

組織的 RDM 支援サービスのためのエンタープライズアーキテクチャの検討

青木学聡¹

概要: 研究データ管理(Research Data Management, RDM)は、研究の開始から終了までを通じ、どのような研究データをどのように扱うかを定め、これを実践することである。この中には、研究データの入手、保存、共有、公開、破棄といった、研究データに対するあらゆる操作が含まれるが、多くの場合、これらは研究者毎の裁量下において選択されている。

この一方、今日のオープンサイエンスの浸透、学術機関におけるガバナンス強化を背景とし、組織単位での RDM 手法の統一、共通化が求められている。組織単位での RDM 支援サービスの導入は、組織と研究者の意向、費用対効果等、様々な対立軸が存在し、これらのギャップを明確にしながら、組織と研究者の合意点を探る必要がある。

本稿では、組織が RDM 支援サービスを導入する場合に想定するビジネスモデルの考え方、またこれを実現するためのエンタープライズアーキテクチャの構成方法について検討する。

キーワード: 研究データマネジメント、ビジネスモデル、エンタープライズアーキテクチャ

Enterprise Architecture Approach for Institutionalized Research Data Management Service

TAKA AKI AOKI¹

Abstract: Research data management (RDM) is the practice of defining what research data is to be handled and how it is to be handled from the beginning to the end of a study. This includes all operations on research data, such as obtaining, storing, sharing, publishing, and destroying, which are often chosen at the discretion of each researcher.

On the other hand, with the spread of open science and the strengthening of governance in academic institutions today, there is a need to unify and standardize RDM methods at the institutional level. The introduction of the institutional-wide RDM support services may include various conflicts, such as the intentions of the organization and researchers, cost effectiveness, etc. It is necessary to clarify these gaps and find a point of agreement between the institution and the researchers.

In this article, we will discuss the concept of the business model assumed when an organization introduces RDM support service, and how to compose an enterprise architecture to realize it.

Keywords: research data management, business model, enterprise architecture

1. はじめに

研究データマネジメント(Research Data Management, RDM)とは、研究の開始から終了までを通して、どのようなデータを利用・生成・収集するか、またこれらのデータをどのように解析・保存・共有・公開するか、等を定め、これを実践する一連の行為である。あらゆる情報がデジタル化され流通する、「デジタル・ファースト」時代においては、RDM も例外ではなく、研究データをデジタル情報として、安全かつ効率的に取り扱うための情報サービス、そしてこれを実現するための情報システムの整備が不可欠である。特に、研究データの公開と相互利用性の拡大は、2021年3月に公開された第6期科学技術・イノベーション基本計画で取り上げられるキーワード「オープンサイエンス」、「Society 5.0」を支えるコンセプトとして、これからの学術研究における重要課題とされている[1]。

デジタル化の程度に関わらず、研究者はデータの収集・

整理・解析・共有等の一連の RDM に対し、RDM という言葉を知らないながらも個別に実践している。一方、大学等学術機関の執行部は機関の基本方針や戦略を基に、研究者と研究データ、並びに支援組織の関係を明確化し、組織的な RDM サービスを構築する必要に迫られている。しかしながら、組織的 RDM サービスの導入は、

- 多様な研究データの取り扱い
- 個別の研究者の慣習を変更することへの抵抗
- サービス間相互連携の拡充
- サービス継続性の確保

等の点において、提供者・利用者との間に大きな隔たりが存在する。組織的 RDM サービスの成功には、このギャップの詳細な分析と段階的な解消手段の提示が必要とされる。本稿では、組織的情報戦略・サービスのモデリング手法である、エンタープライズアーキテクチャ(EA)の技法を援用し、研究者が日常的に実施している、研究データの収集、

¹ 名古屋大学
Nagoya University

保存、加工、公開等の RDM 業務の分類とその動機の明確化を試みる。そのうえで、組織的 RDM サービスが整備すべき事項を、トップダウン、ボトムアップそれぞれの視点から作成した EA モデル図の対照により検討する。

