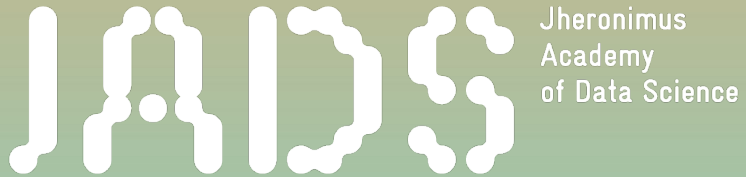
Data Market Design

Data Bites Lunch

24-05-2022

Stefan Driessen, Geert Monsieur, Willem-Jan van den Heuvel



IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 15, 2022, accepted March 12, 2022, date of publication March 22, 2022, date of current version March 30, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3161478

Data Market Design: A Systematic Literature Review

STEFAN W. DRIESSEN, (Graduate Student Member, IEEE), GEERT MONSIEUR, AND WILLEM-JAN VAN DEN HEUVEL

Jheronimus Academy of Data Science, Tilburg University, 5211 DA 's-Hertogenbosch, The Netherlands

Corresponding author: Stefan W. Driessen (s.w.driessen@jads.nl)

This work was supported in part by the European Union Horizon 2020 Project under Grant 825480.

ABSTRACT Data markets are platforms that provide the necessary infrastructure and services to facilitate the exchange of data products between data providers and data consumers from different environments. Over the last decade, many data markets have sprung up, capitalising on the increased appreciation of the value of data and catering to different domains. In this work, we analyse the existing body of scientific literature on data markets to provide the first comprehensive overview of research into the design of data markets, regardless of scientific background or application domain. In doing so, we contribute to the field in several ways: 1) We present an overview of the state of the art in academic research on data markets and compare this with existing market trends to identify potential gaps. 2) We identify important application domains and contexts where data markets are being put into practice. 3) Finally, we provide taxonomies of both design problems for data markets and the solutions that are being investigated to address them. We conclude our work by identifying common types of data markets and corresponding best practices for designing them. The outcome of this work is intended to serve as a starting point for software architects and engineers looking to design data markets.

INDEX TERMS Data market, data marketplace, data product, literature review.

I. INTRODUCTION

Nowadays, data is no longer viewed as an inept byproduct of (business) processes, but rather a valuable resource [1], [2]. A famous analogy proclaims data as the new oil,[1] and, like oil, it can be traded, processed and used in different contexts and applications. Indeed, the last decade has seen an incredible increase in both the amount of data being collected [3], [4], as well as the development of infrastructure necessary to process and share the vast amounts of collected data in new contexts [5], [6].

In the wake of these trends, many data markets have sprung up, facilitating data exchange between data providers and data consumers. These data markets capitalise on the increased appreciation of the value of data, catering to different domains (e.g., IoT [7], medical data [8] manufacturing data [9]) and contexts (e.g., national data [10], [11]). Therefore, it is not surprising that the scientific community has taken an interest in the phenomenon of data markets as well: as fig. 1 shows, there is a definite trend in scientific articles being published that have a term related to data market(places) in their title or keywords. In this work, we analyse the existing body of scientific literature on data markets to provide the first

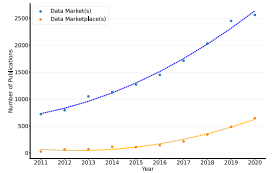
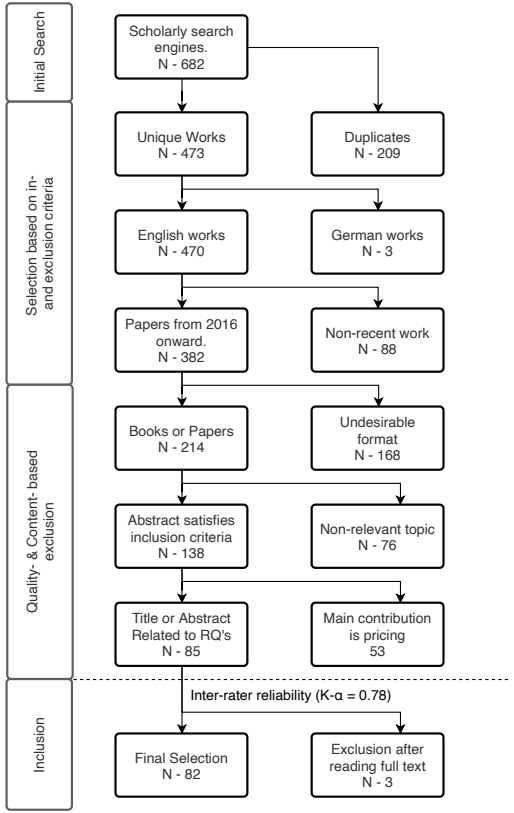


