Smart Contracts for Data Accountability and Provenance Tracking

Section 1 Summary

Use Case Summary			
Use Case ID:	DTM-001	**Use Case Type:**	*Horizontal*
Submission Date:	December 14, 2018	**Is Use Case supporting SDGs**	*Yes*
Use Case Title:	Smart contracts for data accountability and provenance tracking	**Domain:**	*Data processing, storage and management*
Status of Case	*Proof-of-concept*	**Sub-Domain**	
Contact information of person submitting/ managing the use-case	*Full Name: Ricardo Neisse* *Job Title: Scientific Project Officer* *E-mail address: ricardo.neisse@ec.europa.eu* *Telephone number: +39 0332 78 9592* *Social media: https://twitter.com/EU_ScienceHub* *Web site:* https://ec.europa.eu/jrc/en		
Proposing Organization	*European Commission Joint Research Center, Ispra, Italy*		
Short Description	*Smart contracts can be used to track data provenance and encode usage control policies regulating the access and usage (e.g., redistribution) of subject's data by controller and processors.*		
Long description	*The recent approval of the General Data Protection Regulation (GDPR) imposes new data protection requirements on data controllers and processors with respect to the processing of European Union (EU) residents' data. These requirements consist of a single set of rules that have binding legal status and should be enforced in all EU member states. In light of these requirements, this use case propose the use of a blockchain-based approach to support data accountability and provenance tracking. This approach relies on the use of publicly auditable smart contracts deployed in a blockchain that increase the transparency with respect to the access and usage of data. Smart contracts can be used to encode data usage policies and provenance tracking information in a privacy-friendly way.*		
SDG in Focus (when applicable)	*Goal 16: Promote just, peaceful and inclusive societies*		
Value Transfer:	*Fingerprints of digital identity and personal data items*	**Number of Users:**	*Large scale including citizens of many EU countries*
Types of Users:	*Data Subjects, Data Controllers, and Data Processors*		
Stakeholders	*Citizens, enterprises handling digital identity and personal data items, government institutions auditing privacy practices of enterprises.*		

Data:	*Fingerprints of pairs of data type and values exchanged between a data subject and data controller, including an obfuscated usage control policy regulating how the data should be used by the controller/processor.*
Identification:	*The use case proposes a privacy-friendly way of encoding identities, data and policies in a way that is still meaningful for auditability purposes. The only thing that can be learned is the structure of the policy specified by data subjects and no details about the data or restricted activities that can be performed by data processors and controllers.*
Predicted Outcomes:	

Overview of the Business Problem or Opportunity
Subjects provide data to controllers and processors and have no auditable way of verifying how their data is being processed, stored, and redistributed. In case of privacy violations reported by subjects controllers and processors should be able to prove the data is stored and processed according to the subjects' privacy requirements. *Smart contracts can be used as an auditable way of encoding data provenance information and privacy requirements to enable subjects to evaluate who has accessed their data and the conditions for storage, processing, and redistribution of the data. In case subjects believe their privacy requirements are not being fulfilled they can revoke data access and usage rights using the blockchain. This provides a mechanism for legal compliance in the face of the new EU General Data Protection Regulation (GDPR). Since in public blockchains the smart contracts are readable by anyone the data provenance and accountability information should be encoded in a privacy friendly way.*
Why Distributed Ledger Technology?
In traditional centralized ledgers data subjects have no way of auditing and verifying (1) the set of data accessed by data controllers and processors and (2) how the provided data is being used. The use case relies on the immutability, verifiability, and transparency of DLT.

Section 2 Current process

Current Solutions
Not available.