2. アプローチ手法の検討

2.1 エンタープライズアーキテクチャ概念の導入

組織としての情報戦略、サービス、システム、コンポーネントを総括的に扱う手法として、エンタープライズアーキテクチャ(Enterprise Architecture, EA)がある。The Open Group が提唱する EA 記述手法は、TOGAF(The Open Group Architecture Framework)と呼ばれ[2]、世界的に広く利用されている。TOGAF では、一貫した戦略に基づき、提供すべき情報サービスを記述する「ビジネスアーキテクチャ」、情報サービスを実現するための「情報システム・アプリケーションアーキテクチャ」、さらにこれらの具体的実装である「テクノロジアーキテクチャ」という形で EA 要素を分類、整理する。多くの場合、経営陣は戦略と情報サービスが主な検討対象であることに對し、情報基盤部門、協力ベンダは、情報システムやそれらの具体的な実現方法が興味の対象である。TOGAF の階層化されたアーキテクチャ記述は、興味の分離を明確にする一方、それぞれのアーキテクチャ記述間の相互作用を考慮する必要がある。

2.2 バリューストリーム

ビジネスアーキテクチャの構成は、いわゆるトップダウンアプローチといえる。ここでは、組織がなぜそのような情報サービスが必要とするか、またその実現のために必要な機能を整理することが求められる。バリューストリーム(Value Stream)はビジネスアーキテクチャを構成するモデリング手法の一つとして位置づけられる[3]。ここでいう、「バリュー」とは単なる金銭的価値だけではなく、「～が便利になる」「～が実現する」といった、ステークホルダに対する利便の総称とされる。

バリューストリームは、バリューが発生する要件を段階(バリューストリームステージ, Value Stream Stage)に分割し、記述する。それぞれのステージには、

- 当該ステージのタイトルと概要
- 当該ステージに対する関係者(ステークホルダ)
- ステージの発生、終了要件
- 発生する価値

が記述される。これらが相互につながることで、バリューストリーム全体の価値とその実現に必要な要件が具体化される。

2.3 バリューストリームによる研究データライフサイクルの記述

研究データのライフサイクルは、RDM の説明では欠かせない要素であり、様々な形式で図式化されている[4]。これらをバリューストリームの形式により再定義を試みる。バリューストリームの記述では、ステージを進めるための要件、及びこのための関係者の記述に主眼を置くこととなる。「新たな研究データを生成し、これを公開する」ことをゴールとした、バリューストリームステージをまとめると以下ようになる。いずれのステージにおいても、主語は「研究者」、すなわち研究活動は研究者(または研究チーム)の主体的活動であることを改めて意識する必要がある。また、「開始」、「終了」条件は研究者の需要とそれが満たされること、「価値」は各ステージでは研究者の需要を満たすため実現されること、と解釈できる。

1. 入手方法の検討
開始条件: 研究者が学術的興味に基づきデータの入手・生成方法を検討する
終了条件: データ入手・生成方法が決定する
発生する価値: 具体化されたデータ入手・生成方法
2. 生成・入手
開始条件: 研究者が研究目的、経済的合理性等に基づきデータの入手を決定する
終了条件: データを研究者が利用できるようになる
発生する価値: 利用できるデータの存在
3. 整理・保存
開始条件: 研究者が指定したルールにより、データの整理・分類を必要とする
終了条件: データ整理され、有効に利用できる状態と

