D-GLAM における保存向け電子資料

大矢一志
鶴見大学

D-GLAM and Digital Content Edition for Preservation

Kazushi Ohya
Tsurumi University

2024 年 11 月 4 日
Preprint

1　みんぱくへの敬意と感謝

国立民族学博物館の創立 50 周年に心から御祝いをさせて頂きたく存じます．大阪万博の月の石に強烈な印象を持った小学生時代には，まだ国立民族学博物館（みんぱく) の設立と存在意義は理解できないでいましたが，大学時代，それはちょうどマルチメディアの時代からデジタル・インターネットの時代への端境期で，当時の立ち位置を理解するためメディア史を学んだ際にみんぱく初代館長梅棹忠夫先生の著述を読み，みんぱくの先進性に驚いた記憶があります．また言語学の修士号を持つ身でありながら工学系の博士課程に進学し電子図書館の研究を始めた頃には，文理融合と情報のあり方を考える必要に迫られ梅棹先生の著述を再読し，情報の深い理解に驚かされました．また，大学の専任教員として文理融合の研究を始めたばかりの頃，みんぱくにおられた工学系研究者の方から文理融合は不可能という旨の発言を聞かされ，生半可な知識では文理融合は語れず，双方の分野で独立して評価される研究者でなければならないと恐怖した記憶があります．振り返ってみると，みんぱくは，わたくしが異なる文化の境界面で立ち続け，文系・理系の双方の分野で論文を書ける研究者になることを目標に定めた研究生活の精神的礎の一部でありました．50 周年記念シンポジウムにお声がけいただきました光栄と，50 周年を心から祝いたい自分がこの場所に居られる幸せを感じる機会を頂きましたことに，心から感謝を致しますと共に，機関として常に先端を走り続けて 50 周年を迎えられたことへのお祝いをさせて頂きたく存じます．

わたくしが本記念シンポジウムに呼ばれました理由は，現在は Digital Humanities(DH) と呼ばれる研究分野に長らく携わってきたことと推察します．そこで，わたくしからは 50 年後のみんぱくを，DH という千里鏡を通して眺めてみたいと思います．

2　Digital Humanities 史

DH の歴史を紹介する書籍は多いですが，その多くは Hockey(2004)[20] を前提にしながらの補足・解説になっていると思われます．それは，この論文が, それまでは Computational Humanities や Humanities Computing などの多様な名称であったこの分野に今日まで続く Digital Humanities という名称を与え，研究分野を総覧した書籍 Schreibman et.al. eds. (2004)[36] に収録されているからです．またこの書籍は多くの大学で教科書として採用されてきたこともあります．Hockey(2004) では，2004 年までの DH 史を，(1) Beginings: 1949 to early 1970s, (2) Consolidation: 1970s to mid-1980s, (3) New Development: Mid-1980s to Early 1990s, (4) The Era of the Internet: Early 1990s to the Present の 4 期に分けて紹介しています．

(1) Beginnings 期は, R.Busa のコンコーダンス作成 [4] を原初とし, 1970 年に the Association for Literary and Linguistic Computing(ALLC) と the Association for Computers and the Humanities(ACH) がケン

ブリッジ大学で初めて合同で開催した会議 [44] が契機となり，毎年開催される学術集会へと成長してゆくところまでとしています．この時期の記述は，著者である Hockey 氏が関わる人文学を中心とした紹介に留められていることから，自著大矢 (2017)[79] では，この時期を胎動期 (the Quickening Period) と称し，Busa の活動を支えてきた IBM など計算機科学 (computer science) からの視点を含めて歴史をまとめてみました．例えば Hockey(2004) では欠けていた IBM の活動 [41][17] や Yale University (1965)[43], Bowles ed. (1967)[2] の紹介を含めています．この時期は，本来は数値計算用に開発された計算機を記号処理で使う試みとして，人文資料を扱うことが計算機科学と人文学の双方の研究者で模索されていた時代です．

(2) Consolidation 期は，大型計算機 (mainframe computer) を使いながら各研究機関が果敢に人文資料の電子化に取り組んでいた時期で，例えばオックスフォード大学で Oxford Text Archive (OTA) が発足，後に BNC(British National Corpus) を作り上げる組織的土台が確立されていきます．この時期はテキストデータの塊である病院のカルテを扱う MUMPS[29][62] が開発・実用化されるなど，計算機を記号処理分野で使うことが本格化した時代です．

(3) New Development 期は，パーソナルコンピュータ (PC) の誕生により，研究者個人が人文資料の電子化に参加できるようになった時代です．1977 年に Commodore 社から PET が，Tandy Radio Shack 社から TRS-80 が，Apple 社から AppleII が発売され，それぞれが個性的機能を備えながらも魅力的な価格帯で販売され，なにより全てが造形的に美しい商品となっていたことから，当時の PC の登場は，近年の例でいえば iPhone の登場にも似た熱狂的評価を受けました．PC の登場により研究者個人が資料の電子化に参加できる魅力だけでなく，組織的な資料の電子化においても人的資源を大量に分散投入することで (2)Consolidation 期とは比較にならない大量の電子資料が短期間に作られるようになりました．また計算機センターに大型計算機の利用を申請することなく，自分が使いたい時間に使いやすいソフトウェアを選べる便利さは圧倒的でした．PC 誕生までは，資料の電子化そのものが課題でしたが，PC 誕生後は，電子資料が大量に作られ始めたことで，作られた電子資料を使うことが課題となりました．具体的には，それぞれがバラバラに作られてしまっている電子資料の相互利用が課題として認識されます．この問題意識から生まれたのが 1987 年から始まる Text Encoding Initiative(TEI) の活動です [3]．大矢 (2017) では，Hockey(2004) の 2 期と 3 期を合わせて基礎確立期 (前期) (the Dawn Period: Before the Internet) としてまとめています．

(4) The Era of the Internet 期は，インターネットが一般に使われるようになった頃から始まりますが，もう少し厳密に 1993 年にブラウザ Mosaic が登場したときから始まったといってよいでしょう．web 時代の始まりです．日本での事例ですが，今でも強烈に記憶しているのは，1994 年 1 月に日本で最初の民間向けプロバイダ IIJ が発足したお披露目セミナーで，富士通，NEC といった日本の通信業界を担う企業の研究者が挙って全面的に Mosaic を否定し未来はないと断言されていたなか，IIJ の技術者はブラウザにインターネットの未来があると果敢に宣言されていたことです．その後ほどなく現在ある web の時代へと世界は急速に変化してゆきます．インターネットは熱狂的に市場で受け入れられていきましたが，DH ではもう少し冷めた見方を研究者はしていたと思います．それはブラウザで表示されるコンテンツが HTML[54] という，機能が貧弱なマークアップ言語を採用していたからです．HTML は誰でも簡単に web ページが作れることを目的に作られたもので，DH の研究者が求めるような機能を備えていませんでした．DH では，HTML の登場以前から既に，高度なハイパーリンク機能を備えたソフトウェア DynaText[49] を使い，電子資料が作られていました [40][39]．そこで DH の研究者は新しいマークアップ言語の検討に参加し，DH の要求を叶えようとします．例えば，最も機能性の高いリンクを備えるマークアップ言語 HyTime[55] の論議には S.DeRose 氏 [8] が，続いて開発された XML[69] の論議には C.M.Sperberg-McQueen 氏が参加しています．但し，残念なことに高度なリンク機能を備えた XML Link の論議は止まり，最新のマークアップ言語である HTML(Living

Standard)[53] では，1991 年の HTML から規定されてきた 1 組の IDREF-ID で構成される 1 本のシンプルリンクしかサポートされていません．DH の歴史とりわけマークアップ言語を使った電子資料やその技術からすれば，インターネット以前の状況に未だに到達していない事実があります．

Schreibman et.al. eds(2004) 以降，現在までのおよそ 20 年は，DH という名称がこの分野を示す名辞として定着した時期で，これを大矢 (2017) では撹拌期 (the Stirring Period) と称しました．実は執筆当時，混乱期と名付けようか悩んだのですが，今となれば，これから本格的な混乱期に入る様相からすると，この名称を保留にして正解だったと思います．2004 年以降，学問の名称として DH が定着したにも関わらず，DH の定義論は盛んになってゆきます．

