

日本学術会議 課題別委員会
オープンサイエンスの取組に関する検討委員会
(第23期・第4回)
議事要旨

1. 日 時：平成27年7月21日(火) 13:00~15:00
2. 場 所：日本学術会議 5階 5-A(1) 会議室
3. 出席状況
出席者：土井委員長、大杉幹事、吉川委員、岡委員、喜連川委員(スカイプ)、三成委員(6名)
欠席者：杉田副委員長、戸田山委員(2名)
参考人：北村行伸 一橋大学経済学研究所教授
青柳正規 文化庁長官
オブザーバー：渡邊堯 名古屋大学太陽地球環境研究所・客員教授、情報通信研究機構・招聘専門員、真子博 内閣府政策統括官(科学技術・イノベーション担当)付参事官補佐(国際統括)
事務局：盛田参事官、松宮補佐、大西専門職、鈴木専門職付
4. 配布資料：
資料1 第3回議事要旨(案)
資料2 社会科学におけるオープンサイエンスへの取組：社会科学のデータアーカイブの構築
参考1 委員名簿
机上配布資料 人文学とオープンサイエンス

5. 議 事：

(1) 前回議事要旨案の確認

資料1に基づいて、前回議事要旨(案)が確認され、了承された。

(2) 先例紹介：社会科学におけるオープンサイエンスへの取組：社会科学のデータアーカイブの構築

北村行伸 一橋大学経済学研究所教授より、資料2に基づいて、「社会科学におけるオープンサイエンスへの取組：社会科学のデータアーカイブの構築」について説明が行われた。

【社会科学におけるデータ】

社会科学におけるデータとは、統計調査に基づくもの、アンケート、意識調査などの結果、社会実験などに基づく結果、政策立案に関連した資料、文献、国・地方の歴史資料、公的・私的組織によって残された実務資料など様々なレベル、様々な内容の資料を総称したものである。

自然科学では、研究者の研究目的に沿って、計画され実施された実験結果やその実験過程で収集されたデータがオープンサイエンスの主要な対象となるが、社会科学では、地方自治体も含めた政府が最大のデータ作成者であり、調査者である。公的統計データ、行政目的で収集している行政記録(税、社会保障、年金、医療、介護、教育)政府の行政サービスの一環として行っている業務データ(例：失業保険制度におけるハローワークでの業務)など、最大のデータ作成者は政府である。その意味で、オープンサイエンスは、オープン・ガバメントの側面を有していると言える。

また、インターネットで集められているデータが膨大な物になってきており、そうした民間で集められているデータについても考える必要がある。

【大学研究者、学会による調査】

大学研究者や学会等が収集しているデータとしては以下のようなものがある。なお、こうしたデータは、予算の継続的な確保という課題を抱えている。

- ・社会階層と社会移動全国調査(SSM調査 The national survey of Social Stratification and

social Mobility)

- ・慶應義塾大学パネルデータ設計・解析センター (Panel Data Research Center at Keio University)
- ・大阪大学社会経済研究所附属行動経済学研究センター (Research Center for Behavioral Economics)

【大学が調査資料のアーカイブとして機能している例】

統計資料・データをアーカイブとして集めている例として、以下のものがある。

- ・東京大学社会科学研究所のSSJDA (Social Science Japan Data Archive) : 民間研究機関、個人研究者、地方自治体、政府独自調査など多くの調査データ (数百件) がアーカイブ化
- ・一橋大学経済研究所附属社会科学統計情報研究センター (Research Centre for Information and Statistics of Social Science) : 明治維新以後の歴史統計を収集公開し、また、政府統計の報告書、作成に関わる関連資料などを収集。また、政府統計の2次利用の促進の一環として政府匿名マイクロデータの提供などを積極的に推進。

