

# Tension between GDPR and Public Blockchains: A Data-Driven Analysis of Online Discussions

SINCONF 2020

Zeynep Chousein, Hacı Yakup Tetik, Rahime Belen Sağlam, Abdullah Bülbül, Shujun Li

2020.04.11



# GDPR

The General Data Protection Regulation

---

It is a regulation adopted in 2018 on data protection and privacy for individuals within the entire European Union and the European Economic Area.

The GDPR primarily aims:

- Individuals keeping their own data under control
- Compliance of companies in the European Union with this regulation



# GDPR

The General Data Protection Regulation

---

The parties that process personal data cannot take any action regarding the data except the consent of the person and the GDPR regulation. No data can be processed without the consent of the persons and they have the right to revoke their consent at any time.

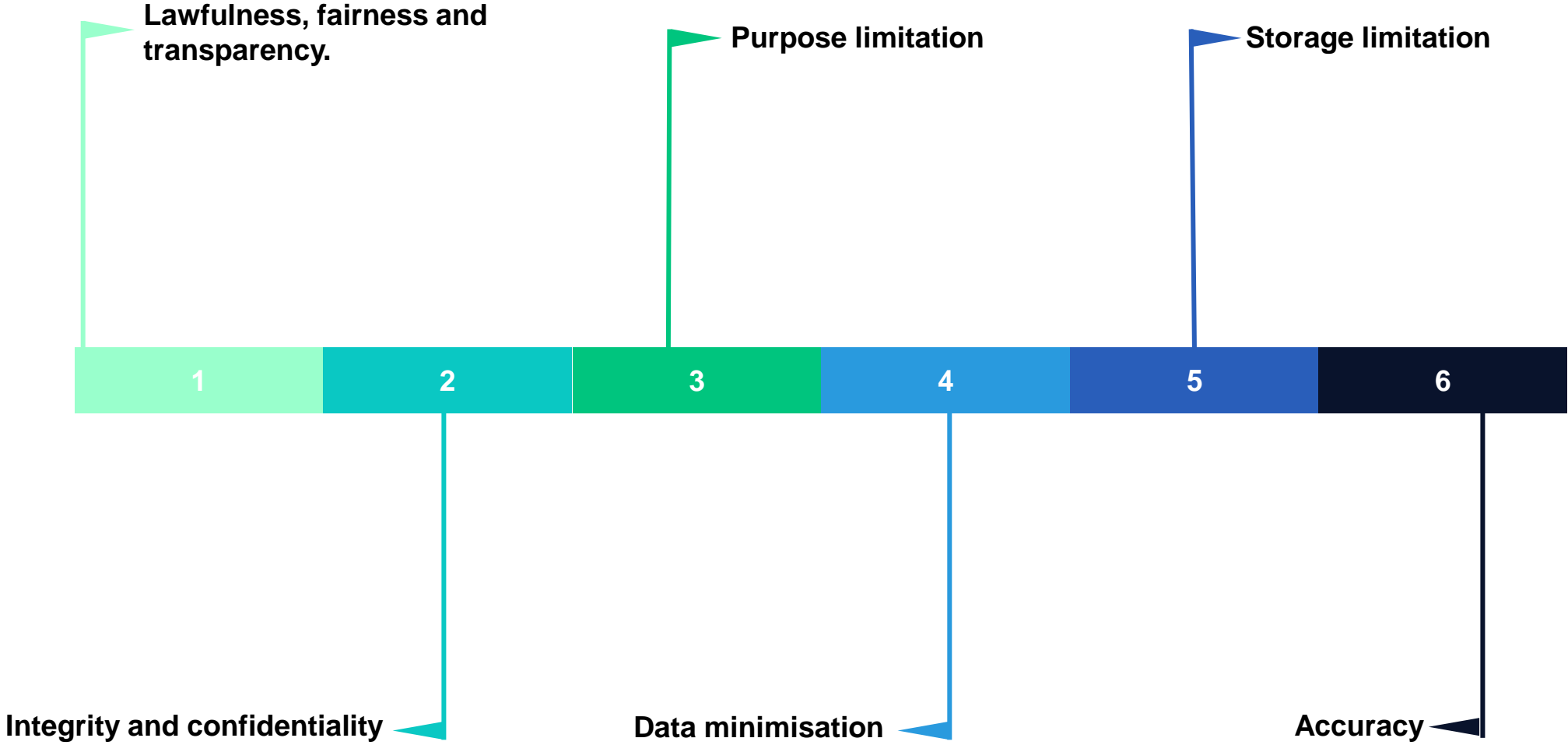
The parties that process personal data must specify the following characteristics;

