Sourcing and managing external data

External data is an underexploited resource

Open or freely available data from the Web or sourced from data providers is growing at a rapid pace. Despite the availability of external data, Competence Center Corporate Data Quality (CC CDQ) research shows that the majority of companies do not systematically screen for relevant external data and lack an overview of which external data sets have already been sourced and by whom.

If this opportunity is seized, external data can complement internal data and help to:

1. enhance analytics (e.g., with market trends, or benchmarks),
2. optimize business processes (e.g., with geolocation, weather, or traffic data),
3. reduce data maintenance efforts, and improve data quality (e.g., with reference data),
4. innovate and create new data-driven services (e.g., through customer insights).

DEFINITION   "External data refers to any type of data that has been captured, processed, and provided from outside the company." CC CDQ

Four external data types

Open Data

Data that is freely available from governments, NGOs and companies, which can be used as well as republished by everyone without restrictions from copyright or patents.

Examples: statistical datasets from national statistical offices or the OECD, traffic information from transport companies, maps from OpenStreetMap

Paid Data

Datasets that are provided at certain cost (e.g., pay for use, subscriptions, freemium) by specialized data providers or brokers, or from data marketplaces.

Examples: market research data from Nielsen, financial data from Bloomberg or Thompson Reuters, company information from Dun & Bradstreet or Bureau van Dijk

Shared Data

Data that is exchanged and shared between companies in business ecosystems, either bilaterally or through platforms.

Examples: product data shared through GS1 data pools, business partner data shared through the CDQ Data Sharing community, aircraft data shared through Skywise
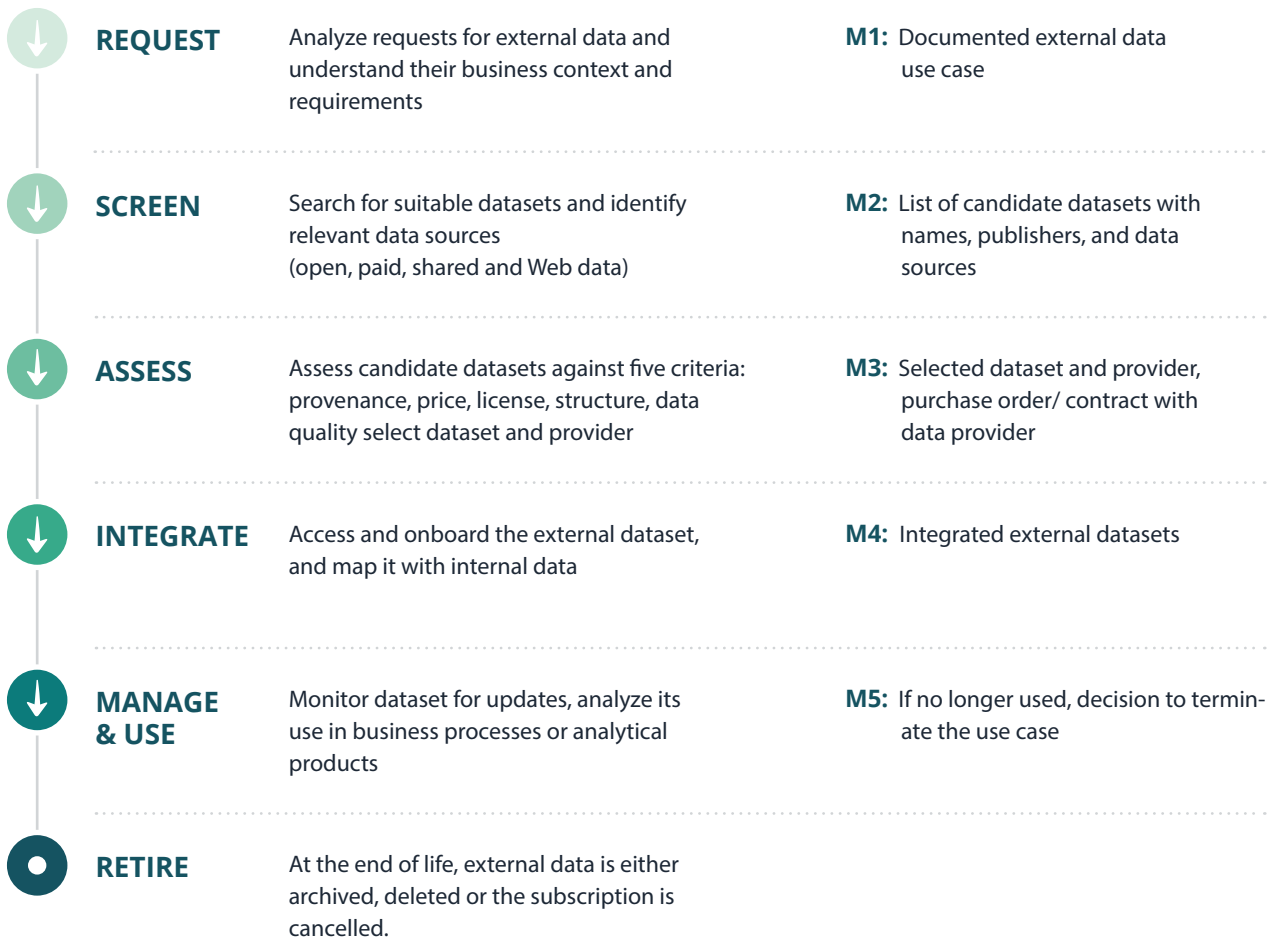
Web Data

