Data Catalogs
CDQ Service Portfolio

St. Gallen, 2019

Jan Oliver Reh
Partner

CDQ
Corporate Data Quality

Data Catalogs - Introduction

CDQ Market Study on Data Catalogs

CDQ Consulting Services

About CDQ

The vision: the data-driven enterprise

PEOPLE

DATA

ORGANISATION

DATA CITIZENSHIP

With data as an asset, **every employee** needs to be **data-savvy** - not only data managers and data scientists.

DATA DEMOCRATIZATION
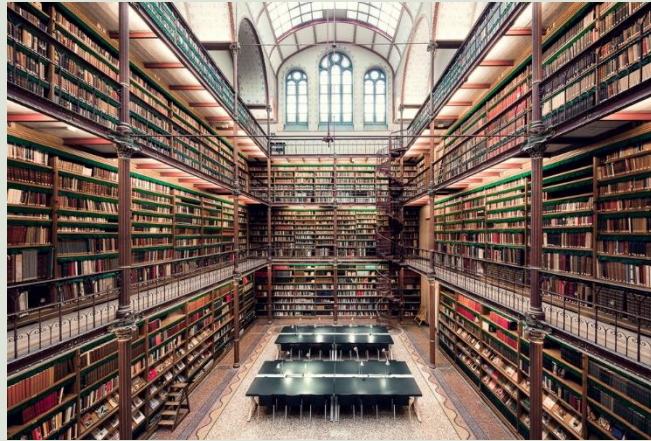
"Data democratization means that **everybody has access to data** and there are no gatekeepers that create a bottleneck at the gateway to the data."[1]

DATA SHARING CULTURE

Sharing data is a data citizen's **key responsibility** and provides benefits through additional uses of the data in other areas of the company.

[1]What is Data democratization? A Super Simple Explanation And The Key Pros And Cons https://www.forbes.com/sites/bernardmarr/2017/07/24/what-is-data-democratization-a-super-simple-explanation-and-the-key-pros-and-cons/#48ad53a06013

Typical questions from data users:

- Where can I find data?
- Who owns them?
- Who maintains them?
- Who can help me explain them?
- How good are they?
- Are they still up-to-date?

Business goals to be achieved with a data catalog

- 🔍 Identify and find data
- 🛠️ Establish a solid data governance foundation
- 🕐 Raise data awareness
- ✓ Increase transparency
- ⚙️ Control and develop data models
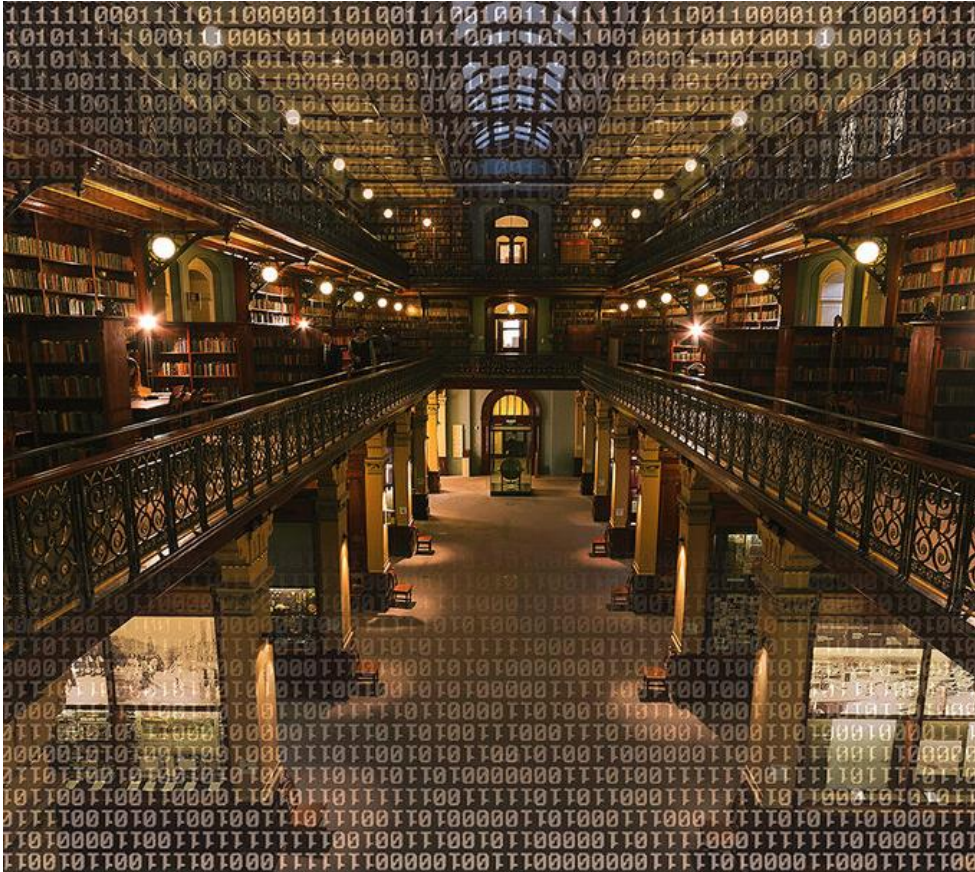- 📚 Document exisiting data
- ⊕ Maintain a single point of truth

