

グローバル企業 におけるオープ ンデータの価値

世界最大のオープン エンティティ
グラフを、コラボレーション、イノ
ベーション、そして変革の原動力へ
と変える

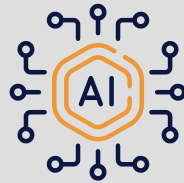
June 2026

Kirsten D. Sandberg

Foreword by Gabriele Columbro, *FINOS*

グローバル企業におけるオープンデータの価値

“「オープンデータは、次世代のエージェント型AIにとって不可欠な基盤となります。」
- Jim Zemlin, CEO, Linux Foundation.



OpenData.Orgのデータセットは、222カ国にわたる**3億2400万の組織**、12億人、5億1200万の場所を網羅しており、そのすべてが10万件以上の検証済みの公開情報源に基づいています。



共有オープン識別子は、サイロ化されたシステム間のエンティティを結びつけ、重複レコードを排除し、データシステムを整合させ、企業に「**単一の信頼できる情報源**」を提供します。
(Gartner).



OpenData.Org 構造化識別子とメタデータにより、LLMは検証済みの事実に基づいて構築されるため、AIエージェントはハルシネーションや誤った関連付けを起こすことなく推論を行うことができます。

エンティティ解決やデューデリジェンスのプロセス全体にエージェント型AIを導入する金融機関は、**生産性を2000%も飛躍的に向上させることができます**。
可能性があります。



エンティティデータを公開することで、法人をその物理的な所在地、サプライヤーネットワーク、および原材料と結びつけることにより、**サプライチェーンおよびESG報告の全体像**を把握できるようになります。



各組織は、同じ中核的なエンティティデータのクリーニングに、総額で数十億を費やしています。オープンデータを通じてリソースを共有することで、**運用コストを削減**し、データの精度を向上させることができます。
(McKinsey/GLEIF)



2023年、承認済みプッシュ決済による不正利用額は1兆ドルを超えました。**エンティティ照合データを共有**することで、金融機関は受取人の本人確認、マネーロンダリングの仲介口座の検知、および異常の特定が可能になります (Experian)



OpenData.Orgは、設計段階から**プライバシーを最優先に考慮**しており、GDPR、CCPA、および世界各国のプライバシー規制に完全に準拠しています。また、ガバナンスプロセスにおいては、「忘れられる権利」に基づく要請にも適切に対応しています。



独自仕様でサイロ化されたエンティティデータの時代は終わりを告げようとしています。OpenData.Orgは、信用情報機関の識別子や商用ファームグラフィック・ベンダーに代わる、オープンな選択肢を提供しています。



「エージェントを把握する」とは、標準化されたオープンなエンティティ識別子が、企業間ワークフローにおける**エージェントの身元認証のための信頼基盤**となる、コンプライアンス分野の新たなフロンティアです。



「オープンライセンスで提供され、高品質かつ共同で維持管理されるデータセットは、あらゆるAI技術プロバイダーや導入企業にとり、まさに信じられないほどの機会となります。」
- Gabriele Columbro, FINOS 事務局長

グローバル企業におけるオープンデータの価値：主な活用事例

データ品質の低さにより、企業は数百万の損失を被っています。**エンティティ解決とマスターデータ管理**により、レコードを関連付け、重複を排除し、統一されたビューを構築することで、より適切な意思決定が可能になります。



オープンデータとAIは、共有エンティティグラフや自動化されたワークフローを通じて、**貿易コンプライアンス**の効率化、オンボーディングコストの削減、企業間での本人確認の向上、および不正検知の迅速化を実現します。



オープンなエンティティグラフを活用することで、銀行のコストを削減し、**資本市場分析**を精緻化し、企業間の関係を整理することで、オープンデータセット全体にわたるリスク検出、ポートフォリオのスクリーニング、および規制に関する知見の獲得が可能になります。



AIエージェントを標準化されたエンティティデータ、ナレッジグラフ、および出所追跡に基づいて構築することで、誤った情報を生成する現象を減らし、精度と検索能力を向上させ、ドメイン横断的な知能を強化することができます。

検索機能を活用した生成技術は、構造化情報と非構造化情報をLLMと組み合わせることで、事実に基づいた文脈豊かな応答の質を向上させ、リアルタイムのエンティティ解決を支援し、リスク領域全体にわたる分析能力を強化します。



マルチエージェント・インフラストラクチャは、エンティティグラフ、エージェントIDフレームワーク、ゼロトラスト型のIDおよびアクセス管理アーキテクチャ、相互運用性プロトコルを統合し、**Know-Your-Agent**アプリケーションのワークフローとガバナンスモデルを強化します。



グラフベースのクレジットモデルは、企業体と財務データを結びつけることで、リスク評価の精度を高め、経営難の兆候を検知し、与信判断の参考とし、市場を横断したクレジットモニタリングを強化します。



単一エンティティの視点では、**顧客生涯価値**を限定してしまいます。統合されたエンティティグラフを活用することで、顧客をより包括的に把握し、成長の可能性を明らかにし、商機を優先順位付けし、長期的かつ収益性の高い顧客関係を深めることができます。



環境・社会・ガバナンス（ESG）報告では、**サプライチェーンのパフォーマンス**を評価するために、サプライヤー、流通業者、労働者、資材、資金の流れを結びつける、位置情報を認識するエンティティグラフが不可欠です。



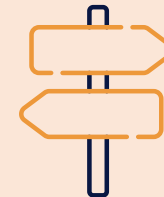
エンタープライズ インテリジェンスは、公開されているウェブ上のシグナルやオルタナティブ・データを活用し、従来の構造化データセットを利用するユーザーよりも早く、合併、新製品の発売、戦略の転換などを検知します。



小売店舗や倉庫の立地選定には、人口統計、移動状況、競合他社の立地に関するデータが必要です。オープンデータを活用することで、立地の一極集中を緩和し、**インフラや物流に関する意思決定**をより的確に行うことができます。



地理空間データや経済データは、**災害対策、復旧、および資源配分**の指針となります。公開データセットやAIモデルは、国境を越えた連携、ガバナンス、そして危機後のレジリエントな開発を促進します。



Contents

序文	6
エグゼクティブサマリー	7
エンティティの世界のマッピング	8
イノベーションに活用可能なオープンなエンティティデータセット	11
エンティティデータの公開が急務に	13
オープンデータプロジェクトを推進する勢力	16
オープンなエンティティデータの活用事例	16
このプロジェクトに参加した理由	26
ご支援の方法	28
新しいデータプロジェクトの立ち上げ	30
オープンデータの価値を認識する	30
インセンティブモデルの確立	32
プロジェクトの使命と価値観の設定	32
適応性を考慮した設計	33
ガバナンスへのユーザーの参画	34
高品質なデータの維持	36
リスクの管理と信頼の構築	36
オープンイノベーションのためのライセンスの選定	37
コミュニティとコミュニケーションをとる	38
世界中の優秀な人材の発掘、育成、および人材確保	39
最小限の実用可能な製品の推進	39
謝辞	41
クレジット	42
著者について	42
付録	43
参考文献	47

序文

専有データやサイロ化されたデータの時代は、「オープンな共有」という新しいパラダイムへと移行しつつあります。Fintech Open Source Foundation (FINOS) のエグゼクティブ・ディレクターとして、私はオープンソースのコラボレーションが金融サービス業界をいかに変革しうるかを目の当たりにしてきました。しかし、皆様のお手元にある本報告書『グローバルエンティティにおけるオープンデータの価値』が示すように、私たちははるかに大きな変革の瀬戸際に立っています。その変革は、ソフトウェアの枠をはるかに超え、データに支えられ、AIによって加速される現代のグローバル経済の基盤そのものにまで及ぶものです。

金融サービス業界において、オープンデータの合理性は否定できません。業界全体で、まったく同じコアエンティティデータを整備・維持するために数十億単位の費用が費やされています。リソースを統合して高品質なデータの「共通基盤層」を維持することで、各機関はオンボーディングや取引処理コストを推定10%削減できます。しかし、このオープンなグローバルエンティティグラフの可能性は、単なる業務効率化にとどまらず、業界そのものを変革するものです。

しかし、エージェントによる企業横断的なワークフローがビジネスのあり方を再定義する未来に向けて進むにつれ、その意義は効率化をはるかに超えたものとなります。AIエージェントが自律的かつ確実に動作するためには、事実に基づいていなければなりません。高品質で標準化された参照データがなければ、大規模言語モデル（LLM）は「ハルシネーション」や誤った関連付けを起こしやすくなります。

3億2400万以上の組織と12億人以上の人々に関する、構造化され検証済みのディレクトリを提供することで、OpenData.Orgのようなイニシアチブは、AIが世界中の市場において正確に推論を行うために必要な「ロゼッタストーン」を提供しています。さらに、「オープンウェブ」の基盤そのものが、LLMのトレーニングの動向によって脅かされているこの時代において、オープンライセンスで提供され、高品質かつ相互に維持管理されるデータセットは、あらゆるAI技術プロバイダーや導入企業にとって、まさに信じられないほどの機会となっています。

これは、あらゆるデータ駆動型産業に波及する機会です。小売や経済開発の分野では、正確にジオコーディングされた空間データにより、起業家は立地を最適化し、政府はサービスが行き届いていない経済回廊の格差を解消することができます。サイバーセキュリティの分野では、ドメインやインフラの所有権を明確にマッピングすることで、運用者は上流の障害や乗っ取られた資産からネットワーク構造を保護できるようになります。環境・社会・ガバナンス（ESG）報告やサステナビリティに配慮した投資の分野においても、このオープンデータセットは、企業とその物理的なサプライチェーンをつなぐ欠けていたリンクを提供し、真のサステナビリティに必要な透明性を実現します。

本レポートで詳述されているユースケースやインセンティブモデルをぜひご検討ください。次世代のAIモデルのトレーニングを行うにせよ、世界貿易における摩擦の解消を目指すにせよ、その道は開かれています。私たちが力を合わせれば、世界最大のオープンエンティティグラフを、グローバルなイノベーションの触媒へと変えることができるでしょう。

Gabriele Columbro

Executive Director, Fintech Open Source Foundation

エグゼクティブ サマリー

組織、業界、サプライチェーンを越えてデータのサイロを打破し、エンティティデータを統合する時が到来しています。多者間連携により、グローバルなエンティティ解決に欠けていた参照キーセットを提供することが可能になります。

Open entity data project. 法律により、組織は取引相手を把握する義務があります。エンティティの身元を迅速に特定するために、OpenData.orgのようなオープンデータ・プロジェクトが提供する「グローバル エンティティ グラフ」は、標準化された識別子とデータの透明性をもって、数百万のエンティティ間の関係性を示しています。このデータセットは検証済みの公開情報源から収集され、毎月更新され、データ駆動型のイノベーションを促進するために柔軟なアクセスと利用を提供しています。

プロジェクトの成功に必要なリソース。データの品質、アクセシビリティ、網羅性、目的適合性、および有効性を長期的に維持するためには、このようなデータプロジェクトには、大規模なガバナンスのための構造的、法的、および財政的な支援が必要です。そこでBrightQueryは、マルチステークホルダー間の協力の基盤として、初期段階のグローバルエンティティデータセットをまとめました。Senzingは、システムを横断してレコードを実世界のエンティティに紐付ける解決パートナーとして参画しています。

オープンデータの緊急の必要性。オープンなエンティティデータは、AIを検証済みの事実に基づいて運用するのに役立ちます。このようなデータセットは、着実に増加するコンプライアンスコストに対処するとともに、組織にプロプライエタリなソリューションに代わる柔軟で透明性の高い選択肢を提供します。共有リソースとして、オープンデータセットはシステムのロックインを軽減し、相互運用性を向上させ、拡大するデータエコシステム全体での共同貢献を通じて着実に進化していきます。

業界を横断する強力なユースケース。オープンエンティティデータセットのユースケースには、エンティティ解決やマスターデータ管理、貿易コンプライアンス、顧客確認（KYC）およびマネーロンダリング防止（AML）チェック、さらにAIエージェントやAIモデルを事実に基づいて定着させ、検索や応答の精度と速度を向上させることが含まれます。その他には、資本市場、企業信用、顧客関係管理（CRM）、サプライチェーン、ESG報告などが挙げられます。

メンテナンスでは協力し、イノベーションでは競争する。起業家や組織は、中小企業が単独では構築できない、差別化につながらないインフラの維持コストを分担することができます。これにより、重複する作業を最小限に抑えつつ、すべての関係者がより強力な機能、より優れたコンプライアンスツール、そしてAIの品質向上を享受できます。このような「協調的競争」は、リソースを共有し、データクレンジングのコストを削減するとともに、人材の育成、発明、そしてより広範な経済成長を促進します。

戦略的洞察と市場の需要シグナル。このようなオープンな企業データセットから最大の恩恵を得られる組織は、貿易コンプライアンスのコストと複雑さを低減し、顧客との連携を効率化してオンボーディングの摩擦を軽減し、信頼性の高い参照データに基づいてAIモデルを構築し、サードパーティのリスクを評価してサプライチェーンの完全性を維持しようとしている組織です。

エンティティの世界のマッピング

早くも紀元前6世紀の頃から、指導者たちは経済活動において人々の身元を特定することの価値を認識していました。ローマの人口調査や『後漢書』から、『Domesday Book』やチューダー朝時代の教区登録簿に至るまで、行政機関は社会において「誰が誰であり、誰が何を所有しているか」を記録しようと努めてきました。¹ 地域の産業経済がグローバルなデジタル経済へと移行するにつれ、不正防止機能を備えた拡張性の高いエンティティ解決ソリューションへのニーズはますます高まっています。不正行為の検知と抑止にかかるコストは、ほぼすべての金融機関で上昇しており、2023年の世界全体の総支出額は2,061億ドルに達した。²

より根本的な問題はアーキテクチャにある。初期のウェブの先駆者や検索エンジンの設計者たちは、情報を文書とその統一リソースロケータ（URL）を中心に整理し、ウェブページの順位付けを主に人気度に基づいて行っていた。30年にわたり、この設計上の選択——特に、頻繁にリンクされているページを権威あるものとみなす点——が、情報検索のあり方を決定づけてきた。Sir Tim Berners-Leelは、この設計が真のユーザー価値よりも、クリックベイトやバイラル拡散を優遇してしまうと警告している³。LLM（大規模言語モデル）は、この同じメカニズムをより洗練された形で表現したものである。文書内で「ABC社」や「ジョン・スミス」といった言及があった場合、LLMは周囲のテキストから確率的な推測によってその参照先を特定する。これは、よく知られたエンティティについては通常正確だが、それ以外の大部分については不確実である。

情報の構成要素は文書ではなく、エンティティ、すなわち組織、人物、場所、政府、製品、そして現在それらに代わって行動しているAIエージェントである。原子が物質を構成するように、エンティティは情報を構成しており、したがって信頼性の高い基盤層はそこから始めなければならない（表1）。

表1：情報の基盤となる層

	THE DOCUMENT WEB (LOCATORS PLUS RANKINGS)	THE ENTITY WEB (KNOWLEDGE GRAPHS)
Its unit of truth	Document container (webpage, PDF, post, etc.)	Fact/relationship (subject-predicate-object)
How users find it	Uniform resource locator (URL)	Uniform resource identifier (URI)
What indicates its value	Popularity (number of pages linking to it)	Consensus and provenance (verifiable lineage)
How AI reads it	Probabilistic guessing (AI predicts words)	Deterministic logic (AI follows exact paths)

オープンなエンティティデータセットに関するグローバルな協働は、こうしたビジネス上の課題に対するオープンソースソリューションを開発する者にとって絶好の機会である。OpenData.Orgで公開されているオープンエンティティグラフの構築は、その代表的な例である。これは、組織、場所、住所、人物といった経済の主要な構成要素（図1）を網羅する世界規模のディレクトリとして機能している。各エンティティには一意の識別子と、スキーマ、フィールドの説明、構造情報、データ定義を含む透明性の高いメタデータが付与されている。⁴

このグラフは、米国だけでなく、世界222の国と地域にわたる3億2400万のエンティティを網羅している（図2、次ページ）。⁵ このプロジェクトは、デラウェア州の企業BrightQueryによるオープンデータイニシアチブとして始まった。同社は、OpenData.Orgは、信用調査機関や企業属性・事業体データを提供する同種のベンダーによるプロプライエタリなIDシステムに代わるオープンソースのデータソリューションとして構想され、事業体に関する法的観点とビジネスの観点の両方を提供することを目指している。”⁶

図1：オープンエンティティグラフにおける経済の主要な要素

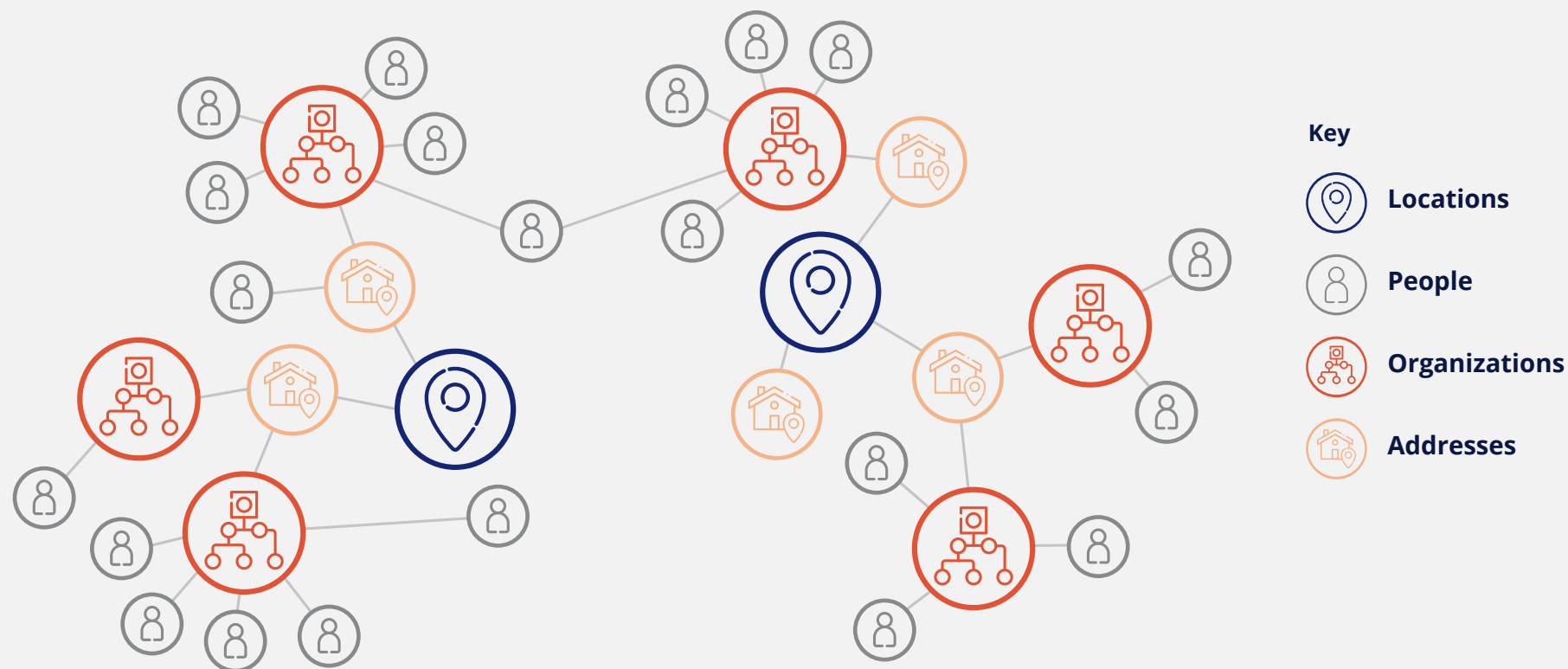
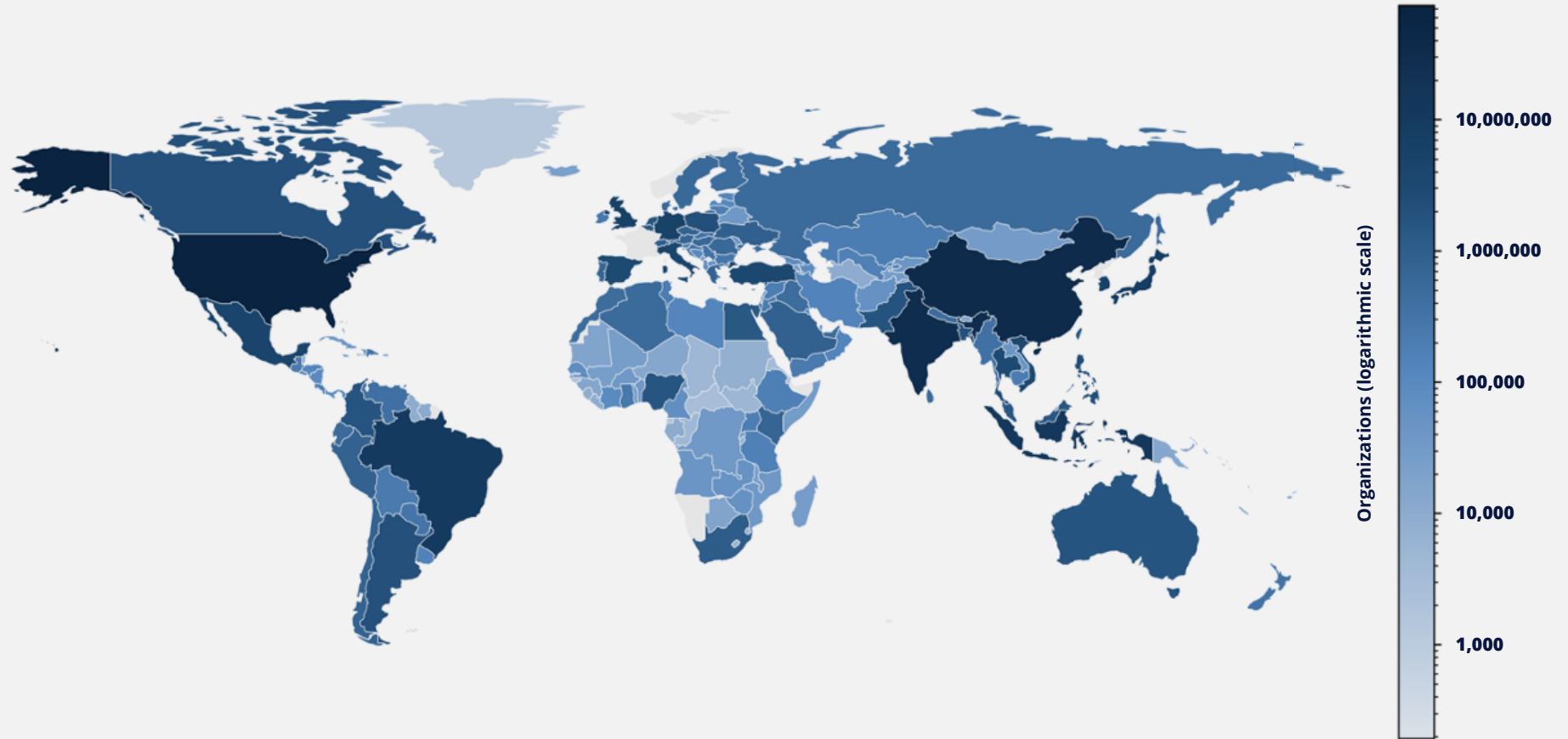


図2：国別の組織の分布

このオープンなエンティティグラフは、世界222の国と地域にわたる3億2400万のエンティティを網羅しています。対数スケールにより、分布のばらつきが激しい場合でも可視性が向上します。



Source of data: Shay Herskovitz, email to Kirsten Sandberg, 31 Mar. 2026. Map created by and used with permission from Brian Malone, who used a navy gradient to shade countries by organization count.