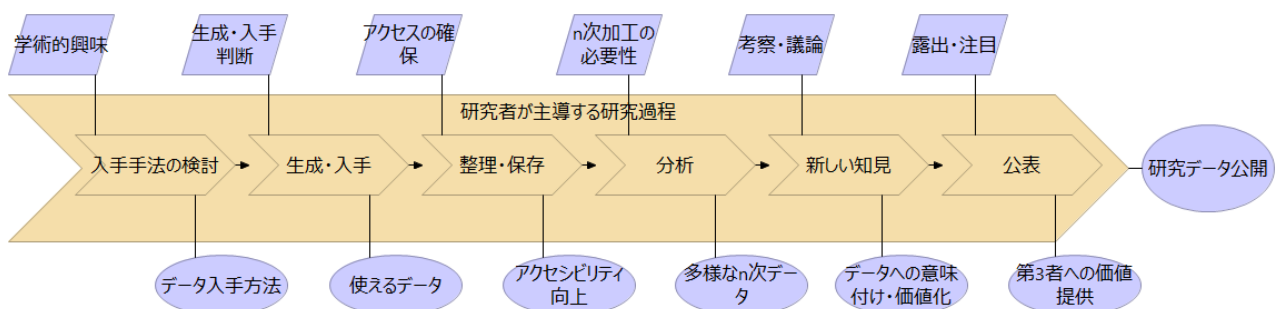


図 1: 「研究データライフサイクル」のバリューストリーム形式による表現

なる

発生する価値: 整理されたデータセット, と保存されるメディア

4. 分析

開始条件: 1次データを元に, それらの加工, 複数データ間の比較を必要とする

終了条件: 新しい知見につながる2次データの生成

発生する価値: 多様な2次データの生成

5. 新しい知見

開始条件: 1次データ, 2次データから新しい知見が見いだされる

終了条件: データを基に論文など, 新しい知見が取りまとめられる

発生する価値: データを解釈する新しい知見

6. 公表

開始条件: 第3者へのデータ提供が求められる

終了条件: データが第3者より参照可能になる

発生する価値: 第3者からの参照による新たなデータ利用, 研究者自身のプレゼンス

このようなバリューチェーンの諸要素を Archimate により図示した結果を図1に示す。Archimate には version 3.1 より Value Stream 要素が追加されたことから, EA 設計においてバリューチェーン図を記載することが容易となっている[5]。図1では, 上記における「開始条件」を「requirement」要素に, 「終了条件」及び「発生する価値」を「value」要素として関連付けている。

なお, 図1の例では, 「研究データの公開」を終点としているが, より多くの研究データ利用者に対し積極的にサービス展開をすることも考えられる。その場合は, 「第3者の研究データ利用者」を主語として, 以下のバリューチェーン

ームステージが新たに構成されることとなる。

1. (第3者利用者が)入手方法を検討

開始条件: 第3者利用者が学術的興味に基づき, データの入手を検討する

終了条件: 第3者利用者が(自身が提供する)研究データを見つけ出す

発生する価値: 研究データの所在をアピールし, 第3者利用者がこれを検討候補に入れる

2. (第3者利用者が)データ入手

開始条件: 第3者利用者がデータへのアクセスまたは入手を検討する

終了条件: データが利用可能になる

発生する価値: 利用に関するライセンス提示, アクセスコントロール等, スムーズなデータ利用

2.4 ケイパビリティマッピング

バリューチェーンの作成の目的は, ある一連の行為が生み出す価値, 及びそれらの行為の開始, 終了条件といった, 動機付け(motivation)の分類・整理することにある。これらの行為を実施するために必要な機能を結びつける必要がある。この機能は, ケイパビリティ(Capability)と呼ばれ, 単に情報システムに限らず, 事業実施に必要とされる要素すべてが該当する。

