

人文学分野とサイバーインフラストラクチャ ～デジタル・ヒューマニティーズにおける現状と課題～

永崎 研宣*

人文学分野においてもサイバーインフラストラクチャは欠かせないものになりつつある。歴史的には 1940 年代から取り組みが行われ続けてきたものであり、人文学研究者自身が積極的に取り組んでいくことがこれまで以上に重要になってきている。国際的にはすでに欧米の関連学会によって設立された ADHO を中心として大きな組織的枠組みができあがっており、我国もそこに参画している。個別のテーマとしては知的所有権をはじめ様々な事項に配慮しながら進めていかなければならない。また、評価や教育の体制等については、特に米国では人文学の側で新しい枠組みの構築に向けての取り組みが進みつつある。そこでは、我国の人文学の貢献の余地も大いにあり得る。

キーワード：人文学，デジタル・ヒューマニティーズ，TEI，ADHO，知的所有権，クリエイティブ・コモンズ，Unicode，Europeana

1. はじめに

紀元前三世紀頃のアレクサンドリアにおけるムセイオンと言えば、当時世界最大の図書館を併設した学芸の一大中心地として有名である。欧州科学財団による人文学研究向けの研究インフラに関する報告書¹⁾において、注目に値する最初の重要なインフラとして挙げられているのはそれである。中国においても高祖劉邦の時代に石渠閣等の図書館が整備された。活版印刷術が出てくるよりもずっと以前から、研究を支えるインフラは着実に展開されてきていたことをまずは押さえておきたい。

さて、デジタルメディアによる情報流通がもはやインフラに匹敵するものとなった昨今、人間の活動に関わる様々な情報が、デジタルデータとして入手でき、デジタルツールを用いて様々な加工でき、それを通じてさらなる価値を生み出し、さらにはそうしたものを基盤として研究を行うというスタイルも徐々に定着しつつある。たとえば米国ではオバマ大統領の主導のもと、効率的なアクセスとそこから創造的な成果が生み出されることを目指し、公衆の参加と協働を促すべく、Data.gov²⁾を通じて連邦政府機関が作成する国勢、気象、経済などのデータセットを提供している。そこでは、単なる統計情報の集計結果だけではなく、元になったデータやツール、地理情報等も提供されており、利用者が自由かつ容易に処理や分析を行えるようになっている。たとえば、ニューヨークタイムズにおいてこうした情報が活用され、美しく見やすく各種統計が公開されていたことが話題になったのは記憶に新しい³⁾。この種のデータは各国政府や各種機関でも様々な公開しており、さらに、それらのデータを集め、容易に視覚化できるツールとして Google Public Data Explorer⁴⁾が提供されている。データが作成され、流通し、再利用されてそれがさらに価値を生

み出していくというスパイラルはこの種のデータにおいてはかなりの程度まで確立されつつある。自由な再利用が知的創造のサイクルを高める事例としてまず想起されるのは Linux だろう。Linux がフリーソフトとして自由に利用可能な形で公開されたことで世界中の開発者が智慧を寄せ合うようになり、やがて主要コンピュータ企業もその開発に参画するようになり、無償で自由に利用できるソフトウェアが今やスパコントップ 10 のすべてのオペレーティングシステムを占めるという状況に至った⁴⁾。そのような自生的とも言える知的生産のサイクルが、人文社会科学分野においても確立されていくことが期待されていると考えてよいだろう。

さて、Data.gov で公開されるような情報の多くは比較的数値化・構造化しやすいものが多いように思われるが、一方で、人文学の多くの分野で基礎データとなるいわゆる文化資料は、利活用しやすい形での数値化や構造化がやや難しいという面がある。こういったデータをどのようにしてより適切にデジタル化・共有・処理・再利用するかということについては、すでに 1940 年代後半、Roberto Busa 神父により中世の神学者トマス・アキナスの著作の電子レクシコン構築が開始されており、その後現在に至るまで、我国も含め世界中で様々な議論され実践されてきた。とりわけ、2000 年代に入るとデジタル・ヒューマニティーズと名付けられた一つの大きな傘の下でさらに大きく展開されつつある。本稿では、情報技術の特性をより生かした形での安定的な情報共有基盤をサイバーインフラと呼ぶことにした上で、特に人文学分野におけるサイバーインフラをめぐる状況をみていくべく、国内外の一連の関連動向を「DH 分野」と仮に名付け⁵⁾、以下に検討を進めていきたい。

2. 歴史的動向

2.1 黎明期～Web 登場以前

上述のように、1940 年代、いわゆるコンピュータの黎明期にはすでに人文学におけるデジタル化は始まっていた。その後、様々な試行錯誤が重ねられる中で、1973 年、英国

*ながさき きよのり 一般財団法人人文情報学研究所
〒113-0033 東京都文京区本郷 5-26-4-11F
Tel. 03-6801-8411 (原稿受領 2013.7.3)

において設立された言語学と文学におけるコンピュータの応用に関する学会、Association for Literary and Linguistics Computing (ALLC, 現在の EADH)⁵⁾へとつながっていく。米国でも、少し遅れて 1978 年、人文学におけるコンピュータの応用に関する学会として Association for Computers and the Humanities (ACH)⁶⁾が設立された。当時はパーソナルコンピュータがなかったため、コンピュータを利用するというだけでもかなりの苦労があったことが想定されるが、大型計算機を有する大学を中心に研究が展開されていたようである。また、我国においても、計量国語学会が 1957 年に設立され、当初は手計算による研究発表が行われていたものの、1950 年代末頃にはコンピュータを利用した研究発表も登場している。そして、1966 年には国立国語研究所に大型計算機が導入され、コンピュータを利用した日本語の研究も徐々に行われるようになっていた。1970 年代には国文学研究資料館や国立民族学博物館等も創設され、当時は目録程度であったものの、現在の文化資料データベースの基盤を成すようなデータベース群の作成が組織的に進められるようになっていったのである。