Picture: unsplash.com

The data catalog links data supply and demand

»A Data Catalog is an integrated platform for data curation; matching data supply and demand. It offers users functions to register data; to retrieve and use data; and to assess and analyze data.«

The need for a data catalog

That's Steve

- Steve used to work as a data scientist for several years
- He **spends up to 80%** of his time hunting for the relevant datasets and prepare them for analytic projects
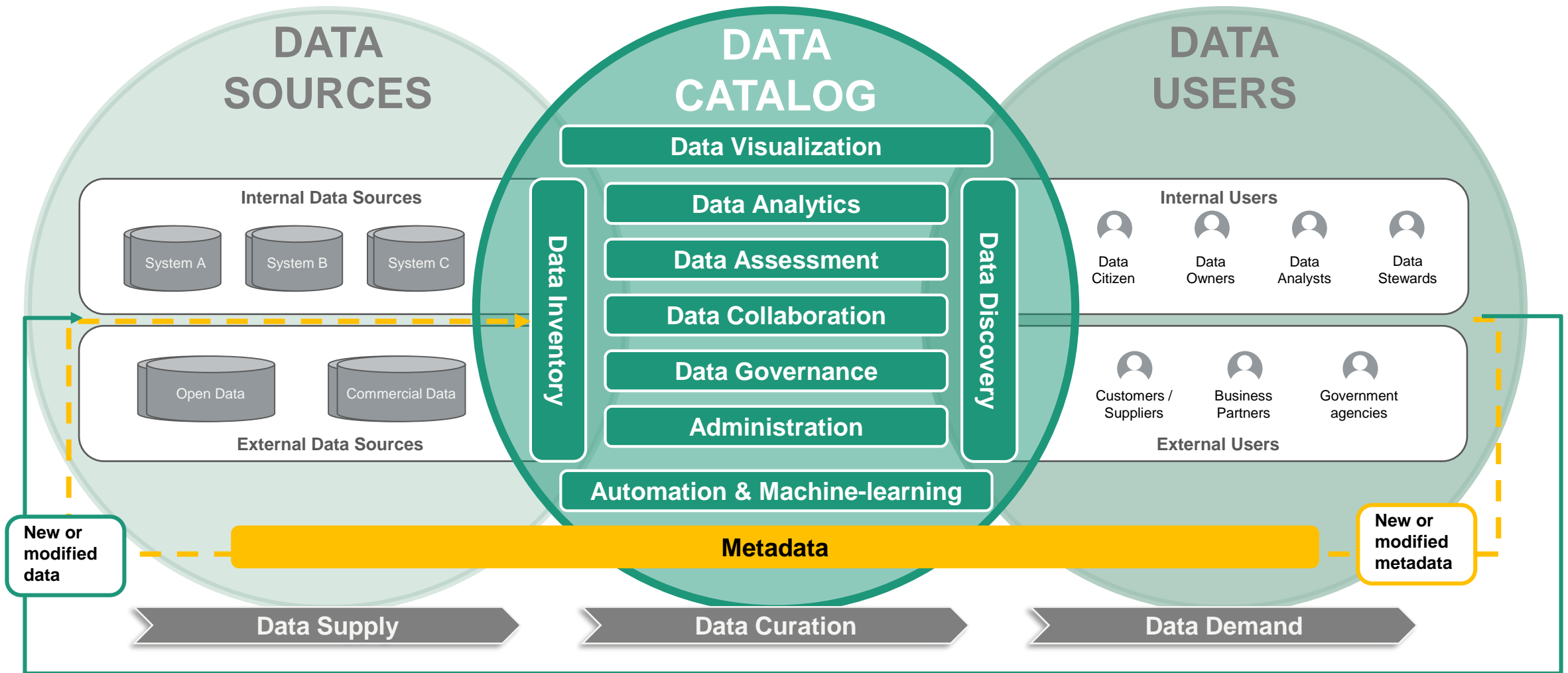- **Only 20%** of his costly time is used for **data consuming**, e.g. gain new insights

