

Auswahl von IT-Anbietern, -Tools und -Angeboten: Partner First, Tool Second

Beratungen im Analytics-Umfeld empfehlen üblicherweise, bei der Entwicklung einer Analytics-Lösung eine systematische, kriteriengeleitete IT--Tool-Auswahl vorzunehmen, die von einer breiten Longlist aus Produkten über eine Shortlist bis zu einem Proof-of-Concept führt. Das IT-Tool steht dabei am Anfang und im Mittelpunkt des Prozesses. Diese Empfehlungen richten sich allerdings an Unternehmen von zumindest mittlerer Größe, die eine eigene IT-Abteilung und idealerweise sogar ein Analytics-Team mitbringen. Unsere Erfahrungen im Projekt legen es nahe, dass diese Voraussetzung gerade in dem betrachteten Umfeld – kleinere Mittelständler mit Potentialen bei der Analyse digitaler Abbilder physischer Objekte – üblicherweise nicht gegeben ist. Trotz oftmals hohem High-Tech-Anteil der produzierten oder verarbeiteten Güter sind umfangreiche IT-Ressourcen und weitergehende IT-Kompetenzen eher die Ausnahme.

Für eine Datengenossenschaft rückt damit die Auswahl eines geeigneten *IT-Anbieters* in den Vordergrund. Dieser muss als kompetenter Partner die Anforderungen der Genossen aufgreifen können und mit geeigneten Tools umsetzen können – er übernimmt in der Genossenschaft letztlich die Rolle einer IT-Abteilung. Die Toolauswahl ist damit verzahnt, da der Anbieter schließlich eine anforderungskonforme, skalierbare und performante Lösung bereitstellen muss. Idealweise befindet der IT-Anbieter auf Augenhöhe, d.h. ist selbst mittelständisch und mit der Kultur und der Branche der Genossenschaft vertraut.

Zusätzlich zu Anbieter und Tool ist zu berücksichtigen, dass ein Anbieter alternative *Angebote* mit unterschiedlichem Leistungsumfang anbieten kann. In der Auswahl hierfür ist neben einem kurzfristigen auch ein langfristiger Zeithorizont zu berücksichtigen. Gerade bei Genossenschaften ohne Analytics-Vorerfahrung stehen zunächst der Aufbau einer belastbaren Datenbasis, deren Management sowie einfache, deskriptive Analysen mit Reports und Dashboards im Mittelpunkt. Mit wachsender Erfahrung und einem schrittweisen Ausbau der Datenbasis werden den Anwendern üblicherweise die Potentiale explorativer und im nächsten Schritt auch anspruchsvollerer KI/Machine Learning-Analysen stärker bewusst. Diese Perspektive sollte in die Anbieter- und Angebotsauswahl einfließen.

Die Rolle des IT-Partners

Unsere Studien legen nahe, den IT-Anbieter nicht selbst zum Genossen einer Datengenossenschaft zu machen. Gerade bei einem wenig IT-affinen Konsortium bündelt der IT-Anbieter Macht, was zu einer Unwucht in der Genossenschaft führen kann: Zum einen konzentriert der IT-Partner als zentraler Lösungsentwickler und -betreiber Kompetenzen, Daten und Wissen und kann so in eine dominante Position gelangen. Zum anderen ist seine Rolle grundlegend verschieden von der aller anderen Partner – und damit unterscheiden sich auch seine Ziele. Unter diesen Voraussetzungen entsteht ein hohes Risiko von Zielkonflikten (beispielsweise durch Nutzung des erworbenen Know Hows bei der Akquise und Betreuung konkurrierender Unternehmen). Mit einem von der Genossenschaft vertraglich beauftragtem IT-Dienstleister lässt sich diese Position einfacher beherrschen.