3　DH の定義

Busa (1951) から Hockey (2004) の頃までの DH は，Computational Humanities や Humanities Computing，Humanistic Computing 等の多様な名称で示されていた，いわば名無しの学問領域でしたが，当時の研究者は，なんとなくではあっても共有していた感覚として，計算機を使う資料の電子化とその応用の為には，工学や計算機科学の知識が必要であるという共通認識があったと思います．ところが，現在では，DH はあくまでも人文学であり，電子資料の再生に必要な知識を学ぶことが DH である，という定義が主張されています [60]．恐らくここでは数学基礎論や形式言語は学ばれることはありません．本シンポジウムのタイトルにある「人文知 (Digital Humanities)」も，これと似た立ち位置にある意味を含めたものと推察します．

この DH の活動事例をめぐる，いわゆる定義論そのものが DH のいち活動領域になってゆきます．その発端はおそらく 2012 年にミネソタ大学から出版された Gold ed.(2012)[12] です．これは，スタンフォード大学で開催された前年の DH2011 で DH を”Big Tent” と称し，多様な活動を包括的に囲い込む，いわば投網のような役割を与えたことを受けて出版されています．Gold ed.(2012) の編集方針は，Schreibman et.al. eds(2004) と同様に，広域に渡る DH の”Big Tent” を肯定的に捉え，新たな DH の案内書 (companion) を作ろうという姿勢を感じます．ところが続く Gold ed.(2016)[13] では，DH を細分化しカオス化させ，定義を混乱させるような編集方針が感じられます．面白いことに Gold ed.(2019)[14] では，編集方針としてひとつの DH を探りたいことが述べられながらも，その内容はむしろ分断・カオス化を促進させるものになっています．極めつけは Fiormonte, Chaudhuri, and Ricaurte(2022)[10] の発行です．DH は文化や地域の多様性を反映したものになるとされています．ここで展開されている “Black DH” という標語には，西欧文化を中心として発展してきた DH とは異なる DH が旧植民地地域から生まれるべきという文化的多様性と，それに伴う研究活動の目的と同時に手法にも多様性が求められるという意味が含まれています．ここでの DH は人文学そのものです．最新の Gold ed. (2023)[15] では，Wuhan から始まったコロナ禍を経て DH はいっそう多様化が進んだとし，Museum との協業も取り入れています．まるでリベラルアーツの紹介をしているような内容です．面白いのは，ミネソタ大学の同シリーズから Jonhson, Mimno and Tilton(2024)[22] が出版されたことです．かつては DH の活動領域を示していた Computational Humanities が，現在では DH という名称からは明示されなくなっている証拠といえるでしょう．

このような DH の定義の変化は，DH と称する活動領域が拡大したこと，DH の専門研究者を育成する大学院が設置され組織的教育が始まったこと，すなわちそれは DH 研究者の学術ポストが誕生したこと，文系研究組織または図書館へ IT 部門の設置を支援する多額の補助金が用意されたこと，結果として参加者が増えてきたことが背景にあります．この状況を”DH Bubble” と称する研究者もいました．PC が誕生した (2) New Development 期でも DH への参加者は増えましたが，当時との大きな違いは DH を人文学とは異なる学術領域ではなく，人文学の一環と捉える研究者の参加が増えたことです．この背景を読み解くことは難しいです

が，わたくしの個人的な感覚では，2 つの理由があったと考えています．ひとつは文理融合の難しさ，ひとつは STEM 領域への忌諱です．

Computational Humanities 等の名称であった時代，計算機科学の知識を学びながら人文学との学際・融合分野を模索してきた研究者同士では，DH を従来の人文学とは異なる科学的な学術研究分野として，その確立を目指していた意気込みがあったと思います．ただしそれがどのような融合の形になるのか，はっきりとは言い切れない時代が長く続いていました．しかしこれは文理融合の難しさそのものの現れであり，現在はこの難しさがよりはっきりと認識されてきたのだと思います．歴史を見れば，科学の前身である自然哲学が神をも恐れぬ魔術的な活動からはじまり，1600 年代にガリレオやニュートンなどが科学の性格を明らかにし，産業革命を経て 1800 年代には科学こそが学術研究の中心であるという評価を世間一般から獲得するに至りました．ただしこれは高々 200 年前のことです．そして現在では科学的手法で作られた人工知能 (AI) に人々は恐怖を感じ，AI の学習には人の学習と同じ権利は与えられないというヒステリックな反応を引き起こしたりと，社会は困惑しています．科学が人文学の領域に近づいた途端，人々は科学を理解できないでいることが露呈したといえるでしょう．人間は，200 年ほどの時間では，文理融合を内省の価値基準として自らのものにすることができないでいます．

このような科学知識への生理的反発は，世界的に STEM 教育の重要性が主張される背景にある，STEM 領域の学習を嫌う学生の増加からも知ることができます．足元を見ても，インドや東アジア系の民族は科学・技術系の学習が得意であるという評価があるなか，日本では科学の学習は敬遠され続け，その傾向は強まる一方です．このような情勢が恐らくは，人文学に計算機を導入する DH の定義にも影響を与えている 2 つ目の要因ではないかとわたくしは感じています．なぜ人文学で計算機を使うのか，それは電子資料の再生に必要だからである，という定義が生まれてきたのは，身近にある道具 (計算機) の存在は疑わない姿勢が増えてきたことが背景にあるのでしょう．故障しなくなった車が当り前になることで，エンジンオイルやエンジンベルト交換の知識がなくとも自動車を運転できるようになったのと同じく，計算機が日常の道具になったことで，その仕組みを知らずに使う利用者が増えたということです．Digital Humanities (DH) という名称は，「今の時代の人文学 (contemporary humanities)」という意味になったのだと思います．

ちなみに，文理融合の難しさは，理系から文系への参入でも見られます．DH と称する活動の中には，計算機科学の手法を使うだけで，その結果や成果が従来の人文学研究にどう貢献するのかよくわからないものがあります．例えば，統計学的手法を使った結果が，人文学上の問題解決にどう関連するのか説明がないケースです．また，人文学で検討されてきた定義や仮説を無視した研究活動があります．例えば，知識表現研究では，言語学の意味論や言語哲学の論議が前提とされていない研究が見られます．もちろん，手法が異なれば求める成果も異なると考えることもできますから，理系でも文系でもない，第三の学術領域を DH が生み出す可能性はあります．このような，手法は理系から，解決すべき問題は文系からという，文理融合の難しさを避けるため，例えば，DH とは別の研究領域として成長してきた言語データ処理の分野 (e.g. LREC[59], COLING[48]) では，DH が処理対象の下位領域として設定され，手法も目的もその分野の文化でまとめられています．ここでは大きな価値観の衝突は起きていません．理系分野でも人間に焦点をあてている研究領域では，比較的容易に人文学と連動した研究活動が始められているようです．

個人的な空想を述べると，ゆくゆく DH という名称は「構造主義」と同じような位置づけの用語になると考えています．先に示した通り，自然哲学は 1800 年代には科学として人文学に取って代わり学術領域の主流となりました．この当時，人文学者は過去にすがる自称学者として自信をなくしかけていました．そこに言語学から音韻変化の法則が報告されると，人文学でも科学的手法，すなわち定義に従う体系的な分析が可能であることがわかり，続いてソシュールがテキストの構成要素である記号を定義化すると，人文学でも体系的分析を

試みる機運が生まれました．これが (ヨーロッパ) 構造主義と自称される活動の発端です．その後，言語研究にとどまったアメリカ構造主義とは異なり，(ヨーロッパ) 構造主義は，境界線の曖昧な人文学のいち分野となりました．DH も今後は構造主義と同じように広く緩やかな人文学のいち分野を示す名称として使われていくと思われます．

4　電子資料 (Digital Content Edition)

DH の定義がどのようであれ，DH の活動に電子資料が必要であることにかわりありません．“No Digital Content, No Digital Humanities” です．電子資料の種類には，(1) マークアップテキスト版 (marked-up text edition), (2) 電子ファクシミリ版 (digital facsimile edition), (3) バイナリ版 (binary edition) があります．Busa が始めた文字入力からコンコーダンス，コーパスが生まれ，そのコーパスにアノテーションを付加したブラウンコーパス (Francis(1964)[11]) から (1) マークアップテキスト版の作成が始まります．マークアップテキスト版の電子資料作成は，現在まで DH の中心的な活動領域です．

電子ファクシミリ版は，図書館に所蔵されている資料としてあったファクシミリ版の電子バージョンです．電子画像の歴史は 1960 年代まで遡ることが可能ですが，静止画像データの実用が始まるのは 1980 年代，一般化したのは 1990 年代に入ってからのことです．本稿では，電子ファクシミリ版の作成については Kenney and Rieger (2000)[25]，金沢 (2004)[24]，大矢 (2010, 2024)[76, 82] を挙げることで解説を省略いたします．

