Open Data Market Architecture and Functional Components

Demchenko, Y.; Cushing, R.; Los, W.; Grosso, P.; de Laat, C.; Gommans, L.

DOI
10.1109/HPCS48598.2019.9188195

Publication date
2019

Document Version
Final published version

Published in
2019 International Conference on High Performance Computing & Simulation (HPCS)

Open Data Market Architecture and Functional Components

Yuri Demchenko, Reggie Cushing, Wouter Los,
Paola Grosso, Cees de Laat
University of Amsterdam, The Netherlands
{y.demchenko, R.S.Cushing, W.Los, p.grosso,
C.T.A.M.deLaat}@uva.nl

Leon Gommans
Air France - KLM
leon.gommans@klm.com

POSTER PAPER

Abstract— **This paper discusses the principles of organisation and infrastructure components of Open Data Markets (ODM) that would facilitate secure and trusted data exchange between data market participants, and other cooperating organisations. The paper provides a definition of the data properties as economic goods and identifies the generic characteristics of ODM as a Service. This is followed by a detailed description of the generic data market infrastructure that can be provisioned on demand for a group of cooperating parties. The proposed data market infrastructure and its operation are employing blockchain technologies for securing data provenance and providing a basis for data monetisation. Suggestions for trust management and data quality assurance are discussed.**

Keywords- Open Data Market, Data Marketplace, Trusted Data Market, Industrial Data Space, Data Economics, STREAM Data Properties

I. INTRODUCTION

A data economy may emerge with effective use of data from multiple sources. Traditional data sources such as company databases and applications related to the main business line are now complemented by non-traditional sources such as social media or sensors network data from devices such as mobile devices, smart meters, cars and industrial machines, commonly referred to as Internet of Things (IoT). An entirely new market of Big Data technologies and services has emerged over half a decade to help organisations capture and extract value from all the data. The revenue from Big Data technologies and services, however, is still small compared to the value that is expected to result in sectors such as trade, manufacturing, finance and insurance, public administration, health and social care. Big Data technology and high-performance computing provides necessary tools to make innovative use of data to drive high-value business and societal outcomes, however they are not yet fully deployed for new data markets [1, 2]. Currently, the large-scale data driven businesses are mostly driven by a few large companies. The whole data potential can only be unleashed when all participants of the emerging data economy are involved, including SMEs whose innovative potential will significantly increase with wider access to variety of data and with the possibility to trade data as product of the innovative activity.

The concept of Open Data Markets (ODM) can be considered as an important enabling technology to facilitate data driven economy. The ODM must be based on the well-defined architecture and operational model and open for all organisation complying with the data market rules and policies. The ODM must be based on relevant industry standards and provide secure and trusted data exchange between data market actors: data producers/owners and data consumers services and applications developers and operators. In fact, an ODM operational model should benefit from the experience of well-established commodity markets and the big data companies' operational models. A functional data market model should include multiple aspects: the definition of data as economic goods, a data market architecture including a data market infrastructure and its functional components, a data market operational model and a regulatory basis.

This paper presents the ongoing research to define data markets architecture and infrastructure components to enable secure and trusted data exchange and support data value creation. The paper refers to earlier authors' work [3] that provided a definition of the data properties as economic goods and defined the main functional components of the future ODM. The proposed ODM infrastructure definition leverages multiple available technologies that are put together to enable secure and trusted data exchange and trading between data market participants

The paper is organised as follows. Section II provides a definition of the data properties as economic goods and identifies the main functional components of ODM. Section III defines in detail the data market infrastructure allowing the provisioning of an instant virtual private data market infrastructure, on-demand for a group of cooperating parties. Section IV provides suggestions for using blockchain in data transaction provenance and trust management issues. Section V provides a summary and suggests future research directions.