1980 年代に入ると、すでに 77 年に登場していた Apple II に加えて 81 年に IBM PC が登場し、個人がコンピュータを利用して様々な研究を行うための環境が人文系研究者にも比較的容易に提供されることとなった。そこでは様々な試行錯誤が行われた模様であり⁷⁾、結果として、欧米諸国では、共通の技術的な基盤の必要性に至る。そして、上述の ALLC, ACH, 及び、1960 年代に発足していた Association for Computational Linguistics が中心となり、欧州、米国、カナダ等の研究助成団体の支援を受けつつ、1987 年、人文科学のためのテキストデジタル化のガイドラインの作成・提供に向けたプロジェクト TEI (Text Encoding Initiative) が開始されたのである。最初の会合が行われた場所にちなんで「ポキプシー・プリンシプル」と呼ばれる基本原則は以下の通りである。

「テキスト電子化のガイドラインへの準備」⁸⁾

1987 年 11 月 13 日、ニューヨーク、ポキプシー。

1. ガイドラインは、人文学研究におけるデータ交換のための標準的な形式を提供することを目指す。
2. ガイドラインは、同じ形式でテキストの電子化をするための原理を提案することも目指す。
3. ガイドラインは、以下のことをすべきである。
 - 1 形式に関して推奨される構文を定義する。
 - 2 テキスト電子化のスキーマの記述に関するメタ言語を定義する。
 - 3 散文とメタ言語の双方において新しい形式と既存の代表的なスキーマを表現する。
4. ガイドラインは、様々なアプリケーションに適したコーディングの規則を提案するべきであろう。
5. ガイドラインには、そのフォーマットにおいて新しいテキストを電子化するための最小限の規則が入っているべきである。

6. ガイドラインは、以下の小委員会によって起草され、主要なスポンサー組織の代表による運営委員会によってまとめられる。

- 1 テキスト記述
 - 2 テキスト表現
 - 3 テキスト解釈と分析
 - 4 メタ言語定義と、既存・新規のスキーマの記述。
7. 既存の標準規格との互換性は可能な限り維持されるだろう。
8. 多くのテキストアーカイブズは、原則として、交換形式としてのそれらの機能に関して、そのガイドラインを支持することに賛成した。我々は、この交換を効率化するためのツールの開発を支援するよう、支援組織に働きかける。
9. 既存の機械可読なテキストを新しい形式に変換することとは、それらの規則を新しい形式の構文に翻訳するということを意味しており、まだ電子化されていない情報の追加に関して何か要求されるということはない。

以上が 1980 年代後半、欧米において DH 分野に関わる人文学研究者の現状認識に基づく解決策であった。ここには、標準規格がないことの問題、テキストのメタ情報についての自覚が十分でなかったという自己認識、それらを決定していくための枠組みの必要性などが見て取れる。とはいえ、この時点ではメタ言語の扱いが色々な意味で容易ではなく、SGML (Standard Generalized Markup Language) を用いて試みられ、いくつかの重要な電子テキストアーカイブズにおいて採用されたものの⁹⁾、当時は SGML の扱いがそれほど容易でなかったため、十分な広がりを見せるには至らなかった。なお、この頃、日本においても TEI に関する若干の試行錯誤が行われた模様だが¹⁰⁾、そもそもまだ共通の文字コードが用いられていなかった時代であったため、テキストデータの互換性にかかなりの問題があり、国際的に共通利用できるアプリケーションの開発が極めて困難であったこと、さらに、インターネットもまだ十分に利用できなかったためデータを実際に交換することも容易ではなく、結果として、この時点では日本ではほとんど普及することはなかった。

一方、我国では、1989 年、人文学のデジタル化に関連する学会・研究会が相次いで設立される。情報処理学会人文科学とコンピュータ研究会 (以下、SIGCH)¹⁰⁾、情報知識学会¹¹⁾、アート・ドキュメンテーション学会¹²⁾の前身となる研究会である。それぞれ現在に至るまで人文学におけるデジタル化の議論の場としても機能してきているが、とりわけ SIGCH はこの方面の一つの核として機能することになる。

2.2 1990 年代・Web の登場

さて、1990 年代は Web の登場と発展の時期であり、新たな関心に基づく様々な研究が展開されるようになる。欧米では、ALLC と ACH が学術大会を共同開催するように

なり、後の DH の世界連合、Alliance of Digital Humanities Organizations (ADHO) の基盤の形成が始まる一方、特に 90 年代後半に行われた XML (Extensible Markup Language) の開発において TEI の主要メンバー¹³⁾が大きな貢献を行った。とりわけ、XML の仕様の一環である Xpointer が、TEI によって作成されてきたガイドラインの仕様から採り入れられたことは人文学研究の知見が情報工学に対して貢献した例として現在でもよくとりあげられる¹⁴⁾。

このころ、我国においても英語コーパス学会、情報文化学会、情報考古学会等、DH 分野の中でも、人文学の個別分野においてデジタル技術を応用することに関わる学会が相次いで設立される一方、SIGCH では 99 年には「じんもんこんシンポジウム」を査読付プロシーディングという形で開催するようになる。

