

新たな情報財検討委員会（第6回）

日 時：平成29年2月28日（火）14:00～16:30

場 所：中央合同庁舎4号館11階 共用第1特別会議室

出席者：

【委員】中村委員長、渡部委員長、飯田委員、今枝委員、上野委員、奥邨委員、喜連川委員、木全委員、瀬尾委員、関口委員、戸田委員、林委員、福井委員、森委員

【関係機関】内閣官房IT総合戦略室 山路内閣参事官  
総務省 技術政策課研究推進室 中川補佐  
経済産業省 知的財産室 後藤補佐  
特許庁 総務部 仁科企画調整官  
文部科学省 研修振興局 榎本参事官  
文化庁 著作権課 水田課長  
国会図書館 電子情報部 川鍋副部長  
情報通信研究機構 鳥澤センター長

【事務局】井内局長、永山次長、小野寺参事官、福田参事官、岸本参事官、大手参事官補佐、松村参事官補佐

1. 開会
2. 残された論点（AI学習用データ及びAI生成物等）について
3. 報告書骨子案について
4. 閉会

○中村委員長 時間がまいりましたので、ただいまから「新たな情報財検討委員会」第6回の会合を開催いたします。

お忙しいところをお集まりいただきまして、ありがとうございます。

今日は、2時間半をいただいています。

今日御出席をいただいている委員の方はお手元にある座席表のとおりですけれども、関係の機関としまして、内閣官房、総務省、文部科学省、文化庁、経済産業省、特許庁、国会図書館、情報通信研究機構、NICTから御出席いただきますけれども、国立国会図書館の川鍋副部長が14時45分ごろ、文化庁の水田課長が15時ごろ、おくれて到着されると伺っております。

早速、井内局長から挨拶をいただきたいと思います。

○井内局長 本日も、お忙しい中、お集まりいただきましてまことにありがとうございます。

前回まで、日本の競争力の強化に向けて、データの利活用の促進、AIの開発利用の促進の観点からいろいろ御議論いただきました。

前回と前々回のデータ取引につきまして、いろいろ利活用につきまして御議論いただいて、取引については、契約がベースとなるというのが基本的な認識として皆さんに共有していただいた上で、何らかのガイドライン的なものが必要なのではないかとか、不正行為の抑止手段のようなものが何かあったほうがいいのではないかとか、そういったことにつきまして、比較的方向性が収れんしてきたかなという感じもしておりますが、他方で、何らかの新たな権利を付与するかどうかということにつきましては、まだ議論に幅があるような感じもいたします。

これにつきましては、データの利活用、ビジネスの状況でございますとか、諸外国の動向を見ながら、引き続きもう少し議論が必要なのかなという感じかと個人的には感じております。

そうした中で、本日でございますけれども、残された論点はいろいろございますので、そこにつきまして、関係省庁あるいは関係機関から御報告をいただきまして、その上でさらに全体を通しての議論もいただいて、次回以降の取りまとめに向けまして、方向性の集約をできればお願いしたいと思っております。

本日は、2時間半の長丁場でございますけれども、よろしく願いいたします。

○中村委員長 議論に入るに当たりまして、事務局から配付資料の確認をお願いします。

○岸本参事官 それでは、お手元の議事次第をごらんいただきたいと思います。

配付資料のところでございますが、資料1から6までが事務局の資料になっております。

資料7-1から資料9までが、関係省庁、関係機関から御提出いただいた資料となっております。

このうち、情報通信研究機構から御提出いただいた資料7-2ですけれども、席上配付のみとなっております。

資料7-3、文部科学省から御提出いただいた資料ですけれども、資料番号なしで配付させていただいている、資料の一番後ろに置かせていただいておりますので、その点、御確認をお願いいたします。

何か不足等がございましたら、事務局までお申し出いただきたいと思います。

○中村委員長 よろしいでしょうか。

議論に入りたいと思いますけれども、今日はテーマが3つありまして、残っている論点が2つと、報告書の方向性です。

今日最初のパートでは、前回までに整理ができなかった論点でありますAI学習用データのうちのオープンデータ、オープンサイエンスなどについて議論をいただければと思います。それが1つ目。

2つ目が、AI生成物とAI学習用データのうちの著作権関連の部分について議論をいただければと思います。

最後のパートでは、この委員会の成果物となる報告書の構成と方向性を示しました報告書の骨子案を事務局でまとめてもらいましたので、そちらについて説明をしてもらいまして、また、事務局で欧州でのAI・データに関する現地調査に行ってもらいましたので、その報告もいただこうと思っています。その上で、これらについて皆さんに議論をしていただければと考えております。