イノベーションに活用可能なオープンなエンティティデータセット

Global Legal Entity Identifiers、OpenData.Org、OpenSanctionsなど、エンティティの特定に役立つオープンデータ資産は、個人や組織が、あるエンティティが自称する通りであるかどうかを判断するのに役立ちます。

例えば、OpenData.Orgのデータセットは、証券マスターデータや価格設定サービス、市販の参照データセット、非上場企業に関するデータなどを補完し、金融機関が重視する中核的なエンティティの大部分を網羅しています。FINOSのゴールドメンバーによって理事会代表に選出されたプレーン博士のリーダーシップの下、BrightQueryは以下の特徴を備えた、高い基準を満たす初期の実体データセットをキュレーションしました。

これまでに10万件以上の検証済み公開情報源：これまでのところ、データは83,800を超える米国の政府機関、32,300を超える米国の地方自治体、23の米国連邦機関、51の米国州務長官事務所、および28のカナダの機関・登録機関から収集されており、民間や専有データのスクレイピングによるものではありません。⁷ データの精度を高めるため、本プロジェクトでは主にオープンガバメント・リポジトリからエンティティデータを取得しています。

継続的な網羅性、月次更新：このデータセットは2017年から現在までを網羅しており、ほとんどのデータは2010年まで遡ることができ、米国国勢調査局の経済センサスが5年ごとに更新されるのに対し、ほとんどのデータは毎月更新されています。⁸

プライバシーコンプライアンス：このデータセット、データ収集方法、およびデータ管理慣行は、一般データ保護規則（GDPR）、カリフォルニア州消費者プライバシー法（CCPA）、および世界的なプライバシー規制に完全に準拠しています。また、ガバナンスプロセスでは、「消去権」および「削除権」の規定に基づく「忘れられる権利」の要請に対応しています。⁹ データは主に政府の情報源から取得されているため、企業が自社のデータの削除や消去を要請することはほとんどありません。

プライバシーと有用性の最適なバランス：個人のプライバシーと安全を保護しつつ、組織に有用な地理的コンテキストを提供するため、このデータセットには国および州または県の情報のみが含まれており、郵便番号や住所は含まれていません。

公開ではなく参照によるプライバシー：共有エンティティ層は、機密性の高い個人データの流通を拡大するのではなく、縮小させることができます。今日、組織は主に本人確認のために個人を特定できる情報を保存・交換しています。もし各エンティティが、暗号ハッシュのように機能する永続的で不可逆的な公開識別子を保持していれば、銀行、病院、雇用主、または取引相手は、あらゆる情報漏洩の危険性がある基礎データを保存したり送信したりすることなく、その識別子を参照することで、取引相手が誰であるかを確認できるようになります。データのトークン化や選択的開示と組み合わせることで、このようなシステムはプライバシーの拡大ではなく、むしろ縮小につながるだろう。¹⁰

AIネイティブかつ開発者重視の製品：OpenData.Orgは、AIワークフロー、アプリケーション、チャットボットに文脈豊かなデータを提供できるようデータを整備しました。「このプロジェクトの当初のきっかけは、対話型LLMで見られた『ハルシネーション（誤った情報生成）』や『ミスアトリビューション（誤った関連付け）』でした」と、プレーン博士は述べています。¹¹OpenData.Orgの構造化された識別子とメタデータは、AIシステムが類似した名称を持つエンティティを区別するのに役立ちます。各識別子は、以下の3つの標準に基づいて構築されています：

- **一意性**（2つのエンティティが同じ識別子を共有しないようにするため）；
- **永続性**：名称の変更、再編成、または移動後も存続すること；および
- **解決可能性**：関連するレコードを確実に取得できるようにすること。

このような識別子は、人、組織、作品の間で信頼性の高いリンクを構築するため、OpenAIのChatGPT、GoogleのGemini、AnthropicのClaudeといった対話型LLMが、応答においてエンティティをより正確に特定できるようになります。「幻覚現象の解消を支援することは、この取り組みの大きな動機であり、大きな利点でもあります」とプレーン博士は述べました。「私たちは、事実に基づいたAIを望んでいます」

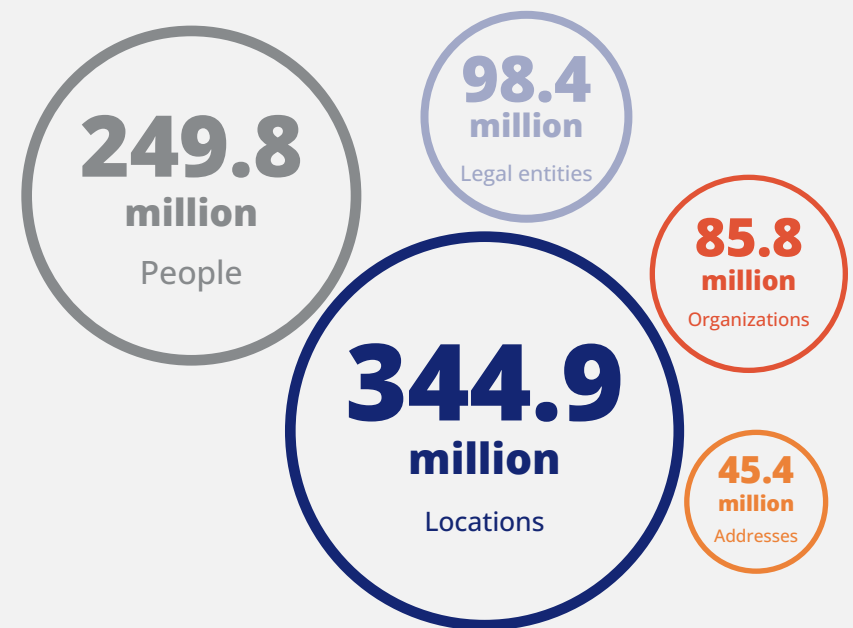
四半期ごとの更新による柔軟なデータ提供：ユーザーは、JavaScriptオブジェクト表記（JSON）、Senzing、その他の形式で利用可能なFTP（ファイル転送プロトコル）による一括ダウンロード、モデルコンテキストプロトコル（MCP）サーバー、アプリケーションプログラミングインターフェース（API）統合、そして将来的にはソフトウェア開発キット（SDK）を利用できます。¹²Senzingの「関連付け可能なデータ（CORD）」構造により、ユーザーはレコードを容易に照合でき、データ提供者はリアルタイムAPIやMCPサーバーにアクセスして、自然言語を用いてプログラマ的に検索できるようになります。

オープンデータとプレミアムデータ：このプロジェクトは、「オープン・コア・データ戦略」を採用しており、オープンデータの中核を基盤とし、その上でBrightQueryやその他の貢献メンバーがプレミアムデータを個別に販売できる仕組みとなっている。（詳細は付録を参照。）

BrightQueryは、Google Cloud PlatformおよびAmazon Web Servicesを通じてデータを提供しています。データセットを常に本番環境対応の状態に保つことを目指しており、そのため、破壊的な変更を導入することなく柔軟性を維持できるよう、データアーキテクチャを慎重に調整しています。また、Linux Foundationエコシステム内のテンプレート、フォーマット、プロセス、およびユーザーエンゲージメントのその他の側面を模倣することで、それらの仕組みに慣れているユーザーがすぐに馴染めるようにしています。

図3：OpenData.Orgによる米国組織の網羅範囲

例えば、OpenData.Orgの公開組織データセットは、米国経済のこれら5つの主要分野を網羅しています。



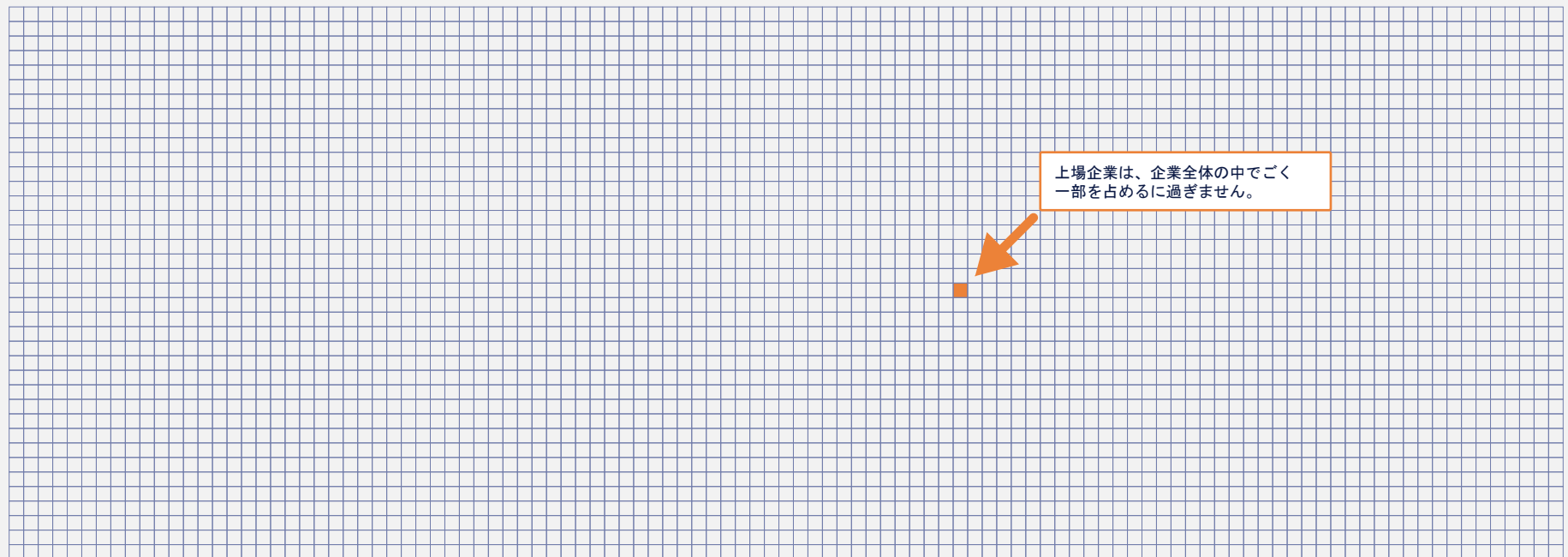
Source of data: "US coverage," BrightQuery documentation, as of 1 May 2026, <https://docs.brightquery.com/dashboard>.

エンティティデータの公開が急務に

人工知能の導入が過去最高を記録していることに加え、世界的な金融規制への準拠コストの上昇や、プロプライエタリなコンプライアンスソリューションの切り替えにかかる高額なコストが相まって、Linux Foundationのオープンアセット群における他のオープンデータプロジェクトと同様に、オープンエンティティデータプロジェクトを立ち上げる緊急性が浮き彫りになっています。¹⁴ 例えば、OpenData.Orgで入手可能なデータは、これら3つの課題に正面から取り組んでいます。

図4：OpenData.Orgによる米国企業の対象範囲

OpenData.Orgのデータセットには、下の図のオレンジ色の四角で示されている、米国の証券取引所にティッカーシンボル付きで上場している企業よりも、約12,000倍もの非上場企業（グリッドで表されています）が含まれています。



Source of data: "Record Counts," BrightQuery Documentation, as of 1 May 2026, https://docs.brightquery.com/record_counts.

AIの基盤化

Linux FoundationのCEOであるJim Zemlin氏によると、オープンデータは次世代のエージェント型AIにとって不可欠な基盤となり、LLMのトレーニングや推論時の幻覚を排除するために必要な高品質なグラウンディングを提供します。オープンなエンティティデータにより、標準化されたデータセット全体にわたる正確かつリアルタイムな事業体クエリが可能となり、信頼性が高く自律的な企業横断的なワークフローの開発が加速されます。¹⁵

世界銀行の見解によれば、データアグリゲーター、再配布業者、およびAI開発者は、共通のメタデータ標準を採用・適用し、「信頼できる情報源を優先し、ライセンスや解釈上のガイダンスを尊重し、すべてのAI出力を、強力な出所保証と、リスク管理、規制遵守、および責任あるAIのための追跡可能かつ監査可能な管理措置を備えた正規データに紐付ける」必要があります。¹⁶

このような信頼性の高いエンティティデータの層があれば、AIエージェントは組織内外のシステムを横断して正確な情報を迅速に取得できます。エンティティデータ標準を厳格に適用することで、エージェント型AIアプリケーションや自律型エージェントのワークフローの相互運用性が向上します。BrightQueryは、エコシステム全体の個人や組織によるデータ提供がもたらすネットワーク効果により、データ品質が継続的に向上すると考えています。Plehn博士は、参照識別子やデータソースのカatalogが拡大するにつれ、このデータセットが世界経済の発展に歩調を合わせていくと予想しています。

「オープンアセットは、AIツールの開発に必要な資本を削減しますが、実行上の非対称性を解消するわけではありません。」

オープンアセットは、AIツール開発に必要な資本要件を圧縮するが、実行上の非対称性を解消するものではありません。統計的なマクロデータを集約し、容易にアクセスできるように公開するGoogleのオープンデータイニシアチブ「Data Commons」の責任者であるプレム・ラマスワミ氏は、LLMとナレッジグラフを組み合わせることで、それぞれを個別に使用するよりも大きな効果が得られると考えています。

「もしLLMをナレッジグラフの情報で裏付けられれば——たとえそれが世界の情報のわずか1%に過ぎないとしても——LLMは不足している情報を推論し、私たちがより良い意思決定を行う手助けをしてくれるでしょう」と彼は述べました。¹⁷

コンプライアンスコスト

顧客確認（KYC）、銀行・事業者確認（KYB）、およびマネーロンダリング防止（AML）規制への準拠は極めて重要です。LexisNexis Risk Solutionsによると、これらへの対応コストはほぼすべての金融機関で上昇しています。欧州、中東、アフリカでは、組織が850億ドルを費やしており、これに対し、カナダと米国では610億ドル、アジア太平洋地域では450億ドル、ラテンアメリカでは150億ドルが費やされています。¹⁸

ブラックボックスとシステムのロックイン

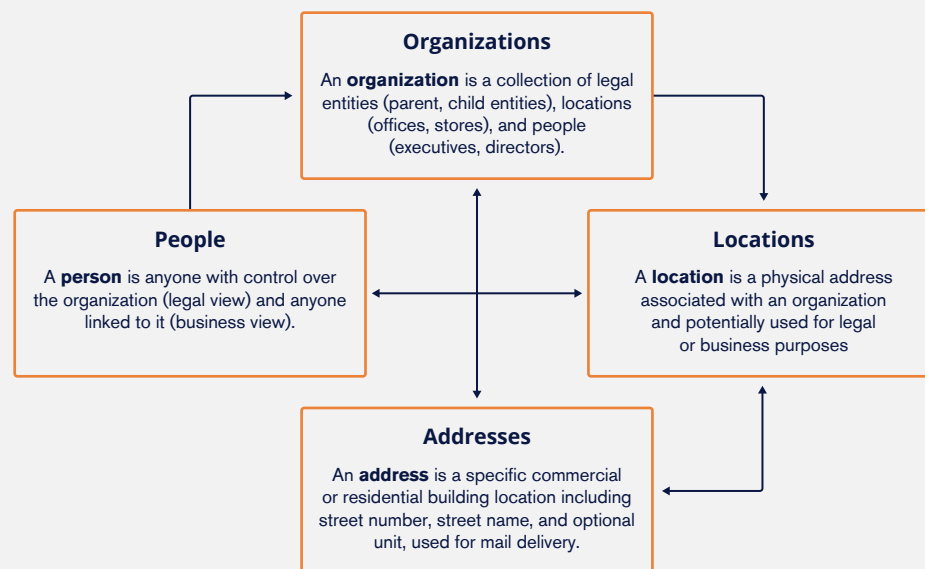
このレベルのグローバルなエンティティ解決を実現するため、組織は通常、プロプライエタリなデータフィードを購読するか、あるいは所有者の利用規約やサービスレベル契約（SLA）に基づき、分析プラットフォームを通じてそれらにアクセスしてきました。しかし、こうした契約では、組織が実行できることが制限される傾向があります。¹⁹一部の企業では、他の場所で使用するたびに所有者と交渉する必要がなく、より拡張性が高く、モジュール化されており、事業全体でより活用しやすいデータセットを必要としています。「Community Data License Agreement Permissive 2.0 (CDLA Permissive 2.0)」などのオープンライセンスは、データ利用者にデータの利用におけるより大きな自由度を与え、データが今後もオープンな状態を維持されるという保証を提供します。²⁰

こうしたサブスクリプションサービスは、通常、ツールやAPIを通じて、組織が閲覧できる範囲を制限する傾向があります。データは不透明な箱の中に閉じ込められています。ユーザーはデータベース内の変更内容を確認することができません。むしろ、ユーザーが見ることができるのは、通常はAPIクエリに基づいて、その時点においてベンダーが提供している情報のみです。例えば、OpenData.Orgのデータであれば、ユーザーは実際に差分比較を行い、データセットの経時的な変化を確認することができます。同様に重要なのは、こうしたオープンプロジェクトには、そのような変更を作成・実施するための明確な参加型プロセスが整備されていることです。

最後に、これらの企業は、ファームグラフィックス、信用履歴、コンプライアンス記録といった独自のデータセットを永続的な識別子と組み合わせ、そのワークフローを加入者の顧客関係管理（CRM）および企業資源計画（ERP）システムに組み込んでいます。しかし、OpenData.Orgとは異なり、これらの企業のデータスキーマは移植性がありません。（OpenData.orgのスキーマについては図5を参照）。いったん統合されると、データは業務の一部となるため、データサービスの切り替えは業務に支障をきたします。契約上の義務から脱却するのは困難です。SLA（サービスレベル契約）は依存関係を固定化し、移行リスクを高め、多くのユーザーによるネットワーク効果によって、独自仕様のデータスキーマが事実上の標準となってしまう。²¹

「経済データの収集・公開に関する代替手法を検討するには、まさに今が絶好のタイミングだ」と、Open Data Exchange（CODE）の代表兼創設者であり、画期的な著書『Open Data Now』の著者であるGurin氏は述べました。「明白な理由から、あらゆる立場の人々が『従来の連邦政府によるデータ収集に代わる方法はあるのか？』と問いかけている。オープンデータ・プロジェクトが、高い水準でそれを実現できるかどうかは、非常に大きな価値を持つだろう」と彼は述べました。²²

図5：オープンデータのスキーマと関連付け



Source: "Open data schema and linkages," OpenData.Org, as of 7 May 2026, <https://opendata.org/>; "Universe data relationship," BrightQuery, as of 7 May 2026, https://docs.brightquery.com/universe_data_relationship_schema. Used with permission.

オープンデータプロジェクトを推進する勢力

オープンデータ・プロジェクトは、何もないところから生まれるものではありません。OpenData.Org にとって、BrightQuery はデータセットの初期の管理者および提供者であり、これは、世界中の上場企業だけでなく民間企業も網羅しようとするBrightQuery の取り組みを反映したものです。また、BrightQueryは、米国国立科学工学統計センター（NCSSE）傘下の「米国国家セキュアデータサービス実証（NSDS-D）」における主要な政府パートナーであり、主要なAI企業への政府データの主要な供給元でもあります。²³ 同社はAIアライアンスの理事会メンバーを務め、エンタープライズ・データ・マネジメント・カウンシル、AIインフラストラクチャ・アライアンス財団、およびオーパチュア・マップスのメンバーでもあります。

BrightQueryとその協力パートナーは、この取り組みを中立的で独立したガバナンス体制を持つ組織に委ねることを提案しています。その中立性を維持するため、BrightQueryは、組織、人物、場所のディレクトリ、およびそれら間の検証済みの関連性を、寛容なライセンスの下で完全にオープンソースとして、自社に特別なガバナンス権限を持たせることなく、中立的な組織に寄贈する計画です。

Senzingは、本プロジェクトのエンティティ解決パートナーである。Senzingは、データプラットフォームが、異なるレコードがいつ同じ実世界のエンティティ（すなわち、人物、企業、住所、その他の対象）を指しているかを判断し、それらのレコードを正しくリンクまたは統合できるよう支援しています。²⁴

AIの基盤構築や、より高い透明性と柔軟性を備えたコンプライアンスコストの管理におけるエンティティデータへの差し迫った需要に応えるため、本プロジェクトには、中立的な拠点、オープンなガバナンス、世界クラスの参加者、そして501(c)の指定と組織構造が必要です。Plehn博士は、自身のオープンデータマニフェストの中で次のように的確に述べています。「人間と機械の知識の基盤となる層は、誰の所有物であってはならない。それはオープンであり、公共の利益のためにガバナンスが行われ、その決定がそれに依存する機関によって構築されるべきです。」²⁵

オープンなエンティティデータの活用事例

Plehn博士によれば、初期段階において最も有力なユースケースは、エンティティの解決と識別、資本市場、企業信用、販売・マーケティングという4つの分野にまたがっており、AIモデルのトレーニングはこれらすべての分野に共通して行われています。²⁶

エンティティ解決とマスターデータ管理

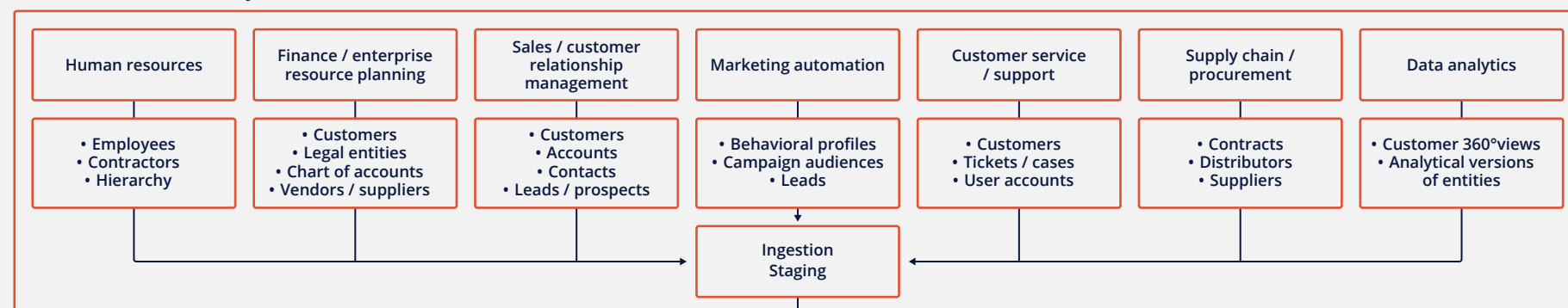
金銭や、データ、知的財産などのその他の貴重品に関しては、企業は取引相手がどの組織であるかを把握しておく必要があります。Gartner社によると、重複レコード、不正確さ、不整合など、コンプライアンスチームが是正しなければならないデータ品質の低さにより、組織は年間1,290万ドルの損失を被っていると推定されています。²⁷ 「意思決定の過程において、誰が誰であるかを正確に把握していないことは、多大なコストを招きます」と、Senzingの創業者兼CEO兼チーフサイエンティストであるJeff Jonas氏は述べています。²⁸ 同氏は、給与保護プログラム（PPP）を例に挙げました。PPPの申請データと連邦政府のデータベースとの間で、重複や不一致を検出するための管理体制が不十分だったため、一部の人が合計40億ドルに上る複数回の給付金を受け取っていたことが明らかになったのです。²⁹

オープンエンティティグラフは、エンティティと関係を明示的にモデル化し、エンティティ解決ワークフローを通じて複数のソースからのレコードをリンクさせることで、これらの問題を軽減するのに役立ちます（図6）。³⁰このようなアプローチにより、手作業によるデータ修正が削減され、システム間の整合性が確保され、個人や組織にとっての「唯一の真実の源（single source of truth）」が維持されます。これこそが、マスターデータ管理（MDM）が達成しようとしていることです。³¹

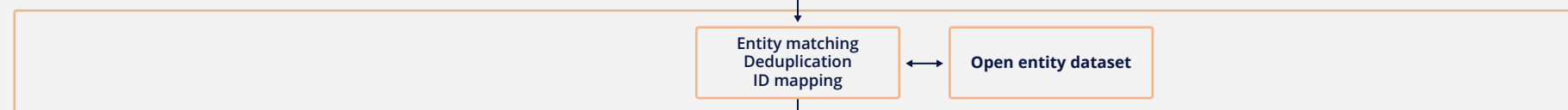
「このプロジェクトは、本質的にエンティティ解決のための『ロゼッタストーン』を作り出すものです」と、ProSight Financial Association のリサーチ、不正対策、およびソートリーダーシップ担当マネージングディレクター、Isio Nelson氏は述べています。「金融機関にデータのサイロがある場合、あるデータセットからエンティティ識別子を取り込み、他のデータセットでそのエンティティを見つけ、それらをリンクさせることで、各組織や個人について把握しているすべての情報を確認することができます。そこから、経営陣は生産性を高め、イノベーションを起こすために何が必要かをより明確に把握できるようになります。」³²

図6：サイロからインサイトへ

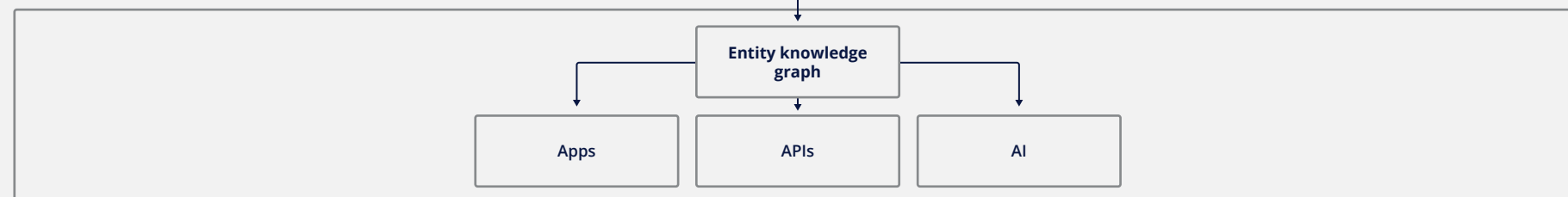
Silos of unresolved entity data



Entity resolution / master data



Resolved entity data in action



貿易コンプライアンス、顧客確認（KYC）、およびマネーロンダリング対策

KYC、KYB、およびAML規制への準拠も極めて重要です。McKinsey Global Instituteの推計によると、顧客登録時にデジタルで身元を特定することで、オンボーディングコストを最大90%削減し、世界全体で給与詐欺を1.6兆ドル削減できる可能性があります。³³

オープンデータは役立つ可能性があります。例えば、英国の企業公開登録簿の管理を担当する政府機関であるCompanies Houseは、企業の透明性を高めるため、2015年にオープンデータポータルを立ち上げました。³⁴ 調査によると、AML規制下にある企業にとって、この登録簿の推定総価値は年間1億7,000万ポンドから4億6,000万ポンドであり、法執行機関のユーザーにとっては、ユーザー1人あたり年間約2,600ポンドの価値を生み出していることが示されています。³⁵ 2024年には、ユーザーによる登録簿へのアクセス回数は165億回に達し、オープンデータの産業的価値は10億〜30億ポンドの範囲に上りました。³⁶

