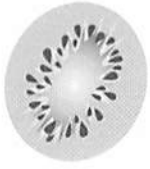
ビューティフルデータ

Toby Segaran
Jeff Hammerbacher 編

堀内孝彦、真鍋 加奈子、苅谷 潤
小俣 仁美、篠崎 誠 訳

O'REILLY®
オライリー・ジャパン

Beautiful Data



Edited by Toby Segaran and Jeff Hammerbacher

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo

はじめに

『ビューティフルコード』に続く本をデータを題材にして作らないかと提案されたとき、それは素晴らしい、と乗り気になった。とても野心的な試みだ。いまやあらゆる専門分野や日常生活の様々なところで、データの収集、可視化、処理が行われている。その幅広さには目を見張る。そこで我々は、賞賛に値する仕事をしている、非常に広範な領域の人々に声を掛けた。予想以上に多くの人々が協力してくれるというので、驚いた。

結果としてできあがったのが本書である。データを扱う仕事はあらゆる分野に存在し、しかもそれは美しいものだということを読者に伝えたい。題材は、役所との戦い、火星探査機、統計プログラムの利用方法、可視化、Radioheadのビデオ編集、地図、DNA、そして"データ哲学"としか言いようのない何かまで、多岐にわたる。

本書（原書）の印税は、データを自由に解き放つことで世界をよりよい場所にしようとする2つの団体、クリエイティブコモンズとSunlight Foundationに寄付される。データとの出会いで世界をどう変えることができるのか、読者自身で考えてもらえれば幸いだ。

本書の構成

各章は、大まかにデータの収集、保存、編成、検索、可視化、分析の順番で並んでいる。

1章「データの中に生活をみる」はNathan Yauによるもので、パーソナルデータ収集という新分野の2つのプロジェクトにおける動機と挑戦について述べる。

2章「ビューティフル・ピープル」はJonathan Follett、Matthew Holmによるもので、ウェブを通して人々からデータを集める際には信頼、説得、テストが大事であると説く。

3章「火星上での組み込み画像処理」はJ. M. Hughesによるもので、宇宙探査という制約下で動作するデータ処理システムの設計における挑戦について議論する。

4章「PNUTShellにおけるクラウドストレージの設計」はBrian F. Cooper、Raghu Ramakrishnan、Utkarsh Srivastavaによるもので、現代的なウェブアプリケーションを支えるために、世界中に分散したデータセンタを一つの共通ストレージプラットフォームへと統合する目的でYahoo!が設計したソフトウェアについて述べる。

5章「情報プラットフォームとデータサイエンティストの登場」はJeff Hammerbacherによるもので、Facebookにおけるデータチームの経験を元に、情報処理ツールの進化とそれを活用する人間を追う。

6章「写真アーカイブの地理的な美」はJason DykesとJo Woodによるもので、人々が無償で提供するありふれたデータが、空間的データとしてカラフルに視覚化したときに持つ力について語る。

7章「データの自己発見」はJeff JonasとLisa Sokolによるもので、多くの人が必要とするであろう、データを全体として取り扱うための新しい考え方を説明する。

8章「リアルタイムのポータブルデータ」はJud Valeskiによるもので、ウェブを通じてソーシャルデータ、ロケーションデータを配布するときに起きている限界に切り込み、この問題の解となりうる方式について議論する。

9章「ディープウェブを活用する」はAlon HalevyとJayant Madhavanによるもので、今はまだウェブのフォームの裏側にとらわれているデータを検索可能にするためにGoogleが開発したツールについて述べる。

10章「Radiohead「House of Cards」のプロモーションビデオができるまで」はAaron KoblinとValdean Klumpによるもので、賞をとったミュージックビデオ制作の裏側における、レーザー、プログラミング、撮影現場についての冒険物語だ。

11章「都市データの視覚化」はMichal Migurskiによるもので、我々の身近な世界で最も重要なあるデータを、オーブンにして美しく加工するまでの経緯を詳しく説明する。

12章「sense.usの設計」はJeffrey Heerによるもので、データの可視化をソーシャルな活動として捉え直し、この新しい視点を元に150年におよぶアメリカの国勢調査データを探求する。