Existing Flow (as-is)		
Step	**User Actions**	**System Actions**

Process scheme (as-is)

Data and information (as-is)		
Data	**Type**	**Description**

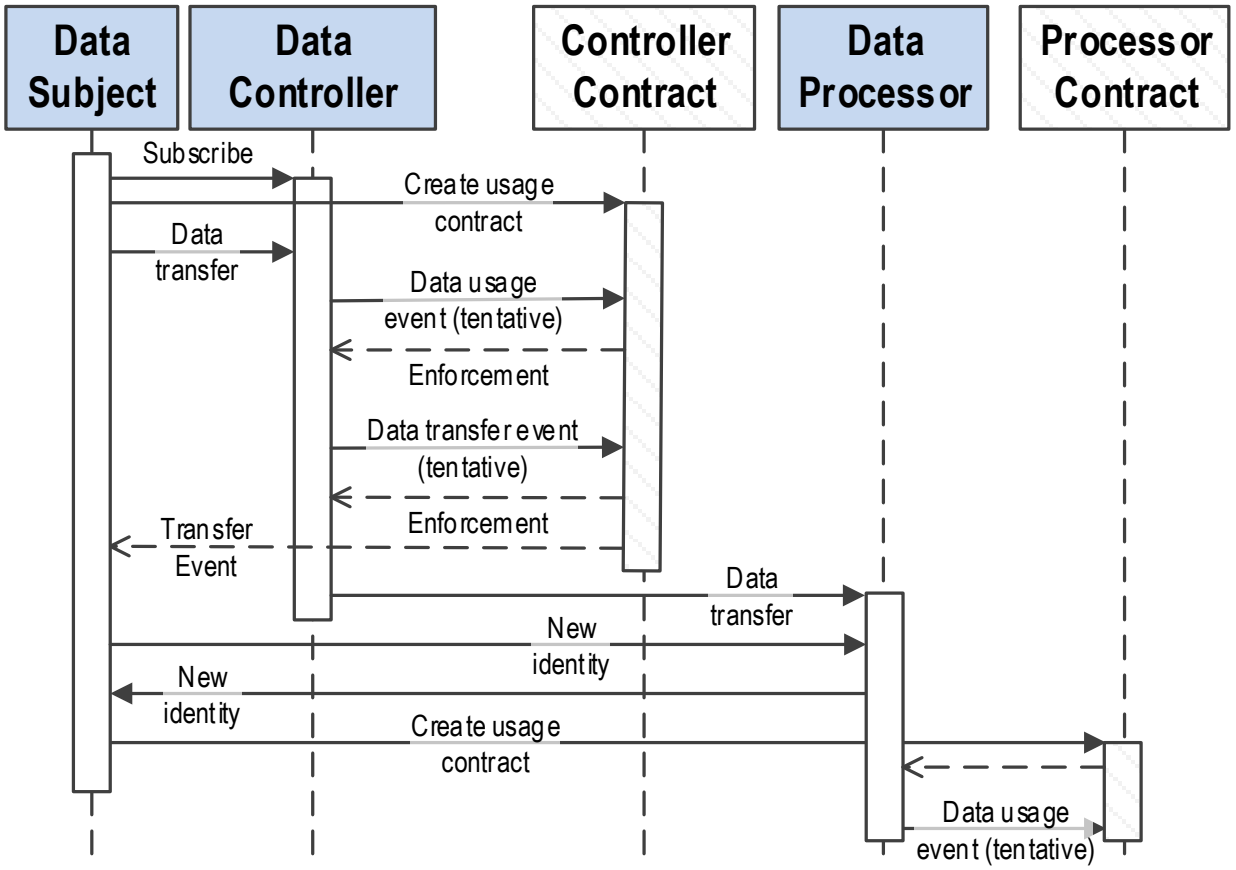
Participants and their roles (as-is)		
Actor	**Type/Role**	**Description**

Other Notes

Section 3 Expected process

Expected Flow (to-be)		
Step	**User Actions**	**System Actions**
1.	The data subject subscribes with the data controller and provides the list of data types/values that are exchanged together with an usage control policy.	A data usage contract is created in the blockchain using a random nonce and fingerprints for the identities, data items (types/values) and usage control policies associated to the data items.
2.	The data subject transfer data to the data controller.	The data controller stores the data and a reference to the smart contract.
3.	The data controller is about to use the data for any internal activity.	The data usage contract is consulted to verify if the activity is allowed to be performed with the data item and an answer is returned (allow/deny/modify/delay data usage).
4.	The data is about to be transferred to a data processor.	The data usage contract is verified and if the transfer is allowed a new contract is created for the specific data processor. The cycle repeats the same for each data processor starting with step 1.

