



Data Sharing

The best way to better data

Data quality is fitness for use. In business context, that means that data must meet business requirements in terms of accuracy, completeness, and up-to-date-ness. Data is *fit for use* if the data-using business processes run smoothly.

For customer and vendor master data, *fitness for use* can be translated to *know your customer* and *know your vendor*. The better your master data represents real-world customers and vendors, the better your procurement, marketing, sales, and logistic objectives are met.

- **Predict delivery shortages:** Knowing production locations of your vendors (today's locations and not addresses from 2 years ago) helps to understand impacts of natural disasters or political unrest on your supply chain.
- **Post-merger success:** Integration of clean customer and vendor data is much easier than being faced with duplicates, diverse address formats, or outdated identifiers.
- **Social compliance:** Identify and replace rotten apples in your supply chain, before press finds out.
- **Sanctions and embargos:** Be aware of all regulation, globally, and identify affected customers before custom authorities do.
- **Avoid payment fraud:** Know the effective bank accounts of your vendors and do not transfer money to fraudsters. Share uncovered attacks with peers to fight back.

In this article, we present *Data Sharing*, a new data management approach to increase your data quality at lower costs. You can calculate your own *Data Sharing* business case on our website, www.cdq.ch/businesscase.

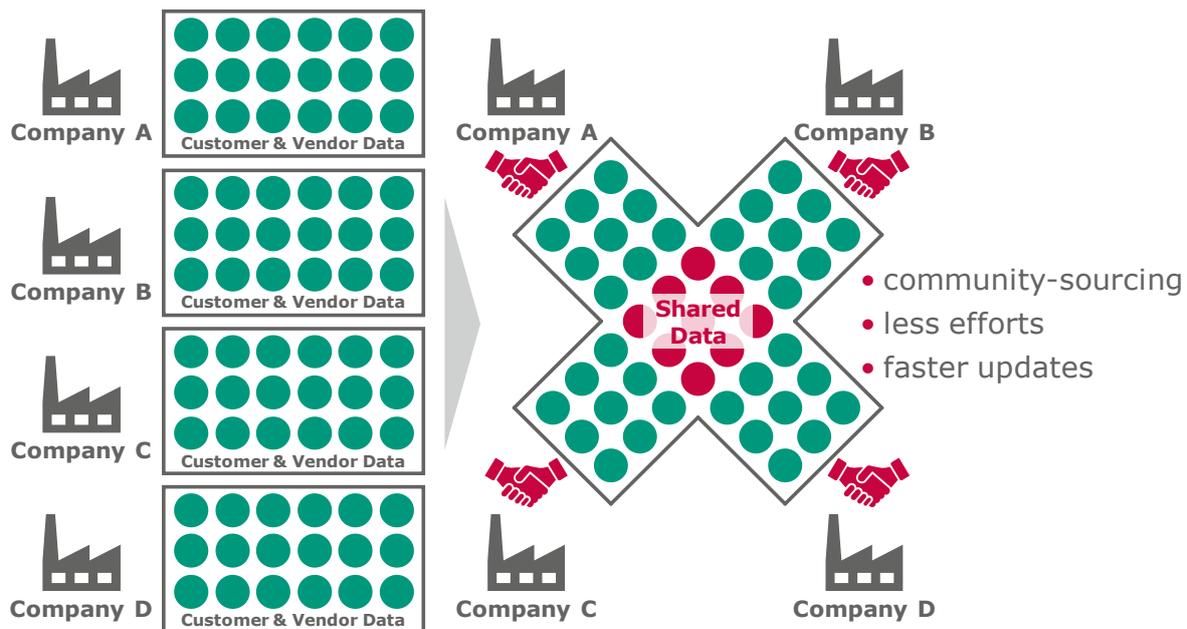


Figure 1: Better data with less effort by community-sourcing

Data Sharing can reduce data management efforts

If you (Company A) have the same customer as another company (Company B), the *know your customer* requirement is the same for you and Company B, at least for many use cases. Beside individual information like payment terms, most attributes do not depend on a specific business relation, e.g. legal name, address, or tax number. Also, further information such as financial stability and social compliance refer to the real-world status of a business partner, and thus are right or wrong, independently from Company A or Company B.