Typen von IT-Partnern und IT-Tools

Im Rahmen unserer Gespräche mit IT-Dienstleistern sowie bei der Suche nach konkreten Partnern für unsere Datengenossenschaften haben wir unterschiedliche Typen von IT-Partnern und IT-Tools identifiziert, die potenziell für eine Datengenossenschaft in Frage kommen. Der Unterschied besteht dabei primär in der *fachlichen Herkunft* und infolge der *inhaltlichen Schwerpunktsetzung*, der *Kompetenzen* und dem *Funktionsumfang der vom Partner mitgebrachten Tools*. Es ist zu beachten, dass jeder Fall eine individuelle Ausrichtung erfordert und mehrere Anbieter auch durchaus

gemeinsam agieren können. Des Weiteren sehen wir die großen IT-Cloudanbieter/Hyperscaler typischerweise eher als mittelbar über einen kleineren IT-Anbieter angebunden, d.h. der IT-Anbieter nutzt die Dienste des Cloudanbieters für den Aufbau der DG-spezifischen Umgebung.

1. IT-Anbieter mit Schwerpunkt Automatisierungstechnik, Produktion und Logistik

IT-Unternehmen mit einer Spezialisierung auf industrielle Anwendungen und Automatisierung bringen wichtige Kompetenzen bei der Anbindung von Maschinen, zu den relevanten Standards und Normen sowie oftmals umfangreiche Kenntnisse der technischen, organisatorischen, kulturellen und rechtlichen Rahmenbedingungen des industriellen Mittelstands mit. Die hierbei inhärente ingenieursorientierte Ausrichtung spiegelt sich oftmals auch im beruflichen Hintergrund der Mitarbeiter sowie den von ihnen umgesetzten Konzepten und Herangehensweisen wider – was ein Vorteil für die Zusammenarbeit ist, wenn die Genossen der DG einen ähnlichen Hintergrund haben. Teilweise verfügen derartige Unternehmen über individuelle, spezialisierte Toolsets für die Digitalisierung und die Prozessunterstützung mit eigenen Datenmodellen und Datenhaltungsansätzen. Diese unbestreitbaren Stärken können in einer analyseorientierten Datengenossenschaft schnell zu Herausforderungen werden, wenn die Architekturen und Modelle nicht mit aktuellen, insbes. cloudbasierten Datenhaltungs- und -analysearchitekturen kompatibel sind. V.a. die Überführung in typische analytische Datenmodelle ist dann schnell mit erheblichem Aufwand verbunden, da Themen wie Historisierung (inkl. Slowly Changing Dimensions), die Schaffung eines Single Point of Truth sowie eine Ausrichtung auf integrierte, zeit- und organisationsübergreifende Analysezusammenhänge nicht vorgesehen sind. Diese Art Anbieter bietet sich damit insofern dann an, wenn parallel zum Aufbau eines Datenraums auch eine Digitalisierungsinitiative mit tiefer fachlicher Einbindung in die Prozesse der Genossen erforderlich wird.

2. IT-Anbieter mit einem Schwerpunkt auf dem Internet der Dinge/dem Internet of Things (IoT)

Breiter und generischer ausgerichtet als die erste Gruppe sind IoT-orientierte Anbieter und deren Lösungen. Hier stehen die Anbindung typischer IoT-Datenquellen (Smarte Sensoren und Aktoren/Smarte Maschinen/RFID etc.) mit anspruchsvollen Anforderungen an die vorzuhaltenden Datenvolumina, die Geschwindigkeit des Datenanfalls sowie den Umgang mit komplex strukturierten Daten im Mittelpunkt. Gerade für cloudbasierte Dashboards mit Real Time-Integration komplexer, verteilter Sensorik sind derartige Anbieter prädestiniert, selbst wenn diese nicht immer den tiefen fachlichen Hintergrund der ersten Gruppe mitbringen. An Grenzen stoßen solche Anbieter unserer Erfahrung nach, sobald auch hier weitergehende Anforderungen an eine analyseorientierte Datenhaltung sowie tiefergehende Analysen gestellt werden.

