

データエコシステム構築論：情報システム開発の変革に向けたアプローチ

Conceptualizing Data Ecosystem Construct: An Approach to Transforming Information System Development

宇田川 佳久
Yoshihisa Udagawa

東京情報大学 総合情報学部
Faculty of Informatics, Tokyo University of Information Sciences

1. まえがき

大規模データの分析が可能になった現在、データ分析は企業経営に不可欠なものとなっている。従来のデータ分析システムは、開発と運用が明確に切分けられ、運用開始後の機能変更を前提としていなかった。これは、大規模業務システムがウォーターフォール・モデルに基づいて開発されていることが影響しているものと考えられる。技術的にも、定型データを対象としたデータベースを中心に業務ソフトウェアを開発することから、頻繁な変更に対応することは前提としていなかった。

一方、デジタルデータの蓄積が質量共に拡大し、加えて、分データの流通や運用環境が著しく変化している。競争力のある企業経営を維持するために、迅速なデータ分析機能の改良開発のニーズも増大している。

従来の業務システム開発・運用サイクルが、年単位で実施されるものであったのに対し、ここ数年、データエコシステムという呼称のもと、従来よりも短い期間での業務機能の改良に対応する開発モデルが提案されて来た[1]-[6]。本文では、これまでの研究論文に基づき、データエコシステムの定義、構成、課題について述べる。

2. データエコシステムの定義

データ分析の結果は、事実を主観的に反映することから、政策・経営判断への貢献度を増している。分析対象とするデータは、クラウドシステムや通信技術の進展により、一組織や一企業で一括管理できるレベルを超え、組織・企業間で共有する環境が整ってきた。このような状況下で、データエコシステムは、自然界の生態系をデータ分析の形態に当てはめた言葉である。データエコシステムの趣旨は、データ分析に係わるすべての利害関係者を考慮したシステムを構築することである。それにより、永続・安定したシステムを構築することを目的としている。

データエコシステムは、このような背景から検討されているデータ分析システムの総称であり、現時点で規格化された定義はないが、以下の定義が一般的と考える[1]-[5]。

データエコシステムは、データ蓄積基盤、分析手法、および、分析した結果を扱うアプリケーションの集まりである。データエコシステムは、顧客分析、商品の価格付けなどのデータ分析により企業の意志決定を支援する。

データ分析機能は、ユーザからの要求に応じて機能強化されて成長してゆく。その成長過程では、より多くのデータを取り込み、分析手法が考案され、最終的には、企業収益に貢献する。この過程は、「環境」あるいは「基盤」という言葉が使われている[4]。

加えて、利用者や経営に視点を置いた次の定義もある[6]。

企業内外データを活用し、あらたなビジネスモデル／収益モデルを創出すべく形成するステークホルダーの集合体

表現は異なるが、総じてデータエコシステムの定義や概念は、社会的な認知に耐えうる段階にあると考える。

3. データエコシステムの構成

データエコシステムの構成については、既存のデータ分析システムの構成を踏襲しているため、概ね、図1に示す構成となっている[3]。

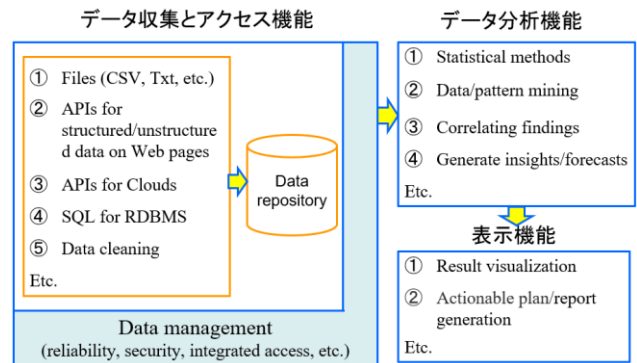


図1 データエコシステムの構成

3.1 データ収集とアクセス機能

多様なデータ供給源からさまざまなデータフォーマットに対応したデータを取得するための機能である。ファイルによるデータの取得は、大量のデータをまとめて取得することができること、CSV形式やText形式を扱うアプリケーションが多いという特徴がある。Webページには自然言語や画像といった非構造化(unstructured)データから構成されているものも多い。画像、音声、自然言語文などの非構造化データから特徴を抽出するといったAPIも普及している。構造化されたデータは、関係データベースに格納することができる。多くの企業・組織で、国際標準言語SQLでアクセスできる環境が運用されている。

3.2 データ分析機能

利用目的に応じたデータ解析を行うために、統計的解析手法、パターンマイニング、相関関係の抽出などの機能を提供する。この機能の出力は、経営的な観点から、診断、因果関係の説明、予測、対処方法などに分類される。

3.3 表示機能

データ分析結果を視覚的に表示する機能で、多くの市販パッケージが利用可能である。個別の視覚化が必要な場合は、独自のアルゴリズムで実装する。

3.4 プロセス構成

Stobierski は、データ収集から加工・分析の観点からデータエコシステムのプロセス構成について論じている[5]。図2に示すように5プロセスから構成されていて、生データをIoT機器などから収集し(Sensing, Collection), 続く解析処理で活用できるよう特徴を抽出する(Wrangling)ことで非構造化データを構造化する。データ分析(Analysis)では、利用者や経営者からの要件に応じて、診断、因果関係の説明、予測、対処方法などを導出する。これら一連のプロセスで使われたデータは、クラウドなどに記憶する。