Process scheme (to-be)



Participants and their roles		
Actor	**Type/Role**	**Description**
1.	*Data subject*	Any person providing data to a data controller/processor.
2.	*Data controller*	Any organization receiving data from a data subject.
3.	*Data processor*	Any organization receiving data from a controller to perform specific data processing activities.

Data and information		
Data	**Type**	**Description**
1.	*Subject, controller, and processor identities*	Unique identities for subjects, controller, and processor that are not re-used for other contracts in order to avoid linkability.
2.	*Data type and value fingerprints*	Fingerprints of data types and values using a one-way hash function in combination with a random nonce to prevent dictionary attacks.

Data and information		
Data	**Type**	**Description**
3.	*Usage control policy*	An Event-Condition-Action policy specifying the data usage event, the condition, and the respective action (allow, deny, modify, delay, or execute).

Security and privacy

The data items and identities stored in the blockchain are obfuscated to allow a privacy-by-design approach for the use case.

Main Success Scenario + expected time line

The main success scenario is a public blockchain where all data subjects are able to record and audit the data exchanged with data controllers and processors including their usage control policies in a privacy friendly way.

There is currently no expected time line since this is a research prototype.

Conditions (pre- or post-)

Subjects provide data to controllers and processors and have no auditable way of verifying how their data is being processed, stored, and redistributed. In case of privacy violations reported by subjects controllers and processors should be able to prove the data is stored and processed according to the subjects' privacy requirements.

Smart contracts can be used as an auditable way of encoding data provenance information and privacy requirements to enable subjects to evaluate who has accessed their data and the conditions for storage, processing, and redistribution of the data. In case subjects believe their privacy requirements are not being fulfilled they can revoke data access and usage rights using the blockchain. This provides a mechanism for legal compliance in the face of the new EU General Data Protection Regulation (GDPR). Since in public blockchains the smart contracts are readable by anyone the data provenance and accountability information should be encoded in a privacy friendly way.

Performance needs

In public blockchains, scalability is an issue considering the amount of data accessed, stored, and processed by many data controllers and processors. Maybe this approach is more viable for very sensitive data, for example, medical records.

Legal considerations

1. The goal of the use case is to support the implementation of the General Data Protection Regulation (GDPR)

Risks

No legal, business and technical risks related to use case were identified.