The availability of an external reference (or *truth*) bears the potential to align, to collaborate, to share efforts of managing business partner data. This is what *Data Sharing* is about: A trusted network of companies who manage business partner data as a shared resource.

To better explain the value of *Data Sharing*, the following chart illustrates the overall effect along the three key dimensions of data management: Time, quality, and effort. From quality perspective, proactiveness can also be implemented with classical data management techniques like data quality metrics, data architecture, etc. However, efforts are significant: Data requirements must be collected continuously and globally, translated into data quality metrics, and implemented in dashboards and workflows. Existing business partners change their addresses, run out of business, or tax numbers expire, and all these changes are to be identified, proactively.

With *Data Sharing*, you get the same results (maybe even more proactiveness), but at lower efforts: Continuous collection of data requirements, monitoring of business partner status, design of data quality metrics, and many more tasks are shared across the *Data Sharing* community and reduce the effort for each company.

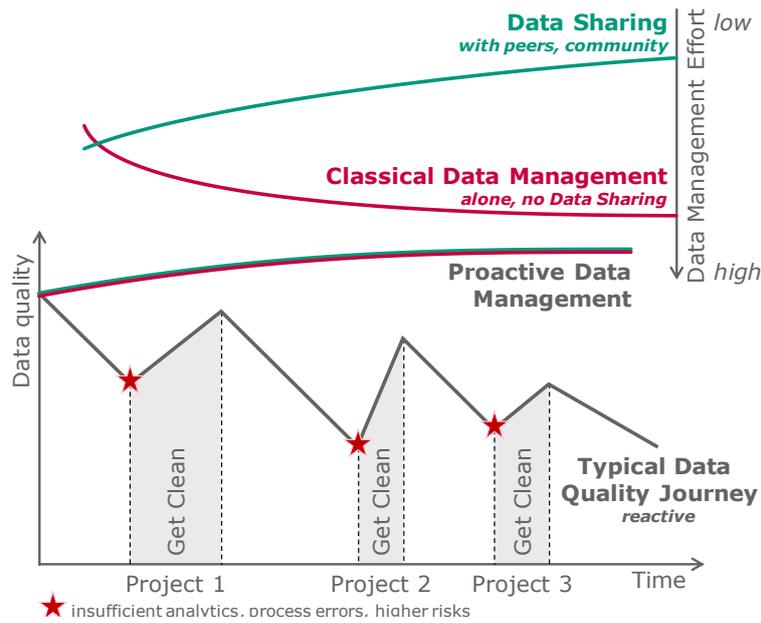


Figure 2: Less data management effort with Data Sharing

Data Sharing reduces data lifecycle costs directly

According to Six Sigma (rule of ten), indirect benefits by higher data quality exceed the direct benefits by 10 times. However, benefits of process failure reduction, good decisions, or more accurate business planning are hard to quantify and influenced by many factors.

To quantify (a part of) the value of *Data Sharing*, we focus on the direct benefits, basically cost reduction due to more efficient data maintenance.

