

商業用不動産デジタルインフラの構築に向けた技術と取り組み

株式会社 estie 代表取締役 平井 瑛
ひらい えい

1. はじめに

2010年代以降、様々な産業をソフトウェアが飲み込む中、不動産業界はその波から確実に取り残されているように見える。もっとも、不動産業界は世界で最も大きなハードウェア産業であり、そのデジタル化には大きな困難を伴うことは想像に難くない。その困難の大きな部分は、間違いなく基礎的なデータ整備の問題に起因する。

巨大で複雑な構造物である不動産は、ハードウェアとしての性質だけではなく、より抽象度の高い性質を持っている。分筆された多数の土地の上にまたがって所在し、それらの土地には借地権、所有権などの関係が入り乱れる。建物自体の権利関係も共有、準共有、区分所有など多岐に渡り、利用者としてのテナントも契約形態は普通借家、定期借家など様々である。もちろん、設備や権利といったストック情報に加え、さらに資産としてのオペレーショナルな性質からくるフロー情報も多く持つ。空室に対して入居者がどのような条件で募集されて、どんな顧客候補と商談を重ねているのか、工事の予定はいつで、キャッシュフローをどれくらい生んでいるのか。

こういった様々な顔を持つ不動産を一意にデジタルデータとして扱い、さらに時系列変化にも耐えうるデータ構造を構築するためには、深い業界知見と、ソフトウェア技術への精通が求められることは想像に難くない。

本稿では、株式会社 estie（エスティ）が実際に開発した技術の解説を通じて、奥深い不動産業界のデータ構築、デジタル化の一端を紹介したい。

2. 株式会社 estie の事業

株式会社 estie は2018年12月に創業した不動産テック企業であり、日本最大級のオフィスビルデータ分析基盤である「estie マーケット調査^{*1}」をはじめとして、物流施設のデータ基盤を提供する「estie 物流リサーチ^{*2}」など、複数の不動産データサービスを提供している。当社のデータを活用する顧客は、東急不動産や野村不動産といった大手不動産デベロッパーや、KJR マネジメントといった大手J-REITの運用会社のおよそ7割にのぼり、そのほかにも金融機関など多くのユーザーに活用されている。

オフィスや物流施設といった、所謂商業用不動産業界では、古くから電話や対面商談での会話を中心とした情報交換を行い、そこで得た非公式な情報に立脚して投資や運用の意思決定を行うことが一般的であった。しかし近年、空室率の悪化や急激な新規賃料の調整を背景として、リアルタイムで網羅的なデータを収集し、賃料設定の継続的改善による収益最大化を模索する動きが広まりつつある。

