

人文学と情報学の界面で

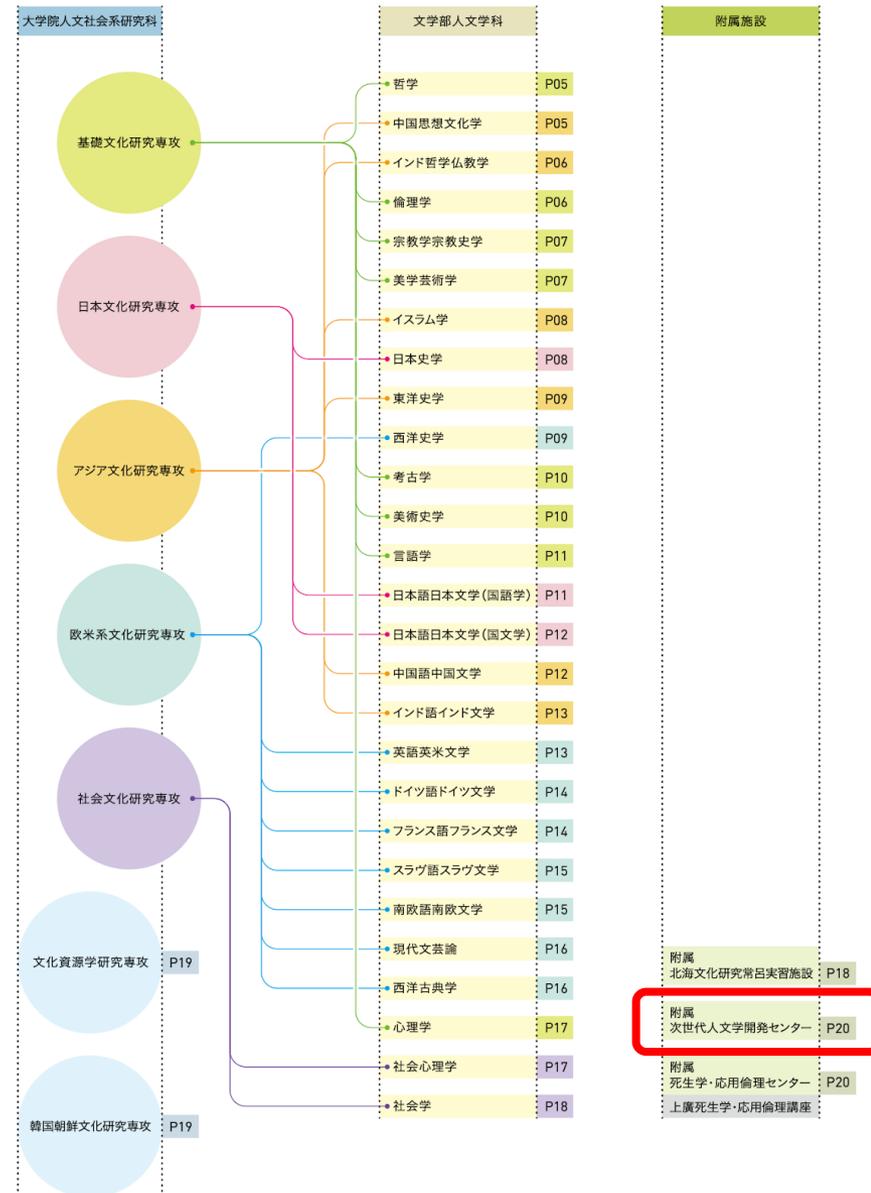
東京大学大学院人文社会系研究科

大向 一輝

2020年12月5日

自己紹介

- 所属 (2019年～)
 - 大学院人文社会系研究科・文学部
 - 附属次世代人文学開発センター
 - 人文情報学部門
- テーマ
 - ウェブ情報学
 - 知識表現
 - ソーシャルメディア
 - オープンデータ
 - 学術コミュニケーション
 - 学術情報システム
 - オープンサイエンス
 - 人文情報学