13章「データでできないこと」はCoco Krummeによるもので、人々が犯しやすいデータについての誤解と誤使用をわかりやすく示した。

14章「自然言語のコーパスデータ」はPeter Norvigによるもので、ウェブから収集した1兆語の自然言語コーパスデータを用いて、いくつかの刺激的な例題を読者に案内する。

15章「データの中の生命：DNA物語」はMatt WoodとBen Blackburneによるもので、DNAの中のデータの美しさと共に、そのデータを生成、収集、加工する巨大なシステム基盤について述べる。

16章「実世界のデータをビューティフルにする」はJean-Claude Bradley、Rajarshi Guha、Andrew Lang、Pierre Lindenbaum、Cameron Neylon、Antony Williams、Egon Willighagenによるもので、クラウドソース化と思い切った透明性の組み合わせにより、新薬発見研究の世界を前進させる試みについて紹介する。

17章「外見のデータ解析：数百万人の社会的ステロタイプ調査」はBrendan O'ConnorとLukas Biewaldによるもので、他人の顔写真を人々に匿名で評価してもらった際に現れる相関とパターンを紹介する。

18章「ベイエリア・ブルース：住宅市場崩壊の影響」はHadley Wickham、Deborah F. Swayne、David Pooleによるもので、ベイエリアにおける近年の住宅市場崩壊をオープンソースソフトウェアと公開データに基づき詳しく追った。

19章「政治に関するビューティフルデータ」はAndrew Gelman、Jonathan P. Kastellec、Yair Ghitzaによるもので、統計というツールとデータの可視化により、社会を組織する政治プロセスについての洞察を得た事例を紹介する。

20章「データをつなぐ」はToby Segaranによるもので、ウェブから得られる多くのデータセットをつなぎ合わせる難しさとその可能性について探求する。

本書の表記

本書では、以下の表記を使用しています。

等幅(Constant Width)
サンプルコードを示す。

サンプルコードの使用について

本書の目的は、読者の仕事の手助けをすることです。一般に、本書に掲載しているコードは各自のプログラムやドキュメントに使用してかまいません。コードの大部分を転載する場合を除き、出版社に許可を求める必要はありません。たとえば、本書のコードブロックをいくつか使用するプログラムを作成するために、許可を求める必要はありません。なお、O'Reillyから出版されている書籍のサンプルコードをCD-ROMとして販売したり配布したりする場合には、そのための許可が必要です。本書や本書のサンプルコードを引用して問題に答える場合、許可を求める必要はありません。ただし、本書のサンプルコードのかなりの部分を製品マニュアルに転載するような場合には、そのための許可が必要です。

出展を明記する必要はありませんが、そうしていただければ感謝します。出展を明記する際には、Toby Segaran、Jeff Hammerbacher編「ビューティフルデータ」(O'Reilly Media, Inc.)のように、タイトル、著者、出版社、ISBNなどを盛り込んでください。サンプルコードの使用について、正規の使用の枠を超える、またはここで許可している範囲を超えると感じる場合は、permissions@oreilly.com までご連絡ください。

意見と質問

本書(日本語翻訳版)に関するコメントや質問は以下に送付してください。

株式会社オライリー・ジャパン
〒160-0002　東京都新宿区坂町26番地27　インテリジェントプラザビル1F
電話　03-3356-5227
FAX　03-3356-5261
電子メール　japan@oreilly.co.jp

本書に関する技術的な質問や意見につきましては、次の宛先に電子メールを送ってください。

bookquestions@oreilly.com(英語)
japan@oreilly.co.jp(日本語)

本書のWebページには、正誤表、サンプルコード、追加情報が掲載されています。以下のアドレスでアクセスできます。

http://www.oreilly.com/catalog/9780596157128/
http://www.oreilly.co.jp/books/4873114897/

オライリーに関するその他の情報(文献、会議、リソースセンター、O'Reilly Network)については、

次のオライリーのWebサイトを参照してください。

http://www.oreilly.com

http://www.oreilly.co.jp

目次