2020年6月にリリースした「estie マーケット調査」は、日本全国8万棟のオフィスビル情報に

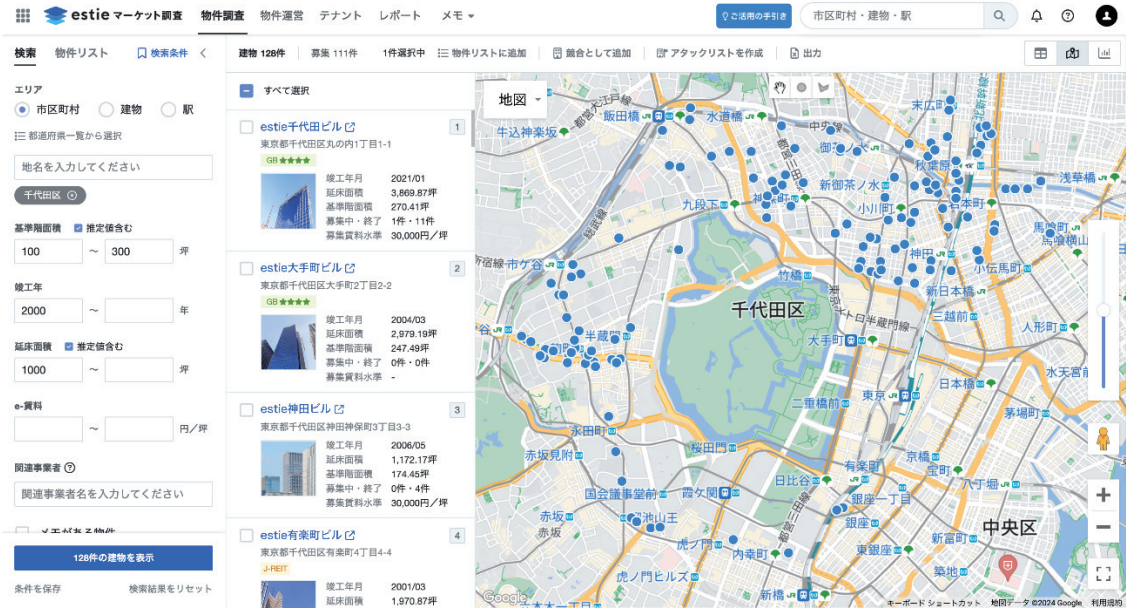


図1: estie マーケット調査 (estie 作成)



図2: estie 物流リサーチ (estie 作成)

加えて、日次で更新される 10,000 件の空室情報、東京・神奈川・大阪・愛知・福岡の主要エリアにおけるオフィス入居テナント情報を提供している。また、特筆すべきはその賃料情報にある。都心 5

区に所在する賃貸オフィスビル 26,770 棟のうち、規模の大小を問わず 87.4%の物件について具体的な賃料のデータを備えていた。インターネット上で取得できる賃料情報は「未定」や「相談」が

多く、賃料開示率が約 6.2%程度であったことを考えると、「estie マーケット調査」の賃料データの網羅性が確認できる（2023 年の当社調査による）。

2023 年 10 月にリリースをした「estie 所有者リサーチ（旧名：estie 物件売買）^{※3}」においては、登記簿情報を大規模にデータベース化することによって、物件ごとの所有権の移転履歴や、事業者ごとのポートフォリオ入れ替え履歴などが相互にリンクした状態で検索可能となっている。2024 年 5 月には「estie 物流リサーチ」にて物流施設データの提供を開始し、本製品は大和ハウス工業や三菱商事都市開発といった物流不動産大手に活用されている。

当社は先述のような非公開情報を扱うだけでなく、ユーザーのニーズに応じて公的データと結びつけることを通じて価値を増大している。2024 年 4 月 1 日、国土交通省の発表した「不動産情報ライブラリ^{※4}」は、同 6 月 23 日 17 時時点でページビュー数が 670 万件を超える公的な不動産情報基盤である^{※5}。このライブラリには、周辺施設情報、ハザード情報、都市計画情報、価格情報、地形、人口といったデータが網羅され、API (Application Programming Interface) による連携も可能なシス

テムである。当社は顧客の物件取得の検討に際して活用される「estie 案件管理」と国土交通省の公表する国土数値情報を連携することで、不動産会社のデューデリジェンス業務の効率化を進めている^{※6}。

2024 年 7 月末現在、当社の扱う不動産データはオフィス、物流施設等と多岐にわたり、土地建物の住所やスペックといった情報だけではなく、所有権や賃料などの動的な情報を搭載した日本最大級のデータ基盤となっている。これらのデータを日次で更新するのは約 20 名のデータエンジニアが構築した巨大なデータパイプラインの仕組みである。次章以降では、不動産データを扱う上での課題に触れつつ、当社がどのような技術を用いて効率的かつ高度なデータ管理の自動化を実現しているのかについて触れたい。

3. 不動産データを扱う上での課題

不動産データを扱う上で最も難しいのが、同一性の判定である。一つの不動産には複数のオーナー、管理会社、仲介会社などが関わり、それらが自社独自の方法で情報の管理を行なっている。その中には手書きの情報や、略称などを含めて、同じ情報が異なるフォーマットで保存されているこ



図 3：estie 案件管理とオープンデータの連携リリース（estie 作成）

とが常であり、マスターデータを構築するための統一的なキーが存在しない。

この課題に対するアプローチとして最も有名なものが、国土交通省などが主導する「不動産 ID」の取り組みである。日本全国に所在する土地や建物に一意の ID を割り振ることで、オープンデータとクローズデータの結合が可能となり、不動産分野における DX が飛躍的に進展し、データに基づく効率的なビジネス展開・行政政策の実施が期待されている^{※7}。

また、不動産データを扱う上でのもう一つの課題は、非構造化データの取り扱いである。非構造化データとは事前に整形されずに元の形式のまま保存されているものであり、機械が読み取ったり分析をしたりすることが困難な情報である。

