SEMCON

Semantic Containers for Data Mobility

Programm / Ausschreibung	IKT der Zukunft, IKT der Zukunft, IKT der Zukunft - Ausschreibung Data Market AT	**Status**	abgeschlossen
Projektstart	01.10.2018	**Projektende**	30.06.2019
Zeitraum	2018 - 2019	**Projektlaufzeit**	9 Monate
Keywords	data provisioning provenance container billing		

Projektbeschreibung

Das wirtschaftliche Potential von Daten als Treiber für Zukunftstrends wie maschinelles Lernen, AI und Big Data und die bedeutende Rolle in modernen Wirtschaftsstrukturen ist heute außer Frage gestellt. Und obwohl Daten oft als das "Öl des 21. Jahrhunderts" bezeichnet werden, gibt es doch einen bedeutenden Unterschied: Daten werden wertvoller je mehr sie genutzt und verknüpft werden. Doch trotz zahlreicher Initiativen und Anstrengungen Datenmärkte in einzelnen Domänen zu etablieren, zögern vielen Firmen und Organisationen ihre Daten aktiv miteinander auszutauschen und zu vernetzen.

Das ist insbesondere deshalb bedauernswert, da der Aufwand zum Austausch von Daten gering, der ökonomische Nutzen dagegen aber unverhältnismäßig groß ist. Trotzdem haben sich aber nur wenige erfolgreiche Datenmärkte etabliert, das sich zum einen auf technische Herausforderungen zurückführen lässt, zum anderen aber auch offenen Fragen der kommerziellen und rechtlichen Umsetzung beim Verkauf von Daten geschuldet ist. Viele dieser Fragestellungen sind auch darauf zurückzuführen, dass der Verkauf von Daten als ein "Kontrollverlust" angesehen wird. Gleichzeitig beklagen potentielle Datenkäufer Schwierigkeiten beim Umgang mit externen Datenquellen auf Grund von uneinheitlicher Datenqualität und schwierigem Nachweis der Herkunft bzw. eventuell bereits durchgeführten Verarbeitungsschritten.

Das beantragte Projekt adressiert diese Probleme indem eine einheitliche Infrastruktur zur Bereitstellung von Daten durch sogenannte "Semantische Container" entwickelt wird. Dieses Konzept erlaubt Datenanbietern eine effiziente und kontrollierte Weitergabe von Datensätzen mit Abrechnungs-Service und einschränkbaren Verwendungszweck der Daten. Den Empfängern von Daten werden klar definierte Mechanismen für den Zugriff und der Integration unterschiedlicher Datenquellen geboten, die eine validierbare und reproduzierbare Verwendung erlauben. Durch die "Containerisierung" von Daten, einer semantischen Beschreibung des Inhalts und der erlaubten Verwendung, und einheitlichem Zugriff entsteht ein "Wirtschaftsgut Daten" mit klar definiertem und nachweisbaren Inhalt, Qualität, Nutzungsrecht und Preis.

In diesem Projekt wird das vorgestellte Konzept anhand eines vollständigen Datenflusses von Datenanbietern (ZAMG und EODC) via Daten Markt Austria zu Datenempfängern (Nutzern von OwnYourData.eu) prototypisch abgebildet und mit Hilfe von LinkedWidgets.org interaktiv visualisiert. Dazu wird Docker als etablierte Container-Technologie eingesetzt, die bereits

jetzt Mechanismen zur Verteilung und zum Einsatz von komplexen Softwaretechnologien bereitstellt und die Basis zum Auffinden, Verteilen und Integrieren von Datensätzen liefert. Um einen Markt um das Wirtschaftsgut Daten aufzubauen, müssen die zugehörigen Regeln semantisch beschrieben, durch kryptographische Methoden abgesichert und mit Blockchain Technologie unveränderbar gemacht werden. Ein vollständiger Audit-Trail von Datenquellen und allen Verarbeitungsschritten erlaubt sowohl lückenlose Dokumentation als auch automatisierte Validierung und Reproduzierbarkeit.

Abstract

Today, the economic potential of data as a driver for technologies such as machine learning, AI, and big data analytics and its role as a cornerstone of modern economies is largely undisputed. Although data is often likened to scarce natural resources such as oil, however, a major difference is that data can become more useful the more it is used. Nevertheless, despite open data initiatives and some efforts to create markets for data in particular domains, many data are monopolized today.

This is unfortunate, given that data can be replicated and moved at very low cost and generate great economic value when shared. Anyhow, only a few successful data markets have been established, which may be attributed partly to technical issues and partly to business rationale and questions over the viability (and sometimes legality) of business models built around the sale of data. A lot of these issues can be attributed to the problem that data providers selling data (rather than just access to data) largely have to give up control under which terms the data will be used. On the other hand, from a data consumer perspective, both technical challenges and limited trust in the quality, completeness, and origin of the available data have contributed to the restricted adoption of data markets today.

The proposed project tackles these challenges by developing a standardized infrastructure for data provisioning based on the concept of Semantic Containers. This concept allows data providers efficient distribution without giving up control over its usage and monetization while providing data consumers with efficient and well-managed mechanisms to obtain and integrate data in a trustworthy and reproducible manner. By packaging data and processing capabilities into reusable containers, describing the semantics of the content and permissible usage, and providing uniform interfaces, a data set becomes a commodity with well-defined content and quality, as well as clear ownership and a price tag.

In this project, a complete data flow from a data provider (ZAMG and EODC) via the Data Market Austria to data consumers (users of OwnYourData.eu) will be implemented as a experimental prototype in an interactive environment (based on semantic mashup concepts developed in the context of LinkedWidgets.org). To this end, we will leverage existing container technologies such as Docker, which already provide scalable mechanisms for deploying complex software assemblies and use them as a foundation for an infrastructure for data discovery, provisioning, and integration. To create a commodity market around data requires a given set of rules that will be captured in semantic descriptions and enforced through cryptographic methods for proving ownership, and blockchain technology to guarantee immutability. Complete audit trails of data sources and processing steps provide gapless provenance and allow full reproducibility.

Projektkoordinator

- Verein zur Förderung der selbstständigen Nutzung von Daten

Projektpartner

- Technische Universität Wien