2.3 2000 年代・Digital Humanities の出現

2000 年代に入ると、いよいよ状況は一気に展開する。人文学研究者が時折議論し想定するサイバー空間における様々な構想や枠組みはそれまで絵に描いた餅でしかなかった。それを実際に試行できる場が用意されるようになったのがこの頃である。当時 Web2.0 と呼ばれた、コラボレーションを軸とするサイバー空間での情報交換の枠組みは人文学のデジタル化においても大きなインパクトを与えることになる。同時に、それまで Humanities Computing あるいは Computers and the Humanities 等と呼ばれていた一連の研究活動が、Digital Humanities という呼称に収斂されていくのである。Humanities、すなわち人文学の一環であるということ強く意識したこの呼称は、急速に関連研究者の間に広まっていったようである。それが如何に急速だったかということは Google books ngram viewer¹⁵⁾において 2000 年頃を境に Digital Humanities という語の用例が急増しているところに端的に表れている。このころ、John Unsworth, Ray Siemens, Susan Schreibman によって共同編集された人文学デジタル化に関する包括的な入門書に Digital Humanities という名称が付けられた¹⁶⁾。同時期に議論されつつあった ALLC と ACH の連合組織もやはり Digital Humanities を冠する組織となり、2006 年にパリ・ソルボンヌ大学で開催された ALLC と ACH 共催の年次国際会議にも Digital Humanities の名が付けられるに至り¹⁷⁾、この名称は決定的なものとなった。

Digital Humanities と呼ばれる以前から、DH 分野に関わる学際研究を如何にして評価し、どのようにして発展させていくかということは重要なテーマであって来たが¹⁸⁾、Digital Humanities と自称するようになったことは、すなわち、主流メディアが紙からデジタルへと移行せざるを得ないことがいよいよ人文学者にとっても真に迫ってくるものとなり、同時に、上述のように理論的な構想を実際に試行できる環境が提供されるようになったことで、人文学研究者自身がより主体的にデジタル化研究に取り組むようになるということを意味している。そこでは研究資金をどのようにして確保するかということに始まり、テニユアをど

うするかということに至るまで、人文学に足場を置きつつ学際的な広がり確保しようとする一つの研究分野としてのフルセットを用意するための場の重要性がますます重視されるようになったのである。米国では 2006 年、全米人文学基金が、Digital Humanities の研究プロジェクトを対象とした研究支援体制を整備し¹⁹⁾、明確に Digital Humanities の本格的な創成へと舵を切り、一方、欧州では欧州全体での研究インフラの整備という文脈のなかで DARIAH (Digital Research Infrastructure for the Arts and the Humanities) の準備が 2005 年に開始された。また、DH 分野の教育研究機関の連携を促進する CenterNet²⁰⁾も 2007 年に始まる等、研究手法だけでなく、研究機関や研究組織を形成していくことも志向されていた。コミュニティを形成し、それを通じて共同研究を展開しつつ相互評価できる体制を確立することが急務であったと言えるだろう。さらにその一方で、2008 年には Google Books や Internet Archive のデジタル化コンテンツを含む大規模協働リポジトリである Hathitrust²¹⁾が設立され、2012 年には 1000 万冊を超える図書を対象としてサービスを開発・提供している。特に注目されるのは、著作権保護期間にある著作物であっても検索や解析ができるサービスである。これは、人が直接閲覧(消費)することなくコンピュータによる解析結果のみが提供される仕組みとなっており「非消費的(non-consumptive)研究」と呼ばれている。

2000 年代には、上述の TEI ガイドラインの XML 準拠も行われた。すなわち、SGML に準拠して 1994 年に作成・公開された P3 ガイドラインを XML に切り替える形で、2002 年、P4 が公開されたのである。しかしながら、TEI ガイドラインは自然文にマークアップを行うことを主眼としていたため、XML へと移行したことで、自然文の要素において木構造を許容するかどうかという問題に突き当たることになる。さらに、XML の高機能なスキーマの扱いにも十分に対応していなかったことや、過去の様々な問題を解決すべく改良を進め、抜本的な改訂を行い XML に十分に対応した P5 ガイドラインを 2009 年に公表する。ここではスキーマが用途に応じてモジュール化されているなど、500 を超えたエレメントを使いやすくするための工夫が行われている。また、この際には、外字を XML エレメントとして記述するための gaiji モジュールも導入された。

日本では 2000 年代、21 世紀 COE プログラムの助成金による大規模研究プロジェクトが幅広く立ち上げられ、その中で DH 分野に関わるプロジェクトがいくつも立ち上がった²²⁾。さらにこれに続く Global COE プログラムにおいても、数は減ったものの、やはり DH 分野に関わるプロジェクトが複数展開された。これらの成果は必ずしも芳しいものばかりではなかったが、一定の成果を残しつつ、この分野に関わる研究者を育成できたことは僥倖である。また、国立国会図書館をはじめ、様々な公共機関でパブリックドメインとなった資料をデジタル化して公開する動きが徐々に本格化していったのも 2000 年代であった。

2.4 2010年代・DH分野の広がり

2010年代に入ると、DH分野でのネットワーク化の流れはより広い規模での大きなネットワークの形成へとつながっていく。ADHOの重視するミッションの一つとしての多言語・多文化主義への注力の成果として、豪州を中心にオーストラリア地域のDH研究者が集い、Australasian Association for Digital Humanities (aaDH)²³⁾を設立、ADHOの構成組織の一つとなった。また、日本においても主にALLCの協力の下、国際的な文脈におけるDHと日本のそれまでの人文学におけるデジタル化研究との橋渡しをすべくコミュニティの形成が進んでいたものが、日本デジタル・ヒューマニティーズ学会(JADH)²⁴⁾として発足するに至り、2012年、ハンブルクでのDH国際会議の期間中に開催された運営委員会において、ADHOへの加入が承認され、構成組織の一つとなった。さらに、ドイツ語圏やスペイン語圏、イタリア語圏に関してもそれぞれDH分野の学会が設立されるなど、文化的多様性を如何にして保ちつつDH分野を展開していくかという議論と実践は世界的にますます広がりを見せつつある。我国においても、DH分野の研究としてはSIGCHをはじめとする関連学研究会を中心として着実に展開してきている一方で、サイバーインフラとしては国会図書館のデジタル化資料公開や国立国会図書館サーチ²⁵⁾、人間文化研究機構の研究資源共有化システム²⁶⁾等が充実しつつあり、さらに、MALUI連携²⁷⁾等を通じて関連組織間の連携も図られつつある。