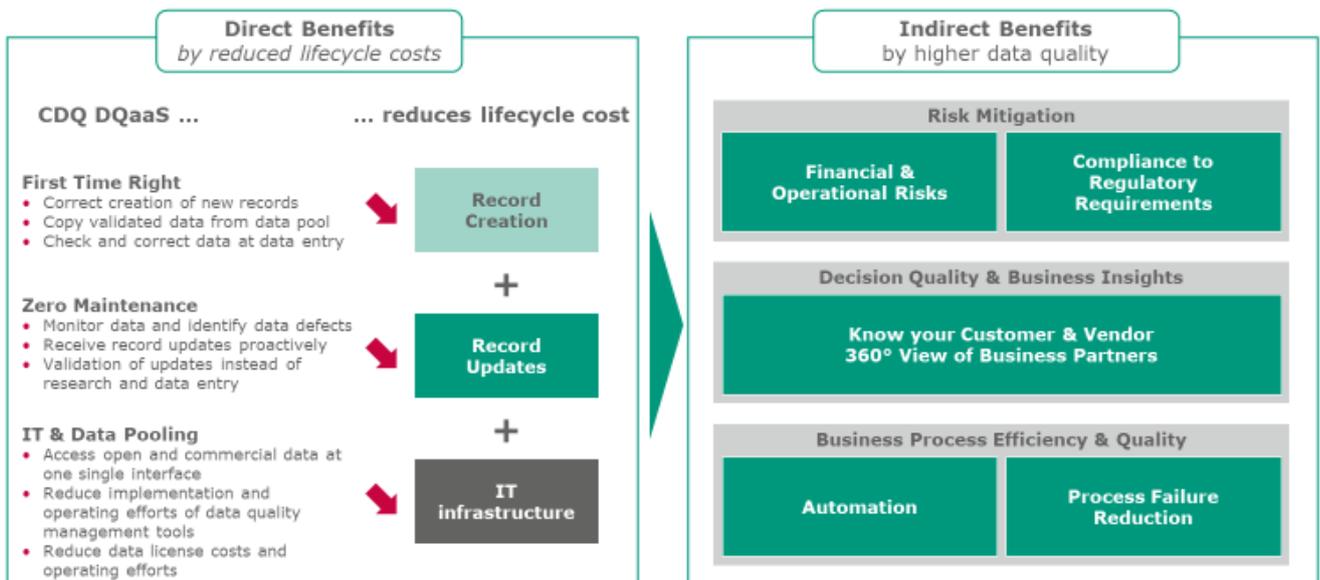


Figure 3: Data Sharing benefit framework

- **First Time Right:** Accurate and complete creation of new records, copy validated data from data pool, check and correct data at data entry.
- **Zero Maintenance:** Monitor all records and identify data defects proactively, receive record updates proactively, validation of updates instead of research and manual data entry.
- **IT & Data Pooling:** Access open data and commercial data at 1 interface, pool operation efforts of data quality management tools, reduce data and software license costs.

Data Sharing effects quantified

To quantify direct benefits of *Data Sharing*, parameters regarding costs, complexity, and effort of your data management process must be quantified.

- **Number of records:** Number of customer and vendor records in your systems. Do not care about legal entities, ship-to and bill-to addresses, or duplicates, just take the records which are effectively used in your processes and managed (i.e. not inactive) in your systems.
- **Data maintenance costs:** Personnel costs of your data workforce with hands on keyboard. This number may differ from region to region, just assume an average hourly rate – full costs, not only salary.
- **Data maintenance duration:** How long does it take to research correct addresses, legal names, or tax number, to validate the information by authority websites, and to enter the data manually? Our research shows 3 minutes on average per attribute.
- **Overlap with Data Sharing Pool:** If you find an up-to-date and validated customer record in the data pool, you do not have to research, validate, and enter data manually, you can just copy it. The Overlap is the coverage or match rate you can expect. The average for the Shared Data Pool is 43%.
- **Created/updated record ratio:** The expected number of created and updated records per year can roughly be derived from the overall number of records. Our statistics show 5% for customer data and 11% for vendor data.
- **Data purchase & IT costs:** Refers to costs of external reference data, data brokers, company profiles, etc., and to license, integration, and operation costs of tools and external services like tax number validation, duplicate matching, or address cleansing. Due to data validation by Data Sharing Community and use of a common platform, these costs can be reduced.

You can calculate your own business case on our website, www.cdq.ch/calculator.