例えば、あるオフィスビルの管理会社が空室のリーシングを行っている場面を想定する。担当者が見込み顧客との打ち合わせでヒアリングした情報を、自然言語で CRM に入力する。それを別の担当者が閲覧しながら PM レポートを作成し、PDF 化して運用会社に送付を行うといった場面は日常茶

飯事である。これら「CRM 上のメモ」「PDF 形式の PM レポート」をリーシングに活かしたいところだが、構造化されていないため機械的に判別したり、示唆を導き出したりすることは困難である。現在は人間がその役割を担っているが、運用物件数が増えれば、同じ分析クオリティを維持することは不可能となる。

これら「同一性判定」および「非構造化データ」といった課題が複合的に絡み合う現実では、不動産データを扱う場合は人間がマニュアル作業をして判断することが唯一解となってしまふ。他の産業で起きているような、大規模データを機械的に統合し、AI に学習させることで新たな示唆を得るといった業務は、不動産事業者にとっては大変骨の折れる仕事となる。

このように、不動産業界の DX を阻害する要因として基盤データの整備というハードルが存在するが、当社はいかにしてその課題を解決し、顧客にデータ提供を行っているのか。次章においてその取り組みの一端を紹介したい。

不動産IDを起点とした情報連携がもたらす社会（多様なデータの連携推進）  国土交通省

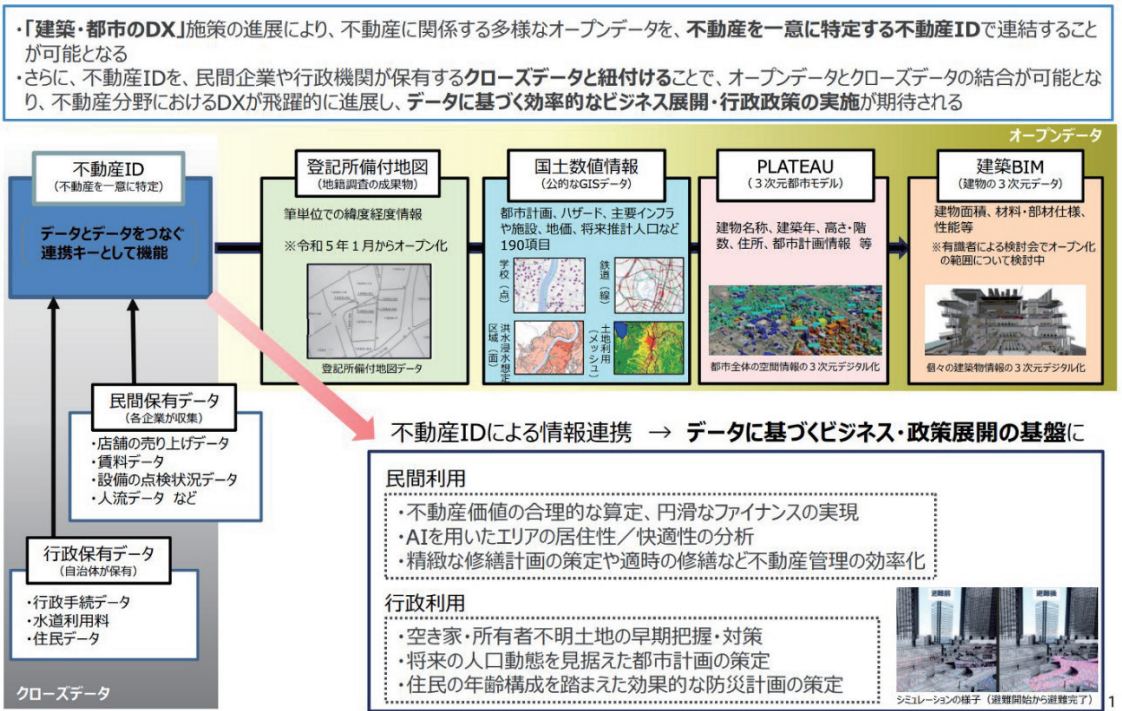


図4：国土交通省「不動産 ID を起点とした情報連携がもたらす社会」

4. estieのデータ構築技術

(1) 統合建物データ基盤

まず「同一性判定」について、様々な顧客業務やアセットタイプごとに複数のサービスを提供している当社の取り組みについて紹介する。例えば「estie マーケット調査」で提供しているオフィス賃料データを、「estie 所有者リサーチ」で活用したいと考えた際、簡単に連携することはできない。もし両プロダクトがそれぞれ建物データを構築していた場合、どのようにして同一性の判定を行うべきであるか。

