

デジタルアーカイブ社会実現に向けたレイヤー構造の必要性和人文学の役割

永崎研宣

1 はじめに

「デジタルアーカイブが日常的に活用され、様々な創作活動を支える社会・学術・文化の基盤となる社会」としてのデジタルアーカイブ社会の実現が提唱されるようになり、デジタルアーカイブ(以下、DA)はいよいよ本格的に社会の基盤の一部として確立されようとしている。たとえば、水道が引かれて水が簡便に使えるようになり、電線が引かれて電気を使えるようになったからといって、水や電気の使い方にまつわるあらゆることがいちいちすべて開示され、すべての人々がそれらを活用する必要がないのと同様に、社会基盤の使い方は人それぞれである。一方で、水は飲んだり洗浄に用いたりすることができるが、電気はコンセントが用意されただけで何かができるわけではない。パソコンやエアコン、冷蔵庫など、何かを介することで我々の活動を支援し、選択肢を増やしてくれる。水も、ただ水をそのまま使用するだけでなく、沸かして風呂にしたりさらに温泉の素を入れてみたり、あるいは、料理に用いたりするなどすれば、水道という基盤に依拠しつつ様々なことが実現できるようになる。社会の基盤となるものは、さまざまな人に自由に利用されることで新たな価値を生み出していくものであり、発展的な利用の仕方のノウハウやツールもまたさまざまな人に開発され共有され、それ

らを通じてさらに新たな価値を生じていく。なかにはビジネスとして展開され多くの雇用を生み出していくこともある。DAもまた、そのような存在としての社会基盤の一つになろうとしているのだとすれば、やはり同様の事態がもたらされることになるだろう。しかし、電気と水のインフラとしての展開がそれぞれの特性に応じて異なるように、DAもまた、その特性に応じて展開していくことになるだろう。

なお、人文学という立場においてはこのような基盤自体やそのあり方から是非に至るまでの様々な事象が、やはり様々な観点から議論の対象となり得る。デジタル・ヒューマニティーズ(以下、DH)においても、日本ではまだあまり見られないものの、海外ではデジタル社会そのものへのラディカルな批判に取り組む流れも出てきている¹⁾。しかしながら、本章では、敢えて議論の範囲を限定し、社会基盤としてのDAの構築を議論の前提とし、それ自体についての議論は別の機会とする。そして、そこにおいてDHを含む人文学が果たし得る役割について、DAを成立させているレイヤー構造を手がかりとして検討する。

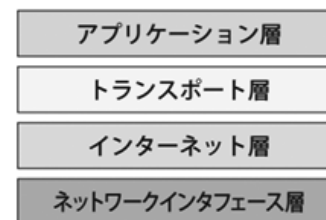
2 レイヤー構造の考え方

ここでレイヤー構造を採りあげるのは、主に二つの理由がある。一つは、インターネットの仕組みがしばしばレイヤー構造として説明され、DAは主にインターネット上で展開されるものであることから、インターネットのレイヤー構造の上にさらにDAのレイヤーを重ねるというイメージは比較的捉えやすいものであると考えたからである。しかしながら、情報工学・インターネット技術に詳しい読者ばかりでない本書においてインターネットのレイヤー構造を引き合いに出すことは適切なかどうか、という疑問が生じるかもしれない。それこそがまさに、これを採りあげるもう一つの理由である。というのは、本稿を執筆している2022年度から、高等学校情報科「情報I」が必修修化され、その内容の一つとして採りあげられているからである。文

部科学省が配布している「高等学校情報科「情報I」教員研修用教材 第四章 情報通信ネットワークとデータの活用²⁾」には、以下のような説明が図とともに登場する(図1)。

(2) 有線LANと無線LANの違い

TCP/IPの4階層モデルについて説明する。インターネットの通信において、通信する際の規定を定めたものを通信プロトコル(略してプロトコルということもある)という。このプロトコルがあることで、相互に通信を行うことが可能になっている。TCP、IPv4、IPv6、DHCP、HTTP、SMTP、POPなど様々なプロトコルがインターネットの通信で用いられるが、これらは、4層のレイヤーに分けて考えることができる。インターネットでアプリケーション同士が通信する際は、上位のアプリケーション層から下位の層に移動することで、通信手順が付加され、ネットワークインタフェース層で物理的な通信を行っていく。



図表4 4層のレイヤー

図1 「情報I」教員研修用教材の一部

すなわち、数年後には、インターネットのレイヤー構造は、高校を卒業したなら誰もが知っているべき知識になっている。筆者も含め、この情報Iを履修していない者がそのような知識体系を学んだ若者達とどのように協働していくべきなのかは今後の検討課題だが、本件に限らず基礎的な情報処理・情報ネットワークについての知識に基づいて思考することが当然となる世代によって、DAの世界は大きくアップデートされざるを得ないだろう。

このようなことから、本章では、この基礎的な情報ネットワークの知識を踏まえてDAのあり方としてのレイヤー構造の検討を試みる。

なお、図1に示した資料では説明がやや簡素すぎるため、もう少し説明を加えておくと、TCP/IPの4階層モデルと呼ばれているインターネットのレイヤー構造には、(1) イーサネットや無線LAN等のハードウェアが通信を成立させる層としてのネットワークインタフェース層、(2) 複数のネットワーク間でデータ伝送を実現するための層としてのインターネット層(ここにはインターネット・プロトコル(IP)が当てはまる)、(3) 信頼性のあるデータ通信を実現するための通信制御やプログラム間の通信を行うトランスポート層(ここにはTCPやUDPが当てはまる)、そして、(4) Webやメールな

どのアプリケーション同士でやりとりを行うアプリケーション層がある。たとえば、Web 頁を見ようとすると、アプリケーション層に属する Web ブラウザが閲覧したい Web 頁のデータを取得しに行こうとするが、そうすると、データをパケット毎に分割し、さらに伝送ロスがあった際には再送の要求もしてくれる TCP(=トランスポート層)に依頼する。TCPは、相手先の Web サイトのアドレスを見つけてデータを伝送しようとしてくれる IP(=インターネット層)に依頼する。そして IP は、実際にデータを伝送してもらうために、イーサネットや無線 LAN 等のネットワークインターフェイスに依頼するのである。データの送信側がそのようにすると、データの受信側でも同様に Web 通信が行われることになる。そして、メールやオンライン会議等、インターネットを利用する様々なサービスの場合にも、Web と同様に、このアプリケーション層において通信を行い、その際に、トランスポート層にデータを委ねることになる。これをトランスポート層の側からみると、アプリケーション層にどの通信手段が来ようとも、同様にデータの受け渡しを行うことになるのである。つまり、このレイヤー構造においては、上の階層と下の階層がそれぞれに独立しつつ相互に連携することで一つの通信を成立させているのである。

DA の多くがインターネットを利用することにより依拠しているこのレイヤー構造は、DA の利用という局面では、アプリケーション層のさらに上にいくつかのレイヤーが重なることになると考えることができる。それによって DA に関する議論がより効果的に整理され、有機的な成果へと結びつくことがより容易になるかもしれない。そこで、本稿では、インターネットの階層モデルの上位層となる DA のレイヤー構造について検討してみたい。

