

## 「知の集積と共有」の変革期に"じんもんこん"がなし得ること

永崎研宣<sup>1</sup> 後藤真<sup>2</sup>

1 山口県立大学情報化推進室 2 花園大学 文学部文化遺産学科

小特集「「知の集積と共有」の変革期に"じんもんこん"がなし得ること」を企画した。本小特集では、今までに「人文科学とコンピュータ研究会」が築いてきた研究の蓄積の再確認と、研究環境の変化などについての検討を行うことを目的とする。「人文科学とコンピュータ」と呼ばれる研究領域が明確なものとなって、20年経過した。その間に、インターネットの普及、爆発的なコンピュータ技術の進歩とともに、この研究会は歩いてきた。同時に、人文科学の研究状況も、社会の変容とともに大きく変遷しつつある。そこで、研究会を「歴史」として振り返ることになり、人文科学の変容を身に激しく受けながら進まざるを得ない世代の研究者として、いくつかの議論提起を試みてみたい。世代を超えた議論の先に、新たな地平を見出すことを期待する。

### Possibility of "SIG Computers and the Humanities" in Change period of

#### "common and accumulate knowledge"

Kiyonori NAGASAKI<sup>1</sup> Makoto GOTO<sup>2</sup>

1 Yamaguchi Prefectural University 2 HANAZONO University

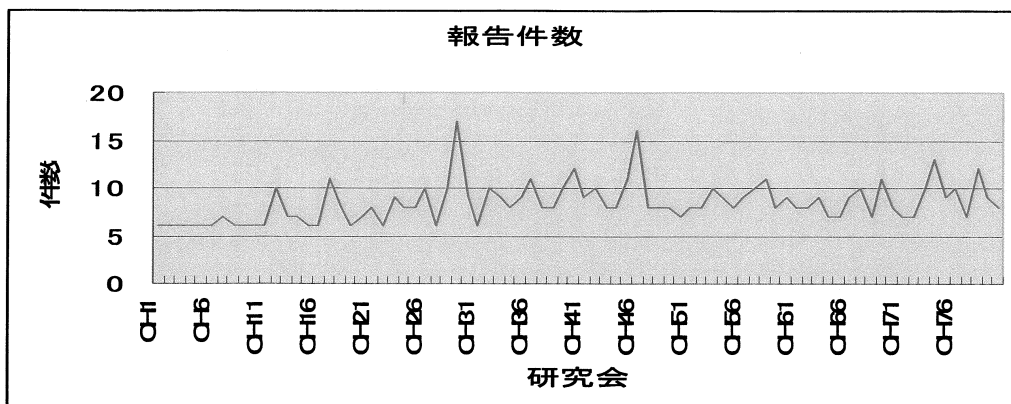
We planned a small special feature. We reaffirm the accumulation of the study that "SIG Computers and the Humanities" built so far. And we perform the examination about changes of the study environment. We are aimed for these both examination.

### 人文科学とコンピュータ研究会—80回の歴史と課題—

後藤 真

本稿は、小特集のための材料とすることを目指した論考である。なお、本稿は分担執筆である。それぞれの原稿の責は、それぞれの執筆者が負う。特に、研究会の集計に関し、集計ミス、分類ミスや、分析はすべて後藤が責めを負うべきものであることをお断りしておく。

人文科学とコンピュータ研究会は1989年5月に第1回研究会を開始し、2009年5月の研究会で、満20歳を迎えることとなった。その間、回数は80回を数えた。そして、人文科学とコンピュータシンポジウム（じんもんこんシンポジウム）は、1999年に国立民族学博物館で第1回を開始後、2008年12月の筑波大学での開催で第10回、満10年を迎えることとなった。この80回の研究会と、10回のシンポジウムにおいて、発表された件数は実に1000件を超える。そこで、報告者の一人として、まずは、この20年の歩みを振り返り、その成果と課題を検討する材料を提供したい。果たして、この20年の間に、何が大きく変わり、何をさらに推し進めるべきなのか。深く掘り下げることはできないが、今後の議論の射程のため、若干の情報を提供することを目指す。



