Why is the EU GDPR so Misunderstood and Unfamiliar to so Many?

Why Compliance isn't just reducing intrusions. What every DPO, CEO, CIO should know about GDPR

In a Dell Press Release "Dell announced results of a global survey on the European Union's new General Data Protection Regulation (GDPR).

- More than 80 percent of global respondents know few details or nothing about GDPR
- Less than one in three companies feel they are prepared for GDPR today
- 97 percent of companies don't have a plan to be ready for GDPR
- Only 9 percent of IT & business professionals are confident they will be fully ready for GDPR"

Big Data Revealed
RADAR FOR YOUR DATA LAKE

Methodology Page 2

"To give you some idea of the punitive impact this can have, in the UK in 2016, Tesco Bank experienced a data security breach that affected 9,000 of its customers. Had GDPR been in force when this breach took place, Tesco Bank would have been fined £1.9 billion ($2.3 billion)"cmswire

GDPR Key Points: 10 key facts businesses need to note about the GDPR – ComputerWeekly.com

1. **GDPR applies to all –** The GDPR applies to all companies worldwide that process personal data of European Union (EU) citizens. **See it in its entirety at** GDPR.Institute
2. **The GDPR widens the definition of personal data**
   The GDPR considers any data that can be used to identify an individual as personal data. It includes, for the first time, things such as genetic, mental, cultural, economic or social information.
3. **The GDPR tightens the rules for obtaining valid consent to using personal information**
   The GDPR requires all organizations collecting personal data to be able to prove clear and affirmative consent to process that data
4. **The GDPR makes the appointment of a DPO mandatory for certain organizations**
   The GDPR requires public authorities processing personal information to appoint a data protection officer (DPO)
5. **The GDPR introduces mandatory PIAs**
   The inclusion of mandatory privacy impact assessments (PIAs) in the GDPR is mainly due to the influence of the UK's Information Commissioner's Office, which has worked a lot with PIAs in the past. The GDPR requires data controllers to conduct PIAs where privacy breach risks are high to minimize risks to data subjects.
6. **The GDPR introduces a common data breach notification requirement**
   The GDPR harmonizes the various data breach notification laws in Europe and is aimed at ensuring organizations constantly monitor for breaches of personal data.
   The regulation requires organizations to notify the local data protection authority of a data breach within 72 hours of discovering it.
7. **The GDPR introduces the right to be forgotten**
   The GDPR introduces very restrictive, enforceable data handling principles.
   One of these is the data minimization principle that requires organizations not to hold data for any longer than absolutely necessary, and not to change the use of the data from the purpose for which it was originally collected, while – at the same time – they must delete any data at the request of the data subject.
8. **The GDPR expands liability beyond data controllers**
   In the past, only data controllers were considered responsible for data processing activities, but the GDPR extends liability to all organizations that touch personal data.
   The GDPR also covers any organization that provides data processing services to the data controller, which means that even organizations that are purely service providers that work with personal data will need to comply with rules such as data minimization.
9. **The GDPR requires privacy by design**
   The GDPR requires that privacy is included in systems and processes by design.
   This means that software, systems and processes must consider compliance with the principles of data protection. However, the proper erasure of information, for example, is not something often seen in software. But in the future, all software will be required to be capable of completely erasing data, which will be a challenge for a lot of software engineers.
10. **The GDPR introduces the concept of a one-stop shop**
    GDPR, which allows any European data protection authority to take action against organizations, regardless of where in the world the company is based. GDPR enforcement is also backed by significant fines of up to €20m or 4% of group annual global turnover.

About Methodology:

As you can see from the above statements, intrusion protection is not the main focus of GDPR regulations. They reach deeper into the true meaning of Data Governance and give the EU citizen the right to remain anonymous, if desired, even from individuals within your own organization.

How do you become compliant with EU GDPR and other Regulatory Agencies without disrupting or degrading your legacy systems or existing Hadoop Clusters? The days of using old generation Data Profiling and Discovery tools are not the desired approach for sensitive production systems and in many cases not even allowed.