それでは、AI学習用データのうちのオープンサイエンスなどの部分について議論に入りたいと思います。事務局で、これまでの会合でお出しいただいた意見を踏まえて、資料を準備してもらいました。また、関係機関、産総研、NICT、文科省からも資料を準備していただいております。

早速、事務局と関係機関から説明をいただきたいと思います。よろしくどうぞ。

○岸本参事官 それでは、お手元の資料2をごらんいただきたいと思います。

こちらは、これまでの会合におきましていただいた意見のうち、本日御議論いただきたい論点に関する部分をまとめたものでございます。パートを分けて議論していただく都合上、まずは4ページまでを御紹介させていただきたいと思います。

まず、AI学習用データ作成の促進に関する意見でございますけれども、データ作成者とAI学習を行う者が異なりデータの提供が行われる場合、1つ目の○のところですが、著作権法47条の7で特定者間は許されるという解釈もあるのではないかという御意見、2つ目の47条の7の明確化で対応できるのではないかという御意見がございました。また、上から4つ目ですが、AI学習用のコミュニティーをつくるという政策的提案を生かしていかないといけないという御意見ですとか、その下のSPVを考えてよいのではないかという御意見がございました。

その一方で、留意点といたしまして、留意点の2つ目、裁判例では特定者についてかなり限定的に解されており、近い関係が必要であるという御意見ですとか、一番下の「譲渡」だけではなくて、ネットでの送信・ダウンロードについてどうするか考えるべきではない

かという御意見がございました。

2 ページ目でございますが、協調領域だからといって何でもフリーとするということではなくて、目的に沿ったフリーの仕組みを決めていくべきだろうという御意見がございました。

3 ページ目でございます。

民間における学習用データの作成・提供でございますけれども、1 つ目の第三者への公開が違法になるのであれば、何らかの特例がないと、我が国だけAI開発がおくれてしまうという御意見がある一方で、慎重な意見のところの2 つ目ですけれども、情報を発信する側の自由な意思表示に基づく提供が本則であるべきであるという御意見、その下の本当に使えるデータを公共でシェアしようとする、プライバシーの問題が出てくるという御意見もございました。

③のところですが、公的機関における学習用データの作成・提供に関する御意見です。1 つ目の官民連携で、ビッグデータのアーカイブの整備を進めるべきであるという御意見、また、その下のAI学習用図書館という小見出しのついている部分ですが、図書館にアーカイブを構築して、AI学習ができるようにしてほしいという御意見などがございました。

4 ページ目でございます。

「オープンサイエンス」という見出しのところですが、AI学習の方法論では差が出ないということで、まず、データを押さえることが極めて重要であるという御意見がございました。

また、これに対して、留意点のところの2 つ目ですが、データを集めるときの法的な整理をどうするかという問題を挙げていただいた御意見。その下の公的機関といっても産学連携が進む中で、どこまでが公的機関とするかは難しいという御意見がございました。また、一番下のアーカイブの問題は、ためることではなくて捨てることであって、有限のスペースでデータをどうマネージするかということも問題であるという御意見がございました。

データについての御意見はここまでといたしまして、資料3をごらんいただきたいと思っております。オープンサイエンス、オープンデータの検討状況についての資料となっております。

めくっていただきまして、知財推進計画2016で関係部分を抜粋しております。現状と課題のところ、アンダーラインのところを拾っていただきたいのですが、企業等におけるオープンデータのような取り組みの一定の範囲内での促進など、データの共有・利活用がなされやすい環境整備について検討を進めていくことが必要であるとしております。

また、一番下の2 行ですが、公的研究資金による研究成果を格納しているデータベースについては、オープンサイエンスに係る動向を踏まえつつ、オープン化と保護のあり方について引き続き検討が必要であると言っております。

その次のページですけれども、今後取り組むべき施策として、アンダーラインのところですが、オープンデータのような取り組みの一定の範囲内での促進のほか、公的研究資金による研究成果や研究データのオープン化と利活用を促進するための著作権や契約のあり方などについて、検討を行う。あるいは、研究データを格納するデータベース構築と情報サービス提供に向けた考え方について、検討を行う。研究データシェアリングのプラットフォーム構築などを行うなどの検討を進める等々、記載がございます。

その次の3ページ目、4ページ目ですが、工程表が載っております、これらの課題に関しまして、内閣府の会議ですとか、文部科学省において検討を行うこととなっております。

5ページ目以降が、内閣府政策統括官からの提出資料となっております。

6ページ目、国際的な動向を踏まえたオープンサイエンスに関する検討会において、一昨年3月に取りまとめられた報告書の概要となっております。国としての基本姿勢でございますが、公的研究資金による研究成果の利活用促進を拡大することを我が国のオープンサイエンス推進の基本姿勢とすること、その下にございますが、オープンサイエンスの推進の目的・意義については、成果の再利用によるイノベーションの加速、新たな産業の創出等とされております。また、公開の範囲につきましては、研究論文及び論文のエビデンスとしての研究データについて、原則公開とすること。研究データの範囲については、(3)にあるとおりでございます。また、(4)研究を実施する機関において、研究データ等の管理に係る規則を定めることとしております。

