

データ品質マネジメント

実践ガイド (ISO 8000 シリーズ)

Managing Data Quality A practical guide

Tim King, Julian Schwarzenbach 著

Andrew King, 苑田義明 訳



日本規格協会

データ品質マネジメント
実践ガイド (ISO 8000 シリーズ)

Managing Data Quality A practical guide

Tim King, Julian Schwarzenbach 著

Andrew King, 苑田義明 訳

著作権について

Tim King と Julian Schwarzenbach 著の『Managing Data Quality: A practical guide』は 2020 年に BCS, The Chartered Institute for IT により発刊されたものです。本書は、出版社である BCS, The Chartered Institute for IT および原作者の許諾を得て、日本規格協会が発行したものです。本書に収録した日本語訳は、著作権法により保護されています。本書の一部または全部について、当会及び BCS, The Chartered Institute for IT の許可なく引用・転載・複製等を行うことを固く禁じます。

“Managing Data Quality: A practical guide”, Tim King and Julian Schwarzenbach, 2020, published by BCS, The Chartered Institute for IT. This work has been published by the Japanese Standards Association (JSA) with licensing approval and rights granted by BCS, The Chartered Institute for IT, the publisher, and the copyright holder of the work in its original. No part of this publication may be reproduced, reprinted or utilized in any form without permission of JSA and the publisher.

目次

図版・表	VIII
著者	X
謝辞	XII
日本の皆様へ	XIII
略語	XIX
用語集	XX
序文	XXIV

第一部 企業データの課題	1
1章 データ資産	2
データとは何か?	2
データ品質とは?	14
データ品質マネジメントとは?	18
まとめ	19
2章 データの利活用とマネジメントにおける課題	20
混沌としたデータの広がり	20
複雑な意思決定	21
好循環か、悪循環か?	22
不明確なデータ所有権	23
バックアップとデータの品質	24

	ビジネスケースにおけるデータ品質と透明性の欠落	25
	データトライアングル	26
	原材料としてのデータ	27
	データマシーン—期待と現実	28
	読者のデータは読者を信頼しているだろうか？	30
	企業のデータ品質マネジメントにおける課題	33
	まとめ	34
3章	データ品質に人々が及ぼす影響	36
	データ品質と安全衛生の比較	37
	人々とデータ	40
	データ動物園	41
	データに対する行動様式がどう相互作用するか	59
	チームの一員としての個人	60
	組織内のチーム	61
	データ劣悪化要因	62
	まとめ	63
4章	ケーススタディと事例	65
	不適切なデータの影響の実例	65
	ケーススタディ—マーズ・クライメイト・オービター (火星探査機)	67
	ケーススタディ—データ品質を低下させるメンテナンス 生産性の目標	68
	ケーススタディ—RAILTRACK	70
	ケーススタディ—法定報告	71
	ケーススタディ—大きすぎた車両	72
	ケーススタディ—小売業の失敗	72
	ケーススタディ—不適切な管理と性急さによるデータ品質の低下	73
	まとめ	74

第二部	データ品質マネジメントの枠組み	75
5章	データ品質マネジメントの目的とスコープ	76
	データマネジメントとデータ品質マネジメントの違い	76
	データ品質マネジメントの重要な原理	79
	まとめ	80
6章	ISO 8000-61 のアプローチ	81
	ISO 8000-61 のスコープ	81
	ISO 8000-61 におけるプロセス群	82
	まとめ	85
7章	データ品質マネジメント能力レベル	86
	能力レベル 1	87
	能力レベル 2	89
	能力レベル 3	90
	能力レベル 4	93
	能力レベル 5	94
	能力モデル全体を通して	95
	まとめ	96
8章	ISO 8000-61 プロセス群	97
	データ処理	97
	データ仕様と作業指示プロビジョニング	99
	データ品質監視と制御	103
	データ品質計画	104
	データ関連支援	108
	資源プロビジョニング	116
	データ品質保証	120
	データ品質改善	125
	まとめ	132