図1で挙げたそれぞれのバリューチェーンステージにおいて必要とされる代表的なケイパビリティをまとめると以下ようになる。

1. 入手方法の検討

- 既存の研究データの所在を検索サービス等により見つけられる
- 研究データの生成に必要な, 観測・実験機器やソフ

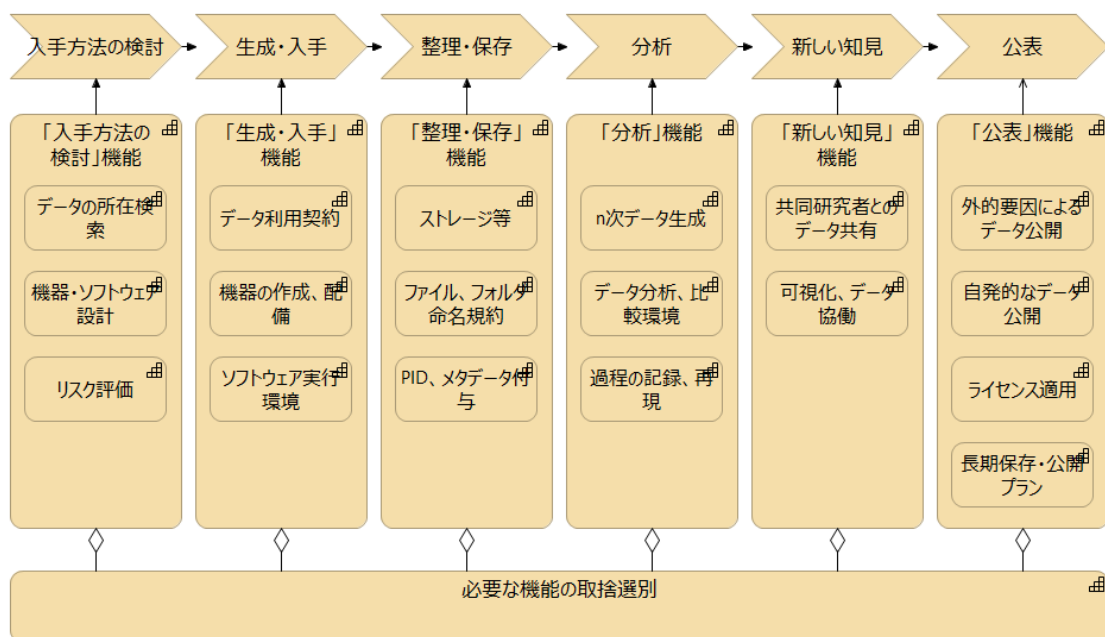


図2: 「研究データライフサイクル」のバリューチェーン-ケイパビリティマッピングの例

トウェアを設計できる

- データの生成・入手に際し、倫理・法的リスクを評価できる
2. 生成・入手
 - 種々の契約、手続きを実施し、データ提供を受けることができる
 - 観測・実験機器を作成、配備しデータを収拾できる
 - ソフトウェア実行環境を整備し、データを生成できる
 3. 整理・保存
 - 研究データへのアクセシビリティを確保するストレージ、ネットワーク、データベースを整備できる
 - ファイル・フォルダに適切な命名規約を導入できる
 - 永続的識別子(Persistent Identifier, PID)を含む、適切なメタデータ記述を導入できる
 4. 分析
 - n 次データを生成するプログラム実行環境を整備できる
 - データを分析、比較できる環境を整備できる
 - 分析過程を記録・再現できる環境を整備できる
 5. 新しい知見
 - 共同研究者等と研究データを安全に共有できる
 - 可視化等、データを中心とした含む多面的な協働を実現できる
 6. 公表
 - 出版社等、外的な要求に応える形式で、研究データを公開できる
 - 自発的に研究データを公開できる
 - 研究データ公開時に、適切な利用条件を設定できる
 - 長期保存、長期公開のための計画ができる

バリューストリームとケイパビリティの対応付け(バリューストリーム-ケイパビリティマッピング)を図示した結果を図2に示す。

図では、それぞれの機能の取捨選択を行うメタ的な機能を追記している。バリューストリーム、ケイパビリティのマッピング過程を詳細に考察することで、バリューストリージステージ及びケイパビリティ技術の粒度をそろえる作業を進めることとなる。