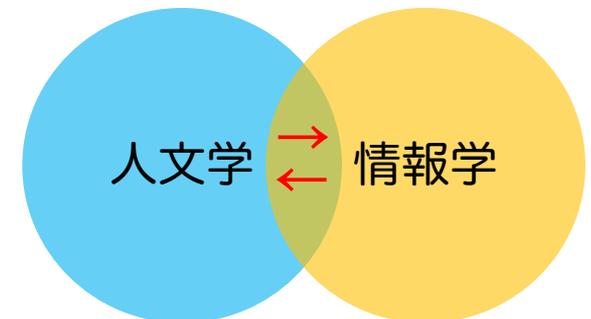
自己紹介

- データベースとの関わり
 - CiNii：論文・図書
 - NACSIS-CAT/ILL：大学図書館の総合目録
 - 文化庁メディア芸術データベース：現代文化
- 社会との関わり
 - 株式会社グルコース取締役
 - 特定非営利活動法人リンクト・オープン・データ・イニシアティブ
副理事長
 - IT総合戦略本部官民データ活用推進基本計画実行委員会オープン
データワーキンググループ構成員
 - 知的財産本部デジタルアーカイブジャパン実務者検討委員会デジタル
アーカイブ全体戦略ワーキンググループ・ジャパンサーチワーキ
ンググループ構成員

デジタル人文学の定義

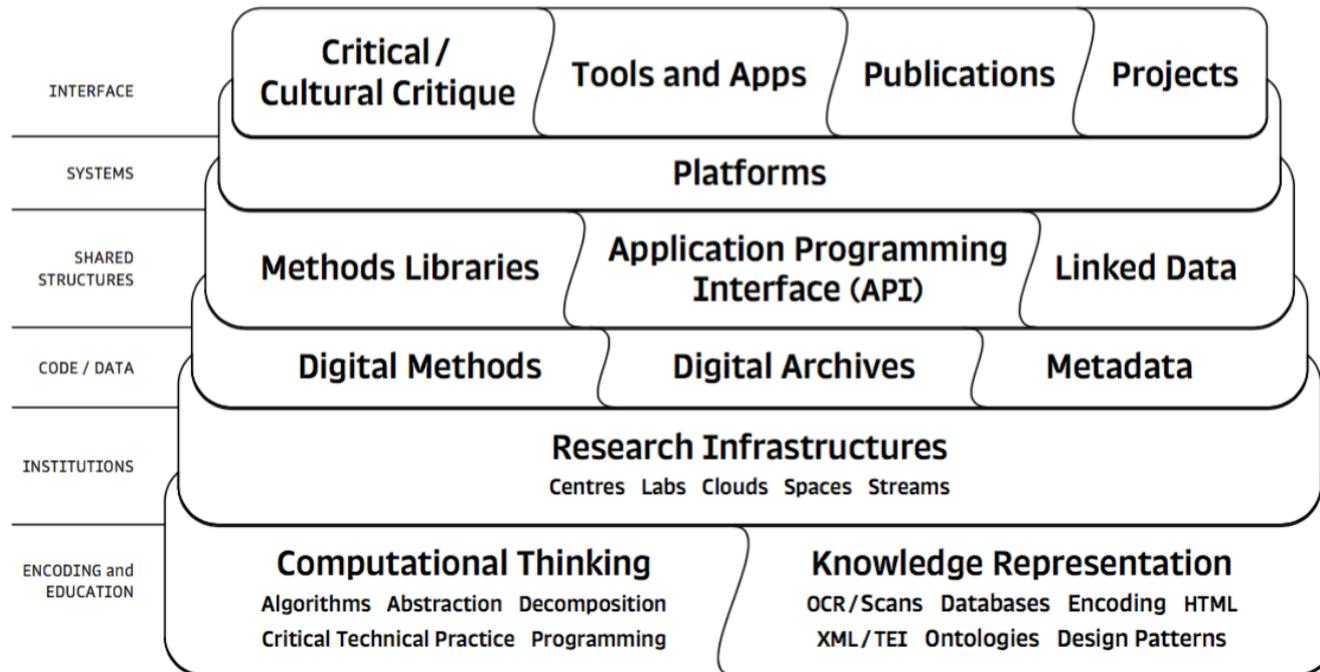
- Digital humanities is work at the intersection of **digital technology** and **humanities disciplines**. [UCLA Center for Digital Humanities]
- The digital humanities sit at the crossroads of **computer science** and **the humanities**. [Stanford Humanities Center]
- Digital Humanities refers to new modes of scholarship and institutional units for collaborative, transdisciplinary, and **computationally** engaged research, teaching, and publication. [Schnapp2012]
- デジタル・ヒューマニティーズとは、**人文学的問題**を**情報学的手法**を用いて解くことにより新しい知識や視点を得ることや、人文学的問題を契機として新たな情報学の分野を切りひらくことなどを旨とする、**情報学**と**人文学**の融合分野である。[北本2015]

