

ボーダレスなデジタルデータの柔軟な利活用に向けた オープンネットワークアプローチ

金子 晋丈^{1,a)}

概要: デジタルデータは一日に 2.5 エクサバイト新たに生成されていると言われている。筆者は、増加し続けていくデジタルデータのボーダレスな利活用に向け、アーカイブ基盤の確立、デジタルデータの仮想化、およびデジタルデータのネットワーク化を中心技術とし、アプリケーションサービスの構築を行いながら、技術開発を行ってきた。本報告では、ボーダレスなデジタルデータの柔軟な利活用を可能にする自律型オープンデータ流通基盤のアプローチ、およびその主たる機能である Virtual File と Catalogue System、これらを基に構築した人文科学的アプリケーションサービスについて紹介する。

Open network approach toward flexible utilization of borderless digital data

1. はじめに

理工系分野に限らず人文科学研究分野においても、研究や調査にデジタルデータが無くてはならない存在になって久しい。このような背景のもと、さまざまな人文科学研究資料がデジタル化され、蓄積されてきている。しかし、蓄積されたデジタル資料（以下、デジタルデータ）が十分に活用されている状況にあるとは言いがたい。具体的には、デジタルデータを閲覧したい利用者が十分な背景知識を有した上で特定のデジタルデータを探索し使用する場合を除いて、利用者は膨大なデジタルデータから自分の興味の範疇にあるデジタルデータに出会い、閲覧し、知的刺激を高め、さらに興味を深めたり広めたりし、あらたなデジタルデータとの出会いに進むことができていない。

蓄積されたデジタルデータの活用の課題は、人文科学研究分野に限定されない。現在、デジタルデータは日々 2.5×10^{18} バイト生成されており、2020 年には世界中の各人が 1.7MB/s でデータを生成すると推定されている。すなわち、これらのデジタルデータと増え続けるデジタルデータ利用者の出会いをどのようなメカニズムで実現するかは、人文科学研究分野にとどまらないデジタル社会にお

ける喫緊の課題となっている。

これまでのデジタルデータの活用手法の多くは、デジタルデータから読み取れる情報や来歴をメタデータとしてできるだけ多く正確に記載し、記載内容をデータベース化し検索、発見するという方法がとられてきた。この方法は、デジタルデータの数がそれほど多くない場合や、利用者が正確な用語や背景知識を有している場合には機能する。しかしながら、デジタルデータの数が増加している現在では、デジタルデータを取り巻く登録側、利用側の利用者数の増加により、登録側利用側を含めた用語の統一や正確な用語の選択が非常に困難であり、結果として、適切な検索、情報発見に結びつかなくなっている。ソーシャルネットワークサービスはこのような検索に依らない情報との出会いを需要とするサービスとして機能していると捉えることができる [1]。すなわち、デジタルデータの数と利用者数の増加は、デジタルデータの全利用者間共通の検索用語や背景知識、時代知識を前提としない分野横断的なデジタルデータの利活用を必要とする。

本稿では、筆者らが検討を進めている自律型オープンデータ流通基盤について述べる。本基盤は、既存の検索とは異なるアプローチでデジタルデータとの出会いを提供するオープンな基盤で、複数の所有者、利用者、サービス提供者の参入を積極的に促し、デジタルデータのボーダレスかつ柔軟に柔軟に容可能な利活用を図る。本基盤は、主に、プ