【社会科学におけるデータの意義再考】

社会科学の研究対象となるのは、社会を構成する国民 (家族、家計) や政府・企業・NGO などの組織、それを運営するためのルールとしての法や政治、市場などである。

特定の政策効果を測定したい場合、条件を管理した社会 (フィールド) 実験などを通してその効果を測定することがあり、その場合は自然科学の実験と同じようなデータを収集することになるが、日常的な社会経済活動を観察するためには、標本理論に基づいて、無作為に抽出したサンプルに調査を行うことで、社会全体 (母集団) の活動を把握するという考え方に従っている。

しかしながら、現在の統計調査では、この無作為抽出サンプルを確保し母集団情報を把握することがますます難しくなっている (調査回答拒否、公的機関の情報漏洩による調査への不信、生活スタイルの多様化による若年単身世帯の把握の困難化)。

代替的に、インターネットでモニターとして登録したサンプルに対してオンライン調査を行ったり、企業の顧客情報やIoT (Internet of Things) を通して集められたビッグデータなどを利用して、社会活動を捉えようという動きもある。

研究者の間で議論になっているのは、これらの代替的情報は母集団情報をどれくらい反映しているのか、セレクション・バイアスはどれくらいあるのかといった実証的課題である。

とはいえ、同時に圧倒的な量のデータがインターネットを通して集められている現実を無視する訳にはいかない (歴史上蓄積されてきた総情報量を上回る情報が単年度で集められるようになった)。Google や Facebook 等で開発されているビッグデータ分析手法について社会科学の研究者が学ぶべきものは多いし、彼らの活動を営利活動と捉え、オープンサイエンスの射程には入らないと考えるのは間違っている。むしろ官民学の協力体制を構築することが必要になっている。

ビッグデータを分析する手法として用いられている人工知能 (machine learning あるいは、deep learning と呼ばれる) の手法・ソフト開発は、人類全体にとっての数々の課題解決に結びつくものであり、オープンサイエンスの中心的課題であるといってもいいだろう。

政府等の収集するデータのみでは十分ではなく、今後一層、民間データを協力的に活用してることが重要となる。実際に、マーケティング会社が日々の物価等をビッグデータとして大学に提供し、これを大学の研究者が統計的に分析してさらに情報提供するといった取組も実際に行われている。とはいえ、民間データや分析手法については、秘匿性が高く不明瞭な部分があることや、データ提供の継続性も確実ではない等の課題があり、これらの点についても、社会科学として関与・検討していく余地がある。

【公的統計の公開】

公的統計をどのように公開していくかということも非常に重要な課題である。

公的統計はその質量ともに、民間統計、研究者コミュニティの生み出す統計調査を遥かにしのいでいる。総務省統計局、統計センターには、多数の統計専門家が所属している。そこでは統計の品質管理に関しても細心の注意が払われている。

公的統計には統計法という法律上の縛りがあり、一定の期間後、調査票原表を廃棄処分するこ

とが行われている。もちろん、データベース化されたデジタル情報は保存されているが、各省庁にとって、一定期間を過ぎた過去のデータを大切に保存しておくインセンティブは低い。将来の世代の研究に役に立つという意識は希薄である。

統計行政の指令塔機能を持たされた統計委員会の下で、公的統計の2次利用の諸施策については議論され、実際、利用が拡大してきている。過去の公的統計の維持管理を目的としたデータアーカイブについても議論はされているが、具体的な動きにはつながっていない。

できれば、政治的なリーダーシップの下に、各省庁縦断的なアーカイブ組織を作るか、国立国会図書館、国立公文書館等にデジタルデータアーカイブ部門を作るか、あるいは国立大学法人の中から、全国共同利用共同研究拠点に認定されている機関にデータアーカイブの運営を委託するなどの方法が考えられる。

【社会科学のデータアーカイブの必要性】

これまで蓄積されてきた社会科学のデータは、当初の収集目的とは異なる目的に利用できる可能性がある。この様な統計の目的外利用を認めることで、新しい知見を得ることが可能になる。とりわけ、コンピュータの情報処理能力が格段に進歩し、情報工学、コンピュータサイエンスの急速な進歩により、過去には考えられなかったようなデータの利用方法が可能になってきた。