McKinsey & Companyによると、金融機関は、潜在的な金融犯罪のケース担当者や調査員を支援するためにエージェント型AIを導入することで、すでに生産性を15〜20%向上させています。もし、AIエージェントがエンティティ・リゾリューションを含むデューデリジェンスの各プロセスで連携し、人間の監督者が例外事項を処理する「デジタルファクトリー」を導入すれば、飛躍的な生産性向上（200〜2000%）を実現できるでしょう。³⁷

次のフロンティアとなるのは、企業間のワークフローです。そこでは、AIエージェントが有価証券の売買や、見積依頼（RFQ）や情報提供依頼（RFI）に関する交渉を支援することになります。エージェントの身元は、規制対象のワークフロー、会話の記録、規制当局への報告などにおいて、極めて重要な要素となるでしょう。³⁸ 「自動化とAIによってますます形作られていくデジタル経済において、信頼性が高く追跡可能なエンティティデータは不可欠である」と、Global Legal Entity Identifier Foundation (GLEIF) のデータ品質管理およびデータサイエンス責任者であるZornitsa Manolova氏は述べています。³⁹

表2：AIエージェント主導のワークフロー

PROCESS	HUMAN WORKFLOW (TODAY)	AI-DRIVEN WORKFLOW (TOMORROW)
Payment across firms	Manual initiation, compliance checks, routing via automated clearing house/ real-time gross settlement, and reconciliation	Smart contract and unified ledger processing, AI compliance and routing, atomic settlement
Financial trade across firms	Order placement, affirmation, clearing, settlement, reconciliation	Automated pre-trade risk, smart routing, and instant post-trade workflows with distributed ledger technology

金融機関やその他のグローバル企業が共同で不正を迅速に特定できれば、それだけ早く不正を阻止し、損失を防ぐことができます。例えば、信用情報機関のエクスペリアン（Experian）は、2023年に1兆ドルを超えた「承認済み送金詐欺」を削減するために、世界の金融業界が取引データやオープンなエンティティ解決データを活用し、金融機関間で受取人の身元確認や「ミュール」口座の検出、行動上の異常の検知を迅速化できると推定しています。⁴⁰

Amazon Neptune Analytics を活用した AWS Entity Resolution のような商用実装例は、調査担当者が標準化され照合済みの顧客データや取引データをグラフ構造に変換し、関係性を追跡することで、「カード非提示（CPNP）」などの不正を示すパターンを、リレーショナル手法よりも効率的に特定できることを示しています。⁴¹ ポーランドに拠点を置く DataWalk SA のようなプラットフォームも同様に、統合されたナレッジグラフが、本来はサイロ化されていた社内外の構造化データおよび非構造化データを、エンティティの 360 度ビューに統合し、AML、不正のトリージ、および規制報告を促進する方法を示しています。⁴²

オープンドメインにおいては、Enterprise Data Management Council (EDMC) が、グローバルな法人参照データを用いてAMLナレッジグラフのプロトタイプを開発しました。これにより、データセット間のエンティティを相互に関連付けることで、AMLおよびKYCプロセスの有効性が向上し、手作業のコストが削減され、検出精度が向上しました。⁴³ 例えば、OpenData.Orgのデータを利用すれば、ユーザーはコンプライアンスのニーズに合わせてナレッジグラフをカスタマイズできます。また、イノベーターたちはトランスフォーマーを活用して、エンティティ解決の精度をさらに高めています。⁴⁴

資本市場

オープンなグローバルエンティティナレッジグラフは、銀行の運用コスト削減に寄与する可能性がある。例えば、McKinsey & Companyによると、銀行が法人識別子を導入すれば、取引処理コストを少なくとも10%削減できるとのことです。これは、投資銀行業界では年間1億5,000万ドル以上、貿易金融のための信用状を発行する銀行では5億ドルの節約につながります。⁴⁵

資本市場の参加者は、オープンな事業体グラフを活用して、上場企業および非上場企業のデータを構造化され相互に関連付けられた形で整理し、異常の検出、競合分析の強化、投資およびリスク管理戦略の洗練に役立てることができます。表形式のデータセットとは異なり、このようなグラフは企業、その財務状況、所有構造、および関係性をモデル化しているため、企業はより詳細な分析を行うことができます。

スウェーデンの研究者らによってMITライセンスの下で公開された、116万9000社以上の実在企業と5080万件の重み付けされた企業間関係からなる、オープンで大規模かつ異種混合のグラフである「CompanyKG」データセットを例に挙げてみましょう。⁴⁶ このデータセットのユーザーは、競合他社群の類似性やリスクをより容易に定量化し、投資ワークフローにおいてポートフォリオのスクリーニングを行うことができます。⁴⁷ もう一つのオープンリソースである FinReflectKGは、S&P 100構成企業の米国証券取引委員会 (SEC) 10-K報告書から抽出された財務ナレッジグラフを提供しており、エンティティ、関係、時間的事実を結びつける構造化されたトリプレットにより、リスクエクスポージャーの追跡や財務開示における異常パターンの検出が可能となっています。⁴⁸

より広範なオープンデータの分野において、Wikidataのナレッジグラフには、パブリックドメイン (CC0 1.0 Universal) の下で1億2,070万の実体とその属性が含まれており、これによりアナリストは、一貫性のある識別子や、業種コードや地理位置情報などの外部データセットへのリンクを用いて、ファームグラフィックを充実させることができます。⁴⁹ 静的なグラフ表現にとどまらず、Agentic GraphRAGのような実装では、構造化された実体間の関係と、非構造化された提出書類 (例: 10-K報告書) を組み合わせています。⁵⁰ オープンなエンティティデータセットと相まって、これらは資本市場アナリストが、S&P 100における規制リスク、サプライチェーンの依存関係、および取引相手先へのエクスポージャーに関する複雑な疑問に答え、独自のライセンス上の制約を受けることなく、投資ダッシュボードや異常検知モデルを強化するのに役立ちます。

エージェントとAIモデルの学習

「長年にわたり、多様なグループ間の連携を促進し、フェデレーテッド・クラウドおよびデータ管理システムの構築に取り組んできましたが、今や、AIシステムをステークホルダーの議論の場に持ち込む段階に至りました」と、Red Hat LLCの米国リサーチ・ディレクターであるHeidi Picher Dempsey氏は、Mass Open Cloudや大型ハドロン衝突型加速器 (LHC) といった共同プロジェクトから科学者や実務家を得た知見について論じた記事の中で述べています。⁵¹ 現代のAIシステムは、信頼性の高い推論を行い、グローバルに相互運用を行うために、高品質なデータに依存しています。これらのユースケースは、エンティティデータソース、抽出フレームワーク、エンティティ解決、ナレッジグラフ、プロバンス追跡、AIモデル、エージェントインフラストラクチャにまたがる階層型アーキテクチャが、分析および運用ワークフローにおいて、信頼性が高く、監査可能で、適応性のあるインテリジェンスをいかに促進するかを示しています。

標準的な参照データに基づく、信頼性の高いAIの基盤化。AIによる「ハルシネーション」を減らし、結果の事実的正確性を高めるためには、開発者はAIシステムを信頼できる知識に基づいて構築する必要があります。⁵² こうした基盤がなければ、LLMは回答を生成する際、2つの名称が同じ実体を指しているかどうかを推測することになりかねません。基盤が確立されていれば、モデルは権威ある実体レコードや関係性を取得することができます。例えば、銀行のAIコンプライアンスアシスタントが、回答を生成する前に、外国資産管理室 (OFAC) の制裁リストや金融犯罪取締ネットワーク (FinCEN) のデータベースなどの情報源から検証済みのレコードを取得する場合がこれに該当します。

企業が製品名、構成部品、ファームウェアのバージョン、およびドキュメントを、製品カタログやエンジニアリングナレッジベースなどの構造化されたリポジトリに保存しておけば、AIサービスエージェントをそれらの標準化された製品エンティティやドキュメントグラフに基づいて構築し、エージェントの精度、有用性、および顧客体験を向上させることができます。⁵³

検索拡張生成 (RAG)。 RAGは、構造化された外部知識とLLMを組み合わせ、より事実に基づき、文脈を認識した応答を提供する、よく研究されているもう一つのアーキテクチャパターンです。⁵⁴ 例えば、BrightQueryは、「ü genticハイブリッドRAGファクチュアルAIプラットフォーム」と呼ばれるものを開発しました。これは、構造化情報と非構造化情報を融合させ、AIモデルを適用してエンティティ間の関係を解釈し、異常を検出し、ビジネスや政府分野における意思決定に役立つ洞察を生成するものです。⁵⁵ このプラットフォームは、RAG技術とエージェントベースのワークフローを併用しており、そのアーキテクチャはモジュール式かつオープンソースであるため、業界を問わず拡張や適応が可能です。MCPサーバーを介してエンティティグラフにアクセスすることで、AIエージェントはエンティティをリアルタイムで特定し、誤情報を減らすことができます。

このプラットフォームは、多様なデータセットを取り込み、統一されたフレームワーク内で整合させることで機能します。資本市場では、財務報告書とリアルタイムのニュースや意見を融合させ、予測担当者、リスクアナリスト、コンプライアンス担当者を支援する。政府の分野では、経済データと世論を統合し、政策分析や国家安全保障上の意思決定に情報を提供します。このプラットフォームは、ユーザーにいくつかの明確なメリットをもたらします。それは、より深い文脈と裏付けのある洞察による高品質な意思決定、複雑なデータ環境における分析の高速化、そして体系的なバイアス低減による信頼性の向上です。このプラットフォームの柔軟な設計は、セクターを問わず一貫したパフォーマンスを維持しつつ、進化するデータニーズに対応します。

AIエージェント向けの、世界規模をカバーする標準化されたエンティティ。 世界中のサービスやワークフローにわたって推論を連携させるため、AIエージェントはエンティティの一貫した意味表現に依存しています。ナレッジグラフは、エンティティと関係を構造化されたセマンティックトリプル（知識表現の基本単位）としてエンコードするための形式的なモデルであり、これによりツールやサービス間でセマンティクスや推論を共有できるようになります。⁵⁶ ポストン・コンサルティング・グループによると、2030年までにAIエージェントは銀行の収益性を30%向上させ、コストを最大40%削減できる可能性があります。⁵⁷



エコシステムへの貢献によるデータの品質管理におけるネットワーク効果。クリーンで標準化されたエンティティデータは、AIの性能と自動化された意思決定の信頼性を向上させます。組織は、エンティティ解決やナレッジグラフを活用して、重複を排除し、現実世界の同一のエンティティに関するレコードを関連付け、関係性を構造化しています。エコシステムの参加者が新しいレコードを提供したり修正を加えたりすることで、データセットは時間をかけて充実していき、こうしたネットワーク効果がトレーニングデータや下流のAIアプリケーションを強化します。⁵⁸

ガバナンス、リスク、コンプライアンス（GRC）の統制は、追跡可能かつ監査可能でなければならない。信頼性を確保するためには、AIモデルはあらゆるデータ変換を追跡可能にし、AIエージェントはベンチマークに対して人間と同等かそれ以上のパフォーマンスを発揮しなければなりません。⁵⁹ データプロヴェナンスおよびリネージシステムは、データの起源と変更履歴を記録するため、エンティティデータがソースから解決および推論パイプラインを経て流れる過程で、ステークホルダーは監査、コンプライアンスの検証、およびインサイトの再現を行うことが可能になります。⁶⁰

権威あるエンティティ解決による効率的な情報検索。ナレッジグラフは、エンティティの構造化された表現とその意味的な関係を統合したものです。開発者が、このように解決済みで標準化されたエンティティIDをこれらのグラフに供給すると、グラフは文脈に応じた正確な検索や関係性に基づく高度な推論を行うことができ、それによって検索の効率とインサイトの抽出が向上します。⁶¹

「エージェントを知る（Know Your Agent）」などの適応型および新興のユースケース。高度なマルチエージェントインフラストラクチャは、正規化されたエンティティ表現を、エージェントIDフレームワーク、ゼロトラスト型のエージェントIDおよびアクセス管理アーキテクチャ、ならびにLOKA（Layered Orchestration for Knowledgeful Agents）、Agent2Agent、MCPなどの相互運用性プロトコルと統合する。これらの技術とナレッジグラフのメモリ層により、相互運用可能なAIエージェントは、アイデンティティの認証、ツールの発見、アクションの調整、構造化されたエンティティに関する推論を行い、「エージェントを知る（KYA）」などの新たなユースケースに向けた適応型ワークフローやガバナンスモデルを強化することができます。

「AIエージェントを開発する人の多くは、必ずしもそれを『エージェント』として認識されたくないと考えています。他のLLM開発者の中には、さまざまなサイトからデータを収集し、トラフィックの発生元を意図的に隠蔽している人もいます」と、GoogleのData Commonsに所属するRamaswami氏は述べました。同氏は、多数のウェブページを保有するサイト運営者が、ある特定の月に突如として莫大な請求額を突きつけられる理由について、「特定の国や特定のLLM開発者からの多数のIPアドレスによるアクセスが集中しているためです」と説明しました。「それらのエージェントの開発者が、人間を装うように設計していたとしても、私は驚くことはありません」⁶²

企業信用

企業信用の分野では、AWS LabsのSageMaker信用スコアリングツール（Apache 2.0ライセンス）のようなグラフベースの手法により、SECの10-K/Q報告書から構築された企業ネットワークグラフを通じてエンティティを結びつけ、財務比率と組み合わせることで、信用格付け予測モデルをどのように強化できるかが示されています。⁶³ OpenData. Orgのようなオープンなエンティティデータセットは、設立詳細や取締役情報など基礎的な属性を提供しており、ユーザーはこれらを信用モニタリングのワークフローに連携させることで、関連企業全体における経営悪化や不正の初期兆候を検知することができます。⁶⁴

ProsightのメンバーであるIsio Nelson氏によると、主に銀行、信用組合、および金融サービスの専門家からなる同組織のメンバーは、「業務をより円滑に進め、既存の取引関係を拡大し、既存のポートフォリオのリスクをより深く理解し、リソースを適切な場所に配分できるよう、自らがサービスを提供している、あるいは提供したいと考えている企業に関するより質の高い情報を求めている」とのことです。⁶⁵ オープンなナレッジグラフデータセット、オープングラフデータベース、およびオープンAI技術を組み合わせることで、これらの実装例は、オープンな事業体グラフが、公開市場およびプライベートクレジット環境の両方において、資金調達評価、リスク管理、およびリアルタイムの信用リスクモニタリングをいかに強化できるかを示しています。⁶⁶

営業、マーケティング、および顧客関係管理

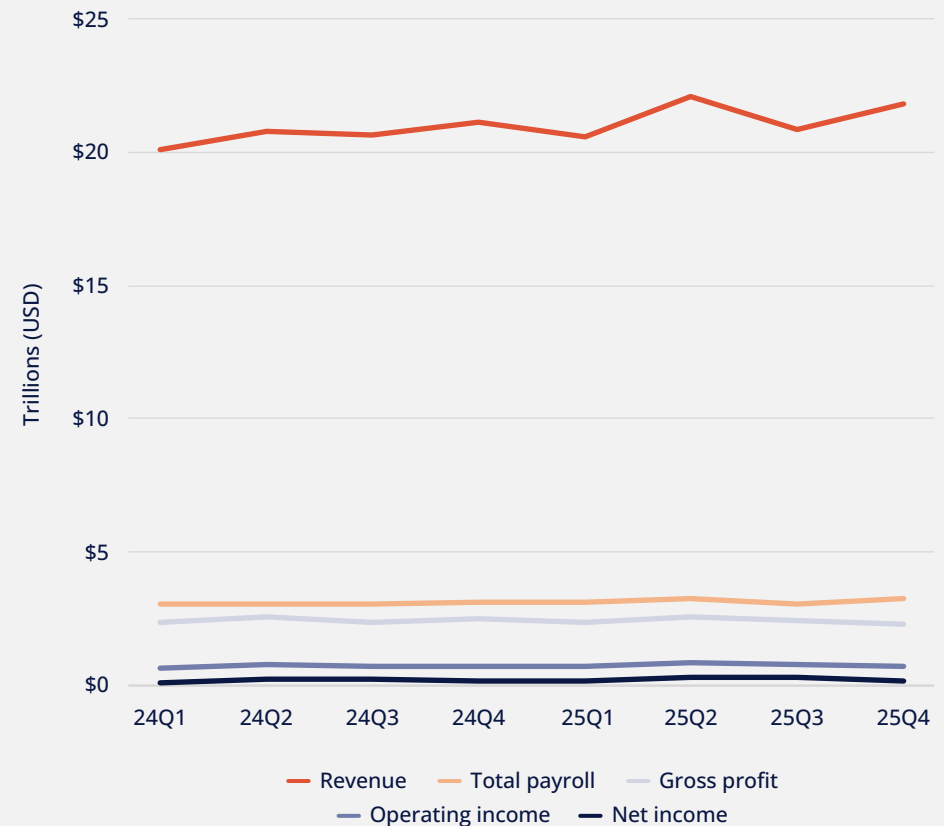
研究者たちは、金融機関が顧客生涯価値（CLV）を算出する際、単一の金融機関のデータに依存することが多いことを明らかにしました。もしオープンバンキングのデータを活用し、顧客の複数金融機関にわたる行動を分析すれば、顧客一人あたりの収益性を高め、長期的な関係を育み、最終的にはCLVを20%以上向上させるような、顧客中心の戦略を策定できることがわかりました。⁶⁷

ProsightのNelson氏によると、このプロジェクトがエンティティに紐づけているその他のあらゆるデータに加え、データの系譜や利用権を考慮すれば、「組織が顧客レコードの一部を遡及的に補完し、既存の顧客をより深く理解するのに役立つ可能性がある」とのことです。例えば、「現在の顧客のうち、どれだけの人が小規模事業を営んでいるか、あるいは商業法人の代表者であるか。個人事業主なのか、従業員を抱える企業なのか。キャッシュフローはどのような状況か」。こうしたデータは、商機の優先順位付けに役立つとNelson氏は述べました。「これらの顧客のうち、成長見込みが最も高い上位10社はこれだ。あるいは、業績は振るわないものの、最も潜在力のあるアカウントはこれだ」⁶⁸

オープンデータソースからのエンティティ解決、関係性、リスク指標を統合されたグラフに組み合わせることで、営業およびマーケティングチームは顧客プロファイルを充実させ、見込み客にアプローチする前に、絶好のビジネスチャンスやリスクを明らかにすることができます。例えば、組織はOpenData.Orgのデータセットと、Wikidataナレッジグラフ（共同でキュレーションされたエンティティのグローバルグラフ、パブリックドメインのCC0ライセンス）を組み合わせることができます。これを企業、個人、および属性に関する初期データセットとして活用し、SPARQLを介してクエリを実行することで、見込み客レコードを充実させ、セグメント化することができます。⁶⁹ 同様に、OpenSanctionsからの制裁やリスク関連データを追加し、高リスクまたはコンプライアンス上注意が必要なターゲットをフィルタリングすることも可能です。⁷⁰

図7：正社員を雇用する企業の成長

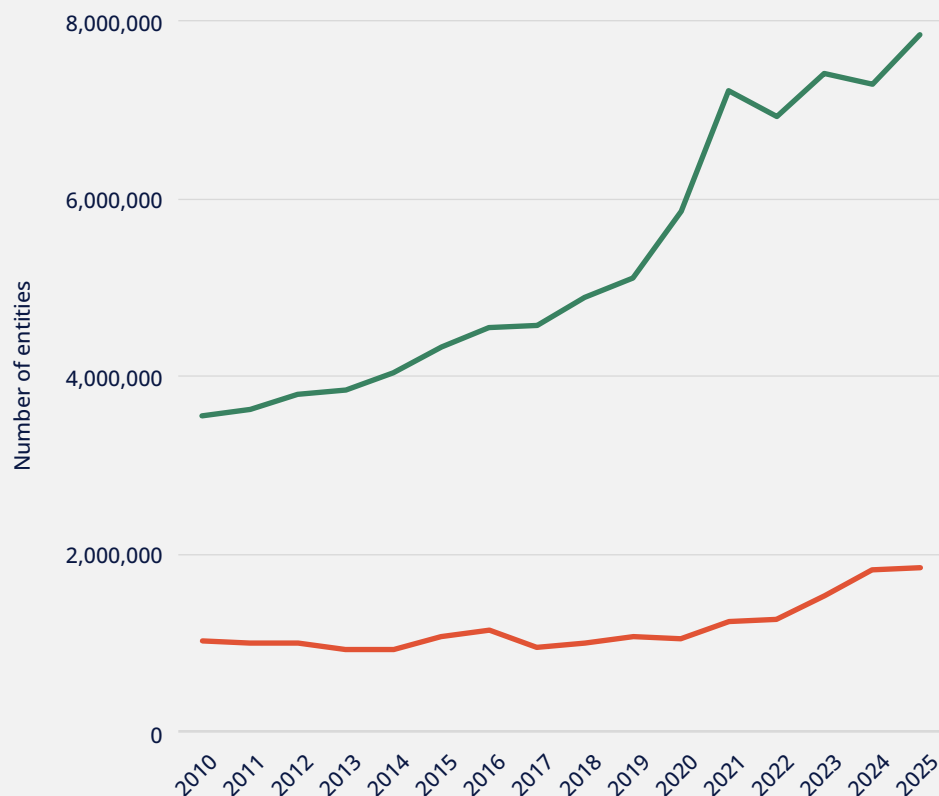
米国の約800万社の中小企業を対象とした分析によると、売上高（6.43%）、給与総額（6.24%）、および雇用者数（0.64%）の増加が見られました。



Source of data: BrightQuery Inc. and ProSight Financial Association, "Companies with full-time employment," *BrightQuery-ProSight Economic Analysis Report*, as of 1 Jan. 2026, <https://docs.brightquery.com/bq-prosight-economic-analysis-report>.

図8：米国における法人の設立と廃業の推移

営業担当者は、担当エリア内の郵便番号や建物の住所ごとに、新規事業設立の状況を分析することができます。緑の線は新規事業設立を示し、赤の線は廃業を示しています。



Source of data: BrightQuery Inc. and ProSight Financial Association, "Annual legal entity new formations and dissolutions," *BrightQuery-ProSight Economic Analysis Report*, as of 1 Jan. 2026, <https://docs.brightquery.com/bq-prosight-economic-analysis-report>.