バイナリ版の電子資料は，博物館等の展示を中心に開発され，例えばみんぱくが先進的に導入してきたビデオテーク，フランスから発信されたメディアテーク (Médiathèque; multimedia library) といった，いわゆるマルチメディアの時代に生まれた展示様式を電子化したものです．近年はバイナリ版に注目が集まっていますが，DH では長らくバイナリ版は研究対象から外れていました．理由は，再利用性が低く研究利用に向かないこと，メディアの寿命が極端に短いことです．時間依存コンテンツの場合は技術的難しさも研究対象から外れていた理由です．

本稿では，マークアップテキスト版の開発史を紹介します．

4.1　マークアップテキスト版と TEI

現在まで続くマークアップテキスト版の電子資料は，SGML[16] を採用した TEI の結成 (1987 年) 以降，本格的な制作が始まります．TEI の活動は，DH を理解する上で欠くことのできない活動です．TEI は 1987 年にマークアップテキスト版の電子資料の作成方針を”The Prouhkeepsie Principles” と”Design Principles”[63] としてまとめ，これを元にタグの種類とその関係 (構造) を TEI Guidelines で規定します．残念ながらこのガイドラインは定義集に留まるもので，なぜこのタグが必要なのか，このタグはなぜ他のタグとの関係 (構造) がこのようになっているかの説明がありません．それらは Ide and Véronis (1995)[21] で解説されています．この 2 つを合わせて読むことで，TEI スキーム (scheme; タグとその関係の定義で別名はアプリケーション，データベース系では schema と表現) の全容を理解することができます．ところが TEI はもうひとつ，TEI スキームを理解する上で欠かせない第三の資料をまとめてきませんでした．TEI スキームの使い方を案内し，マークアップテキスト版作成の実例を示す資料が作られていません．“TEI by Example”[65] という情報源が作られたことはありましたが，規模は小さいものでした．但しこれは，TEI スキームの成立背景を知れば，ある意味仕方のないことで，また当時の TEI 活動の良心の表れであることも分かります．

TEI スキームは，top-down 式に作られたものではありません．それぞれの研究領域の専門家が必要としたタグを持ち寄り bottom-up 式に束ねてまとめられたものでした．これはちょうど DH の定義に似たものがあります．結果として，似たような役割のタグが複数あったり，似た内容が複数の書き方で表現できたりと，統

一性・一貫性のない歪な設計になっています．TEI ガイドラインの改定の歴史は，この混濁したスキームを見通しよく整理してきた歴史です．例えば，”class” の抽出もそのひとつです．TEI でいう”class” とは，計算機科学における”class” とは似て非なる概念で，TEI スキーム中にある似た役割 (意味) の記述を”class” としてまとめ定義を独立させたものです．これにより同じ内容がどこで書けるのか分かりやすくなりましたが，書けてしまうことに変わりありません．実際，TEI スキームは多様なインスタンス (実際のマークアップテキストのこと) を生み出し，利用者に混乱を与えました．この記述の多様性を制御してきたのが TEI のもう一つの活動です．TEI スキームを使う人々に，タグの使い方や，その背景にある哲学を示し，利用者が同じようなインスタンスを書くよう案内・誘導してきました．残念ながら，これらの案内が資料としてまとめられていません．論議の記録が公開されているだけです．これを検索するだけでは，TEI スキームを自らの資料に当てはめてマークアップテキスト版を作成することは困難でした．結果として，TEI スキームを使う人々は，TEI コミュニティのメンバーに使い方を信託のように伺うという，一種の教団化現象が生まれます．これを口の悪い研究者は”TEI Police” と揶揄してきました．但し，TEI を擁護すれば，これは歴史的には仕方のないことです．先に紹介したように，統一性のない雑多なスキームを” 規格” として運用するため，TEI コミュニティは実例に向き合う都度，試行錯誤しながら運用を模索してきたのです．コミュニティの中で公開で統一的な運用を模索していたと理解するのが，歴史的には正しいと，わたくしは考えています．この模索の様子は，まさに「記述実験 (metawriting experiment)」そのもので，マークアップ言語研究の重要ないち側面です．例えばその重要性は，XML の策定記録 [66] を読むとよく分かります．

TEI の利点は多くの文献で紹介されていることから，ここでは自明のものとして省略します．本稿では英語の資料でも知ることが難しい TEI の欠点を紹介します．具体的には，(1) 記述の多様性，(2) 深い構造，(3) 単一文化性，(4)stand-off スタイル，(5) 拡張性，(6)XML 採用の 6 点です．

(1) 記述の多様性は先に紹介した歴史的背景が原因です．統一した記述にまとめる努力はされていますが，なぜそうするのかという理屈が上手く説明できないケースが多く，その解説を理解する労力が無駄になります．記述実験は重要な活動ではありますが，これはわたくしのようなマークアップ言語の研究者にとっては重要という意味で，当該資料の専門家にとっては，無駄な時間でしかありません．しかも，当該資料が持つ重要な情報の記述が疎かになる可能性があります．

(2) 深い構造は，TEI の欠点として古くから批判されてきました．SGML 系のマークアップ言語は，単位間の関係性を木構造として表現します．従って，関係性をより詳細に記述しようとすると，自然と木構造の階層は深くなります．これを避けるには，単位間の関係を，木構造ではなく，別の手段で示す必要があります．例えば，要素 (タグ) の意味・概念を別に定義したり，リンクを使い意味 (抽象) 構造として表現するなどです．どの手法が正しいということはなく，バランスのとり方が論点となります．従って，TEI スキームを使いながらも，浅い構造で記述すれば，この欠点は解消されます．

(3) 単一文化性とは，TEI スキームの要求仕様は全て西欧文化の研究者から出されてきたことが背景にあります．これは歴史の先駆者として当然のことで，これを欠点と指摘することには躊躇するのですが，わたくしのように日本文化の資料の電子化を考えてきた研究者にとっては大きな問題でした．“Black DH” の主張もこの文脈から生まれています．他文化への拡張の試みは，日本文化を例として Ohya (2014)[78] にまとめてあります．この研究成果以降，わたくしは TEI 以外の選択肢を模索し始めました．

(4)stand-off スタイルとは，インスタンスの構造に加えて，リンクを使い意味構造を表現する書き方のことです．これがもたらす技術的問題の理解には，マークアップ言語の詳細な知識が必要になることから，本稿では解説を省きます．この問題の詳細は大矢 (2009)[75] に，また解決方法とそれを支援する Python ツールの公開については Ohya (2022)[80] にまとめてあります．

(5) 拡張性の問題には，TEI の哲学的な矛盾が含まれています．TEI スキームの哲学となった “Design Principles” には，“Since research necessarily involves the asking of questions that have not been asked before, a research-oriented encoding scheme must also be extensible. “とあります．研究活動の必然である新しい情報の発見に対応するため，TEI スキームは固定されたものではなく拡張できるという方針です．これはある意味，TEI が目指す「規格 (standard)」としての役割と矛盾しています．規格は相互利用を促進するものですが，拡張性は相互利用を阻害します．TEI の発足は，相互利用の実現が目的でしたから，この哲学的矛盾を取らざるを得なかった，とするのが現在からみた歴史的評価です．当時作られていた電子資料の多様性は，研究成果の多様性の反映ではなく記述の混乱とみなし，まずは相互利用を目指した，ということです．XML では，この矛盾に対応するため，“valid document”（妥当な文書) と”well-formed document”（整形式文書; 整 + 形式 + 文書) の 2 種類の文書を導入しました．前者は，データ単位が事前に分かっているケースでの文書，後者はデータ単位は事前には分からず，作業を通しながらデータ単位を発見してゆく文書です．人文資料の電子化は「発見的行為 (heuristic process)」と呼ばれることがあります (MacCarty 2005[28])．作品を電子化するとき，正確な入力を心がけると自然とテキストを正しく読み取るようになり，結果として作品を正確に読み解くことになります．電子化の作業を通して，作品と正対し，新しい発見につながることを DH 研究者は経験してきました．これを電子化に伴う発見的行為と呼んでいます．テキストの入力と同時に発見的にタグを入力できる整形式文書は，DH 研究者にとり理想のデータ形式です．TEI スキームは SGML を基に設計さたことから，妥当な文書しか書けませんでした．しかしその後の XML で採用された，人文学研究に相応しい整形式文書に対応した TEI スキームは今日まで開発されていません．発足時に，それまでの”declarative markup(宣言的タグ付け)” に留まるメタ記述手法ではなく，”descriptive markup(記述的タグ付け)” が可能な SGML を採用した TEI が，より自由な記述を支える整形式文書に対応してこなかった事実は，不思議としか言いようがありません．これは TEI の致命的な欠陥であるとわたくしは考えています．恐らくは，規格としての性格を優先させたこと，処理系専門家の意見が人文学研究者の意見より尊重されたことが背景にあるのでしょう．結果として，TEI スキームにある拡張性と，その実効性が著しく削がれています．