978-1-7281-4484-9/19/$31.00 ©2019 IEEE

II. Data Properties as Economic Goods and Data Market Requirements

A. STREAM Data Properties

Emerging data driven economy and modern Big Data technologies facilitate interest to making data as a new economic value (data commoditisation) and consequently the identification of new properties of data as economic goods. The importance of making data exchangeable has been recognised by the research community in defining the FAIR data principles: to become a tangible research product, data must be Findable, Accessible, Interoperable, Reusable [4]. The following properties are leveraging the FAIR data principles and defined as STREAM properties for industrial and commoditised data [2]:

[S] Sovereign
[T] Trusted
[R] Reusable
[E] Exchangeable
[A] Actionable
[M] Measurable

Other data properties important to enable data commoditisation and to allow for data trading and exchange of goods include: Quality, Value, Auditability/Trackability, Branding, Authenticity, besides the original FAI(R) properties Findability, Accessibility, Interoperability, Reusability. All data transfers and transformations require managed data ownership, IPR and privacy. The data property originated from its digital form of existence defined as not-Rivalry, on one hand, makes data exchange (copying, distribution) easy, but on the other hand, creates problem when protecting proprietary, private or sensitive data or IPR.

B. Characteristics of emerging data markets

Data markets should be able to support all major properties of Big Data from/for all application domains, allow data exchange and integration at different stages while preserving data provenance and auditability. Below is non-exhaustive list of the general data market properties powered by cloud technologies [2]:

- support for heterogeneous data exchange at different processing stages;
- cloud powered/integrated;
- customer-centric, open for all market participants;
- automated, smart;
- regional/sectoral specialized;
- collaborative;
- governed;
- secure, trusted;
- auditable;
- transparent;
- enabling data commoditization and operating with monetized data values.

The above listed properties provide a suitable framework for the definition of the ODM architecture. It will also contribute to a better definition of data as economic tradeable goods.

III. Data Market infrastructure and components

The proposed trusted ODM architecture follows and advances the Industrial Data Space Association (IDSA) Architecture 2.0 that is built on the foundations of data sovereignty and the IDS Connector architecture [5]. We propose to extend it with infrastructure elements to support secure and policy enforced trusted data exchange.

Figure 1 illustrates the generic Data Market infrastructure that primarily relies on the end-to-end data exchange and contracts with limited additional infrastructure and third-party services. The following functional components are defined: Provider and Consumer (or data producer and data user), Directory (or catalog) as a simple service to publish information about data, Data Exchange as an important third party infrastructure service, optional Data Storage, and Data Intelligence provided as part of infrastructure services to support business scenarios for data trading and value services, Data Broker, and Data Connectors operating at the producer and consumer sides.

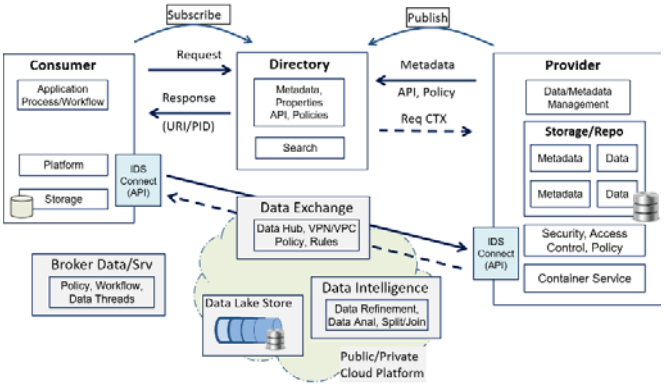

Figure 1. Data Market infrastructure components leveraging IDS Architecture 2.0.

The following data market operational scenarios are considered:

a) 1 to 1 – individual contracts (including intermediaries) that may also support automatic contract negotiation and policy enforcement;

b) 1 to n – data webshop providing data services to many customers, including IoT sensor data streaming; reflects a most common business and operational model for data services subscription based on a standard Service Level Agreement (SLA). Data producer/provider may employ special infrastructure for data services delivery such as Content Distribution Network CDN) or edge computing infrastructure, typically available from public cloud services providers.

c) n to 1 – one data consumer/processor collecting data from many data sources/providers with the goal to drive own data-based applications or provide value added data services. This model in fact is similar to the previous model but may require building a special infrastructure by the data collector adopting to the collected data sources and data types.

d) n to m – full data market mesh that should allow provisioning the Virtual Private Data Market (VPDM) infrastructure instance for a group of cooperating organisations. The VPDM infrastructure should be based on the negotiated agreement and support a set of standard services to ensure

technical, business and regulatory requirements on data exchange.