7ページ目でございますが、これは関係省庁や各研究機関において実施方針や計画を定める際の留意点についてでございますが、1つ目の>にございますように、オープンサイエンスにおける「無料」と「自由」に関する定義を明確化し、関係者間で共通認識を醸成すること。2つ目にありますように、著作権ポリシーを定めること。3つ目にありますように、大規模データセットの所有権について定めておくことのほか、上から5つ目にありますように、例えば、オープンサイエンスを進めるために研究者へのインセンティブを高めることなどが留意点として挙げられております。

その下の8ページでございますが、今後の検討課題、今後、中期的に取り組むべき課題として、例えば、メタデータの統一など、論文、研究データの公開・共有化に関する検討ですとか、オープン化に係る著作権の問題ですとか、研究データの保存・整備の仕組みに関する検討などが挙げられております。

9ページでございますが、このような長期的課題や各機関における実施状況のフォローアップについて検討を行うフォローアップ検討会についての資料となっております。後ほどごらんいただきたいと思います。

10ページ以降が、IT総合戦略室の提出資料となっております。

12ページでございますが、こちらは昨年12月に成立しました官民データ活用推進基本法の概要となっております。右側の一番上の枠にございますように、政府による官民データ

活用推進基本計画の策定ということが第8条として盛り込まれておりますし、また、その下の枠、基本的施策の一つでございますが、11条といたしまして、国・地方公共団体・事業者によるみずから保有する官民データの活用の推進や関連する制度の見直しについて規定が盛り込まれております。

13ページ目でございます。この11条を受けまして、国・地方公共団体がオープンデータを進めていくことになるわけでございますが、ピンクの枠のところを見ていただきたいのですけれども、政府が進めるオープンデータというのは、公的機関が保有するデータを、例えば、CC BYなど「二次利用可能なルールの下」で、「機械判読に適した形」で公開されることとなっております。このように民間が編集・加工しやすい形でインターネットで公開することによって、新事業の創出ですとか、行政の透明性・信頼性の向上等をさせていくというものでございます。

以下、未来投資会議での総理の御発言の抜粋ですとか、データカタログサイトのデータセット登録数の推移などの表が添付されておりますので、後ほどごらんいただきたいと思います。

以上でございます。

○中村委員長 ありがとうございます。

次に、産総研の関口様から説明をいただきたいと思っております。よろしく申し上げます。

○関口委員 それでは、資料7-1「産総研データバンクについて」ということで、こちらのほうは、私ども研究機関で、今、データをどのように扱っているかというところに関して、御紹介申し上げるといところから御説明申し上げたいと思っております。

最初に、「産総研データバンク構想」というページがございますけれども、こちらでは、我々の研究活動の中で、いろいろな結果としてもしくはその過程としていろいろな取得されたデータ、それらがオンラインでアクセスできるものを御提供していくというデータと、それをサポートするシステムという形での構想でございます。

現在、下記にありますような4つのデータバンクを整備しているところでございます。1つ目は、地理空間・地図系ということで、これは私どもの傘下でございます地質調査所が特に提供しております地質関連もしくは衛星画像などのデータでございます。

2つ目は、物質・材料系と呼んでおりますけれども、さまざまな物質の特性に関するデータ、それは論文等で発表されたものを我々のところで確認したものをこういうデータバンクとして提供しています。

3番目は、人体系ということで、たんぱく質のデータベース等は必ずしも全て私どもが持っているわけではないのですけれども、そういうものに対するリンク、いろいろな人の形に関するさまざまなデータ、私どもでは「デジタルヒューマン」と呼んでおりますけれども、そういう人間の機能に関するデータというもの。

4番目としては、情報系。これは今回も特に議論になっておりますような、いろいろなところで得られるセンサーの情報でありますとか、人流や行動観測などを行ったデータ、

ある種の経済活動などのデータみたいなものを「情報系」と呼んでいます。

ここのプラットフォームの上では、私どもは、まずはこういうメタ情報、どういう情報があるかということを中心に整備することで、これらの認知度を上げること。それから、単にウェブ上で公開するというのではなくて、いろいろと有償、無償の利用を将来的に実現するために、例えば、SNSなどのアカウントとのリンクをつくっていく。我々のところで常にこういうデータを提供することは、必ずしも運用上好ましくないといえますか、サービスレベルが一定の保障はできませんので、例えば、クラウド上などに移行することで常に例えば商業利用などができるようにシステムをつくっていく。複数のクラウドなどとのリンクを実現しまして、例えば、大容量のデータ処理が必要になった場合には、こういうところのデータをそういう別のクラウドなどに提供できるようにしていくということ。このようなデータの利用に関するライブラリーなどを整備して、利用する人たちのノウハウが蓄積されるような環境をつくっていくという構想でございます。