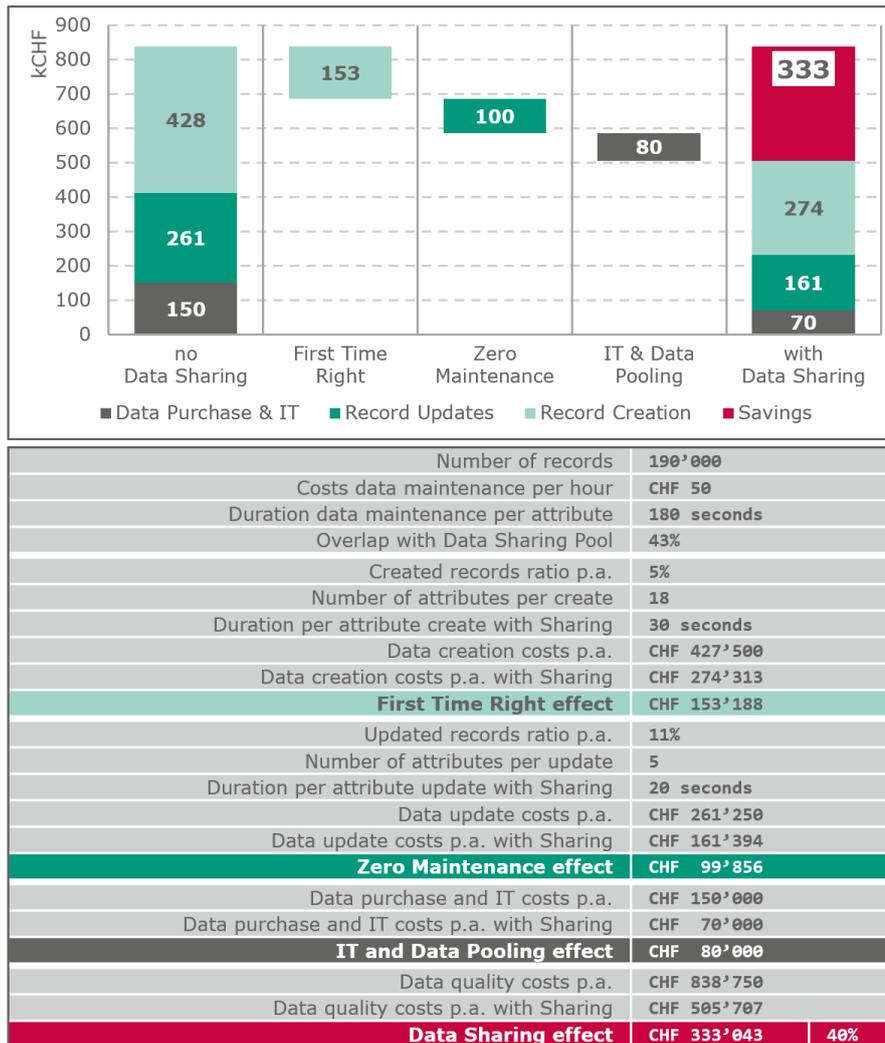


Figure 4: Business case for Data Sharing

The business case considers the three direct benefits of *Data Sharing*:

- **First Time Right effect:** Lead time reduction of data creation processes, calculated by the difference between data creation costs with and without Data Sharing. Creation costs without Data Sharing are just calculated by the number of created records (5% of the overall number of records, see above), the number of maintained attributes per new record (18), duration of data maintenance per attribute (180 seconds), and personnel costs (50 CHF per hour). With Data Sharing, data maintenance is faster (30 seconds) because data can be copied from the Data Sharing pool instead of manual data entry. However, this effect is only calculated for records which match with the data pool (43%).
- **Zero Maintenance effect:** Like the *First Time Right* effect, just for data update costs with updated records ratio (11%) and less attributes per updated record (5). The default assumption for maintenance duration per updated attribute (20 seconds) is lower than for record creation because pre-validated updates may come proactively from the Data Sharing community.

- **IT and Data Pooling effect:** Straight forward, just the cost savings due to replaced data management tools (e.g. address cleansing, duplicate matching) and lower costs of external reference data.

Again, the quantified *Data Sharing* effect is only due to reduction of data maintenance efforts. Indirect benefits due to higher data quality would exceed these numbers by 10 times (rule of ten, Six Sigma).

Data Sharing cases

Regarding the effect of *Data Sharing* on data maintenance costs, size matters. Another important factor is the maturity of data management processes, i.e. the lead time of data record maintenance due to automation and software integration. The following three cases illustrate how the number of records and the maintenance duration per record affect *First Time Right* and *Zero Maintenance*.

Case A: Low integration depth but benefits from day 1

The company from Automotive sector has 70k customer records and 50k vendor records of business partners in 100+ countries. Data is maintained centrally in a shared service center in Eastern Europe.