¹ 慶應義塾大学

Keio University, Hiyoshi, Yokohama 223-8522, Japan

^{a)} kaneko@inl.ics.keio.ac.jp

レビュー時にデジタルデータの価値を鮮明に利用者が捉えることを可能にする Virtual File と、メタデータやデジタルデータの内容解析に依らずデジタルデータの意味集合を形成する Catalogue System によって形成される。いずれも、デジタルデータに含まれる情報の価値を人間が抽出し、価値付けに寄与しないところでコンピュータ処理を最大限活用する。具体的には、まず、デジタルデータ一つ一つに注目し、そこから得られる複数の意味や価値をそれぞれ異なる Virtual File にすることで、意味・価値の顕在化を行う。そして、意味や価値、コンテキストを共有する複数の Virtual File を Catalogue として集合化し Virtual File (デジタルデータ) 間の相対的な距離を与えた上で、複数の Catalogue 情報を組み合わせ計算処理することで、意味集合の動的な形成を図る。最終的には、利用者にとって既知のデジタルデータ群を起点に、計算で得られた未知の Virtual File 情報を利用者に提示し、デジタルデータとの出会いを提供する。本基盤では、デジタルデータ、Virtual File, Catalogue, およびサービスアプリに対して、それぞれに明確な所有者の概念が存在し各所有者による管理を徹底する。これにより、デジタルデータの所有者でなくても自由に Virtual File や Catalogue, サービスアプリの生成を可能にし、本基盤を用いる利用者の拡大をすること、デジタルデータの利活用の幅を広げ所有者の保有コストの低減と保有モチベーションの向上を図ること、利活用の拡大に貢献した Virtual File や Catalogue, サービスアプリの作成者・所有者に経済的なフィードバックを与えることを可能にする。

以下では、2章で自律型データ流通基盤構築に向けた課題について述べ、3章で自律型オープンデータ流通基盤のモデルを示す。4章および5章で、Virtual File と Catalogue System を紹介し、6章でサービス事例を紹介し、7章で本稿をまとめる。

2. 自律型データ流通基盤に向けた課題

2.1 ボードレスなデジタルデータ利活用に向けた課題

現在検討を進めている自律型データ流通基盤は、所有者横断的、サービス提供者横断的、利用者横断的なシステムを目指している。すなわち、どこかひとつの事業主体が主導的な役割を果たして実現するデジタルデータ流通基盤ではなく、デジタルデータの保存側から活用側までをそれぞれ複数の事業主体が構成する自律分散型の基盤の確立を目指している。必然的に扱うデジタルデータは、デジタルデータの分野や種類、フォーマットに限定されず、また利活用の方法に限定がない。このような自律型データ流通基盤における課題は以下の4点である。

- ★ デジタルデータの所有者が異なり、また一般的にデータは利用者にとって遠隔地にあるため、データ転送にコストがかかり利用者の計算機を用いて直接デジタル

データを解析することは困難である。

- ★ デジタルデータの所有者が作成したメタデータや特徴量を得て検索対象にできる可能性はあるが、利用者と所有者に背景知識や時代知識の差による言語・用語・特徴量選択の差異があると、発見に至らない。
- ★ テキスト、音声、画像などクロスメディア環境が想定されるが、メディア毎に異なるメタデータや特徴量を用いて検索することが一般的であり、それらをシームレスに探索することができない。
- ★ 異なる言語や用語、メタデータ、特徴量をつなぎ合わせるには、中間的な語彙データや辞書データが必要になるが、万人が満足するような辞書を構築することは困難である。

いずれも、デジタルデータに含まれる意味を利用者が検索語や特徴量、もしくはデジタルデータ自体を指定して検索を行おうとすることによって発生する課題である。

2.2 柔軟に変容可能なデジタルデータ利活用に向けた課題

自律型データ流通基盤を用いたデジタルデータの利活用の方法や形態は、固定的であってはならず、利活用を経る毎に柔軟に変容可能でなければならない。すなわち、本基盤が扱うデジタルデータを容易に追加できるだけでなく、すでに登録されたデジタルデータの新たな側面や意味づけ、価値付けが発見された際には、これらの価値が容易に基盤に追加され、発信されなければならない。さらには、利用者とデジタルデータの出会いの提供方法も自由に定義でき、変更でなければならない。このような柔軟性に対する課題は、以下の4点である。

- ★ デジタルデータの流通基盤に登録されるデジタルデータの意味づけとも言えるメタデータ項目は一度定義されると、ほとんど更新がなされず変化が発生しない。
- ★ デジタルデータのメタデータ付けは所有者によってのみ行われ、デジタルデータの実際の利用者の思考を経た新たなメタデータの追加は行われず、デジタルデータの追加的な価値付け、意味づけが行われない。
- ★ デジタルデータと他のデジタルデータの関係を利用者が見いだしたとしても、それら関係情報の保存共有は行われず、複数のデジタルデータを一堂に提示することで得られる価値観の発信が行われない。
- ★ デジタルデータ流通のポータルサイトは、バックエンドのデータベースと一体化していることが多く、利用の仕方を柔軟に変更することは困難で、デジタルデータと利用者の出会いのパターンを固定化してしまっている。