3. IT-Anbieter mit Schwerpunkt Business Intelligence und Analytics (BIA)

„Klassische“ BIA-Anbieter haben ihre Wurzeln in einem unternehmensweiten Reporting sowie im Aufbau einer integrierten analyseorientierten Datenhaltung (Data Warehouse/Data Lake). Derartige Anbieter sind v.a. für den initialen Aufbau einer Datenplattform interessant. Andererseits sind bei den entsprechenden, hier favorisierten Lösungen anspruchsvollere AI- und Machine Learning-Lösungen oftmals eher ergänzt als integrativ vorgesehen. Schwerer wiegen eine möglicherweise fehlende Erfahrung mit der Anbindung technischer Systeme sowie einem mangelnden tieferen Verständnis für die Zielgruppe, speziell im Umfeld *kleinerer* Mittelständler.

4. IT-Anbieter mit Schwerpunkt Artificial Intelligence und Machine Learning

Der Vorteil spezialisierter AI/ML-Anbieter ist, dass diese Kompetenzen mitbringen, die insbes. für weitergehende Ausbaustufen einer Datengenossenschaft relevant sind – methodisch, mathematisch-statistisch, werkzeug-, entwicklungs- und betriebsbezogen. Oftmals sind diese Anbieter gleichzeitig mit den Grundlagen allgemeiner BIA-Lösungen vertraut. Selbst wenn dies nicht die Kernkompetenz ist, sollten entsprechende Anbieter bei einer perspektivisch stärkeren Ausrichtung auf KI/ML-Lösungen berücksichtigt werden.

5. Allgemeine IT-Anbieter mit Mittelstandsbezug ohne spezifische Schwerpunktsetzung

Auch nicht primär auf Automatisierungstechnik, IoT, BIA und/oder AI/ML spezialisierte Partner können in interessante IT-Partner werden, speziell wenn sie mit den Grundlagen cloud- und datenorientierter Lösungen vertraut sind, ein insgesamt breites und aktuelles IT-Know How vorweisen können und sich durch eine starke Mittelstandsorientierung auszeichnen. Hier können geringe Reibungsverluste bei der Kooperation mit der Genossenschaft etwaige Spezialisierungsdefizite potenziell ausgleichen. Dennoch sollte geprüft werden, inwiefern diese Anbieter der vorliegenden Aufgabe gewachsen sind. Insbes. sollte deshalb auf Erfahrungen beim Aufbau von Cloudlösungen mit starkem Datenanteil sowie Vertrautheit mit typischen Datenhaltungs- und -analysesystemen geachtet werden.

Kriterien für die integrierte Auswahl von IT-Partnern, -Tools und -Angeboten

Basierend auf diesen Vorüberlegungen wurde im Laufe des Projektes ein Katalog von Kriterien für die integrierte Auswahl von IT-Partnern, -Tools und Angeboten entwickelt. Dieser Katalog ist bereits sehr umfangreich, kann aber durchaus auch erweitert werden.

So wird bei den *toolbezogenen Kriterien* neben übergreifenden Kriterien das potenziell relevante Funktionsangebot sowohl auf der Datenhaltungs- und der Analyseseite berücksichtigt. Bei den *anbieterbezogenen Kriterien* stehen v.a. die Kompatibilität mit dem Konsortium sowie Angebote zu Schulung, Betreuung und Beratung im Mittelpunkt – Fragen, die v.a. bei Mittelstandskonsortien ohne eigene IT-Abteilung kritisch werden. Die *angebotsorientierten Kriterien* berücksichtigen neben allgemeinen Informationen zum Preis- und Servicemodell sowohl Fragen der Entwicklung und Bereitstellung der Lösung wie auch solche zu deren Betrieb. Hierbei wird jeweils differenziert nach Datenanbindung und -bereitstellung, Datenanalyse und Frontend/Dashboard.