Special Requirements

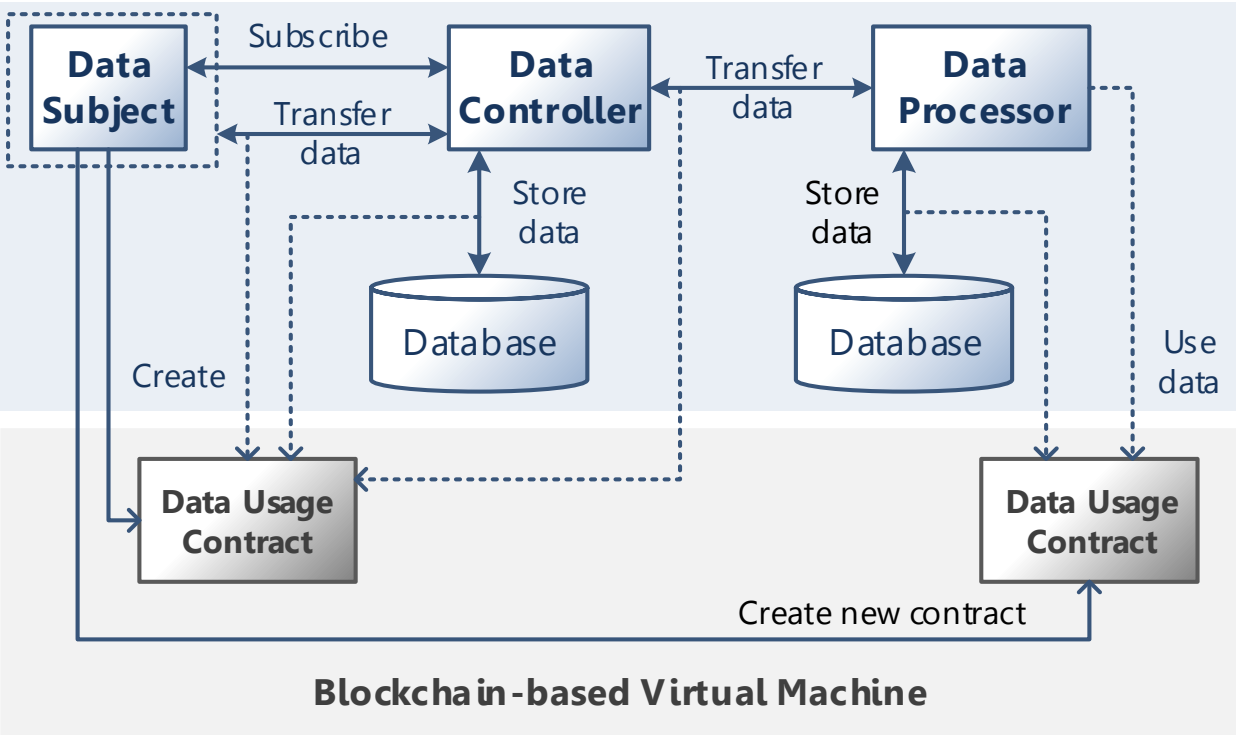
No special business and technical requirements of use case were identified.

External References and Miscellaneous

Ricardo Neisse, Gary Steri, and Igor Nai-Fovino. 2017. A Blockchain-based Approach for Data Accountability and Provenance Tracking. In Proceedings of the 12th International Conference on Availability, Reliability and Security (ARES '17). ACM, New York, NY, USA, Article 14, 10 pages. DOI: https://doi.org/10.1145/3098954.3098958

Other Notes

The following figure presents the high-level architecture of the data accountability and provenance tracking model proposed in this use case. In this architecture, three main entities are depicted following the GDPR terminology: the Data Subject, the Data Controller, and the Data Processor. When the subject subscribes with a controller, which is typically the role of a service provider, it creates a policybased Data Usage Contract specifying constraints on the usage and redistribution of any data obtained explicitly or implicitly by the controller. Explicit data is any data provided directly through interactions with the subject such as the e-mail addresses or birth date. Implicit data is any data acquired automatically, for example, sensor data from IoT devices in the environment surrounding the subject, data acquired by apps installed in mobile devices, or even server log files registering details of the network interactions between subject and controller services (e.g., IP addresses). The contract in this model acts as a data provenance tracker, policy evaluation entity, and event logger that allow the subject to easily check all data transfers and usage transactions providing assurance that only transactions conforming to the contract policies are authorized and registered in the blockchain.



Appendix 1

Domains and subdomains for use cases categorization

Vertical:

1. Finance
    a. Financial management & accounting
    b. International & interbank payments
    c. Clearing and settlement
    d. Reduction of Fraud
    e. Financial messaging
    f. Asset lifecycles and history
    g. Trade finance
    h. Regulatory compliance & audit
    i. AML/KYC
    j. Insurance
    k. Peer-to-peer transactions
2. Healthcare
    a. Pharma
    b. Biotechnology
    c. Medicine
3. Industries
    a. Manufacturing
    b. Energy
    c. Chemical
    d. Retail
    e. Real estate
    f. IT and telco
    g. Supply chain management
    h. Transportation
    i. Agriculture
4. Government and public sector
    a. Taxes
    b. Government and non-profit transparency
    c. Legislation, compliance & regulatory oversight
    d. Voting
    e. Taxation and customs
    f. Intellectual property management
    g. Land Registries

Horizontal:

1. Identity Management
2. Security Management
    a. Public Key Infrastructure

3. Internet of Things
4. Data processing, storage and management
    a. Data Validation (includes provenance)