SUPPORTING DATA PORTABILITY IN THE CLOUD UNDER THE GDPR

Yunfan Wang and Anuj Shah
Carnegie Mellon University
Privacy Engineering Capstone Project, Fall 2017

Executive Summary

Background

The EU's General Data Protection Regulation (GDPR), which will repeal and replace the Data Protection Directive of 1995 in May 2018, provides people within the EU the right to data portability under Article 20. This right allows a data subject to request a controller to whom they have provided personal data to transmit those data in a structured, commonly used, and machine-readable format to another controller or to themselves. Thus, it extends beyond the access rights allotted in most existing privacy frameworks and lowers the barrier for consumers to try a variety of competing services. However, this right has generated immense concern among data-driven enterprises – a recent survey of privacy professionals indicates that portability is the most difficult challenge facing those responsible for building a GDPR compliance program. Responding to portability requests will become even more complex for companies that distribute their services across cloud infrastructure. Cloud service providers should consider how to support data portability for their clients.

Problem Statement

1. Investigate how controllers' and processors' GDPR obligations to data subjects define what cloud service providers need to offer.
2. Identify the technical solutions that cloud service providers can leverage to support data portability for their clients.
3. Demonstrate these technical approaches to GDPR compliance through hypothetical use cases.

Significant Findings

1. **Controllers bear the majority of responsibility for GDPR compliance and the protection of data subject rights.** Controllers also have specific portability obligations depending on whether they are the sending or receiving party. While Article 20 does not require controllers to maintain fully compatible formats purely to enable data portability, it does encourage the development of interoperability standards.
2. **Processors are required to assist controllers in their fulfillment of data subject rights**, even though processors are not explicitly mentioned in GDPR Article 20.
3. **Which personal data fall under Article 20 obligations remains unclear.** While regulatory bodies within the EU agree that companies must port data explicitly provided by the data subject in response to a portability request, they disagree about whether Article 20 also covers data generated during a user's activity with their service.
4. **Identifiability of a dataset may partially determine whether it falls under Article 20 obligations.** GDPR Article 11 relieves controllers of their obligations under Article 15

through 20 if they can demonstrate they are not in a position to identify the data subject. Recital 26 states that the GDPR does not apply to anonymous information.

5. **The recipient of data from a portability request informs which portability method is most appropriate.** For example, access to commonly used and easy-to-store file formats through a usable web interface would be most appropriate for a data subject who wants to receive data directly. Transmission to another controller would be best achieved via industry-wide standard migration protocols, which would tolerate data loss.
6. **Cloud service providers cannot directly support Article 20 compliance solutions for business customers of Infrastructure-as-a-Service (IaaS) products.** Cloud providers only have coarse-grained visibility of IaaS data. Therefore, customers must implement their own portability methods for their end users.
7. **Cloud service providers can support migration at the IaaS and Platform-as-a-Service (PaaS) level for the narrow case in which customers are natural persons.** They may export and import virtual machine instances and databases.
8. **Cloud service providers have the opportunity to support Article 20 compliance in PaaS and Software-as-a-Service (SaaS) products.** Cloud providers have finer-grained visibility of PaaS and SaaS data, and can therefore assist with implementing portability methods. However, this is not required under the GDPR.
9. **GDPR-specific metadata classification provides a solution to Article 20 compliance that is flexible to various understandings of the GDPR and data types.** A useful metadata scheme would allow business customers to classify data types, sources of data, types of identifiers, whether a data type required consent of the data subject, identifiability of a dataset, and applicable data subject rights. This solution is demonstrated with a PaaS-level use case involving fitness tracking companies and a SaaS-level use case involving personal assistants.
10. **Our metadata classification proposal is unlikely to bring greater legal exposure to cloud providers.** While some cloud providers may be wary of greater involvement in a business customer's stored data, the customer is ultimately deciding how to classify their datasets and in turn deciding how to process those data. Thus, the cloud provider remains a processor.
11. **There is precedent in the cloud market for sensitive data classification and management.** Our proposal is a minor extension of existing services from Google Cloud Platform and Microsoft Azure.

Conclusion

If cloud providers want to continue providing competitive products to both their individual and corporate clients, they must consider how the right to data portability translates to their operations as controllers and processors. Their compliance solutions must also remain flexible due to the lack of certainty in GDPR enforcement. We present a set of recommendations for meeting this demand. The recommendations are based on careful consideration of how the recipient of data from a portability request, the cloud service level, and one's interpretation of the GDPR determine the appropriate portability solution. Most importantly, they rely on existing technical methods and build upon precedent in the marketplace, meaning that a compliance program for GDPR Article 20 is well within reach.