Figure 2 illustrates the whole set of composable infrastructure components that will allow customised Virtual Private Data Market provisioning on demand for contractual groups/entities to enable secure and trusted data exchange and trading.

Technologies as Virtual Private Cloud (VPC) and Virtual Private Networks (VPN) provide highly secure environments for interacting members that allows easy setup and configuration for a given group of interacting parties [6, 7].

For the better definition of the ODM technological aspects we introduce the Trusted Data Market concept and platform that generalises the technology to enable secure and trusted data exchange on the ODM. The proposed ODM infrastructure should be based on the TDM platform and allow all described above operational scenarios, it should allow composition of infrastructure provisioned on demand for each configuration of provider-consumer relations.

The following are the main building blocks of the composable VPDM infrastructure:

- Trusted Data Market platform which is cloud based on trusted and compliant to industry security and compliance standards (e.g. Open Telecom Cloud by T-Systems, or one of public cloud platform)
- Catalog/Directory of data providers, algorithms or application providers, TDM platforms – with related information, including API and SLA, allowing options assessment and further/initial services and infrastructure composition
- Data Market Portal as the main way of communicating with customers, including the web-based Management console, that should allow all services search, review, request, and management
- Set of templates and images (VM or containers) for deploying VPDM instances on demand, including Virtual Private Cloud (VPC) and Data Container(s)
- Secure Data Connectors to enable sovereign end-to-end data provider and consumer connection, leveraging the IDS Connectors architecture and blockchain based computational enforcement of the market policies and rules
- Secure Data Container (SDC) which is blockchain enabled and deployed/invoked as part of instant VPDM/VPC
- Secure and Trusted Data Storage which is blockchain enabled for data provenance support (data transaction, copies, replication, etc), and typically implemented as a Data Lake storage that allow fast storage of multiple data formats (without prior schema definition) and later access by multiple applications in a number of formats (schema on read).
- Data Intelligence Hub (DIH) allowing data analytics processing in the trusted environment of VPDM/VPC
- Smart Contract Management (SCM) system, blockchain enabled and based on the private blockchain platform (combining industry standard Hyperledger and Solidity contract language)
- Smart Contract Virtual Processing Environment (SCVPE4TDM) that leverages the Ethereum Virtual Machine approach and implementation
- Service Level Agreement Management (SLAM) that provides information to define instant Smart Contracts and relies on the SCVPE4TDM functionality

Security and Trust management components that include Federated Access Control (Authentication, Authorisation, Identity Management) Infrastructure (FedAAI) and Virtual Infrastructure Trust Bootstrapping Protocol (VITBP) that is based on the Trusted Platform Module (TPM) and allows trusted bootstrapping of the on-demand provisioned virtual security infrastructure to VPC and hardware and TPM based trust anchors.

Trusted Broker (TB) is operated as a 3rd party service and has important role in many possible scenarios of the practical use of TDM and enabling multiple business models

The practical implementation of the proposed TDM infrastructure will also include a number of trials implementing core use cases that can be run as part of the Data Market Intelligence Hub (DMIH). DMIH will provide secure environments such as Sealed Room or Data Harbor for demonstrating and trying out the proposed TDM platform.