そのために、社会科学の諸データをどこか一か所のデータアーカイブに集めて、専門家（データ・アーキビスト、データ・キュレーター、データ・サイエンティスト）によって維持・管理・公開される仕組みを作ることで、特定の研究テーマに関して利用可能なデータを比較検討しながら研究計画を具体化していくことが可能になる。過去の利用実績に関しての情報蓄積や情報共有も必要である。

現状では、専門家を育成しても、彼らのポストがない、安定していないという問題がある。ポストに維持に必要な資金は課題である。

これらの人材を確保するためには、データベース形成やデータベースの公開に対して、学術的な評価を与えるインセンティブが必要である。また、データベースの維持・管理・公開にはコストがかかることも事実であり、それに対する財源は広く利用者コミュニティによって分担される必要がある。

【社会科学におけるオープンサイエンス：まとめ】

子供の頃、「アラビアン・ナイト」の中の『アリババと40人の盗賊』の話で、『開け、ゴマ!』（open sesame!）という呪文を唱えると財宝の蓄えられた洞窟の入り口が開くシーンに心踊らされたものだが、今はopen data!（『開け、データ!』）、open science!（『開け、科学!』）と唱えて、情報の宝の山の入り口を開こうということのようである。

実際には、残念ながら、社会科学の資料室のイメージは、黄金に輝く財宝の山ではなく、かび臭い資料と干乾びたデータテープと段ボールの山といったものだろう。誰が見ても宝であると思えるような資料はほとんどない。社会科学の素養のない盗賊にはゴミの山にしか見ない。

すなわち、アリババ達との違いは、情報の山の中には宝は含まれているが、大半はゴミであって、その中から宝を探し出すには、それなりの知識と技能が必要であり、それを行うことができる専門家がいなければ、宝は見つからないということである。

社会科学のデータアーカイブを宝の山とするのは、データ分析の専門家であり、そのデータ分析から生み出されるエビデンスベースの素晴らしい政策提言である。我々に必要な呪文は Open Data! Open Science! And Work Hard!! だろう。

説明に続いて、意見交換が行われた。主な発言は次の通り。

【データ・サイエンティストの評価、ポスト】

・今回、「データ・サイエンティストとしての就職先がない」という話があったが、これまで他の分野に対して行ったヒアリングでは、「データ・サイエンティストが不足している」という話があった。統計ができる人材ならば、就職ができるのではないか。

→データ・サイエンティストが不足しているという面があることは同じであるが、データ・サイエンティストとしての能力があったとしても就職先に困るため、育成にも課題があるというの

が現状である。もちろんすごい能力がある人は、就職先があるだろうが、それは一部である。社会科学のデータ分析の技術水準を高めて、国際的な競争力を持つような人材を育てないとならない。そのためには裾野を広げないとならないが、裾野を広げすぎると中途半端な状況の人が出てきて、就職問題が深刻になってしまう。

→Google 等に就職できるような人と、就職できない人（中途半端な人）の違いは何か。価値基準は何か。

→おそらく人工知能の研究をやっているような人材は民間で引く手数多なのであろう。しかし、社会科学ではデータを扱っているレベルは必ずしも最先端なテクノロジーではなく就職は難しい。また数量経済史のような研究テーマが領域にまたがるような場合では、史学分野、経済学分野、いずれの分野でもアカデミックな就職は難しく行き場がない。

- 例えば、史学等の分野では、文章そのものを読むことが大変なテクニックで、データの収集そのものが非常に困難であるが、それをきちんと評価しポストにつながるような状況にはなっていないと思う。

【分野融合研究、学際的研究の必要性】

- 感染症の伝播状況等を地域間の違いを見ると、そこには自然科学の統計や微生物学等では完結し得ない要因、例えば、政治・経済・宗教・文化などが寄与しているが、実際には、そうした社会科学的な視点を加味した分析は行われにくい。しかし、そこを解析しないと、有効な対策にまでは結びつかない。領域を超えたサイエンスが行われにくいという面がある。そういう意味で、社会科学データのアーカイブスは、自然科学の分野にとっても有用であり、領域を超えたオープンサイエンス、コラボレーションができれば、プラクティカルな意味でのサイエンスとして再利用できるように思う。