Vor der eigentlichen Anbietersichtung sollte dieser umfangreiche Katalog auf die für die vorgesehenen Wertschöpfungsszenarien relevanten Kriterien reduziert werden. Hierbei kann ggf. auch nach einer kurzfristigen und einer langfristigen Perspektive differenziert werden, beispielsweise wenn weitergehende KI/ML-Funktionalität erst für eine spätere Ausbaustufe der Datengenossenschaft vorgesehen ist.

Die selektierten Kriterien können im nächsten Schritt zum Screening einer größeren Longlist an Anbietern genutzt werden, beispielsweise durch ein punktebasiertes erstes Scoring, wobei die Kriterien gewichtet werden sollten (etwaige KO-Kriterien sollten separat und vorweg geprüft werden). Die resultierenden Punktwerte erleichtern eine erste Vorselektion, die Ergebnisse sollten aber auch qualitativ plausibilisiert werden. Gerade in Grenzfällen empfiehlt sich eine kritische Analyse der Resultate.

Auf dieser Basis kann dann auch ein erstes Lastenheft vorbereitet werden, das im Rahmen von Vorgesprächen genutzt wird. Spätere Schritte würden gemeinsame Workshops, Demonstrationen

der Werkzeuge, Prototypen und Proof of Concepts vorsehen, um schrittweise einen oder mehrere geeignete Partner zu identifizieren.

Werkzeug

Für die entsprechenden Selektionsprozesse wurde im Projekt ein Spreadsheet-basiertes Werkzeug entwickelt, mit dem zum einen die Kriterienselektion vorgenommen werden kann und zum anderen das Scoring/Screening. Hierfür bietet das Werkzeug zwei Vorlagen, die jeweils individuell angepasst werden können. Das Werkzeug findet sich in GitHub unter der Adresse

https://github.com/Projekt-Datengenossenschaften/Werkzeugkasten/blob/main/05_AuswahlVonAnbieternToolsUndAngeboten_VorgehenUndTool/02_DG_Werkzeugkasten_TOOL_Angebotsauswahl_KriterienselektionUndAngebotsbewertung.xlsx

Datenmodelle für dynamische Datengenossenschaften:

Zum Umgang mit anspruchsvollen Anforderungen an Agilität und Vertraulichkeit

Eine Datengenossenschaft erfordert eine agile und skalierbare technische Basis für das von ihr unterstützte Daten-Ökosystem. Hieraus resultiert die Forderung, dass die zugrundeliegende Plattform und hier insbesondere die Datenhaltung dynamisch erweitert werden können, sei es zur Ergänzung neuer Partner, einer neuen Datenquelle (etwa eine neue Maschine), eines neuen Sensors oder Sensortyps, einer neuen Kennzahl oder auch zusätzlicher Kontextdaten (etwa Daten zur Prozesskonfiguration).

Abstraktionsschritte für eine problemlose Erweiterbarkeit

Eine Lösung, die im Projekt gewählt wurde, ist das Einziehen von Abstraktionsschritten in das Datenmodell: Statt mit konkreten Tabellen zu arbeiten (etwa „Luftdrucksensor bei Schneidemaschine x“), werden Tabellen für *generische Objekte* geschaffen (DT_Objekt), denen dann jeweils ein oder mehrere *Sensoren* (DT_Sensor) zugeordnet werden. Jeder Sensor spezifiziert wiederum Zeilen einer Tabelle „Messung“ (FT_Messung), wobei eine Messung neben dem Sensor noch eine Messzeit zugeordnet ist (erfasst in „DT_Uhrzeit“ und „DT_Datum“ erfasst). Diese Tabellen können bei Bedarf durch weitere, beschreibende Tabellen spezialisiert oder erweitert werden (vgl. Abbildung 1).