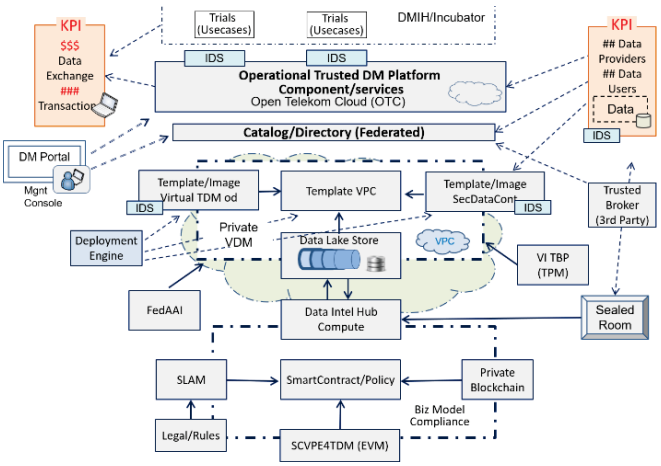

Figure 2. Trusted Data Market functional components

IV. ENSURING ODM PLATFORM TRUSTWORTHINESS

A. Security and trust management

The data sovereignty and trustworthiness of the data market platform are considered as key to enable data exchange and integrating/involving data markets into B2B relations. The proposed trusted ODM provides foundation for enabling data sovereignty by providing mechanisms for data owners/producers and data driven service providers to create data centric digitally enforceable rules embedded into mutual

contracts between B2B and B2B2C partners. The following technologies and mechanisms are used:

- VPDM instance for cooperating parties provisioned on demand with embedded digitally enforced rules that ensure compliance with inter-party contacts and data exchange and use policies. Security of the TDM instance is ensured by the proven Virtual Private Cloud technology available from the major public cloud platforms (e.g. AWS and Azure) and supported by open cloud management platforms Open Stack and CloudStack [6, 7].
- Virtual trusted infrastructure provisioning and trust management and access control using trust bootstrapping mechanisms and Trusted Platform Module (TPM) [7, 8].
- Secure Data Containers (SDC) that re leveraging the advanced container platforms and extending existing data container and distributed Big Data files system with rich metadata definition and data provenance functionality
- Blockchain enabled smart contracts used for policy/contracts enforcement and data provenance/accountability [9, 10]

Data trustworthiness (also defined as Veracity in the Big Data properties definition [11]) is ensured primarily by the data owner/producer and must be supported by the ODM infrastructure, in particular providing mechanisms and/or services to monitor the data market members' reputation which solutions can be adopted from other technology domains both peer-to-peer networks and blockchain [12, 13].

The proposed ODM architecture and operational model requires a variety of technologies to work together realizing data centric data exchange and transformation to enable data-based applications and services and added value data services creation. New functionality and technology combination will require re-thinking existing concepts and models, extending usage scenarios, in particularly for distributed file systems integrity and security [14].

B. Data assets management and cost model

Making data a valuable asset requires mature Data Management and Governance adoption by organisations that must be based on the well-defined industry best practices [15, 16]. Data traded as economic goods must possess/demonstrate such properties as measurable quality, identifiability, veracity, non-rivalry/re-usability, privacy, compliance and comply with the FAIR principles.

The Data Management and Governance is an important part of the whole Data Market ecosystem that enables the data producers and data driven application providers/businesses to bring data to the market as economic goods.

The trusted ODM implementation must include Data Management and Governance (DMG) framework to guide the organisation on implementing best practices/standards and assessing DMG maturity required for managing data as economic goods and operating data related services.

V. SUMMARY AND FUTURE RESEARCH

The paper presented the conceptual view on the Open Data Markets to enable secure and trusted data exchange and trading as economic goods. The presented work is a result of the active research at the System and Networking Lab at the University of Amsterdam. The proposed ODM architecture is based on the modern cloud technologies that allow provisioning on demand complex infrastructures which is secured with the Virtual Private Cloud mechanism. The main focus of current development is integration of multiple technologies and testing their interoperability in basic data exchange scenario, in particular blockchain technologies to support complex data exchange, provenance and monetization. This includes private permissionable blockchain, inter-blockchain communications, smart contract languages and their interoperability and security.

The ongoing technical development is also supported by the conceptual research on the data economics as the part of the activities at the Research Data Alliance (RDA) Interest Group on Data Economics (IG DE) [17] which involves wide research community of practitioners and targets at delivering major definitions related to the data properties as economic goods and best practices recommendations on data management, economic and business models.