- 法律などの制度を設計した際、例えば食品安全基本法等においてもその法律を作り色々なことを義務づけた時に、それを支える人等（ポスト・受け皿、支える人を育てる人を含む）が不足するという問題がある。ニーズがあってそのための人材を育てたところで、ポスト・受け皿がなく、建前通りにならないということが実際にはあるように思う。

→文部科学省の情報科学委員会で、データ・サイエンティストに関する議論を行っており、現状では人数が少ないという話をしている。人を育て、必要なところに育てた人がいけるような仕組みについても考えないとならない。

- 学際的・分野横断に関する話として次のようなこともある。厚生労働省は「21世紀出生児縦断調査」という21世紀の初年に出生した子の実態及び経年変化の状況を継続的に観察する調査を行っている。対象の子供は、今では15歳にまで成長している。最初は、身長・体重、子育て等についての調査をしていたが、だんだん学校教育に関わることを調査するようになってきた。この統計は厚生労働省の予算で行っているのであって、誰がどのように責任を持ってこの調査を追いかけるのか、上手く整理をする必要がある。本当は、社会科学のデータなので、各省庁が集まってオールジャパンで行うことが良いように思うがそういう風にはなっていない。また、Googleのような企業が行っていることは、ビジネス・利益追求であり社会科学的な側面があるが、そこで働く先端的な仕事をしているのは自然科学のサイエンティストが多い。つまり、利益追求という社会の仕組みを元にしながらテクノロジーを駆使するという文理融合の課題が存在している。

- 今後、社会分析を行う際に、研究サイドとしてどのようにビックデータと向き合っていくのか。例えば、Google等の企業の膨大なデータ・情報収集力と緊密な関係を築いていくのか、人工知能による翻訳等の新しい解析システムの導入をどうするのか、などあるか。

→ビックデータの勢いがものすごく大きい。例えばGoogleは日々の物価指数を提供することを目指しているが、一方で政府統計の物価指数は月に1度が限度だろう。Googleの方は、現状ではブラックボックスが大きくすべてについて合意できるものではない。重要なのはGoogle等の民間のビッグデータと伝統的な政府統計の違いを認識し差を説明できることだと思われる。なお、現状ではGoogleの物価指数等は、母集団が不明瞭なままインターネットで売り買いされている

財をカウントしており統計的には意味を持たないという面があるが、ただ将来的には、世界中の人の購買に近いデータを収集でき、速報性もあり正確なデータ提供できる可能性はある。現在では、まだ人工知能・ロボットにデータを読ませて解析するということが可能な段階には至っておらず、人間が介在せざるを得ない。人間が地味にデータを集めて整理する部分が必要であると思われるが、そうした行動については評価が高いとは言えないのが問題である。そうした地味な研究をしている研究者は、小さなコミュニティの中では生きられても、日本全体、世界全体で見た時にマイノリティになって評価がされない。

【公的データの2次利用】

- ・公的データの公開に当たり、プライバシーの問題についてはどのように対応しているのか。
→統計法に基づいて運営されており、基本的には統計を集める際の当初の目的に基づいてデータは利用されるが、2次利用として大学研究者等が利用する場合にも個人情報秘匿した上で提供される。
- ・イギリスの古いセンサスデータを今の人活用できるという例があったが、後世の人が、例えば100年後に活用するためには、そのための仕組み（個人情報の取り扱いも含めて）について色々考えておかないと、仮に100年取っておいても結局利用できないということになりかねないのではないか。
→例えば、戦前に集めたデータについては、戦前の法律に基づくので、結果的に今利用しやすいということがある。ただし戸籍については公表できない。指摘の通りプライバシーの部分も含めて、公的データを後世の人が利用できる権利を確保しておくことを考えることも大切である。
- ・社会科学系の場合、デジタル保存のみならずモノ保存が必要な部分があり、その場合場所が必要になるが、何か考えがあるか。
→書籍も含めた分散管理をして、ある程度消失等に備えておく必要があるが、コンピュータで分析することが主流となっており、そもそもデジタル化していないと使い物にならなくなっている。