例えば、「田中ビル」「タナカビル」「Tanaka Building」は全て同じ建物である可能性がある。同じように、「東京都 XX 区 YY 2-3-4」「東京都 XX 区 YY 二丁目-3-4」「東京都 XX 区 YY 2 丁目 3 番 4 号」も同じ建物かもしれない。これらはシンプルだが、現実には同じ住居表示にたくさんの異なる物件が所在していたり、同じ丁目に「田中ビル」が3棟存在していたりすることもよくある。これらを同一物件とそうではない物件として、機械的に判別することが不動産データ活用の一丁目一番地である。

当社では、様々なデータソースから自動連携される情報の同一性判定を毎日自動で行うことで、日本最大級の不動産データベースを日次更新することが可能となっている。その一つの成果が統合建物データ基盤の構築である。

物件名や住所だけではなく、その他様々な物件に付与された情報をキーとすることで、複数物件の中から同一のものを抽出して紐付けることを可能としている。またこれらは物件だけでなく、とある物件の中でフロアが空室募集に出れば、他データソースの空室データと照合することで漏れなくダブりのないマクロな募集状況がデータ化できる。この技術により、一部のポータル等で課題となっている重複募集を解決できる可能性がある。このように当社が開発した同一性判定ロジックによって作成した estie 独自の統合建物はいわば「estie 版不動産 ID」というべきものであり、当社の不動産データ基盤の技術的基礎となっている。

統合建物と外部データの紐付け精度は高く、J-REIT 物件を対象とした紐付け試験を行ったところ、誤った紐付けとなってしまったものはオフィス物件で0.1% (4/5,068 件)、住宅物件で0.4%

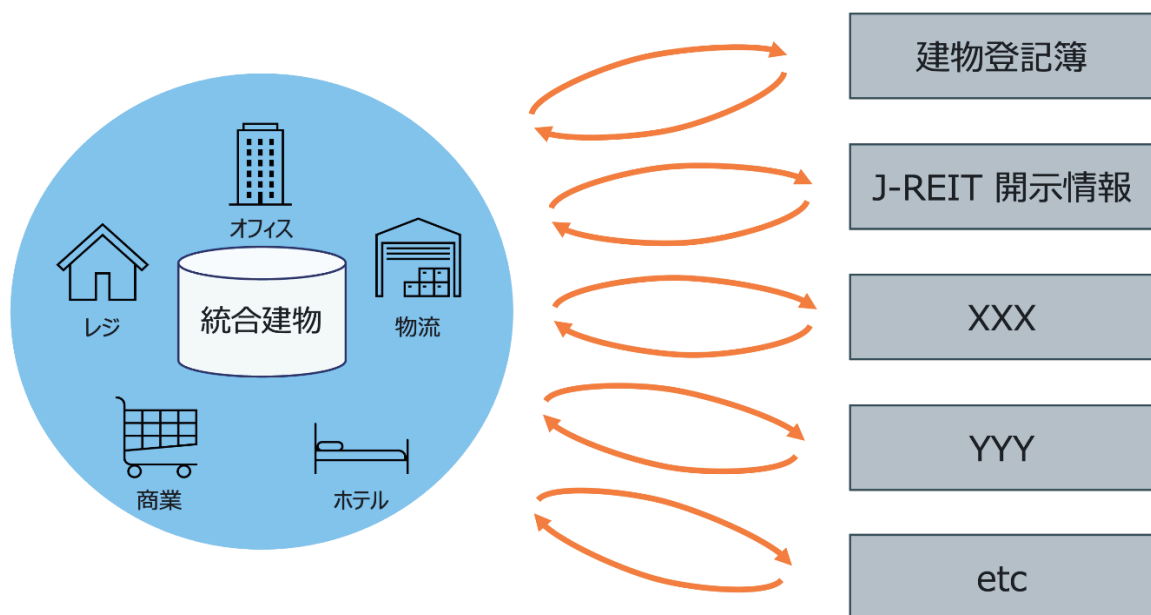


図5：統合建物データイメージ (estie 作成)