1. 人文学研究への情報技術の適用
2. 情報学研究への人文学の知見の適用



デジタル人文学の要素技術

- **UCLA Introduction to Digital Humanities**
 - **蓄積**：データベース・メタデータ・オントロジー
 - **可視化**：地理情報システム・仮想空間
 - **応用**：データマイニング・テキスト分析
- **The Digital Humanities Stack**



デジタル人文学の事例 (1/6)

- Google Books Ngram Viewer
 - 北米を中心とした大学図書館所蔵資料のデジタル化 (2003~)
 - 資料群に含まれる単語の頻度分布を時系列で表示する
 - 「遠読」 (Distant Reading) による新たな知見の抽出

Google Books Ngram Viewer

Graph these comma-separated phrases: (The United States is + The United States has)/The United States,(T case-insensitive

between 1800 and 2000 from the corpus English with smoothing of 3

 共有

 Tweet

[Embed Chart](#)



デジタル人文学の事例 (2/6)

- 潜在的トピックによる史料分類と関連史料の提示
 - 史料群に含まれる単語を機械学習を用いて自動分類
 - トピックごとに役割が類似した単語が集約
 - トピックの分布が類似した文書を関連史料として推薦

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
1 彌宜之	大津	入道々	御房事	コノ一行挿 入セラ	一々	一々	謹言上	阿闍梨御房 上下二	如件也	住人百	下村事
2 神主十	入寺供	合戦之	執金	号文書ト同 筆ニカ	不可也	百文之	恐クハ	僧都也	執金	房寺務之	覚円内
3 足利将軍家 執事高	阿闍梨御房 上下二	日々	如件也	文書トシテ 収メタ	本所之	同々	之由事	法印也	依会	有識免一 宇中揖里	行清寄附 徳禅寺之
4 度例	法師三	行損	仍うりけん の状如	年預之	住持之	三々	之闕	権少僧都法 眼和尚	雑掌之	垣内券	観空状
5 第二二	年預之	元弘三年七 月十七	謹下	五師代	前机	十二月中可 致沙汰	一々	権大僧都法 眼和尚	在判下	入寺免一 宇下居又	次徒弟院 事
6 義満公 7 九号文	入か 寺々	御奉書 進上之	候了 僧正之	アリシ ナリト	同前一 吉村	二〇 三十ケ日	候あ 候哉い	権懇懃之 護摩師	可被下 状如此候	有麦作人 八騰三	左中將 田村々
8 尊氏判	一自	状如此候	如此之	コノ文書ノ 差出者	二〇	五十八	御々	御々	殿いれ候	反定田五 斗真長房	花幔
9 参重陸文	集会事	如件也	天氣	第一三	可為三分之 百人		可被下	僧正之	之理	四騰持之	名田庄内 田村下村
0 文敷	行事也	同々	之由事	ホア	新条	五十枚初後 夜導師	之処二	金剛定院御	左衛門大尉 中原朝	六興	鶴鳴保

デジタル人文学の事例 (3/6)

- 時間基盤情報
 - 史資料に記載された日付の構造化とID付与
 - 異なる暦法間の相互変換
 - あいまいな時間表現の明示的な記述規則を定義

About: 2019-01-14

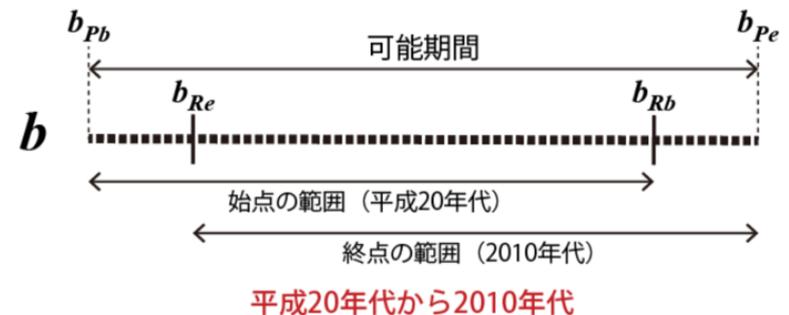
<http://datetime.hutime.org/date/2458497.5>

C.E. 2019 January 14
(Gregorian Calendar)

C.E. 2019 January 14
(Julian/Gregorian Calendar (reformed in 1582))

08 5779 שבת
(Hebrew Calendar-TEST)

平成31年1月14日
(Japanese Calendar (Southern Court))



デジタル人文学の事例 (4/6)

CiNii Books

- データベース同士をつなぐ
 - IDと名寄せ
 - NDLデジタルコレクション・新日本古典籍総合データベース・HathiTrust

The image illustrates the integration of three digital library systems: CiNii Books, NDL Digital Collection, and HathiTrust. Yellow arrows point from the CiNii Books interface to the NDL Digital Collection and HathiTrust interfaces, indicating data flow and linking between these platforms.