(3) 先例紹介：人文学とオープンサイエンス

青柳正規 文化庁長官より、机上配布資料に基づいて、「人文学とオープンサイエンス」について説明が行われた。

【サイエンスの分類とトレンド】

人文学も含めてサイエンスを考えると次の3つのように分類することができる。

- ・curiosity drive：広義の認知科学 科学（理学、知りたいということが原動力となっている）
- ・genealogical restriction：拘束された知 学問（人文学、一種の系統学でこれまでどのようなことが行われてきているか、その積み重ねの上に構築する）
- ・mission oriented：設計科学 科学技術（工学、経済学）

系統学・系譜学的な性格が強い人文学について、今後どういうあり方で進んでいるかについて様々なところで議論が行われている。例えば、平成18年度文科省委託研究「今後の人文・社会科学系分野教育の在り方に関する調査研究」では、次の3点が指摘されている。

- ・継続性が重要で短期間のプロジェクト型研究になじまない
- ・「多文化の共存」が人類の緊急課題であり、人文学の需要は増大しているはず
- ・社会の関心がかわった → 実益主義、自由市場主義・・・・

近年では、社会の関心実益主義、自由市場主義が広がってきたため、サイエンスの3つの分類のうち、mission oriented以外のもの（curiosity drive、genealogical restriction）が無視されつつあると言える。

【構造主義と人文学：人文学の分水嶺】

過去50年間の大きなインパクトとしては、1960年代以降のフランスで隆盛となった構造主義

の広がりがあり、これは、人文学を超えて他の分野にも大きな影響を与えることとなった。しかしわが国においては、フランス等のラテン系ほど構造主義の影響は大きくなかった。具体的には、次の通り。

- ・人文学の最近 50 年の大きなインパクト：理論、方法論は構造主義（1960 年代～）
- ・言語学の構造主義を、レヴィ＝ストロースはその理論を文化人類学に採用。その有効性を証明
- ・その結果、構造主義は、人文学の他の分野にも影響を与える
- ・わが国でも文化人類学の中核的拠点として国立民族博物館が 77 年に創設
- ・とはいえわが国では、他分野への構造主義影響はフランス等のラテン系ほどに浸透することはなく、アングロ・ゲルマン系の状況と類似していた。構造主義の短所長所をきちんと理解して取り入れるということはあまり行われなかった。
- ・それよりもむしろ、ベトナム戦争や大学闘争の影響が大きい

【人文学におけるデータとニーダムの意義】

Joseph T. M. Needham “History of science and technology in China”（1954～）が書かれたあたりから、人文学でも大規模なデータが重要視されるようになってくる。

そもそも、データを扱うサイエンスの分野としては、人文学が最も古い分野と考えられる。古くはアレクサンドリアで図書館ができたことから始まり、紀元前 6 世紀頃からギリシャでは情報の生まれた年代と場所をデータに付与することが行われるようになっており、これが現在において書誌情報における出版地と出版年を記載することにつながっている。この他にも辞書や百科事典が生み出されてきたというのも、データを扱うサイエンスの先駆けと言える。

Needham が行ったのは、データを単に量のみとして示すのみならず解釈も示すことであり、卓越した “science manager” であると言える。ここに人文学のデータの扱い方の意味があると言える。

【機能主義的権威主義の崩壊とアナール派】

1960 年代以降権威主義が崩壊していく中で、フランス国立社会科学高等研究学院を中心としたアナール派が、身近ではあるがそれまで見過ごされ埋もれていたデータの採集とそれを生かした成果をあげるようになっていった。