FIGURE 1. Research Trends for Data Markets, an exponential growth is observed. Source: Number of results for each query in google scholar.

The associate editor coordinating the review of this manuscript and approving it for publication was Kostas Kolomvatsos.

[1]The Economist, "The world's most valuable resource is no longer oil, but data," may 2017



[2]

Our contributions

1. An overview of the state of the art on data markets in literature and comparison with industrial trends.
2. Investigation of application domains and impacts of the domain on the design and implementation of data markets.
3. Taxonomies of problems and solutions for data market design and implementation

[3]

The Economist

Menu | Weekly edition | Search

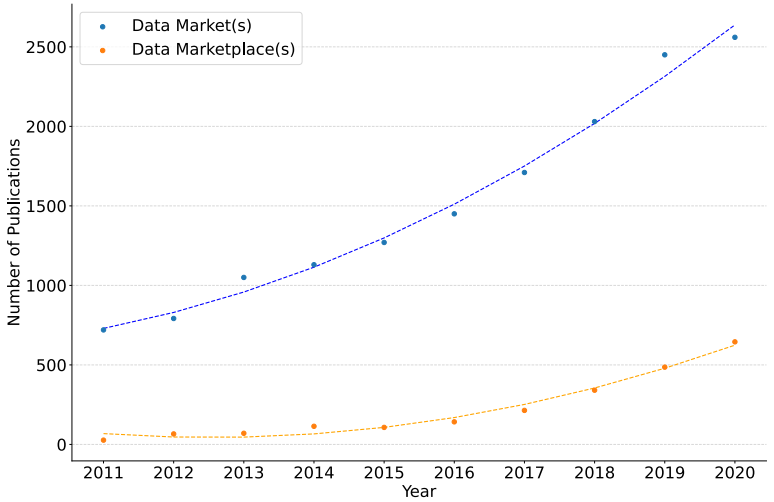
Leaders | Regulating the internet giants

The world's most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules

[4]

A data market is a platform that provides the necessary infrastructure and services to facilitate the exchange of data products between data providers and data consumers from different environments.