TEIに関して付言しておく、2011年末、さらに大きな変革を迎える。それまでは、木構造を前提として構造化されたテキストに対してマークアップするというモデルが基本であり、版面・記述面の様態、すなわち、紙面上のどの箇所にどのような文字がどのように書かれたか、といったことはあくまでも付随的なものと位置づけられていたが、この時に公表されたver. 2においては、テキストの構造として木構造を採り入れなくとも版面・記述面上のテキストの様態をそのままにデジタル化できるという枠組みが正式に採り入れることとなった。これはTEI内の手書き資料のデジタル化に関する研究グループManuscript SIGが長らく検討してきた、手書き資料の変遷過程を適切に記述するための枠組みに対応するための措置であり、TEIの応用範囲を飛躍的に拡張するものでもあった。また、これに加えて、ISO TC37/SC4²⁸⁾との連携も含めた言語資源としての扱いについての議論やCIDOC-CRM²⁹⁾との連携に関する議論など、テキスト資料を軸としつつ様々な方向への展開が進められつつある^{注4)}。

3. DH分野における近年のトピック

さて、DH分野、あるいは人文学デジタル化に関わる全体的な動向は以上のように展開してきたが、それでは個別のトピックについてはどのように扱われているのか。以下、それに関する近年の動向を見てみよう。

3.1 評価の問題

ここで見てきたように、DH分野においては、ただ便利なツールや使いやすいデジタル化資料が出てくるのを待っているだけでなく、その仕様策定や、設計、さらには作成そのものに参画していくことで人文学研究者自身の必要性を様々な形で反映させていくという流れが、すでに本格化しつつある。しかしそこには、評価の問題という大きな壁が立ちはだかっている。研究者のキャリアにおいては、成果に対してかけた労力に見合う評価が得られないことには取り組むことも継続することも困難である。しかし、人文学研究者が情報技術に関わる事柄について注力したとしても人文学の成果として評価されることは極めて稀である。そして、人文学研究者が取り組むような技術的テーマでは、情報学・情報工学等の分野において評価されることも容易ではない。また、情報技術に関わる専門家についても、人文学研究者の個別のニーズを満たすようなシステム開発などを行ったとしても研究成果として認められることはあまり容易ではないようである。ではビジネスとして成立させれば良いのかと言えば、直接のマーケットは大きいものではなく、また、人文学研究者の個別分野毎のニーズに踏み込めば踏み込むほどマーケットは小さくなっていくため、これもまた極めて難しい。では共同研究として行えばよいのかと言えば、どのような成果を目指すべきか、という点で齟齬が生じることもあり、なかなか一筋縄ではいかない。このようなことから、DH分野、あるいは人文学デジタル化に関わる研究に取り組んでもかけた労力に見合う評価が行われないのではないかと、という問題は、現在も十分には解決されていないものの、すでに様々な議論されている³⁰⁾。特に、ADHOではこのことは主要な問題の一つと捉えられており、毎年これに関わるセッションがいくつも開催されている。その結果として、人文学あるいは情報学等のいずれかの分野として評価するだけでなくDH分野という学際領域として評価するという形態が一つの有力な選択肢としてある程度広く認知されつつある。ADHO主催の年次国際学術大会をはじめ、様々なDH分野の国際会議が査読付国際会議として開催されており、さらに、様々な査読付学術誌³¹⁾が刊行されつつある等、DH分野として査読を経た成果を生み出せる体制が広がりつつある。なかでも、1986年よりオックスフォード出版局から発行されている査読付学術誌『Literary and Linguistic Computing』(LLC)³²⁾は、現在ではトムソン・ロイターのArts & Humanities Citation Index(R)及びSocial Sciences Citation Index(R)にリストされており、インパクトファクターも付与されるようになっている。我国でも、SIGCHが主催する年次学術大会「人文科学とコンピュータシンポジウム」では査読付プロシーディングを刊行しており、他の関連学会では査読付学術誌を刊行している。このような方向での解決はひとつの有力な解決策と言えるだろう。もちろん、DH分野で評価されるということがすぐに何らかのポスト等につながるかどうかという点は依然として難しいと言わざるを得ないが、これについても、学際領域が世

界的に広く振興されつつある現在では、そのことがプラスに働く局面が多くなっていると考えてよいと思われる。さらに付言しておく、特に米国の DH 分野では、評価のためのガイドラインを作成・公開しつつ³³⁾、現在のアカデミズムにおける評価の枠組みを分析して新しい評価の仕組みを検討する動きも盛んとなっており³⁴⁾、こちらにも注目しておく必要があるだろう。

3.2 知的所有権

DH 分野にわたるサイバーインフラを考える場合、知的財産権をどのように扱うかという問題は避けて通れない。これには DH 分野のなかでも様々な立場があり、また、基盤ソフトウェアの変遷ともあいまって大きな問題となることもある。傾向としては、まず、一次資料を所蔵し、その保管に公費を用いていないところはそもそもデジタル公開したがらず、デジタル公開したとしても強い利用制限をかけたりすることがある。また、二次資料についても、作成に手間がかかっている場合にはデジタル化資料であっても無償公開とされない場合も少なくない。また、無料公開ではあるものの「学術利用限定」という形での公開となる場合もある。加えて、データベースに関して、作成に時間と手間がかかったために無償公開しないという場合もある。このような場合には、もちろん、有料であればそもそも閲覧すること自体の敷居が高くなってしまい、また、学術利用限定という場合には、再利用の際には学術利用であったとしても、その成果が学術目的以外の利用も想定されるような場合には再利用ができなくなってしまうため、結果として活用されにくいという問題がある。