CiNii Books Interface (Top): Shows search results for '古今和歌集' (Kokin Wakashū). The interface includes a search bar, navigation tabs (All records, Images, Metadata, Content), and detailed information for the selected item, such as NII accession number (8814780383) and publication details.

NDL Digital Collection Interface (Bottom-Left): Displays the digital image of the '古今和歌集' manuscript cover. The interface includes a search bar, navigation tabs (All records, Images, Metadata, Content), and detailed information for the selected item, such as NII accession number (8814780383) and publication details.

HathiTrust Interface (Bottom-Right): Shows the digital image of the '古今和歌集' manuscript cover. The interface includes a search bar, navigation tabs (All records, Images, Metadata, Content), and detailed information for the selected item, such as NII accession number (8814780383) and publication details.

デジタル人文学の事例 (5/6)

- 日本古典籍くずし字データセット
 - 古典籍15点の3999文字種・403242文字の画像データを提供
 - 画像認識・機械学習のための教師データ
 - くずし字認識アルゴリズムのコンテスト開催

くずし字データセット



MNISTデータセット (手書き数字)



デジタル人文学の事例 (6/6)

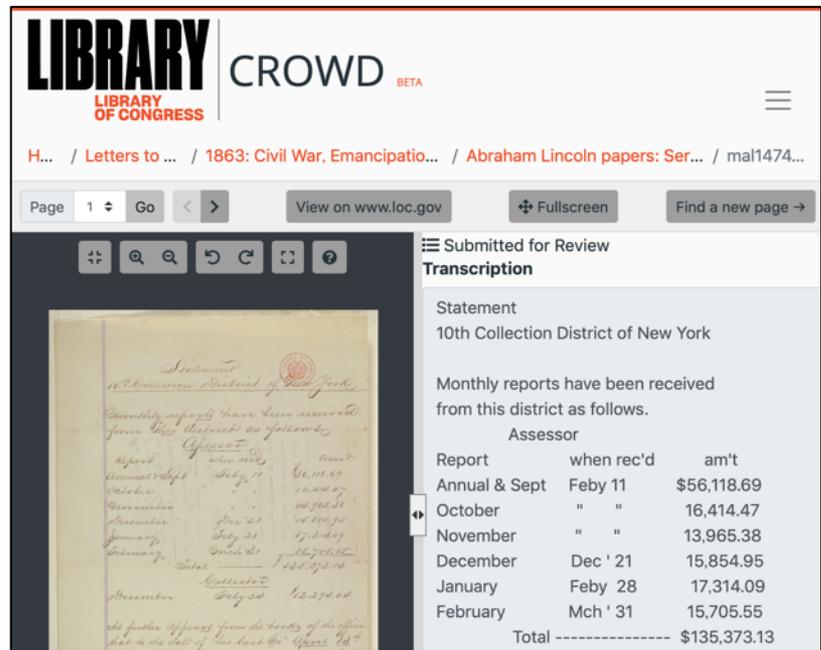
- クラウド翻刻
 - ユーザ参加による史料画像のテキスト化
 - 教材の開発とゲーミフィケーション
 - 参加者4000名超・入力文字数500万超 (みんなで翻刻)

みんなで翻刻 [京都大学古地震研究会]

米国議会図書館



<https://honkoku.org>



<https://crowd.loc.gov>

デジタル人文学の研究コミュニティ

• 国内コミュニティ

- 情報処理学会人文科学とコンピュータ研究会
 - 研究会：年3回
 - 人文科学とコンピュータシンポジウム（じんもんこん）：年1回
 - 2020年12月12日～13日
 - 論文誌
- Japanese Association for Digital Humanities (JADH)
 - 国際会議：年1回
 - 論文誌：日本語・英語

• 国際コミュニティ

- Alliance of Digital Humanities Organizations (ADHO)
 - 国際会議：年1回 **2022年東京開催**
 - 論文誌：Digital Scholarship in the Humanities