そしてデータ量が増えるにつれて、共時的研究（シンクロニック・スタディ）が重要となっていった。それまでは、データ量が少ないので、ある分野・項目について違う時代のデータを持ってきてそこにエボリューション・変化などを見ていくことが歴史研究の中心であった。しかし、データ蓄積が増大し、コンピュータが進歩するようになると、同時代に様々な分野でどのようなことが起きているか実相を明らかにした上で、さらに時代を積み重ね重層的に変化を見るといった立体的にエボリューション・変化を見ることが可能になった。

- ・共時的研究 synchronic study
経時性を本質的に含む「記述」「言説」を基本とする文字資料を中心とした研究
- ・経時的研究 diachronic study
歴史的個体は継時的であると同時に共時的空間を持つ。この共時的空間の復元と構造の解明空間性を有するが故に本質的に共時的である象形資料がこの分野の中心的研究資料
ただし、静態的 static となるおそれがある

しかしながら、そうした立体的に実相を把握するためには、データの蓄積方法や、マンパワー等の大きな問題が生じるようになった。

【人文学の可能性】

台湾中央研究院、古典文献全文検索資料庫「漢籍電子文献」、正史「二十五史」という中国の歴史書（100 冊ぐらいの規模）が、デジタル化されデータベース化されている。1970 年以前においては中国史研究者は 30 年をかけてこの書物を読み、例えば、「塩」という字をピックアップして論文にまとめて成果としておりそれが評価された。しかし今では、10 秒程度でピックアップすることができるようになった。それによって学問の性格が変わってきたと言える。

韓国では「朝鮮実録」という李朝の書物を、英訳も含めてデジタル化することが行われており、自分達の祖先が書いたものをハングルが読めない人も含めて誰でもアクセスできるよう取組が進んでいる。

日本でも東大の SAT 大蔵経テキストデータベース研究会で「大正新脩大蔵経」のデジタル化が進められている。

こうした取組の際に、「フォントの問題」が生じる。すべてデジタル化するためには、まずフォントを作るところから始めなければならない、そのフォントは、韓国、台湾、中国本土と共通するものではないという、効率の悪さがある。

こうしたことから、実際にデジタル化している古文書等は、一部であって、色々とでこぼこが大きい状況と言える。

また、「維持費の問題」もある。せっかく大きなお金をかけてデータベース化してもそれが継続的な資金でない場合が多く、その場合は、サーバーに載せ続けることが困難となってしまう。各大学で地域の情報をデジタル化しているものも多々あり、死屍累々がそこそこにあるとも言える。そうしたデータが日本には本当に多く、一部は、デジタル文化財創出機構で凸版印刷のデータベースに載せるということが行われている。

【モノと文字記録の対応問題】

人文学では、データ保存の際に、モノと文字記録をどう対応させるかという問題がある。例えば、裁判の物証等についてはかなり確実な証拠資料であるが、それだけでは何も語るものではない。一方で、証言文書などの文献は自らを語っているが、バイアスがかかっている可能性もあり注意する必要がある。この物証と証言文書をどのように組み合わせていくかということが非常に重要である。しかし、物証を画像化するとデータ量が膨大になり大きな費用が必要となる。

【デジタルデータの保存・分類】

また、印刷物とデジタルデータでは、その保存分類方法が確立しているかどうかについて大きな差がある。印刷物は、保存分類方法が確立されている。実際には数百部が印刷されあちこちに分散されて分類されて保存されるので消失リスクも少ない。ところがデジタルデータでは、保存分類方法が確立されているとは言えない。またサーバーも1か所～2か所程度で、ちょっとしたことで焼失してしまうリスクがあると言える。

こうしたデジタルデータ化に関わる様々な問題（保存分類方法、消失リスク等）は様々あるが、まだそれらに正面から対応しているとは言えない。社会的にも議論をしていかないとならないが、単なるユーザーはまだ危機感を感じていない。

