
Über dieses Buch

Dieses Buch richtet sich an alle, die mit Daten arbeiten – sie sammeln, integrieren und auswerten müssen oder sie analysieren möchten – und dazu die Hilfsmittel der Informationstechnologie einsetzen. Es ist von Insidern geschrieben, gewissermaßen von Datenexperten für Datenexperten. Denn unser Geschäft ist die industrielle Erhebung und Verarbeitung großer Datenmengen zur Informationsgewinnung und die Bereitstellung der Ergebnisse als Basis für wichtige Entscheidungen.

Den überwiegenden Teil unserer beruflichen Erfahrungen haben wir im Bereich der statistischen Informationssysteme der Deutschen Bundesbank gesammelt. Das in diesem Buch präsentierte Gedankengut bezieht sich jedoch auf die globale Rolle der Statistik beim Aufbau umfassender Datenwelten und bei der Bereitstellung von Information als öffentliches Gut, nicht auf die speziellen Merkmale der Notenbankstatistiken. Deshalb stammen die im Folgenden aufgeführten Beispiele mehrheitlich auch bewusst nicht aus der Welt der Finanzwirtschaft oder einer anderen wissenschaftlichen Spezialdisziplin. Wir wollen anhand von Beispielen mit Alltagsrelevanz zeigen, wie die Konzepte der Statistik die Basis für eine universelle und standardisierte Bereitstellung von beliebigen Informationen bieten und sozusagen einen *barcode of information* liefern können, wie er zum Beispiel aus der Warenwirtschaft bekannt ist. Wir möchten also zeigen, dass für beliebige Informations- und Datenpunkte, nehmen wir als Beispiel die „durchschnittliche Schneehöhe in Garmisch-Partenkirchen im Januar 2016“, genauso eine Identifikation gebildet werden kann wie für „einen Liter Fair-Trade-H-Milch, Fettanteil 1,8 %“.

Für alle Informationsdienstleister stellt sich seit Jahren immer wieder die Herausforderung, explosionsartig wachsende Datenwelten erforschbar und nutzbar zu machen. Dies ist aber nicht nur ein Problem großer Datenmengen und zahlreicher Quellen. Tatsächlich liegt in der korrekten Verknüpfung von Daten unterschiedlicher Quellen und Themengebiete sowohl die größte Herausforderung als auch das größte Potenzial zum Aufbau von Wissen. Dieses Vorgehen wurde schon früh in der Kriminalistik eingesetzt und mit dem negativ belegten Begriff der „Rasterfahndung“ assoziiert. Inzwischen wird in nahezu allen Unternehmen und Institutionen versucht, unter dem – inzwischen positiv belegten – Begriff der „Datenintegration“ den Wissensaufbau durch Zusammenführung von Informationen aus unterschiedlichen Geschäftsbereichen oder Wissenschaftsdisziplinen voranzutreiben.

Die Vision, die wir in diesem Buch beschreiben, besteht in einer standardisierten Datenwelt, vergleichbar mit einem Baukastensystem, also einem „Lego für Daten“. Darin liegen Daten zu unterschiedlichsten Themengebieten in einer leicht zugänglichen und zueinander passenden Form vor. Diese Datenwelt kann kontinuierlich durch Einfügung neuer Inhalte erweitert werden. Je nach Erfordernis kann sie zur Problemerkennung und -lösung sowie zum Wissensaufbau verwendet werden, indem die Inhalte (Bausteine) flexibel miteinander verknüpft und inhaltlich und technisch zu neuen „Informationsgebilden“ geformt werden.

Es geht also um eine geordnete Sammlung relevanten Wissens, ein Datenkompendium, das gewissermaßen als „öffentliches Gut“ genutzt werden kann. Dabei mag sich der Begriff „öffentlich“ auf ein spezifisches Unternehmen, eine Branche, eine Wissenschaftsdisziplin oder – noch globaler gedacht – auf eine buchstäblich öffentliche Kollektion aus geografischen, meteorologischen, medizinischen, finanziellen, verkehrstechnischen, landwirtschaftlichen, versorgungstechnischen, pädagogischen und psychologischen Datenbeständen beziehen, möglicherweise ergänzt um Registerdaten wie Kataster- oder Unternehmensverzeichnisse (sofern zugänglich).

