Data Processing Inventory

— a solution for compliance of GDPR Article 30

Carnegie Mellon University
Privacy Engineering Capstone Project, Fall 2017
Team members: Lidong Wei, Bill Quan, Jun Ma
Sponsor: Citi
Advisor: Nicolas Christin

The General Data Protection Regulation (GDPR) [1] is a regulation that requires organizations to protect the personal data and privacy of EU citizens for transactions that occur within EU member states. As one of the world's most expansive data privacy laws, it takes effect on May 25, 2018 [2]. The GDPR aims primarily to give control back to citizens and residents over their personal data and to simplify the regulatory environment for international business by unifying the regulation within the EU. And failure in compliance could cost companies up to 20 million Euro or 4 percent of global annual turnover [1].

Article 30 of GDPR [3], also named as Records of Data Processing Activities, requires both controller and processor to include detailed information of their profile as well as the processing activity. It particularly specifies that all the major organizations should maintain a record of data processing activities and provide to supervisory authority when necessary. Article 30 details what information should be included before a data processing activity can happen, and thus all the major organizations should seek to comply with it as data processing activities are happening every day.

This project is designed to organize the data processing activities of Citi. It includes the following phases as major components:

- Analysis of GDPR Article 30 requirements. Analysis of publicly available third-party descriptions of Article 30 compliance, and publicly available information on vendor tools designed to meet Article 30 compliance objectives.
- Design prototypes (including design work and coding) for areas related to GDPR Article 30 inventory of data processing inventories based on previous analysis. The working prototype needs to include any design work or coding for extracting, organizing and/or presenting the Article 30 data processing inventory.

The research phase includes the introduction of other general data protection regulations and the comparison between these regulations and GDPR. All the researched regulations have some similarities with GDPR to some extent, however, GDPR defines much more strict requirements that cover almost all the aspect of data's life cycle. Therefore, GDPR is a brand new regulation and requires organizations to pay much more attention to it than ever. Besides, the research work also looks for the publicly available third-party guidance. The Belgian Data

Protection Authority's guidance [4] stands out for providing a detailed explanation and clarification of the GDPR Article 30. The guidance and the template aim to assist data controllers and data processors in putting in place the records of processing activities as required by Article 30 of the GDPR. Last but not least, the research work also found some available vendor tools which are designed to meet Article 30 requirements. Among which, the Oracle's Audit Vault and Database Firewall [5], Veritas' Global Data Visibility [6], and OneTrust's Privacy Management System [7] are some example tools that we have looked into. However, those tools have some deficiencies when it comes into Citi's use case, such as the constraint of database, third-party data sharing and access control mechanism. To help Citi to comply with GDPR Article 30, it is necessary to develop a working prototype which specifically applicable to Citi's use case.

The development of prototype (include design work and coding) is for areas related to GDPR Article 30 inventory of data processing inventories [3], including how data can be collected, organized, presented and kept up to date to meet requirements. This work includes data management, data analytics and data visualization components and original, innovative thinking regarding how to setup and maintain a compliant Article 30 inventory. This project acts as a prototype of Article 30 data inventory for Citi internal use. It summarizes the requirements of Article 30 and leverages major features of publicly available tools. It can be used as an original tool to meet the Article 30 requirements and can be further extended to an integrated tool combining Citi internal techniques. It gives a good start for Citi to explore the approaches to manage data inventory to comply with Article 30.

The criteria of this prototype are:

- Compliance of GDPR Article 30: The most important criteria of this prototype, since its main function is for Citi to make sure all the data processing activities are compliant to GDPR Article 30.
- Data inventory management: Including retrieve, edit, delete, insert of a record through user interface.
- Data inventory export: Export the result of inventory search result to present to Data Protection Authority.
- Role-based access control [8]: Separate the privileges of different roles.
- User interface: A user friendly interface for user to interact with.
- Data statistics: Present the data processing history of a user, includes data visualization and detailed statistics.

To meet those criteria, this system leverages some developing technologies and mechanisms:

- This system supports import data from existing questionnaire and online form fill in. By setting some of the fields as required fields, a user must fill in some content before they can submit the form. Besides, there will be Data Protection Officers who are responsible for reviewing the submitted form and approve it accordingly.
- Users have the privileges of editing, deleting and retrieving the records that initiated by themselves.

- This system utilizes the exceljs package in NPM [9] library to achieve the import/export the records to/from the data inventory. The imported data will be automatically populated into corresponding fields and users can review the form before they submit it. As of the searching result from the data inventory, users can export into a formatted excel file and download it.
- There are three levels of user group: super admin, regional admin and normal users. Each level of user has different privileges and acts as different roles in the data processing activities. Specifically, super administrator is a global administrator in charge of manage the system. Regional administrators are the administrator role in their region. Normal user could be the branch privacy officer who is responsible for submitting a record when there is a data processing activity happening.
- This system uses Nodejs [10] as a web framework to present web service. It leverages the advantages of large NPM library to build an integrated system. With the deployment of this service, user can access the web page wherever there is an internet connect and modern web browser. To build a user-friendly graphic interface, this system uses Bootstrap [11] as the frontend library to present well-organized and pleasing web page. Each component on the web page is well arranged and easy for user to understand.
- The statistics of this user's data processing history are displayed in the dashboard page once the user login. This system uses Google chart API [12] to display live data on the web site and can connect to the data source and update the data in real time whenever the page is loaded.

This project is compelling because it is designed to organize the data processing activities of Citi organization and assist it in becoming compliant to Article 30 of the GDPR regulation which comes into effect on May 25th, 2018. This project allows the group to apply our technical skills and knowledge of privacy preserving techniques to a real-world setting. Our goal of this project is to help Citi to manage the data processing activities records efficiently and acts as a prototype of data processing inventory. As there could be more complicate cases for Citi's internal system, this prototype is just an entry point of how the data management system should work to comply with Article 30. More future works need to be finished if it is integrated into Citi's internal system.

[1] Michael Nadeau. General data protection regulation (gdpr) requirements, deadlines and facts. https://www.csoonline.com/article/3202771/data-protection/general-data-protection-regulation-gdpr-requirements-deadlines-and-facts.html, Nov 2017.
[2] EU gdpr portal. www.eugdpr.org.
[3] Nicholas Vollmer. Article 30eu gdpr "records of processing activities". https://www.privacy-regulation.eu/en/30.htm, Dec 2016.
[4] Matt Urglavitch. Belgian dpa guidance on gdpr article 30 records of processing requirements. https://onetrust.com/belgian-dpa-issues-guidance-article-30-records-processing-requirements, Aug2017.

[5] Oracle audit vault and database firewall. https://www.oracle.com/database/security/audit-vault-database-firewall/index.html.
[6] Veritas. 360 data management for gdpr. https://gdprtech.com/wp-content/uploads/2017/03/veritas-360dm-gdpr-brochure-en-public.pdf, Mar 2017.
[7] The ten step guide to meeting gdpr article 30 record keeping requirements. http://onetrust.com/pdf/20170203-10-Step-Guide-for-Data-Mapping.pdf, June 2017.
[8] Sandhu, Ravi S., et al. "Role-based access control models." Computer 29.2 (1996): 38-47.
[9] “About Npm.” Npm, www.npmjs.com/about.
[10] Foundation, Node.js. “About.” Node.js, nodejs.org/en/about/.
[11] Mark Otto, Jacob Thornton, and Bootstrap contributors. “Introduction.” · Bootstrap, getbootstrap.com/docs/4.0/getting-started/introduction/.
[12] “Using Google Charts | Charts | Google Developers.” Google, Google, developers.google.com/chart/interactive/docs/