こういった動きの一方で、クリエイティブ・コモンズ・ライセンスに象徴されるような、ある程度自由に資料やツールを再利用できるようにしようとする動きも活発である。公的予算で保管されている文化資料を公的予算でデジタル化した場合や公的助成金でソフトウェアを開発した場合等にはこの傾向が強い。また、DH 分野のなかでもツール開発や分析手法を研究対象にしている人々の間では、成果物の自由な再利用を望む声が特に強い³⁵⁾。実際の所、クリエイティブ・コモンズ・ライセンスの元で配布されるデジタル化文化資料が増加しつつあるのは確かである。

その例の一つ挙げるとするなら Europeana³⁶⁾だろう。Europeana は、欧州の文化資料に関するサイバーインフラの代表格と言えるものであり、欧州中の、デジタル化された 2700 万件以上の図書や絵画、映画やその他博物館・美術館の展示物と、文書館の記録のインターフェイスとして機能するインターネットのポータルサイトである。「モナリザ」やダーウィンやニュートンの著作、モーツァルトの音楽などが特に有名なものである。アムステルダム国立美術館や大英図書館、ルーブル美術館から地方の文書館や地域の美術館・博物館まで、全欧州で 2000 以上の機関が、すべての EU 加盟国から参加しており、すべての時代にわたって、欧州の文化的科学的遺産を見ていくことができる。この圧倒的な情報量や、あるいは、使いやすい検索機能だ

けでなく、SNS へのリンクや Wikipedia に簡単に引用できるボタンも含めた洗練されたインターフェイスなども注目に値する。さらに、この Europeana ではすべてのメタデータをすべてクリエイティブ・コモンズの CC0、すなわちパブリックドメインとして公開しており、自由な再利用によって文化の発展を促進するという姿勢を明確にしている。ただし、このような場合には、プロジェクトの開始当初から無償公開を前提とした上で展開していかねば後々難しいことになりがちのため、十分な予算かもしくは自己犠牲の精神、あるいはその両方が必要となり、また、出版編集の専門家であり多くの経験的蓄積を有する出版社の協力が得にくい状況になるという問題もある。

また、上述の Hathitrust の取り組みと同様に、Google Books Ngram Viewer が、やはり著作権問題を回避する形で書籍のテキストの統計情報のみを公開し、利用者が視覚的に把握しやすいようにグラフ化できる仕組みを提供している。2000 万冊以上の本のテキストデータから語彙の利用状況が容易に把握できるサービスはまさに革新的なものであり、もはや、避けて通ることは難しいものとなっている。一方で、統計情報の元となる OCR には誤読み取りを完全に避けることは困難であり、それゆえ、特に人文学に多くみられるような、依拠するテキストの正確性・厳密性が求められるがちな分野、あるいはそういった分野をも含む学際的研究においては、そのようなデータの取り扱いをどのように位置づけるかということについての方法的な枠組みが必要であると思われる。

3.3 人文学の枠組みの捉え直し

サイバー空間における人文学の活動として注目されているものの一つに Transcribe Bentham プロジェクト³⁷⁾がある。功利主義の提唱者として有名な Bentham の思想が、当時の編集者による改変のために正確な形で公表されなかったことを明らかにするため、University College London に所蔵された自筆原稿に基づきより適切な形で著作集を刊行する Bentham project³⁸⁾が、より広範な協力体制を得るために DH 分野の研究者と共同で開始したものが Transcribe Bentham である。ここでは MediaWiki を改良し、世界中の誰もが、Web 上で Bentham の自筆原稿の画像を見ながら TEI によるマークアップをしつつ翻刻を完成していくという環境を構築し、実際に、世界中の研究者だけでなく、一般の人々も含めて広く参加者を集め、着実に進行しつつある。かなり読みづらい書き癖であるとは言え、英語で書かれているため、海外からの参加者を募りやすいという面があり、専門研究者だけでなく広く様々な立場の人が人文学研究の営みに参加できる枠組みとしても注目を集めている³⁹⁾。

人文学への幅広い参加を集めるという点で注目されるのは DH award⁴⁰⁾である。DH 分野では、ADHO をはじめとして様々な団体によって専門家による専門的な評価に基づく表彰が行われており、とりわけ、若手研究者の表彰には国内外の学研究会が力を入れているところだが、オックス

フォード大学の James Cummings の呼びかけによって開始された DH award では、世界中の一般の人々からの推薦・投票によって Web 上で公開されている DH 分野の成果の表彰を行うという新しい方向性を持つものであった。推薦されたもののうち、最低限の要件を満たしたものはすべて候補となり、60 件以上の候補が挙げられ、その中には日本からの数件の Web サイトも含まれていた。いくつかの観点ごとに賞が設けられたが、なかでも、「一般利用者向けのベストプロジェクト」「DH の楽しい活用」といった賞が特徴的であり、前者として表彰されたカンタベリー大地震のデジタル・アーカイブプロジェクトでは 3000 票以上を集めるなど、専門家の枠を超えて DH 分野が広く注目され、それを通じて人文学への注目もまた広がったであろうことがみてとれる。

さらに、Wikipedia と人文学の研究成果との間をつなぐとするプロジェクトも始まっている⁴¹⁾。このような流れは Public Digital Humanities, あるいは Public Humanities, という枠組みで新たな課題として捉える向きもあり、今後が注目されるところである。