【古文書の学際的利用の意義】

変体仮名の自動読み取りプロジェクトが話題となっている。古文書はこれまでは、国文学の研究者が中心となって分析されてきた。しかし古文書の情報は、例えば、地震、津波、といった違う分野に役立つ情報も含まれている。

国文学、歴史学以外の他分野の人々が共通に使えるように、データが構築されていくと、具体的な課題解決につながる研究に寄与するだろう。

また、かつて日本では、フィールドサイエンスが力を持っていたが、だんだんシミュレーションの方にシフトしてしまった。そうすると、データ収集の際に、古文書のようなあまりバイアスのかかっていないものからデータを集めることが今まで以上に重要になってくる。

【経済効率と学術データ：地質学の例】

地質学の例であるが、かつては、石油会社が石油探索の際に地層から抜き取った「ボーリングコア」を地下数十キロにわたり残して、学術に提供しており、そのボーリングコアが地質学に大変寄与していた。しかし、現在では経済効率を優先するため、石油が出る付近以外のボーリングコアは残さなくなってしまうと、地質学では大変困った状況になっている。この様に、経済効率によって、データの材料を確保することが困難になっているところもある。

【人文学のオープンサイエンス】

オープンサイエンスは、学術の進捗に対して重要な方法であると考えられる。

これまで科学の進捗に従って、細分化され、サイエンスの全体的な調和を誰も見ることができなくなり、神頼みになっている。19世紀には全体論・ホーリズムがあって、様々な警告を与えて

いたが、還元論と比較した場合の厳密さの欠如から後ろに追いやられていった。この部分をオープンサイエンスでビッグデータにすることによって、定量的な物から定性的な物に転換できる可能性があるのではないかと考える。

これまで人類が作ってきた、辞書、事典、叢書、コルプス等の編纂・公開については、今までのように印刷物として残すと同時に、一方でオープンサイエンス化するためにデジタル化する必要がある。

アーカイブスは、日本においては国立公文書館があるものの、公文書館の構築は先進国の中では遅れてできている。先進国では公文書館等のアーキビスト達がデータを抑えていわば権力化していることを反面教師として、日本においては徹底的に国立公文書館の権限を抑制したものになっておりそれがまた別の問題を生んでいる。日本では国立公文書館は、独立行政法人から国立に戻そうという良い動きもあるので、それを実施した後に、より権限を広げていくことが必要だと思われる。

国立国会図書館では、4年ほど前より予算180億円程度を投入してデジタル化を進めているが、残念ながらその半分ぐらいは著作権者が不明でネット上に載せられないという問題が生じている。一方で韓国の中央図書館と協定を結び精度の高い翻訳ソフトを用いてお互いの国の文献を自国の言葉に翻訳できるような取組もある。こうした動きが一層進むことが期待される。

Googleのような民間も大変優れたソフトを作っているが、商業ベースであり、いつやめるかもわからない。この問題に対応するため、EUではヨーロッパナ (Europeana) という電子図書館ポータルサイトを立ち上げている。日本もこれに先駆ける形で、文化庁が「文化遺産オンライン」を立ち上げている。立ち上げ当初は非常に評価が高かったが、その後この事業は予算が減少されることによりなかなか進まない。

知の共和国を早く作らなければならないというのもあり、例えばポンペイにいけば、世界中のポンペイ研究者の情報（論文、異動他）が手に入る。こうした知の共和国のようなものが日本の人文学の中にどの程度あるかと言えば心もとない。

【国際学士院連合】

それぞれのアカデミーが提案して、学術編纂プロジェクトが行われ、これまで3000程度が公表されている。こうした継続性のある活動を蓄積していくことが、オープンサイエンスにつながっていくと思われる。

説明に続いて、意見交換が行われた。主な発言は次の通り。

【フィールドサイエンスの可能性、保存するデータとは】

・例えば材料の分野では、シミュレーションによる有望な組み合わせを元にして効率的な実験を行うようになっているようである。またデータ公開の前に知財を抑えるということも行っている。この様な程度経済的にも成り立つやり方でデータベースの構築等が行われている。また情報の分野では、IoTで様々なデータが取れるようになりその情報を活用することによって、ようやくフィールドサイエンスができるようになってきたということがある。