### 1. 人文科学とコンピュータ研究会の「サービス」

ここでは、まず、研究会報告を、「集計・分類」した結果を見てみる。研究会報告の件数は676件である。ここで、分類のために特に着目した点は、発表において検討した題材が人文科学のいかなる分野にあたるものなのか、である。人文科学と一口に言っても、その学のある方はさまざまである。この研究会はその多くの学の中で、どのような学問を主たる対象としてきたのか、概括ではあるが、確認してみたい。

分類した体系は表1のとおりである。分類は、論題を中心に行った。論題のみで判別できないものは、できる限りにおいて、その内容を確認した上での分類を行った。分類は自動的判別などを用いず、報告者による判断である。そのため、詳細なルールを現状では説明できるものではない、暫定的なものであることを、最初に断っておく。

分類のルールは以下のとおりである。

1. 技術的な研究が圧倒的に主であり、ほぼ人文科学の資料を用いていないと思われる研究は、技術研究として、無理に分類を行わなかった。
2. ただし、技術開発研究でも「資料となる対象」がはっきりしているものは、その資料に則して分類を行った。
3. デジタル・アーカイブに関わる総括的な研究は「総論」とした。
4. 文字研究は、その件数が多いため、独自に項目を作成した。
5. それ以外の分類しがたい研究は「その他」に分類した。

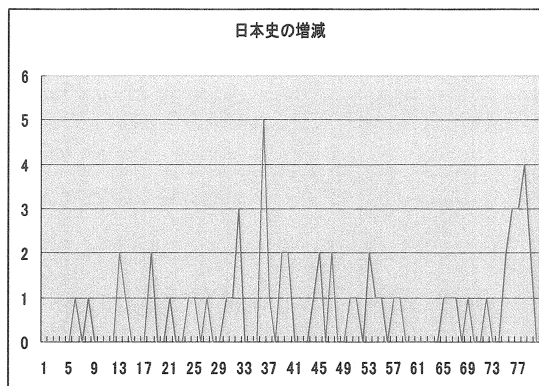
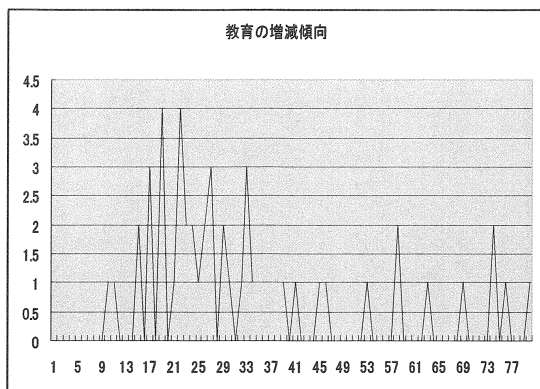
報告論題名が、資料のみの場合、厳密に対象となる学問が明確でないものもある。たとえば、仏教文献は、歴史学とも仏教学とも、はたまた書誌学とも、その題目のみではわからないものも多い。また、報告自体が、学の方向性を明示していないものもある。また、古文書研究などにおいては、その研究を文字研究とすべきか、日本史学の研究とすべきかなど、判断に迷うところも多い。そのため、その報告者の意図と分類が一致せず、不完全な集計であることも、前提としていただいたうえで、あくまでも、大まかな傾向としてみていただければ幸いである。その成果のうち、一部をまとめて、表示したものが、表1である。

大きな傾向としては、以下のようなものが見てとれる。

1. 文字研究は、近年増加傾向を見せる。
2. 教育関連は、2000年前後までは非常に多くを占めるが、近年漸減傾向にある。
3. モーションキャプチャを中心とする、「舞踊」など、芸術分野の増加傾向が顕著。
4. 歴史研究分野は、CH35 近辺と、近年に大きな山がある。「東洋史」の研究分野がほとんどない。