3.4 ツールの開発

また、単に、資料をデジタル化したり、それを活用して成果を出したりするというだけでなく、それを誰もが同様に行えるようにするためのツールを開発することも特に米国では様々に行われている。特に、ジョージ・メイソン大学で開発された、いわゆるメタデータ CMS である Omeka⁴²⁾はまさにその典型である。誰もが簡単にインストールできて CMS 上でメタデータを公開・共有し編集もできる環境を提供するだけでなく、プラグインを追加して様々な機能拡張を行えるようにすることに力を入れている。特に、時空間情報を Web 上でマッピングできる Neatline⁴³⁾や、Web 上で協働翻刻ができる Scripto⁴⁴⁾等のプラグインが有名である。我国では、WEKO⁴⁵⁾や Next-L Enju⁴⁶⁾等がこうしたツールに近いものと言えるだろう。また、歴史資料の翻刻支援ツールである Smart-GS⁴⁷⁾や文字の関係情報を扱うツールも含めた総合環境として CHISE⁴⁸⁾が開発・公開されている。

3.5 文字の問題

Unicode の普及によって、文字の問題はほぼ解決したかに思える。しかし、TEI P5 ガイドラインにおいても gaiji モジュールが存在することは、この問題が依然として容易ではないということを示している⁴⁵⁾。外字と言えば漢字の外字が想起されるが、人文学分野からのアプローチとしては、SAT 大蔵経テキストデータベース研究会により、ISO/IEC10646 で漢字の符号化提案を扱う組織 IRG において、『大正新脩大蔵経』の外字 3000 文字以上が新規登録提案されている⁴⁹⁾。また、欧州でも、Medieval Unicode Font Initiative⁵⁰⁾が、主に中世の文献に出てくる文字の符号化を目指して活動している。微細な違いについては IVS あるいは VS を利用することで、文字としては同一である

もの同士の字形を区別することや、Unicode 標準に影響を及ぼすことなく暫定的に字形の違いを記述・記録しておくことが可能となっているが、その場合には研究者コミュニティ等において可能な限り完全で共有可能性の高い枠組みを構築しておく必要があるだろう。

3.6 教育に関して

サイバーインフラを活用するためにはその活用手法を広めていく必要がある。そこで、教育に関しても DH 分野では様々な工夫が行われている。DH 分野ではすでに我国を含めて世界各国で学部・大学院における教育カリキュラムが用意されつつあるものの、進歩する情報技術に対応していくためには継続的に学習を続けることが必要であり、一度学位を取得しただけで学習がすべて終了するわけではない。しかし、規定の教育カリキュラムだけでは時間がかかりすぎる。そこで、1 週間程度のサマースクールを開催して集中的に学習すると修了証が出るようにする、という動きがカナダのヴィクトリア大学⁵¹⁾と英国のオックスフォード大学⁵²⁾を中心として始まり、米国、スイス、ドイツなど、世界各地にネットワークが広がりつつある。

4. 終わりに

ここまで、DH 分野におけるサイバーインフラにまつわる動向についてみてきた。とてもすべての動きを網羅し切れたとは言えないが、昨今の大きな動向がある程度までは描写できたのではないかと思う⁵⁶⁾。特に、本稿を起点としてさらなる確認をすることが容易になるように、URL を可能な限り付記しておいた。

サイバーインフラを整えたとしても、その利用者が増加し、そこから適切な成果が作り上げられていかないことにはその意義が十分に達成されたとは言えない。その点については、日本だけでも 3 万人超いとされる人文学研究者がサイバーインフラ上の文化資源を生かし、さらに意義のある発信を行ったとしたら、その相乗効果は相当なものとなることが期待される。そうだとしたなら、人文学研究者が、より適切に、より効率的にデジタル化文化資源を活用できる道を用意していくことはいっそう重要になっていくだろう⁵⁷⁾。そして、そのためには、人文学研究者自身がサイバーインフラ上での自らの活動にとって何が必要となるのかをより一層明らかにしていくことが必要となるだろう。さらに、その活動自体が適切に評価される枠組みを作り、そこに向けた教育カリキュラムも整備していく必要があるだろう。情報技術が言語文化を超えて標準化されていくなか、そこに載せられる文化資料に関わるサイバーインフラ、そしてそれを活用する枠組みや技術もまた、相当部分まで国際的に議論され、部分的には標準化されつつある。この動きは「方法論の共有地」として関連分野全体で協働すべきであるという考え方⁵³⁾が特に欧州を中心に広まっている。さらに、DH 分野は人文学にとって「自然科学に共有言語を提供してきた「数学」と同じ役割をもつものとみなされてよいかもしれ」⁵⁴⁾ないとする考え方もある。そのよ

うな方向付けがグローバルに急速に行われつつある中であって、我国でもそこに適切な形で参画していくことは急務である⁵⁵⁾。そして、ただ受容して追従するだけでなく、古来より積み重ねられてきた日本文化そのものに加えて、非欧米圏として明治期以来人文学の伝統を作り上げてきた蓄積に基づく独自性をも踏えた形で適切に発信していくことで、グローバルな世界と向き合おうとする世界の DH 分野の在り方により良い形で貢献していくことができるだろう。

ムセイオンや石渠閣等を起点とするなら研究インフラの歴史は 2000 年を超えるものであり、サイバーインフラはまだ端緒についたばかりである。しかも、現段階では、デジタル技術の進歩にあわせてインフラとしての要件もどんどん変化していくため、技術的側面だけでなく教育や情報共有など、様々な面での即応性と柔軟性が求められる状況となっており、安定化するには今少しの時間も必要だろう。そのようにして様々な紆余曲折を経ていくなかで、人文学研究のためのサイバーインフラは、やがて磨かれ、鍛え上げられ、連綿と続いてきた研究インフラの系譜に連ねるにふさわしいものとなっていくことだろう。

註・参考文献

Renear, Allen H. (2004), "Text Encoding", *A Companion to Digital Humanities*, Blackwell Publishing, 2004, pp.218-239.