(6)XML の採用は，TEI のみならず，DH 全体の大きな問題です．計算機科学の世界で XML は熱狂的に受け入れられ，多くのソフトウェアで採用されました．例えば，Microsoft の Word や Excel のデータ形式として採用されています．また図書館のカタログデータの記録形式としても採用されています [61]．ところが XML はこの 10 年で急速に使用頻度を落としています．テキストデータの表現形式として，現在は圧倒的に JSON[58] が採用されています [67]．また XML 関連規格の殆どは開発が止まっています．XML データを使う人が少なくなれば，それに対応したツール類の開発も止まってしまいます．TEI が 1987 年当時に SGML を採用したことは，ある意味，挑戦であったかもしれません．それと同じく，現在 2024 年に TEI スキームを策定するとすれば，XML ではなく HTML(Living Standard)[53] を採用するのでしょう．わたくしのようなマークアップ言語を研究してきた古い研究者からすると，整形式文書をサポートしない HTML(Living Standard) を電子資料に使うことには不満を感じます．但し，出版形態として web 版コンテンツはおそらく唯一の選択肢です．また，XML データが，DH の世界だけで使われてゆく，という可能性は低いと思われます．ソフトウェア開発を維持するマーケットが DH 周辺だけでは小さすぎるからです．また例え一部のソフトウェア開発が継続したとしても，特定アプリケーションに依存するスタンスは，TEI も発足時の哲学で否定したように，DH では避けられてきた選択です．欧米文化の人文資料の多くは既に TEI データとしての電子化が済んでいることから，XML を使う重い慣性に対抗する動きは鈍くなることが予測されます．XML を安定した枯れた技術と評価する人もいますが，それをサポートするツールのアップデートが止まり，OS のアップデートが一方的に進んでゆく現状は，もはや枯れたではなく廃れた技術とするのが正しい観察です．これか

ら電子資料を作るケース，特に DH を人文学と考える研究者は，この事実に十分に注意して下さい．ちなみに，わたくしは XML でも HTML でもないマークアップ言語を提案していますが [80]，これは研究レベルの仕様で実用上の仕様ではありません．

4.2 機能と表現

マークアップテキスト版を中心に電子資料が作られたきた DH では，世間一般で享受されている web 版コンテンツが次第に多機能・高表現力を持つに従い，表現力を高めた電子資料の作成が目標になってゆきました．DH 史で見たように，DH で必要とされる高機能のリンクは，現在の HTML ではサポートされていません．HTML の機能拡大は DH の強い要望でした．実は，世間一般もブラウザの登場以降，ネット回線の速度向上と，ブラウザの機能追加を熱望するようになります．これに対して Microsoft は，1996 年自社ブラウザ Internet Explore に ActiveX controle という機能を導入して応えます．ところがこれは，とんでもない勇み足でした．Microsoft は，当時ブラウザの処理系に求められていたセキュリティ上の制限 SandBox 型を無視し，ActiveX controle を備えたブラウザを介して Windows 95 が持つ機能 (COM) を呼び出すことができるようにしました．結果，ActiveX controle は発表と同時に悪意あるコンテンツで利用され，重大なセキュリティ問題を引き起こします [9]．例えば，悪意ある web 版コンテンツは，ハードディスク上にあるファイルを自由に削除できました．この事件以降，インターネット世界では，ブラウザに HTML の規定を超える機能を備えることに躊躇します．web 版コンテンツ停滞の時代です．web 版コンテンツへの機能追加は，Java や JavaScript を使うプログラミング言語処理系 (programming language processing) ではなく，マークアップ言語の多機能化 (markup language processing) で実現することが検討されてゆきます．この流れの中で生まれたのが 1996 年末公開，1998 年に正式化された XML です．また XML の機能を向上させる関連規格が同時に策定されました [71]．ところが，XML 関連規格の論議は難航し，それを実装するブラウザは殆どありませんでした．ブラウザ作成者側からすれば，規格として意味が決められた宣言型命令文よりも，素直にプログラミング言語を規定しソースコードから柔軟に処理・機能を追加できる方が良かったのです．但し，Microsoft が自社利益を優先させた勇み足の所為で，ブラウザとプログラミング言語の融合は敬遠され続けます．このトレンドを変えたのは，Google が始めた Google Map(2005 年) のサービスでした．ここではプログラミング言語 JavaScript[57] を web 版コンテンツと融合させ，いわゆる動的な web 版コンテンツを実現しました．この成功を受け，動的 web 版コンテンツの制作は，2006 年に Ajax[45] として規格化されます．Microsoft の大失態から 10 年の停滞期を経て，ようやく安全な動的 web 版コンテンツが作られはじめます．そして，機能と表現力の向上がプログラミング言語処理系で実現できる環境が整うと，XML 関連規格への要求は減り，XML の利用も衰退してゆきます．

現在では，世間一般からの高機能・高表現力の要求は，サーバ側とブラウザ側の両処理系で満たされ，ブラウザは OS のような位置づけ，すなわち，計算機でできることはブラウザ上でも全てできる環境になりました．DH でも，HTML より高機能なマークアップ言語を求めた当初の立場を変え，現在の web 版コンテンツで採用されている処理系を使ったコンテンツ作りが始められています．この高機能・高表現力の電子資料をもとめる流れは，DH2023 のキーノート Kenderdine (2023)[51] で象徴的に示されました [52]．DH では，人文学を超えて，博物館の展示やアートの世界との接点を深めるまで，表現力を求める傾向は強まっています [46][47].

4.3 情報と表現; 記号の定義

高い表現力とそれを支える高度な機能が求められるようになった電子資料ですが，この高い表現力には注意が必要です．表現力と電子資料の関係は，ソシュールの記号の定義から理解を深めることができます．ソ

シュールは，現実世界で観察される意味の現れ方を，記号の側面から説明しました．ソシュールの定義に従えば，記号 (sign) は，意味 (signified, meaning) と表現記号 (signifier, expression) の組で定義されます．従って，意味または表現記号のどちらかが異なっていれば，記号も異なります．これを電子資料に当てはめると，電子資料 (Digital Content Edition) は，内容 (content, meaning) と表現 (expression) から成り立つと説明できます．例えば，作品「星の王子さま (The Litter Prince; Le Petit Prince)」は，書籍版や kindle 版，web 版，app 版が存在していますが，それぞれの版で媒体独自の表現記号を使います．kindle 版と web 版は共に電子資料でも，利用者の体験的情報は異なることから，異なる記号 (電子資料) になります．この記号定義の注意点は，意味や表現記号は必ずしも既定である必要はないことです．例えば，「DH」という名前 (表現記号) があっても，その意味は未定 (TBD) ということがあります．逆に，名前がなくとも，DH ともいえる学術活動の意味概念は存在できます．言語接触の歴史では，このように概念はあっても名称がない場合には，他言語から名前を借用し，記号を成立させてきました．例えば，日本語の「システム」（英語の system から）や英語の”emoji”（日本語の絵文字から）などです．このような意味と名前 (表現記号) の緩やかで自由な関係をソシュールは「恣意性 (arbitrariness)」と表現しています．この記号の定義はチューリング機械の限界や知識表現研究を理解する上でも重要な役割を果たします．チューリングは，ヒルベルトの問いかけに対して，問題を解くことは表現記号の操作であるとし，計算可能性をチューリング機械として提示しました．ここでは意味を伴う記号は扱われていません．また Semantic Web[1] の知識表現で使われる RDF[64] では，表現記号が意味ラベルになっていますが，これでは同じ表現記号でも異なる記号 (=意味が異なる) が存在することや，表現記号がない記号 (=意味はある) は記録できません．

この定義からわかるように，表現力は記号 (または版) の存在を弁別します．すると，現在 DH のトレンドである高表現力を伴う電子資料は，同じコンテンツでも，従来の表現力でしか作られていない電子資料とは異なる版になります．この高い表現力で生まれた異なる版を，現在の技術では将来に残す手段がありません．

5　電子資料と保存

電子資料一般 (digital objects) の保存では, その保存方法と保存対象を素性に, 保存の仕様を分析します．保存方法としては, 例えば, (1) 動態保存 (living preservation), (2) 機能保存 (functional preservation), (3) 内容保存 (content preservation), (4) 外観保存 (appearance preservation), (5) 存在保存 (existence preservation) があります.