	件数	総論	技術	文学	歴史	文字	語学	宗教	哲学	考古	地図	民族	民俗	芸術	教育	心理	メタ	博物	他
1	6	2	0	0	1	0	0	0	0	1	0	0	0	2	0	0	0	0	0
2	6	1	3	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
3	6	1	1	1	0	1	0	0	0	2	0	0	0	0	0	0	0	0	0
4	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0
5	6	0	1	0	1	0	0	0	0	0	2	1	0	1	0	0	0	0	0
6	6	0	0	0	3	0	0	1	0	0	0	0	1	1	0	0	0	0	0
7	7	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	1	0
8	6	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	1	2	0
9	6	1	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	1	1
10	6	0	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	1	1
11	6	1	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	1
12	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0
13	7	0	0	0	4	0	0	0	0	0	0	1	0	0	0	0	0	0	2
14	7	0	2	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0
15	6	0	2	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0
16	6	0	1	1	0	0	2	0	0	1	0	0	0	0	0	0	0	0	1
17	11	1	0	2	0	1	0	0	1	0	0	0	0	0	3	0	0	2	1
18	8	0	2	1	2	1	1	0	0	0	0	0	0	0	0	0	0	1	0
19	6	1	0	0	0	0	1	0	0	0	0	0	0	0	4	0	0	0	0
20	7	0	2	0	0	0	0	0	0	0	0	0	0	3	0	1	0	1	0
21	8	0	4	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	1
22	6	0	0	0	0	0	1	0	0	1	0	0	0	0	4	0	0	0	0
23	9	0	0	2	0	1	0	0	0	2	0	0	0	2	2	0	0	0	0
24	8	0	2	0	1	0	0	0	0	1	0	0	0	1	2	1	0	0	0
25	8	2	1	0	1	0	1	0	0	1	0	0	0	1	1	0	0	0	0
26	10	0	1	2	0	2	0	0	0	0	0	0	0	0	2	0	0	2	1
27	6	0	0	0	1	1	0	0	0	0	0	0	0	1	3	0	0	0	0
28	10	0	3	0	0	2	1	0	0	0	0	0	0	0	0	0	1	2	1
29	17	0	5	1	0	0	4	0	0	1	0	0	0	1	2	2	0	0	1
30	9	0	1	0	1	2	2	0	0	0	0	1	0	1	1	0	0	0	0
31	6	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2
32	10	0	2	1	3	2	0	0	0	1	0	0	0	0	1	0	0	0	0
33	9	0	1	1	0	0	1	0	0	1	0	0	0	1	3	1	0	0	0
34	8	1	0	0	1	0	0	0	0	1	0	0	0	2	1	0	0	1	1
35	9	0	1	0	0	2	1	0	0	0	1	0	1	2	1	0	0	0	0
36	11	0	0	0	5	1	0	0	1	0	0	0	0	2	1	0	0	0	1
37	8	0	0	2	1	0	0	0	1	0	0	0	2	1	1	0	0	0	0
38	8	0	4	0	0	0	1	0	0	0	0	0	0	2	1	0	0	0	0
39	10	0	1	2	2	1	1	0	0	0	0	0	0	1	1	0	0	1	0
40	12	0	2	1	2	0	2	1	1	1	0	1	0	1	0	0	0	0	0
41	9	0	1	2	0	0	1	0	0	0	0	0	0	4	1	0	0	0	0