(なお、掲載 URL はすべて 2013 年 6 月 20 日に確認した。)

注 1) Marja Makarow, et. al. ed. "Research Infrastructures in the Digital Humanities", Science Policy Briefing no.42, European Science Foundation, 2011/9, http://www.esf.org/fileadmin/Public_documents/Publications/spb42_RI_DigitalHumanities.pdf

注 2) この種の動向については我国では様々な呼称がある。単にカタカナ表記をした「デジタル・ヒューマニティーズ」や、筆者の確認した限りでは国立国会図書館において用いられ、2013 年から書籍のタイトルにも登場するようになった「デジタル人文学」、あるいは同志社大学の学部名に用いられた後に大きく広まるようになった「文化情報学」、これに加えて「人文情報学」「人文科学とコンピュータ」「文化とコンピューティング」等がある。ただし、重なるところは大きい、それぞれに射程が少しずつ異なっているところにも注意されたい。

注 3) この頃の状況は、以下の資料等にて垣間見ることが出来る。「インタビュー Susan Hockey 文書処理の一環としてのテキスト DB インフラの開発」『SuperASCII』vol.2, 1991 年 9 月, pp.136-140.

注 4) さらに、2013 年の年次国際シンポジウムのテーマは「Linked TEI」となっており、Linked data との関係についての議論も活発化している。

注 5) TEI P5 Guideline, 5 Non-standard Characters and Glyphs <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/WD.html>

注 6) なお、近年の米国の DH 分野の動向については「A Short Guide to the Digital Humanities」の和訳が公開されているのでこちらも参照されたい。「デジタル・ヒューマニティーズ入門」(日本語訳) <http://21dzk.l.u-tokyo.ac.jp/CEH/index.php?sg2dh>

注 7) このための情報共有を支援するためにメールマガジン『人文情報学月報』が刊行されている。なかでも特に、菊池信彦氏による国内外の DH 分野の動向紹介の連載記事は有益と思われる。<http://www.dhij.jp/DHM/>

1) 米国データポータルサイト DATA.GOV <http://www.data.gov/>

2) たとえば、オバマ大統領の 2013 年予算案視覚化したものを参照されたい。

<http://www.nytimes.com/interactive/2012/02/13/us/politics/2013-budget-proposal-graphic.html>

3) "Statistics for a changing world: Google Public Data Explorer in Labs", Google Official Blog <http://googleblog.blogspot.jp/2010/03/statistics-for-changing-world-google.html>

4) EADH - The European Association for Digital Humanities <http://www.top500.org/>

5) TOP500 <http://www.allc.org/>

6) The Association for Computers and the Humanities <http://www.ach.org/>

7) Renear (2004) は、この背景について「不必要な多様性と、貧弱なエンコーディングシステムの蔓延が人文学研究を支援するコンピュータの潜在能力を妨げるおそれがあると感じられていた」と述べている。また、Willard McCarty は、「TEI 以前に他の皆と同様に独自のコーディングシステムを開発した」と "Questioning, Asking and Enduring Curiosity: an Oral History Conversation between Julianne Nyhan and Willard McCarty", <http://www.digitalhumanities.org/dhq/vol/6/3/000134/000134.html> において述懐している。

8) Design Principles for Text Encoding Guidelines <http://www.tei-c.org/Vault/ED/edp01.htm>

9) たとえば、ヴァージニア大学の日本テキストイニシアチブには TEI に準拠してマークアップされた日本文学電子テキストが含まれている。

<http://etext.virginia.edu/japanese/index.euc.html>.

10) 情報処理学会人文科学とコンピュータ研究会 <http://jinmoncom.jp/>

11) 情報処理学会 <http://www.jsik.jp/>

12) アート・ドキュメンテーション学会 <http://www.jads.org/>

13) イリノイ大学 (当時) の C. M. Sperberg-McQueen がエディタを務めていた。

<http://www.w3.org/TR/WD-xml-961114.html>

14) Melissa Terras, Present, not voting: Digital Humanities in the Panopticon: closing plenary speech, Digital Humanities 2010, Lit Linguist Computing (2011) 26 (3): 257-269.

15) <http://books.google.com/ngrams>

16) Susan Schreihman, Ray Siemens, John Unsworth ed., *A Companion to Digital Humanities*, Blackwell, 2004.

17) Alliance of Digital Humanities Organizations <http://www.adho.org/>

18) Deborah Lines Andersen ed., Digital Scholarship in the Tenure, Promotion, and Review Process, M.E. Sharpe, 2004.

19) National Endowment for the Humanities, Office of Digital Humanities

<http://www.neh.gov/divisions/odh> なお、Data.gov を通じて公開されているデータセットによれば、DH 分野に特化された助成金交付決定額と NEH 全体における割合の推移は、2007 年: 916,785 ドル (0.8%) / 2008 年: 1,972,995 ドル (1.4%) / 2009 年: 3,193,657 ドル (3.8%) / 2010 年: 4,878,731 ドル (3.8%) / 2011 年: 4,342,009 ドル (3.2%) / 2012 年: 5,535,311 ドル (8.0%) となっている。また、ここに含まれない助成研究の中にも DH 分野的なテーマを扱うものが少なくない。

20) CenterNet <http://digitalhumanities.org/centernet/>

21) Hathitrust <http://www.hathitrust.org/>

22) 各種の COE プロジェクト関連の Web サイトを参照されたい。21 世紀 COE プログラム <http://www.jsps.go.jp/j-21coe/> グローバル COE プログラム <http://www.jsps.go.jp/j-globalcoe/>