(1) 動態保存とは，そのままが存在している保存のことです．シアトルにあった Living Compomputer Museum + Labs は，計算機を動く状態で保存していた博物館でした (残念ながら今年 2024 年に閉館してしまいました)．(2) 機能保存は，レプリカを作り，同じように動く状態にしておいたり，エミュレーションでソフトウェアを動かせるようにしておく保存方法です．利用者側からすると電子資料の (1) 動態保存と (2) 機能保存からは，同じ体験的情報が得られます．(3) 内容保存は，人が享受するコンテンツを保存しておく方法です．例えば，Web Archiving などの多くはこの保存方法を採ります．この場合 (2) 機能保存とは異なり，操作手順などはそのまま同じではありませんが，得られるコンテンツは同じです．(4) 外観保存は，静止画や動画で電子資料を記録する保存方法です．例えば，ゲームのプレイを録画したり，web ページのスクリーンショットを保存したりする方法です．(5) 存在記録は，いわゆる書誌データ，メタデータの記録による保存です．電子資料のコンテンツは記録されませんが，電子資料の存在そのものは記録保存されます．

保存の仕様作成にはこの 5 種類の保存方法と合わせて，電子資料が生成される段階を基準とした保存対象の分類が必要です．例えば，(i) 一次データ (raw data)，(ii) 素材データ (data on content factory)，(iii) 公開データ (data on a public platform)，(iv) 複合コンテンツ (data on a service platform) です [81][82].

(i) 一次データとは，はじめに作られるデータのことです．例えば，デジタルカメラで撮られた画像データ，エディタ上で入力されたテキストデータなどが相当します．(ii) 素材データとは，一次データを加工して作られた，電子資料を作成する際の素材となるデータのことです．例えば，ファイル名が変更された画像データ，ページやレイアウトを整形したワープロデータや PDF ファイル，前後をカットされた音声データ等です．(iii) 公開データは，素材データそのもの，または web 版コンテンツであれば複数のファイルから構成される web ページとして公開されたものがこれに相当します．(iv) 複合コンテンツは，ブラウザ上では (iii) 公開データと同じ印象ですが，その元となる素材が複数のサービスから提供されているものです．データマート (data mart) またはデータウェアハウス (data warehouse) と呼ばれるサービスから部分素材の提供をうけ作られた 1 つの web 版コンテンツです．現在の新聞社の web ページは，自社記事に加えて，広告，株式情報，天気情報など，他のサイトから提供されているコンテンツをまとめた複合コンテンツがほとんどです．(iii) 公開データと (iv) 複合コンテンツはそれぞれ，単一サイトから提供されるコンテンツ (sigle site content)，複数サイトから提供されるコンテンツ (multiple sites content) ともいえます．

これらの保存に関する分類を使い，具体的に電子資料の保存のケースを観察してみます．例えば，(1) 動態保存が可能である対象は (i) 一次データのみで，(ii) 素材データは作成に利用したアプリケーションの寿命によっては保存されません．すると (iii) 公開データも (iv) 複合コンテンツも保存の保証はありません．(2) 機能保存でも (i) 一次データは保存される可能性は高く，(ii) 素材データと (iii) 公開データは依存する処理系が特殊なものでない (過去でいえば Flash 等の非推奨アプリや，後述する非推奨データベースを使わない) 場合には保存の可能性は高くなります．(3) 内容保存は，(i) 一次データ，(ii) 素材データ，(iii) 公開データで保存の可能性は高く，(iv) 複合コンテンツの保存はほぼ不可能です．著作権処理も困難のひとつですが，利用者別コンテンツ提供が可能となった現在では，体験的情報を保存することができません．(4) 外観保存と (5) 存在保存は，全ての対象で保存が可能です．メタデータは存在保存に加えて外観保存の役割も担うようになってきました．

ここで非推奨データベースの補足をしておきます．電子資料と関連したデータベース (DB) の使い方には，(a) 検索 (retrieval)，(b) 制作 (assembly)，(c) 隠蔽 (hiding)，(d) データ整理 (organizing) の 4 種類があります．DB はもともと (a) 検索の為に作られたソフトウェアですが，電子資料では (b) 制作の為にも使われます．web サーバ側の処理系と連動して DB を使い，web 版コンテンツを構成する部分情報を検索して取り出し，それを素材として web 版コンテンツを表示させるケースです．このような (b) 制作で DB を使うケースでは，コンテンツを (c) 隠蔽する為に DB が使われることもあります．部分コンテンツの元となるデータ全てを利用者には公開したくない時，それを (c) 隠蔽するために検索結果だけを表示させます．オープンデータのような隠すつもりはない場合でも，インタフェースとして検索メニューしか用意されていないケースでは，実質 DB は (c) 隠蔽のために使われています．先の保存の分析で，非推奨データベースと紹介したのは (c) 隠蔽のために使う DB のことです．電子資料の保存で (c) 隠蔽目的で使われる DB は，(i)-(iv) の全てにおいて障害となります．(d) データ整理として使える DB は，残念ながらまだありません．技術的な話をすると，この実現には，スキーマレス (schema-less) で，かつ検索以外で収蔵データの観察手段が提供される必要がありますが，これが実用化されている DB がありません．もし (d) データ整理として使える DB があれば，個人がアイディアを練り情報を集め整理する道具として使うことができます [80].

5.1 保存に向けて

電子資料の表現が，ブラウザとサーバ側の両処理系で実現されている現在，電子資料の保存は極めて難しい状況です．特に，世間一般が享受する (iv) 複合コンテンツの保存は技術的にも社会的にも不可能な状態で，こ

のままでは未来から見た歴史的今日は，無記録の時代になるかもしれません．それはともかくとしても，DH や博物館などが扱う学術的資料性が高い電子資料においても，高い表現力を備えた電子資料の保存は危機的な状況にあります．これはちょうど DH 史においてインターネット以前に作られていた電子資料が，HTML でサポートされなくなった結果，再生できなくなり，実質的に存在しなくなった歴史を，より規模を拡大して繰り返すことになります．わたくしたちは，電子資料を儚い消費型情報として位置づけ，書物が担ってきた知的資産の継承という役割を電子資料で失わせる絶滅事 (extinction event) に加担しています．

電子資料の保存が完璧にはできない技術の限界に贖うには，銀の弾を求めるのではなく，泥臭い地道な工夫で対応するしかありません．例えば，先の観察から得られる教訓には以下のようなものがあります．(I) 学術的な電子資料は (iv) 複合コンテンツにしない（外部サービスやサイトに依存しない）．(II)(3) 内容保存，(4) 外観保存，(5) 存在保存は常に実現する．(III)(2) 機能保存をできるだけ実現する．(IV)web 版コンテンツでは (1) 動態保存を目指し，(V)web 版コンテンツは (iii) 公開データを原則とする．(VI)DB を (c) 隠蔽には使わない．これに，従来までの保存の原則を付け加えると，例えば以下のようなものがあります．(VII) 特定アプリケーションに依存したコンテンツ制作はしない（但し，HTML ブラウザを除く）．(VIII)web 版コンテンツを構成する全ての素材を公開する．

これらは全て経験から得られた帰納的な命題で，永続性のある真理ではありません．生のことばを使えば，かっこ悪い方法です．それでも，現世に生まれた責任として，後世に伝えるためにやれることはやるべきではないでしょうか．

6 D-GLAM

図書館，文書館，博物館・美術館をまとめて，人類の文化資産を扱う機関として LAM または GLAM と表現することがあります．GLAM はこれまで，所蔵する文化資産の保存については独自に対応していました．GLAM の中で電子資料の扱いにいち早く取り組んだのは図書館です．1994 年からアメリカでは電子図書館プロジェクト (Digital Library Initiative)[35] が始まり，現在ある web 版コンテンツの原型と，その有効性が確認されます (DLI1 は 1999 年に終了)．この成功を受けて，web システムの発明者である Bernars-Lee 卿は，ネット世界を単なる雑多な情報の置き場所ではなく知識や知恵の収蔵庫にすることを 1999 年に提唱し [1]，現在まで続く知識表現研究の源流を作ります．その後の電子図書館 (DL) 研究は，1998 年の XML と 2004 年の Goole Print Library Project(後の Goole Books) の登場以降，電子資料の定義が拡散し，一般的なインターネット研究に吸収されてゆきます [68]．DL 研究では，電子資料の利用促進が主たる目的で，所蔵文化資産の保存については Digital Archive を含む Digital Preservation[6] という別の活動として取り組まれてきました．Digital Preservation の取り組みは，図書館だけでなく文書館や博物館からも参加する GLAM 全体の活動になっています [50][33]．博物館は，インターネットが大衆に利用されるようになっても長い間 web 版コンテンツを重要視していませんでしたが，資金難により入場料収入への期待度が高まると，宣伝媒体として web 版コンテンツを見直し始めます．収蔵コレクションを web 版コンテンツとして紹介する動きが活発化し [23]，さらにコロナ禍を経た現在は，オンライン展示 (Digital Exhibition) としての web 版コンテンツが重要視される気配です [38][5].