ある学問が隆盛な時には予算が付くが、しばらくすると別の学問が隆盛になり、以前ほど予算が付かなくなってしまうということがある。それによって今までコツコツと積み重ねていたことが継続できなくなるということが散見される。その対応の策として、国際学士院連合の取組が位置づけられるのだろうが、結局のところすべてのデータを残していくということとはできないと思われるが。

→すべてのデータを取っておくことは出来ないだろう。とはいえ、日本ぐらいの経済規模あるところにおいて人文系がどんどん弱くなっているということ、その例の一つとして、国語辞典・百科事典の編纂が困難になっていることがあげられる。また、例えば住友金属鉱山が鹿児島県の菱刈鉱山という金山を採鉱しているがこれは、海外の指導者を育成するために行っているということだ。それは一つの現場経験が色々な形で展開するという側面に着目したためであるが、この様にフィールドサイエンスも色々な可能性を持っている。指摘の通り、フィールドサイエンスも色々な方法があり、またある程度絞り込みが必要な面もあるだろうが、「フィールドサイエンス」の可能性を十分に認識した上で考える必要があると思う。また、日本学術振興会が

特別な奨学金を分類学に対してかなり長い間提供していたことがあった。しかしながら、それもいつの間にかなくなり、例えば東大の解剖学においてさえも分類学ではない方法に移っているように、ある学問分野の中で体系を守っていく動きは以前のようにはない。

【分野融合】

- ・自然科学でも他分野との融合の動きをやって行かないと成り立たない時代であるが、人文学ではどのような状況なのか。
- 海外の場合、戦後、地域を限定してその中で他分野が相互研究をしていくというやり方が行われた。その他デジタルヒューマニティの分野ではデジタル化した情報を元に新しい分野を開いていこうという動きがあった。しかし、日本ではそうしたことはあまり定着していったいない。あくまでもディシプリンを守るという姿勢で取り組まれてきた。今、文部科学省等から今後のあり方についての提案が示されつつあるが、これに対してアカデミックの側から声を大にして議論を行うことが必要な段階だと考えている。

【デジタル情報のリスク管理】

- ・デジタル情報のリスク管理の体制が整っていないという指摘があったが、それに対して具体的に管理していく方策などはあるのか。
- そうしたものの管理をしていく際に一番近い分野は、人文学では図書館学であるが、そこは本が中心である。モノをデジタル化してデータを蓄積するということに関しての協力は、なかなか得にくいのが現状である。結局自前でやらないとなくなる。ただ、今は以前と比べて機器類が安価に手に入る時代ではあるので工夫次第ではないか。

【強調すべき提言の種】

- ・学術会議で提言を取りまとめることを前提とした場合に、オープンサイエンスに関して特に強調すべき点があれば提示願いたい。
- オープンサイエンスに当たっては、産官学と連携していかないとならない。その際に学は、産官が行わないことを担うが、その資金源が安定していないと難しい。またそれを維持するための人材育成へも資金的配慮が必要である。それがないと上手くいかないと思われる。
- 大学を串刺しにするようなデータベースを集める委員会、残すデータを取捨選択するような組織を作り進めていく必要があると思われる。それを行うと著作権等の問題が出てくるだろうが、EU Direct のように研究者の善意をベースに訴訟になった際に考えるという方法を参考にしながら、著作権で思考停止しないようにすると良いのではないか。

(4) その他

【次回の開催について】

次回の委員会は9月以降に開催予定とした。次回委員会では、「オープンデータ、オープンサイエンスにかかる構築及び維持に関わるエコシステム」「『オープン』の意味」等の論点を分野の違いなどに着目しながらまとめて議論を行い、それを踏まえて今後の委員会の方針を相談する。加えて、本委員会の議論の公表方法についても相談することとした。

以上