表 題 部 (主である建物の表示) 調製 余 白 不動産番号 9876543210987	
所在図番号 余 白	
所 在 A市B町一丁目 2番地 余 白	
家屋番号 2番の3 余 白	
1 種類 2 構造 3 床 面積 m ² 原因及びその日付〔登記の日付〕	
共同住宅 鉄骨造陸屋根地下1階付4 1階 123:45 令和3年2月1日新築	
車庫 階建 2階 123:45	
店舗 3階 123:45	
	4階 123:45
	地下1階 123:45
権 利 部 (甲区) (所有権に関する事項)	
順位番号 登記の目的 受付年月日・受付番号 権利者その他の事項	
1 所有権保存 令和3年3月3日 所有者 東京都A区B町一丁目1番1号	
	第12345号 株式会社 A 不動産
2 所有権移転 令和4年5月6日 原因 令和4年4月31日売買	
	第23456号 所有者
	東京都C区D町二丁目3番Eビル405号室
	B 建物株式会社

図6：登記簿イメージ (estie作成)

The screenshot shows a search interface with a map on the left and a list of results on the right. The map displays various buildings labeled 'サンプルXビル' (Sample X Building) in different colors. The list on the right contains the following data:

建物名/住所	名称	所有者	割合
サンプルAビル 東京都千代田区一番町XX-X	サンプル株式会社	サンプル株式会社	100%
サンプルBビル 東京都千代田区一番町XX-X	サンプル商事	サンプル商事	-
サンプルCビル 東京都千代田区神田XX-X	株式会社サンプル製作所 sample株式会社	株式会社サンプル製作所 sample株式会社	65.42% 34.58%
サンプルDビル 東京都千代田区神田X-XX	サンプル信託銀行	サンプル信託銀行	100%
サンプルEビル 東京都千代田区神田X-X	株式会社サンプルジャパン	株式会社サンプルジャパン	100%
サンプルFビル 東京都千代田区若本町X-X	サンプルコーポレート株式会社	サンプルコーポレート株式会社	100%
サンプルGビル 東京都千代田区若本町X-X	estie 花子 様 estie 一郎 様	estie 花子 様 estie 一郎 様	50% 50%
サンプルHビル 東京都千代田区船場町XX-X	サンプルコンサルティング 様 サンプル附図 様	サンプルコンサルティング 様 サンプル附図 様	4.32% 95.68%

図7：estie 所有者リサーチ (estie作成)

(8/1,947件)に過ぎなかった。それぞれ94.9%、93.4%は完全自動で正しい紐付けが行われ、残りは自動でいくつかの選択肢が提示され、目検で同一判定を行うことができた。

(2) 非構造化データの自動構造化

次に「非構造化データ」を自動で構造化する当社技術の一例を紹介したい。不動産業界において、手書きのメモやPDF化された物件概要書など、非

構造化データが巷に溢れていることは先述の通りである。中でも多くの事業者にとって馴染み深いのが登記情報ではないだろうか。

不動産登記は広く一般に請求・取得ができる不動産情報の宝庫である。しかしPDF形式で提供されるそれは、個別物件についての情報として価値があるものであったとしても、構造化されたデータとして検索可能な形で提供したり、他のデータと紐づけたりするには、多くの場合人手を介したデータ打ち込みなどが必要不可欠である。

当社は2023年10月に取得した特許技術によりこの問題を解決している。登記情報を機械が読み取り、estieのマスターデータベースに自動で紐付け、「estie所有者リサーチ」を契約するユーザーが物件軸で所有権異動履歴を検索するだけでなく、オーナー軸で物件の売買履歴を検索するなどの逆引きが可能となっている。

人間が登記簿を閲覧しながら手作業で情報を入力する手間は既に不要となり、estieの開発した技術によって自動の紐付けを行うことが可能である。一部の顧客企業では、過去に取得したまま死蔵する登記簿を大規模にデータベース化しなおすために当社技術を活用するなど、まさに社内に眠るデータ油田を掘り起こす活動が可能となっている。