- the purpose of data processing,
- its legal basis,
- how long the data is retained,
- whether it is shared with any third party



# GDPR Principles

Data protection principles under the GDPR



# Blockchain

---

It is a decentralized peer-to-peer network that acts as a distributed database by holding it to store transaction data on each node in the blockchain. It is also called distributed ledger technology (DLT) because each node keeps a copy of the distributed database and is at the center of this blockchain





## Blockchain vs GDPR

- The general perception about the blockchain that the processed data cannot be changed and can be monitored by everyone has led to the fundamental argument that GDPR compliance cannot be achieved.
- Because some of the rights supported by the GDPR, such as the right to delete personal data, contradict the unchangeable principle of the blockchain.
- Although solutions such as hiding personal data with an intermediate layer without disrupting the basic structure of the blockchain in the studies performed for this purpose, the current blockchain structure is not compatible with GDPR.



## How to collected data

In this study, we examined how much data owners are informed about this issue in online public communications with GDPR:

1. source code repositories (Github),
2. blogs and web forums, and
3. public tweets of a much more diverse group of Twitter accounts



## How to collected data

---

In our study, we looked at 50 cryptocurrencies with a market value of more than \$ 150 million in Coinmarketcap data.





# Github

---

Among the selected cryptocurrencies, 35 of them have github accounts. Thanks to the Github API search methods used, 970 repos were downloaded to our local



## Blog and Forum Posts

---

Blog and forum data were analyzed for

- Finding blockchain service providers interaction with their users in terms of GDPR
- Discovering the users have what kind of concerns in these interaction



## Blog and Forum Posts

Blogs and forums of top 50 cryptocurrencies were handled.

We developed tools with using two different libraries for gathering data:

- BeautifulSoup
- Selenium



## Blog and Forum Posts

We collected posts of blogs and forums which contains the word GDPR or the full title “General Data Protection Regulation“ in order to find relevant posts

Type	Number
Total number of web forums covered	18
Total number of blogs covered	38
Total posts collected	17,821

# Twitter

---

We used official twitter accounts of 39 cryptocurrencies.

We collected tweets which contains hashtags of official twitter accounts published between 1 January 2018 and 1 May 2020 with using library `GetOldTweets3`.