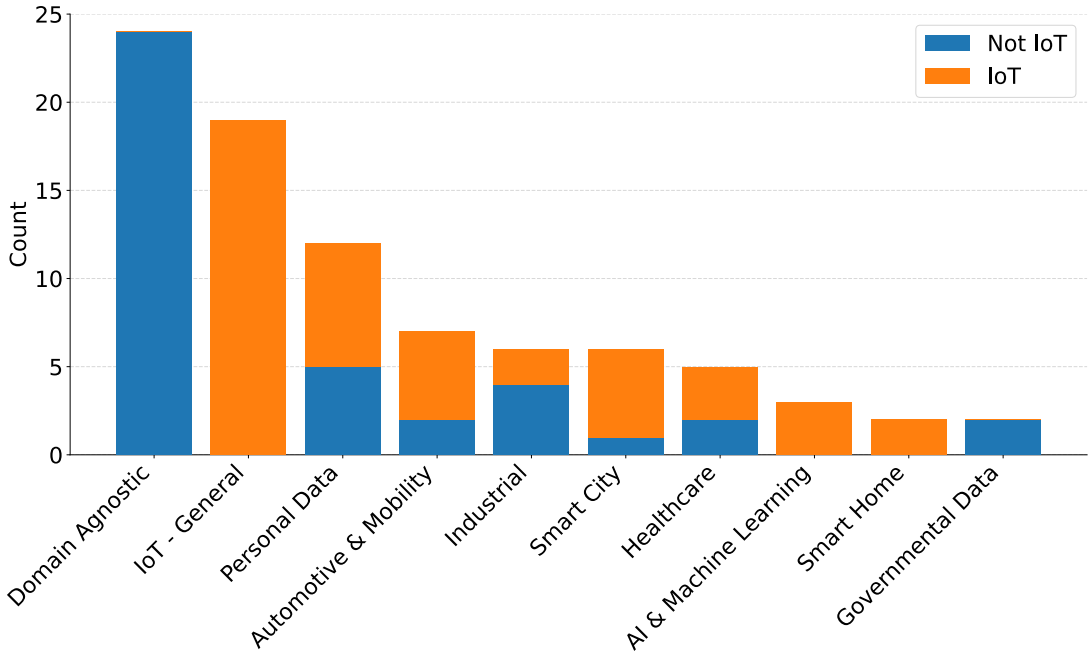


[5]

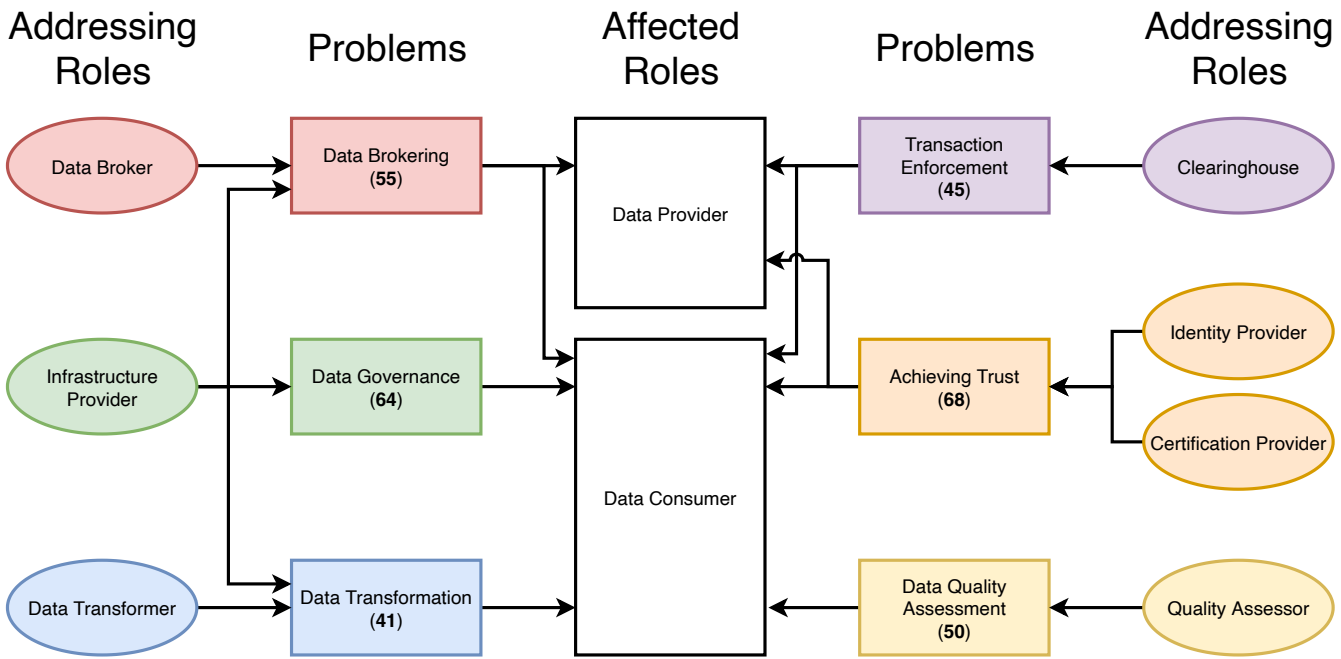
Examples

- Social Media Platforms consume your data in exchange for services
- European Initiative GAIA-X and Nokia Data Marketplace facilitate B2B data exchange.
- Decentralised Data Marketplaces allow individuals to exchange data.
- Internal Data Marketplaces facilitate data exchange inside organisations.