For record creation, process specialists can use an external website to lookup business partner data in the shared data pool. If they find the business partner in the pool (match rate: 52%), they can manually copy the data to their workflow tool and save significant time, i.e. 60 seconds per attribute compared to 180 seconds for manual address research, tax number validation, etc.

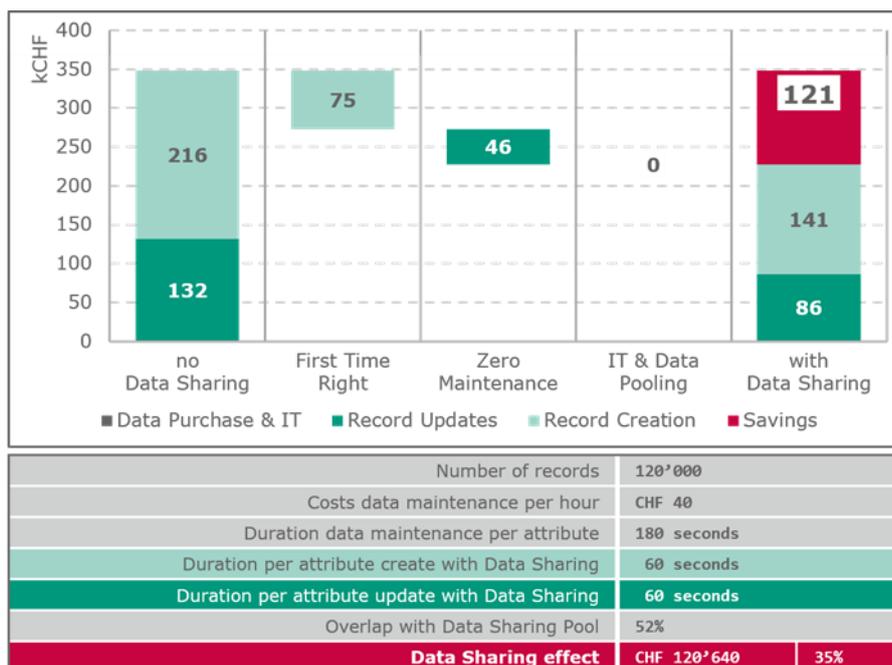


Figure 5: Case A, low integration depth but benefits from day 1

For record updates, the same website is used, with same benefits. However, due to lower update efforts in general (5 attributes versus 18 attributes, on average), *Zero Maintenance* effect is lower than *First Time Right* effect.

Case B: Deep integration and full-service use

The company from Pharma sector has 280k customer records and 210k vendor records of business partners in 150+ countries. Data is maintained in a global *follow-the-sun* Shared Service Center at three locations in Southeast Asia, Eastern Europe, and Middle America.

The company has deeply integrated Data Sharing services into their workflow tools: Process specialists can search the data pool directly from their workflow, manually entered data (if not found in the pool) is automatically validated and cleansed, and updates from the Data Sharing community pop up in the workflow, all in the same tool.

Beside *First-Time-Right* and *Zero-Maintenance*, the company could also phase-out an existing tool for address cleansing and duplicate matching, and they could reduce costs of external reference data (see *IT & Data Pooling* effect).