情報学から見るデジタル人文学

大規模化

- デジタル情報の保存コストの低減による多様な史資料の蓄積
- 複数の史資料群を横断した検索・データマイニング

人文学
人文知

多層化

- 史資料に内在する知識の明示化
 - 時間・空間・人・対象…
- 知識の同一性・関連性に基づく体系化

共有と協働

- インターネットを通じた分散的な情報源への透過的なアクセス
- 資源共有から共同作業へ

知識インフラ

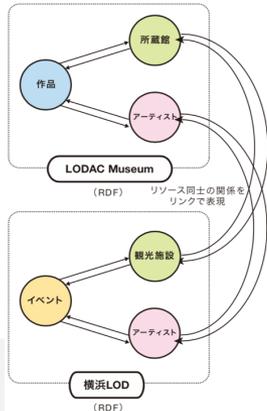
大規模デジタルアーカイブの構築

研究活動のネットワーク化

デジタル人文学の課題

大規模化

- 意味を捨象した「パターン認識」
- 解釈可能性・説明可能性の欠如



人文学
人文知

多層化

- 明示化を誰が担うのか？
- 暗黙的な知識の関連性は多様かつ膨大

共有と協働

- ユーザ参加のインセンティブ設計
- 信頼性が担保できない

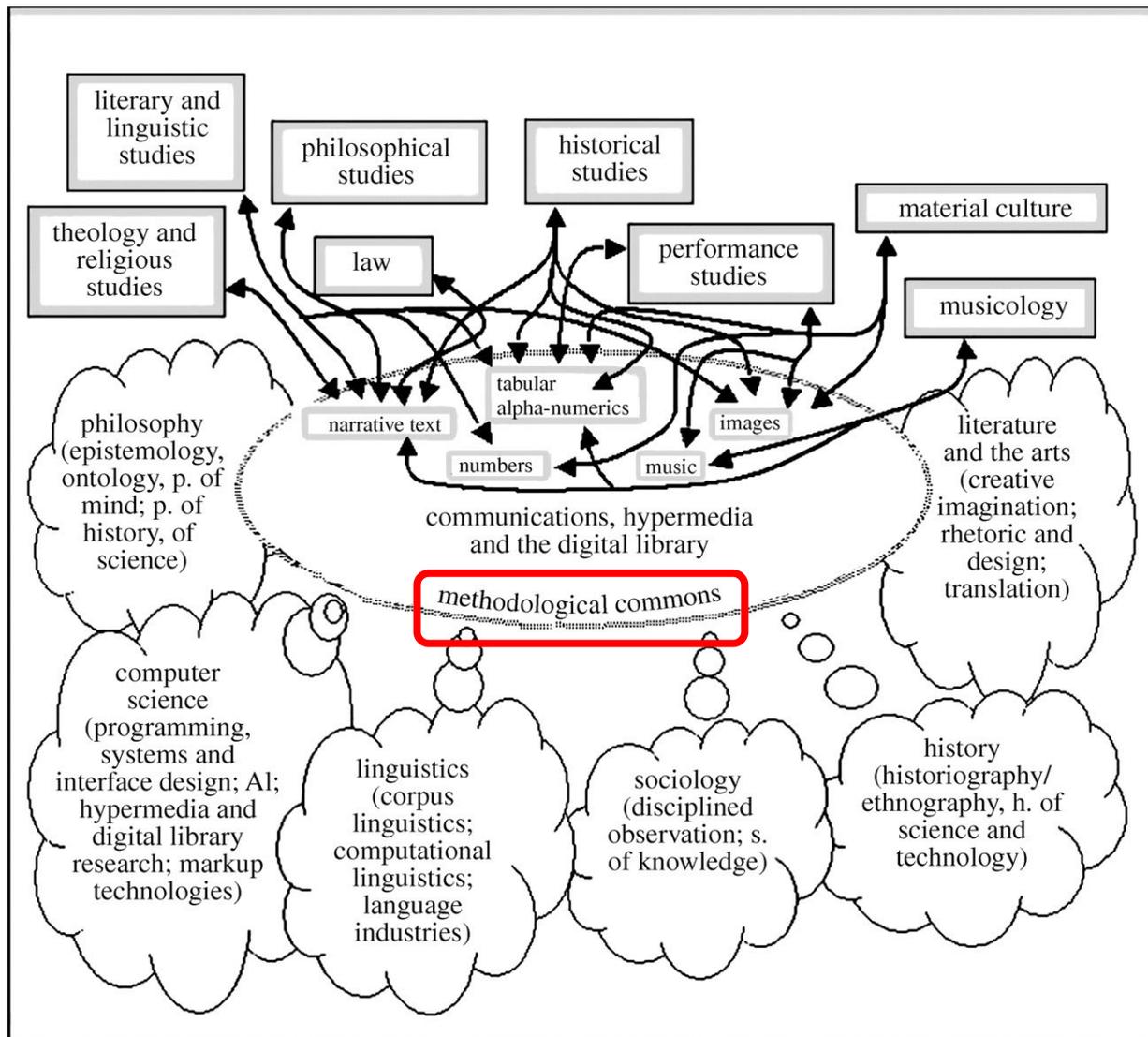


知識インフラ

継続性の担保 社会に対する説明責任

道具として情報学を用いるだけでは解決できない

人文学から見るデジタル人文学



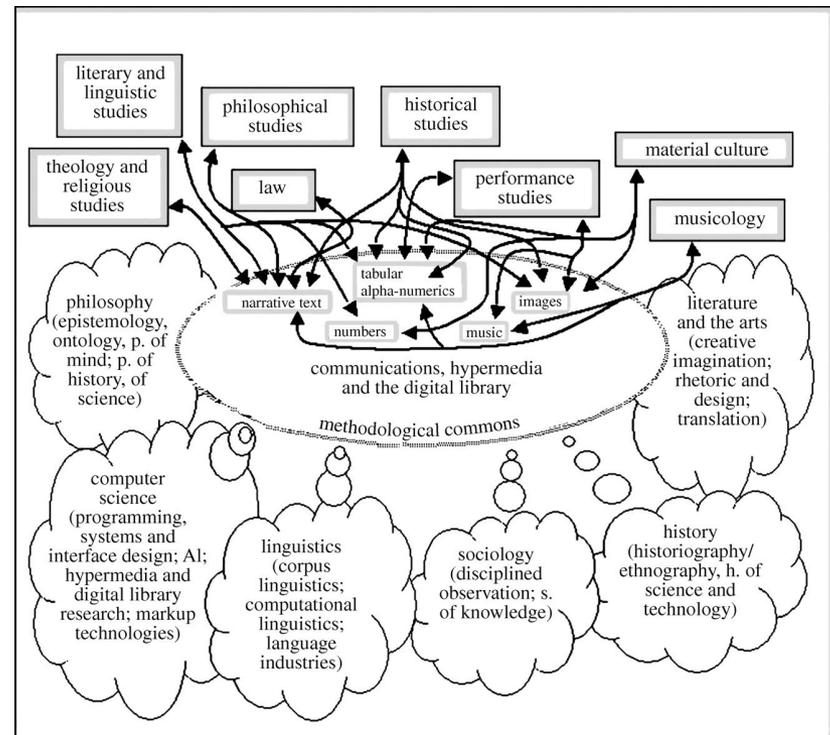
人文学から見るデジタル人文学

「方法論の共有地」

- 研究目的の多様性を前提として
 - 手法の高度化・複雑化
 - 大量の資料
 - 情報技術の導入
 - 手法の共通化に向かう

「他者」との分業

- 異なる職能で構成されるチーム
 - ツールの制作者
 - 資料の提供者