[6]


Not IoT
IoT
Count
25
20
15
10
5
0
Domain Agnostic
IoT - General
Personal Data
Automotive & Mobility
Industrial
Smart City
Healthcare
AI & Machine Learning
Smart Home
Governmental Data

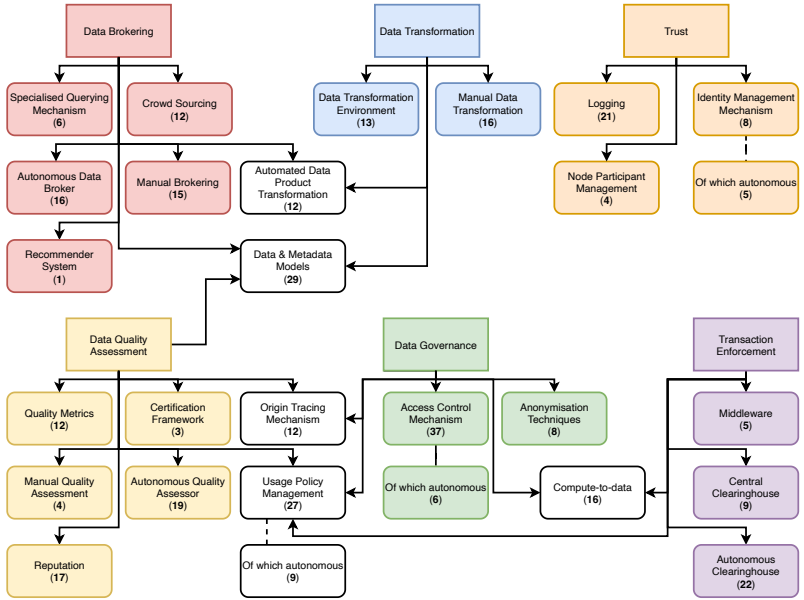
[7]


Addressing Roles
Problems
Affected Roles
Problems
Addressing Roles
Data Broker
Infrastructure Provider
Data Transformer
Data Brokering (55)
Data Governance (64)
Data Transformation (41)
Data Provider
Data Consumer
Transaction Enforcement (45)
Achieving Trust (68)
Data Quality Assessment (50)
Clearinghouse
Identity Provider
Certification Provider
Quality Assessor

[8]

Interesting Observations

- Mix of technical and organisational solutions
- Some solutions are more important than others.
- Blockchain and autonomous actors are popular for decentral platforms.



[9]

The five data market archetypes

	Generalist	Specialist	Industry Exchange	Enabler	Aggregator
Defining Characteristics	- Heterogeneous data - Domain agnostic - Many-to-many matching	- Homogeneous data - Single Domain	- Providers & Consumers are companies/organisations - Data from one domain but heterogeneous structure - Decentral architecture - Consortium-owned - Specialised Software	- Many-to-many matching - Small data products	Many-to-one + one-to-many matching - Extensive control of all processes - Monopoly
Central Roles	- Data Broker - Clearinghouse	- Domain Dependent - Data Transformer	- Infrastructure Provider - Identity Provider - Certificate Provider	- Clearinghouse - Infrastructure Provider	- Data Transformer
Critical Problems	- Data Brokering - Transaction Enforcement	Domain Dependent - Data Transformation	- Data Governance	- Data Transformation - Transaction Enforcement	- Data Transformation - Data Governance
Typical Solutions	- Central Clearinghouse - Specialised Querying Mechanism - Manual Actors	- Quality Metrics - Automation - Compute-to-data	- Identity Management Node Participation Management - Certification Framework - Usage Policies	- Middleware - Central/automated clearinghouse - Manual transformation - Transformation Environment	- Anonymisation Techniques Data & - Metadata Models - Usage Policies
Example Works	Hayashi & Ohsawa [131], Spiekermann [15], Nguyen & Won [154]	Ahmed & Shabani [8], Sakr [65], Sajan et al. [52], Alsharif & Nabil [90]	Llewelyn et al. [50], Munoz-Arcentales et al. [111], Pillman et al. [82], Radhakrishnan & Das [96]	Cao et al. [11], Jeong et al. [87], Figueredo et al. [88], Perera et al. [20]	Eng et al. [60], Niu et al. [36], Thomas & Leiponen [12], Liang et al. [155]

[10]

JADS

Jheronimus
Academy
of Data Science

The founders of JADS

TU/e Technische Universiteit Eindhoven University of Technology

Tilburg University

's-Hertogenbosch

Provincie Noord-Brabant