いずれも、デジタルデータの流通に所有者のみが強く関わっており利活用における所有者の負担が高いため利活用が促進されない。また、利用者の知見が流通に組み込まれていかないことに課題がある。

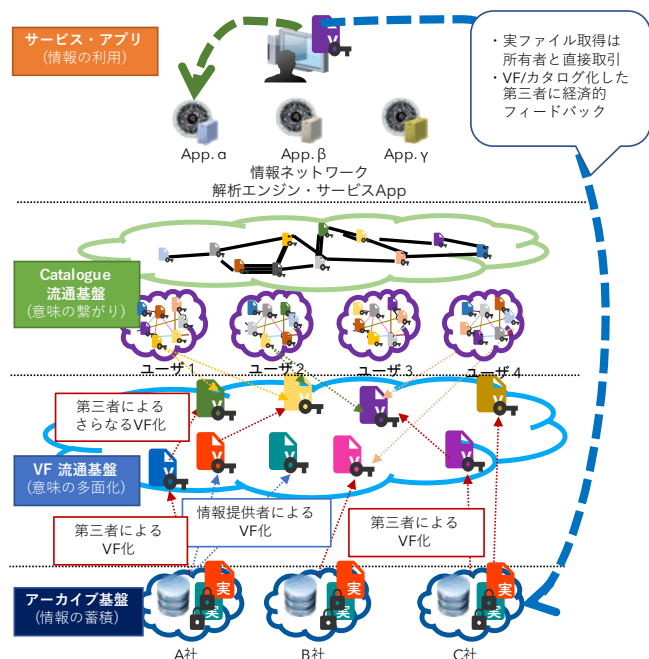


図 1 自律型オープンデータ流通基盤のモデル。

Fig. 1 Model of autonomous data circulation platform.

3. 自律型オープンデータ流通基盤

筆者が検討する自律型オープンデータ流通基盤は、デジタルデータの価値付けを Virtual File と Catalogue によってオープンにし利用者貢献型のデジタルデータの流通基盤とするとともに、Catalogue が形成するハイパーグラフ空間を用いることで計算機処理可能な形でデジタルデータの価値を柔軟に捉えようとするものである。

本基盤は、デジタルデータをノード、デジタルデータ間の関係をリンクとするネットワーク構造を核としている。より正確には、このネットワーク構造においてノード (頂点) は、筆者が Virtual File と呼ぶ仮想化されたデジタルデータである。デジタルデータは潜在的に複数の意味や価値を有しており、これらの価値が顕在化することによって新たな Virtual File が生成される。すなわち、Virtual File はデジタルデータが本来有する複数の価値のうちの一つに注目しそれに特化して顕在化、表象化がなされたものである。したがって Virtual File とデジタルデータの関係は、基本的には多対一の関係となる。また、デジタルデータの所有者が少なくとも一つの Virtual File を生成し、本自律型オープンデータ流通基盤のノードとして登録する。そして、利用者がデジタルデータに別の意味や価値を見いだした際には利用者が新しい別の Virtual File を生成し基盤に登録する。

一方リンク (枝) は、筆者が Catalogue と呼ぶ Virtual File 間の関係の有無に対応する。Catalogue は何らかの関係性を共有する 2 つ以上の Virtual File 群を束ねるものである。身近な例では、デスクトップコンピュータ環境におけ

るディレクトリ (フォルダ) と似た概念である。利用者は複数の Virtual File 間になにかしらの共有している意味やコンテキストを見いだすと、これらをひとつの Catalogue にまとめる。1 つの Virtual File が異なるコンテキストを有する場合は、各コンテキストに対応する複数の Catalogue に同一の Virtual File を登録することが可能である。Virtual File の作成者/所有者と Catalogue の作成者/所有者が合致する必要は無い。数理的には、Catalogue は Virtual File を頂点とするハイパーグラフ (Hyper Graph)^{*1}となる。したがって、2 頂点間の関係というよりも、複数頂点間の関係、いわば集合の概念と捉えることができる。また、Catalogue のハイパーグラフにおける枝の属性に、集合の意味は含まれない。ただし、Catalogue の補足情報として言語化した意味を保持することを否定するものではない。