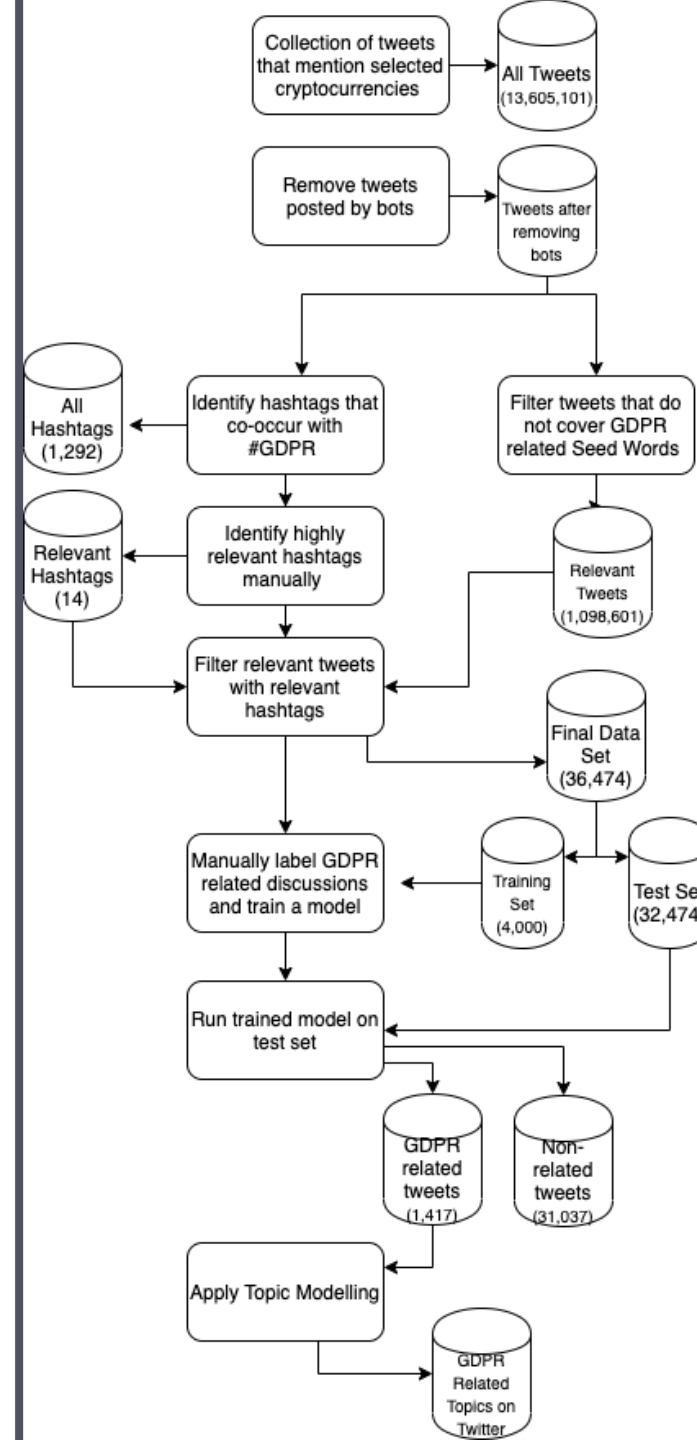


# Twitter

We used official twitter accounts of 39 cryptocurrencies.  
More than 13 million tweets were collected in that study.

Preparation of data has some steps:

- Finding and eliminating tweets which were posted by bots.
- Filtering tweets by using seed words.
  - Seed words were decided manually by visiting GDPR document and the ICO's guide
- After elimination and filtering tweets we have 1,606,269 tweets.



Text

# Twitter

Preparation of data has some steps:

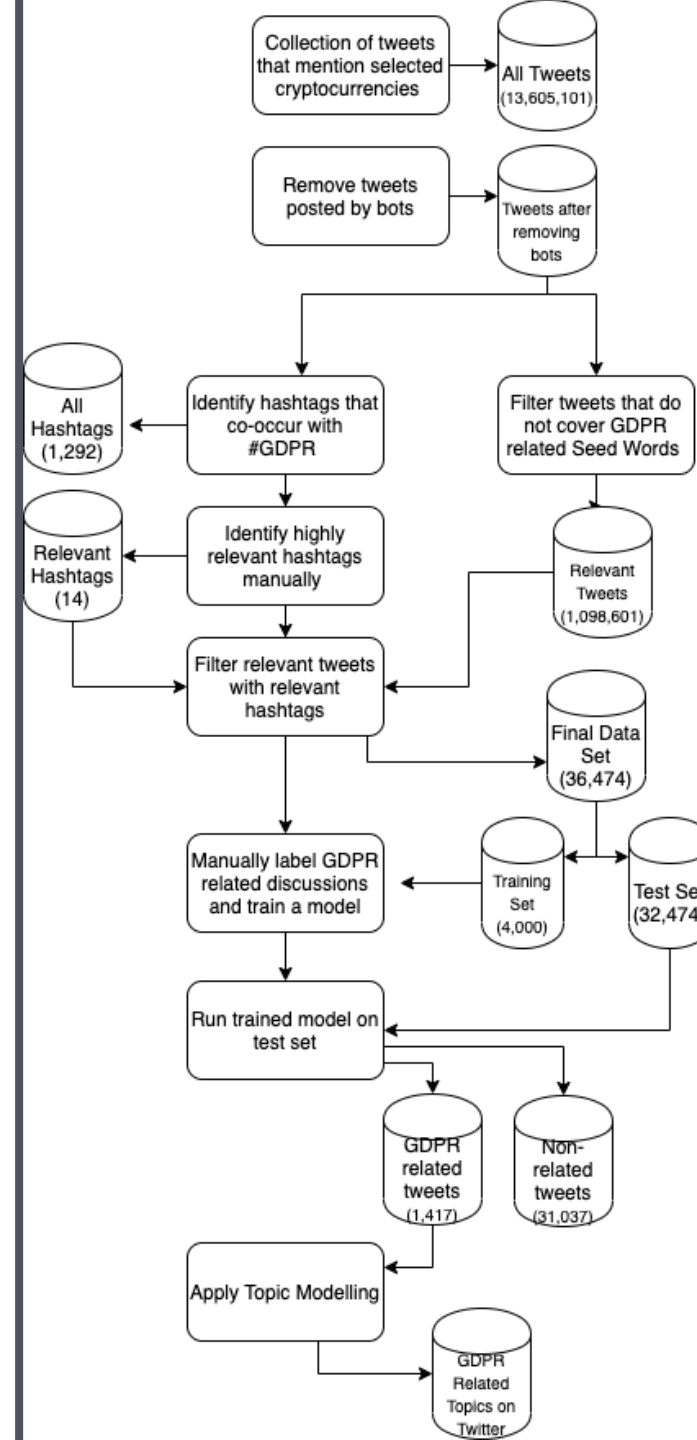
- Ranking hashtag frequencies for finding most popular ones.
  - 14 most popular hashtags were selected for future analysis.
  - 36,482 tweets were found which contain those hashtags.