3 DAにおけるレイヤー構造

ここでは、DA をレイヤー構造として捉えようとする場合の全体像を提示した上で、それを構成する各レイヤーについて説明する。

3-1 DAにおけるレイヤーの全体像

DA は、インターネットの4階層モデルにおいてはアプリケーション層に位置するものである。しかし、DA の成り立ちや運用を考慮すると、その上にいくつかのレイヤーを見いだすことができるだろう。DA の基層となるのはデジタルコンテンツであり、これが一つ目のレイヤー(DA-L1)になる。デジタルコンテンツはそのままでは単なるバイナリー(2値化)データであったり、あるいはテキストデータだとしてもそれだけでは活用が難しいため、メタデータや内容を説明したり注釈したりするためのデータが必要になる。これらをまとめてアノテーションと位置づけ、二つ目のレイヤー(DA-L2)とする。そして、アノテーション(DA-L2)を介してデジタルコンテンツ(DA-L1)の閲覧や検索・分析等ができるような仕組みとしてのインターフェイスが必要であり、これをインターフェイスレイヤー(DA-L3)とする。最上位のレイヤーとしては、技術的な事柄とは必ずしも関連するとは限らず、むしろ技術的でない事柄が重要であることも多いはずだが、多様な関心を持つ DA ユーザに向けて伝えるための様々な手立てをコミュニケーション・レイヤー(DA-L4)として立てておきたい。

また、DA を構成する要素としては、法や規範、技術的な規格なども必須である。それらはいずれも、いずれのレイヤーにおいても常に参照されるべきものであることから、各レイヤー構造の外に置かれる参照要素とする。

以上を図式化したものが図2である。これを踏まえて以下に各レイヤーについて検討していきたい。

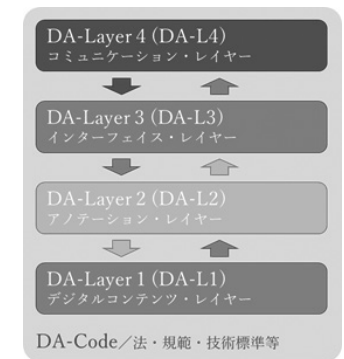


図2 デジタルアーカイブを構成するレイヤー

3-2 デジタルコンテンツのレイヤー(DA-L1)

DAにおいて基礎となるのはデジタルコンテンツである。これには、物理的な資料としての書籍をはじめとする様々な物品をデジタル化したものと、最初からデジタルデータとして作成されたものがある。前者に関しては、デジタル化の手法や粒度に応じて様々な種類のコンテンツがあり得る。たとえば古典籍や古文書をデジタル化する場合には、平面のデジタル撮影画像とすることもあれば、テキストデータとすべく文字起こしをすることもある。DA登場の文脈に鑑みるなら前者が中心になりそうではあるものの、アナログな資料を符号化し離散的な情報として保持したデータ、つまり、二値化された情報という観点ではその二者の間に差異を見いだすことは難しく、DAの対象としては同列に扱うべきだろう。しかし一方で、データとしての扱いやすさやアナログの再現性という観点では大きく異なることも確かである。そこで、デジタルコンテンツとしてのテキストデータ、デジタル画像、3Dや動画、についてそれぞれ検討してみよう。なお、デジタルコンテンツに付与される基本的なメタデータまではこのレイヤーに含めるとする。

3-2-1 DA-L1におけるテキストデータ

テキストデータに関しては、元の資料において文字が書かれていると判断されたものを、任意の文字コード(近年はUnicodeがよく用いられる)に基づいて可能な限り文字をそのままに記録したものがこれに該当する。操作がしやすく処理にかかる負担もそれほど大きくないため、情報技術が現在ほど発達していなかった時期にはテキストデータがほとんど唯一の選択肢であり、テキストデータベース、テキストアーカイブなどと呼ばれ、世界各地で作成・公開されてきた。しかしながら、「文字をそのままに記録」することには様々な課題がある。これは、デジタルでなくともアナログの時代から存在した問題であり、デジタル媒体に移行したことで顕在化したと言えるだろう。デジタル化の本質的な問題とも関わるため、ここで少し検討しておきたい³⁾。手書きの文字は、どんな形であれ、記述することが可能であり、それにど

のような内実を見いだすかは受け手の側に委ねられる。書き手と同じ文字体系を学習した読み手であればほぼ正確な伝達が可能であるはずだが、手書き文字の形状の多様性は必ずしもそれを保証せず、受け手次第で理解のされ方が変わってしまう場合もある。一方、コンピュータにおけるテキストデータとしての文字は、文字コードという表のなかで定義された文字を使うことになる。そして、処理速度を向上させつつ互換性を確保するために、文字コード表は書き手と受け手が共通のものを使用することが強く期待される。たとえば、65番はアルファベットのA、66番はアルファベットのB、といった案配である。しかしながら、書き手と受け手が不特定多数であり、その人々が各自のコンピュータを用いてテキストを閲覧・利用する可能性がある場合には、不特定多数が共通の文字コード表を利用することを志向することになる。結果として、一定の信頼性が認められる組織が定めた文字コード表を何らかのコミュニティが共有するという形になる。たとえば、米国であれば、1963年にASA(American Standards Association, 現在は American National Standards Institute)という国家の規格を定める団体により制定されたASCII(American Standard Code for Information Interchange)という文字コード表を用い始めた。すなわち、米国の規格制定団体という、一定の信頼性が認められる組織が決めたことを米国内で利用し始めたのである。あるいは、日本であれば、1978年にJIS(日本工業規格、現在は日本産業規格)によってJIS漢字コードが制定されたことで、日本で用いられる漢字の番号が共通化され、日本国内では安定して漢字をコンピュータ上で利用できるようになった。それぞれの政府等の関連機関によって管理された文字コード表は、それぞれの範囲においては安定的な運用が可能となったのである。

一方、各国で文字コード表が別々に策定されることは、同じ文字番号が国によって異なる文字として扱われる可能性を生じさせることであり、実際のところ、文字の表示がおかしくなり、結果として、文字コードが異なるとプログラムもうまく動作しないという問題を抱えていた。インターネットが普及し始めたことによりこの問題は一気に顕在化した。すなわち、それま

では設置された地域で用いられている言語を表示できれば十分だったコンピュータが、インターネットに接続されたことで、世界中の文字をうまく表示するという使命を課せられたのである。政府や言語・文字体系を共有する地域のコミュニティによる文字コード表の共有は、かくして、グローバルなコミュニティに依拠することになる。そこで中心となっていくのがUnicode⁴⁾である。

Unicodeは、国際標準化機構(ISO)等が定めるISO/IEC 10646⁵⁾という規格に準拠する形で定められており、文字コード表そのものは国際標準化機構が策定している。つまり、165カ国(2021年12月時点)が加盟する標準規格策定組織が制定したものが世界中のコンピュータや情報機器を利用する人々の間で共有される形になっているのである。これが普及したことにより、デジタルコンテンツ層(DA-L1)におけるテキストデータは国際的に広く利活用可能なものとしてDAのレイヤーの一つとなったと言える。すなわち、テキストデータは、一つ上位のアノテーション層(DA-L2)からアノテーションを付与される基盤となるのである。