2.5 研究者と組織間のギャップ分析にむけた考察

ケイパビリティは、原則として「組織が持つべき機能」である。しかしながら、今回の研究データライフサイクルに関しては「既に研究者がほぼすべての機能を実現している」ことを前提とし、モデリングを実施している。したがって、組織的 RDM サービス環境の議論においては、「どの機能を共通化することで効果が期待できるか」の課題に取り組むこととなる。ギャップ分析においては、「組織が想定する情報システム・サービス」というトップダウン志向と、「研究者が今実施している各種 RDM の実態把握」というボトムアップ志向との突合せを行うことになる。

図3は、研究データ管理プロセスを、研究者視点に基づき、

- 研究の実施
- 研究成果の保存と公開
- 論文、研究データの公開
- 研究公正対応のためのデータ保管

の分類で Archimate により整理したものである。これは、オープンサイエンス、オープンデータ活動とともに、研究データに対する研究公正強化対策を取り入れて事前で作成されたものである。さらに、研究者が選択的に利用する情

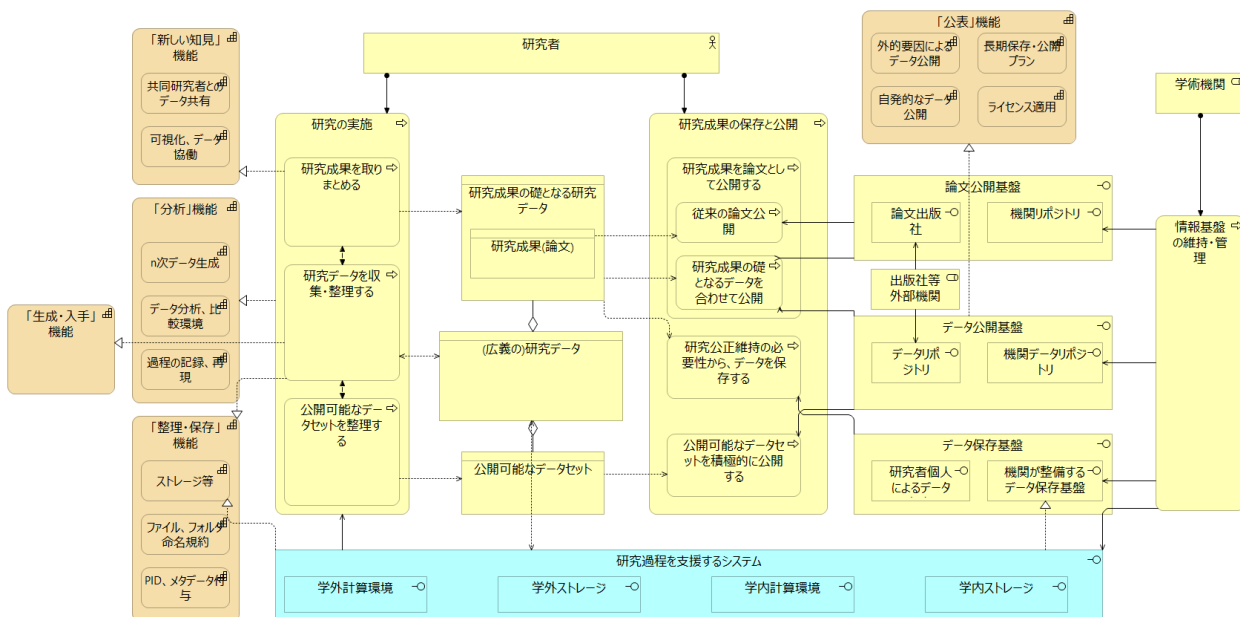


図3: 研究者視点から見た研究データの収集、保存、分析、公開の EA モデルに図2のケイパビリティ対応関係を当てはめた例

報システムとして、学内外のストレージシステム、及び分析・シミュレーション基盤の利用を簡易的に掲載している。

このモデル図に対し、図2で提示したケイパビリティ要素の対応付けを試みたが、図示される通り簡潔な対応とならないことが明確に理解される。図1, 2で検討した、オープンサイエンスを志向した一貫した研究データ管理と、現場で実際に行われている、日常的なRDM作業、特に研究データの分析と、研究公正のための保管に対する要件の理解が不十分であることが想像される。この一方で、研究データの公開に携わるための、機関リポジトリ、学外のリポジトリについては、明確な対応付けが可能なことも示唆されている。