GLAM がそれぞれの所蔵品を電子化し，それを広報や展示 (Digital Exhibition) として公開している様子は，利用者側からも，その情報を提供するシステム側からも，GLAM の違いなく同じようなものとして見えてきます．このような同じように見える電子資料の公開の様子は，DH や DL などの名称で扱われていますが (e.g. [33],[56])，本稿ではこれを便宜上 Digital GLAM(D-GLAM) と称しておきます．D-GLAM は web 版コンテンツとメタデータでその存在が定義されます．

6.1 D-GLAM の web 版コンテンツ

D-GLAM の web 版コンテンツでは，それが展示向けか保存向けかを明示すべきです．D-GLAM では，保存版は必須で，展示版が作られる場合には同時に保存版も作られるのが機関としての責任でしょう．先の電子資料の歴史で見たように，現在は，電子資料の保存が技術的に難しい情勢です．文化の過渡期に生きるわたくしたちには，スマートな方法を選択する術は残されておらず，将来は否定されるであろう現時点でしか通用しない対症的方法を使ってでも，文化資産を後世に伝える努力をすることが求められます．例えば，そのようなポリシーとして，以下のようなものが考えられます．

1. 保存版，または保存版と展示版が作られる．
2. (iii) 公開データとして作成する．
3. DB を (c) 隠蔽目的で使わない．
4. web 版コンテンツを構成する全ての素材は独立したファイルとする．
5. 特定アプリケーションに依存した素材制作はしない．
6. 相互利用を急がない．

6 番目の相互利用を急がないについては，6.3 節で解説します．

6.2 D-GLAM のメタデータ

図書館では既にメタデータの横断検索が実現される一方で，博物館ではまだ自館の所蔵物を記録することにも解決すべき課題があるとして，図書館のメタデータを参考にしながら課題に取り組まれていますが，これには注意が必要です．図書館が記録するメタデータは工業製品である印刷本を対象にしています．すなわち，規格化された製品のメタデータを記録しています．これは LC Cataloging-in-Publication Data や British Library Cataloguing in Publication Data の有無に関わらず，印刷本として誕生してから今日までの商業活動として，印刷機の製造，紙の製造，装丁機械の製造という，産業化に伴う自然な結果です．工業製品である書籍の存在記録は，定型化が容易です．これに対して，同じ印刷本でも日本の版本 (woodcut print) や稀覯本となった印刷本，印刷本ではない写本，博物館で扱う様々な現物は 1 点ものであるケースが殆どで，その弁別を可能にするメタデータの策定では，図書館のメタデータは参考になりません．メタデータの記述方法においては図書館のメタデータが特異解 (singular solution) であり，博物館を含む Digital Preservation で求められるメタデータの方が本来のメタデータの姿なのです．本来のメタデータの検討では，図書館のメタデータにあるような規格やスキームに従う規範的な姿勢でデータを求めるのではなく，記録される対象を観察した結果得られた，必要と思われる情報を全て記録する，いわば記述的姿勢でデータは作られるべきです．結果として作られたメタデータは，他館との互換性はないことからすぐには横断検索や相互利用はできませんが，互換性を目標としてメタデータを作るのは危険な選択です．優先されるのは互換性ではなく記述の方です．保存の対象物に対して，研究者や所蔵館の視点で必要と判断された情報が十分に記述的に記録されていれば，その情報は未来において必ず相互利用が実現されます．一方で，現時点でしか書けない，現在だから書ける情報が十分には記録されていない場合，未来においてそれが補足・修正されることはありません．現世のわたくしたちの義務は，十分に記述し，それを記録し，公開することです．データが残されていれば，そして将来そのデータが必要とされていれば，その相互利用は現在よりも一般には低コストで実現されます．メタデータを作成するポリシーとしては，例えば以下のようなものが考えられます．

1. メタデータは紙上に記録された情報とは別に (再) 作成する．

2. 項目・構造は独自に決めて良い (規格は気にしない).
3. データ変換が容易にできることを確認しておく.
4. 必ず入力者が記名する.
5. 全データを公開する.
6. 相互利用を急がない.

6.3 相互利用を急がない

現在の DH では，作成した電子資料を自らのプロジェクトで公開することが殆どです．また，既に研究成果として発表された電子資料よりも高い表現力を持たせた電子資料を新たに作り，別の研究成果として発表されることがあります．わたくしはこの公開方法に執着した研究スタイルは，少し欲張りすぎていると感じます．電子化した資料は URL を持たせた静的なコンテンツとして全てを公開し，(iii) 公開データの web 版コンテンツの素材として使うべきです．結果として作られた web 版コンテンツの表現力は低く，それは世間一般の web 版コンテンツと比べると利用者にとり魅力は低いかもしれません．しかし研究者がやるべきことは，資料の電子化とそのメタデータの記述で，保存性をなくしてまで表現力の高いコンテンツを作ることではありません．もし利用者が公開されている情報をもっと便利に使いたい，新しい発見に使いたいと思うようであれば，そのような利用者に，より表現力の高い web 版コンテンツを作ってもらえばよいのです．オープンデータの哲学は，利用者側に利用形態を選択させることです．研究者の役割は，素材として公開する資料が学術的に間違っていないこと，後世に伝えるべき情報を漏れなく記述し公開することと，わたくしは考えています．研究者と研究機関の役割はデータを作成し公開するまで，その相互利用・再利用は利用者に委ねます．もちろん，この利用者に研究者も含まれています．すると，保存版を制作した後に自らが求める表現や操作性を実現する展示版を作るメタ研究ができることになります.

7 50 年後のみんぱくに向けた期待

みんぱくはとても難しい仕事をしていると思います．扱う対象は紙媒体や物にとどまらず，無形の音や踊り，さらには社会や風景までも対象とするように，万物を扱う責任があります．またその研究方法はフィールドワークという，数値化も定型化も何もないところから情報を取り出す，まさに情報の 1 次産業ともいえる現場での作業が主体です．その学術活動を担う研究者には，強い好奇心と，それを満たす行動を支える体力，フィールドワークの現場でも流されない明晰な頭脳が求められます．またフィールドから上がる情報は，計算機科学や IT 業界の常識では扱えないことが多く，研究者自らがデータの整理・入力まで参加することが求められてきます．みんぱくがこの 50 年間，フィールドワークを伴う研究分野への技術導入で常に先端におられたことには心から敬意を表します．これからの 50 年においても，フィールドワークからの情報収集，全国の人文学者に対する電子化支援，新しい記録技術の導入実験など，これまでの活動の継続を期待しております.

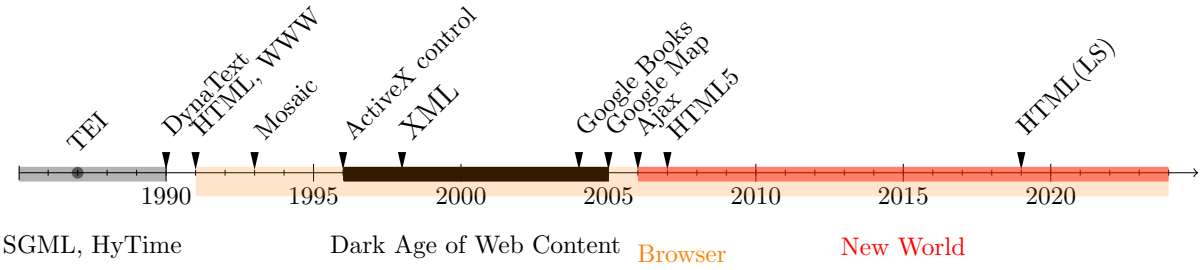
本稿では，DH と電子資料の歴史から web 世界での博物館の今後の役割，D-GLAM の姿を予測・提言してみました．みんぱくも D-GLAM に参加することになる学術機関と推察します．そこで，機関内部での電子情報の管理体制からは離れ，web 版コンテンツについてみんぱくに期待したいことを以下に挙げてみます.