#GDPR	#DataProtection	#DataPrivacy
#privacy	#compliance	#trust
#regulation	#PersonalData	#Regulations
#biometrics	#law	#security
#legal	#brokageofpersonaldata	

# Twitter

Preparation of data has some steps:

- Topic modelling ran on this set for finding main themes.
- We randomly selected 4,000 tweets out of 36,482 tweets because result is too broad and meaningless.
- 4000 tweets were labelled manually.
- Rest of tweets were trained a classifier by using Naives Bayes algortihm.
- After classifiying 1,417 GDPR-related were detected.



Text



# Results



## Blog and Forum Posts

- 
- Blog posts are used for publishing news, announcements or entertainment by systems.
  - Forum post are used by the users of systems for finding answers of their questions.
  - We analyzed 56 posts which were GDPR-related by manually.



## Blog and Forum Posts

---

- 33 posts gives generic information given about the GDPR.
- 9 posts are related about right to be forgotten (RTBF).
  - 3 of them gives very brief statements.
  - Other 3 posts gives information about users of their systems have this right, without any further explanation .
  - Last 3 posts of them says that user data cannot be deleted on their systems

## Blog and Forum Posts

---

- 4 posts are related about the right to access.
  - 2 of them has vague information.
  - Other posts are not helpful to understand how users can exercise it.
- 2 posts are related to tranfer of personal data.
  - There is one question about it but it is not answered.
- Only 1 forum post is related to the right to withdraw consent

- 
- Only 1 post is related to data minimization.
  - Again only 1 post is related to data protection by design and default.

After pre-processing data, we found different number of topic with different number of words by using LDA algorithm.

Topic	Words
Topic 1	essential ico decentralized tokensale crypto dapps invest crowdfunding access right
Topic 2	security privacy ico bit data project new pany GDPR capitaltechnologiesresearch
Topic 3	data personal pdata secure crypto opiria yourblock io brokageofpersonaldata opirium
Topic 4	ico crypto news io tokensale smart contract technology platform right
Topic 5	ico crypto tokensale project join forget dont token secure security



# Conclusion



## Conclusion

---

There is a lack of explicit acknowledgment and warnings to users on the legal challenges introduced by the nature of the blockchain technology.

There is no enough warning in Github repositories and Forum and Blog posts. We could not find any proper communications of blockchain systems and service providers on GDPR.



## Conclusion

---

Blockchain systems are not active on Twitter in terms of GDPR

**Thanks for  
listening 😊**

**Mail: [zozkalkan@gmail.com](mailto:zozkalkan@gmail.com)**