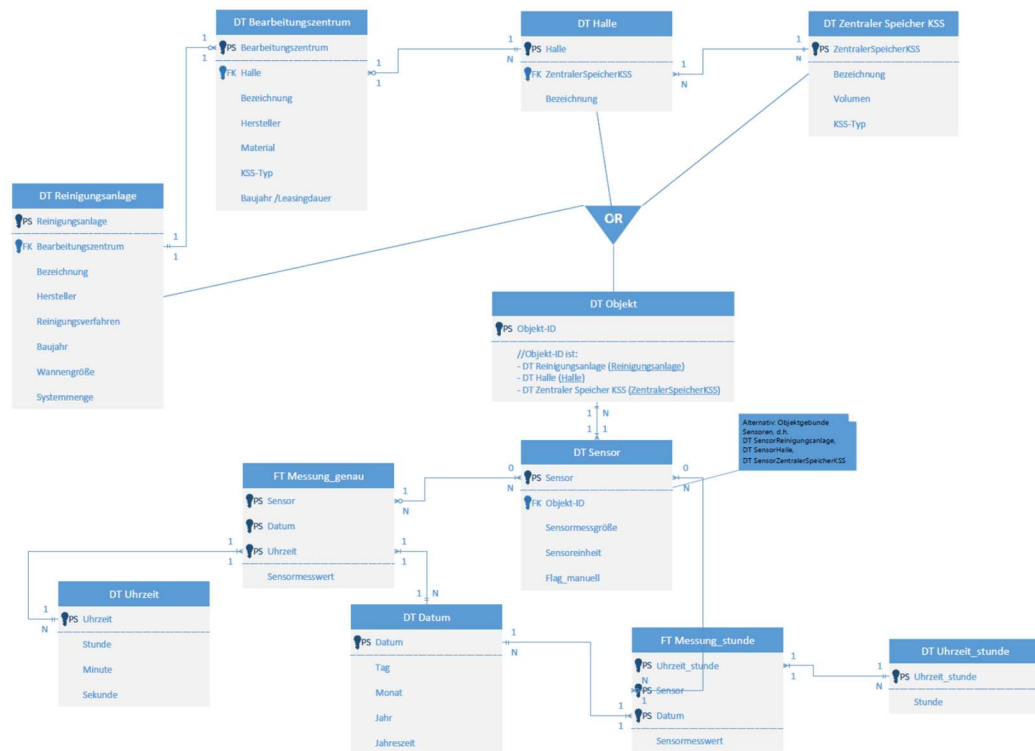


Abbildung 1: Datenmodell im Kühlschmierstoff-Fall

Die Abkürzung „DT“ steht für „Dimensionstabelle“, da die entsprechenden Einträge der Tabelle zur Selektion und Aggregation der Messwerte/Fakten der „Faktentabelle“ (FT) *Messung* genutzt werden können und damit zum Navigieren im Datenraum. Beispielsweise können über die Dimensionstabellen alle Messungen eines Sensors an einer bestimmten Maschine ausgewählt werden, die jeweils montags zwischen 13 und 14 Uhr erfasst wurden.

Die Einbindung eines neuen Sensors und ggf. eines neuen Sensortyps erfordert bei diesem Ansatz keine Modifikation an bestehenden Tabellen: Es wird lediglich die Sensortabelle um eine neue Zeile ergänzt, in der neben einer neuen Sensor-ID die Messgröße und die Maßeinheit des Sensors festgehalten werden. Dadurch ist dann auch spezifiziert, welche Semantik diejenigen Zeilen in der „Messung“-Tabelle haben, die diesem Sensor später zugeordnet sind.

Jeder Sensor ist des Weiteren einem Objekt zugeordnet, in dem er verbaut ist und dessen genaue Ausprägung variieren kann. Im Kühl-/Schmierstoffszenario wären das beispielsweise eine *Reinigungsanlage*, ein *Schmierstoffspeicher* oder eine *Halle*.