1. 保存版と展示版の両方を作る.
2. 保存版では DB を隠蔽には使わない.
3. 保存版では全ての素材は独立したファイルとする (Open Data).
4. 保存版では全ての素材は機関が決めたひとつの名前付与規則でファイル名が付与される.

5. データ公開よりもデータ作成の優先順位が高い.
6. フィールドワーク研究はリアルに限定される (digital research, digital realm との線引 [18][30]).

わたくし個人の研究活動 (言語ドキュメンテーション) の経験からもう一点，希望を申し上げれば，人文学研究で扱う情報の電子化は，各所で進められるようになりましたが，そのデータの将来の管理先が不確実であるのが現在の状況です．これは日本に限ることではありませんが，現在は作るまでの予算はあるものの，継続した公開までを安定して実現できているプロジェクトがほとんどありません．共同研究機関であることを利点として，また災害に強い千里丘陵の立地を活かして，みんぱくには日本の電子化拠点になって欲しいと願っています．とりわけ，データの保存について期待させてください．個人研究者のデータだけではなく，プロジェクトベースで作成されたデータや，倒産した大学が公開していた研究データなどを受け入れる体制がみんぱくに作られますことを期待しております.

付録 A　マークアップ言語とプラットフォーム年表



付録 B　発表スライド

Happy 50 years Anniversary!

D-GLAM and Digital Content Edition for Preservation
Lessens from a History of Digital Humanities(DH)

Kazushi Ohya
Tsurumi University

2024-11-17

*Meaning of Digital Humanities(DH) is “contemporary humanities”.

Digital-GLAM(D-GLAM)

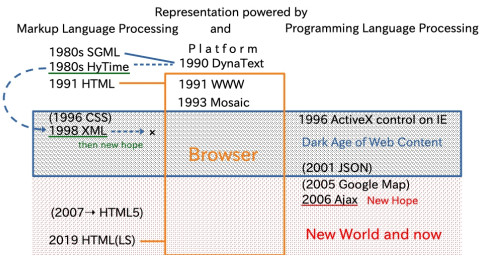
(α) Digital Web Content
for Exhibition ----- make it as you like
for Preservation – (1) Living Preservation (iii) data on public platform
a)do not use DB for hiding content
b)do not rush to reciprocal use or shares of content
c) all of the resources have own URL

(β) matadata
a)do not use DB for hiding metadata
b)do not rush to reciprocal use or shares of metadata
c)do not care about standards
d)Not a normative but a descriptive way is important.

DB can be used for
1)retrieval
2)assembly
3)hiding
4)organizing.

Making content and metadata is prior to making them published. Do not persist in making data on attractive web pages. Making all the data be open, i.e. open data. If the data is open to users who want to use them, the users would make the data something more accessible and attractive instead of researchers.

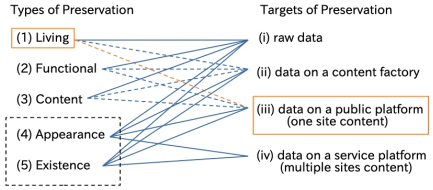
History of Web Content and Representation



Suggestions for 2074

1) making Digital Content Edition(DCE) for preservation, or preservation and exhibition.
2) not using DB for hiding content
3) all the resources of DCE have own URL (Open Data)
4) all the resources (files) of DCE are named according to one rule in the institution.
5) the primary task is making digital content, not making them be public.
6) restricting fieldwork research to a real world(not digital research or digital realm)

Analysis of Preservation of Digital Content Edition



参考文献

[1] **Berners-Lee,T.** (1999) *Weaving The Web*, Harper Business

[2] **Bowles, F.A. ed.**(1967) *Computers in Humanistic Research*, Prentice-Hall

[3] **Burnard,L.** (1988) "Report of Workshop on Text Encoding Guidelines", *Literary and Linguistic Computing* vol.3, issue 2, Oxford University

[4] **Busa, R.** (1951) *Sancti Thomae Aquinatis hymnorum ritualium varia specimina concordantiarum – A First Example of Word Index automatically Compiled and Printed by IBM Punched Card Machines –*, Bocca

[5] **Cempellin,L. and Crawford,P. eds.** (2024) *Museum Studies for a Post-Pandemic World*, Routledge

[6] **Corrado,E.M. and Sandy,H.M.** (2017) *Digital Preservation for Libraries, Archives, and Museums*, Rowman & Littlefield Publishers

[7] **CPA and RLG** (1996) *Preserving Digital Information*, CLIR Report pub63, CLIR

[8] **DeRose,S.J. and Durand,D.G.**(1994) *Making Hypermedia Work*, Kluwer Academic Publishers

[9] **Dormann,W. and Plakosh,D.**(2008) "Vulnerability Detection in ActiveX Controls through Automated Fuzz Testing", Cert Coordination Center, Carnegie Mellon University

[10] **Fiormonte,D., Chaudhuri,S. and Ricaurte,P. eds.**(2022)*Global Debates in the Digital Humanities*, University of Minnesota Press

[11] **Francis,W.N.** (1964) *Manual of Information to accompany A Standard Sample of Present-Day Edited American English, for Use with Digital Computers*, Brown University

[12] **Gold,M.K.ed.**(2012)*Debates in the Digital Humanities*, University of Minnesota Press

[13] **Gold,M.K and Klein, L.F. eds.**(2016)*Debates in the Digital Humanities 2016*, University of Minnesota Press

[14] **Gold,M.K and Klein, L.F. eds.**(2019)*Debates in the Digital Humanities 2019*, University of Minnesota Press

[15] **Gold,M.K and Klein, L.F. eds.**(2023)*Debates in the Digital Humanities 2023*, University of Minnesota Press

[16] **Goldfarb,C.F.** (1991) *The SGML Handbook*, Oxford University Press

[17] **Greenberger,M. ed.**(1962) *Computers and the World of the Future*, MIT Press

[18] **Hargittai,E. and Sandvig,C.** (2015) *Digital Research Confidential*, MIT Press

[19] **Hindley,M.**(2013) "The Rise of the Machines", *Humanities* Vol.34, No. 4, the National Endowment for the Humanities

[20] **Hockey,S.**(2004)"The History of Humanities Computing", *A companion to Digital Humanities*, Blackwell Publishingor

[21] **Ide,N. and Véronis,J. eds.**(1995)*Text Encoding Initiative*, Kulwer Academic Publishers

[22] **Johnson,J.M., Mimno,D., and Tilton, L. eds.** (2024) *Computational Humanities*, University of Minnesota Press

[23] **Jones,T. and Simpson,D.** (2020) "Websitea as a publishing platform", *The Routledge International Handbook of New Digital Practices in Gallaries, Libraries, Archives, Museums and Heritage*

Sites, Routledge

[24] **金沢勇二** (2004) 「宮内庁正倉院事務所所蔵「聖語蔵経巻」のカラー CD-R 化」 『日本写真学会誌』67 巻 2 号，日本写真学会 (Kanazawa,Y., 2004, "Color Digital Recording on CD-R of Shogozo Scrolls owned by the Office of the Shosoin Treasure House, Imperial Household Agency, Japan", *Journal of The Society of Photography and Imaging of Japan* Vol.67, The Society of Photography and Imaging of Japan)

[25] **Kenney,A.R. and Rieger,O.Y.** (2000) *Moving Theory into Practice*, RLG

[26] **Lancashire,I. ed.** (1991) *The Humanities Computing Yearbook 1989-1990*, Clarendon Press

[27] **Lewi,H., Smith,W., Lehn,D., and Cooke,S. eds.** (2020) *The Routledge International Handbook of New Digital Practices in Gallaries, Libraries, Archives, Museums and Heritage Sites*, Routledge

[28] **MacCarty,W.** (2005) *Humanities Computing*, Palgrave

[29] **Miller,R.R., Causey,J.P., Moore,G.W., and Wilk,G.E.** (1988) "Development and Operation of a MUMPS Laboratory Information System: A Decade's Experience", *Proceedings, Symposium on Computer Applications in Medical Care*, American Medical Informatics Association

[30] **Noiret,S., Tebeau,M. and Zaagsma,G. eds.** (2022) *Handbook of Digital Public History*, Walter de Gruyter GmbHp