なお、Unicodeが導入されたことですべての問題が解決したのかと言えばそうではなく、むしろ、グローバル化されたことにより、文字をコード化することやそもそも文字を共有すること自体に内在する課題もまたグローバルな議論の対象となった。文字のバリエーションが多すぎて文字コードでの表現が困難なのは漢字文化圏だけではなく欧州中世文献でも課題を抱えていることや⁶⁾、組み合わせ文字はハンゲルのように組み合わせパターンをすべてコード化すべきなのか、あるいはインド系文字のように組み合わせられる前の音素などをコード化して表示時点で組み合わせるべきかどうか、等々、さまざまな課題が広く議論されるようになっていく。学術的観点からの文字の扱いについても発展途上ではあるものの、Unicodeに関連するコミュニティではこれに近年大きく力を入れており、研究者の側で継続的に関わるのが可能であれば要望を通すことはそれほど難しいことではなくなっている⁷⁾。

これをDA-L1のレイヤーとして位置づける場合、上位層にあたるアノテーション層(DA-L2)から参照され利用されるものである。技術的にはXMLやJSON等の何らかの広く普及しているフォーマットを用いて、人文学テキスト資料のためのXMLに準拠した記述手法を提示するTEI(Text Encoding Initiative)ガイドライン⁸⁾やWeb画像相互運用の国際的な枠組みであるIIIF(International Image Interoperability Framework)⁹⁾をはじめ、なんらかの規格や手法に基づいてアノテーションが付与されることになる。テキストデータの場合、どこまでがデジタルコンテンツでどこからアノテーションかの区別は必ずしも明確ではなく、アノテーションも含めてデジタルコンテンツであるという位置づけがなされることもあり得る。むしろ、テキストデータのうちで、アノテーションではなくその対象として位置づけられる部分、として捉えるとよいかもかもしれない。

3-2-2 DA-L1におけるデジタル化資料画像

DAといえば、デジタル撮影された書籍やその他様々な収蔵品の資料のデジタル化画像を公開するWebサイトの一般名称として用いられることが多い。デジタル化資料画像はデジタルコンテンツ層における主要な要素の一つである。デジタル化にはスキャナーで読み取る場合やデジタルカメラで撮影される場合などがあるが、いずれにしても、対象物をセンサーで読み取りそのデータを二値化することになり、そのバイナリー(二値)データがDA-L1レイヤーとして位置づけられる。文字化という人為的なプロセスが介在するテキストデータに比べると、データ作成時には機械に依存する部分が大きく、一方で、データを解釈するにはテキストデータほどの一意性はなく、データの解像度に応じて多様な解釈に向けた素材を提供することになる。文字資料であれば本文さえ読み取ればよいのか、あるいは返点を始めとする訓点まで子細に読む必要があるのか、等々、ここには大きな幅がある。そして、それゆえに、アノテーションがなくては共通理解を持つことは容易ではなく、バイナリーデータに対して説明のテキストを付与する形でのアノテーション

は、様々な場面で有効になる。

デジタルコンテンツのフォーマットには、画像以外にも、音声、動画、3D等、様々なタイプのデータがあり得る。技術の進展次第ではさらに様々なものがDAで用いられることになるだろう。個々のフォーマットにはそれぞれ様々な議論や留意事項があり、広く用いられているものの多くは国際標準規格として制定されているが、本稿の目指すところはこれをレイヤーの一部として取り扱うことであり、深くは立ち入らない。多くの場合に、アノテーションを付与することでコンテンツへの理解を深め、共有することができる。

3-3 アノテーションのレイヤー(DA-L2)

デジタルコンテンツに対して様々な付帯情報を付与するのがアノテーション(注釈)のレイヤーである。かつては、デジタル化に多大なコストを要したこともあり、資料それ自体が誰から見ても価値を持つような一部の優品のみがDAの収録対象となっていたこともあったが、デジタル技術のコモディティ化が進化した現在、DAにおいては、ごく一部を除き、デジタルコンテンツそのものが価値を持つことは必ずしも多くない。むしろ、そこにあるコンテンツがどのようなものであるかが説明されることで価値を持つことになる。コンテンツ単独で価値の高いものの場合には、ただそのコンテンツについての説明が用意されれば十分だが、多くのDAでは、むしろコンテンツ群がなんらかのまとまりを持っている場合や、コンテンツに含まれる何らかの共通の要素で横断的に集約されることによって価値を持つことがある。特定の人物や地域に関連するもの、同じような形や色彩を持つもの、似たような語彙を含むものなど、様々な観点があり得る。そのようにしてコンテンツのまとまりや共通要素を見いだせるような説明が付けられることがコンテンツの有用性を引き出すために極めて重要なものとなる。

アノテーションは人がつけるものもあればコンピュータが自動的に付与するものもある。近年は人名や地名をディープラーニング技術でテキストから

自動的に取り出すソフトウェアが普及しており、正確性には一定の疑問符がつく場合もあるにせよ、OCR(光学文字読み取り)と組み合わせることで、画像ファイルから人名や地名を取り出すことも相当程度可能になってきている。正確性が足りない部分のみを手で補うことができれば、より効率的により妥当なアノテーションを付与することができるだろう。

アノテーションという概念は相対的なものであり、アノテーションとデジタルコンテンツをまとめたものに対してさらにアノテーションを付与したり、アノテーションそのものに対してアノテーションを付与するといったこともあり得る。そのため、このレイヤーもまた、あくまでも状況を理解し共有するための主観的なものであり、あくまでも説明概念のようなものとして理解していただきたい。

デジタルコンテンツにアノテーションを付与する技術は様々な存在する。Webコンテンツに対するアノテーションとしては、近年はIIIF対応のDAが普及したことに伴い、画像や音声、動画等のIIIF対応コンテンツを対象としたアノテーションが広く利用されるようになってきている。すでに多くのDAがこの種のアノテーションを実装して高い利便性を提供しているが、その仕組みをデータ提供者やDA構築者など、誰もが提供できるようにした人文学オープンデータ共同利用センター(CODH)によるIIIF Curation Platform¹⁰⁾は日本発でありながら国際的にも先進的な事例である。各地のWebサイトにおけるIIIF対応のWebコンテンツに対して自由にアノテーションを付与するため、日本のコンテンツでこれを利用しているものは少なくない。レイヤーとして捉える場合にも、DA-L1レイヤーに対し、それがIIIFに対応してさえいれば、自由度の高い形でDA-L2レイヤーと連携できることになる。

また、テキストデータに対するアノテーションにも様々なものがある。なかでも前出のTEIガイドラインは、人文学向けに策定されたルールとして国際的に広く用いられており、テキストデータへのアノテーションに際して必要とされるようなことは概ね網羅している。人名・地名等の固有名詞にはじまり、書簡の送受信情報や草稿の執筆・修正履歴など、テキストを相手に研