Alle Objekte eines Wertschöpfungsszenarios werden zunächst ohne weitere Beschreibung in der DT_Objekt-Tabelle mit reinen Identifikationsdaten („ObjektID“) erfasst. Diese wird jedoch bei Bedarf um weitere Tabellen ergänzt, in denen spezifische Angaben zu den konkreten Objekttypen festgehalten werden können (etwa in „DR_Reinigungsanlage“ Daten zum Baujahr und zum umgesetzten Reinigungsverfahren). Auf diese Weise werden einerseits alle Objekte einheitlich verwaltet, andererseits hat das Ergänzen einer Tabelle für einen neuen Objekttyp keinerlei Auswirkungen auf bestehende Tabellen. Es ist dabei z. B. ohne weiteres möglich, Daten zu einem Schmierstoff-Entsorger mit *Entsorgungsbehältern* zu ergänzen. Jeder Entsorgungsbehälter wäre wieder ein Objekt mit einer eigenen Objekt-ID, zu dem ggf. zusätzliche Informationen in einer neuen Tabelle „DT_Entsorgungsbehälter“ erfasst werden können. So kann das Wertschöpfungsszenario dynamisch erweitert werden.

Da der Sensor selbst ein Objekt ist, können einem Sensor zudem auch selbst Sensoren zugeordnet werden, etwa zur Überwachung des Sensors oder zur Abbildung eines sog. „virtuellen Sensors“, der aus den Messungen mehrerer physischer Sensoren einen neuen, kombinierten Messwert errechnet.

Möglichkeit zur Separierung von Datenbereichen

Die Trennung von unverfänglichen Objekt-Identifikationsdaten und spezifischen Kontextdaten in eigenen Tabellen erleichtert außerdem die Aufteilung des Datenraums in allgemein für die Genossen zugängliche Tabellen und Kontexttabellen, die jeweils spezifischen Data Ownern zugeordnet sein können. Hierbei zu beachten ist, dass einem Objekt ohne Weiteres auch mehrere Kontexttabellen zugeordnet werden können (etwa bei einem Hallen-Objekt „Halle“ neben einer Tabelle „DT_Halle“ auch eine zusätzliche zur „DT_Elektroinstallation_Halle“ oder eine zur „DT_Netzanbindung_Halle“).

Beachtet werden muss, dass auch die Zustandsdaten vertraulich sein können, da hieraus etwa Beschäftigungsintensitäten oder -zeiten abgeleitet werden können. Um dem zu begegnen, werden die Messungen in unterschiedlichen Tabellen für unterschiedliche Verdichtungsstufen festgehalten, für die jeweils wieder andere Zugangsregeln definiert werden können. Im Beispiel erfasst die Tabelle „FT_Messung_genau“ Messungen auf Sekundenebene, die Tabelle „DT_Messung_Stunde“ jedoch nur auf Stundenebene aggregierte Daten. Weitere Aggregationsstufen sind hier ebenfalls denkbar.

Weitere Besonderheiten: Messmodalitäten, Historisierung und weitere Abstraktion

In dem Datenmodell lassen sich des Weiteren besondere Erfassungsmodalitäten festhalten. So war es im Kühlschmierstoff-Fall erforderlich, zwischen automatisch und manuell erfassten Daten unterscheiden zu können, zum einen zur Einschätzung der Datenqualität, zum anderen zur Kalibrierung von Sensoren (mit manuellen Daten als Referenzdaten). Hierfür wurde eine entsprechende Angabe („Flag“) in die Sensortabelle eingezogen („Flag_manuell“), mit der sich manuelle Daten markieren oder auch in Gänze ein- oder ausblenden lassen. Dieser Ansatz ist erweiterbar etwa für die Kennzeichnung weiterer Eigenschaften der Messerte: etwa bei korrigierten Daten, rein synthetischen Daten, Daten, die in einer Simulation in einem Digitalen Zwilling erzeugt wurden, etc.