9 章	成熟度向上への道すじ	133
	成熟度向上のための計画	133
	成熟度のアセスメント	134
	まとめ	136
第三部	データ品質マネジメントの実践	137
10 章	データ品質マネジメントのための組織の準備	138
	データ活用組織とはどのようなものか?	138
	一般的な組織における改善の視点	141
	データ品質マネジメントへの道のり	145
	変革のためのケース	147
	変革し続ける組織	151
	最高データ責任者 (Chief Data Officer) の役割	153
	組織の準備	154
	まとめ	156
11 章	データ品質マネジメントの実現	157
	データ品質マネジメント実現に向けた取組みの全体像	157
	経営層による支援	158
	文脈の理解	160
	相乗効果の確認	161
	実現にむけた取組みの選択	162
	「対象範囲」の合意	163
	変更管理	165
	データの倫理的活用	167
	課題と問題の処理	168
	既存プロジェクトのリスク排除	168
	予算とリソースの確保	170
	実装の開始	171

まとめ	173
12 章 人的要素—組織員によるデータ品質マネジメントの確保	174
人こそが解決策である	174
行動様式と文化	175
被雇用者のデータ契約	177
データに対する行動様式改善のための戦略	179
行動様式に対する組織的影響	182
まとめ	185
結 論	186
参考文献	188
索 引	191

図版・表

図 1.1	事業活動の構成要素	5
図 1.2	一般的なデータの典型的ライフサイクル	9
図 1.3	ドキュメントの典型的なライフサイクル	12
図 2.1	データ品質の好循環	23
図 2.2	データトライアングル	27
図 3.1	データ動物園の概要	42
図 6.1	ISO 8000-61 プロセスモデル	82
図 7.1	データ品質マネジメントの能力レベル 1	87
図 7.2	データ品質マネジメントの能力レベル 2	89
図 7.3	データ品質マネジメントの能力レベル 3	91
図 7.4	データ品質マネジメントの能力レベル 4	93
図 7.5	データ品質マネジメントの能力レベル 5	94
図 7.6	データ品質マネジメントのための能力モデル全体	96
図 8.1	能力レベルごとの ISO 8000-61 プロセス群	98
図 8.2	概念データモデルの例	110
図 8.3	論理データモデルの例	110
図 8.4	データ品質マネジメントの改善における計測基準の役割	123
図 8.5	石川ダイアグラムの例	128
表 1.1	データセットの例	17
表 3.1	現実世界と情報世界の行動の比較	39

表 5.1	DAMA-DMBOK（第二版）の知識領域	77
表 5.2	ISO 8000-61 が規定するデータ品質マネジメントプロセス群	78
表 9.1	組織のデータ品質マネジメントのための成熟度アセスメント尺度	135
表 10.1	人員に関する改善機会	143
表 10.2	技術的な改善機会	143
表 10.3	プロセスに関する改善機会	144
表 10.4	良いデータ・悪いデータの影響	149
表 11.1	データ品質マネジメントに関する考察	160

著 者

ティム・キング (TIM KING)

ティムは、ひょんなことからデータ品質の分野でリーダーとして活躍することとなった人物である。彼が、国際標準化機構 (ISO) によって新たに設立された作業グループ「産業用データ品質」(ISO/TC 184/SC 4/WG 13) の主査 (convenor) に任命されたのは、2006 年のまさにそのときその場所にいたからである。その後、150 人以上の国際的な専門家の協力のもと、データ品質に関する国際標準「ISO 8000」を開発してきた。

実際のところをいえば、ティムは、過去 15 年にわたる彼自身の取組みの中で、データ交換のための規格の開発と実装を通して、上記に関する自身の専門知識を築いてきた。彼はバブコック・インターナショナルに勤務しており、規格化の仕事と並行して、データの価値を高めるためのコンサルティングプロジェクトを数多く手がけてきた。これらのプロジェクトの多くは、高付加価値で複雑な資産の所有者や運営者を対象としている。それらの組織には、NATO, Shell, ロールス・ロイス, Network Rail, 英国国立原子力研究所, 英国国防省などが含まれる。

これらのコンサルティングプロジェクトをサポートするために、ティムはデータのマネジメントと活用における組織の成熟度をテストするための取組み方法を開発してきた。彼は英国コンピュータ協会 (BCS) および英国機械学会のフェローでもある。

仕事と家族の生活以外では、ティムはもっぱらクロケットというスポーツに

情熱を傾けており、国際レベルの試合で活躍中である。

ジュリアン・シュワルツェンバッハ (JULIAN SCHWARZENBACH)

ジュリアンは、データ管理者であり「データ・エバンジェリスト」であり、長年にわたり、様々な業界や組織において、データを活用して組織の成果を上げ、その経験を積み重ねてきた。

ジュリアンのキャリアは、エンジニアとしての業務を皮切りに、製鉄、重工業、自動車部品製造、採石、水関連の組織でのその役割を通じ、徐々にその焦点はデータへと変わってきた。コンサルタントとしての役割もまた、鉄道、水道、送電、社会住宅、石油化学、古代遺跡など多様な業界に広がってきた。ジュリアンは、そのデータ管理に関する関心の多くを、インフラストラクチャの効果的な資産管理や保守管理を手助けすることに費やしてきた。

加えて彼は、2010年以降BCSデータマネジメント専門家グループの議長を務め、英国規格協会(BSI)によって開発された、ビッグデータにヒントを得た一対の規格の開発におけるBCS代表をつとめてきた。ジュリアンは、アセットマネジメント協会(Institute of Asset Management)の資産情報ガイダンスと需要分析ガイダンスの開発プロジェクトをマネジメントしていた。彼が開発貢献した規格群には、ビルディングインフォメーションモデリング(BIM)のためのPAS 1192規格群と、それに続くISO 19650シリーズへの国際標準化作業がある。彼はまた、ISO 8000、BS 10102-1(ビッグデータ：データ駆動型組織のガイダンス)、BS 10102-2(ビッグデータ：データ集約型プロジェクトのガイダンス)にも貢献している。

ジュリアンは、データおよび資産に関するテーマで定期的に学会発表を行うとともに、データ関連会議の議長を務めている。

謝 辞

著者らは、本書の内容の多くにインスピレーションを与えてくれたすべての人々や組織に感謝したい。本書の多くは、これらの人々がデータにおける課題や取組みを通じて生み出してきた逸話や解決策が元となっている。また、ISO/TC 184/SC 4/WG 13（自動化システムおよび統合/産業データ/産業データ品質）に従事し、本書の中核をなす ISO 8000-61 を開発した専門家の皆様に感謝の意を表す。Data and Process Advantage Limited は、データ品質の行動的側面を説明するために、「Data Zoo（データ動物園）」のコンセプトを再利用することを許可してくれた。また「読者のデータは読者を信頼しているだろうか？」（※2章のコラム）の事例の背後にあるインスピレーションを与えてくれたイアン・ラッシュに感謝する。

日本の皆様へ

1994年10月1日の土曜日、初めて日本の地を踏んだそのときの記憶は、今でも鮮やかに覚えている。成田空港のターミナルビルを一步踏み出した瞬間の、日差しの強さと湿度の高さは、相対的に穏やかな英国の気候に比して、何よりも強烈な印象として刻み込まれた。私の人生におけるエキサイティングな時間が、まさに始まった瞬間だった。

私が非常に幸運だったのは、日本政府から2年にわたるポストドクター奨学生に選ばれたことだった。この奨学金のおかげで、私は国立金属材料技術研究所（現在は、東京から筑波市へ移転）で日本人の研究者らとともに研究に従事する機会を得た。その間、文化、社会、そして技術的精緻さと高みを目指す日本人の姿勢を学んだ。

最終的に、日本における私の滞在はトータルで4年に及んだ（日本製鉄の関連会社と高知工科大学での研究開発も含む）。その後、新たな挑戦の旅へ向けて日本から飛び立ったのである。ただ、心だけはこの「日出づる国」に残しつつ。

データについては、日本での研究に従事する前から関心を持っていた。そして日本に着いた後も、特に材料データベースを中心としたこの興味について思いを巡らせ続けていた。

この間の思索がたどり着いた先は、デジタルデータが普遍的言語であるという点であった。この言語はまた、人類の発展の新たな段階における核心である。

旧来のビジネスモデルは、デジタルトランスフォーメーションの力により変

革を迫られつつある。この力は、組織における意思決定の新たな能力を与えてくれるデジタルデータの役割に関するものだ。

技術革新のこれまでの流れの中で、日本は産業オートメーションを含む技術の現実社会における実装の高度化において重要な役割を果たしてきた。日本は再び、デジタルトランスフォーメーションにおいて同様な役割を果たす機会に直面している。その役割を担えるかどうかは、データの本質的性質と、その性質が製造製品のための品質マネジメントに比して、データ品質マネジメントにいかに関与する相違をもたらしているかを理解できるかどうかにかかっている。

日本の読者が本書へ関心を持ってくださったことに感謝し、データを資産化する実ビジネスにおける課題取り組みの中で、著者らのアイデアの有効性を理解してくださることを願ってやまない。このことは人類の進歩に対する読者一人ひとりの貢献を後押ししてくれるはずだ。

ティム・キング

I still have vivid memories of arriving in Japan for the very first time on Saturday 1 October 1994. As I stepped out of the terminal at Narita Airport, the strength of the sunshine and the degree of humidity seemed so much more intense than the more gentle climate of England. I had begun an exciting period of my life.

I was lucky enough to have been awarded a two-year post-doctoral fellowship from the Japanese government. This fellowship gave me the opportunity to work alongside Japanese colleagues at the National Research Institute for Metals (initially in Tokyo and then in Tsukuba City). I learnt about culture, society and Japanese commitment to technical precision and excellence.

Eventually, my stay in Japan extended to four years in total (with time working at a subsidiary of Nippon Steel and at Kochi University of Technology). I then departed for new adventures but definitely left a part of

my heart in the “Land of the Rising Sun”.

I was interested in data before working in Japan, where I continued to explore this interest, in particular focusing on materials databases. I have since gone on to work with experts from across the world in developing ISO 8000.

My experiences tell me that digital data are a universal language. And a language that is at the heart of a new phase of human development.

Traditional business models are coming under pressure from the power of digital transformation. This power is about the role of digital data in giving organisations new capabilities for decision making.

During previous waves of technological revolution, Japan has played a key role in refining practical implementation of technology, including industrial automation. Japan now has the opportunity to play a similar role for digital transformation. This role depends on understanding the fundamental nature of data and how this nature makes data quality management have important differences from quality management for manufactured products.

I thank you for your interest in this book and I hope you find our ideas useful as you grapple with your practical challenges to make data an asset that helps you in delivering your own contribution to human progress.

Tim King

長年にわたって、私は様々な分野においてデータに関わる仕事に従事してきた。その多くは、エンジニアリングの視点、つまり「データは目的達成のための手段であって、目的そのものではない」という視点である。そのため、利用されているソフトウェアや技術についての関心をほとんど持たず、データが支援する組織の成果物に注力してきた。

1999年の後半、私は膨大な資産登録の機能を含む、巨大なエンタープライズシステム全般にわたる責任を担うことになった。このとき初めて、データ品質に関する関心が芽生えたのである。それは例えば、「データの品質とはなんだろうか?」「どうして巨大なデータのクレンジングプロセスの成果物に価値があるのだろうか?」「どうしてユーザは、悪いと知っているデータを、最初に作る時に変えないのか?」と言ったことである。この瞬間こそ、データに対する振る舞いとしてのデータ・無政府主義種やデータ・リス種などを「Data Zoo (データ動物園)」(※3章参照)として最初に整理し始めたときでもあった。

それから数年が経ち、優れたデータと効率的なデータ利活用に関する私の関心は飛躍的に高まることとなった。そして規格開発における活動と並行してコンサルタント業務を進める傍ら、BCS データマネジメント専門家グループ (BCS Data Management Specialist Group) と、特に資産管理運用が重要な組織などの多くの産業分野横断活動の議長を務める中で、いつしか「優れたデータ」に関するエバンジェリストになっていた。

データは普遍的なものである。いかなる企業も優れたデータなしに競争優位性を獲得できない。データには分野特有の考慮すべきいくつかの要素がある一方で、データ品質はそのほとんどの局面に関係する。加えて、すべての従業員は、データ品質の維持と向上という、優れたデータの生成に関するなにがしかの役割を演じている。そのデータは、生産環境における品質マネジメントと表裏一体なのである。最後に申し上げたいのは、技術こそが答えではない、という点である。なぜなら、あらゆるソフトウェアソリューションの寿命に比して、データは圧倒的に長期にわたって利用可能である必要があり、それも優れた品質である必要があるからである。

本書は、読者らのデータ品質マネジメントに役立つ幅広い現実的アプローチを提供することで、様々な立ち位置にいる読者とそのコンテキストに参考となるよう著した。本書の日本語訳によって、我々の考え方がより多くの人々に届けられることとなった。読者の皆様が本書を楽しみ、また役に立つと感じてくれることを願っている。

For many years I have worked with data in a variety of contexts, usually with an engineering mindset – that the data is a means to an end, not the end itself. I tend to be less concerned about the software and technologies used, than the organisational outcomes that the data supports.

In late 1999 I took over responsibility for a range of large enterprise systems, including a large asset register. This was the point where my interest in data quality arose – What was the quality of the data? Why were the outputs of a large data cleansing process variable? Why were users not initiating changes to data that they knew was wrong? This was also the time when the Data Zoo concept first started to develop with the Data Anarchist and Data Squirrel being some of the first behaviours identified.

Fast forward many years and my interest in good data and effective data exploitation has grown significantly. Through my consultancy work combined with involvement in standards development work, chairing the BCS Data Management Specialist Group and a variety of cross-industry initiatives, particularly relating to asset intensive organisations, I have become an evangelist for good data.

Data is universal – few organisations can excel without good data. Good data principles are also universal – whilst there are some sector specific factors to consider, many aspects of data quality apply in most contexts. Additionally, all employees have a role to play in creating good data, sustaining and improving data quality – a message that strongly parallels quality management in production environments. Finally, technology is not the answer, since in many contexts data will need to be available and of good quality for far longer than any software solution exists.

This book was written to be accessible to a range of readers and contexts

by providing a range of pragmatic approaches to help readers manage their data quality. The Japanese translation of this book allows us to share this message with a wider audience. We hope you enjoy the book and find it useful.

Julian Schwarzenbach

略 語

BIM	ビルディング インフォメーション モデリング
BSI	英国規格協会
CDO	最高データ責任者
CIO	最高情報責任者
CTO	最高技術責任者
DMBOK	データマネジメント知識体系
EDMS	電子文書管理システム
GDPR	EU 一般データ保護規則
HUMS	健全性 / 利用監視システム
IoT	Internet of Things
ISO	国際標準化機構
IT	情報技術
JPEG	Joint Photographic Experts Group
MDM	マスターデータマネジメント
NASA	アメリカ航空宇宙局
NHS	英国国民保健サービス
PAF	小口利用者郵便番号住所ファイル
PoD	運命の預言者
SEP	他人事, 自分の知ったことではない (Somebody Else's Problem)
USPS	アメリカ合衆国郵便公社 (United States Postal Service)

用語集

正確性 (Accuracy)：データ項目とそれが表すエンティティとの間の一致。参考までに、正確性は次のことを確認する必要がある。各データ項目が特定のエンティティにリンクしていること、各エンティティはそれに関連するデータ項目を持っていること。

属性 (Attribute)：実体の特性を記録するために使用されるデータフィールド。ある文脈において不可分とされるデータの単一単位 (ISO/TS 21089:2018)。

最高データ責任者 (Chief Data Officer: CDO)：データの効果的な仕様、取得、活用、およびガバナンスを促進するために、組織の上級レベルで任命された個人。CDO はチーフ・デジタル・オフィサーを指すこともあるが、この役割は通常、デジタル技術を通じたデータの活用に重点を置いている。

完全性 (Completeness)：すべてのエンティティに対してデータレコードが格納されていること、およびエンティティのすべての属性が入力されているという品質。

一貫性 (Consistency)：異なるデータセット間で同じエンティティに関連するデータを正しくリンクする能力。

データ (Data)：通信、解釈、あるいは処理に適した、形式化された形態をとる情報の再解釈可能な表現 (ISO 8000-2:2020)。

データ管理者 (Data custodian)：データスチュワードを参照。

データガバナンス (Data governance)：データの管理に関連する方針の策定と実施 (ISO 8000-2:2020)。

データマネジメント (Data management)：一つ以上の情報システムにおけるデータと関連するプロセスの定義，作成，保存，維持，アクセスの提供の活動 (ISO/IEC TR 10032:2003)。