サービス提供者は、様々な利用者が作成した Catalogue を組み合わせて数値処理を行い、Virtual File 間の関係を把握し利用者に Catalogue や Virtual File の提示を行う。例えば、十分な数の Catalogue 群が形成するハイパーグラフ (複数のハイパーグラフの組み合わせもハイパーグラフになる) を解析して粗密を判断することで、Virtual File 間の近傍性を明らかにすることができる。利用者がよく似たデジタルコンテンツと出会いたい場合には同じ密な空間に位置する Virtual File を提示したり、少し違うデジタルコンテンツと出会いたい場合には、粗な領域を経た先にある Virtual File を提示するといった方法が考えられる。利用者の情報入手におけるコンテキストと Catalogue によって構築されるハイパーグラフの数理解析の橋渡しの方法は一通りではないため、多くのサービス提供者が競争的に機能を開発することが求められる。

以上の機能をモデル化し、記載したのが図 1 である。以下では、Virtual File と Catalogue についてより詳細に説明する。

4. Virtual File

Virtual File は、デジタルデータの意味を顕在化し、その顕在化した意味を先鋭に表現し利用者に訴えかけ、その価値を明確化する。例えば、1965 年に配給された映画 Sound of Music は、様々な視点から語ることができる。例えば、映画コンテンツという視点 (図 2(a))、Julie Andrews という主演女優の視点 (図 2(b))、ドレミの歌の視点 (図 2(c))、第二次世界大戦の視点 (図 2(d))、さらには、65 ミリフィルムで撮影されたシーンの 8K スキャニングという視点 (図 2(e)) である。いずれの表象も、Sound of Music の映画の 1 シーンを静止画として切り出すとともにその静止画に見合うテキストタイトルを与えたものである。しかし、

^{*1} ハイパーグラフとは、2 頂点を 1 つの枝で接続するいわゆる通常のグラフを拡張した概念で、1 つの枝に接続する頂点数が 2 に限定されない。



図 2 20th Century Fox によって 1965 年に配給された映画 Sound of Music の様々な表象。

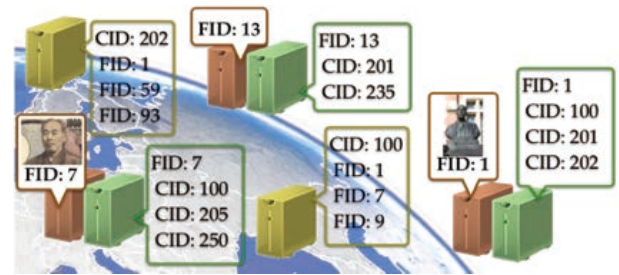
Fig. 2 Various representations of Sound of Music distributed by 20th Century Fox in 1965.

漫然と Sound of Music という名前やタイトルシーンを見せられるのと、上述の表象がなされたものでは、その表象の受け取り手の印象は大きく異なり、各利用者の興味関心への訴求力は異なってくる。Virtual File は受け取り手の興味関心を刺激し、いずれの Virtual File の利用者も Sound of Music の映画の全編、もしくは一部に対応する同一のデジタルデータにたどり着くことになる。

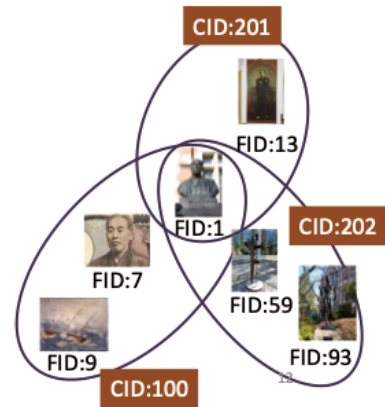
Virtual File は、具体的には、デジタルデータの意味や価値を先鋭に捉えた情報だけをメタデータとして保持し、加えて Virtual File の作成者情報、参照するデジタルデータへのアクセス情報を有する。Virtual File はデジタルデータに比べて軽量で、可搬可能なフォーマットである。また、Virtual File は Virtual File の作成者がデジタル署名し、その出所を明らかにするとともに、認証情報等を加えて、当該 Virtual File を介してデジタルデータへのアクセスが発生した場合には、デジタルデータの所有者から何かしらの経済的フィードバックを得ることを可能にする [2]。

