer
- Application Gateway
- Data Sources

- Archive Platform
- ECM
- Forensic Images/Tapes
- File Servers
- Email Servers
- Desktops/Mobile
- SharePoint & Enterprise Collaboration
- Cloud
- Social Networks
- Media
Obiettivo, verificare la capacità dello strumento di:
• Lavorare nell’ambiente del cliente
• Indicizzare full text il contenuto dei documenti ed i metadati al fine di:
  • Identificare informazioni sensibili (i.e. GDPR)
  • Identificare i dati candidati alla cancellazione
**PoC Environment**

- Appliances StoredIQ configurate nell’infrastruttura VmVare
- Copia di dati significativi su di una share in rete
- ............. i data erano già stati “ripuliti” teoricamente “no ROT data”
- 1,15TB distributed in varie cartelle
WorkBench – PoC Data Map 2
PoC use case: Clean Up

Filtr basati sulla data di ultima modifica e data di ultimo accesso.
Overall Summary Table

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of data objects in the infoset</td>
<td>191445</td>
</tr>
<tr>
<td>Total size of data objects in the infoset (MB)</td>
<td>269002.04</td>
</tr>
</tbody>
</table>

**Distribution by Category**

- **WORDPROCESSING**
- **SYSTEM**
- **SPREADSHEET**
- **PROGRAMMING**
- **PRESENTATION**
- **OTHER**
- **MULTIMEDIA**
- **IMAGES**
- **EMAIL**
- **CONTAINER**
- **COLLABORATION**

**Example of Major Contributors**

- BIZ\EggenJur
- BIZ\Darg\Han
- BIZ\DraxSte
- BIZ\BozDen

**RISULTATO:**
34% del totale dei documenti non modificati negli ultimi 6 anni 66% di questi non sono stati acceduti da Luglio 2011.

Significa che almeno il 23% dei documenti non sono più stati utilizzati e sono ottimi candidati per l’archiviazione su nastro o la cancellazione.
PoC use case: eDiscovery 1

Ricercare dati sensibili PCI (Payment Card Industry)

Filtri basati su regular expression per cercare codici IBAN all'interno di documenti.
PoC use case: eDiscovery

StoredIQ presenta la lista dei documenti “incriminati”

<table>
<thead>
<tr>
<th>File name</th>
<th>File size</th>
<th>File path</th>
<th>Created</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auftragsbuch.accdb</td>
<td>668.55 KB</td>
<td>customs/1. IMPORT/Auftragsbuch</td>
<td>2016-08-03</td>
</tr>
<tr>
<td>Auftragsbuch_Backup_1.accdb</td>
<td>668.55 KB</td>
<td>customs/1. IMPORT/Auftragsbuch</td>
<td>2016-11-23</td>
</tr>
<tr>
<td>DEfekte alte jobbook.accdb</td>
<td>24.55 MB</td>
<td>customs/1. IMPORT/Auftragsbuch/Archiv</td>
<td>2013-11-04</td>
</tr>
<tr>
<td>Auftragsbuch_Backup.accdb</td>
<td>371.81 KB</td>
<td>customs/1. IMPORT/Auftragsbuch/Archiv</td>
<td>2016-01-26</td>
</tr>
<tr>
<td>Auftragsbuch20130925.accdb</td>
<td>12.48 MB</td>
<td>customs/1. IMPORT/Auftragsbuch/Prüfungen_Bereinigungen</td>
<td>2013-09-25</td>
</tr>
<tr>
<td>Jobbook_restriction.accdb</td>
<td>12.46 MB</td>
<td>customs/1. IMPORT/Auftragsbuch/Prüfungen_Bereinigungen</td>
<td>2013-09-25</td>
</tr>
<tr>
<td>ALT1_Auftragsbuch.accdb</td>
<td>771.55 KB</td>
<td>customs/1. IMPORT/Auftragsbuch/Archiv</td>
<td>2015-08-13</td>
</tr>
<tr>
<td>0030282.pdf</td>
<td>52.21 KB</td>
<td>customs/2. EXPORT/Archiv 2016/DEB/ABD</td>
<td>2016-01-27</td>
</tr>
<tr>
<td>0030268.pdf</td>
<td>81.18 KB</td>
<td>customs/2. EXPORT/Archiv 2016/DEB/ABD</td>
<td>2016-01-26</td>
</tr>
</tbody>
</table>
Applicando i filtri IBAN come “overlay” viene creata una heatmap per capire le aree di maggior esposizione.
Selezionando documenti a campione si può verificare la presenza

RISULTATO
Opertura di un documento con codice IBAN.
Ricercare informazioni personali come gli indirizzi.

Creato un filtro utilizzando il wizard StoredIQ per cercare nei fogli di calcolo parole con indirizzo e cliente.
Anche qui è stata verificata l'effettiva presenza