up to
80%

Data Search & Preparation

20%

Data Consuming

Typical roles and user groups

"As business increasingly relies on data, I need an **overview of all of our enterprise's data assets in order to assess their value**, to track the progress made and make decisions."

"I am responsible for the data in my domain. As other departments rely on my data, I need to **take care on data quality** and to **control the access** to it.**"**

"Working with data is part of my **everyday work**. I depend on data provided by other departments".

"It is my job to **identify new use cases** for data analytics, **develop proof-of-concepts**. I use **all data sources** available to me (internal & external) and implement new technologies to provide **business value**."

"With high pressure on achieving GDPR compliance **I need to be able to identify and flag all fields in all systems that contain personal** information and control and protocol access to them."

"I need to make sure that the data I am responsible for corresponds to the agreed data quality. Therefore, I need to **assess the data quality continuously** and implement improvement measures."

"It is my job to **implement business requirements as software solutions**. A lot of my work is to to align my concepts with **enterprise and data architecture** as well as security."

"I define **how data should be stored and consumed** across enterprise systems. For this, I need to **provide data models and metadata to different stakeholders."**

Chief Data Officer · **Data Owner** · **Data Protection Officer** · **Data Citizen** · **DATA CATALOG** · **Data Steward** · **Data Scientist** · **Data Architect** · **Solution Architect**

The data catalog links data supply and demand

The market study evaluates data catalogs along the CDQ capability maps

DATA VISUALIZATION

DATA INVENTORY

- Data Registration
- Metadata Management
- Business Glossary
- Data Dictionary
- Data Lineage
- Data Access
- Upload / Link Content

DATA ANALYTICS

Documentation / Data Stories	Data Application Repository
Scripting / Data Pipeline Mgmt.	Data Lake Status and Health Monitoring

DATA ASSESSMENT

Data Usage	Data Risk	Data Profiling
Data Quality	Data Valuation	Benchmarking

DATA COLLABORATION

Tagging	Following / Updates	User Communication
Rating	Commenting	Sharing

DATA GOVERNANCE

Workflows	Rules & Policies	Handling Sensitive Data
Roles & Responsibilities	Data Authorizations	