5. Catalogue System

Catalogue System は、コンテキストを共有する Virtual File の集合を明示する Catalogue を管理、共有するシステム [3] である。前章で述べたように、Virtual File は潜在的に複数の価値を有するデジタルデータから先鋭に意味や価値を切り出す役割を果たす。このように各デジタル



(a) Catalogues stored distributedly.



(b) Hyper graph view after Catalogues retrieval from distributed Catalogue Servers.

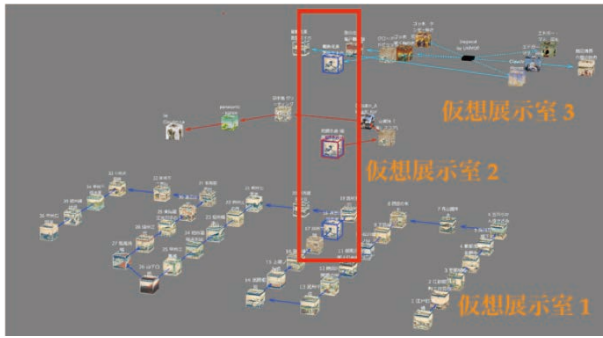
図 3 カタログシステム。

Fig. 3 Catalogue System.

データの意味を先鋭化した上で、同じ意味を有する複数の Virtual File を組み合わせ提示することで、受け取り手はその意味を強く意識することが可能となる。このような複数の Virtual File を組み合わせることで意味や価値を浮き上がらせる機能を果たすのが Catalogue である。

一方で、このようにして形成された Catalogue を複数重畳していくと、複数の意味や視点に基づいた Virtual File 間の関連性の強弱が浮き上がってくる。例えば、ある Virtual File は 10 個の Catalogue に組み込まれる一方で、別の Virtual File は 3 個の Catalogue に組み込まれる。また、1 個の Catalogue にしか組み込まれない Virtual File も存在するであろう。10 個の Catalogue に含まれる Virtual File は意味理解の中心的な役割を果たしており、多くの利用者に認知されていることが明かであり、1 個の Catalogue にしか含まれない Virtual File は余り知られていないと判断することができる。このように、デジタルデータの直接的な解析をせずに、Catalogue による Virtual File 間の繋がり数理的な解析を可能にするすることが Catalogue の重要な機能である。

Catalogue System は Catalogue を自律分散的に管理する機構である。Catalogue の管理とは、Catalogue の作成者/所有者を指定すると、その作成者/所有者が作成/所有する全 Catalogue を入手でき、また各 Catalogue に含まれる Virtual File の一覧を入手可能であること、および、ある



(a) MoSaIC viewer: virtual museum.



(b) MoSaIC viewer: content.

図 4 MoSaIC における可視化。

Fig. 4 Visualization by MoSaIC.

Virtual File が指定されたときにその Virtual File が含まれる全 Catalogue の情報を速やかに入手可能であることである。前者は Catalogue の作成者による Catalogue の管理であり、後者は Virtual File の作成者による Catalogue の管理である。Catalogue System では、前者を Catalogue Server と呼び、後者を Graph Manager と呼ぶ。利用者は、Graph Manager を用いた Virtual File をキーにした Catalogue の取得、Catalogue Server を用いた Catalogue をキーにした VirtualFile の取得を繰り返しながら、関連するデジタルデータを把握する。図 3 は、図 3(a) に自律分散的に管理される Catalogue の状況を示しており、図 3(b) に利用者が各サーバから収集してきた Catalogue 情報を組み合わせて形成したハイパーグラフ空間を示している。図 3(a) では、黄色のサーバが Catalogue Server を、緑色のサーバが Graph Manager を示している。