42	10	1	3	0	3	0	1	0	0	1	0	0	0	1	0	0	0	0	
43	8	0	3	1	0	0	0	1	0	1	1	0	0	0	0	0	1	0	
44	8	0	3	2	1	0	1	0	1	0	0	0	0	0	0	0	0	0	
45	11	0	2	1	3	0	1	0	0	2	0	0	0	0	1	0	0	1	
46	16	0	0	1	4	0	0	0	0	0	2	0	0	6	1	0	1	0	
47	8	0	1	2	2	1	0	0	0	1	0	0	0	0	0	0	0	1	
48	8	0	0	5	1	0	0	0	0	0	1	0	0	1	0	0	0	0	
49	8	0	2	3	0	0	1	0	0	0	0	0	0	0	0	2	0	0	
50	7	2	0	0	1	0	0	0	1	1	0	0	0	1	0	0	0	1	
51	8	0	2	1	1	0	0	1	0	0	0	0	0	0	0	0	3	0	
52	8	1	0	2	0	1	0	0	0	0	1	0	0	2	0	0	0	1	
53	10	0	0	1	2	0	0	0	1	2	0	0	0	1	1	0	0	1	
54	9	0	2	2	1	0	0	1	1	2	0	0	0	0	0	0	0	0	
55	8	0	1	3	1	0	0	0	0	2	0	0	0	0	0	0	0	1	
56	9	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	6	
57	10	0	1	2	1	0	0	0	0	2	0	0	0	4	0	0	0	0	
58	11	0	2	1	1	0	0	0	0	1	0	0	0	3	2	0	0	1	
59	8	1	0	1	0	0	0	0	0	0	0	0	0	2	0	0	0	2	
60	9	1	0	1	1	1	1	0	0	0	0	0	0	4	0	0	0	0	
61	8	0	0	0	0	0	1	1	1	0	0	0	0	4	0	1	0	0	
62	8	0	1	1	0	2	1	0	0	2	0	0	0	1	0	0	0	0	
63	9	0	0	2	0	1	1	0	0	1	0	0	0	1	1	1	0	1	
64	7	0	0	2	1	1	0	0	0	1	0	0	1	0	0	0	0	1	
65	7	0	0	2	1	0	0	0	0	1	0	0	0	3	0	0	0	0	
66	9	1	1	2	1	1	0	0	0	0	2	0	0	1	0	0	0	0	
67	10	0	1	1	1	2	0	0	0	0	0	0	0	1	0	0	3	0	
68	7	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	2	0	
69	11	0	1	1	1	1	0	2	0	2	0	0	1	1	1	0	0	0	
70	8	0	0	1	0	1	0	1	0	1	0	0	0	3	0	0	0	1	
71	7	0	0	1	1	0	0	1	1	1	0	0	0	2	0	0	0	0	
72	7	0	0	1	2	1	0	0	0	1	0	0	0	1	0	0	1	0	
73	10	0	0	0	2	3	1	0	1	0	0	0	0	2	0	0	0	1	
74	13	0	1	1	2	1	0	0	1	0	2	0	0	3	2	0	0	0	
75	9	0	0	1	3	0	1	1	1	1	0	0	0	1	0	0	0	0	
76	10	0	2	0	4	2	0	0	1	0	0	0	0	0	1	0	0	0	
77	7	0	0	0	4	3	0	0	0	0	0	0	0	0	0	0	0	0	
78	12	0	1	1	4	1	0	0	0	1	1	0	0	3	0	0	0	0	
79	9	1	0	0	3	2	1	1	0	0	1	0	0	0	0	0	0	0	
80	8	0	2	0	0	0	0	0	0	0	0	0	0	5	1	0	0	0	
	676	21	76	73	87	43	35	13	14	49	15	4	6	94	51	10	13	48	24



5. 社会学は皆無と言ってよい。散発的に存在するが、技術に即した分析のみで、いわゆる社会調査の研究成果を載せたものは、ないに等しい。また、民族学・民俗学・地域研究分野も少ない。

少し詳しく見ていく。まずは、文字研究の多さに驚かされる。やはり、人文科学研究、とりわけ漢字文化圏に属する社会を対象とした研究においては、文字の問題を避けて通ることができない長期的課題であったことを示しているであろう。また、文字の問題はテキスト解析の問題にもつながるという点も挙げられる。しかし、漢字文献ではない資料の文字解析も一定程度存在する。