Content that companies or users share on their websites or on social media platforms, including the metadata (e.g., location, time, language, biographical data).

Examples: product reviews, Twitter feeds, Facebook user profiles and posts

Six steps for sourcing and managing external data

A systematic approach to sourcing and managing comprises six core phases (M=milestone):

REQUEST
Analyze requests for external data and understand their business context and requirements

M1: Documented external data use case

SCREEN
Search for suitable datasets and identify relevant data sources
(open, paid, shared and Web data)

M2: List of candidate datasets with names, publishers, and data sources

ASSESS
Assess candidate datasets against five criteria: provenance, price, license, structure, data quality select dataset and provider

M3: Selected dataset and provider, purchase order/ contract with data provider

INTEGRATE
Access and onboard the external dataset, and map it with internal data

M4: Integrated external datasets

MANAGE & USE
Monitor dataset for updates, analyze its use in business processes or analytical products

M5: If no longer used, decision to terminate the use case

RETIRE
At the end of life, external data is either archived, deleted or the subscription is cancelled.

KEY TAKE-AWAYS

1 Incorporating external data not only can enhance internal data, but it can additionally optimize business processes and improve innovation.

2 Instead of using external data ad-hoc, companies need to develop their own standardized approach to data sourcing.

3 Six steps for sourcing and managing external data should be embedded into the data management framework.

Further readings

Krasikov, P., Eurich, M., & Legner, C. (2022)
Unleashing the Potential of External Data, a DSR-based Approach to Data Sourcing.
In Proceedings of European Conference on Information Systems

Krasikov, P., Legner, C., & Eurich, M. (2021)
Sourcing the Right Open Data: A Design Science Research Approach for the Enterprise Context.
In Proceedings of International Conference on Design Science Research in Information Systems and Technology
(pp. 313-327). Springer, Cham.

Krasikov, P., Obrecht, T., Legner, C., & Eurich, M. (2020)
Open Data in the Enterprise Context:
Assessing Open Corporate Data's Readiness for Use. In International Conference on Data Management
Technologies and Applications (pp. 80-100). Springer, Cham.

Contact

Dr. Markus Eurich	Prof. Dr. Christine Legner	Pavel Krasikov
markus.eurich@unil.ch	christine.legner@unil.ch	pavel.krasikov@unil.ch

About the CC CDQ

The Competence Center Corporate Data Quality (CC CDQ) is a research consortium and Europe's leading expert community for data management. Members of the CC CDQ benefit from a cross-industry network, knowledge sharing, and research & co-innovation.

We bring together data management experts from practice and academia: Corporate partners are more than 15 renowned companies from different industries. The research team is located at the Faculty of Business and Economics (HEC – University of Lausanne) and headed by Prof. Dr. Christine Legner. The CC CDQ was founded in 2006 at the Institute of Information Management (IWI – University of St. Gallen).

Today, it is operated by CDQ AG.

CDQ
COMPETENCE CENTER