「プラットフォームのユーザーは、リスクが低い、あるいはリスクのない見込み客に向けたアプローチキャンペーンをカスタマイズすることができます。」

TerminusDBや JanusGraph などのオーブングラフデータベースは、ユーザーがこうした統合されたエンティティグラフを保存・検索するための実用的な手段を提供し、企業ネットワーク内の二次的な連絡先、共通の所属先、あるいは共有されるリスク要因を発見できるようにします。⁷¹ 例えば、これらの技術に基づいて構築されたマーケティング分析プラットフォームでは、WikiDataの識別子、OpenSanctionsのリスクフラグ、BrightQueryの一意のエンティティID、および企業のCRMシステムからの顧客IDを統合することで、見込み客に関するリスクを考慮した360度の全体像を把握することができます。プラットフォームのユーザーは、リスクがゼロまたは低い見込み客に対して、アプローチキャンペーンをパーソナライズすることが可能になります。⁷²

こうしたオープンなナレッジグラフアーキテクチャは、例えば自然言語クエリ層や埋め込みモデルなどを通じて、フェデレーテッドAIワークフローも支えています。これにより、営業チームは「業界Xに関連し、かついかなる制裁リストにも掲載されていないすべての連絡先を表示してください」といった高レベルの質問を行い、キャンペーンのセグメンテーションやコンプライアンスに関する洞察を得ることができます。⁷³

サプライチェーンとESG報告

環境・社会・ガバナンス（ESG）に関するあらゆる取り組みにおいて、最も欠けている要素の一つが「場所」です。「企業、そのサプライヤー、販売業者、従業員、そして原材料の産地といった要素間の関係性を示すナレッジグラフを構築しようとしているのです」。その目的は、その結果として形成されるサプライチェーンが環境にどのような影響を与えるかを理解することにあります。⁷⁴ このオープンデータプロジェクトは、そうしたあらゆる情報を一元化する一つの手段です。「これは実に強力な取り組みです」と、Overture Maps Foundationの最高技術責任者（CTO）であるAmy Rose博士は述べました。「これは2022年に立ち上げられた共同のオープンデータイニシアチブであり、Overture Maps Foundationの数十社の加盟企業から集まったソフトウェア開発者、データ専門家、地図作成エンジニア、プロダクトマネージャーらが主導しています。」⁷⁵ 「サプライチェーンに関わる事柄は、どれも非常に重要な課題です」と彼女は語りました。⁷⁶

「ESG要因にわたる企業の業績を分析するには、投資家はまず、対象となる事業体を明確に特定する必要があります」と、ブルームバーグLPのオープンデータおよび標準化担当チーフストラテジスト、リチャード・ロビンソン氏は述べています。⁷⁷ そのためには、単一の法人を追跡するために使用されるさまざまな識別子を照合し、マッピングする必要があるかもしれません。そして、LEIやFIGIといったオープンデータ標準や、OpenData.Orgのオープンエンティティグラフは、まさにその作業を支援するものです。そうでなければ、特定の法人のESG実績を長期的に分析することは、「投資家や規制当局にとって極めて困難なものとなりかねません」。⁷⁸

エンタープライズ インテリジェンス

公開されているウェブ上のシグナルやオルタナティブデータから企業情報を導き出すことも、もう一つのテーマでした。⁷⁹ アナリストは、Common Crawlなどの公開されているウェブデータのパターンを分析することで、合併、新製品の発売、企業方針の転換といった戦略的な動向を推測することができます。⁸⁰ 例えば、ウェブサイトのコンテンツの変更、求人情報、特許出願、あるいは企業発表などの変化は、買収や新たな取り組みの初期の兆候を示すことがあります。⁸¹ 自動車業界や金融セクターの分析から医薬品の価格設定に至るまで、さまざまな業界でこうしたアプローチが試みられています。ウェブシグナルは多くの場合、従来の構造化されたビジネスデータセットよりも早く、かつ詳細に現れるため、この種のウェブ由来のインテリジェンスは、従来のデータセットを補完し、時にはそれを凌駕することもあります。

小売店舗および倉庫の所在地

実店舗やフランチャイズ店にとって、立地はすべてを左右します。最適な場所を特定するには、営業時間中のターゲット層の人口密度や世帯所得の中央値、徒歩や車でのアクセスしやすさを示すスコア、そして近隣にある補完的な事業や競合他社のデータが必要となります。

「例えば、私がピザのフランチャイズ店を開業するとしましょう。どこに出店すべきか検討しているところです。そのためには、『他のピザ店はどこにあるのか』や『ターゲットとする地域の人口統計はどうか』といった疑問への答えが必要です。人口統計データは米国国勢調査から無料で入手できます。しかし、他のピザ店の所在地はどこで入手すればよいのでしょうか？」と、Data CommonsのRamaswami氏は尋ねました。⁸² その情報がないと、起業家は既存のピザ店の近くに新しいピザ店を開店してしまう可能性が高いでしょう。「だからこそ、ガソリンスタンドが4軒も隣り合わせに建ってしまうのです」と彼は述べました。⁸³

ウェブ標準の先駆者であるDr. Ramanathan V. Guha氏は、同様の課題について次のように説明しました。非営利団体「Feeding America」は、スーパーマーケットから余剰在庫を回収し、そのネットワークを構成するフードパントリーに再分配しています。「同団体は、国内のフードパントリーの約半数のバックエンド運営を担っています」とGuha氏は述べました。⁸⁴ 倉庫の設置場所を決定するためには、全国的なネットワーク全域にわたる交通システムや、貧困率および人口動向に関するデータが必要でした。

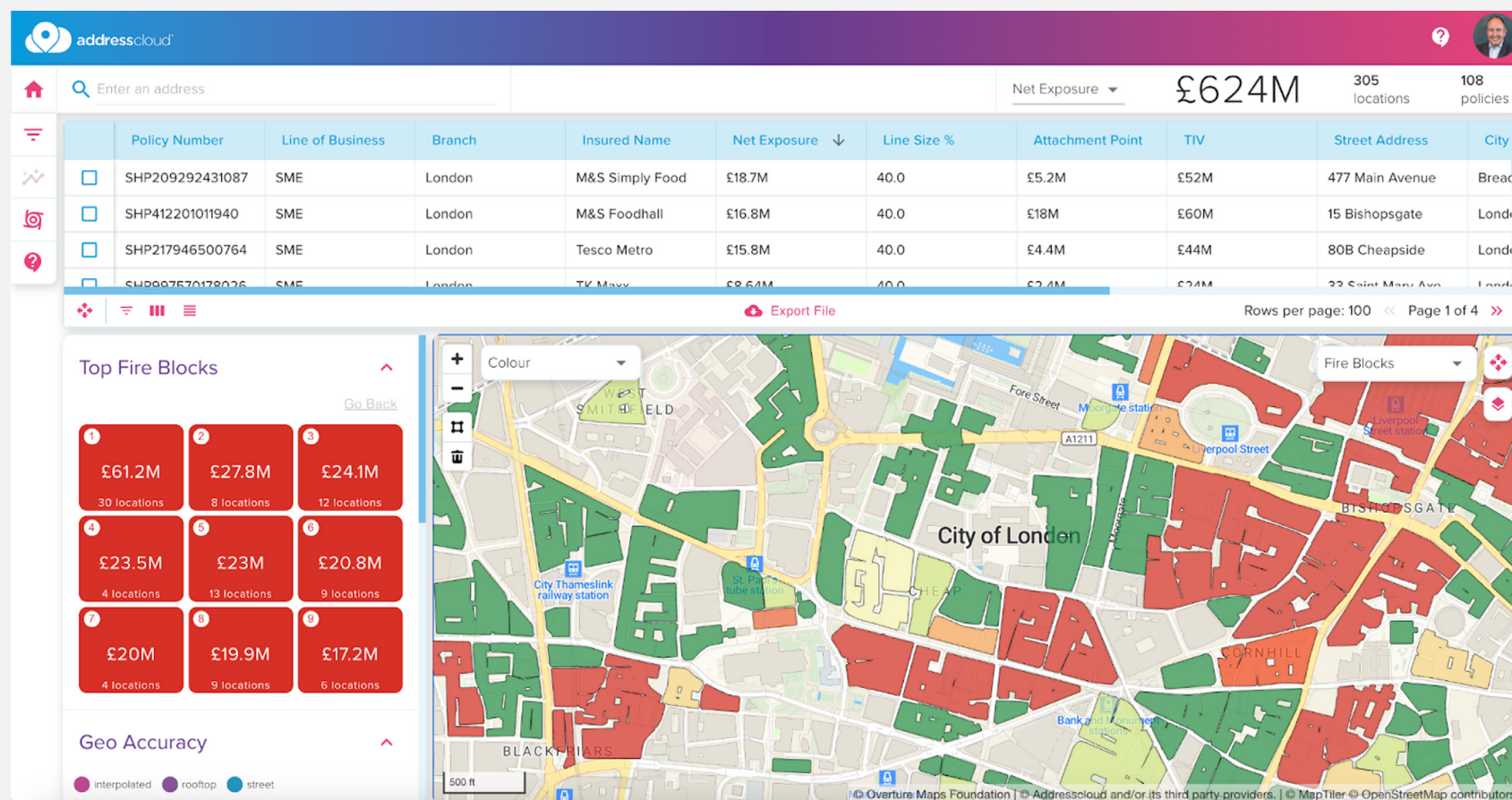
彼らは、OpenData.Orgのようなオープンデータセットを活用することで、より優れた意思決定ができると見えています。「オープンデータが増えれば、より効率的なシステムを構築できるようになり、そうした効率的なシステムは、長い目で見れば私たち全員の役に立つでしょう」とRamaswami氏は述べました。⁸⁵

災害リスク管理と経済開発

Overture Maps社のRose博士は、地理空間データと経済開発データを組み合わせることで、サービスが過剰な地域や不足している地域を特定するのに役立つと説明しました。⁸⁶ 政府は、データとAIを、災害リスク管理や対応計画の策定・実施だけでなく、状況の変化に応じてそれらを的確に調整するためにも活用できるでしょう。⁸⁷

自然災害や戦争の後、地域社会は単に復興の方法だけでなく、産業と立地間の関係をどのように活用してリスクを軽減し、公共資源をより効果的に配分し、より効果的な官民パートナーシップを構築し、経済的格差を解消できるかを検討することができます。Addresscloudの事例については、図9をご覧ください。

図9：Overture Mapsのデータに基づいて火災発生傾向を表示したAddresscloud Riskプラットフォーム



Source: "How Addresscloud Scaled Geospatial Risk Analysis with GERS," Case Study, Overture Maps Foundation, 10 Sep. 2024, <https://overturemaps.org/case-study/2024/how-addresscloud-scaled-geospatial-risk-analysis-with-gers/>. Used with Addresscloud's permission.



「私たちが非常に価値あると考えているのは、国境を越えた問題解決です」と、United Nations International Computing Centre (UNICC) のAIハブ責任者であるAnusha Dandapani氏は述べました。「オープンデータセットを共有することで、相互運用性が非常に容易になり、まず第一にガバナンスが強化されます。しかし、単にオープンデータやナレッジグラフがさらに増えればよいというわけではありません。私たちに必要なのは、オープンデータとオープンAIモデルなのです」。同氏は次のように説明しました。「オープンデータとオープンモデルを組み合わせれば、特にグローバル・サウスにおける国境を越えた能力構築において相乗効果が生まれるでしょう。ただし、政府や公共部門の文脈でこれらを一体として捉え、オープンデータが比例的に普及し、かつ主権データセットの利用・再利用に関する明確な政策が整備されていることが条件となります」⁸⁸

このプロジェクトに参加した理由

個人や組織が、Linux Foundationにとって新たなオープンデータプロジェクトを利用したり、それに貢献したりする理由はいくつかあります。「資本主義において、情報は自由であり、いかなる摩擦もなく流通するものであるべきです」と、Data CommonsのRamaswami氏は述べました。「情報は、最も自由な市場を生み出すものであるべきです。しかし、実際にはそうではないことは周知の事実です。」

Ramaswami氏は、組織の規模や大きさ、支払い能力に伴う利点と、中小企業、スタートアップ、起業家が直面するコスト面の障壁について指摘しました。「資本主義の観点から言えば、このオープンデータ・プロジェクトはある程度、競争の公平性を確保するものです」と彼は述べました。「また、これまでは繁栄できなかったであろう企業にも、新たな機会をもたらすことになります。」⁸⁹

組織能力の強化

「金融サービス業界では——他の業界でも状況は非常によく似ていると思いますが——人々は、その取り組みが変革をもたらすものであったり、戦略的であったり、あるいはコスト削減につながったりするからこそ、それに参加したいと考えるのです」と、FINOSの最高執行責任者（COO）であるJane Gavronsky氏は述べました。FINOSは、金融サービス業界の会員やユーザーの間でオープンイノベーションを推進しており、50のプロジェクトやワーキンググループを展開しています。⁹⁰

大規模で高品質なデータセットや堅牢なデータインフラを維持することは、中小企業にとって大きな負担となり得ます。例えば、OpenData.Orgプロジェクトに参加することで、地域銀行などの**意欲的な中堅・中小企業**は、データ活用の取り組みを推進し、高品質なデータセットにアクセスし、協働の力を活用し、規制当局と連携することが可能になります。特に、公共部門がオープンデータやオープンガバメントの取り組みを非常に好意的に捉えている管轄区域においては、その効果が顕著です。また、このプロジェクトが世界の規制当局と連携できる範囲において、それらの規制当局もこの種のデータセットから恩恵を受けることになるでしょう。

財務面でのプラスの影響

このようなプロジェクトに貢献するためには、組織は収益やコスト削減の大きな可能性を見出さなければなりません。「データは一般的に、その両方に非常に大きな効果をもたらしますが、メッセージでは、それが運用上のものか、変革的なものか、あるいはその両方であるかを明確にしなければなりません」とGavronsky氏は述べました。「このプロジェクトのターゲット層は、オープンなコラボレーションが有益であることを理解し、その製品が自分たちにとって有用だと感じ、実際にそれを利用し、貢献し、他の人々にも広めてくれる人々です。それがFINOSの役割です。私たちはコミュニティの構築を支援しています。」⁹¹

アクセスの拡大

「非営利団体の分野では、一般的にエンタープライズグレードの実体解決ソフトウェアを利用できる人はおらず、技術インフラを運用するための十分な予算を持っている組織もほとんどありません」と、「政治的、犯罪的、あるいは経済的に注目される個人や企業の国際データベース」であるOpenSanctionsプロジェクトの創設者、Friedrich Lindenberg氏は述べています。⁹² 同氏は、OpenData.Orgプロジェクトを、学術界や市民社会分野のためのリソースと捉えています。これらの分野では、収集した有用なデータを活用してさらに多くのことを成し遂げたいと考えているものの、そのためのリソースが乏しいのが現状です。「データの世界の多くは極めて封建的であり、データによる『領地』が存在し、それぞれが独自の命名規則を持っています。今では、CRM（顧客関係管理システム）をお持ちの方なら誰でも、OpenData.Orgを利用してデータを充実させることができます。予算の限られた企業、スタートアップ、事務所も、このオープンデータセットを大いに活用できるでしょう」と彼は述べました。

さらなる経済成長

研究によると、オープンデータや空間データは、大きな経済的価値をもたらすものの、その価値は十分に活用されていないことが示されています。コンサルティング会社Lateral Economicsの推計によると、G20諸国はオープンデータから年間約2.6兆ドルの利益を得られる可能性があるとして示されています。一方、テクノロジー企業のSpatineoの調査では、フィンランドは130億ユーロに上る潜在的な利益のうち、25%未満しか獲得できていないことが明らかになりました。また、国家空間データシステムの活用を拡大することで、実質的な直接的・間接的な便益がさらに得られることが示されています。⁹³

より最近の調査において、McKinsey Global Instituteは、オープンな金融データ・エコシステムを広く導入することで、欧州連合（EU）、英国、米国などの高所得国グループは2030年までにGDPの最大1.5%に相当する経済成長が見込め、インドのような中低所得国では4~5%の成長が見込めると推計しています。⁹⁴

重点的なイノベーションと競争優位性の確立

Bloomberg LPのRobinson氏によると、「オープンデータにより、企業やプロバイダーは付加価値の創出や、その分野での競争に注力できるようになります」とのことです。同氏は*data.world*に対し、「たとえ『標準』であっても、オープンでないものが義務付けられると、ロックインが生じ、イノベーションが阻害されます」と語りました。また、「オープンデータの基盤」は、「企業に事業規模ではなく、サービスの質や幅広さにおいて競争することを迫ることで、競争を促進する」と述べました。⁹⁵

これは、人工知能（AI）分野での競争において極めて重要となります。AIにおける差別化の鍵はデータにあります。最先端のAIモデルのコモディティ化が進む中、イノベーションはAIスタックの上位層へとシフトしつつあります。このような、差別化要因とならず、したがって競争上の優位性をもたらさないデータセットについては、各企業がコストを分担することで、その分をモデルのトレーニングやAIエージェントの開発、その他のソリューションの構築に投資できるようになります。そこそが、真の競争上の差別化が生まれる場となるのです。

一例として、CODEのGurin氏は、自動運転車（AV）の開発に必要なデータに関して、米国運輸省が主要自動車メーカーと進めている取り組みを挙げました。⁹⁶ CODEはこれらの協議を促進し、その結果、自動運転車の安全性を目的とした

「Work Zone Data」イニシアチブが誕生しました。また、CODEは、「自動運転車の統合に向けたデータへのアクセスを拡大し、実行可能な優先事項と実施における明確な役割を導き出す」ことを目的とした、より広範な「Data for AV Integration」イニシアチブも支援しました。⁹⁷ これは「協競（coopetition）」のモデルであり、企業は「自社の事業にとって不可欠な政府の安全イニシアチブを実現するために、データを提供し、基準を策定する」べく協力しました。⁹⁸

資源を統合することによる可能性

「データの収集には費用がかかります。そのデータを活用可能にするのにも費用がかかります」と、RSS、RDF、Schema.orgといった広く利用されているウェブ標準の考案者であるR. V. Guha博士は述べました。「人々がデータセットを利用するためには、容易に活用できるデータが一定量以上必要です。そして、データセットが一定量に達するように人々がデータを寄贈するためには、その運営主体が中立であり、今後も中立であり続けると信じられる必要があります」と、同博士は語りました。⁹⁹

Gavrinsky氏はその考えに付け加えました。「はい、どのようなプロジェクトであれ、通常は独自にプロジェクトを進めるような大規模な機関が必要です」と彼女は述べました。「そうした機関は、必要に迫られてプロジェクトに投資することになりますが、それは彼らの競争上の優位性にはなりません。」もしそれらの投資がオープンで、誰もが利用できるようになれば、市場にとって価値あるものとなり、間接的に彼らも恩恵を受けることになります。彼女は次のように説明しました。「私たち全員がより質の高いKYCデータを持てれば、誰もが恩恵を受けます。社内でデータを整理・活用するためのコストを回避でき、より質の高いデータをより迅速に入手できるのです。」¹⁰⁰

ご支援の方法

新しいプロジェクトコミュニティでは、通常、データ品質、データセットの活用、新たなユースケースといった主要分野に焦点を当てます。例えば、初期のワーキンググループでは、以下のようなプロセスの概要を策定することができます：

このプロジェクトへのオープンデータの提供について。 Plehn博士は、協働の力と「群衆の知恵」を信じています。「データの世界に100%の正確さなどありません。「80%の精度に達すれば、私は大変嬉しく思います」と彼は述べました。「しかし、世の中のあらゆる企業の集成的な知識と知恵を結集できれば、力を合わせることで、100%にかなり近い精度に到達できると思います。」¹⁰¹ BrightQuery チームは、誰もがウェブサイト、LinkedInアカウント、電話番号、住所などに関する情報を更新したり、誤りを修正したりできるオープンなコミュニティの実現を目指しています。

透明性が高く、そのまま置き換え可能な機能を備えた**マッピングおよびテストツールを開発**し、貢献者が既存のデータ標準やIDにマッピングできるようにすることで、既存のデータセットをこのオープンデータセットと置き換える際の負担を軽減します。

「もし彼らの投資がオープンで、誰もが利用できるものであれば、それは市場にとって価値あるものとなり、彼らは間接的にその恩恵を受けることになるでしょう。」

エンティティグラフの拡張。 オープンデータのエコシステムでは、貢献者は自身の業務にとって最も意義のあるエンティティタイプ（例：出版物、有価証券など）を追加することができます。例えば、学術界や文学界には、著者、著作権、出版物があります。製薬業界には、特許、製品ライン、ロット、成分があります。建設会社やエネルギー企業、海運・航空事業、さらには国連開発計画（UNDP）のような国際的な経済開発機関でさえ、その膨大な有形資産を特定し、追跡するための手法を持っています。

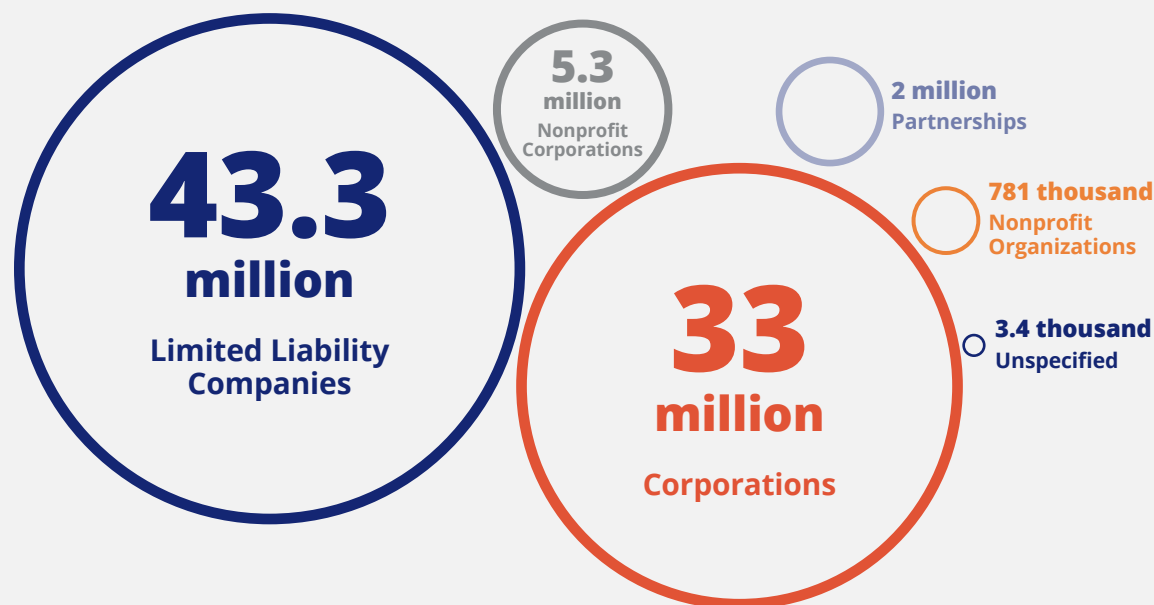
証券データ。 OpenData.Org ではすでに、エンティティレベルの情報を、構造化された規制書類や企業提出書類に記録された証券・資本情報と統合しており、これによりユーザーはリアルタイムの証券データを紐付けることができます。証券データは、最も当然の追加項目となるでしょう。Massachusetts Institute of Technology のデジタル経済イニシアチブに所属する研究員である Frank Nagle 博士は、米国証券取引委員会（SEC）に提出され、同委員会を通じて公開されている年次報告書（Form 10-K）および四半期（フォーム10-Q）報告書は、米国証券取引委員会（SEC）に提出され、同委員会を通じて公開されているため、有用であると考えました。¹⁰² Nagle博士は、これらのエンティティ間の関係——財務関係、所有関係、顧客・サプライヤー関係——に加え、取締役や役職レベルでのつながりも把握したいと考えていました。¹⁰³

サステナビリティに関するデータ。また、特に欧州連合（EU）の企業サステナビリティ報告指令（CSRD）や、オーストラリア、カナダ、中国、インド、インドネシア、日本、韓国、シンガポール、南アフリカ、スイス、英国、ベトナムにおけるその他のサステナビリティおよび気候変動に焦点を当てた報告イニシアチブに関する、統合されたサステナビリティ情報も欠けています。¹⁰⁴

非営利団体のデータ。GoogleのData Commonsを率いるRamaswami氏は、非営利団体のデータ市場における非効率性を指摘し、慈善団体や非政府組織に関する情報を含めることを提案しました。¹⁰⁵ 本稿執筆時点で、OpenData.Orgは600万を超える非営利団体を網羅していました（図10）。

詳細については、データセットのダウンロードは <https://opendata.org/download> をご覧ください。オープンデータ・コンソーシアムへの参加についてご相談の際は、partners@opendata.org までメールでお問い合わせください。本論文の次のセクションでは、オープンデータ・プロジェクトを立ち上げるプロセスについて解説します。

図10：OpenData.Orgにおける各種組織のカバー範囲



Source of data: "Organization counts," BrightQuery Documentation, as of 1 May 2026. https://docs.brightquery.com/org_summary.



新しいデータプロジェクトの立ち上げ

オープンデータの価値を認識する

「なぜLinuxカーネルはソフトウェアとしてこれほど重要なのでしょうか？」と、Linux Foundationの法務・戦略プログラム担当上級副社長であるMichael Dolan氏は尋ねました。「それは、世界のほとんどのソフトウェアがサポートしている、実績のあるオペレーティングシステム上で、誰でもいつでも即座に何かを構築できるよう、オープンに提供されているからです。」¹⁰⁶ 例えば、Googleは自社のAndroidオペレーティングシステム（OS）を、アップストリームのLinux長期サポートカーネルをベースに開発しました。また、NASAは、パーセベランス探査機と共に火星へ送られた小型ヘリコプター「Ingenuity」に、Linux OSを搭載しました。¹⁰⁷

「オープンライセンスの下でテクノロジースタックの基盤となるレイヤーを利用でき、即座に何かを作り始められることは、大きな相乗効果をもたらします」とDolan氏は述べました。「マクロ的な観点から見ると、オープンソースを通じて技術を公開的に共有することは、既存ソフトウェアの更新と新機能の革新の両方を促進します。この原則はデータにも当てはまり、データが公開的に共有・改善されることで、継続的な更新や新たな活用事例が可能になります。」¹⁰⁸

図11：オープンソースのリファレンスアーキテクチャ：データからインサイトへ

組織は、データをスタックの基盤層として活用し、さまざまなユースケースに対応したエンティティ解決・ナレッジグラフシステムを構築する際に、オープンソースソリューションを活用することができます。

ANALYTICS	Apache Superset, Linkurious, Metabase, etc.
Users gain insights from tools that analyze relationships and patterns to identify trends and anomalies.	
APPLICATIONS	Cytoscape, Gephi, Jupyter, Streamlit, etc.
Tools use entity data to power workflows such as risk scoring, compliance, and customer intelligence.	
SERVING / APIS	GraphQL (Ariadne, Graphene, Hasura), REST, etc.
Application programming interface services expose resolved identities through queryable interfaces.	
STORAGE / QUERY	JanusGraph, TerminusDB, ArangoDB, etc.
Users gain insights from tools that analyze relationships and patterns to identify trends and anomalies.	
ENTITY GRAPH	Python, Apache Spark, etc.
Users gain insights from tools that analyze relationships and patterns to identify trends and anomalies.	
ENTITY RESOLUTION	Splink, Dedupe, Febrl, etc.
Users gain insights from tools that analyze relationships and patterns to identify trends and anomalies.	
NLP EXTRACTION	spaCy, Hugging Face transformers, Stanford CoreNLP, etc.
Users gain insights from tools that analyze relationships and patterns to identify trends and anomalies.	
INGESTION / ORCHESTRATION	Apache Airflow, Apache NiFi, Dagster, Prefect, etc.
Users gain insights from tools that analyze relationships and patterns to identify trends and anomalies.	
DATA SOURCES	OpenCorporates, OpenData, OpenSanctions, etc.
Users gain insights from tools that analyze relationships and patterns to identify trends and anomalies.	