ACKNOWLEDGMENT

The research leading to these results has received funding from the Horizon2020 project FAIRsFAIR (grant number 831558), and builds upon the work done within the Dutch NWO Research project 'Data Logistics for Logistics Data' (DL4LD, www.dl4ld.net), supported by the Dutch Top consortia for Knowledge and Innovation 'Institute for Advanced Logistics' (TKI Dinalog, www.dinalog.nl) of the Ministry of Economy and Environment in The Netherlands and the Dutch Commit-to-Data initiative (https://commit2data.nl/)

REFERENCES

[1] A Digital Single Market Strategy for Europe, EC Communication, 2015 [online] http://europa.eu/rapid/attachment/IP-15-4919/en/DSM_communication.pdf

[2] [EC-DSM, 2018] Towards a Common European Data Space', European Commission, April 25th 2018, https://ec.europa.eu/digital-single-market/en/news/communication-towards-common-european-data-space

[3] Demchenko, Yuri, Wouter Los, Cees de Laat, Data as Economic Goods: Definitions, Properties, Challenges, Enabling Technologies for Future Data Markets, ITU Journal: ICT Discoveries, Special Issue "Data for Goods", To be published December 2018

[4] Barend Mons, et al, The FAIR Guiding Principles for scientific data management and stewardship [online] https://www.nature.com/articles/sdata201618

[5] IDS Reference Architecture Model: Industrial Data Space, Version 2.0 [online] https://www.internationaldataspaces.org/wpcontent/uploads/2018/04/InternationalDataSpacesAssociation_ReferenzArchitecture2.0.pdf.

[6] Amazon Virtual Private Cloud. User Guide [online] https://aws.amazon.com/vpc/

[7] Yuri Demchenko, Fatih Turkmen, Mathias Slawik, Defining Intercloud Security Framework and Architecture Components for Multi-Cloud Data Intensive Applications. Sixth IEEE International Workshop on Cloud Computing Interclouds, Multiclouds, Federations, and Interoperability (Intercloud 2017), In Proc. 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. Madrid, Spain, May 14-17, 2017

[8] Membrey, P., K.C.C.Chan, C.Ngo, Y.Demchenko, C. de Laat, Trusted Virtual Infrastructure Bootstrapping for On Demand Services. The 7th International Conference on Availability, Reliability and Security (AReS 2012), 20-24 August 2012, Prague.

[9] Elena Karafiloski and Anastas Mishev. Blockchain solutions for big data challenges: A literature review. In IEEE EUROCON 2017-17th International Conference on Smart Technologies, pages 763–768. IEEE, 2017.

[10] Drasko Draskovic and George Saleh. Decentralized data marketplace based on blockchain. 2017.

[11] Demchenko, Yuri, Fatih Turkmen, Ching-Hsien Hsu, Christophe Blanchet, Charles Loomis, Cees de Laat, Cloud Computing Infrastructure for Data Intensive Applications, Book chapter, In book "Big Data Analytics for Sensor-Network Collected Intelligence", Elsevier, 2017

[12] A.A. Selcuk, E. Uzun, M.R. Pariente, A reputation-based trust management system for P2P networks, Published in: IEEE International Symposium on Cluster Computing and the Grid, 2004, ISBN: 0-7803-8430-X

[13] Sherman Lee, A Decentralized Reputation System: How Blockchain Can Restore Trust In Online Markets, Aug 13, 2018 [online] https://www.forbes.com/sites/shermanlee/2018/08/13/a-decentralized-reputation-system-how-blockchain-can-restore-trust-in-online-markets/

[14] Demchenko, Y., C. de Laat, O. Koeroo, D. Groep, Re-thinking Grid Security Architecture. Proceedings of IEEE Fourth eScience 2008 Conference, December 7–12, 2008, Indianapolis, USA. Pp. 79-86. IEEE Computer Society Publishing. ISBN 978-0-7695-3535-7 / ISBN 978-1-4244-3380-3

[15] DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition). DAMA Intrnational, 2017.660 pp.

[16] DNVGL DNVGL-RP-0497, Data Quality Assessment Framework [online] http://rules.dnvgl.com/docs/pdf/dnvgl/rp/2017-01/dnvgl-rp-0497.pdf

[17] RDA Interest Group on Data Economics (IG-DE) [online] https://www.rd-alliance.org/groups/data-economics-ig