Wir glauben, dass sich diese Vision nur mit der disziplinierten Einhaltung bewährter Prinzipien und unter Nutzung neuester Methoden und Technologien verwirklichen lässt. Die von uns präferierten Ansätze stellen wir in unserem Buch vor. Dabei gehen wir besonders auf die beiden für die Datenarbeit relevanten Erfolgsfaktoren ein: die Einführung eines universell nutzbaren Ordnungssystems verbunden mit dem gemeinsamen Willen zur Standardisierung. Wir stellen den in der internationalen öffentlichen Statistik genutzten Standard *SDMX* (*Statistical Data and Metadata Exchange*) vor und zeigen auf, welche tiefgreifenden Veränderungen durch die Einführung dieses Standards und des damit verbundenen Ordnungssystems für die Arbeit der internationalen Statistikcommunity möglich waren. Wir glauben, dass der Schritt zur Standardisierung der Durchbruch zur gewinnbringenden Nutzung großer Datenmengen für vielfältige Themen ist. Und dass daher die Nutzung eines weltweit etablierten ISO-Standards wie *SDMX* auch für andere Themenbereiche der entscheidende Erfolgsfaktor sein kann. Dabei sind stets die ersten Schritte eines Standards von der Idee einiger weniger bis zum marktbeherrschenden Selbstläufer die schwersten. Auch diese beschreiben wir in den folgenden Kapiteln.

Die Motivation zum Schreiben dieses Buches besteht in diesem Werben für Standardisierung, für ein universelles Ordnungssystem für Daten und für die Verwendung der Konzepte und Standards der Statistik zum Aufbau dieser Datenwelt. Wir möchten damit alle die erreichen, die zum Ziel der Information als öffentliches Gut beitragen können, von der Wissenschaft über die Softwareindustrie bis hin zu intensiv datennutzenden Unternehmen und Institutionen, von der Leitungsebene bis hin zur Arbeitsebene. Wir richten uns daher sowohl an den Professor als auch an den wissenschaftlichen Mitarbeiter, sowohl an den Softwarearchitekten als auch an den Programmierer, sowohl an den *CIO* (*Chief Information Officer*) als auch an den *Data Analyst*. Dies soll aber nicht bedeuten, dass wir unseren Leserkreis auf die genannten Rollen beschränkt sehen möchten. Vielmehr schließen wir alle diejenigen mit ein, die sich für unser Thema interessieren.

Dieses Buch ist bewusst kein Fachbuch, kein „SDMX für Einsteiger oder Experten“ geworden, es soll vielmehr eine Hinführung zum Aufbau eines Ordnungssystems für die Datenwelten leisten. Andererseits dürfte es ohne eine konkrete Vorstellung davon, wie SDMX aufgebaut ist, schwer zu vermitteln sein, warum ausgerechnet dieser Standard das Potenzial für die Umsetzung unserer oben genannten Vision hat. Deshalb wird in diesem Buch nach den grundsätzlichen Ausführungen von Teil 1 auch eine gut verständliche, aber dennoch mit der erforderlichen Detailtiefe ausgestattete Beschreibung des Standards SDMX in Teil 2 erfolgen.

Wir hoffen, dass unsere Gedanken inspirierend und hilfreich sind, nicht zuletzt, weil ein Leser sich darin vielleicht wiedererkennen kann. Und wir möchten mit unseren Überlegungen dazu einladen, in der aktuellen *Big-Data-Bewegung* eine andere Perspektive einzunehmen, statt immer weniger über Daten nachzudenken und immer mehr auf die Schlagkraft der IT-Systeme zu bauen. Bei allem technischen Fortschritt stehen nach unserer Ansicht immer noch die Intelligenz, der Ideenreichtum und die Erfahrung des Menschen, der die Technik nutzt, im Vordergrund. Denn auch in „*Think Big*“ steht immer noch das „*Think*“ an erster Stelle.

In diesem Buch vertreten wir unsere persönlichen, aus unserer langjährigen Praxis und aus dem intensiven Gedankenaustausch mit Kollegen von anderen Organisationen, Herstellern, Software- und Beratungshäusern gewonnenen Ansichten. Diese spiegeln selbstverständlich nicht zwangsläufig die Ansicht der Deutschen Bundesbank oder ihrer Mitarbeiterinnen und Mitarbeiter wider.

Frankfurt am Main, Februar 2017

Reinhold Stahl, Dr. Patricia Staab

Die Vermessung des Datenuniversums

Datenintegration mithilfe des Statistikstandards SDMX

Stahl, R.; Staab, P.

2017, XIII, 108 S. 39 Abb. in Farbe., Softcover

ISBN: 978-3-662-54737-3