オープンデータセットは、オープンな人工知能（AI）モデルやエージェント型AIと同様に、このスタックのもう一つの基盤層を構成しており、前者の品質は後者の品質向上にとって極めて重要です。（前ページの図11を参照してください。）これらはすべて、世界でも最も熾烈な競争を繰り広げる企業や、厳しい要求を突きつけるイノベーター——企業、非営利団体、政府機関、学術機関——が、それぞれのステークホルダーにより良いサービスを提供するために生み出した製品やサービスの、競争前の段階にある構成要素です。こうした組織の多くは、これらの基盤層を維持・強化するために協力し合っています。¹⁰⁹

多くの商用ユースケースにおいて、既存のオープンデータセットだけでは不十分です。「科学研究から得られるオープンデータは数多く存在しますが、それらは一般的に、商用や企業環境において必要とされるデータではありません」とDolan氏は述べています。その結果、イノベーターは、必要なデータを誰が保有しているかを特定し、適用される利用規約を理解した上で、それらの規約が他のデータセットとの統合、再配布、および意図したビジネスモデル内での展開を許可しているかどうかを判断しなければなりません。「一般的に、データ提供者はアクセスや利用を制限することが多いため、こうした回答はビジネスのニーズを満たしません」と同氏は述べました。「歴史的に見て、データそのものは通常、排他的な所有権の対象とはならないことを認める長年の法的枠組みが存在するにもかかわらず、データは厳格に管理され続けてきました。」¹¹⁰

そのギャップを埋めるため、Dolan氏は「Overture Maps Foundation,」の取り組みを挙げました。同Foundationのメンバーは、開発者が利用できる「信頼性が高く、使いやすく、相互運用性のあるオープンな地図データ」を作成することで、「現在および次世代の地図ソリューションを支えている」とのことです。¹¹¹

「成功するプロジェクトには、運営委員会や支援者からの支持が得られているため、発表すれば地域の人々が集まってくるのです」と、MITのFrank Nagle博士は述べました。「それは単なる一つのプロジェクトではなく、成長する準備が整っており、長期的にプロジェクトを育てていくことに注力しているコンソーシアムなのです。」¹¹²

AFINOSおよびOverture Mapsプロジェクトのメンバーとして、BrightQueryチームは、Linux Foundationが技術コミュニティの運営・育成において卓越した手腕を発揮し、高い運用基準に基づいてオープンプロジェクトを管理するためのプラットフォームを構築していることを、身をもって体験しました。

このチームは、Linuxがもたらす以下の点を重視しています。すなわち、プロジェクトを中立的な立場で管理し、公平なマネージャーがプロジェクトの推進を支援すること、導入と成長を加速させるための商用連携、貢献者やユーザーの知的財産権に関する明確な規定、そしてプロジェクトコミュニティにおいて実行者が意思決定者となるガバナンスモデルです（図12）。¹¹³ また、そのDevOps環境やインフラは最先端であり、セキュリティや保守ツールも同様に最先端のものです。プロジェクトの運営担当者は、プロジェクトの健全性を測定するための指標やダッシュボード、およびコンプライアンスの監視と維持のためのガイドを備えています。

図12：新規プロジェクトの立ち上げに向けた実績のあるシステム



Data Foundationの社長兼CEOであるNick Hart氏は、「初期段階で正確なユーザー層の予測を行う」ことや、ターゲットユーザー層とその主要なユースケースに合わせてシステムを設計するだけでなく、「そのユーザー層に対する実際の価値提案の方向性を明確に把握しておく」ことも重要だと指摘しました。¹¹⁴

インセンティブモデルの確立

オープンデータ エンタープライズ センター (CODE) のJoel Gurin氏は、自身の経験に基づき次のように述べています。「オープンデータの推進者にとって、大規模なデータ共有に真に有効なモデルを見出すことは困難な場合があります。こうした取り組みは、主に企業自身の利益のために必要とされる場面で実現されるものです。」¹¹⁵ したがって、最優先事項は、このデータセットに適したインセンティブモデルを確立し、対象となるエンティティの幅と深さ、および各エンティティに関するデータの完全性と品質の両面で、データセットが継続的に進化するようにすることです。Linux Foundationには、オープンソースプロジェクト向けの確立されたモデルがあり、そこでは、プロジェクトへの貢献に対するインセンティブは、企業や個人がそのプロジェクトを採用することで得られる価値から生じます。

個人にとっても、情熱を注ぐこと、次の仕事につくこと、あるいは公の場で技術力をアピールすることなど、貢献する動機は十分に確立されています。組織のユーザーにとっては、プロジェクトのバグを可能な限りなくすこと、独自のデータを提供すること、データの品質を維持すること、あるいは必要な機能を追加することなどに、強い利害関係があります。プロジェクトがユーザーに金銭的な支援、新しいデータの追加や検証という形の労力、あるいはその両方を求めるという、この多面的なインセンティブモデルを的確に構築することが、プロジェクトの成功に向けた最初の重要な判断ポイントとなります。

OpenData.Orgのようなプロジェクトにおいては、金融サービス業界に膨大なデータセットを供給しているBloombergs、FactSets、MorningStarsといった既存のデータベンダーに対し、このようなプロジェクトを競合ではなく付加価値のあるものと捉えてもらえるよう、インセンティブモデルを構築することが含まれます。これにより、「BloombergがDNAに刻み込まれている」ような人々にとって、慣性の壁を乗り越えたり、新しいデータパイプラインへの切り替えや統合にかかるコストを吸収したりするのに役立つかもしれません。¹¹⁶

OpenSanctionsプロジェクトの創設者であるFriedrich Lindenberg氏は、OpenSanctionsを営利企業であると説明しましたが、同社が販売しているのはデータではなくライセンスです。「制裁データを必要とする人々は、単なるデータファイルを必要としているわけではありません。彼らに必要なのはベンダーとの関係です。商業的な利用において、データが利用可能であり、更新され、一定の品質が保たれていることを保証してくれるサプライヤーが必要なのです」と彼は述べました。「この分野では、サービスを販売できるのであれば、データを販売する必要はありません。当サイトはオープンデータポータルとして機能しており、ジャーナリスト、メディア機関、学者、市民社会が無料で利用しています。」¹¹⁷

プロジェクトの使命と価値観の設定

オープンソースコミュニティにとって新たなプロジェクトであるため、MITのNagle博士は、明確なミッションと範囲、プロジェクトの境界線の明確化、そして明確なリーダーシップを求めています。同博士は、単一の組織が「すべてを決定する」体制から、運営委員会による統括体制へと移行したプロジェクトが直面する成長の苦難を目の当たりにしてきました。

例として、彼は機械学習フレームワークのPyTorchを挙げました。「PyTorchはすでにオープンソースであり、幅広い貢献者コミュニティも存在していましたが、Metaが所有していました。2022年にMetaがPyTorchプロジェクトのガバナンスを非営利団体であるPyTorch Foundationに移管した際、貢献が増えるのではないかと推測されていました。」¹¹⁸ Nagle博士とその同僚の調査によると、Meta自身によるプロジェクトへの貢献は大幅に減少した一方で、他の営利企業、特に半導体メーカーのような補完企業からの貢献は増加しました。一方、アプリ開発者やクラウドプロバイダーからの貢献率は横ばいのままでした。¹¹⁹ Nagle博士は、この現象を、プロジェクトの主要なステークホルダーが抱く明確なビジョン、彼らが重視すること、そしてプロジェクトを通じて達成しようとしている目標に起因すると指摘しました。¹²⁰



Wikimedia Enterpriseの社長であるLane Becker氏は、明確なミッションに体现された価値観、「私たちが遵守すべき原則」、そして「私たちにそれらを遵守させる」という呼びかけの重要性を強調しました。¹²¹ 同氏は、「深いレベルでコミットし、『これが私たちの信条であり、これに基づいて私たちを評価してください』と明確に書き出し、それを公表することで、実際にその価値体系に縛られるようにすべきだ」と提言しました。Wikimedia Foundationについて、同氏は次のように述べました。「私たちは、その時点の状況に合わせて価値体系をさまざまな形で表現することもあります。価値体系そのものへのコミットメントは比類のないものです。私たちはこれほど長い間それを守り続けてきたため、他の何らかの目的のためにそれを放棄することなど、組織内では考えられないことです。」¹²²

Common Crawl FoundationのRich Skrenta事務局長は、倫理的な観点、すなわち私的利益と公共の利益のバランスについて指摘しました。¹²³ オープンデータ・プラットフォームを構築する組織は、大規模なアクセス可能性がもたらす公共の利益が、技術的にはアクセス可能ではあるものの検索が困難な紙媒体の公的記録が持つ「事実上の非公開性」の喪失に伴う潜在的な私的損害を上回るかどうかを自問しなければなりません。¹²⁴

データ対象者、ならびにデータ収集者やデータ充実化担当者を倫理的に採用し、適切に扱うことは極めて重要です。オープンデータプロジェクトの運営者は、対象者のプライバシー権および労働者の労働権を保護・尊重する、責任あるデータ収集の原則と実践を遵守しなければなりません。¹²⁵

適応性を考慮した設計

Data FoundationのHart博士は、「適応性があり、柔軟性さえ持ちつつ、基盤となるデータに影響を及ぼしうる政策や政府・企業の運営上の変化にリアルタイムで対応できるほど機動性の高いインフラを整備すること」を推奨しました。同氏は、「優れたデータアーキテクチャを構築し、これらすべてに備えて計画を立てるようにしてください」と述べました。同氏の見解では、経済分析局（BEA）と国勢調査局の統計インフラが長きにわたり維持されてきたのは、「非常に明確な一連のプロセスを有し、専門家から公的な助言を得ており、その手法について透明性を保っている」からであるとのことです。¹²⁶

Hart博士は、データ利用者が対応しなければならない「完全に公開されているもの」から「完全に特権的なもの」までのライセンスの連続性を認めました。「私たちの組織モデルでは、『オープンさ』という言葉を用いています。これは、一部のデータは常に完全に制限されたり専有されたりしているものの、それがよりオープンなデータの形成に寄与する『パイプライン』として捉えているからです。」同博士は、GDPを、さまざまな公共データ、専有データ、および非公開データで構成されるパブリックドメインのデータ例として挙げました。¹²⁷ 「これは、さまざまな種類のデータへのアクセスポイントのスペクトルを認識する必要がある理由の一例です」と彼は述べました。「一部のデータは本質的に機密かつ非公開として扱い、他のデータはGDPへと集約されるにつれてよりオープンなものとして扱います。これにより、産業別や経済のさまざまなレベルごとにデータを提示し、投資から政策立案に至るまでの意思決定に役立てることができます。」¹²⁸

「有用な情報を得たいのであれば、データガバナンスが鍵となります」とHart博士は付け加えました。「初期段階でデータを適切に管理できなければ、データに関するあらゆる問い合わせに対応するコストが高くなってしまいます。」同博士は、データを適切に管理できる形でサービスを設計・提供し、データを改善するための仕組みを整えることを提案しました。「データガバナンスは誰かの優先事項であると同時に、サービス提供そのものに組み込まれているものです。」提供チームは、それをワークフローに組み込んでいます。

BrightQueryのようなプロジェクトの規模と範囲を踏まえ、Hart博士は、高度な技術を持つユーザーや、極めて専門的な研究を想定して設計することには警鐘を鳴らしました。同博士は、このプロジェクトには「Amazonの評価スケールやコメントシステムのようなもの」が必要だと考えていました。それは、コミュニティレベルで人々がリアルタイムかつ積極的にフィードバックを行える仕組みであり、例えば、他の100人のユーザーが、自分の管轄区域や自社の業種に関してデータ品質の問題を抱えていたことを、ユーザーが確認できるようなものです。¹²⁹ 彼は、このシステムを「データを作成している主体だけに限定されない、非常に迅速な品質保証管理」として構想しており、プロジェクトチームが「迅速な仕組みとして構築できる」と考えていました。

ガバナンスへのユーザーの参画

「このプロジェクトがどのようなガバナンスモデルを採用するかは、差別化要因となるでしょう」と、GoogleのData Commonsに所属するPrem Ramaswami氏は述べました。ユーザーや貢献者にとっては、データの更新、追加、削除、および主張や情報源の裏付けを行うための透明性のあるプロセスが必要です。「組織に対して、自組織のエンティティを管理するよう呼びかけることは、データを最新の状態に保つための最良の動機付けとなります」と同氏は述べました。ただし、プロジェクトの拡大に伴い、それらの更新が審査待ちの状態で放置され、古くなってしまわない限り、という条件付きです。¹³⁰

OpenSanctionsのLindenberg氏によると、運営者たちが「包括的なガバナンスの仕組み」を構築するのに時間を費やしたにもかかわらず、「結局、誰もそのデータを利用しなかった」というオープンデータプロジェクトを数多く目にしてきました。同氏にとって重要なのは、「人々に自由に試してもらうこと」です。人々がデータを自由に試したり、成果を共有したり、自由に組み合わせたりできる場を作ることが大切なのです。そうすることで、プロジェクトの発起人は学ぶのです。「データセットに対するニーズが分かれば、それに基づいて仕組みを構築し、『このデータセットがその役割を果たすためには、こうした仕組みが必要だ』と明確に示せるようになります。」¹³¹

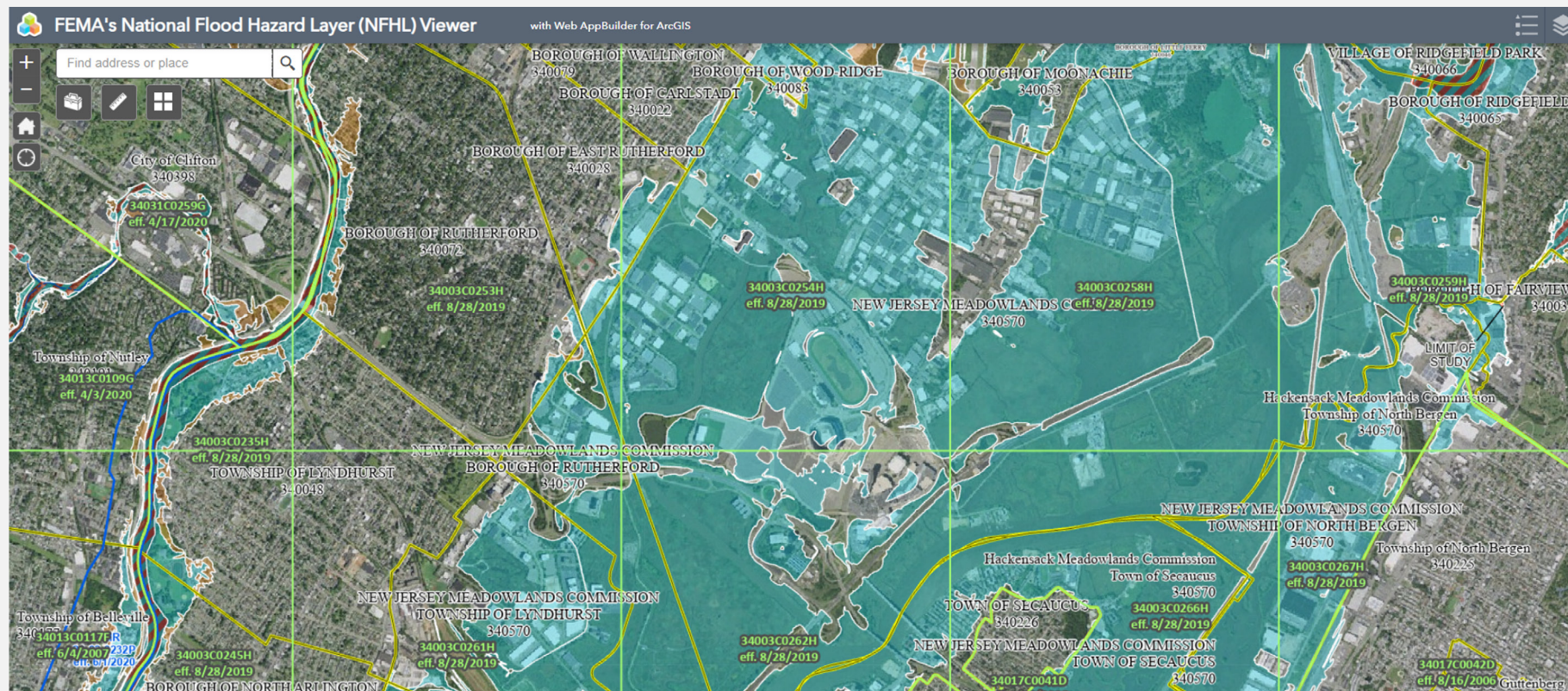
「有用な情報を得たいのであれば、データガバナンスが鍵となります。初期段階でデータを適切に管理できなければ、データに関するあらゆる問い合わせに対応するコストが高くなってしまいます。」

例えば、CODEは連邦緊急事態管理庁（FEMA）と協力し、最前線の地域社会を巻き込んで、FEMAの洪水ハザードマップの価値を評価しました。¹³² このプロジェクトを通じて、同庁に対して重要な利用者からのフィードバックが寄せられたほか、洪水リスクの軽減に役立つデータ、地図、その他の資産をまとめたCODEの公開リソースハブ「Flood Information Products Resource Hub」が構築されました。¹³³ CODEのJoel Gurin氏にとって、この経験はオープンデータの管理における重要な原則を浮き彫りにするものでした。「データセットの公開レビューやガバナンスについて考える際、データの質や正確性だけでなく、そのデータが表す人々や組織に与える潜在的な影響についても考慮しなければなりません」と彼は述べました。¹³⁴

Linux Foundationの下では、多くのプロジェクトがメンテナ主導のガバナンスモデルを採用しており、このモデルを通じて、誰でもパッチやプルリクエスト、マージリクエストを通じて貢献することができます。¹³⁵ コミッターやメンテナがそれらの変更をレビューし、権限を持つメンテナ、特別関心グループ、または技術運営委員会が、どの変更を採用するかを決定します。¹³⁶ 各プロジェクトは独立しています。Linux Foundationは、プロジェクトを法人として設立・運営するための支援を行い、プロジェクトのライセンス、貢献者契約、知的財産（商標やブランド名など）、インフラ（継続的インテグレーション、セキュリティ、コンプライアンスなど）、およびイベントやコミュニティの運営を支援しています。¹³⁷

図13：全国洪水ハザードレイヤービューア

2026年FIFAワールドカップ決勝戦の開催地であるNew Jersey Meadowlandsを中心としたこの図では、青・水色が洪水リスクの高い氾濫危険区域、黄褐色・オレンジ色が中程度のリスクのある洪水区域、色付けされていない部分がリスクの低い区域を示しており、赤い縞模様は洪水流路または水流が最も強い区域を表しています。



Sources: Federal Emergency Management Agency, National Flood Hazard Layer Viewer, US Dept. of Homeland Security; basemap imagery: US Department of Agriculture and US Geological Survey, accessed 24 May 2026 and used under 17 U.S.C. § 105. <https://hazards-fema.maps.arcgis.com/apps/webappviewer/index.html?id=8b0adb51996444d4879338b5529aa9cd>.

大規模なオープンデータプロジェクトでは、複雑な法的およびコンプライアンス上の義務も考慮しなければなりません。例えば、ウェブデータの収集、変換、配信を行う際、Common Crawl Foundationのような組織は、「一連の法的責任」を管理する必要があります。¹³⁸ これには、出所の追跡、個人を特定できる情報や違法なコンテンツなどの機密情報の検出・削除、著作権上の制限の遵守、ウェブクロールングおよびアーカイブに関するオプトアウト設定の尊重、デジタルミレニアム著作権法（DMCA）に基づく削除要請、ならびに欧州連合（EU）の一般データ保護規則（GDPR）などの規制に基づく「忘れられる権利」の要請への対応が含まれますが、Common Crawlはこれらすべてを実施しています。¹³⁹ 新しいオープンデータイニシアチブにとっての教訓は、法的ガバナンスが技術的ガバナンスと並行して進化しなければならないということです。安全策を講じずに急ぎすぎると、重大な法的リスクを招く恐れがあります。

高品質なデータの維持

「データの出所に関する完全な透明性は、極めて重要です」と、Wikimedia EnterpriseのBecker氏は述べています。「この情報はどこから来たのでしょうか？ このプロジェクトは、データそのもの、その出所、構造、誰がどのように貢献したかについて、どの程度まで完全な透明性を確保できるのでしょうか？ Wikipediaの観点から見れば、これは極めて明白なことです。」¹⁴⁰

同様に重要な優先事項は、高品質なデータです。「なぜ誰もが『データを正しく保つ』 ことについて語り続けるのでしょうか？ それは難しいことだからであり、データはあっという間に『正しくない』状態になってしまうからです」とGavrinsky氏は述べています。「たとえ一瞬だけ正しくても、あっという間に『正しくない』状態になってしまいます」。特に非参照型データについてはそうです。「誰もが、ある程度は自分たちのためにデータを『さらに正確』なものにし、そして非常に重要なこととして、データが引き続き正確であることを絶えず検証しなければなりません。」¹⁴¹ データの利用者でさえ、自らのニーズに合わせてデータを整備し、資産として積極的に管理しなければなりません。

「データを販売しようとするベンダーは、『当社には最高品質かつ最も大量のデータがあります』と言うでしょう。しかし、私はそれほど心を動かされません。なぜなら、重要なのは『品質』であり、品質は測定するのが難しく、証明するのはさらに難しいからです。しかし、そこにこそ真の価値があるのです」とGavrinsky氏は述べました。彼女は、ベンダーが効果的に収集・再編成・整理を行い、その努力を評価する顧客に販売している公共データのクラウドソーシングに、真の機会を見出しています。「顧客は最新のデータに対して対価を支払っているのかもしれませんが。しかし、先週のデータは公開されており、無料で入手可能であり、そうでなければ購入する余裕のない組織にとっては非常に価値のあるものになり得ます。もし私たち全員が公共データを統合すれば、その管理コストは誰にとっても安くなり、誰もがより質の高いデータセットを手に入れることができるでしょう。これこそがチャンスです。それが、私たちがFINOSを立ち上げた理由なのです」とGavrinsky氏は語りました。

リスクの管理と信頼の構築

プロジェクトの責任者や理事会には、最初からリスクについて率直に伝える必要があります。「リスクを直視すれば、それに対処できるようになります」とGavrinsky氏は述べました。「製品に潜むあらゆる潜在的な問題について、今考えておくほうがよいでしょう。そうしなければ、いずれ誰かがそれを指摘することになるからです。」¹⁴² 例として、彼女は、フィールドの密度と継続的な正確性から成る、公開の範囲とデータの正確性について言及しました。

「まず、何が公開情報で、何が無料、何が有料なのかを明確にしましょう。」例えば、これら300のフィールドがある企業に関する情報だとしましょう。このオープンデータモデルでは、ユーザーは最初の100項目を無料で利用できます——最初の100項目は公開されています——残りの200項目については、ユーザーが料金を支払うことになります。

第二に、すべての項目が完全に記入されているレコードがいくつあるかを明確にしましょう。「例えば、5,000万件の企業レコードがあるとします。その5,000万件のうち、無料の100項目がすべて入力されているレコードはいくつあるのでしょうか？ これを私は「フィールド密度」あるいは「データ密度」と呼んでいます。もしLEIが登録されている企業がわずか10万社しかないとしたら、それはあまり興味深いことではありません」と述べています。GLEIFが発行したLEIは323万件あり、そのうち299万件が有効であることを考慮すると。¹⁴³

第三に、それらの項目のうち、いくつかが正しいのかを明確にしておくことです。「仮に、その100の項目がすべて記入されているとします。それらは今日、正しいでしょうか？ 明日、明後日、その翌日はどうでしょうか？」¹⁴⁴

最後に、プロジェクトの責任者たちは、これらの項目が正しいことをどのように確認しているのでしょうか？どのような定番のテストを実施しているのでしょうか？どのような万全なプロセスを構築しているのか、そのプロセスを管理しているのは誰なのか、その担当者にはどのような研修を行っているのか、担当者が確実にプロセスに従っていることを確認するためにどのようなチェック体制を整えているのか、といった点についてです。

「金融サービス企業は、データベンダーに対してだけでなく、自社に対してもこうした質問を投げかけています。なぜなら、自社の責任をベンダーに転嫁することはできないからです。規制当局が罰金を科すのは企業側であり、ベンダー側ではないからです」とGavrinsky氏は述べました。¹⁴⁵

不測の事態に備えた計画が鍵となります。CODEのGurin氏は、「気候と経済的正義スクリーニングツール」をめぐり、オープンデータ・コミュニティが総力を挙げて取り組んだ様子について説明しました。環境品質評議会は、このツールを完全に開発・導入し、「Justice40イニシアチブ」に基づく連邦政府の投資指針として活用していました。¹⁴⁶ その後、2025年1月、連邦政府の措置により、このツールは一般公開から削除されました。わずか数日のうちに、2016年に環境データの監視・保存ネットワークとして設立された「Environmental Data and Governance Initiative」は、「Public Environmental Data Partners」との新たな連携を通じて、アーカイブされたデータセットからツールを再構築し、独立したプラットフォーム上でオンライン公開を再開しました。¹⁴⁷ 「非営利団体がツールを復活させたのです」と、Gurin氏は述べました。「彼らはそのために大規模なコミュニティ参加型プロセスを経て、非常に効果的にツールを再現しました。」¹⁴⁸

オープンイノベーションのためのライセンスの選定

OpenData.Orgのオープンエンティティグラフが提供するグローバルなエンティティ解決の精度を実現するため、組織は従来、プロプライエタリなデータフィードを直接購読するか、所有者の利用規約に基づき分析プラットフォームを介してアクセスしてきましたが、こうした方法では組織の活用範囲が制限されがちです。CDLA Permissive 2.0のようなオープンライセンスがあれば、ユーザーは、例えばOpenSanctionsなどの他のデータセットと組み合わせ、貿易コンプライアンス