究を行おうとする場合に必要なのがXMLのタグとして利用できるようになっており、これを容易に利用するための環境も国際的なコミュニティにより着実に構築・提供され、アップデートが続けられている。上述のIIIFとの連携も各地で取り組まれており、画像・テキスト双方へのアノテーションも実現されている¹¹⁾。

地図、あるいは座標情報に対するアノテーションも考えられるだろう。これにも様々なフォーマットがある。他にも、コンテンツに応じてアノテーションのフォーマットが様々な容易されている。3Dへのアノテーションも近年は注目されつつあるが¹²⁾、これはまだ規格としては十分に固まっていないため、今後の動向に注目する必要がある。

ここで対象としているのはあくまでもレイヤーとしてのアノテーションであり、何らかのフォーマットに限定されるものではない。いずれのフォーマットであっても、下位レイヤーであるデジタルコンテンツに付与されるものであり、そして、上位レイヤーであるインターフェイスのレイヤー(DA-L3)へとこのアノテーションを渡して利用者に供されることになる。

3-4 インターフェイスのレイヤー(DA-L3)

DAに対して利用者がアクセスするためにはインターフェイスが必要である。アノテーションを介してデジタルコンテンツにアクセスしたり、そういったものがない場合にはデジタルコンテンツに直接アクセスして自ら解釈することになる。そして、デジタルコンテンツとアノテーションで構成されるデータを利用者の個々の操作とつなげる役割を果たす。そのつながり方には様々なものがある。たとえば以下のようなものが考えられる。

1. 表示／閲覧
2. 検索
3. 分析・視覚化
4. 文脈化・キュレーション

まずはこれらを一ずつつみていこう。表示／閲覧は、コンテンツとアノテーション、あるいはコンテンツのみをどのようにして閲覧しやすい形で提供するか、ということになる。コンピュータやスマートフォンでの利用を想定するのであれば、それぞれに一定の制約を前提として見やすい形を検討することになる。技術的なことは極めて重要だが、一方で、何を見やすくすべきか、ということも議論の対象となる。IIIFであればいくつかのViewerの中から選択できるが、既存のもので目的を十分に達成できるのかどうか、といったことも検討の必要があるかもしれない。結果として、新たなViewerが開発提供されることもあれば、既存のViewerが改良されて利用に供されることもある。IIIF対応コンテンツに限らず、あらゆるコンテンツにおいてこの点は様々な立場からの検討が必要である。

検索についても様々な観点があり得る。自らの組織で所有・公開するコンテンツの検索だけでも、表記揺れへの対応や検索を通じたコンテンツの探索を容易かつ効果的にする仕組みなど、検討事項は様々だが、さらに、余所のコンテンツとの横断検索も検討の余地がある。日本では現在ジャパンサーチがこの点について精力的に活動しているが、そこに提供したコンテンツの情報(メタデータ)がどのように検索されればよいのか、あちらで検索してこちらにたどり着いたときに導線をどう設計するか、といったことは、検討する意義が大いにある。さらに、多様なアノテーションも検索対象とできるなら、その可能性は大幅に広がることだろう。

分析・視覚化は、表示／閲覧の一部であると言えるが、コンテンツ群への俯瞰的な把握やより深い理解を志向する取組みである。具体的には、コンテンツを地図上にマッピングしたり、コンテンツを数量化してグラフで表示するもの等が典型的である。テキストデータや深いテキストアノテーションが用意されているものであれば、様々なコンテンツの特徴や傾向を分析して視覚化することが可能だが、そうしたものがなくても、メタデータを対象としてコンテンツ群全体の傾向を把握できるようにするといったことも十分に有用である。近年は、画像分析技術の高度化が進み、画像を対象とした分析と

視覚化も徐々に行われるようになってきている。

文脈化・キュレーションは、何らかの文脈に沿ってコンテンツ同士のつながりを提示するものである。このつながりは、説明の文章であったり、グラフであったり、Webのリンクや1画面中のデザイン、動画によるものなど、様々な形で提示され得るものである。自らのコンテンツを対象として行うことにも十分な意義があるが、IIIF対応コンテンツであれば、世界中のサイトのコンテンツを対象とした文脈化が可能になる。コンテンツがただ提示されている状態では、そのコンテンツに詳しい人でなければ意義を見いだすことはなかなか容易なことではないが、複数のコンテンツにつながりがあり、そのつながりのなかで意義を理解できるようにすると、理解度は深まり、利用者の関心も高まることだろう。このようなことは、DAでなくても図書館や博物館・美術館等でこれまで様々に行われてきたことであり、そういった営みと対比することでその可能性と制約が見えてくることもあるだろう。

インターフェイスのレイヤーでは、技術的には前出のIIIFが重要な役割を担いつつある。IIIFはデータフォーマットを規定しているものの、保存や共有を目的としたものではなく、相互運用を主眼とするものであり、むしろこのレイヤーに親和性が高い。Web annotationという国際標準規格に準拠してWebのマルチメディアコンテンツを相互に自由に利活用できる規格を目指しているだけでなく、それらを用いたインターフェイスの開発も活発である。上述のIIIF Curation PlatformはIIIFに特化された代表的な例だが、特定のコンテンツにより特化してIIIFの利活用を実現したものや、逆に、Drupal¹³⁾やOmeka¹⁴⁾等のWebコンテンツマネジメントシステムにIIIFを導入して、既存の優れたインターフェイスの有用性をより高めようとする例もある。いずれにおいても、IIIFが、より広い文脈化を実現するための鍵として大きな期待を背負っていることがうかがえる。

このインターフェイスのレイヤーは、下位のレイヤーの仕組みがなんであれ、何らかの形でコンテンツ(DA-L1)やアノテーション(DA-L2)を受け取ってインターフェイスとして提示する役割を担っている。そして、これを社会

につなげるのが、さらに上位のレイヤーであるコミュニケーションのレイヤー(DA-L4)である。

3-5 コミュニケーションのレイヤー(DA-L4)

コミュニケーションのレイヤーは、DAのあらゆる側面を関係者や利用者、そして広く社会に対して伝える(これはアウトリーチと呼ばれる活動を含む)とともに、そこからのフィードバックを受け、場合によってはそれに基づいて改善をも行う活動全般を指す。DAに限らず多くの取組みには本来そのような活動が必要であり、DAの固有性を考慮せずともこのレイヤーは十分に成立する。しかし一方で、どのような取組みにも固有の側面があり、それに応じてコミュニケーションの必要性や在り方は少しずつ異なるものになる。そのような観点から、DAにおけるコミュニケーションのレイヤーについて検討してみよう。

コミュニケーションのレイヤーにおいてDAの社会へのアピールを実施する手段としては、Webサイトの作成やSNSでの宣伝、マスメディアを利用した告知などがあるだろう。既存の組織によって構築・運用されるDAの場合には、マスメディアの利用を含め、すでに蓄積されたノウハウがあることも多く、それに従えば相当程度のアピールが可能な場合もあるだろう。そのようなノウハウを持たない任意団体や個人によるDAの場合にはWebやSNSの活用が有効だろう。また、近年ではメタバースなどの仮想空間を活用する道も拓かれてきている。インターネット上には、効果的なアピールを可能とする新たな環境が着々と提供されており、活用の仕方次第でより有効なアピールが可能となるだろう。