6. サービス構成例

筆者は、慶應義塾大学デジタルメディア・コンテンツ統合研究センター（以下、DMC）において、上述のモデルに基づいたシステムの構築に取り組んでいる。システムの構築にあたっては、モデルの確立とともに、アプリケーションサービスを通したモデルの有用性の検証が必要である。本稿では、DMC において構築した 2 つのアプリケーションサービスを紹介する。

図 4 は、Museum of Shared and Interactive Cataloguing



図 5 キャンパスミュージアム。

Fig. 5 Web service of Campus Museum.

プロジェクト（以下、MoSaIC）における仮想美術館のビジュアライゼーション [4] の様子を示したものである。仮想美術館は、ひとつの Catalogue が一つの仮想展示室を構成し、仮想展示室を共通のデジタルデータを介して連続的に積み上げることで、単一のデジタルデータを異なる視点で捉えることを可能にする。図 4(b) は、本システムで表示するコンテンツ表示である。

図 5 は、慶應義塾大学日吉キャンパスに存在する様々な人文的資料を多元的につなげて表示する Web サービス、キャンパスミュージアムのスクリーンショットである。このサービスでは、ひとつひとつの Catalogue に含まれる複数のデジタルデータ（Virtual File）を横方向に並べ、縦方向にメインとなるデジタルデータが含まれる Catalogue をリストすることで、一つ一つの資料を異なった視点から捉えることを可能にしている。

アプリケーションサービスの構築とその動作検証は、デジタルデータ利活用の原理的モデル像とサービスにおける使い勝手という現実的側面のギャップを明らかにする。例えば、原理的モデル像では、すべてのデジタルデータは対等に扱われるべきであるが、重み付けをした情報提示がなされなければ人間は情報を理解することができない。このような体系的なデジタルデータの対等性と意味理解におけるデジタルデータの不均衡性をどのように達成していくかは、これから解決していかなければいけない課題である。

7. おわりに

デジタルデータの流通の困難性は、デジタルデータの所有者と利用者によるデジタルデータに対する価値付けの偏りにある。現在のデジタルデータの流通基盤は、この価値の偏りをデジタルデータの所有者が解消することを求めている。加えて、デジタルデータの所有と維持には膨大なコストがかかっており、結果的にデジタルデータの利活用が滞る結果となっている。

筆者は、デジタルデータの利活用に向けて、自律型オープンデータ流通基盤の検討を行ってきた。具体的には、デジタルデータの価値付けを Virtual File と Catalogue によってオープンにし利用者貢献型のデジタルデータの流通基盤とするとともに、Catalogue が形成するハイパーグラフ空間を用いることで計算機処理可能な形でデジタルデータの価値を柔軟に捉えようとしてきた。また、その可能性と有用性を検証してきた。今後も、アプリケーションサービスを含めて、基盤の検証と機能の開発を継続する予定である。

参考文献

- [1] 総務省情報通信国際戦略局情報通信経済室：スマートフォン経済の現在と将来に関する調査研究の請負報告書，総務省 (2017)，
- [2] 荻谷凌，金子晋丈：Virtual File：認証認可とファイル取得の独立制御を可能にする広域データ流通フォーマット，信学技報，vol. 118, no. 466, IN2018-127, pp. 259-264, 2019 年 3 月.
- [3] Y. Miyashita, H. Ishikawa, F. Teraoka, and K. Kaneko : Catalogue: graph representation of file relations for a globally distributed environment. In Proceedings of the 30th Annual ACM Symposium on Applied Computing (SAC '15). ACM, New York, NY, USA, pp. 806-809. DOI: <https://doi.org/10.1145/2695664.2696047>
- [4] H. Ishikawa, K. Kaneko : Polymorphic cataloguing and interactive 3D visualization for multiple context of digital content: MoSaIC. In Proceedings of the 14th International Conference Advances in Computer Entertainment Technology (ACE 2017), 2018. pp. 172-186. DOI: https://doi.org/10.1007/978-3-319-76270-8_13