ソリューションを開発することが可能になります。¹⁴⁹ AIモデルの学習においては、ライセンスの選択が特に重要です。データ プロヴェナンス イニシアチブ（DPI）によると、開発者がAIシステムの学習に使用するデータセットの多くは、明確かつ一貫したライセンスや出典表示が欠けています。¹⁵⁰ その結果、モデル開発においてデータが実際にどのように使用されているかについて、法的および倫理的な不確実性が広く生じています。また、DPIの調査によると、AI企業がウェブデータをスクレイピングし、商用モデルの学習に利用していることに対応して、ウェブサイト所有者が自社のウェブデータへのアクセスを制限し始めていることが判明しました。そして、データに対する管理権を取り戻そうとする所有者の取り組みにより、AIの学習に利用できるオープンウェブの範囲が縮小しつつあります。¹⁵¹ トレーニングデータの透明性、説明責任、および合法的な利用を確保するため、DPIは、AIコミュニティ（研究者、企業、政策立案者）に対し、より適切な文書化、標準化されたライセンス、監査可能なデータパイプラインといった、体系的なデータ出所管理の実践を採用するよう求めています。¹⁵²

CDLA Permissive 2.0 のようなオープンライセンスを採用する目的は、異なる提供者からのデータを可能な限り自由に組み合わせられるようにし、利用時の障壁を最小限に抑えることにあります。「オープンライセンスを採用していても、プロジェクトによってはライセンスに関する障壁が生じる可能性があります」と、Overture Maps Foundation の Rose 博士は述べています。¹⁵³ 例えば、合衆国法典第17編第105条に基づき、米国政府の職員や従業員が公務の一環として作成した米国政府の著作物（テキスト、データ、またはソフトウェア）には、著作権保護が適用されません。¹⁵⁴ その代わりに、これらの著作物は、米国内で公表された時点でパブリックドメインとなります。¹⁵⁵ したがって、連邦政府機関は、「ShareAlike」やその他の制限的なライセンス条項が付いたデータセットを基に新たなデータセットを作成することはできません。また、その結果として生成される統合データセットにそのような条項を課すようなプラットフォームを通じて、自らが保有するパブリックドメインのデータを提供することもできません。Googleのような商用ユーザーも、「ShareAlike」ライセンスには手を出さないでしょう。なぜなら、彼らが作成したものを常に共有することを保証できないからです。¹⁵⁶

「データのライセンス供与方法、および他のデータセットと組み合わせた際に下流にどのような影響が及ぶかについて、細心の注意を払ってください」とRose博士は述べました。¹⁵⁷ オープンデータ協定の下で明示的にライセンス供与されていないデータの取り込みを防ぐためのポリシーとプロセスを整備し、提供されたデータについては、利用者が利用規約を理解できるよう、明確な出典情報およびライセンス情報を提示してください。¹⁵⁸

「データをどのように活用し、どのように活用しないべきかについて、批判的に考えてください」

と、Wikimedia Enterpriseの社長、Lane Becker氏は述べました。「私たちは設立当初から、人々が自由な知識にアクセスできるよう、あらゆる道を可能にするために、ライセンスを極めて寛容なものに保つという意識的な選択をしてきました。これにはコンテンツの分岐（フォーク）も含まれますが、これは私たちの『オープンレッジ』という理念に沿って容認されるものであり、たとえそれが必ずしも当組織の使命や目標と一致しない場合でも同様です。」¹⁵⁹

例えば、Wikipediaは、そのテキストをCreative Commons Attribution-ShareAlike 4.0 License (CC BY-SA 4.0)の下で公開しています。¹⁶⁰ 2023年、あるロシア人編集者がWikipediaをフォークし、ロシア連邦の政治的優先事項やメッセージ戦略により沿ったバージョンである「Ruwiki」を作成しました。¹⁶¹ もう一つの例として、Elon Musk氏が2025年にWikipediaから分岐させて立ち上げた「Grokikipedia」が挙げられます。このサイトでは、人間の執筆者や編集者の代わりに、xAI社の言語モデル「Grok」が使用されています。参考文献の比較を行った研究者たちは、Grokikipediaが、企業、政府、シンクタンク、およびユーザー生成コンテンツをより多く引用していることを発見しました。その中には、陰謀論サイト「Infowars」、ネオナチ系フォーラム「Stormfront」、白人至上主義サイト「VVDare」など、Wikipediaコミュニティにおいて信頼性に欠けると見なされている情報源も含まれていました。¹⁶²

「ライセンス体系の見直しは、オープンソース界における『触れてはならないタブー』でした」とBecker氏は述べました。「まるで『ハリー・ポッター』シリーズの第5巻に飛び込んだかのように。子供たちは『ヴォルデモートが戻ってきた！』と叫びながら走り回っているのに、大人たちは『いや、そんなことはない』と応えているような状況です。」 学生たちには目撃証言と証拠があり、教授たちは所属する機関を信頼しています。「事態は決着の段階へと向かっています。状況があまりにも深刻で、もはや誰も無視できなくなっているからです。」¹⁶³

コミュニティとコミュニケーションをとる

「ここでの『Velocity』とはネットワーク効果のことです。つまり、志を同じくする人々がリソースを共有し、互いの取り組みを相乗効果で高め合うことで何が起るかということです。物事はより速く形になり、クリエイターたちはより速く創作活動を進められるようになります」と、SenzingのJeff Jonas氏は述べました。¹⁶⁴

例えば、Wikimedia Enterpriseの構想は、2年間にわたる公開協議プロセスを通じて、Wikimediaコミュニティからの提言として生まれました。¹⁶⁵ コミュニティは、「Wikimedia運動の持続可能性を高める」こと、および「プロジェクトから得られる有用な情報を、インターネット上で永続的に無料で利用可能にし、維持すること」という財団の使命を実現するためには、収益戦略が必要であると認識しました。¹⁶⁶

コミュニティは、エンタープライズグレードのAPIの開発と、それを収益源へと転換することを最優先課題としました。Wikimedia Enterpriseは、それを実現するための仕組みでした。¹⁶⁷ 2021年にWikimedia Foundationの完全子会社である有限責任会社として設立されたWikimedia Enterpriseは、WikimediaおよびWikimediaのプロジェクトデータを大量に、商用顧客向けに有料サービスとして提供しています。¹⁶⁸



この参加型プロセスを支えていたのは、Wikimediaが「ムーブメント・コミュニケーション」と呼ぶ活動に専念するチーム全体でした。¹⁶⁹ 「私たちのチームは、コミュニティとどのように関わっていくのが最善かについて多くの時間を費やし、私たちの取り組みについて早い段階でフィードバックをいただくようにしています」とBecker氏は説明しました。「それは、フィードバックが製品やサービスを適切に作り上げる上で不可欠であるだけでなく、何かを公表する前に、コミュニティが私たちの計画の背景を理解できる場を設けたいと考えているからです。これは業務の大きな部分を占めています。」

「この組織で行うすべての活動は、何らかの形で地域社会と連携して行われています」とBecker氏は述べました。「その緊張関係は、概して建設的なものです。たとえ最も否定的な意見であっても、このエコシステムにおいては有益なものとなります。結局のところ、それはあくまで人々が互いに語り合っているに過ぎないのです。」

世界中の優秀な人材の発掘、育成、および人材確保

データの利用、とりわけAIモデルでの利用には一定のリソースが必要となるため、UNICCのDandapani氏は、「現時点では、オープンデータが、より大きなAI能力や最大の演算能力を持つ主体に不釣り合いなほど恩恵をもたらしている」として、世界的な公平性を提唱しました。¹⁷⁰ 国際通貨基金（IMF）によると、低所得国では、ハイパースケール・コンピューティング施設、超高速ネットワーク、大容量の発電・冷却システム、そしてそれらを運用・保守するための熟練した人材へのアクセスが不足しています。¹⁷¹ 「グローバル・サウス（南半球の途上国）の主体が持つ能力を向上させることは不可欠です」とDandapani氏は述べました。

彼女の同僚であり、Digital Public SolutionsおよびUNICC Open Source Program Officeの責任者であるEmily Bennett氏も、この見解に同意しました。「おそらく最大の課題は、能力と知識でしょう。ソリューションの現状から恩恵を受けるためには、特定の技術だけでなく、人材を育成する基盤となる教育システムにも目を向ける必要があります」と彼女は述べました。「データ全般に関してグローバルコミュニティと協力する中で、各組織には何らかのデータガバナンスは存在しているものの、機密性の分類についてはあまり真剣に検討されていないことがわかります。」¹⁷² 彼らは、そもそもデータを公開すべきかどうかという点よりも、データの出所や保存場所の方をより重視しているのかもしれません。

両者は、学术界と国際機関との間で、目的を明確にした教育面での連携が役立つという点で意見が一致しました。例えば、UNICCとNew York University School of Professional Studiesは提携し、オープンデータ、生成AI、アジャイル手法を活用して、国連の6つの公用語およびさまざまなメディア形式にわたる有害コンテンツを検出・評価・分類するためのマルチモーダルなプロトタイプを開発しました。¹⁷³ もう一つのアプローチとして、オープンデータセットの活用事例をテーマにしたハッカソンに学生や教授陣を参加させ、各自のコミュニティに関連し、かつ固有のデータを追加・更新するよう呼びかけることが挙げられます。

最小限の実用可能な製品の推進

Linux Foundation傘下でのプロジェクトの立ち上げにあたり、インタビューに応じた専門家たちは、プロジェクトの責任者に対し、その「最小限の実用可能な製品（MVP）」を前面に打ち出すよう推奨しました。

独自性を際立たせましょう。「人々が可能性を見出せるよう、十分な詳細情報を提供してください」と、Overture Maps FoundationのRose博士は述べています。¹⁷⁴ 独自性のポイントは、単に事業名、正確な所在地、業種分類コードといった単純なものでも構いません。例えば、OpenData.Orgのデータセットは、米国だけでなく、世界222の国と地域の事業体を網羅しています。

課題を特定します。エンティティ解決データに関する既存の選択肢の多くは、サイロ化された独自仕様となっています。本プロジェクトは、すべてのデータサイロを結びつける、堅牢なオープンソースソリューションを求める業界のニーズに応えるものです。「ビジネスエンティティデータに、正確なジオコーディングが施された空間情報を追加することは、非常に刺激的であり、極めて有用です」と、Overture Maps FoundationのRose博士は述べています。¹⁷⁵

技術的負債が過度に蓄積されるのを避けましょう。技術的負債とは、例えばコードの堅牢性、最適化、再現性、あるいはドキュメントの作成において、開発者が行うトレードオフのことを指します。¹⁷⁶「他の人々をプロジェクトに巻き込みたいからこそ、迅速に進めたいと思うものです。人々に見ていただけるよう、何かを世に出したいのです」とRose博士は述べました。「しかし、技術的負債については非常に慎重になる必要があります。オープンなプロジェクトでは、その持続可能性を確保しなければならないからです。時には少人数で、時には大勢で、そしてほとんどの場合、メンバーが入れ替わりながら進めていくことになります。ですから、技術的な観点からは、プロジェクトに参加し、作業を行い、離れて、また戻って来ることが容易であるようにすべきです。そのためには、しっかりとした基盤が必要なのです。」¹⁷⁷

対話、合意形成、そして協働を重視しましょう。競争前の協業におけるタスクフォースやワーキンググループといったフォーラムは、創造性を飛躍的に高める原動力となります。それらは資金を有効活用できるだけでなく、参加者が他者がすでに成し遂げたことを繰り返したり、一から作り直したりする手間から解放し、新しいアイデアや解決策を導き出し、形作ること集中できるようにするのです。¹⁷⁸「異なる企業のメンバーを一堂に集めて話し合い、誰もが納得できる解決策にたどり着くのです。その解決策は合意形成によって導き出されたものであり、プロジェクトが適切な検討を経たものであることは誰もが理解しています」とRose博士は説明しました。「外部のステークホルダーであれ、データを提供する人々であれ、データを利用する組織であれ、あるいはプロジェクトのメンバーであれ、そうした対話に積極的に取り組むようにしてください。そうすることで信頼が築かれます。また、プロジェクトがどのように進化すべきかを理解する助けとなり、誰もがそのプロジェクトに意義を見出し続けられるようになります。」¹⁷⁹

すべてのデータが公開される必要はないことを覚えておいてください。「私たちは、エンティティを一意に識別するのに役立つデータは確保したいものの、詳細な情報は必要としません。その理由の一つは、そうしたデータが実際には汎用化されていない、つまり競争上の優位性を持つデータだからです」とローズ博士は説明しました。「私たちが実際に行っているのは、誰もが同じ言語でコミュニケーションをとれるよう基盤を築くことです。また、ユーザーがエンティティに、他のオープンデータであれ、非公開データ、機密データ、さらには機密指定されたデータであれ、あらゆる種類のデータを非常に簡単に紐付けられるようにしています。」¹⁸⁰



謝辞

オープンデータの管理と活用に関するご経験を、率直かつ明るい雰囲気の中で、そして冷静な楽観主義をもって共有して下さった以下の専門家の方々に、心より感謝申し上げます：

- Lane Becker, president, Wikimedia Enterprise
- Emily Bennett, head, Digital Public Solutions and the United Nations International Computing Centre Open Source Program Office
- Gabriele Columbro, executive director, Fintech Open Source Foundation
- Anusha Dandapani, chief, AI Hub, United Nations International Computing Centre
- Heidi Picher Dempsey, US research director, Red Hat LLC
- Michael Dolan, senior vice president of legal and strategic programs, The Linux Foundation
- Dr. Jose M. Plehn, founder and CEO, BrightQuery Inc.
- Jane Gavronsky, chief operating officer, Fintech Open Source Foundation
- Eloísa Granado, enterprise account executive, Wikimedia Enterprise
- Dr. Ramanathan V. Guha
- Joel Gurin, president and founder, Center for Open Data Exchange
- Dr. Nick Hart, president and CEO, Data Foundation
- Dr. Shay HersHKovitz, vice president of research, BrightQuery Inc.
- Dr. Jeff Jonas, founder, CEO, and chief scientist, Senzing Inc.
- Friedrich Lindenberg, founder, OpenSanctions

- Brian Malone, head of partnerships and product strategy, BrightQuery Inc.
- Dr. Frank Nagle, advising chief economist, The Linux Foundation; and research scientist, Initiative on the Digital Economy, Massachusetts Institute of Technology
- Isio Nelson, managing director of research, fraud, and thought leadership, ProSight Financial Association
- Prem Ramaswami, head, Data Commons, Google LLC
- Dr. Amy Rose, chief technology officer, Overture Maps Foundation
- Rich Skrenta, executive director, Common Crawl Foundation

本調査およびレビューの全過程において、独創的かつ多大な貢献をして下さった、Linux Foundation Researchの上級副社長であるHilary Carter氏、ならびに同組織の上級研究員兼エコシステム・マネージャーであるAnna Hermansen氏に、心より感謝申し上げます。

本研究にご支援をいただいた以下の団体に感謝申し上げます：



クレジット

図1および図3には、gungyoga04による位置情報アイコン、Smashiconsによる銀行(法人)アイコン、Ahmad Roaayalalによる(自宅)住所アイコン、Freepikによるユーザー(人物)アイコン、およびSmashiconsによるカレンダー(組織)アイコンが使用されており、これらはすべてMagnificのフリーライセンスの下で使用されています。図12には、Michael Irigoyen氏による電球のアラートアイコン、Michael Irigoyen氏によるロケット打ち上げアイコン、Simran氏による天秤のアイコン、Googleによるコンパスアイコン、GreenTurtwig氏によるアカウントグループのアイコン、およびAustin Andrews氏による現金倍率アイコンが使用されています。これらはすべてMDI/Pictogrammersから提供され、Apache License 2.0の下で使用されています。

著者について

Kirsten D. Sandberg氏は、研究者、ライター、編集者として、知的財産、イノベーション、科学、技術の各分野における研究プロジェクトに定期的に携わっています。同氏は、ブロックチェーン・リサーチ・インスティテュートの編集長を務めるとともに、ペース大学の大学院出版プログラムの非常勤講師も務めています。10年以上にわたり、ハーバード・ビジネス・パブリッシングにおいて、戦略、マーケティング、金融を専門とするエグゼクティブ・エディターを務めていました。



付録

表A1は、公開データセットに含まれる初期項目を示しており、表A2は公開データセットに含まれるデータ型について詳しく説明しています。また、表A3はBrightQueryプレミアムデータセットで利用可能な項目を示しており、表A4はプレミアムデータ型について説明しています。

表A1：経済の主要分野に関する公開データ

このオープンなエンティティデータプロジェクトは、222の国と地域を対象としています。

FIELD TYPE	ORGANIZATIONS	LOCATIONS	ADDRESSES	PEOPLE
Global	324 million	512 million	197 million	1.2 billion
United States	85.8 million	344.9 million	45.4 million	249 million
Open data	- Company ID - Company name - Legal name - D/B/A - Website URL - Company LinkedIn URL - Wikipedia URL - Crunchbase URL - Phone number - Number of locations - Ticker symbol - Capital IQ ID - CIK - FIGI ID - Set of LEIs - Set of ISINs - NPI ID - OSM ID - PERM ID - SAM ID Placekey - Google Plus Code IDs - HQ address with latitude-longitude and administrative area	- Location ID - Company ID it relates to - Location name - Website - Location phone number - Location address with latitude-longitude and administrative area and postal code - Placekey - OSM ID - GERS ID - Google Plus Code ID	- Building ID - Company IDs it relates to - Number of organizations - Legal entities - Locations and units associated with the address - Address with latitude-longitude and administrative area - Placekey - GERS ID - OSM ID - Google Plus Code IDs associated with the address	- Person ID - Company IDs it relates to - Person full name - State or province - Country - Person LinkedIn URL - Instagram URL - Wikipedia URL

Sources of data: “The World’s Largest Open Global Entity Graph,” Open Data Flyer, 10 Oct. 2025, p. 3, <https://opendata.org/BrightQuery-Open-Data-Flyer-Oct-2025.pdf>; Dashboard, <https://docs.brightquery.com/dashboard>; Data Dictionary, <https://opendata.org/#datadictionary>; Data feeds, <https://docs.brightquery.com/data-feeds>; and Dr. Shay Hershkovitz, vice president of research, BrightQuery, email to Kirsten Sandberg, 31 May 2026.

表A2 : OpenData. Orgにおけるデータ型の詳細

ABBREVIATION	FULL NAME	PURPOSE	STEWARD	SOURCE LINK
Capital IQ ID	Standard & Poor's (S&P) Capital IQ Identifier	Tracks entities and securities within the S&P database.	S&P Global Market Intelligence	https://www.spglobal.com
CIK	Central Index Key	Identifies corporations and individuals who file with the SEC.	US Securities and Exchange Commission	https://www.sec.gov
FIGI ID	Financial Instrument Global Identifier	An open standard for identifying financial instruments.	Object Management Group; Bloomberg LP (Registration Authority)	https://www.openfigi.com
GERS ID	Global Entity Reference System	A legacy identifier for global corporate entity mapping.	Overture Maps Foundation	https://overturemaps.org/gers/
IRS	Internal Revenue Service	US agency responsible for tax collection and law enforcement.	US Department of the Treasury	https://www.irs.gov
ISIN	International Securities Identification Number (ISO 6166)	Uniquely identifies a specific securities issue (stocks, bonds).	International Securities Identification Numbers Organization	https://www.isin.org/about/
LEI	Legal Entity Identifier (ISO 17442)	A global 20-character code to identify distinct legal entities.	Global Legal Entity Identifier Foundation (GLEIF)	https://www.gleif.org
NPI ID	National Provider Identifier	A unique 10-digit identification number for covered health care providers.	Centers for Medicare and Medicaid Services (CMS)	https://www.cms.gov
OFAC	Office of Foreign Assets Control	Administers and enforces economic and trade sanctions.	US Department of the Treasury	https://home.treasury.gov
OSM ID	OpenStreetMap Identifier	A unique ID for features (buildings, roads) in the OSM database.	OpenStreetMap Foundation	https://www.openstreetmap.org
PERM ID	Permanent Identifier	An open, permanent ID for entities, people, and organizations.	London Stock Exchange Group plc (LSEG) (formerly Refinitiv/Thomson Reuters)	https://www.permid.org
Placekey	Placekey	A universal standard identifier for physical places/points of interest.	Senzing acquired Placekey in 2025	https://www.placekey.io

表A2 : OpenData. Orgにおけるデータ型の詳細（続き）

ABBREVIATION	FULL NAME	PURPOSE	STEWARD	SOURCE LINK
PLEI	Proto Legal Entity Identifier	A global alphanumeric identifier for legal entities	OpenCorporates	https://p-lei.com/
RDI	Residential Delivery Indicator	Identifies whether a specific address is residential or commercial.	United States Postal Service (USPS)	https://www.usps.com
REG ID	Legal Entity Registration ID	Unique number issued by the jurisdiction with which the Legal Entity is registered.	Secretary of State	https://www.sos.ca.gov/
SAM ID	System for Award Management ID	Required for entities doing business with the US Federal Government.	General Services Administration (GSA)	https://sam.gov ; https://www.ecfr.gov/current/title-2/subtitle-A/chapter-I/part-25

In 2022, the US General Services Administration switched from Dun & Bradstreet’s proprietary Data Universal Numbering System (DUNS) to its own System for Award Management (SAM) ID, an open identifier standard with a unified tracking system for contractors and other third parties doing or seeking to do business with the government. All the US government’s other integrated award environment systems—such as CPARS, eSRS, FPDS, and FSRS—use the SAM ID for KYC and AML, and so the open solution has lowered the barriers to entering the federal marketplace.¹⁸¹

表A3 : BrightQueryから入手可能なプレミアムデータ

FIELD TYPE	ORGANIZATIONS	LOCATIONS	ADDRESSES	PEOPLE
Premium data	• Company type • Structure • Employer identification number (EIN) • NAICS, SIC, and IRS primary and all codes • names and sector data • OFAC and tax lien indicators • Year founded • BQ credit and confidence score • Most recent and historic revenue and employment numbers along with income statement and balance sheet items	• Organization name and website tied to the location • Location type • Class location categories (over 10,000) • Location confidence score • Location active indicator	• Building address type • Active indicator • State • Address RDI details • Building unit type • Address valid indicator • Building unit record type and carrier route • Legal entity and location level mapping to the address	• Legal, business, and consumer • Organization name • Website and ticker associated with person • Person's title • Person's role • All publicly available email and phone numbers associated with person • Complete address (often residential)

Source of data: “The World’s Largest Open Global Entity Graph,” Open Data Flyer, 14 Oct. 2025, p. 3, <https://opendata.org/BrightQuery-Open-Data-Flyer-Oct-2025.pdf>.

表A4 : プレミアムデータセットに含まれる追加のデータ型の詳細

ABBREVIATION	FULL NAME	PURPOSE	STEWARD	SOURCE LINK
Capital IQ ID	Standard & Poor’s (S&P) Capital IQ Identifier	Tracks entities and securities within the S&P database.	S&P Global Market Intelligence	https://www.spglobal.com
EIN	Employer Identification Number	Used by the IRS to identify business entities for tax purposes.	Internal Revenue Service (IRS)	https://www.irs.gov
NAICS	North American Industry Classification System	Classifies business establishments by their primary economic activity; replaced SIC.	US Census Bureau	https://www.census.gov
SIC	Standard Industrial Classification	A four-digit code used to identify the primary business of an entity; replaced by NAICS.	US Department of Labor (OSHA)	https://www.sec.gov/search-filings/standard-industrial-classification-sic-code-list ; https://www.osha.gov/data/sic-manual

参考文献

1. "How the Roman Census Worked (and Why It Mattered)," *UNRV Roman History*, n.d., <https://www.unrv.com/government/how-the-roman-census-worked-and-why-it-mattered.php>; Jeremy M. Norman, "The Domesday Book, Recording the First English Census," *History of Information*, ID 221, n.d., <https://www.historyofinformation.com/detail.php?id=221>; Ban Gu, "Han Shu," Chinese Text Project, n.d., <https://ctext.org/han-shu>; "Nonconformist and Non-Parish Registers: Births, Marriages and Deaths, 1567–1969," National Archives, UK Government, n.d., <https://www.nationalarchives.gov.uk/help-with-your-research/research-guides/nonconformist-non-parish-births-marriages-deaths-1567-1969/>.
2. LexisNexis Risk Solutions Study Reveals Global Financial Crime Compliance Costs for Financial Institutions Totals More Than US\$206 Billion," Press Release, LexisNexis Risk Solutions, 26 Sep. 2023, <https://risk.lexisnexis.com/global/en/about-us/press-room/press-release/20230926-global-financial-crime-compliance-costs>.
3. Tim Berners-Lee, "30 Years On, What's Next #ForTheWeb?" open letter, World Wide Web Foundation, 11 Mar. 2019, <https://medium.com/@webfoundation/30-years-on-whats-next-fortheweb-30a1d1ea034c>.
4. BrightQuery Open Data Flyer, OpenData.Org, 10 Oct. 2025, <https://opendata.org/BrightQuery-Open-Data-Flyer-Oct-2025.pdf>. See also R.V. Guha and Vineet Gupta, "Reference by Description," *arXiv*, 7 Mar. 2016, <https://arxiv.org/pdf/1511.06341>; Nishadi Kirielle, Peter Christen, and Thilina Ranbaduge, "Unsupervised Graph-Based Entity Resolution for Complex Entities," *ACM Transactions on Knowledge Discovery from Data* 17, no. 1 (20 Feb. 2023): 1–30, Association for Computing Machinery, <https://doi.org/10.1145/3533016>; and Mohammad Hossein Moslemi et al., "Heterogeneity in Entity Matching: A Survey and Experimental Analysis," *Data and Knowledge Engineering* 164 (Jul. 2026), <https://doi.org/10.1016/j.datak.2026.102575>.
5. Shay Hershkovitz, vice president of research, BrightQuery Inc., email to Kirsten Sandberg, 31 Mar. 2026.
6. Jose M. Plehn Dujowich, interviewed via Zoom by Kirsten Sandberg, 12 Jan. 2026.
7. Sources of data in BrightQuery's open data project: US government offices, https://docs.brightquery.com/government_offices; US local jurisdictions, https://docs.brightquery.com/local_jurisdictions; US federal agencies, https://docs.brightquery.com/federal_agencies; US Secretary of State offices, https://docs.brightquery.com/secretary_of_state_jurisdiction; and Canadian agencies and registries, https://docs.brightquery.com/canadian_agencies_registries, as of 24 Feb. 2026.
8. Shay Hershkovitz, vice president of research, BrightQuery Inc., email to Kirsten Sandberg, 31 Mar. 2026.
9. "What does GDPR stand for?" <https://gdpr.eu/what-does-it-stand-for/>; "Right to erasure," <https://gdpr-info.eu/art-17-gdpr/>; California Consumer Privacy Act of 2018, <https://oag.ca.gov/privacy/ccpa>; and "Right to delete," <https://oag.ca.gov/privacy/ccpa#sectiond>.
10. Tokenization here refers to substituting a sensitive data element with a nonsensitive surrogate value that can be re-associated with the original data only through separately secured mapping. See ISO/IEC 20889:2018, *Privacy Enhancing Data Deidentification Terminology and Classification of Techniques* (Geneva: International Organization for Standardization, 2018), <https://www.iso.org/standard/69373.html>; and Simson Garfinkel et al., NIST SP 800-188, *Deidentifying Government Datasets: Techniques and Governance* (Gaithersburg, MD: National Institute of Standards and Technology, Sep. 2023), <https://doi.org/10.6028/NIST.SP.800-188>.
11. Jose M. Plehn Dujowich, PhD, interviewed via Zoom by Kirsten Sandberg, 12 Jan. 2026. Cade Metz and Karen Weise, "AI Is Getting More Powerful, but Its Hallucinations Are Getting Worse," *New York Times*, 5 May 2025, <https://www.nytimes.com/2025/05/05/technology/ai-hallucinations-chatgpt-google.html>.
12. See, for example, https://docs.brightquery.com/bulk_data_feed and https://docs.brightquery.com/business_identity_api.
13. Jose M. Plehn, "Laying the Groundwork for Open Data with the Linux Foundation," PowerPoint presentation, BrightQuery, Nov. 2025, p. 52.
14. ChatGPT reached 100 million users in about two months after its launch in late 2022, whereas TikTok took ~9 months and Instagram, ~2.5 years, in part because it leveraged existing Internet, cloud, and smartphone infrastructure, and technology vendors integrated generative AI into their offerings. Krystal Hu, "ChatGPT sets record for fastest-growing user base," Analyst Note, *Reuters*, 2 Feb. 2023, <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>; Ravi Mayuram, "Six tips for combatting vendor lock-in and the data gravity challenge," *VentureBeat*, 24 Apr. 2023, <https://venturebeat.com/data-infrastructure/6-tips-combat-vendor-lock-in-data-gravity>; and Sagar Joshi, "AI Adoption Statistics," *G2 Lean*, G2.com Inc. 28 May 2025, <https://learn.g2.com/ai-adoption-statistics>.

