Blockchain and the GDPR

How to block and delete person related data in a technology that is designed to be unchangeable

Requirements of the GDPR

The GDPR, which is designed to protect the privacy of a natural person, requires the possibility to block and delete data which belongs to a natural person amongst other specifications. These two claims are to the two most difficult in the context of blockchain based applications.

1) Blocking in this context means that the data related to this person will not be processed for the time being marked as blocked.
2) Deletion can also be logical, which means that the related data itself does not have to be deleted, but it cannot be linked to the related person anymore

There are several legal cases in which data cannot be deleted, even if the person asks for it. But hardly any application with users living in the European Union can surely exclude that it has to be able to fulfill the above requirements.

This article is not intended to dive into details about legal situations. For specific questions a team of legal and technical experts have to work on answers. Here are only technical aspects considered.

Deletion of Data in Blockchain

Blockchain is designed to be unchangeable. This is done by two basic approaches:

1) The exact content of each block in the chain and the order of the whole chain itself is verified by hashes ensuring the integrity of the data. Changing a single character would change the hashes of the block in which it has been changed as well as the hashes of all following blocks.
2) So called “nodes” compete in solving a mathematical task (“proof of work”). The winning node determines the order of the data within a block. *
3) A lot of copies of the blockchain are supposed to be distributed on machines all over the world. *

* The nodes/machines are originally intended to belong to various independent persons and institutions. That is where the term “democratic” comes from.

So we see, data cannot be deleted from a blockchain without messing up the whole concept of the technology itself. But with a sensible planning it can be prepared for logical deletion and the possibility of blocking the processing of a users data for a certain time.

Schematic Explanation of the Blockchain Technology

A block in blockchain is a sequence of transactions. In the original blockchain a transaction stands for an actual transaction of cryptocurrency, but in other blockchain-based applications it can be any kind of event.

header hash 0: **hash of** block 0	**header hash 1:** **hash of** header hash 0 + block 1	**header hash 2:** **hash of** header hash 1 + block 2	**header hash 2:** **hash of** header hash 2 + block 3	⋯
transaction n_0	transaction n_1	transaction n_2	transaction n_3	
⋮	⋮	⋮	⋮	
transaction 1_0	transaction 1_1	transaction 1_2	transaction 1_3	
block 0	block 1	block 2	block 3	

And the hashes as functions of inputs are computed like this:

header hash 0: hash(block 0)

header hash 1: hash(block 1 + header hash 0)
= hash(block 1 + hash(block 0))

header hash 2: hash(block 2 + header hash 1)
= hash(block 2 + hash(block 1 + header hash 0))
= hash(block 2 + hash(block 1 + hash(block 0)))

header hash 3: hash(block 3 + header hash 2)
= hash(block 3 + hash(block 2 + header hash 1))
= hash(block 3 + hash(block 2 + hash(block 1 + header hash 0)))
= hash(block 3 + hash(block 2 + hash(block 1 + hash(block 0))))

⋮ ⋮

Now it is important to understand, that a hash function cannot be inverted. This means that computing forward is easy, but computing backward is not possible.
The only way to get the original input of a hash is to try all possible inputs, compute their hashes and compare them to the given hash. This can take seconds or billions of years. It depends on the number of possibilities and the performance of the hash function itself.

This is why a blockchain looses its proven integrity if you cut off the first block or any number of blocks from the beginning!

Logical Deletion in Blockchain

Logical deletion in the context of the GDPR means to unlink the person related data from the person. It should not be possible anymore to assign the data to a natural person. This can only be done by an ID or key in the transactions on blockchain that is linked to an account of a natural person, where all other personal data is stored an can be deleted.

For the decision which data is considered personal in which context, legal experts have to be engaged. This should happen before the publishing of a blockchain based application.

Because as soon as the first block is up there, you cannot change any data anymore which already has been uploaded!

A Remark on Data Retention

For everybody who is concerned about data privacy I need to emphasize though, that even if the law accepts logical deletion of data in the context of the GDPR, anybody with access to the blockchain can make their own documentation of IDs/keys belonging to natural persons and still relate the transactions to the person whose account has been deleted.