データオーナー (Data owner)：データ資産の説明責任を負う個人。

データ品質 (Data quality)：データの固有の特性のセットが要求事項を満たす度合い (ISO 8000-2:2020)。

データ品質基準 (Data quality criteria)：データの品質の性質を理解するためにデータに適用できる具体的なテスト。これには，品質の評価に使用する方法も含まれる。

データ品質マネジメント (Data quality management)：データ品質に関して組織を指揮し，管理するための調整された活動 (ISO 8000-2:2020)。

データセット (Data set)：データの論理的に意味のあるグループ化 (ISO 8000-2:2020)。

データスチュワード (Data steward)：特定のデータリソースのセットを管理する責任を委任された個人または組織 (ISO 8000-2:2020)。

ISO 8000：データ品質のためのマルチパート構成の ISO 規格。

ISO 9000：品質マネジメントの様々な側面に対応した規格群で，製品やサービスが顧客の要求事項を一貫して満たすことを保証し，品質を一貫して向上させることを望む企業や組織に指針とツールを提供している。

メタデータ (Metadata)：他のデータを定義し，記述するデータ (ISO 8000-2:2020)。

精度 (Precision)：データエントリに対する特異性の程度 (ISO/IEC 11179-3:2013 の定義を変更)

構造化データ (Structured data)：データセットにおいて，データの明示的な要素 (リレーショナルデータベース内のテーブル，カラム，キー，XML ファイル内のタグなど) によってカバーされる意味。

適時性 (Timeliness)：データ項目がどれだけ最新のものであるかを示す指標。

一意性 (Uniqueness)：データセットの中で，エンティティがそれに関連す

る単一のデータエントリを持っているかどうかの指標.

非構造化データ (Unstructured data): データセットにおいて, データの構造的な要素でカバーされていない意味のレベル (例えば, 診断的な医療スキャンの画像における脳の特徴).

妥当性 (Validity): データの構文と構造を定義する規則へのデータの適合性.

バリュー (Value): エンティティまたはアクティビティを表すためにデータフィールドに格納されている数字または文字.

翻訳にあたっての諸事項

- ・本文中、利用されている用語が特に ISO 9000 シリーズの文脈におけるものであることが明らかな用語については、JIS Q 9000:2015 (品質マネジメントシステム－基本及び用語) を参考に行っている。
- ・ISO 8000 シリーズについては JIS 版がないため、ISO/TC 184/SC 4/WG 13 および WG 23 での議論をもとに、翻訳者らの判断で訳をあてている。
- ・ISO/IEC 33020 に関する訳は、JIS X 33020:2019 (情報技術－プロセスアセスメント－プロセス能力のアセスメントのためのプロセス測定フレームワーク) を参考に行っている。
- ・5 章の DAMA-DMBOK に関する用語の日本語訳は、DAMA 日本支部監訳『データマネジメント知識体系ガイド 第二版』による。
- ・一部、英国の社会・歴史的背景に基づく用語や説明に関して日本の読者に補足が必要と思われたものについては、翻訳者らの判断で最小限の補足を行った (〈※〉において付記している)。
- ・用語の用途が明確でないものについては、著者らと協議し、翻訳において一部修正したものがあつる。

序 文

データ¹は、我々の周りにあまねく存在する。データの量は指数関数的な勢いで増加しており、我々の生活はデータの活用によって支えられ、成立しているのである。その一方で、多くの組織では、データの効果的管理とデータの品質の向上に苦慮している。

組織（および社会）のデータへの依存は比較的新しい現象であり、データを効果的に管理するための技術はまだ発展途上にあって、これらのアプローチに対する認知度は一般的に低い。

本書は『データ品質マネジメント—実践ガイド (Managing Data Quality: A Practical Guide)』と銘打って、努めてわかりやすく著した。本書の焦点は、データ品質を管理する方法を理解し、それを達成するための実践的なガイダンスを提供することである。

企業データの品質

本書は、単一のデータベースやデータストアの品質問題だけを検討するものではない。それだけではなく、ほとんどの組織で発生する、より広範な問題群にも注目している。例えば、それは複数のデータストアが存在する問題である。それらデータストアは必ずしも正式に管理されておらず、異なる時期に開発さ

¹ 読者の中には、一般的に「data」という単語を単数形で使う方もあるだろうが、BCSの用例では複数形として扱う点に注意していただきたい。