Nicht im in Abbildung 1 gezeigten Datenmodell abgebildet ist eine weitergehende Historisierung der Daten, um etwa Änderungen von Eigenschaften oder Zuordnungen über die Zeit verfolgen zu können, beispielsweise bei der Umpositionierung einer Maschine. Dies lässt sich jedoch problemlos durch die Ergänzung weiterer Attribute für Gültigkeiten („gültig_von“ und „gültig_bis“, „aktuell“) realisieren. Zu beachten dabei ist, dass die identifizierenden Schlüssel grundsätzlich unabhängig von den konkreten Identifikatoren der physischen Objekte gewählt werden sollten. Beispielsweise sollte eine ObjektID keine „reale“ Maschinenummer sein, da diese im Laufe der Zeit neu vergeben werden kann.

Ergänzung und Fazit

Wir haben bewusst davon abgesehen, das Modell noch weiter zu abstrahieren, wie es etwa die Datenmodellierungsansätze „DataVault“ (Lindstedt et al. 2009, S. 1 ff., S. 56 ff.; Hultgren 2012, S. 19 ff., Hahne 2014) oder das „Anchor-Modell“ (Rönnebeck et al., 2010) vorsehen, die unser Modell maßgeblich inspiriert haben. Der Hintergrund ist, dass ein Datenmodell für eine Datengenossenschaft trotz aller Flexibilität für alle Genossen verständlich bleiben muss – auch für solche ohne vertiefte Datenmodellierungskenntnisse. Wir haben uns deshalb hier für einen Kompromiss aus Flexibilität und Zugänglichkeit entschieden.

Abgesehen davon halten wir in einer Datengenossenschaft ein Datenmodell für erforderlich, das einerseits ohne Modifikations- und Neuentwicklungsaufwand erweitert werden kann, andererseits aber eine gezielte Aufteilung in verschiedene Datenbereiche für unterschiedliche Data Owner, Verwendungskontexte und Vertraulichkeitsstufen erlaubt. Ein entsprechend agil handhabbares Datenmodell ist ein weiterer Baustein für eine agile Datengenossenschaft auch für schnellwachsende und dynamische Datenökosysteme.

Quellen

- Hahne, M.: Logische Modellierung mehrdimensionaler Datenbank-systeme. Deutscher Universitätsverlag, Wiesbaden, 2002, S. 237 f.
- Hultgren, H.: Modeling the Agile Data Warehouse with Data Vault. New Hamilton, Denver, 2012.
- Lindstedt, D., Graziano, K., Hultgren, H.: The Business of Data Vault Modeling, 2. Aufl., lulu.com, Morrisville, 2009.
- Rönnbäck, L., Regardt, O., Bergholtz, M., Johannesson, P., Wohed, P.: Anchor modeling – agile information modeling in evolving data environments. Data Knowl. Eng. 69(12), 2010, S. 1229-1253

Glossar

Für ein einheitliches Verständnis der in den entwickelten Werkzeugen entwickelten Konzepte wurde im Projektkonsortium ein Glossar entwickelt und als Wiki veröffentlicht.

Die Definitionen sind bei gängigen Begriffen mit Literaturquellen unterlegt. Es wird jeweils gekennzeichnet, inwiefern es sich um einen allgemeinen Begriff (z. B. Data Governance) oder um ein spezifisch auf Datengenossenschaften ausgerichtete Konzept handelt (z. B. Digitales Abbild). Des Weiteren sind die Begriffe hierarchisch geordnet, über Synonymbeziehungen verknüpft und einer Priorität zugeordnet (1 – Kernbegriff, für das Verständnis des Konzepts zwingend, 2 – Begriff, der für die Zusammenarbeit im Konsortium erforderlich ist, 3 – primär für ein Teil-Team im Konsortium relevant und 4 – Hilfsbegriff).

Das Glossar ist über die folgende URL erreichbar:

<https://wiki.datengenossenschaft.com/>

Login: Gast

Passwort: VW49rLxapK

Darüber kann es als XML-Export aus dem Datengenossenschafts-Repository bezogen werden.

Anleitung zum Aufsetzen: