# データ活用型
# プロジェクトのマネジメント

日下部 貴彦【著】

コロナ社

# ま え が き

本書は，データサイエンティストや研究者等が実施するデータ分析の領域だけでなく，これを包含するプロジェクト全体に焦点を当て，データ分析・活用を行う際に必要な基礎知識を整理したものである。昨今，産官学のさまざまな組織がさまざまなプロジェクトでデータを収集し，蓄積することが活発に行われており，大学での学術研究だけでなく企業等でのデータ活用の機会も一般化しつつある。一方で，データ活用の際に留意すべき個人情報保護法をはじめデータ利用に関わる制度も整備されてきており，データそのものに対する技術的視点以外にもデータ活用に必要な知識の領域が広がってきている。このような知識が不十分である場合には，コンプライアンス違反や手戻り等のプロジェクト遂行上のリスクに直結し，プロジェクトの成功の妨げとなることが想定される。このような背景から本書では，データ分析や活用を行うプロジェクトを成功に導くためにメンバーが共通的に持っておくべき基礎知識を網羅的に取り扱うことに主眼をおいた。

これまで筆者は，東京大学空間情報科学研究センターでの専任講師・准教授の在任中に，人の流れをはじめとした都市・交通に関わるデータの収集やこれらのデータを利用した研究を実施してきた。この際には，都市・交通分野のデータがさまざまな企業や行政組織によって分散的に収集されていることに起因したデータ入手手続きの煩雑さやデータクレンジング等の技術的課題等，多くの課題に直面してきた。その後，2021 年 4 月より，「阪急阪神ホールディングス（株）の本社に『データ分析ラボ』を創設し，DX プロジェクトの推進を支援する。」という企業内のプロジェクトのもと，ラボ組織を立ち上げ，チームをリードし，データ分析を業務として確立することをミッションとしたディレクターとして活動している。このプロジェクトの当初は，筆者の大学でのデータ分析の経験を活かしてデータサイエンティストを育成することが，ディレクターの業務として期待されていた。しかし，このための環境を整えるため

には，データの収集や管理の手順書や業務フロー等を整備し，データ基盤の企画・運営，データ分析を用いた事業部門へのコンサルティング等を実施する必要があった。このように，企業内でデータサイエンスの知見を活かす環境を構成するための一連の業務を担ったが，このために必要な知識は既往の資料等では整理されたものが少なく，知識の習得やステークホルダーへの展開のために，自ら体系的な整理を行うことが必要であった。また，実際のプロジェクトを推進する過程では，ラボ内でのデータ分析に費やす時間と比べて，データの契約や実装，ステークホルダーへの説明に必要な時間が多く，プロジェクトの構成上は，データサイエンティストが分析を行う工程と同等にこれらの周辺にあるプロセスが重要なことを改めて実感した。

そこで本書では，筆者の上記のような経験で得た知見を中心に，データ活用を行うために必要な一連のプロセスに着目し，実践に必要となる知識を整理した。特に，データ分析を担当するデータサイエンティストだけでなく，プロジェクトオーナー，プロジェクトマネージャー，チームリーダー等のプロジェクトを統括する方にも読んでいただくことで，データ活用・分析を伴うプロジェクトの特性や進行方法等についての共通的な知識を形成し，プロジェクトのより効率的な遂行やより確実な成功に寄与したいと考えて執筆を行っている。なお，本書で示しているプロジェクト構成等の例は，企業でのプロジェクトを中心としているが，プロジェクトの進行については筆者が経験した大学での研究室運営からもヒントを得ており，研究室の主宰者の方などにも是非読んでいただきたい。

本書の執筆にあたっては，山本隆弘氏をはじめ阪急阪神ホールディングス（株）グループ開発室 DX プロジェクト推進部およびデータ分析ラボのメンバー，関連する事業会社との取り組みを通じて得た知識・経験が大きな糧となっている。また，（株）ソーシャル・デザイナーズ・ベースの山田菊子氏に，執筆にあたってのきめ細やかなアドバイスをいただいた。親愛なる家族，活動を共にした皆様，コロナ社の皆様をはじめ執筆を支えてくださった多くの皆様に謝意を表します。

2023 年 11 月

日下部　貴彦

# 目　　次

## 1章　序



## 2章　データ利用の類型と必要なスキルセット



## 3章　データ活用に関連する法令

## 4章 データ活用型プロジェクト



## 5章 データ活用型プロジェクトのマネジメント

## 6章　データ分析の実践

## 7章 集計・可視化の実践例



## 付録：データ活用に関連した法令に関する条文

# 1章

# 序

データサイエンス[1)†]のコミュニティ（例えばGoogleが運営しているKaggle[2)]）やさまざまなオープンデータ[3)]の取り組みの広がりにより，多様な分野で収集される実際のデータを用いて，データサイエンスの経験を積み，技術力を研鑽する環境が整ってきている。このような環境により，データサイエンティストがさまざまな手法の試行錯誤を繰り返すことが可能になり，手法の高度化や精度向上などを行うことで，データ活用の領域が加速度的に広がってきている。

データ活用が行われる現場は，研究者の自由な発想に基づいて行われる学術研究での高度な手法の開発や実装を目指したものだけでなく，企業でのマーケティング，製品開発，経営企画等や，行政組織での施策設計，EBPM（evidence based policy making）[4)]等をはじめとして多方面に展開されることが期待されている。このようなデータ活用を前提としたプロジェクトの展開では，データ分析等を実施するデータサイエンティストだけでなく，プロジェクトの企画者や統括者，データの提供者等を含むプロジェクトメンバーが，データ活用を行うプロジェクトの特性を理解し，実行することがプロジェクトを成功に導く鍵となる。このような背景から，本書は，データ活用・分析の技術的な内容だけでなく，法令や契約，プロジェクト管理，情報技術等を含む全体像を俯瞰し，データ活用を行うプロジェクトメンバーが円滑にプロジェクトを推進するうえで必要となる知識をまとめたものである。

---

† 肩付き数字は，巻末の引用・参考文献番号を表す。

## 1.1　データサイエンティストを取り巻く状況とプロジェクト

初めに述べたように，さまざまな分野でデータ活用を加速させるうえでのデータサイエンティストの役割は，特定の AI（artificial intelligence）や機械学習等の手法の高度化や精度向上を行ったり，提供されたデータに対して数理的な手法を適用したうえで解釈を得ることだけでは，満たされなくなってきている状況がある。つまり，データ活用の目的を達成するために的確なデータを収集したうえで，収集データをはじめとしたさまざまな情報を用いて課題設定を行い，その課題に対して適切な手法を選択し，データ解析とその活用を通じて目的を達成する一連のプロセスのなかでのさまざまな役割が重要となっている。

一連のデータ活用のプロセスを実行するためには，これまでデータサイエンスのコミュニティで利用できたオープンデータだけでなく，各専門分野の研究や企業での研究開発に必要なクローズドなデータセットを対象とすることも必要となる。このためには，分析手法の側面だけでなくデータの契約やプロジェクト管理，要件定義等のプロセスの側面からもデータを適切に取り扱うための取り組みが必要となる。特に都市や交通，マーケティング等の活動を対象として，民間企業をはじめとした多くの事業者等が収集するビジネスデータを取り扱うためには，個人情報の取り扱いをはじめとしたデータに対する制度や考え方を，データサイエンティストだけでなく，データの収集や契約，活用に携わるプロジェクトメンバーが共通に理解したうえでプロジェクトを推進する必要がある。さらに，プロジェクトのアウトプットを用いて意図したデータ活用が実施できるよう，適切なスコープの設定やプロジェクトの管理も重要な要素となる。（**図 1.1**）。