また、有名なコンテンツであればそれを丁寧に紹介するだけでも有用なアピールになり得るが、DAのコンテンツの中には、知名度が高くなく、コンテンツ単体、あるいはそのコレクションだけでは価値がわかりにくいものが多い。そのような場合には、単にコンテンツを公開するだけでなく、それを説明する文脈を含めたアピールが有用であり、自らのコンテンツだけでな

く関連する他のコンテンツも含めた横断的な大きな文脈を示すことも選択肢となり得る。

コンテンツに応じたターゲットを意識することも重要である。古典籍や古文書の画像を公開しただけでは読める人が多くないため、広く読んでもらいたいのであれば翻刻や解説が必要なこともあるだろう。そういった付加的なコンテンツやその見せ方については下位レイヤーで提供されるものだが、DA-L4ではそのようにして用意されたものを効果的に提示することを目指すことになる。あるいは、そういったものが用意されていない場合には目的に応じて下位レイヤーで作成することを検討すべきだろう。

DAの場合には、そのコンテンツの活用を活発に行ってくれるユーザへの告知を手厚く行うことが効果的な場合もある。特に古文書や古典籍、古写真等であればそれを扱っている学会・研究会等の研究者コミュニティへの丁寧な連絡は有効だろう。また、下位レイヤーにおける丁寧なメタデータの記述や、TEIやIIIF等の標準的なフォーマットに対応したデータ公開などがあれば、それを強調するような告知も重要である。

このレイヤーにおけるDAにとって重要な課題としては、その意義や必要性についての適切な社会的合意を得るという点がある。本書の読者であれば、DA自身の潜在的可能性について疑いを差し挟むことはないかもしれないが、広く社会全体の合意形成という観点でみた場合、DAの意義は前提として語れるものではなく、むしろそれ自体を説明しなければならない段階である。

たとえば書籍の場合、出版社を通じて刊行すれば専門家や一般の人々にまで届く流通経路とそれに見合う人手と費用が確保されており、さらに、保存と長期的利用に関しても、最終的には国立国会図書館に納本することで制度的に保障される形になっている。書籍の作成、すなわち、執筆・組版・印刷製本は、それぞれが生業として成立し得る分業可能なものとして遂行されている上に、それらのいずれもが、保存と長期的利用について何かを自ら能動的に負担しなくとも、公的資金によって国立国会図書館がそれを実施してくれるのである。これに比べると、DAは、構築・運用から保存や長期的利用

に至るまで、業務と費用に関する社会的コンセンサスがまだ十分ではなく、システムとしても制度的に用意されているものは決して多くはない。したがって、構築・運用・保存・長期的利用等に関する制度作りと社会的合意を今後形成していく必要がある。そして、それが実現するまでは、作成や運営に関わる者が自ら取り組まざるを得ない。すなわち、運営のための費用を継続的に投入すべき価値があることをアピールし、出資者を納得させ続けなければならない。それを達成するための様々な活動もまた、このレイヤーに位置づけられる。

そして、そのようにして告知したものに対しては様々なフィードバックがあるだろう。そのなかには自らの組織や事業、あるいは担当者個人で対応できるようなものもあれば、専門家に依頼しなければならないもの、専門家コミュニティに参画しなければ対応できないものなど、様々なものがあるだろう。このレイヤーで受けたフィードバックは、下位のレイヤーのいずれかにおいて対応することになる。

いずれの場合にも、下位レイヤーにおいて用意されたものを十全に告知すること、そして、それに対するフィードバックを受けることがこのDA-L4の役割であり、これを通じてDAの意義や必要性についての社会的合意を形成するとともにDAをよりよいものにしていくことが期待される。

3-5-1 DA-Code／法・規範・技術標準等

法律や規範、技術標準等は、4つのレイヤーのそれぞれにおいて、それらの在り方を規定する形で深く関わってくる。敢えて述べるまでもないかもしれないが、コンテンツ(DA-L1)、アノテーション(DA-L2)、インターフェイス(DA-L3)、コミュニケーション(DA-L4)のいずれにおいても法律としての著作権や規範としての利用条件はそれぞれに作成から公開・頒布に至るあらゆる局面でその活動を規定する。ここでは、各レイヤーに関わりを持つ法律や規範、技術標準等を総称する枠組みとして、DA-Codeを提起する。

たとえば、著作権法改正により2019年から可能となった「非享受利用」と

「軽微利用」により、DA-L1において、著作権保護期間中のコンテンツに関して内容を享受しないデジタルコンテンツの作成が可能となり、DA-L2においてはコンピュータを用いた様々なアノテーションを自動生成できることとなり、DA-L3においては軽微利用の概念を踏まえた提示ができることになった。これがDA-L4に様々な新たなものをもたらすであろうことは言うまでもないだろう。

あるいは、技術標準についても、上で触れたUnicodeやTEIガイドライン、IIIFに限らず、様々な技術標準がそれぞれのレイヤーでできることを拡張すると同時に制約もすることになる。たとえば、Unicodeに使いたい文字、あるいは文字体系が用意されていなければ、DA-L1の時点から大きな制約を受けることになるが、逆に、すでに用意されていたなら、相当にマニアックな文字、あるいは文字体系であってもコンテンツの作成が可能となる。同様に、DA-L2においては、アノテーションでのそれらの文字・文字体系の利用可能性や、アノテーション自体をそれらで記述できるかどうかの問題となり、DA-L3のインターフェイスではそれらを他の言語・文字体系とあまり変わらないレベルで処理できるのか、あるいは特別な処理が必要になるのか、ということが問題になる。DA-L4はそうした状況を受けた活動をするようになる。特にUnicodeについて留意すべき点は、そこで文字が提供されていなければWeb検索しても実質的には見つからないということになり、知識流通基盤としてのDAの遡上に載ることすら難しい。技術標準は多かれ少なかれ各レイヤーにおいて可能なことを規定してしまうことになる。

このように、DA-Codeは、すべてのレイヤーにおいてDAを形作る基盤となるものである。本書を読むような方々には自明のことだが、DAをよりよいものとしていくためにはDA-Codeへの関わりは不可欠である。法制度だけでなく、技術標準に関しても対応が必要な場合があり、それを調査検討し、必要に応じて新設や改訂などを提案できる仕組みを作っていくことが日本のDAの発展において今後は重要になっていくだろう。

4 DAのレイヤー構造におけるDH

4-1 DHの概要

4-1-1 DHのこれまでの流れ

DHは、人文学においてデジタル技術を応用することに関する研究領域を指すものであり、一般に、1940年代にロベルト・ブサ神父が始めたトマス・アクィナスの電子索引が嚆矢とされる。この動向は、デジタル技術の発展とともに形と呼称を変えていき、2004年に刊行されたDHの総合的な入門書A Companion to Digital Humanities¹⁵⁾において用いられた「Digital Humanities」という名称が国際的な学会連合やその学術大会の名称にも用いられるようになり、国際的に広く推進されるようになる¹⁶⁾。