図2 データエコシステムのプロセス[5]

このプロセスモデルでは、プログラミング、API、ツール、クラウド基盤、アルゴリズムを使って、新しい要件や分析を柔軟に実施することを特徴としている。

4. データエコシステムの課題

データエコシステムの課題については、経営的な観点と技術的な観点があると指摘されている[7]。

4.1 経営的観点からの課題

- I. データの提供者と利用者で、Win-Win の関係を構築できるか。
 - ① 現状では実施例が少なく、多くの事例を集める必要がある。
 - ② データエコシステムにする理解、効果、構成の合意が十分ではない。
- II. ビジネスの効率化に貢献できる分析結果や提案をデータエコシステムから導出できるか。
 - ① 組織内外のデータを組み合わせて組織化する方法論の確立が待たれる。
 - ② 非構造化データから構造化データを生成する機能(Wrangling)が未成熟である。

4.2 技術的な観点からの課題

- I. データエコシステムの開発方法論の検討
 - ① 従来のウォーターフォール・モデルに基づく情報システム開発方法論が適用できるか。できない場合は、どのような方法論が考えられるか。
 - ② データエコシステムの開発者が備えるべき技量と教育方法に関する検討が待たれる。
- II. 提供者が異なるデータの活用に関する課題
 - ① 提供されるデータの価値・対価の評価方法の検討が待たれる。
 - ② 偽データを発信した場合の責任に関する取り決めが必要である。
 - ③ データエコシステムの利害関係者の組織運営・費用分担に関する検討が待たれる。

5. データエコシステムの開発方法論の検討

データエコシステムは、オープンデータや第3者組織が保有するデータなどを対象とすること、データから業務の拡大につながる結果を短い周期で求めることが、従来の情

報システムとの大きな違いと考える（攻めのデータ活用）。一方で、クラウドに代表されるデータ流通の基盤やデータ分析方法が進展してきたことから、数か月といった短い周期でのリリースに対応できるような開発が可能になっている。

図3は、従来の情報システム開発を前提とした場合の開発工程であり、主要なシステム開発を完了した後に、業務アプリケーションの適合化開発を行うものである。一般に、適合化開発は、主要なシステム開発に比べて小規模である。ベンダーとしては、要員の確保、負荷の平準化といった課題を内在している。

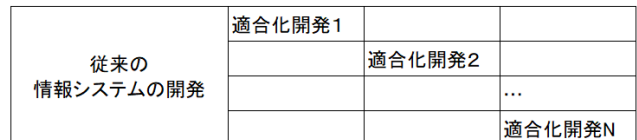


図3 データエコシステム開発工程案1

図4は、データ管理基盤の開発に続いて、業務アプリケーションを開発するモデルである。データ管理基盤は、DBMS、クラウドAPIなどを活用して構築する。適合化開発で業務アプリケーションを整備するモデルであり、ベンダーにとっては、図3のモデルより多くの開発要員を投入することから、対応しやすいと考える。

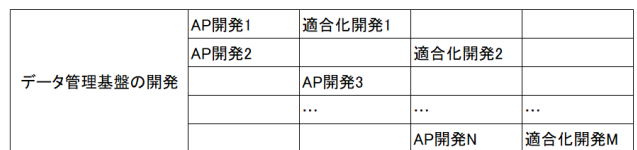


図4 データエコシステム開発工程案2

6. おわりに

ビジネス環境の変化に追随した情報システムの開発に着目した「データエコシステム」について調査結果を述べた。今後は、開発に関わる要員の種別[8]と保有すべき技量を考慮した開発方法論の検討を予定している。

参考文献

- [1] What is a data ecosystem, 2021, <https://mixpanel.com/topics/what-is-a-data-ecosystem/>
- [2] Data Ecosystem, 2021, <https://www.cognizant.com/glossary/data-ecosystem>
- [3] Modern Data Ecosystem: Introduction to Data Analytics, <https://www.coursera.org/lecture/introduction-to-data-analytics/modern-data-ecosystem-OJQ9T>
- [4] D. Stodder, Recommendations for a Modern Data Ecosystem, 2017, <https://tdwi.org/articles/2017/04/11/4-recommendations-for-a-modern-data-ecosystem.aspx>
- [5] T. Stobierski, Key Elements of Data Ecosystem, Harvard Business School Online, <https://online.hbs.edu/blog/post/data-ecosystem/>
- [6] Teradata, ビジネス「生態系」データエコシステムが新しい収益構造を導く, 2020年5月, <https://www.teradata-jp.com/post/202005-ecosystem>
- [7] Y. Udagawa, A. Walek, and P. Dini, Panels: The Eco-System of Data and Software, Apr. 2021, <https://www.iaria.org/conferences2021/ProgramALLDATA21.html>
- [8] Key Players in the Data Ecosystem, <https://www.coursera.org/lecture/introduction-to-data-analytics/key-players-in-the-data-ecosystem-ZyoZ9>