[31] **O'Sullivan,J. ed.** (2022) *The Bloomsbury Handbook to the Digital Humanities*, Bloomsbury Academic

[32] **Parry,R.** (2007) *Recoding the Museum*, Routledge

[33] **Sabharwal,A.** (2015) *Digital Curation in the Digital Humanities*, Chandos Publishing

[34] **Saracevic,T. and Dalbello,M.** (2005) "Digital library research and digital library practice: how do they inform each other?", unpublished, e-LIS, `http://eprints.rclis.org/6706/`

[35] **Schats,B. and Chen,H.** (1996) "Building Large-Scale Digital Libraries", *Computer* vol.29 no.5, IEEE

[36] **Schreibman,S. Siemens,R. and Unsworth,J.eds.**(2004)*A Companion to Digital Humanities*, Blackwell Publishing, `https://companions.digitalhumanities.org/DH/`

[37] **Schreibman,S. Siemens,R. and Unsworth,J.eds.**(2016)*A New Companion to Digital Humanities*, Wiley Blackwell

[38] **Smith,W., Lehn,D., Lewi,H., Constantinidis,D., and Best,K.** (2020) "The experience of using digital walking tours to explore urban histories", *The Routledge International Handbook of New Digital Practices in Gallaries, Libraries, Archives, Museums and Heritage Sites*, Routledge

[39] **Solopova,E. ed.**(2000) *The General Prologue* on DynaText, Cambridge University Press

[40] **Sperberg-McQueen,C.M. and Burnard,L. eds.**(1997) *Guidelines for Electronic Text Encoding and Interchange*, Electric Book Library Vol.2, Electronic Book Technologies

[41] **Tasman, P.** (1957) "Literary Data Processing", *IBM Journal of Research and Development* vol.1, issue 3, IBM

[42] **Terras,M., Nyhan,J., and Vanhoutte, E. eds..**(2013)*Defining Digital Humanities*, Ashgate

[43] **Yale University** (1965) *Computers for the Humanities ? – A Record of the Conference Suponsored by Yale University on a Grant from IBM, January 22-23, 1965*, Yale University

[44] **Wisbey,R.A. ed.**(1971) *The computer in literary and liguistic research*, Cambridge University

Press

[45] Ajax, 2005, https://web.archive.org/web/20150910072359/http://adaptivepath.org/ideas/ajax-new-approach-web-applications/

[46] Ars Electronica Conference, https://ars.electronica.art/news/en/

[47] Digital Research in Humanities and the Arts, https://drha.tech/

[48] ACL COLING, https://aclanthology.org/venues/coling/

[49] DynaText, WikiPedia, https://en.wikipedia.org/wiki/Dynatextl

[50] DLF-Forum(Digital Library Federation Forum), https://www.diglib.org/

[51] Kenderdine,S. (2023) DH2023 Keynote, https://dh2023.adho.org/opening-keynote-july-11/

[52] DH2023 Keynote Video, https://vimeo.com/847301655

[53] HTML Living Standard, whatwg, https://html.spec.whatwg.org/multipage/

[54] HTML2.0, 1995, RFC1866, IETF, https://datatracker.ietf.org/doc/html/rfc1866

[55] HyTime 2d (working edition, not authorized ISO version), ISO/IEC JTC 1/sc 18 WG8 N1920rev, https://web.archive.org/web/20070430101241/http://www1.y12.doe.gov/capabilities/sgml/wg8/document/n1920/pdf/n1920.pdf

[56] International Conference on Theory and Practice of Digital Libraries (TPDL), http://www.tpdl.eu/

[57] Javascript, 1995, https://web.archive.org/web/20070916144913/https://wp.netscape.com/newsref/pr/newsrelease67.html

[58] JSON, https://www.rfc-editor.org/rfc/rfc8259.txt

[59] LREC Conferences, http://www.lrec-conf.org/

[60] the Open University, "Digital Humanities: humanities research in the digital age", https://www.open.edu/openlearn/history-the-arts/digital-humanities-humanities-research-the-digital-age/

[61] METS; Metadata Encoding and Transmission Standard, The Library of Congress, https://www.loc.gov/standards/mets/

[62] MUMPS, 1977, http://71.174.62.16/Demo/AnnoStd?Frame=Main&Page=a100002&Edition=1977

[63] The Poughkeepsie Principles, TEI, https://tei-c.org/Vault/ED/edp01.htm

[64] Resource Description Framework(RDF), https://www.w3.org/TR/PR-rdf-syntax/Overview.html

[65] TEI by Example https://teibyexample.org/exist/

[66] Reports from the W3C SGML ERB to the SGML WG and from the W3C XML ERB to the XML SIG, W3C, https://www.w3.org/XML/9712-reports.html

[67] "The Rise and Rise of JSON", 2017, Two-Bit History, https://twobithistory.org/2017/09/21/the-rise-and-rise-of-json.html

[68] The WEB Conference, ACM, https://thewebconf.org/

[69] XML1.0 (1998) "Extensible Markup Language (XML) 1.0", W3C, https://www.w3.org/TR/xml/

[70] XMLHttpRequest, 2006, W3C, https://web.archive.org/web/20080516060525/http://www.w3.org/TR/2006/WD-XMLHttpRequest-20060405/

[71] XML Families, W3C, https://www.w3.org/TR/?filter-tr-name=xml

[72] **Ohya,K.**(1999) “Introduction to new text-based data management – For those who are tired of re-writing a list of corpora on the occasion of presentation–”, *Journal of Chiba University Eurasian Society* No.2, Chiba University

[73] **大矢一志，土屋俊** (2000)「システムが決まらなければデータベースが出来ないというのは本当か –テキストベースデータモデル利用の提案–」『第 2 回アートドキュメンテーションフォーラム報告書』 アートドキュメンテーション研究会 (Ohya,K. and Tutiya,S., 2000, “A Proposal to Use Text-based Data Models for Pre-processes of Making Databases and Data Preservation”, *Art information towards the next millennium : proceedings of the 2nd forum on art documentation*, Japan Art Documentation Society)

[74] **大矢一志** (2006)「マークアップの課題を syntax から見た分類と解決のステップ」『TEI Day in Kyoto 2006 報告書』, 京都大学 (Ohya,K., 2006, “Markup problems:Syntactical analysis and steps to their resolution”, *TEI Day in Kyoto 2006: Abstracts*), Kyoto University)

[75] **大矢一志** (2009) 「少数言語コーパス向け記述データの構造」『情報処理学会シンポジウムシリーズ』 Vol.2009, No.16, 情報処理学会 (Ohya,K., 2009, “Data Structure for Minority Language Corpora”, *IPSJ Symposium Series* Vol.2009, No.16, IPSJ)

[76] **大矢一志** (2010)「大判資料 (古地図等) の分割撮影向け簡易撮影台の作成」『鶴見大学紀要』Vol.47 part 4, 鶴見大学 (Ohya,K., 2010, “ Portable Camera Frames to Take Multiple Shoots for Large Documents like Maps”, *The Bulletin of Tsurumi University* vol.47,part 4, Tsurumi University)

[77] **大矢一志** (2011)『人文情報学への招待』 神奈川新聞社 (Ohya,K., 2011, *Introduction to the Digital Humanities*, Kanaga Syinbun)

[78] **Ohya,K.**(2014)“Unit-based Scheme Connection Between TEI and Original Scheme To Promote Data Sharing Beyond Cultural Diversities” TEI 2014 `https://docsci.infon.org/stack/ohya2014.zip`

[79] **大矢一志** (2017)『人文情報学読本 –胎動期編–』神奈川新聞社 (Ohya,K., 2017, *Digital Humanities Reader – The Quickening Period –*, Kanagawa Shimbun)

[80] **Ohya,K.** (2022) “An Architecture of resolving a multiple link path in a standoff-style data format to enhance the mobility of language resources”, *Proceedings of the 13th Conference on Language Resources and Evaluation(LREC2022)*, European Language Resources Association(ELRA) `https://aclanthology.org/2022.lrec-1.307`

[81] **Ohya,K.** (2023)“Criteria to emancipate content providers from obsession with specifications for content preservation and propositions as guidelines on making content for easy reuse in the future”, ICOM-CIDOC2023, `https://docsci.infon.org/stack/ohya2023.pdf`

[82] **大矢一志** (2024)「でんしかしよう！」『書物学』Vol.25, 勉誠社 (Ohya,K.,2024,”Denshika Shiyou(double meanings of ’Requirements for making digital content’ and ’let us do making digital content’, *Syomotsugaku* Vol.25, Benseisya)”)