BigDataRevealed (BDR) is a software product specifically designed to address the same issues that GDPR is now mandating. BDR identifies and encrypts Personal Data by using Pattern Recognition algorithms and features of our Intelligent Catalogue for maximum Personal Data discovery. BDR is written entirely within the Hadoop framework for enhanced efficiency and allows BDR to process streaming data.

BDR is able to use the massive processing power of Hadoop to discover exposed PII data in legacy systems by following the easy to use instructions attached to this document.

For a more in-depth description of BigDataRevealed's (BDR) features please read the following.

What I suggest is to create a BigDataRevealed Hadoop Cluster by exporting data to Hadoop using either sqoop or our Spark/Kafka connectors. Name the Folders and Files as closely to the names of the source files brought into Hadoop.

As data gets loaded into Hadoop;

1. Allow BDR to discover and build Catalog/Metadata from the Hadoop Files copied from your legacy data.
2. BDR will identify the most likely business columnar classifications and even give you the percentage of data in that column that matches a pattern. As an example, a column may contain the following:
   a. Email 85%
   b. Phone 10%
   c. Unknown 2%
3. Then Run detailed Discovery on the files that searches for Personally Identifiable Information using algorithms found in BDR's Collaborative Library. The following patterns represent a small sample of the many contained in BDR's Library:
   a. Names
   b. Addresses
   c. Zip Code
   d. Social Security
   e. Credit Card Info
   f. User-Defined / Collaborative
   g. And so on ….
4. Store the Folder/File/Column path and the Discovery Pattern Name where found
5. Use BDR's Drilling capabilities to view the source data and verify results.
6. BDR will allow history to be kept every time these processes are run along with time stamps. This will help identify when other jobs where run that may have contained PII and other sensitive data that subsequently entered your databases and file systems.

For Hadoop, BDR-**S**ecure**S**equester**E**ncryption**:**

1. If you wish to **secure Data within Hadoop**, follow the process described below to Secure and Sequester PII and other sensitive data:
    a. Create a safe and secure Encrypted Hadoop Cluster off the Grid, off of the Web, so that NO outside hacker can reach the data during this process. Or if protecting a current Hadoop Cluster, create a large Encrypted Zone and use this during the BDR-**S**ecure**S**equester**E**ncryption Processes.
    b. Copy the original files with Personally Identifiable Information into a Hadoop Encrypted Zone
    c. Instruct BDR to encrypt the PII and sensitive data and write a new safe, secure compliant file.
    d. BDR will Store the "AES 256" Encryption keys off the Grid and outside of the Edge Node and use User Role Based Authority to access those keys when needed.
    e. Delete the old files from HDFS and Hbase that contained PII and other sensitive data.
    f. Now you have preserved your Original Files in an Encrypted Zone if desired.
    g. And you have created new files with encrypted PII data while leaving other data in the files for Data Scientist and others to use.
    h. And you have deleted all files and iterations of files that contained exposed PII data from HDFS and Hbase.

    - This complete series of 1-6 above and 1 a-h can be continually processed at whatever latency you desire to maintain ongoing compliance.

2. **For Legacy Databases such as SQL, Mainframe, AS400, Documents and more …,**
   BDR-Secure**-S**equester**-E**ncrypt**:**

    a. Using the Intelligent-Catalog/Metadata created in steps 1-6 above and stored by BDR.
        a. Review each file processed by BDR in your Hadoop Cluster.
            i. BDR identifies each column containing PII or other sensitive data.
            ii. Identify what Files and columns PII or other sensitive data was found.
            iii. Backup and secure Original Files/Data.
            iv. Use SQL tools or queries to encrypt or remove PII or sensitive data from legacy system.

No matter to what degree you feel ready to pass GDPR audits, you can quickly and easily use BDR to verify your readiness and determine your real exposure. Just a few resources and several days could give that peace of mind you are looking for.

Email or call us and feel comfortable you are on your way to GDPR compliance. For more information about our discovery process and the **BDR**-**S**ecure**S**equester**E**ncryption process that is facilitated using our intelligent catalog and Metadata, contact us at info@bigdatarevealed.com or 847-791-7838.
https://vimeo.com/216064822 Video of recent Webinar