一般に，データを活用するプロジェクトでは

- データ活用を含むプロジェクトスコープの明確化
- 新たなデータの収集や第三者が収集したデータの受領を含むデータ取得

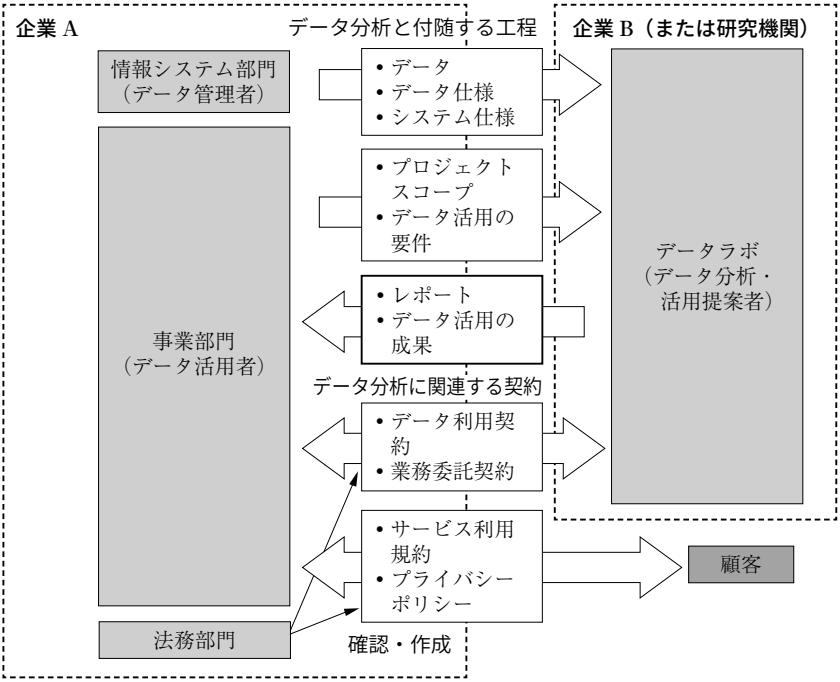


**図 1.1**　データ活用のためのプロジェクト概要

- 個人情報や機密情報へのアクセス制限等適切なデータの取り扱い環境の整備
- 分析可能なデータへの変換・加工

等の上流側のプロセスを経て初めて，データサイエンティストが分析手法の選択や構築，分析や検証，レポーティング等が実施できる。さらに，分析結果の活用や社会実装のためには，情報システムへの実装や関係するステークホルダーへの説明，費用や収益性等のビジネス的な観点での判断も必要となる。つまり，データ活用を行うのはデータサイエンティストではなく，その上流プロセスを行うステークホルダーであり，プロジェクトの中でデータサイエンティストが担う役割はデータ分析やモデル構築を実施する等プロジェクトの中では限られた部分である。したがって，それ以外の業務を担うシステムエンジニアや法務担当者，ビジネスや研究の企画担当者等の各プロセスのメンバーも

データを活用したプロジェクトが具備するべき条件を体系的に理解する必要がある。**図 1.2** および**図 1.3** では，上記に示したデータ活用の実装までの流れの一例と，このようなプロジェクトに関わるステークホルダーの関係を示している。これらの図に示されるように，データ活用に異なる役割を持った多数のプロジェクトメンバーが関わる。このことから，各プロセスに関わるメンバーの