23) オーストラリア地域デジタル・ヒューマニティーズ学会 <http://aa-dh.org/>

24) 日本デジタル・ヒューマニティーズ学会 <http://www.jadh.org/>

25) 国立国会図書館サーチ <http://liss.ndl.go.jp/>

26) 人間文化研究機構・研究資源共有化システム

- <http://www.nihu.jp/sougou/kyoyuka/system/>
- 27) MALUI は, Museum, Archives, Library, University, Industry の頭文字をとったもの。同種のものに日本では MLA, MLAK (K は公民館) などと呼ばれるものもあり, 海外では GLAMs (G はギャラリー) と呼ばれるものがある。NPO 知的資源イニシアティブ編『デジタル文化資源の活用地域の記憶とアーカイブ』勉誠出版, 2011 年。
 - 28) ISO TC37/SC4 は, ISO における専門用語, 言語, 内容の情報資源を扱う技術委員会の中の, 言語資源マネジメントに関わる事項を扱う分科会である。 <http://www.tc37sc4.org/>
 - 29) CIDOC-CRM は, 文化遺産の情報を記録する際に用いられる, 概念と関係を記述するための定義と構造を規定する参照モデルである。 <http://www.cidoc-crm.org/>
 - 30) Deborah Lines Andersen ed., Digital Scholarship in the Tenure, Promotion, and Review Process, M.E. Sharpe, 2004.
 - 31) 後述の LLC 以外には, Digital Humanities Quarterly (DHQ, <http://www.digitalhumanities.org/dhq/>), Journal of Digital Humanities (JDH, <http://journalofdigitalhumanities.org/>), Journal of the Text Encoding Initiative (JTEI, <http://journal.tei-c.org/journal/index>) 等が特に注目される。
 - 32) Literary and Linguistic Computing <http://llc.oxfordjournals.org/>
 - 33) 米国の現代語学文学協会 MLA (Modern Language Association) では DH 分野の評価に関するガイドラインを公表している。 http://www.mla.org/guidelines_evaluation_digital
 - 34) ごく最近, 次のようなエッセイ集が刊行されたので参照されたい。Daniel J. Cohen and Tom Scheinfeldt ed., Hacking the Academy New Approaches to Scholarship and Teaching from Digital Humanities, Univ. of Michigan Press, 2013.
 - 35) Digital Humanities Manifesto 2.0 にはこういった背景が強くあるものと考えられる。 http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf
 - 36) Europeana <http://www.europeana.eu/>
 - 37) Transcribe Bentham <http://www.ucl.ac.uk/transcribe-bentham/>
 - 38) フィリップ・スコフィールド (著), 川名 雄一郎, 小畑 俊太郎 (訳)『ベンサム—功利主義入門』, 慶應義塾大学出版会 (2013 年 1 月)。
 - 39) Tim Causer, Justin Tonra, and Valerie Wallace Transcription maximized; expense minimized? Crowdsourcing and editing The Collected Works of Jeremy Bentham, Lit Linguist Computing (2012) 27 (2): 119-137
 - 40) Digital Humanities Awards <http://dhawards.org/>
 - 41) "Bridging the gap between academia and Wikipedia" <http://www.jisc.ac.uk/news/bridging-the-gap-between-academia-and-wikipedia-27-jun-2013>
 - 42) Omeka <http://www.omeka.org/>
 - 43) Neatline <http://www.neatline.org/>
 - 44) Scriptorio <http://www.scriptorio.org/>
 - 45) Weko <http://weko.at.nii.ac.jp/>
 - 46) Project Next-L <http://www.next-l.jp/>
 - 47) Smart-GS <http://sourceforge.jp/projects/smart-gs/>
 - 48) CHARACTER INFORMATION SERVICE ENVIRONMENT (CHISE) <http://kanji.zinbun.kyoto-u.ac.jp/projects/chise/>
 - 49) IRG REPORTS <http://appsrv.cse.cuhk.edu.hk/~irg/irg/>
 - 50) Medieval Unicode Font Initiative <http://www.mufi.info/>
 - 51) Digital Humanities Summer Institute <http://www.dhsi.org/>
 - 52) Digital Humanities@Oxford Summer School <http://digital.humanities.ox.ac.uk/dhoxss/>
 - 53) Willard McCarty and Harold Short 「Methodologies」 Pisa, April 2002 <http://www.allc.org/node/196>
 - 54) 下田正弘「人文情報学を遠望する」 Digital Humanities 2012 基調講演和訳 <http://21dzk.l.u-tokyo.ac.jp/CEH/index.php?DH2012%20keynote%20address>
 - 55) ADHO においても Global Outlook DH という新たなプロジェクトが開始され, 先進国にとどまらず世界中の DH 分野を広く連携させていこうとする動きが進められている。

Special feature: e-Science: Present and future. Cyber infrastructure for the humanities. Kiyonori NAGASAKI (International Institute for Digital Humanities, 5-26-4, Hongo, Bunkyo-ku, Tokyo, 113-0033 JAPAN)

Abstract: Although cyber infrastructure has been a topic of concern since the 1940's, its position as a forefront issue has recently been increasing, along with the importance of the engagement of researchers of the humanities. One of the main venues for these discussions is the main international collaborative organization, ADHO, in which Japanese researchers have now become participants. In the field of the digital humanities, there remain various issues to be resolved such as intellectual property. Evaluation and pedagogy are addressed to create new frameworks in the US. Japan should contribute to the situation to a certain degree.

Keywords: humanities / digital humanities / TEI / ALLC / ADHO / intellectual property / creative commons / Unicode / Europeana