15. Jim Zemlin, email to Hilary Carter, 17 Mar. 2026.
16. World Bank, “AI-Ready Official Statistics: Opportunities, Challenges, and Recommendations,” UNSTATS, Dec. 2025, https://unstats.un.org/UNSDWebsite/statcom/session_57/documents/BG-5h-CCSA_AI_Readiness_Official_Statistics_v2-E.pdf.
17. Prem Ramaswami, interviewed via Zoom by Kirsten Sandberg, 2 Feb. 2026.
18. “LexisNexis Risk Solutions Study Reveals Global Financial Crime Compliance Costs for Financial Institutions Totals More Than US\$206 Billion,” Press Release, LexisNexis Risk Solutions, 26 Sep. 2023, <https://risk.lexisnexis.com/global/en/about-us/press-room/press-release/20230926-global-financial-crime-compliance-costs>.
19. Craig Hale, “Businesses are being ‘locked in’ to all-in-one platforms, and it’s costing them growth and adaptability,” *TechRadar*, 20 Mar. 2026, <https://www.techradar.com/pro/businesses-are-being-locked-in-to-all-in-one-platforms-and-its-costing-them-growth-and-adaptability>.
20. For other data licenses such as C-UDA-1.0, O-UDA-1.0, ODbL-1.0 ODC-By-1.0, and PDDL-1.0, see <https://spdx.org/licenses/>.
21. Ravi Mayuram, “Six tips for combatting vendor lock-in and the data gravity challenge,” *VentureBeat*, 24 Apr. 2023, <https://venturebeat.com/data-infrastructure/6-tips-combat-vendor-lock-in-data-gravity/>; and Superblocks team, “What Is Vendor Lock-in? Five Strategies and Tools to Avoid It,” *Superblocks Blog*, DayZero Software Inc., 21 Mar. 2025, <https://www.superblocks.com/blog/vendor-lock>.
22. Joel Gurin, interviewed via Zoom by Kirsten Sandberg, 5 Feb. 2026.
23. “National Secure Data Service Demonstration,” National Center for Science and Engineering Statistics, as of 20 Apr. 2026, <https://nces.nsf.gov/initiatives/national-secure-data-service-demo>.
24. “BrightQuery and Senzing Partner to Deliver Faster, Smarter Business Data Resolution,” Press Release, Senzing Inc., n.d., <https://senzing.com/bright-query-partnership-data-resolution/>.
25. Jose M. Plehn, “A Shared Entity-Based Infrastructure for the Modern AI Economy: Building the World’s Knowledge Graph at the Entity Level,” 1 May 2026. Shared with the research project team.
26. Jose M. Plehn Dujowich, PhD, interviewed via Zoom by Kirsten Sandberg, 12 Jan. 2026.
27. A couple of interviewees described how their respective organization’s database had multiple instances of a single person or a company. The most common reasons were (1) the lack of a universal unique identifier for each entity across an organization’s relational and enterprise systems, (2) variations in data entry and formats such as “IBM,” “I.B.M.,” or “International Business Machines,” and (3) multiple input systems and integration processes that create redundant and sometimes conflicting entries across operational, compliance, and analytics databases. See Gartner Research, “Data Quality: Best Practices for Accurate Insight,” Gartner Inc., n.d., <https://www.gartner.com/en/data-analytics/topics/data-quality>, accessed 1 Mar. 2026.
28. Jeff Jonas, interviewed via Zoom by Kirsten Sandberg, 30 Jan. 2026.
29. Dan Ennis, “\$4B in PPP Loans Were Duplicates or Had Mismatched Data, Analysis Finds,” *Banking Dive*, TechTarget Inc., Informa PLC, 2 Sep. 2020, <https://www.bankingdive.com/news/paycheck-protection-program-mismatched-data-SBA/584578/>; and Office of Inspector General (Hannibal Ware), Report 21-09: *Duplicate Loans Made Under the Paycheck Protection Program*, US Small Business Administration, 15 Mar. 2021, <https://www.sba.gov/sites/default/files/2021-03/SBA%20OIG%20%20Report%2021-09.pdf>.
30. George Anadiotis, “What Is a Knowledge Graph?” Ontotext, n.d., <https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/>; and Google Data Commons, “About Data Commons,” n.d., <https://datacommons.org/about>, both accessed 1 Mar. 2026.
31. Henrik Liliendahl Sørensen, “MDM and Knowledge Graph,” *Liliendahl on Data Management*, 28 Nov. 2021, <https://liliendahl.com/2021/11/28/mdm-and-knowledge-graph/>; Nishadi Kirielle, Peter Christen, and Thilina Ranbaduge, “Unsupervised Graph-Based Entity Resolution for Complex Entities,” *ACM Transactions on Knowledge Discovery from Data* 17, no. 1 (Jan. 2023), <https://doi.org/10.1145/3533016>; and Prashanth Rao and Paco Nathan, “From Data to Insights: Entity-Resolved Knowledge Graphs with Kùzu and Senzing,” *Kùzu Blog*, Kùzu Inc., 9 Jun. 2025, <https://blog.kuzudb.com/post/entity-resolved-knowledge-graphs/>.
32. Isio Nelson, interviewed via Zoom by Kirsten Sandberg, 13 Feb. 2026.
33. Olivia White et al., *Digital Identification: A Key to Inclusive Growth* (San Francisco: McKinsey Global Institute, 17 Apr. 2019): vi, <https://www.mckinsey.com/~media/mckinsey/business-functions/mckinsey-digital/our-insights/digital-identification-a-key-to-inclusive-growth/mgi-digital-identification-report.pdf>.
34. “Companies House celebrates 10 years of open data,” UK Government press release, GOV.UK, published 22 Jun. 2025, <https://www.gov.uk/government/news/companies-house-celebrates-10-years-of-open-data>.

35. Department for Business and Trade and Companies House, *Value of Corporate Transparency in Tackling Crime: Policy Summary* (London: UK Government, 15 Oct. 2024), https://assets.publishing.service.gov.uk/media/670e554d366f494ab2e7b88c/policy_summary_report_value_corporate_transparency_tackling_crime_october_2024.pdf.
36. Department for Business, Energy, and Industrial Strategy and Companies House, *Valuing the User Benefits of Companies House Data: Policy Summary*, BEIS Research Paper Number 2019/015 (London: UK Government, 27 Sep. 2019), <https://assets.publishing.service.gov.uk/media/5d8a299aed915d5cff89a4a1/valuing-benefits-companies-house-data-policy-summary.pdf>.
37. Alexander Verhagen et al., "How Agentic AI Can Change the Way Banks Fight Financial Crime," *Insights*, McKinsey & Co., 7 Aug. 2025, <https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/how-agentic-ai-can-change-the-way-banks-fight-financial-crime>.
38. Hui Gong, "AI Agents in Financial Markets: Architecture, Applications, and Systemic Implications," *arXiv*, 14 Apr. 2026, <https://arxiv.org/pdf/2603.13942>.
39. Zornitsa Manolova, "Transforming Data into Opportunities: Metric of the Month: Legal Entity Events," *GLEIF Blog*, Global Legal Entity Identifier Foundation, 5 Dec. 2025, <https://www.gleif.org/en/newsroom/blog/transforming-data-into-opportunities-metric-of-the-month-legal-entity-events>.
40. Experian, "Global Identity and Fraud Report 2024," *Experian Blog*, Experian Information Solutions Inc., 5 Nov. 2024, https://www.experian.com/blogs/global-insights/wp-content/uploads/2024/11/Global_Fraud_Trends_Report_2024_FinalV.pdf.
41. Jessica Hung, Isaac Owusu, and Ross Gabay, "Build Fraud Detection Systems Using AWS Entity Resolution and Amazon Neptune Analytics," *AWS Database Blog*, Amazon Web Services, 5 Feb. 2026, <https://aws.amazon.com/blogs/database/build-fraud-detection-systems-using-aws-entity-resolution-and-amazon-neptune-analytics>.
42. "Entity Resolution Software," product description, DataWalk, accessed 1 Mar. 2026, <https://datawalk.com/product/entity-resolution/>.
43. Ian Hook, "OpenCorporates' Data Powers Anti-Money Laundering Knowledge Graph Prototype," *OpenCorporates Blog*, OpenCorporates, 15 Jul. 2021, <https://blog.opencorporates.com/2021/07/15/opencorporates-data-powers-anti-money-laundering-knowledge-graph-prototype>; and Data Innovation Lab, EDM Council, as of 1 Mar. 2026, <https://edmcouncil.org/innovation/innovation-lab/>.
44. Mingming Geng, "Enhancing Entity Resolution with Multichannel BERT [bidirectional encoder representations from transformers]: A Comprehensive Approach," *Proc. SPIE 13171*, Third International Conference on Algorithms, Microchips, and Network Applications (AMNA 2024), 1317126 (8 Jun. 2024), <https://doi.org/10.1117/12.3031934>; and Jwen Fai Low, Benjamin C.M. Fung, and Pulei Xiong, "Better Entity Matching with Transformers Through Ensembles," *Knowledge-Based Systems* 293 (2024), <https://doi.org/10.1016/j.knosys.2024.111678>.
45. Philip Bruno et al., "The Legal Entity Identifier: The Value of the Unique Counterparty ID," Global LEI Foundation and McKinsey & Co., 15 Oct. 2017, p. 4, <https://www.mckinsey.com/~media/mckinsey/industries/financial%20services/our%20insights/the%20legal%20entity%20identifier%20the%20value%20of%20the%20unique%20counterparty%20id/legal-entity-identifier-mckinsey-gleif-2017.pdf>.
46. EQT Partners, "CompanyKG: A Large-Scale Company Relation Graph for Investment Industry," GitHub, v1.0, <https://github.com/EQTPartners/CompanyKG>; and Lele Cao et al., "CompanyKG: A Large-Scale Heterogeneous Graph for Company Similarity Quantification," *arXiv*, 18 Jun. 2023, <https://arxiv.org/abs/2306.10649>.
47. LLC Research, "A Large-Scale Heterogeneous Graph for Company Similarity Quantification and Relation Prediction, accepted by KDD'2024," GitHub, v2.0, <https://github.com/llcresearch/CompanyKG2>; and <https://zenodo.org/records/11391315>.
48. Abhinav Arun et al., "FinReflectKG: Agentic Construction and Evaluation of Financial Knowledge Graphs," in *Sixth ACM International Conference on AI in Finance* (ICAIF '25), Singapore, 15–18 Nov. 2025 (New York, NY: Association for Computing Machinery, 2025): 283–290, <https://doi.org/10.1145/3768292.3770363>; and the README page of the Financial Knowledge Graph Dataset: S&P 100 Companies, <https://web.archive.org/web/20250826030522/https://anonymous.4open.science/api/repo/KG-Financial-Datasets-SP-100-529B/file/README.md>.
49. "Wikidata," Wikipedia, last edited 22 Jan. 2023, https://www.wikidata.org/wiki/Wikidata:Main_Page, accessed 1 Mar. 2026.
50. John Walsh and Ruben Falk, "Agentic GraphRAG for Capital Markets," *AWS for Industries Blog*, Amazon Web Services Inc., 24 Feb. 2026, <https://aws.amazon.com/blogs/industries/agentic-graphrag-for-capital-markets/>.
51. Heidi Picher Dempsey, "From Particles to Prototypes: What We Learn from Managing Open Clouds," From the Director column, *Red Hat Research Quarterly*, Nov. 2024, <https://research.redhat.com/blog/article/from-particles-to-prototypes-what-we-learn-from-managing-open-clouds/>.

52. Prem Ramaswami, interviewed via Zoom by Kirsten Sandberg, 2 Feb. 2026.
53. See ServiceNow, “Create a Knowledge Base,” Extend ServiceNow AI Platform Capabilities, 31 Jul. 2025, <https://www.servicenow.com/docs/r/servicenow-platform/knowledge-management/create-a-knowledgebase.html>.
54. Andrew Brown, Muhammad Roman, and Barry Devereux, “A Systematic Literature Review of Retrieval-Augmented Generation Models,” *Journal of Intelligent Systems* (8 Aug. 2025), <https://arxiv.org/pdf/2508.06401v1.pdf>; and Ehlullah Karakurt and Akhan Akbulut, “Retrieval-Augmented Generation and Large Language Models for Enterprise Knowledge Management and Document Automation: A Systematic Literature Review,” *Applied Sciences* 16, no. 1 (29 Dec. 2025), <https://doi.org/10.3390/app16010368>. See also DEEP-PolyU, “Awesome-GraphRAG,” GitHub, last updated 3 Mar. 2026, <https://github.com/DEEP-PolyU/Awesome-GraphRAG>.
55. “Agentic Hybrid RAG Factual AI Platform,” BrightQuery AI Inc., as of 10 Apr. 2026, <https://brightquery.ai/agentic/>.
56. Shaoxiong Ji et al., “A Survey on Knowledge Graphs: Representation, Acquisition and Applications,” *IEEE Transactions on Neural Networks and Learning Systems* 33, no. 2 (2022): 494–514, <https://arxiv.org/abs/2002.00388>.
57. Stiene Riemer et al., “How Retail Banks Can Put AI Agents to Work,” Boston Consulting Group, 9 Mar. 2026, <https://www.bcg.com/publications/2026/how-retail-banks-can-put-agentic-ai-to-work>.
58. Peter Christen, *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection* (Berlin: Springer, 2012), <https://doi.org/10.1007/978-3-642-31164-2>; and Stefano Marchesin, Gianmaria Silvello, and Omar Alonso, “Large Language Models and Data Quality for Knowledge Graphs,” *Information Processing and Management* 62, no. 6 (Nov. 2025): 104281, <https://doi.org/10.1016/j.ipm.2025.104281>.
59. Jam Kraprayoon, Zoe Williams, and Rida Fayyaz, “AI Agent Governance: A Field Guide,” Institute for AI Policy and Strategy, *arXiv*, 27 May 2025, <https://arxiv.org/pdf/2505.21808>.
60. Elevate, “AI Data Governance: Provenance, Quality, and Model Lineage,” *Elevate Consult Blog*, Elevate Consult LLP, 30 Dec. 2025, <https://elevateconsult.com/insights/ai-data-governance-provenance-quality-and-model-lineage/>.
61. WNS Analytics, “Information Retrieval Using Knowledge Graphs,” *WNS Analytics AI Lab*, WNS Holdings Ltd., n.d., <https://www.wns.com/capabilities/analytics/wns-ai-lab/enhancing-information-retrieval-using-knowledge-graphs>; Ciyuan Peng et al., “Knowledge Graphs: Opportunities and Challenges,” *Artificial Intelligence Review*, 3 Apr. 2023: 1–32, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10068207/>; and Yilun Zheng et al., “Less Is More: Denoising Knowledge Graphs for Retrieval-Augmented Generation,” *arXiv*, 16 Oct. 2025, <https://arxiv.org/pdf/2510.14271>.
62. Prem Ramaswami, interviewed via Zoom by Kirsten Sandberg, 2 Feb. 2026.
63. AWS Labs, “AWS SageMaker Graph-Based Credit Scoring,” GitHub, last updated 27 Feb. 2023, <https://github.com/aws-labs/sagemaker-graph-based-credit-scoring>, accessed 1 Mar. 2026.
64. OpenCorporates Ltd., “OpenCorporates,” OpenCorporates, accessed 3 Mar. 2026, <https://opencorporates.com/>.
65. Isio Nelson, interviewed via Zoom by Kirsten Sandberg, 13 Feb. 2026.
66. See, for example, Wikimedia contributors, “TerminusDB,” Wikipedia, last modified 18 Jan. 2026, <https://en.wikipedia.org/wiki/TerminusDB>; and Wikimedia contributors, “JanusGraph,” Wikipedia, last modified 10 Feb. 2026, <https://en.wikipedia.org/wiki/JanusGraph>.
67. João B.G. de Brito et al., “Potential Customer Lifetime Value in Financial Institutions: The Usage of Open Banking Data to Improve CLV Estimation,” *arXiv*, 28 Jun. 2025, <https://arxiv.org/abs/2506.22711>.
68. Isio Nelson, interviewed via Zoom by Kirsten Sandberg, 13 Feb. 2026.
69. Last edited on 19 Feb. 2026, <https://en.wikipedia.org/wiki/Wikipedia:Wikidata>.
70. Friedrich Lindenberg, “Inside OpenSanctions’ Open Source Approach to Compliance,” *21 Analytics Blog*, 21 Analytics AG, 8 Feb. 2026, <https://www.21analytics.co/blog/opensanction-open-source-approach-compliance/>.
71. Wikimedia contributors, “TerminusDB,” Wikipedia, last modified 18 Jan. 2026, <https://en.wikipedia.org/wiki/TerminusDB>; Apache 2.0 License, <https://github.com/terminusdb/terminusdb/blob/main/LICENSE>; Wikimedia contributors, “JanusGraph,” Wikipedia, last modified 10 Feb. 2026, <https://en.wikipedia.org/wiki/JanusGraph>; and Apache 2.0 License, <https://github.com/JanusGraph/janusgraph?tab=License-1-ov-file#readme>.

72. "Improving Different Ultimate Beneficial Ownership Investigations," *GraphAware Blog*, Graph Aware Ltd., 24 Jun. 2024, <https://graphaware.com/blog/ubo-investigation-with-graph-technology-copy/>.
73. Wikimedia Deutschland, "Wikidata: Embedding Project," Wikidata, Wikimedia Foundation, last modified 27 Jan. 2026, https://www.wikidata.org/wiki/Wikidata:Embedding_Project. See also Dianne Esber et al., "Reinventing marketing workflows with agentic AI," *McKinsey Insights*, 21 Apr. 2026, <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/reinventing-marketing-workflows-with-agentic-ai>.
74. Amy Rose, interviewed via Zoom by Kirsten Sandberg, 22 Dec. 2025.
75. "Who We Are," Overture Maps, as of 16 Feb. 2026, <https://overturemaps.org/about/who-we-are/>.
76. Amy Rose, interviewed via Zoom by Kirsten Sandberg, 22 Dec. 2025.
77. Richard Robinson, "How Two Unique Identifiers Can Unlock the Value of ESG Data Faster," *Chief Data Officer Magazine*, 29 Apr. 2025, <https://www.cdomagazine.tech/opinion-analysis/how-two-unique-identifiers-can-unlock-the-value-of-esg-data-faster>.
78. Richard Robinson, "How Two Unique Identifiers Can Unlock the Value of ESG Data Faster," *Chief Data Officer Magazine*, 29 Apr. 2025, <https://www.cdomagazine.tech/opinion-analysis/how-two-unique-identifiers-can-unlock-the-value-of-esg-data-faster>.
79. Israel Griol-Barres et al., "Detecting Weak Signals of the Future: A System Implementation Based on Text Mining and Natural Language Processing," *Sustainability* 12, no. 19 (20 Sep. 2020): 7848, <https://doi.org/10.3390/su12197848>.
80. Christian Mühlroth, Laura Kölbl, and Michael Grottke, "Innovation Signals: Leveraging Machine Learning to Separate Noise from News," *Scientometrics* 128, no. 5 (12 Apr. 2023): 2649–2676, <https://doi.org/10.1007/s11192-023-04672-y>.
81. Yuchen Zhao, Xiaogang Bi, and Qing-Ping Ma, "Predicting Mergers and Acquisitions: A Machine Learning-Based Approach," *International Review of Financial Analysis* 99 (Mar. 2025), <https://doi.org/10.1016/j.irfa.2025.103933>; Karen S. Markel, Mihir Tale, and Andrea Belz, "JobPulse: A Big Data Approach to Real-Time Engineering Workforce Analysis and National Industrial Policy," *arXiv*, 14 Aug. 2025, <https://arxiv.org/pdf/2508.11014>; Sebastian Heinrich, "Deriving Technology Indicators from Corporate Websites: A Comparative Assessment Using Patents," *Applied Economics Letters* 32, no. 1 (14 Aug. 2023): 28–41, <https://www.tandfonline.com/doi/full/10.1080/13504851.2023.2244228>.
82. Prem Ramaswami, interviewed via Zoom by Kirsten Sandberg, 2 Feb. 2026.
83. Harold Hotelling, "Stability in Competition," *Economic Journal* 39, no. 153 (1929): 41–57, <https://doi.org/10.2307/2224214>; and Anthony Downs, *An Economic Theory of Democracy* (New York: Harper & Row, 1957), <https://archive.org/details/economictheoryof00down>.
84. Ramanathan V. Guha, interviewed via Zoom by Kirsten Sandberg, 30 Jan. 2026.
85. Prem Ramaswami, interviewed via Zoom by Kirsten Sandberg, 2 Feb. 2026.
86. Amy Rose, interviewed via Zoom by Kirsten Sandberg, 22 Dec. 2025.
87. Anna Hermansen and Kirsten Sandberg, "The Value of Open Source AI for APEC Economies: A Review of Industry, Academic, and Open Source Evidence," The Linux Foundation, 31 Oct. 2025, https://www.linuxfoundation.org/hubfs/Research%20Reports/LFResearch_OSAI_APEC_Version2_103125.pdf.
88. Emily Bennett and Anusha Dandapani, interviewed via Zoom by Kirsten Sandberg, 10 Feb. 2026.
89. Prem Ramaswami, interviewed via Zoom by Kirsten Sandberg, 2 Feb. 2026.
90. Jane Gavronsky, interviewed via Zoom by Kirsten Sandberg, 4 Feb. 2026; FINOS Landscape of projects, <https://landscape.finos.org/stats>.
91. Jane Gavronsky, interviewed via Zoom by Kirsten Sandberg, 4 Feb. 2026.
92. Friedrich Lindenberg, interviewed via Zoom by Kirsten Sandberg, 2 Feb. 2026; and "About OpenSanctions," <https://www.opensanctions.org/docs/about/>, as of 14 Mar. 2026.
93. Nicholas Gruen and John Houghton, *Open for Business: How Open Data Can Help Achieve the G20 Growth Target* (Port Melbourne, VIC: A Lateral Economics report commissioned by Omidyar Network, Jun. 2014), https://lateraleconomics.com.au/wp-content/uploads/omidyar_open_business.pdf; and Jaana Mäkelä and Luukas Raatikainen, *The Economic Value of Spatially Enabled Services in Finland* (Helsinki: Ministry of Agriculture and Forestry and Finnish Geospatial Research Institute, 2018), <https://geoforum.fi/wp-content/uploads/2023/04/The-economic-value-of-spatially-enabled-services-in-Finland.pdf>.
94. James Manyika, Sven Smit, and Jonathan Woetzel, "Financial Data Unbound: The Value of Open Data for Individuals and Institutions," McKinsey Global Institute, McKinsey & Co., Jun. 2021, p. 6, https://www.mckinsey.com/~media/mckinsey/industries/financial_services/our_insights/financial_data_unbound_the_value_of_open_data_for_individuals_and_institutions/financial-data-unbound-discussion-paper-june-2021.pdf.