(3) 不動産データにおける生成AIの活用

最後に紹介する取り組みが、不動産データ抽出にあたっての生成AIの活用である。例えば、当社

が行った「大規模言語モデルを用いた物件概要書（マイソク）PDFからの情報抽出」という研究では、不動産業界で流通している物件情報が記載されたPDFから情報を効率的に抽出するために、大規模言語モデル（LLM）とOCR（光学文字認識）を組み合わせた手法やマルチモーダルLLMによるEnd-to-Endの情報抽出手法を検討し、その効果の測定および評価を実施した。

実際の物件概要書PDFを用いて評価を行った結果、OCR+LLMの手法は全て手作業の場合より約65%の時間を削減することができた。本研究は、不動産情報の抽出作業の効率化と精度向上に大きく貢献する手法を提案しており、実際の業務に適用可能であることを示している^{※8}。

5. おわりに

本稿では、不動産業界のデジタル化にあたって必要不可欠となるデータ構築技術について、株式会社estieにて実際に開発し、実用化している手法について紹介を行った。ハードウェアとしての土地建物を適切にデジタルデータとして扱うことは、日本の都市・地域の進化、ひいては不動産業界の更なる発展のために避けては通れない。

三井不動産、三菱地所などがスポンサーとなり2024年4月7日に当社が主催した「不動産DXカンファレンス2024」において、多くの不動産事業者や不動産テックが業界の革新に向けて新たな取り組みに着手していることが確認できた。一方で、

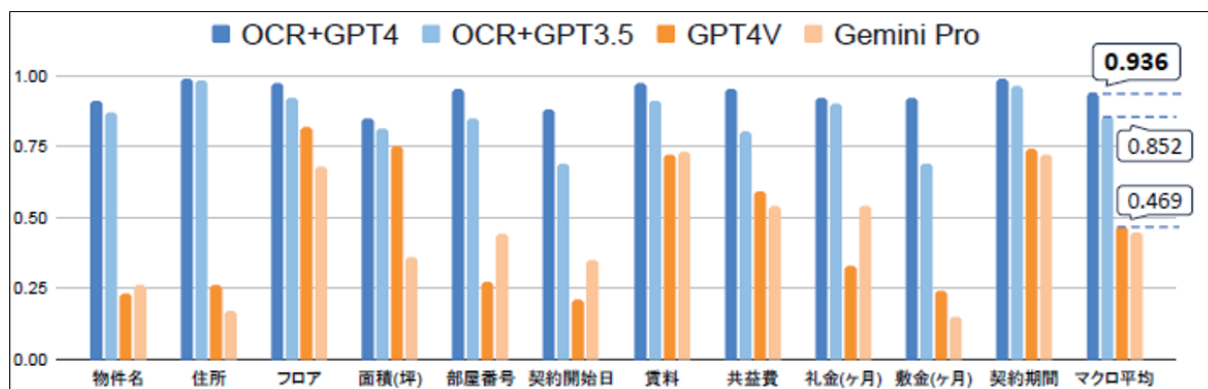


図8：抽出手法の違いによる抽出情報の精度検証結果（estie作成）

全てが順風満帆ではなく多くの課題が残されていることも明らかになった。今後も、行政、産業界、アカデミアが一体となって不動産業界の課題解決に取り組み、より一層のデータ活用を通じたデジタルトランスフォーメーションが実現されるよう、当社としても技術開発を進めていきたい。

参考文献

※¹ <https://www.estie.jp/products/research/>

※² <https://www.estie.jp/products/logistics-research/>

※³ <https://prtimes.jp/main/html/rd/p/000000082.00040115.html>

※⁴ 国土交通省, 不動産情報ライブラリ, <https://www.reinfolib.mlit.go.jp/>

※⁵ 国土交通省, 不動産IDを起点とした情報連携がもたらす社会, https://www.mlit.go.jp/tochi_fudousan_kensetsugyo/content/001750674.pdf

※⁶ <https://prtimes.jp/main/html/rd/p/000000094.00040115.html>

※⁷ 国土交通省, 不動産IDの2024年度の検討内容について, https://www.mlit.go.jp/tochi_fudousan_kensetsugyo/content/001750673.pdf

※⁸ estie, 大規模言語モデルを用いたマイソク PDF からの情報抽出, https://www.anlp.jp/proceedings/annual_meeting/2024/pdf_dir/P10-7.pdf