3. まとめ

EAによる情報戦略、サービス、システムの記述技法や、有力な手法と考えられるが、トップダウンと、ボトムアップそれぞれのモデリング方針を拡大する際の、前提条件のすり合わせ、といったより基本的な合意事項、設計ルールの共通が必要であることが、今回のモデリング試行により明確となった。

2021年現在、RDMに関わる諸事は研究者、特にPIの裁量により決定され、運営されている部分が多い。これに対し、研究において扱うデータ量の拡大、研究ドメインや資金配分機関による、各種データ取り扱いルールの明確が進んでいることから、もはや研究者個人で情報システムレベルからすべてのRDMツールをデザインすることは困難になっており、大学等学術研究機関が漸進的とはいえ、RDM支援サービスを充実させる必要がある。

組織が取り組むべきRDM支援サービスは断片的ではあるが、検討が進められている。例えば、オープンアクセスリポジトリ推進協会は、2017年の日本語での本格的なRDM教材発表以来、図書館のオープンアクセス機能の教科を推進している。また2021年にはRDM人材のためのスキルセット作成プランを発表し、組織におけるRDM機能の整理を進めている[6]。一方、大学ICT推進協議会は、2019年に「学術機関における研究データ管理に関する提言」を

発表し、大学における組織的RDM環境整備の必要性と、サービス内容の概要を提案している[7]。

日本国内でDXの掛け声が叫ばれているものの、特に大学における情報基盤整備は、予算的な課題もあり戦略的な計画を立て、遂行することが困難になっている。上記の活動とEAによる組織、情報サービスデザイン技法を連携することで、組織的RDMにおける効率的な整備プランの立案に寄与できるよう、組織的RDMのためのEAモデリング技法の検討を継続する必要がある。

謝辞

本研究は、科研費基盤研究(A)「多様な学術研究活動を育むアカデミックデータ・イノベーション成熟度モデルの開発」(20H00099)、及び、2021年度国立情報学研究所公募型共同研究「組織的な研究データマネジメントサービス導入プロセスの検討」の支援の下、実施されている。

参考文献

- [1] 内閣府, “第6期科学記述・イノベーション基本計画”, (2021)
<https://www8.cao.go.jp/cstp/kihonkeikaku/index6.html> (参照 2021-08-09)
- [2] The Open Group, “The TOGAF® Standard, Version 9.2” (2018) ISBN:1-947754-11-9,
<https://pubs.opengroup.org/architecture/togaf9-doc/arch/index.html>
- [3] The Open Group, “TOGAF® Series Guide: Value Streams” (2017) ISBN:1-947754-02-7, G178,
<https://publications.opengroup.org/guides/togaf/togaf-series-guides/t180>
- [4] 吉田 幸苗, 天野 絵里子, 松本 侑子, 西菌 由依, 山地 一禎, 南山 泰之, 尾城 孝一, 常川 真央, 大園 隼彦, “教材「研究データ管理サービスの設計と実践」第2版”(2021)
<https://jpcoar.repo.nii.ac.jp/records/607>
- [5] The Open Group, “Archmate® 3.1 Specification” (2019) ISBN:1-947754-30-0, <https://publications.opengroup.org/c197>
- [6] 國本 千裕, “研究データ管理支援者に求められるスキル: 「標準スキル (Ver.0.0)」の内容”, NII学術情報基盤オープンフォーラム2021 (2021/7/8, オンライン)
<https://opac.ll.chiba-u.jp/da/curator/900119527/>
- [7] 大学ICT推進協議会, “学術機関における研究データ管理に関する提言”(2019) <https://rdm.axies.jp/sig/57/> (参照 2021-08-09)