ADMINISTRATION

Configuration	User Management

DATA DISCOVERY

- Search
- Recommendation
- Data Subscription
- Data Delivery

AUTOMATION & MACHINE-LEARNING

Data Supply → Data Curation → Data Demand

Market study: Data Catalogues

Product	DATA INVENTORY	DATA ANALYTICS	DATA COLLABORATION	DATA ASSESSMENT	DATA GOVERNANCE	DATA DISCOVERY	DATA VISUALIZATION	AUTOMATION & ML
Adaptive Metadata Manager™	●	○	●	◐	◕	◔	●	◐
Alation Data Catalog	●	●	◔	◐	●			
Cambridge Semantics Anzo® Smart Data Lake 4.0	●	●	◔	◔	◕			
Collibra Data Governance Center	●	◔	●	◐	●			
Datum Information Value Management®	●	○	◐	◐	◕			
IBM Watson® Knowledge Catalog	●	◔	◐	◐	◕			
IBM InfoSphere IGC	●	○	○	◐	●			
Informatica Enterprise Data Catalog	●	◔	◐	◐	●			
Informatica Axon Data Governance	◔	◔	○	◐	●			
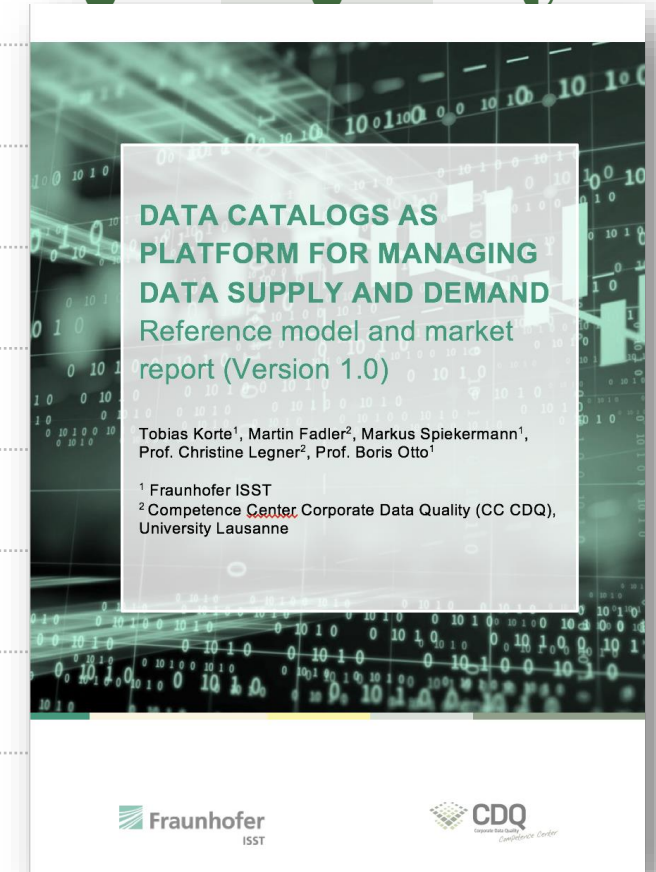
DATA CATALOGS AS PLATFORM FOR MANAGING DATA SUPPLY AND DEMAND
Reference model and market report (Version 1.0)

Tobias Korte[1], Martin Fadler[2], Markus Spiekermann[1], Prof. Christine Legner[2], Prof. Boris Otto[1]

[1] Fraunhofer ISST
[2] Competence Center Corporate Data Quality (CC CDQ), University Lausanne