この時期のデジタル研究活動のための環境整備という観点で興味深いのは、1883年に設立され会員数2万人を超える米国の文学・語学の学会であるMLA(Modern Language Association)が2000年にはデジタル研究活動を評価するためのガイドライン¹⁷⁾を制定している。上述の入門書の3名の編集者も文学研究者であり、米国では文学研究が中心となってDigital Humanitiesの動向が支えられていくことになる。

一方、この動向に呼応した学術政策がこの頃から展開されることになる。米国政府の人文学研究助成機関NEHでは2008年にOffice of Digital Humanitiesを設置して専門的に助成を行うこととなり、欧州ではESFRI Roadmap(研究インフラに関する欧州戦略フォーラム)において人文学のための研究データインフラ事業DARIAH(Digital Research Infrastructure for the Arts and the Humanities)が2006年にリストアップされ、2008年には準備フェーズが始まり、2014年には欧州研究インフラコンソーシアム(ERIC)の一環として正式に設立される。こうした動向も含みつつ、様々な事業や研究プロジェクトが相互に連携しつつ幅広く推進されてきている。

日本国内でも1980年代終わりには情報処理学会に人文科学とコンピュータ研究会¹⁸⁾や情報知識学会¹⁹⁾が設立され、その後もこのテーマを扱う学

会・研究会が様々に設立され、着々と研究が進められてきている。2012年には日本デジタル・ヒューマニティーズ学会²⁰⁾が設立されて国際DH学会連合(ADHO, Alliance of Digital Humanities Organizations)に加盟し、国際的な潮流の中に日本の活動も位置づけられるようになり、その後、2022年にはオンラインではあるが東京大学が主催してアジア初のADHOの年次国際学術大会が開催された²¹⁾。この大会では、AI技術を応用した研究が目立ったものの、人文学研究者が自らデータを構造化することに関する発表や、そのようにして作られた歴史史料や文学作品等のテキストデータを人文学の文脈でデータマイニングをするなど、現在のデジタル技術を人文学資料に適用することで得られる新たな成果が様々な形で発表されていた。人文学と情報学の研究者による共同研究だけでなく、2000年代に広まった欧米のDHの教育課程²²⁾を経て人文学のコンテキストでプログラミングやデータ構築を自ら実施できるDH研究者達による発表も多く見られた。

日本でのこうした動向への学術政策的な対応としては、文部科学省大規模学術フロンティア促進事業の「日本語の歴史的典籍の国際共同研究ネットワーク構築計画」²³⁾が、国文学研究資料館によって推進されており、大規模な日本語歴史的典籍のデジタル画像リポジトリを中心としたデジタル研究資源インフラが構築されつつある。また、2018年に開始された日本学術振興会の人文学・社会科学データインフラストラクチャー構築推進事業²⁴⁾では、デジタル媒体を前提とした人文学の研究データインフラの整備が徐々に進められつつある。

4-1-2 DHにおける研究の状況

DHにおける研究は、その資料と研究手法の関係から、構築系・共有系・解析系に大別することができる。構築系は、データの構築手法に関する研究であり、近年の動向としては、テキストエンコーディング、文字エンコーディングなどがよくみられ、3D学術編集版に関する議論も始まっている。共有系は、データ共有に関するものであり、データ公開のための手法や手続

き論、法制度、技術などが近年はよくみられる。解析系には、データマイニング・テキストマイニング等を通じて人文学の文脈での分析を行うものが以前からよく見られるが、手書き文字認識(HTR)²⁵⁾やくずし字OCR²⁶⁾のように画像を解析して必要なデータを抽出しようとするものなどが近年は増えてきている。さらに、最近では、このようなデジタル技術を直接適用する研究以外にも、デジタル技術を適用すること自体の意味やそれが社会にもたらす影響、とりわけ多様性と平等の問題への関わりを扱う研究が国際的なDHのコミュニティにおいて広まりつつある。

4-1-3 DHにおいて個別分野をつなぐ理念的背景

DHは人文学を対象とする領域だが、人文学と一口に言っても、そこに含まれる歴史学・文学・哲学・言語学・社会学・心理学・人類学などのそれぞれの分野には固有の方法論があり、通常の研究活動においては相互に接点を持つことすらそれほど容易なことではない。上記のようなDHにおける研究の多くは、人文学の各分野においてデジタル技術を応用する研究活動であり、それぞれの個別分野で議論すれば十分であるということにもなりかねない。その壁を乗り越えるための理念的背景としてDHにおいて広く共有されているのが「方法論の共有地(Methodological Commons)²⁷⁾」(図3)という考え方である。人文学に含まれる様々な分野はそれぞれに独自の方法論を有しており、それをを用いて研究成果を出すことになるが、DHにおいてはそれらの方法論に対してデジタル技術を適用することになる。その際、テキスト分析や画像分析、マッピング、データ構築等々、デジタル技術としては同じものを用いることになる。方法論はそれぞれに異なっているとしても、デジタル技術を適用する場面においては共通の場での議論が可能であり、それを共有地と位置づけた上で、各分野が相互に協力しつつ研究を展開し、成果を出していくとしたら、それはその技術の応用という点で各分野がそれぞれに益を得るだけでなく、その際に相互に見えてくる方法論の違いは、相互の対話可能性への意識とともに、自らの方法論的反省の場が提供されることにもなり

得る。DHはそのようにして人文学全体への広がりを持つべくADHOを中心として活動を続けてきている。

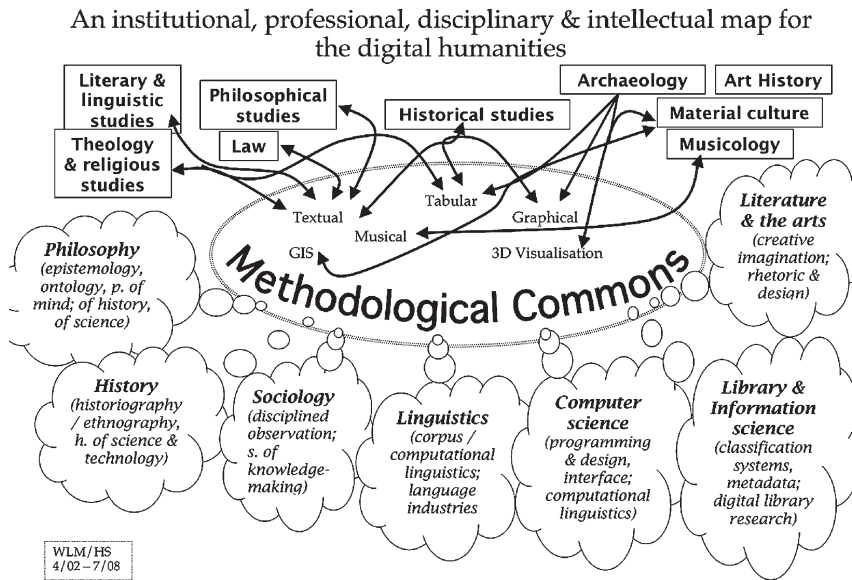


図3 方法論の共有地(Methodological Commons)