また、**e-learning** も、この研究会の主要なテーマであったと言ってよい。特に人文科学そのものの継承において必須の教育に関わることも、主要なテーマであったのであろう。ただし、**e-learning** が、一定の「評価」を得たためか、近年は減る傾向にある。

芸術という分類は、いささか乱暴にも感じたが、いわゆる「人文科学」分野において、芸術系を美術・舞踊・音楽などと持っている研究組織は多くないとの判断が働いた。そのため、あえて芸術とまとめた。しかし、結果的には、その分野研究の多さに意外と驚かされる結果となった。多いのは、絵画研究であり、モーションキャプチャ、音楽研究と続く。

私が個人的に注視すべきであったのが、歴史学分野の移動である。特に、初期の研究が決して多いとはいえない。古文書読解ソフトを中心とする研究が、多くの回数を占め、それ以外は、数件程度しかない。近年、増加傾向を見せているが、これは「アーカイブズ学」をなかば強引に日本史の中に入れた要因もある。これは歴史学の方法論の側に問題があるようにも思われ、今後の私の課題となりうるところであろう。東洋史研究は、10件しかない。しかし、10件のうち7件が、CH71~76に集中的に行われているので、今後発展の余地はあるであろう。

また、同様に社会学分野も少ない。社会学では、関連の研究会が存在することも、その大きな原因であり、かつ、統計処理などの手法は、社会学内部で古典的に用いられてきたことが、主たる原因であろう。しかし、彼らの調査手法、フィードバックのありようなど、社会学者の方法論から得るところは、特にこの人文科学とコンピュータ研究に関して多いように見られる。この分野の研究者に対し、今後研究会がどのようなアプローチをする（しない）のかは重要な課題である。もう少し、詳しい分野ごとに見ていきたい。以下、思いついた順ではあるが指摘していく。

1. 考古学は、遺跡研究はコンスタントに多いが、遺物研究が減っている。

これは、遺物の計測方法などが確立してきたことなどが主たる原因であろう。また、情報考古学会や、動態計測研究会など、他にも研究発表の場が増えてきていることも大きい。遺跡研究は、小澤一雅氏の精力的な研究が全体を押し上げていることも付け加えておく。ほかに、一部の研究者（グループ）による積極的な研究発表によって、件数が増えているものは多い。

2. 文学は、世界各地を対象とした研究が多いが、英語文献を対象としたものは少ない。



また、CH1~20 と、CH65~80 において、報告全体に占める割合を検討した。紙面の関係上、表を載せることはできないが、ご了承願いたい。CH1~20 は報告件数 135 件、CH65~80 は 144 件である。増加率が顕著なのは、文字 (3%→12.5%)、芸術 (8.9%→18.8%) である。一方、減少が顕著なのは、特集『博物館』を組んだ博物館学、小特集『アーカイブズとデジタル技術の未来を考える』を組んだ日本史学を除くと、教育 (8.1%→3.5%) がもっとも目立つ結果となった。

## 2. じんもんこんシンポとの関わり

一方、じんもんこんシンポジウムは通常の研究会とはまったく異なる様相を見せる。特に、希少であった分野が多く見られるようになり、かつ研究分野の偏りもそれほど大きく見られないという特徴をみる (詳細は研究発表時の参考資料を参照のこと)。このことは、じんもんこんシンポが、この分野の「大会」としての位置づけを持っていることと同時に、査読のある研究会であることが大きな影響を与えていることは想像に難くない。

この両者の関係をいかに理解すべきであろうか。潜在的な層が多いとみるべきか。今後の課題であるといえる。

## 3. 課題

以上のことを見ると、いくつかの課題が見えてくるように思える。特に、いくつかの分野はまったく未開拓に近いと言ってよく、それらの研究をどのように進めていくべきか、今後の課題が見えてくる。さらに言い換えれば、研究会にそれらの研究と、情報学のコラボレーションを担える人材がいまだに現れていないといえる。本研究会で、必ずしもやらなければならないか否かは、検討の余地があるにせよ、「人文科学」であることには間違いない。