Market study Data Catalogues

Product	DATA INVENTORY	DATA ANALYTICS	DATA COLLABORATION	DATA ASSESSMENT	DATA GOVERNANCE	DATA DISCOVERY	DATA VISUALIZATION	AUTOMATION & ML
Oracle Enterprise Metadata Mgmt.	●	○	◐	○	◐	◔	●	◐
Podium Data Market Place	●	◐	◔	◐	●			
SAP Information Steward	◔	○	○	◔	●			
SAP Data Hub	◔	◔	○	◐	◔			
Waterline Smart Data Catalog	●	◔	◐	◔	●			
Zaloni Data Management Platform	●	◐	●	◔	●			
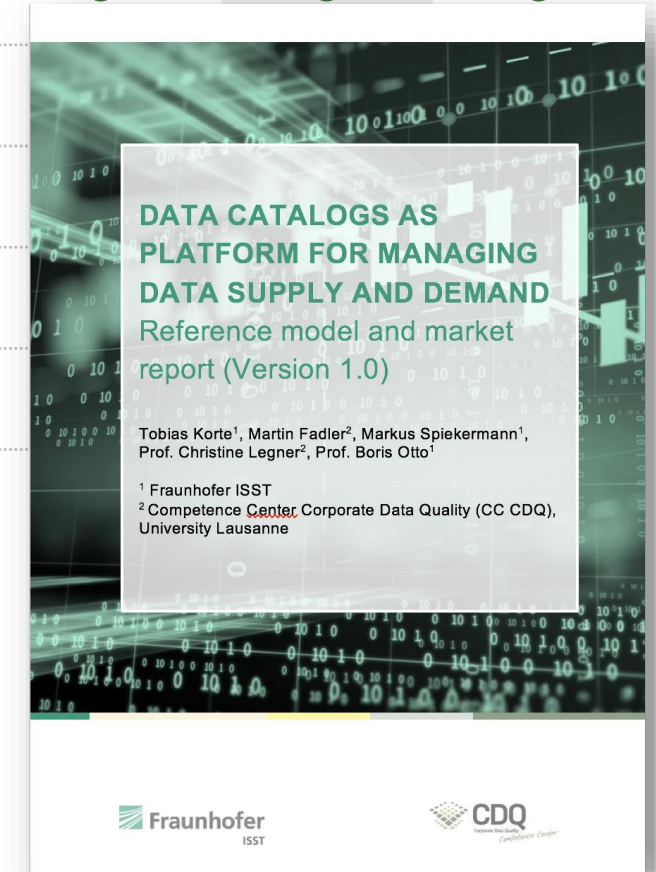
DATA CATALOGS AS PLATFORM FOR MANAGING DATA SUPPLY AND DEMAND
Reference model and market report (Version 1.0)

Tobias Korte[1], Martin Fadler[2], Markus Spiekermann[1], Prof. Christine Legner[2], Prof. Boris Otto[1]

[1] Fraunhofer ISST
[2] Competence Center Corporate Data Quality (CC CDQ), University Lausanne

Fraunhofer ISST

CDQ

A data catalog implementation must ensure to deliver immediate business value
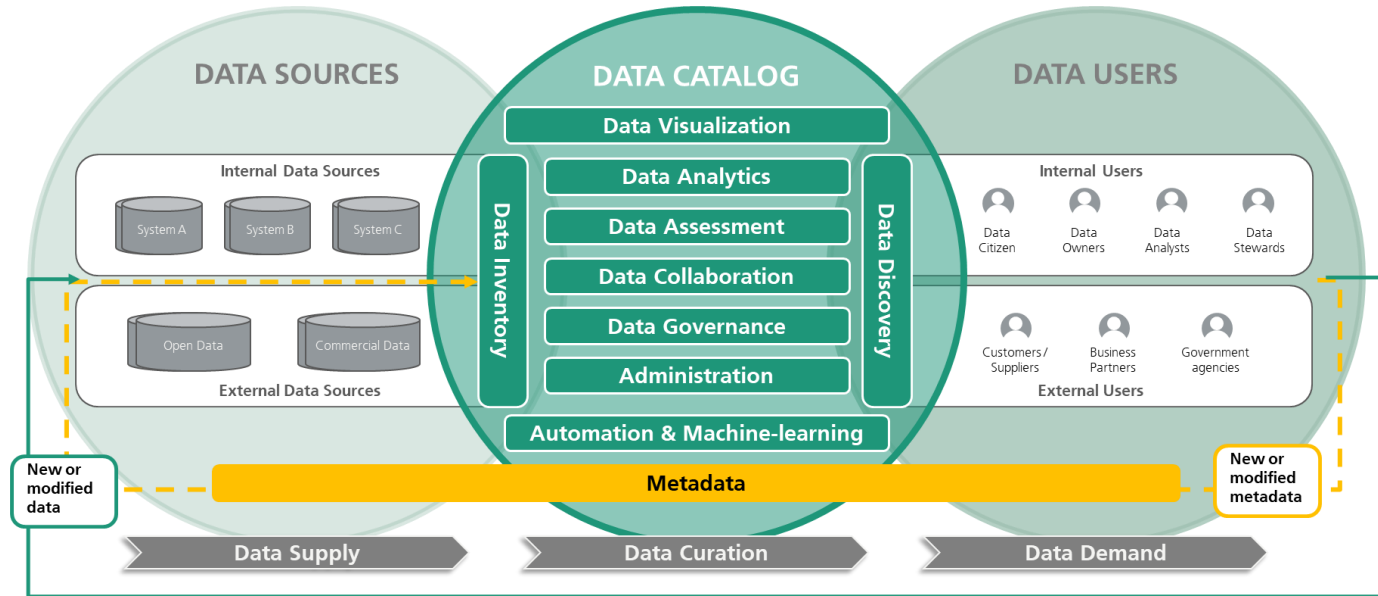
- Provide segment-specific business value.