4-2 DAとDHの接点

以上のようなDHがDAとどのような関係であり得るかについて検討してみよう。DAには人文学の研究対象となる資料をデジタル化したものが多く含まれる。人文学には上述のように幅広い分野が含まれるが、それぞれにおいて何らかの形でDAに研究資料を求めるようになりつつある。図書資料や古文書、古記録・古典籍等はその典型であり、絵画をはじめとする美術品や工芸品、出土品なども含まれる。アンケート調査等の結果もある。そうした資料はアーカイブズ・ミュージアム・図書館等では着々とデジタル化が進められており、人文学の教育や研究において広く活用されるようになってきている。近年は、伝統的な人文学においてもデジタル化資料を用いる例は多く見られ、資料だけでなく、研究成果としての論文も日本であればJ-STAGEや各大学

機関リポジトリを通じてデジタル化と公開が着々と進められている。こうした論文リポジトリをもDAに含むと考えるなら人文学はDAの有力なステイクホルダーとしてもはや抜き差しならぬ関係になっている。なお、人文学として見た場合には幅広く深くDAのコンテンツが関連することになるが、個別の研究者に着目した場合には、自らの専門分野に近ければ近いほど深い関わりを持つが、逆に専門から離れれば離れるほど深い関係ではなくなっていくという点にも留意しておきたい²⁸⁾。

4-3 DAのレイヤー構造におけるDH

それでは、DHの営みをDAのレイヤー構造を手がかりとして検討してみよう。DA-L1においては、研究利用に耐え得るデジタルコンテンツを要求することになる。これは、再現性の高さを実現するためにより高精度なものが求められると考えてしまいがちだが、一方で、研究方法論として必要な精度が実現されていればよいという考え方もある。資料に内在する情報は、デジタルデータにした時点で何らかの捨象が生じてしまうのであり、特にテキストデータの場合には文字の形や色等の微細な違いも十分に再現できない場合が多い。それにも関わらずテキストデータでよしとすることがあるという状況は、許容される捨象の範囲を考慮する上で重要だろう。平面画像であれば、解像度の必要性は分野や研究手法によって異なり、たとえば、テキスト本文の内容を読み取ればよい分野と、漢文写本の角筆点まで読み取りたい分野では要求される解像度や照明の当て方へのこだわりも大きく異なる。音声や動画の再現度についても分野によって異なる場合があるだろう。立体物を3Dで再現するのであれば、技術的には研究上必要な観察範囲を超えたすべての側面の再現が可能だが、あまりにも微細な情報まで再現しようとするとコスト面での困難に直面するだろう。しかしながら一方で、これまではなかった新たな研究手法が編み出された際に、これまでよりも高い再現度が必要になってしまうことがあり、現在の必要性のみに立脚したコンテンツ作成は将来的な再作成を必要とすることになりがちなため、やはりデジタル化

の際の資料の再現度はコストの許す範囲で可能な限り高い方がよく、何らかの配慮は必要だろう。

DA-L2のアノテーションは、DHにおいてはコンテンツに関する専門知を踏まえた解釈の提示とその共有、そして、さらなる研究の発展のために鍵となるものである。これは、人手で作成されることもあれば、コンピュータ分析によって作成されること、あるいは両者を組み合わせて対処されることもある。画像や音声・動画等に付したアノテーションをWeb上でより有効に機械処理するためには、前出のIIIFが有効である。そして、知識処理しやすいアノテーション作成のためには前出のTEIガイドラインが提供する文書構造のモデルが有用である。あるいは、より高次の知識ネットワーク構築を目指してRDF(Resource Description framework)²⁹⁾に準拠したアノテーションの作成に取り組む研究も取り組まれてきている。

DA-L3のインターフェイスに関してしてみると、DHにおける典型例としては、必要な情報にどのようにアクセスできるようにし、コンピュータによる分析結果をどのように視覚化すべきか、ということについての研究が行われている。前出のIIIFやTEIについても、DA-L2で構築されたアノテーションを適切に提示するための研究開発が様々に進められている。目指すところは資料に対して直感的に容易にアクセスし理解できるようにするということになるだろう。そして、こうした研究を包括する概念としては、たとえばScholarly Primitivesという概念が提唱され議論されている。これは、人文科学研究において採られる手法の原初的なものを突き詰めた時にどのような行為に収斂可能か、ということについての具体性を持った提案であり、DHにおけるアプリケーション開発にも応用される。この語には定訳がまだないのだが、仮に原初的研究手法と訳しておこう。2000年にこの原初的研究手法を提唱したJohn Unsworthがその構成要素として挙げたのは発見・注釈・比較・参照・サンプリング・例示・表現である³⁰⁾。人文学における研究はこれらに拠って行われており、人文学をデジタル媒体上で展開するために必須であるとした。この原初的研究手法は、その後も様々な議論を経て発展し、

2020年には欧州の人文学向けデジタルインフラDARIAHの会議のテーマとして採りあげられている。

DA-L4については、DHに取り組む大学教員であれば大学生・大学院生への教育を通じて、そして、大学教員ならずとも、学会における論文発表や書籍の刊行、マスコミ等での露出などを通じて情報提供をするとともにフィードバックを受け取ることになり、幅広く様々な人々とコミュニケーションをとることになる。近年は、欧米先進国を中心に、オープンアクセス・オープンデータ・オープンソース等と呼ばれる自由な研究資源へのアクセスが求められるようになり、世界各国からそうした資源が提供され、研究活動への敷居を下げることに貢献している。ここでは、人文学の特性をうまく活かした研究データインフラストラクチャーの構築と運用が効果的であり、今後重要なテーマの一つとなっていくだろう。

最後に、DHにおいてDA-Codeにあたる研究としては、法制度や技術標準だけでなく、社会正義に関わる研究もこれに含まれるだろう。前述のように、欧米先進国ではその点についての研究も盛んであり、ジェンダーの不平等や南北問題も含む構造的な格差や差別についての研究も進められている。なお、欧米先進国でのDH関連学会の運営においてはジェンダーバランスの確保についてADHO設立時点から実現されており、単なる議論で終わるものではなく実践も伴った営みとして研究活動が推進されている点も、興味深くかつ参考にすべきだろう。

以上のように、DAのレイヤー構造の枠組みからDHの研究活動をみると、DAが持つ広汎な対象範囲からすれば当然のこととは言え、DHはDAの一部として位置づけ得るものであり、それぞれのレイヤーにおいてDHの活動がDAに資するものであることもみてとれる。

5 DAとDHがもたらし得る未来

DAがデジタル時代に即した新たな知識基盤であるなら、デジタル技術を

用いて人間文化を研究しようとするDHがそこに含まれることになるのはある意味で当然のことである。一方で、DHを含む人文学が持つ視座は、DAそれ自体やそれを構成する各レイヤーを問題化し、ラディカルな変革を求めることもある。

現在は総合知という言葉に象徴されるように、複雑化した知の世界において横断的に全体を俯瞰できる優れた解決策が求められるようになってきている。DHはそれを実現する可能性を期待されるものであり、そのためにはより良質なDAを媒介とすることが必要となる。構築者としても利用者としても、DHはDAに貢献し得る存在であり、その貢献によって、DH自体もより高い評価を得られることになる。この相互に有益な関係を持続的に発展させていくことができれば、両者の未来は明るいものとなっていくことだろう。