また、分類が「難しい」研究が多いことが、そもそもの課題であるといえる。技術的研究も無理やりに分類をしている、私の手法の問題によるところが多であることは間違いない。しかし、既存の人文科学の手法のいかなる部分を担うのか、もしくは、既存の研究手法から、どの方向性で逸脱を試みるのか、この点を前提にした研究が減っているのではないだろうか。たとえば、一つの資料を扱った研究でも、その手法の先にあるのは、人文科学のどの系統の研究と接続を試みるのか、わかりにくいものが多かった。

初期のころは、技術と明確な問題解決を目指したものが多く。それは、「人文科学の問題解決のためのそもそものインフラを」という意識が強いものが多いように見られる。しかし、近年は、分類は困難であるが、インフラ整備であるともいえない、というものが多い。これは、人文科学側の「細分化」を批判しつつも、実は技術の応用手法を行っている側の研究も細分化しつつあるのかもしれないという危惧を持つ。

既存の方法論からの逸脱は、無論、今後の研究にとって重要である。研究会の中にもそれを意識的に狙った研究が多いことは間違いない。しかし、過去の研究史のどの部分を踏まえているのか、どの部分が課題であるから、そこから逸脱しようとするのか、そこがあいまいである研究が多い。人文科学研究で用いる資料をコンピュータに適用することと、人文科学研究に情報科学を応用することとの間には懸隔があり、その差こそが、今の困難な状況を生み出しているように感じられる。

以上、乱暴な分類と、乱暴な集計、そして乱暴な仮説を立てた。あくまでも、議論のための、スタートラインとして考えていただければ幸いである。

### 参考文献

『今日からはじまるじんもんこん』(2007年1月)ほか、SIG-CH 研究報告 1~81 を随時参照。

## 「人文科学」と「コンピュータ」における多様な界面と標準化

永崎 研宣

人文科学とは人類の文化の諸相を対象とする様々な学問の総称であり、人文科学と名づけ得る確固とした共通の方法論が存在するというわけではない、ということについてはある程度の賛同を得ることができるだろう。人間の世界理解の仕方の広がりやを反映するがごとくに、人文科学には様々な学問領域・方法論が存在し、そしてまた常に方法論の革新が行われている。それゆえ、「人文科学とコンピュータ」というテーマは人文科学の多様性に応じて様々な界面を提示することになる。「人文科学とコンピュータ」研究会の歴史をひもといてみたとき、そのことは自ずと明らかであり、さらに、この研究会だけでなく、日本情報考古学会、計量国語学会等においてもそれぞれの学問分野に特化する形でコンピュータを利用した研究が行われており、英語コーパス学会、アート・ドキュメンテーション学会、日本アーカイブズ学会等でもそういった研究発表が行われている。さらに近年はそれぞれの学問分野における基幹的な学会においてもそのような研究発表が散見されるようになってきている。また、情報文化学会、情報知識学会、情報メディア学会等といった、より広い枠組みでデジタル時代の情報のあり方を問い直そうとする学会も活動を行っている。その一方で、さらに特定のテーマを扱おうとする研究会や協議会等も一部には活発に活動を展開している。「人文科学」と「コンピュータ」における界面の多様性とは、まさにこの状況そのものに端的に現れていると言えるだろう。すでによく知られているように、声から文字、書写から印刷、輪転機による高速な活版印刷といった新しいメディアの登場とそれによる情報伝達から意識の枠組みに至るまでの様々な面での変容は、コンピュータによってもたらされたデジタル情報という新しいメディアにおいてもまた、新たな形で極めて多様な側面において進行しつつある。「人文科学」と「コンピュータ」における界面には、そうした変容における諸相さえも含まれているのである。