- Serve multiple "customer segments" with different goals.

- Quickly achieve critical mass (content, user base).

→ Build the data catalogue along clearly defined business cases

CDQ supports in all phases of a data catalog implementation

Concept
- Identification of business and data management goals
- Definition of required data catalog capabilities
- Development of data catalog strategy

Design
- Metamodel design
- Definition of operating model
- Tool Selection

Implementation
- Business value assurance
- Tool implementation (e.g. Collibra, Semantic Wiki)
- Conceptual and logical data modeling
- Implementation of governance workflows
- Development of data quality rules and dashboards

Agenda

Data Catalogs - Introduction

CDQ Market Study on Data Catalogs

CDQ Consulting Services

About CDQ

CDQ IS
THE LEADING
DATA MANAGEMENT
CONSULTANCY
IN THE INDUSTRY

CDQ AG
Sharing Data Excellence

Community services: We host the most mature community for data management in Europe, drive innovation in research and practice and train CDOs, data managers and data scientists.

Consulting services: We successfully completed more than 200 data management projects on strategy, governance and architecture.

Cloud services: We offer cutting-edge business data services (e.g. data health checks, GDPR assessments, address validations and enrichments).

We started as a spin-off from the University of St. Gallen

2006

Creation of the **Competence Center Corporate Data Quality** at the University of St. Gallen

Founding partners are Bayer, Daimler, Deutsche Telekom and IBM

2009

Integration of CDQ's **data quality mgmt. approach** into **EFQM standard** (European Foundation for Quality Mgmt.)

2011

Competence Center Corporate Data Quality grows to **over 10 members**

2013

Foundation of the **CDQ Academy** to train future data managers

Initiation of **CDQ Good Practice Award**

2015

Start of the **Corporate Data League** (Collaborative Business Partner Management)

2016

Competence Center Corporate Data Quality grows to **over 15 active members**

CDQ's data management consulting services deliver impact and create business value

Business Impact

We enable digitization initiatives, improve business process efficiency and reduce compliance risks through effective data management.

Data Impact

We clean up, de-duplicate and enrich your business partner data based on our cutting-edge data quality services.
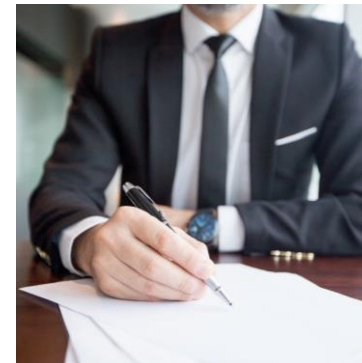
People Impact

We train the next generation of Chief Data Officers and Data Managers based on CDQ's Data Excellence Framework.

Knowledge Impact

We run a network of firms that exchange data management best practices and drive innovation together with the University of Lausanne.