注

- 1) これについて日本語で読めるものとしては、横山説子「デジタル時代における人文学者の社会的責任」(『欧米圏デジタル・ヒューマニティーズの基礎知識』(文学通信, 2021年) 掲載)を参照されたい。
- 2) https://www.mext.go.jp/content/20200722-mxt_jogai02-100013300_006.pdf(最終アクセス: 2022年9月28日)
- 3) デジタルアーカイブにおけるテキストデータの在り方や構築の考え方については、永崎研宣「人文学のためのテキストデータの構築とは」『人文学のためのテキストデータ構築入門 TEIガイドラインに準拠した取り組みにむけて』(文学通信, 2022)を参照されたい。
- 4) Unicode; The World Standard for Text and Emoji <https://unicode.org/>(最終アクセス: 2022年9月28日)
- 5) ISO/IEC 10646:2017 Information technology — Universal Coded Character Set (UCS) <https://www.iso.org/standard/69119.html>(最終アクセス: 2022年9月28日)
- 6) MUFI: The Medieval Unicode Font Initiative <https://mufi.info/>(最終アクセス: 2022年9月28日)
- 7) デジタルアーカイブにおける学術的な文字コードの課題については、詳しくは、下田正弘・永崎研宣編『デジタル学術空間の作り方 仏教学から提起する次世代人文学のモデル』(文学通信, 2019)における下田正弘・永崎研宣「デジタル学術

空間の作り方—SAT大蔵経テキストデータベース研究会が実現してきたもの」[6-1, Unicodeへの登録]及び王一凡「慧琳撰『一切経音義』の符号化をめぐる」を参照されたい。

- 8) P5: TEIガイドライン <https://tei-c.org/release/doc/tei-p5-doc/ja/html/index.html>(最終アクセス: 2022年9月28日)
なお、日本語による紹介としては一般財団法人人文情報学研究所監修『人文学のためのテキストデータ構築入門』を参照されたい。
- 9) IIIF | International Image Interoperability Framework <https://iiif.io/>(最終アクセス: 2022年9月28日)
なお、日本語での紹介としては、永崎研宣『日本の文化をデジタル世界に伝える』(樹村房, 2019)の第五章を参照されたい。
- 10) IIIF Curation Platform | ROIS-DS 人文学オープンデータ共同利用センター(CODH) <http://codh.rois.ac.jp/icp/>(最終アクセス: 2022年9月28日)
- 11) TEIとIIIFの双方に対応している例としては廣瀬本萬葉集のビューワが挙げられる。永崎研宣, 乾善彦, 菊池信彦, 宮川創, 小川歩美, 堀井洋, 吉賀夏子. 「万葉集伝本研究のためのデジタル基盤構築 廣瀬本『万葉集』の構造化とビューワの開発」. 研究報告人文科学とコンピュータ(CH), 2021-CH-125, no. 2 (2021年2月6日): 1-7.
- 12) Schriebman, Susan, Costas Papadopoulos. 「Textuality in 3D: Three-Dimensional (Re) Constructions as Digital Scholarly Editions」. International Journal of Digital Humanities 1, no. 2 (2019年7月1日): 221-33. <https://doi.org/10.1007/s42803-019-00024-6>.
- 13) Drupal - Open Source CMS | Drupal.org <https://www.drupal.org/>(最終アクセス: 2022年9月28日)
- 14) Omeka <https://omeka.org/>(最終アクセス: 2022年9月28日)
- 15) Edited by Susan Schriebman, Ray Siemens, and John Unsworth, A Companion to Digital Humanities, Blackwell Publishing, 2004, <https://companions.digitalhumanities.org/DH/>(最終アクセス: 2022年9月28日)
- 16) 世界各地で開催されるDH学会の発表論文の横断検索を提供するサイトが運用されており、関心がある方はぜひ参照されたい。The Index of Digital Humanities Conferences, <https://dh-abstracts.library.virginia.edu/>(最終アクセス: 2022年9月28日)
- 17) <https://www.mla.org/About-Us/Governance/Committees/Committee-Listings/Professional-Issues/Committee-on-Information-Technology/Guidelines-for-Evaluating-Work-in-Digital-Humanities-and-Digital-Media> このガイドラインは2012年に改訂されている。なお、本Web頁では初版の承認は2000年と書いてあるものの、その公開については2002年とする資料が散見される。(最終アクセス: 2023年5月31日)

- 18) 情報処理学会人文科学とコンピュータ研究会, <http://jinmoncom.jp/>(最終アクセス: 2022年9月28日)
- 19) 情報知識学会, <https://www.jsik.jp/>(最終アクセス: 2022年9月28日)
- 20) 日本デジタル・ヒューマニティーズ学会, <https://www.jadh.org/>(最終アクセス: 2022年9月28日)
- 21) DH2022, <https://dh2022.adho.org/>(最終アクセス: 2022年9月28日)
- 22) 欧州におけるDH教育に関しては、長野壮一「フランスのDH——スタンダール大学における教育と研究」『欧米圏デジタル・ヒューマニティーズの基礎知識』(文学通信, 2021)にてフランスの事例が紹介されている。また、各国のDH教育課程についてはたとえば以下のURLにリストされている。 <https://github.com/dh-notes/dhnotes/blob/master/pages/dh-programs.md> (最終アクセス: 2022年9月28日)
- 23) 日本語の歴史的典籍の国際共同研究ネットワーク構築計画, <https://www.nijl.ac.jp/pages/cijproject/plans.html> (最終アクセス: 2022年9月28日)
- 24) 人文学・社会科学データインフラストラクチャー構築推進事業, <https://www.jsp.go.jp/j-di/torikumi.html> (最終アクセス: 2022年9月28日)
- 25) Transkribus | AI powered Handwritten Text Recognition, <https://readcoop.eu/transkribus/> (最終アクセス: 2022年9月28日)
- 26) くずし字OCR(AIくずし字認識), <http://codh.rois.ac.jp/char-shape/OCR/>(最終アクセス: 2022年9月28日)
- 27) Willard McCarty, Harold Short, *METHODOLOGIES*, Pisa, April 2002, <https://eadh.org/methodologies>. なお、日本語訳と簡単な解説が以下のブログ記事に掲載されているのであわせて参照されたい。「Methodological Commons: デジタル人文学で昔から定番の話」 <https://digitalnagasaki.hatenablog.com/entry/2020/12/20/182659>
- 28) DAコンテンツと専門分野の関係に関しては以下の論文で整理したことがあるので参照されたい。永崎研宣. (2007). 人文科学のためのデジタル・アーカイブにおけるステイクホルダー—仏教文献デジタル・アーカイブを手掛かりとして—. じんもんこん2007論文集, 2017, 347–354.
- 29) Resource Description Framework (RDF), <https://www.w3.org/RDF/>(最終アクセス: 2022年9月28日)
- 30) John Unsworth, "Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this?", Part of a symposium on "Humanities Computing: formal methods, experimental practice" sponsored by King's College, London, May 13, 2000. <https://johnunsworth.name/Kings.5-00/primitives.html> (最終アクセス: 2022年9月28日)