 - それぞれが独自の視点と意思を持つ
- 他者とのコミュニケーション・コラボレーション
 - 相互理解と共通の目的の設定



デジタル人文学の教育

- **カリキュラムの整備**
 - 大学院横断型教育プログラム「デジタル・ヒューマニティーズ」
 - 2012年度より
 - 修了証を発行
 - 内容
 - コア科目
 - 人文情報学概論
 - デジタル・ヒューマニティーズ入門
 - 基礎科目
 - アーカイビング
 - アナリシス
 - プレゼンテーション
 - 関連科目

デジタル人文学の教育

- 課外活動の必要性
 - 附属施設の制約をカバーする
 - 研究指導ができない
 - 授業期間以外の長期的な接点を持つ
- UTDH
 - デジタル人文学を主題とするゆるやかなコミュニティ
 - 方法論の共有・多様な視点の導入
 - 人文情報学勉強会
 - 研究紹介・イベント紹介・報告
 - 共同作業（調査・ツールの評価）
 - 論文執筆の支援・発表練習
 - YouTubeチャンネル
 - 要素技術の解説・イベントの配信

デジタル人文学のこれから

- 教育

- カリキュラムの（再）整備
- 「Invisible College of Digital Humanities」
 - 基本リテラシーに関わる内容は大学間で共有する
 - ツールキットの整備（技術面・制度面）
 - 個別の課題は大学横断型で対応する

- 研究

- 各専門分野への組み込み
 - DH・デジタルアーカイブの普及状況に合わせた対応
 - 例：画像化→テキスト化→知識化
 - それぞれの段階に合わせたデータ構築技術と分析技術の提示