Our sole focus on data management makes us unique in the market

Data Management Strategy

If your data needs to support digital initiatives, process efficiency or regulatory compliance, CDQ helps you define the right data management strategy to support your business goals.

Operating Model

If you want to define or optimize your operating model for data mgmt., CDQ can design the required processes, define central and decentral responsibilities and determine an optimal cost structure.

Data Quality Management

If poor data quality limits your analytics' ambition or creates unnecessary costs and risks in your business processes, CDQ can help you define the right data quality mgmt. approach and support data cleansing.

Data Governance

If your data creation and maintenance activities produce inconsistent data of poor quality, CDQ supports the definition of guidelines to govern your data effectively across organizational units and systems.

Data Architecture & Tools

If you need to design or optimize the data flows across your IT-systems (e.g. CRM, ERP, MDM), CDQ develops the required data models and lifecycles to ensure that you have the right data in the right place.

Business Data Services

If you want to identify duplicates in your business partner data, validate postal addresses or enrich geo coordinates, CDQ can help you cleanse and enrich your data based on our state-of-the-art data quality services.

CDQ's state-of-the-art knowledge base in data management is the foundation for our impact advisory

Data Managers Trained

400+

Company Case Studies

300+

Consulting Projects

200+

Maturity Assessments

80

CC CDQ Workshops

60+

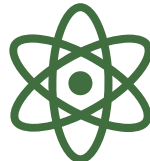**Scientific Publications**

60

CDQ Employees

40+

PhD Dissertations

12

Research Projects

7

Companies are trusting CDQ for years

"As a trustful and reliable partner, CDQ did an excellent job to support the development and establishing our master data management strategy."

Roger Kipfer, Head of Data Quality Management Swisscom

"CDQ has supported us with excellent expertise to build up our Global Master Data Management organization. We also appreciate the active exchange at the Competence Center."

Timo Neumann, Head of Master Data Governance ZF Friedrichshafen

"CDQ helped us to set up data quality management in our organization. They have exceeded our expectations."

Ben Hallez, Head Center of Excellence for Master and Meta Data Management Bayer HealthCare

Overview comprises both current and former partner members of the Competence Center Corporate Data Quality

We are looking forward to hearing from you!

Jan Oliver Reh

CDQ AG
Partner
oliver.reh@cdq.ch
+49 177 272 1911

www.cdq.ch

www.linkedin.com/company/cdq-ag

www.twitter.com/cdq_ag

PMI's Data-Driven Transformation

Tobias Pentek, Martin Fadler, Prof. Dr. Christine Legner
CC CDQ Case Study
2018

University of St.Gallen | UNIL | Université de Lausanne | CDQ

Boris Otto
Hubert Österle

Corporate Data Quality

Voraussetzung erfolgreicher Geschäftsmodelle

OPEN | Springer Gabler

MANAGING DATA AS AN ASSET WITH THE HELP OF ARTIFICIAL INTELLIGENCE

Martin Fadler, Prof. Dr. Christine Legner
Competence Center Corporate Data Quality (CC CDQ) in collaboration with SAP SE

RESEARCH PARTNERS
UNIL | Université de Lausanne | Universität St.Gallen | CDQ

DATA CATALOGS

Integrated Platforms for Matching Data Supply and Demand

Reference Model and Market Analysis
(Version 1.0)

Tobias Korte[1], Martin Fadler[2], Markus Spiekermann[1], Prof. Dr. Christine Legner[2], Prof. Dr.-Ing. Boris Otto[1]

[1] Fraunhofer Institute for Software and Systems Engineering (ISST)
[2] Competence Center Corporate Data Quality (CC CDQ), University of Lausanne

FRAUNHOFER VERLAG

Fraunhofer ISST | CDQ

Data valuation

Andreas Zechmann, Prof. Dr. Christine Legner, Dr. Dimitrios Gizanis

White Paper (draft)
April 2016 – V 1

University of St.Gallen | UNIL | Université de Lausanne | CDQ