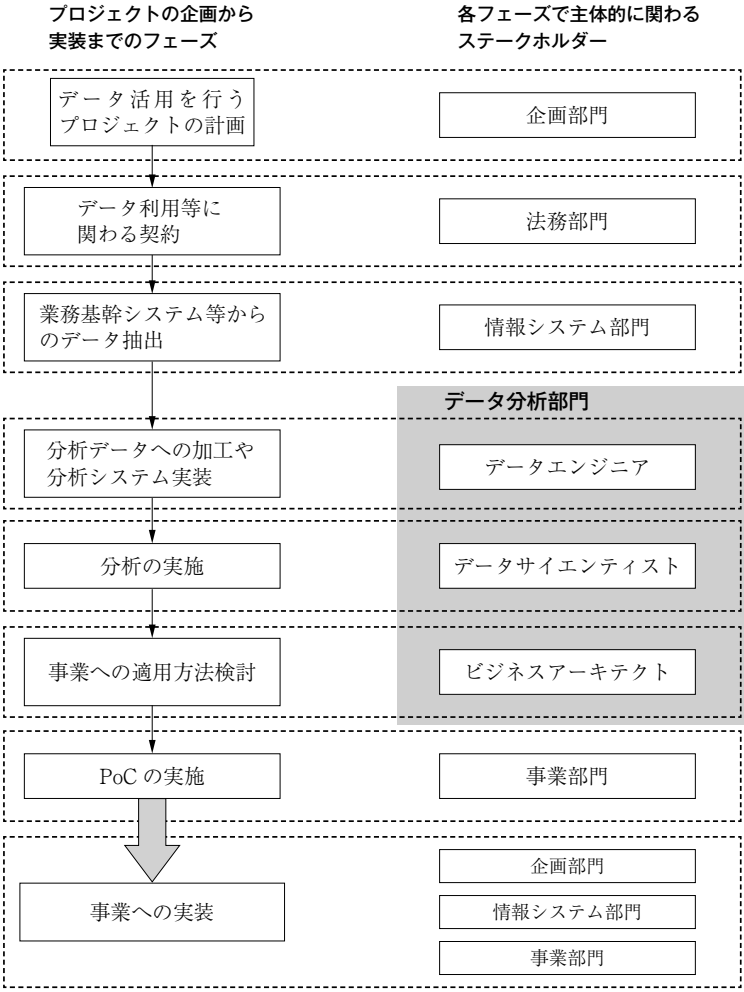


**図 1.2**　データ活用の実装までの流れとステークホルダーの例

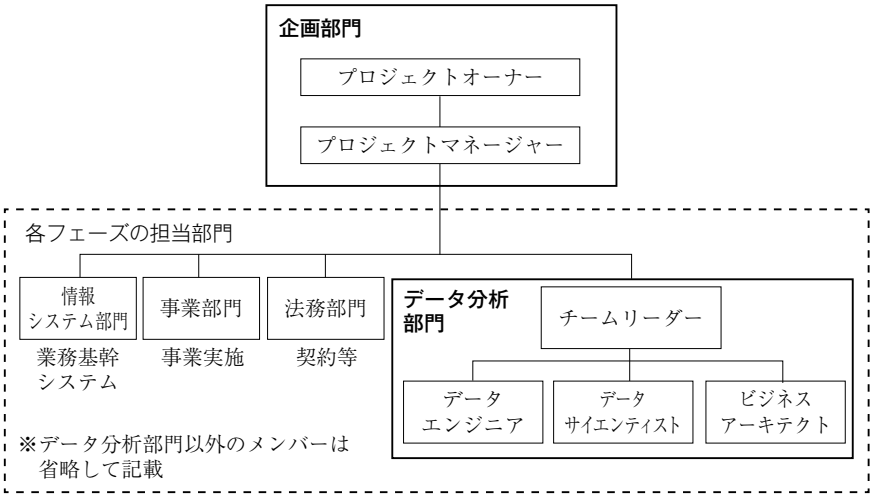


**図 1.3**　データ活用に必要なプロジェクトチームの例

データの活用に対する理解が十分でない場合，データが分析要件を満たさないなどの技術的な問題が発生したり，契約上の要件が不十分で不適切な個人情報の利用や使用条件の違反などのコンプライアンス上の問題が発生し，プロジェクトの失敗につながる恐れもある。

## 1.2　従来型プロジェクトとデータ活用型プロジェクトの定義

IoT（internet of things）を活用して安価に収集できるデータや，さまざまな業務システム等によって収集され蓄積されたデータなどのいわゆる「ビッグデータ」として活用されるデータの取り扱いは，従来の調査，実験での観測で得られたデータの取り扱いと本質的な違いがある。このことを明確にするために本書では，調査・実験での観測によってデータ収集を行う「従来型プロジェクト」と，いわゆるビッグデータ等の既存データを中心としたデータを用いる「データ活用型プロジェクト」という区分を定義し，その違いを説明する（**表 1.1**）。

従来型プロジェクトでは，実験計画や調査設計，データ収集機器の開発などのデータ収集のためのフェーズをプロジェクト内に内包していることを前提と

# 索　　引

―― 著 者 略 歴 ――

2006 年　神戸大学工学部建設学科卒業
2008 年　神戸大学大学院自然科学研究科博士前期課程修了（建設学専攻）
2010 年　神戸大学大学院工学研究科博士後期課程修了（市民工学専攻），博士（工学）
2010 年　日本学術振興会特別研究員 PD
2011 年　東京工業大学助教
2016 年　東京大学講師
2021 年　東京大学准教授
2021 年　阪急阪神ホールディングス株式会社データアナリシスディレクター
　　現在に至る

2022 年　東京大学特任准教授（～ 2023 年）
2022 年　東京大学空間情報科学研究センター客員研究員（～現在）
2023 年　株式会社ソーシャル・デザイナーズ・ベース代表取締役社長（～現在）


# データ活用型プロジェクトのマネジメント

Project Management for Data Utilization and Application




2024 年 3 月 27 日　初版第 1 刷発行　★

検印省略

著　者　日下部（くさかべ）　貴彦（たかひこ）
発行者　株式会社　コロナ社
　　代表者　牛来真也
印刷所　壮光舎印刷株式会社
製本所　株式会社　グリーン

112-0011　東京都文京区千石 4-46-10
**発 行 所**　株式会社　**コ ロ ナ 社**
CORONA PUBLISHING CO., LTD.
Tokyo Japan
振替00140-8-14844・電話(03)3941-3131(代)
ホームページ　https://www.coronasha.co.jp

ISBN 978-4-339-05281-7　C3051　Printed in Japan　（新井）I