コンピュータの登場によるデジタル情報の流通がもたらした人間・社会への様々な面での影響については、すでに多くの議論が行われており枚挙に暇がない。しかしながらそういった議論の主戦場はおそらく情報処理学会・人文科学とコンピュータ研究会とは別のところにある。したがってここでは、人文科学の研究手法においてコンピュータを利用することでそれに何らかの効率性や新発見、あるいはその可能性をもたらすという面に主な焦点をあてることにする。

人文科学の研究手法においてコンピュータを利用するという試みはコンピュータの歴史のごく初期の段階にその淵源をたどることができる。そしてその流れは 1962 年には ACL(The Association for Computational Linguistics)の前身である「AMTCL」の設立につながり、機械翻訳とコンピュータ言語学の研究のための場が形成されることになった。組版や印刷を目的としたコンピュータの利用が始まったのもこの頃である。さらに、1973 年には、言語や文献資料の研究におけるコンピュータ利用を研究するために ALLC (The Association for Literary and Linguistic Computing)が設立されており、人文科学全般を対象としたコンピュータ支援研究のため、1978 年には ACH (Association for Computers in the Humanities)が設立された。そして、1989 年には、我が国においてもコンピュータ民族学や考古学データベースをはじめとする様々な先行する動きを集約・発展させる形で 1 情報処理学会に人文科学とコンピュータ研究会が設立され、その後さらに様々な動きへとつながっていく。一方、上述の ACH, ALLC, ACL は共同で人文科学のためのテキストエンコーディングのガイドラインを作成するための支援を開始し、1990 年にはその最初のドラフト TEI P1 が公開された。これは現在では TEI P5 となり、さらに改良と議論が進められている。また、ACH と ALLC は 2002 年から ADHO (The Alliance of Digital Humanities Organizations)設立への取り

組みを開始し、1989年より共同開催している国際シンポジウムには2006年パリ開催時より「DIGITAL HUMANITIES」の名を冠するようになり、人文科学へのコンピュータの利用に関する取り組みをさらに広範に展開している。

DIGITAL HUMANITIES のシンポジウムにおいて顕著なのは標準化への取り組みである。これは必ずしも標準化に従うというだけではなく、標準化における問題点を指摘しその改善策を提案するということも含んでいる。標準化への志向は、人文科学におけるデータの交換可能性や持続可能性を確保するために不可欠なものであり、同時に、ユーザに対してはデータの利用を容易にさせるだけでなく、共通化されたツールを開発しやすくすることで、より利用しやすい状況を作り出すといったことも視野に入れた上でのことである。これらはいずれも標準化に関して一般的に理解されているメリットであるに過ぎないが、それが意識的・明示的に共同作業として進められているのである。ダブリン・コア等のメタデータをはじめ、すでに様々な標準が存在しており、我が国においても特に博物館・美術館や図書館を中心にかなり広まっているが、メタデータが付与されている個々の資料の内容までも標準に従わせようとする動きはまだ稀である。後者については、そのための標準として作成されつつある上述の TEI が有効だが、我が国ではまだあまり普及していない。理由としては、TEI について解説しているドキュメントがほとんど英語であることや、そもそも欧米の文献資料を念頭において作られている標準であるために我が国でよく扱われるような典型的な東洋的・日本的文献資料に適用するには多少なりとも問題があるといったことが考えられる。しかしながら、我が国においても、データの交換や持続の可能性を確保するためには、何らかの形で、たとえば TEI のような標準にコミットしていくような形で資料の内容の記述に関しても標準的な手法を採用していくようにすることが必要であり、それによって、最終的には全体のコストを下げつつよりよい成果が公開されるようになり、結果として、人文科学の世界がより豊かなものになっていくことが期待できるだろう。そしてそれを実現するためには、資料の内容も含めた標準化についてきちんと議論するための場が今後は必要となってくるだろう。

#### 参考文献

Rehm, G. and Witt, A.. Aspects of Sustainability in Digital Humanities, *Digital Humanities 2008*, June 2008, pp. 21-29. (2008).

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