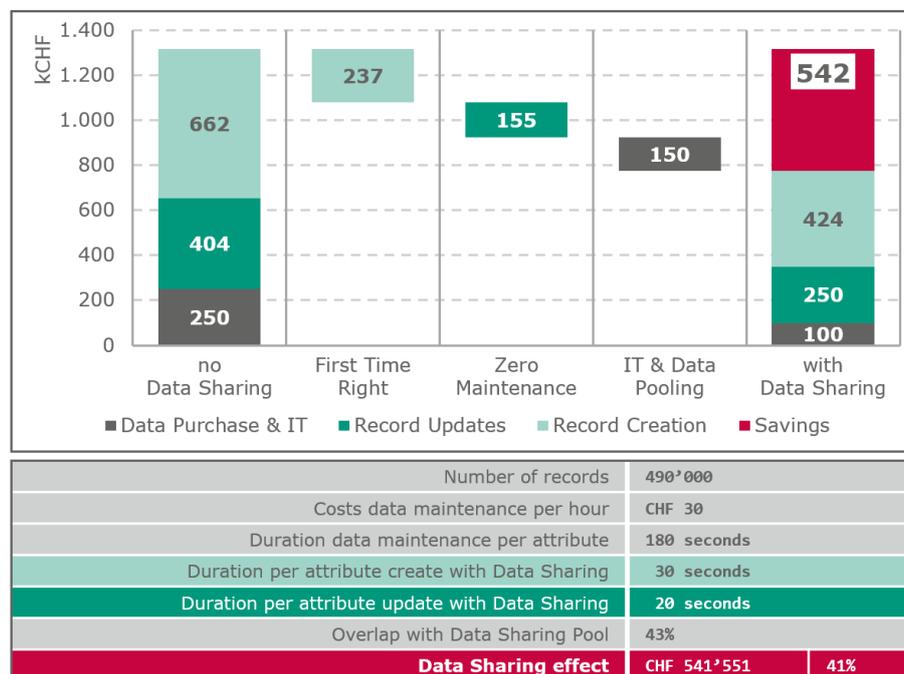


Figure 6: Case B, deep integration and full-service use

Case C: High overlap with data pool

The Machinery company has 60k customer records and 90k vendor records. 80% of their business partners are from European countries and the US with open data access to the national commercial registers. Hence, the overlap with the Data Sharing pool is quite high for this company, 72% overall. Data is managed by process experts within the business processes, supported by an end-to-end workflow, but without a central data management team (i.e. higher personnel costs, higher lead time per attribute).

For data maintenance, data pool lookup, address cleansing, tax number checks etc. are integrated in the workflow tool. However, updates for addresses and legal entity data are not yet fully integrated but downloaded from a website on a weekly basis.

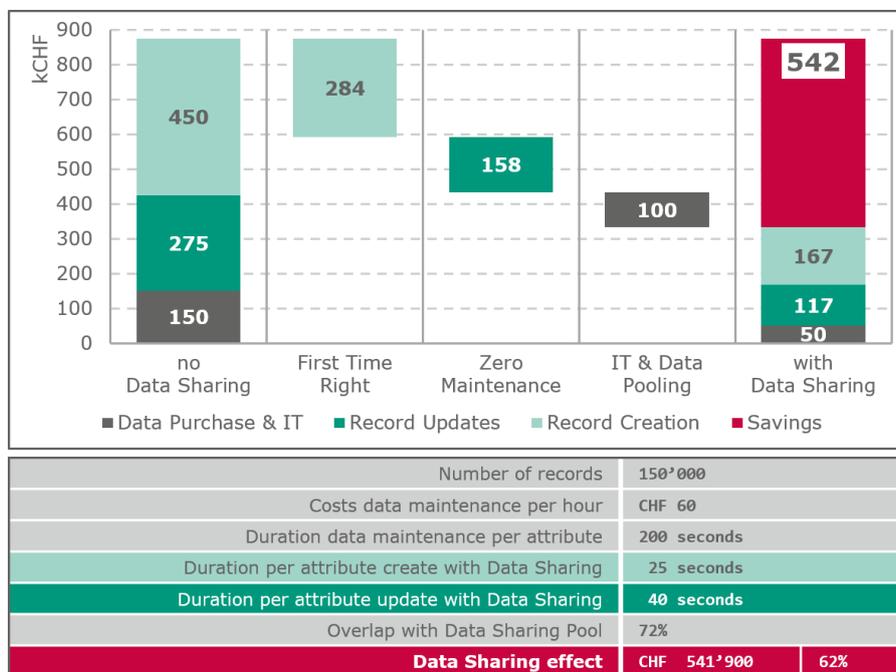


Figure 7: Case C, high overlap with data pool

Conclusion and further information

Data Sharing is the best way to better data. Quantifying the direct benefits of *Data Sharing* shows only a part, according to Six Sigma (rule of ten) only 10%, of the effective benefits. However, even this part should be enough to argue for the invest of joining a *Data Sharing* community. In the *Sharing Economy*, benefits for each community member grow by the network effects if the community grows.

Dare to share!

You can calculate your own business case on our website, www.cdq.ch/business-case.