

科学技術情報整備審議会基本方針検討部会（第2回） 議事要録

日 時：

令和元年 12 月 12 日（木）午前 10 時～正午

場 所：

東京本館総務課第一会議室

出席者：

基本方針検討部会員 4 名

竹内比呂也部会長、佐藤義則部会員、生貝直人部会員、北本朝展部会員

館側出席者 8 名

（幹 事）利用者サービス部長、電子情報部長

（事務局）利用者サービス部副部長、同部サービス企画課長、同部科学技術・経済課長、同課課長補佐、電子情報部副部長（電子情報企画課長事務取扱）、同課課長補佐

次 第：

1. 前回議事要録等の確認
2. データ駆動型社会からデータ神格化社会へ [北本部会員発表]
3. 知識基盤社会におけるデータ駆動型研究の動向
4. その他

配付資料：

（資料 1）科学技術情報整備審議会基本方針検討部会関係者名簿

（資料 2）基本方針検討部会のスケジュール

（資料 3）科学技術情報整備審議会基本方針検討部会（第 1 回）議事要録

（資料 4）基本方針検討部会の検討テーマ（案）について

（資料 5）データ駆動型社会からデータ神格化社会へ

（資料 6）部会（第 2 回）の論点について

（参考資料 1）学術情報流通の整備を巡る諸外国の状況調査（1）～（3）

（参考資料 2）科学技術・イノベーション創出の総合的な振興に向けた科学技術基本法等の在り方について（案）（総合科学技術・イノベーション会議基本計画専門調査会制度課題ワーキンググループ（第 5 回）資料 1-1（令和元年 11 月 20 日））

（参考資料 3）科学技術・学術審議会における第 6 期科学技術基本計画に係る検討状況

（参考資料 4）研究目的に係る権利制限規定の創設に当たっての検討について（案）（文化審議会著作権分科会法制・基本問題小委員会（第 3 回）資料 2（令和元年 10 月 30 日））

（参考資料 5）研究データに関する NII・JST の取組動向

（参考資料 6）国立国会図書館が提供するデータの種類

（参考資料 7）国立国会図書館が提供するデータ利活用の事例等

## 議事概要：

### 1. 前回議事要録等の確認

案のとおり、了承された。

### 2. データ駆動型社会からデータ神格化社会へ〔北本部会員発表〕

### 3. 知識基盤社会におけるデータ駆動型研究の動向

北本部会員が資料5を基に発表した。事務局が資料6について説明した。

主な発言は次のとおりである。

#### ○データ駆動型研究

- ・ 現状のAIによるくずし字認識は、くずし字を翻刻する（現代の文字に置き換える）ことまでであり、現代人が読解するためには古文を現代の言葉で理解できる必要がある。著作権保護期間中の資料に掲載されている古文と現代文の対訳コーパスが利用できれば、データ駆動型人文学のインパクトを高めることができるのではないかと。
- ・ データ偏重からデータの価値判断や倫理など人間側の問題に焦点が当たる中で、データ駆動型人文学を進めるだけでなく、AIの倫理の問題など人文学の知見を他の領域に広げることにも求められている。これら二つの方向性における基盤として、国立国会図書館（以下「NDL」という。）が過去の蓄積をいかしてどのような役割を果たせるのかは検討すべき課題だと思う。

#### ○デジタル化

- ・ 「デジタル化」という言葉が指す内容をはっきりさせた方がよい。現代におけるデジタル化では、画像化だけにとどまらず、テキスト化、構造化が当然に求められるのではないかと。学術雑誌では、全文XML化も広がりつつある。
- ・ テキスト化の研究やその実現に当たっては費用が問題となる。全ての資料をデジタル化、テキスト化することが望ましいが、分野や資料群等による優先順位付けも必要になるのではないかと。そのための検討には、デジタル化やメタデータの整備状況や方向性を明らかにしておく必要がある。今回の提言で方針を示し、NDLで具体化していくと考えている。
- ・ NDLの資料のデジタル化は、テキスト化したものを検索して必要な知識を得ることを考えると、ここ2、30年間の新しいものが重要となることを考慮すべきである。既にあるテキストデータも含め、他機関との連携・分担を図りつつ、どの時代のものまでデジタル化が必要かと併せて、利用の仕組みを考えないといけない。
- ・ HathiTrustもOCR技術自体は研究開発しておらず、Googleに頼っている。他方で、日本語資料のテキスト化の精度は十分ではなく、日本が主体的に取り組むべき課題でもある。画像データからのテキスト化を精緻化するためには、大規模かつ長期の取組が必要であり、NDL単独では難しい。オールジャパンの体制を考えないといけない。

#### ○デジタル資料の収集・整備

- ・ 現在も未来から見れば過去であることを考えれば、デジタル資料について、現在の収集が社会の方向性と一致しているか考えなければならない。法定納本図書館としての役割、また研究プロセスのデジタル化については、次々回の「データ利活用社会における基盤整備」（個別テーマ（3））で議論を深め、今回の提言で扱う必要がある。
- ・ ウェブアーカイブの担い手は日本ではNDLしかない。著作権の問題もあるが、民間のものも含めアーカイブしたデータをどのように利活用できるか、公開できるかを考えなければならない。

### ○研究データ

- ・ 研究データについては、保存の重要性を NDL が周知に努めたこと等もあり、この 5 年間で社会の理解が深まった。NII や JST の取組方針を尊重し、NDL は、両機関でカバーできないもの、例えば、引き取り手のいないデータの救済、地域資料など特定の研究分野に結びつかない横断的なデータに目を向けるとよい。ただし、NDL が全てを預かればよいものでもない。「ひなぎく」のように基盤となるネットワークをどう構築するかが重要となる。
- ・ 地域資料のネットワーク化を進めるとき、基礎となるのは時系列、地理的マッピング、人物である。このうち、人物には典拠データが役立つ。時系列や地理的マッピングには課題がある。地理的マッピングには過去の地名に関する情報が必要だが、我が国ではこれが欠落しており諸外国よりも遅れている。

### ○倫理など

- ・ 現代の倫理観からは到底認められない内容を含む資料、事件報道や個人の住所を含む資料など、デジタル化に当たって問題となり得るものもある。単に検索結果から除外すればよいのか、またオプトアウトの手続きなど、対応を議論する必要がある。

### ○国立国会図書館の役割

- ・ データ駆動型社会において、NDL の役割は広く人々をエンパワーメントすることにある。データの利用者を、一般市民、専門家・学生、分野専門家に分類すると、「専門家・学生」は自らツールを活用する意欲がある人たちである。これらにはデータを利活用できるツールを提供することが考えられる。「一般市民」は、例えば、気象観測データならば、天気予報など、データそのものではなく、直接的な情報を得ることを望む人たちである。例えば、データリテラシーの涵養など、一般市民を対象とした取組も何か考えられるのではないか。
- ・ シチズンサイエンスとの関わりについては、NDL が積極的に何かを支援するというものではなく、広いネットワークの中で考えればよい。
- ・ イノベーションの創出との関わりについては、知識インフラの基盤である図書館が直ちに貢献できるかといえば難しい。NDL が資料の収集、デジタル化やメタデータの整備を進めることで利活用できなかったものを利用可能にすることは、広くイノベーションへの貢献と言えるのではないか。