95. Len Fishman, "How Open Data Induces Financial Firms to Compete on Quality: An Interview with Bloomberg's Richard Robinson," *data.world blog*, data.world Inc., 15 Aug. 2017, <https://data.world/blog/how-open-data-induces-financial-firms-to-compete-on-quality-an-interview-with-bloombergs-richard/>.
96. Joel Gurin, interviewed via Zoom by Kirsten Sandberg, 5 Feb. 2026. See also "Roundtable on Data for Automated Vehicle Safety Summary Report," foreword by Derek Kan, *Transportation.gov*, US Dept. of Transportation, 12 Feb. 2018, <https://www.transportation.gov/sites/dot.gov/files/docs/policy-initiatives/automated-vehicles/304471/roundtable-data-automated-vehicle-safety-report.pdf>; and ITS Joint Program Office, "ITS DataHub: About," *Intelligent Transportation Systems*, US Dept. of Transportation, as of 15 Mar. 2026, <https://its.dot.gov/data/about>.
97. "Work Zone Data Exchange Overview," *Transportation.gov*, US Dept. of Transportation, 7 Nov. 2018, <https://www.transportation.gov/sites/dot.gov/files/docs/policy-initiatives/automated-vehicles/325341/work-zone-data-exchange-overview.pdf>; and "Data for Automated Vehicle Integration (DAVI)," *Transportation.gov*, US Dept. of Transportation, n.d., <https://www.transportation.gov/av/data>.
98. Joel Gurin, via email to Kirsten Sandberg, 31 Mar. 2026.
99. Ramanathan V. Guha, interviewed via Zoom by Kirsten Sandberg, 30 Jan. 2026.
100. Jane Gavronsky, interviewed via Zoom by Kirsten Sandberg, 4 Feb. 2026.
101. Jose M. Plehn Dujowich, PhD, interviewed via Zoom by Kirsten Sandberg, 25 Nov. 2025.
102. US Securities and Exchange Commission, Form 10-K (Annual report pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934), *SEC.gov*, <https://www.sec.gov/files/form10-k.pdf>; SEC, Form 10-Q (Quarterly report pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934), *SEC.gov*, <https://www.sec.gov/files/form10-q.pdf>; and SEC, "Webmaster Frequently Asked Questions," *SEC.gov*, last updated 23 Aug. 2024, <https://www.sec.gov/about/webmaster-frequently-asked-questions#reuse>. These reports synthesize corporate financial performance and include audited financial statements, risk factors, disclosures about operations and governance, and management discussion and analysis.
103. Frank Nagle, interviewed via Zoom by Kirsten Sandberg, 7 Jan. 2026.
104. Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 Dec. 2022 ... as regards corporate sustainability reporting, *EUR-Lex*, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2464>; "Corporate sustainability reporting," *European Commission: Finance*, 9 Dec. 2025, https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en; OECD, *Board Responsibility and Sustainability-Related Disclosure in Asia*, OECD Publishing, Paris, 27 Nov. 2025, <https://doi.org/10.1787/8d2672e7-en>; and Samantha J. Rowe et al., "Worldwide Sustainability Reporting: Current State of Play," *Debevoise & Plimpton Insights*, Debevoise & Plimpton LLP, 24 Mar. 2026, <https://www.debevoise.com/insights/publications/2026/03/worldwide-sustainability-reporting-current-state>.
105. Prem Ramaswami, interviewed via Zoom by Kirsten Sandberg, 2 Feb. 2026.
106. Michael Dolan, interviewed via Zoom by Kirsten Sandberg, 15 Jan. 2026.
107. "Kernel Overview," Android Open Source Project, as of 16 Jan. 2026, <https://source.android.com/docs/core/architecture/kernel>; Evan Ackerman, "How NASA Designed a Helicopter That Could Fly Autonomously on Mars," *IEEE Spectrum*, 17 Feb. 2021, <https://spectrum.ieee.org/nasa-designed-perseverance-helicopter-rover-fly-autonomously-mars>; and Taylor Hill, "Meet the Open-Source Software Powering NASA's Ingenuity Mars Helicopter," *JPL News*, NASA Jet Propulsion Laboratory, 8 Jul. 2021, <https://www.jpl.nasa.gov/news/meet-the-open-source-software-powering-nasas-ingenuity-mars-helicopter/>.
108. The Linux kernel is available under GNU General Public License version 2.0 (GPL-2.0), <https://docs.kernel.org/process/license-rules.html>.
109. Members, LF AI & Data Foundation, <https://lfai.foundation/about/members/>. For more on precompetitive collaborations, see Institute of Medicine Roundtable on Translating Genomic-Based Research for Health, "Requisites for Successful Precompetitive Collaboration," *Establishing Precompetitive Collaborations to Stimulate Genomics-Driven Product Development: Workshop Summary* (Washington DC: National Academies Press; 2011). <https://www.ncbi.nlm.nih.gov/books/NBK54320/>; and Jeffrey S. Barrett, "The Precompetitive Space for Drug or Vaccine Development: What Does It Look Like Now and What Could It Look Like in the Future?" *Journal of Pediatric Pharmacology and Therapeutics* 28, no. 5 (2023): 465–472, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10731930/#>.
110. Michael Dolan, interviewed via Zoom by Kirsten Sandberg, 15 Jan. 2026. See *Feist Publications Inc. v. Rural Tel. Service Co.*, 499 U.S. 340 (1991), <https://www.loc.gov/resource/usrep.usrep499340/>.
111. "Who is Overture for?" Overture Maps Foundation, as of 15 Feb. 2026, <https://overturemaps.org/>.
112. Frank Nagle, interviewed via Zoom by Kirsten Sandberg, 7 Jan. 2026.
113. "Why Host Your Project at the Linux Foundation?" The Linux Foundation, 2023, <https://www.linuxfoundation.org/hubfs/Why%20Host%20a%20Project.pdf>.
114. Nick Hart, interviewed via Google Meet by Kirsten Sandberg, 19 Feb. 2026.

115. Joel Gurin, interviewed via Zoom by Kirsten Sandberg, 5 Feb. 2026.
116. A phrase quoted by Frank Nagle, advising chief economist at the Linux Foundation and a research scientist at the Initiative on the Digital Economy at the Massachusetts Institute of Technology. Frank Nagle, interviewed via Zoom by Kirsten Sandberg, 7 Jan. 2026. See also <https://ide.mit.edu/people/frank-nagle/>.
117. Friedrich Lindenberg, interviewed via Zoom by Kirsten Sandberg, 2 Feb. 2026.
118. Frank Nagle, interviewed via Zoom by Kirsten Sandberg, 7 Jan. 2026. See also “Meta Transitions PyTorch to the Linux Foundation, Further Accelerating AI/ML Open Source Collaboration,” press release, The Linux Foundation, 12 Sep. 2022, <https://www.linuxfoundation.org/press/press-release/meta-transitions-pytorch-to-the-linux-foundation>.
119. Daniel Yue and Frank Nagle, “Igniting Innovation: Evidence from PyTorch on Technology Control in Open Collaboration,” Harvard Business School Working Paper No. 25-013, 10 Sep. 2024, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4960578.
120. Frank Nagle, interviewed via Zoom by Kirsten Sandberg, 7 Jan. 2026.
121. Please find Wikimedia Foundation’s mission here, as of 27 Feb. 2026, <https://wikimediafoundation.org/who-we-are/mission/>; and its principles here, last revised 16 Sep. 2025, <https://en.wikipedia.org/wiki/Wikipedia:Principles>.
122. Lane Becker, interviewed via Zoom by Kirsten Sandberg, 4 Feb. 2026.
123. Rich Skrenta, interviewed via Zoom by Kirsten Sandberg, 3 Feb. 2026.
124. See the landmark case of US Department of Justice, *Reporters Committee for Freedom of the Press v. Department of Justice*, 489 US 749 (1989), <https://supreme.justia.com/cases/federal/us/489/749/>; and David S. Ardia, “Privacy and Court Records: Online Access and the Loss of Practical Obscurity,” *University of Illinois Law Review* 2017, no. 5 (4 Aug. 2017): 1385–1454, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3013704.
125. Mohammad Amir Anwar, “Africa’s Data Workers Are Being Exploited by Foreign Tech Firms: Four Ways to Protect Them,” *The Conversation*, The Conversation US Inc., 31 Mar. 2025, <https://theconversation.com/africas-data-workers-are-being-exploited-by-foreign-tech-firms-4-ways-to-protect-them-252957>; and Michelle Du and Chinasa T. Okolo, “Reimagining the Future of Data and AI Labor in the Global South,” Brookings Institution, 7 Oct. 2025, <https://www.brookings.edu/articles/reimagining-the-future-of-data-and-ai-labor-in-the-global-south/>.
126. Nick Hart, interviewed via Google Meet by Kirsten Sandberg, 19 Feb. 2026.
127. See US Bureau of Economic Analysis, “The Making of GDP,” *BEA in Brief*, US Dept. of Commerce, rev. 2 Apr. 2020, <https://www.bea.gov/sites/default/files/2020-04/BEA-in-Brief-GDP-Final.pdf>; and US Bureau of Economic Analysis, “Input-Output Accounts: Who Sells What to Whom,” *BEA in Brief*, US Dept. of Commerce, 23 Feb. 2021, <https://www.bea.gov/sites/default/files/2021-02/BEA-Input-Output-Accounts.pdf>.
128. Nick Hart, interviewed via Google Meet by Kirsten Sandberg, 19 Feb. 2026.
129. Nick Hart, interviewed via Google Meet by Kirsten Sandberg, 19 Feb. 2026.
130. Prem Ramaswami, interviewed via Zoom by Kirsten Sandberg, 2 Feb. 2026.
131. Friedrich Lindenberg, interviewed via Zoom by Kirsten Sandberg, 2 Feb. 2026.
132. Joel Gurin, interviewed via Zoom by Kirsten Sandberg, 5 Feb. 2026.
133. Center for Open Data Enterprise, “CODE’s Flood Resource Hub: A Guide for Flood Information Products,” *Open Data Enterprise Blog*, 14 Jun. 2023, <https://odenterprise.medium.com/codes-flood-resource-hub-a-guide-for-flood-information-products-8ed26b36273d>; and Center for Open Data Enterprise, “Flood Resource Hub,” *Airtable*, as of 15 Mar. 2026, <https://airtable.com/app7R95PbZ6EDQDUn/shrvjvJPxBYwTmP5/tblilyvW01TuPdnUi/viwTJ0UJ86nLjuqPQ>.
134. Joel Gurin, interviewed via Zoom by Kirsten Sandberg, 5 Feb. 2026.
135. Stormy Peters and Nithya Ruff, “Participating in Open Source Communities,” *Open Source Guides*, The Linux Foundation, n.d., <https://www.linuxfoundation.org/resources/open-source-guides/participating-in-open-source-communities>, as of 1 Mar. 2026.
136. See, for example, “Contributor Guide,” The Kubernetes Authors, <https://www.kubernetes.dev/docs/guide/>; and “Special Interest Groups,” Community Groups, The Kubernetes Authors, <https://www.kubernetes.dev/community/community-groups/>, as of 1 Mar. 2026.
137. “Why Host Your Project at the Linux Foundation?” The Linux Foundation, 2023, <https://www.linuxfoundation.org/hubfs/Why%20Host%20a%20Project.pdf>.
138. Rich Skrenta, interviewed via Zoom by Kirsten Sandberg, 3 Feb. 2026.

139. Internet Engineering Task Force, "RFC 9309: Robots Exclusion Protocol," *RFC Editor*, Sep. 2022, <https://www.rfc-editor.org/rfc/rfc9309.html>; US Copyright Office, "Digital Millennium Copyright Act," <https://www.copyright.gov/dmca/>; European Parliament and Council of the European Union, "Right to be Forgotten," <https://gdpr.eu/right-to-be-forgotten/>; "General Data Protection Regulation," *EUR-Lex*, as of 4 May 2016, <https://eur-lex.europa.eu/eli/reg/2016/679/oj>; and Rich Skrenta, "Setting the Record Straight: Common Crawl's Commitment to Transparency, Fair Use, and the Public Good," *Common Crawl Blog*, Common Crawl Foundation, 4 Nov. 2025, <https://commoncrawl.org/blog/setting-the-record-straight-common-crawls-commitment-to-transparency-fair-use-and-the-public-good>.
140. Lane Becker, interviewed via Zoom by Kirsten Sandberg, 4 Feb. 2026.
141. Jane Gavronsky, interviewed via Zoom by Kirsten Sandberg, 4 Feb. 2026.
142. Jane Gavronsky, interviewed via Zoom by Kirsten Sandberg, 4 Feb. 2026.
143. LEI Statistics, as of 3 Mar. 2026, <https://www.gleif.org/en/lei-data/global-lei-index/lei-statistics>. See also the Global LEI System Statistics Dashboard, <https://www.gleif.org/assets/components/global-lei-system-statistics-dashboard/tableau-dashboard.html>.
144. Jane Gavronsky, interviewed via Zoom by Kirsten Sandberg, 4 Feb. 2026.
145. Jane Gavronsky, interviewed via Zoom by Kirsten Sandberg, 4 Feb. 2026.
146. Sara L. Jordan, "Climate and Economic Justice Screening Tool: Frequently Asked Questions," *Biden White House Archives*, Council on Environmental Quality and the White House, Feb. 2022, rev. 14 Mar. 2022, <https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/02/CEJ-ST-QandA.pdf>.
147. Public Environmental Data Partners, "Climate and Economic Justice Screening Tool," *CEJST 2.0*, GitHub.io, n.d., <https://public-environmental-data-partners.github.io/j40-cejst-2/en/#3/33.47-97.5>.
148. Joel Gurin, interviewed via Zoom by Kirsten Sandberg, 5 Feb. 2026.
149. "About OpenSanctions," as of 2 Mar. 2026, <https://www.opensanctions.org/docs/about/>.
150. Shayne Longpre et al., "The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing and Attribution in AI," arXiv, 25 Oct. 2023, <https://arxiv.org/abs/2310.16787>.
151. Shayne Longpre et al., "Consent in Crisis: The Rapid Decline of the AI Data Commons," arXiv, 24 Jul. 2024, <https://arxiv.org/pdf/2407.14933>.
152. Shayne Longpre et al., "A Large-Scale Audit of Dataset Licensing and Attribution in AI," *Nature Machine Intelligence* 6 (30 Aug. 2024): 975–987, <https://doi.org/10.1038/s42256-024-00878-8>.
153. Amy Rose, interviewed via Zoom by Kirsten Sandberg, 22 Dec. 2025.
154. "17 U.S.C. § 101 (Definitions)," Legal Information Institute, Cornell Law School Legal Information Institute, <https://www.law.cornell.edu/uscode/text/17/101>.
155. "17 U.S.C. § 105 (Subject Matter of Copyright: US Government Works)," Legal Information Institute, Cornell Law School Legal Information Institute, <https://www.law.cornell.edu/uscode/text/17/105>.
156. Prem Ramaswami, interviewed via Zoom by Kirsten Sandberg, 2 Feb. 2026.
157. Amy Rose, interviewed via Zoom by Kirsten Sandberg, 22 Dec. 2025.
158. See, for example, Overture Maps' "Attribution and Licensing" page, last updated 21 Jan. 2026, <https://docs.overturemaps.org/attribution/>.
159. Lane Becker, interviewed via Zoom by Kirsten Sandberg, 4 Feb. 2026.
160. Creative Commons Attribution-ShareAlike 4.0 License Deed, as of 10 Apr. 2026, <https://creativecommons.org/licenses/by-sa/4.0/deed.en>.
161. Smallbones, "Knowledge Manipulation on Ruwiki, the Russian Wikipedia Fork," in *Wikimedia Research Newsletter* 15(7) Jul. 2025, last edited 26 Jul. 2025, https://meta.wikimedia.org/wiki/Research:Newsletter/2025/July#Knowledge_manipulation_on_Ruwiki,_the_Russian_Wikipedia_fork; Yiriy Marin, "How Russia Is Creating Its Own Alternative to Wikipedia," *Russia Post*, 25 Jun. 2024, https://russiapost.info/society/alt_wikipedia; "Russian Wikipedia's Top Editor Leaves to Launch a Putin-Friendly Clone," *Bloomberg*, Bloomberg LP, 12 Jul. 2023, <https://www.bloomberg.com/news/articles/2023-07-12/russian-wikipedia-editor-leaves-to-launch-a-putin-friendly-clone>; and Sergey Volkov, "At the Russian Internet Forum, Wikimedian Vladimir Medeiko presented an alternative to the Russian Wikipedia," *News, World of Encyclopedias*, 24 May 2023, <https://www.encyclopedia.ru/news/enc/detail/82756/>.
162. Tilman Bayer, "Comparing Comparisons of Grokipedia vs. Wikipedia by Three Different Research Teams," in *Wikimedia Research Newsletter* 15(12) Dec. 2025, last edited 3 Feb. 2026, <https://meta.wikimedia.org/wiki/Research%3ANewsletter/2025/December>; Reece Rogers, "Elon Musk's Grokipedia Pushes Far-Right Talking Points," *WIRED*, 27 Oct. 2025, <https://www.wired.com/story/elon-musk-launches-grokipedia-grokipedia-competitor/>; Harold Triedman and Alexios Mantzarlis, "What Did Elon Change? A Comprehensive Analysis of Grokipedia," *CorarXiv*:2511.09685, 12 Nov. 2025, <https://arxiv.org/pdf/2511.09685>; and David Ingram, "Elon Musk's Grokipedia Cites Neo-Nazi Website 42 Times—Study," *NBC News*, 20 Nov. 2025, <https://www.nbcnews.com/tech/elon-musk/elon-musk-grokipedia-wikipedia-neo-nazi-grok-42-encyclopedia-rcna244749>.

163. Lane Becker, interviewed via Zoom by Kirsten Sandberg, 4 Feb. 2026.
164. Jeff Jonas, interviewed via Zoom by Kirsten Sandberg, 30 Jan. 2026.
165. "About," *Movement Strategy*, Wikimedia Meta-Wiki, last edited on 25 Jun. 2024, https://meta.wikimedia.org/wiki/Movement_Strategy/About.
166. "Increase the Sustainability of Our Movement," *Movement Strategy*, Wikimedia Meta-Wiki, last edited on 18 Jan. 2026, https://meta.wikimedia.org/wiki/Movement_Strategy/Recommendations/Increase_the_Sustainability_of_Our_Movement; and "Mission," Wikimedia Foundation, as of 27 Feb. 2026, <https://wikimediafoundation.org/who-we-are/mission/>.
167. "Revenue Streams," *Movement Strategy*, Wikimedia Meta-Wiki, last edited 21 Apr. 2023, https://meta.wikimedia.org/wiki/Movement_Strategy/Recommendations/Iteration_2/Revenue_Streams/1-8; and "Wikimedia Enterprise," Wikimedia Meta-Wiki, last edited 24 Feb. 2026, https://meta.wikimedia.org/wiki/Wikimedia_Enterprise.
168. "Wikimedia Foundation launches Wikimedia Enterprise," *News*, Wikimedia Foundation, 25 Oct. 2021, <https://wikimediafoundation.org/news/2021/10/25/wikimedia-foundation-launches-wikimedia-enterprise-the-new-opt-in-product-for-companies-and-organizations-to-easily-reuse-content-from-wikipedia-and-wikimedia-projects/>; "About Wikimedia Enterprise," Wikimedia Enterprise, as of 26 Feb. 2026, <https://enterprise.wikimedia.com/api/>.
169. Lane Becker, interviewed via Zoom by Kirsten Sandberg, 4 Feb. 2026.
170. Emily Bennett and Anusha Dandapani, interviewed via Zoom by Kirsten Sandberg, 10 Feb. 2026.
171. Eugenio Cerutti et al., *The Global Impact of AI: Mind the Gap*, IMF Working Paper No. 25/76 (Washington, DC: International Monetary Fund, 8 Apr. 2025), <https://www.imf.org/-/media/files/publications/wp/2025/english/wpia2025076-print-pdf.pdf>.
172. Emily Bennett and Anusha Dandapani, interviewed via Zoom by Kirsten Sandberg, 10 Feb. 2026.
173. Anusha Dandapani, Shambhavi Mohan, Devyani Rastogi, and Andres Fortino. *Responsible AI Innovation for Social Impact: A Case Study in Multilingual Media Moderation*. United Nations International Computing Centre and New York University School of Professional Studies, 8 Sep. 2025, <https://www.unicc.org/wp-content/uploads/2025/09/UNICC-White-Paper-Responsible-AI-Innovation-for-Social-Impact-1.pdf>.
174. Amy Rose, interviewed via Zoom by Kirsten Sandberg, 22 Dec. 2025.
175. Amy Rose, interviewed via Zoom by Kirsten Sandberg, 22 Dec. 2025.
176. "Shipping first time code is like going into debt. A little debt speeds development so long as it is paid back promptly with a rewrite," said Ward Cunningham, computer programmer and inventor of the wiki. "The danger occurs when the debt is not repaid. Every minute spent on not-quite-right code counts as interest on that debt. Entire engineering organizations can be brought to a stand-still under the debt load of an unconsolidated implementation." See Ward Cunningham, "The WyCash Portfolio Management System," *OOPSLA '92 Experience Report*, ACM Conference on Object-Oriented Programming, Systems, Languages, and Applications (26 Mar. 1992), C2.com, <https://c2.com/doc/oopsla92.html>.
177. Amy Rose, interviewed via Zoom by Kirsten Sandberg, 22 Dec. 2025.
178. Paul A. David and Francesco Rullani, "Dynamics of Innovation in an 'Open Source' Collaboration Environment: Lurking, Laboring, and Launching FLOSS Projects on SourceForge," *Industrial and Corporate Change* 17, no 4 (Aug. 2008): 647–710, <https://doi.org/10.1093/icc/dtn026>; Sheen Levine and Michael Prietula, "Open Collaboration for Innovation: Principles and Performance," *arXiv*, 8 Jan. 2014, <https://arxiv.org/pdf/1406.7541>; Xiaohong Chen and Yuan Zhou, "Open-Source Collaboration and Technological Innovation in the Industrial Software Industry: A Multi-Case Study," *Systems* 13, no. 6 (3 Jun. 2025): 433, <https://www.mdpi.com/2079-8954/13/6/433>; Hongbo Fang, et al., "Weak Ties Explain Open Source Innovation," *arXiv*, 29 Oct. 2025, <https://arxiv.org/pdf/2411.05646>.
179. Amy Rose, interviewed via Zoom by Kirsten Sandberg, 22 Dec. 2025.
180. Amy Rose, interviewed via Zoom by Kirsten Sandberg, 22 Dec. 2025.
181. US General Services Administration, "Unique Entity ID Is Here," *Integrated Award Environment Systems Information Kit*, last updated 23 Jul. 2025, <https://www.gsa.gov/about-us/organization/federal-acquisition-service/fas-initiatives/integrated-award-environment/iae-systems-information-kit/unique-entity-id-is-here>; and Office of Management and Budget, "Universal Identifier and System for Award Management," 2 C.F.R. pt. 25, *Electronic Code of Federal Regulations*, 14 Sep. 2010, <https://www.ecfr.gov/current/title-2/subtitle-A/chapter-I/part-25>.



BrightQueryは、222カ国にわたる3億2400万の組織、5億1200万の拠点、12億の連絡先を網羅し、政府データに基づく世界で最も包括的なグローバル企業情報を提供しています。非上場企業のデータに特化したBrightQueryは、10万以上の連邦・州・地方自治体の政府機関から、公式の公開書類を通じて情報を収集しています。BrightQueryは、米国国家セキュアデータサービス（NSDS）の主要な政府データパートナーとして機能しており、主要なAI企業向けの主要な政府データサプライヤーでもあります。また、BrightQueryは、AI Alliance, Linux Foundation, FINOS, Overture Maps Foundation, Agentic AI Foundation, Computing Research Association, EDM Council, Data Foundationなど、多くのAIおよびオープンソースコミュニティのメンバーでもあります。



2021年に設立されたLinux Foundation Research は、拡大を続けるオープンソースの協業の規模を調査し、新たな技術トレンド、ベストプラクティス、そしてオープンソースプロジェクトが世界にもたらす影響に関する洞察を提供しています。プロジェクトデータベースやネットワークを活用し、定量的・定性的手法におけるベストプラクティスを徹底することで、Linux Foundation Researchは、世界中の組織の利益となる、オープンソースに関する知見の頼れる情報源を構築しています。

 x.com/linuxfoundation

 facebook.com/TheLinuxFoundation

 linkedin.com/company/the-linux-foundation

 youtube.com/user/TheLinuxFoundation

 github.com/LF-Engineering



Copyright © 2026 The Linux Foundation

本報告書は、Creative Commons Attribution-NoDerivatives 4.0 International Public Licenseの下で提供されています。本著作物を引用される際は、以下のとおり記載してください：Kirsten D. Sandberg, “The Value of Open Data on Global Entities: Turning the World’s Largest Open Entity Graph into a Catalyst for Collaboration, Innovation, and Transformation,” foreword by Gabriele Colombo, The Linux Foundation, June 2026.

この日本語文書は、英語版を機械翻訳し、参考訳として、The Linux Foundation Japanが便宜上提供するものです。

翻訳協力：吉田 行男