めくっていただきまして、地質・地図系データバンクというのが、1つ、現在、我々のところでは非常に進展しているものではございますが、地質図でありましたり、衛星データ、これは生データというよりはむしろ画像のデータです。それから、岩石や断層などの情報などなど、それに関連する文献などを対象としております。基本的には、こういうデータが一つの地図上で統合的に検索・表示するということは、我々はサンプルのユーザーインターフェースとしては提供しておりますが、もちろん生データに対しても一定のOGC標準という国際的な標準に基づく規約でアクセスしていただければ、そういうデータが提供できるということになっています。

ライセンスに関しては、先ほど御紹介がありましたように、これを統一することを目指しております。ウェブサイトで掲載されているコンテンツに関しては、政府標準利用規約、オープンデータの規約でございますが、第2.0版に準拠し、または印刷物などに関しては、CC BYもしくはCC BY-NDなどで提供するというを行っております。

次のページは、物質・材料系のデータベースで、これも一つのユーザーインターフェースを御提示しているだけでございまして、必ずしもこういう形ではなくても、さまざまなクエリを出していただければ、統合的な検索は可能となっております。特に熱物性のデータベースとか、有機化合物のスペクトルデータベースに関しては、これは非常に多いアクセスをいただいています。

最後のページ、研究用データ公開の推進ということで、我々としては、研究関連のデータに関する引用の仕組みというものをきちんと構築していかないと、なかなか公開の促進ができないだろうと考えています。これらのデータは、必ずしも論文を書いた人ではなくて、データを作成した人、また、データベースを作成した人が正当に評価されるための基礎として、このような引用の仕組みということを中心に構築していかないと考えています。こういうものがデータに基づく科学の促進の第一歩と考えています。

その例としては、DOIというものを一つの例として、今、我々のところでは試みで使っ

ておりますが、そういうそれぞれのデータに番号、IDをつけまして、それを参照することによるデータサイテーションの仕組みということで、どれとどれのデータを組み合わせるかとか、そのデータがどこから来ているのかみたいな、そういうトレーサビリティについて基本的な枠組みを構築しています。

もちろん技術とかポリシーとか制度とかたくさん問題はあるのですが、現状では、JSTさんとかJalCさんなどと議論を行って、そういうところの技術委員会に参加し、国際的な枠組みへの加入を検討しているところでございます。また、研究データ利活用協議会、NIIの武田先生が主宰されておられますが、そういうところとも協力しています。

国際的には、RDAといわれるResearch Data Allianceという枠組みで、研究用データの利用促進をやっているところでございます。産総研としても組織加入を行っていたり、データサイテーション・ワーキンググループにもコミットしておったり、私自身もカウンシルのメンバーとして貢献を行っています。それ以外にも、学術会議とかWDS、これはNICTさんが今はやっておられますが、DIASさんなどの研究データを扱うコミュニティーと問題意識を共有しております。

以上でございますが、最後のページ、我々は何でこんなことをやっているかというのは、第5期科学技術基本計画にいろいろと知的基盤に関する記載があったり、それから、研究開発力強化法などに、そういう知的基盤の整備ということが記載されているということを一つの我々の根拠としております。

以上でございます。

○中村委員長 ありがとうございます。

次に、今日は、情報通信研究機構、NICTの鳥澤さんにお越しいただいていますので、御説明いただければと思います。

○鳥澤センター長 NICTの鳥澤でございます。

「AI研究者から見た著作権法」ということで、お話しさせていただきます。

まず、1枚めくっていただきまして、我々が、今、何をやっているかということなのですが、我々は自然言語処理の研究をしております。例えば、ウェブ数十億件の情報をもとにして質問に回答するとか、あるいはツイッター上の災害関連情報を大規模災害が起きたときに分析してお見せする。そういった技術を開発しております。これらに関しましては、民間企業へのライセンスも開始しているところでございます。

余り人工知能等に詳しくない方向けに書いた資料でございまして、ここでは釈迦に説法かと思っておりますので、1枚めくっていただきまして、我々は何を困っているかということをお願いしたいのでありますが、ここでも何遍か学習データという言葉が出てきておりましたけれども、我々は、なぜ空は青いのかといった質問に答えるシステムをつくっているわけではありますが、それをつくるに当たりましては、なぜ空は青いのかという質問と、さらにそれに対して回答の候補をウェブからもとってきてまして、それが正解であるか不正解であるかといったマークを人手で付与した、そういう学習データを大量に準備して機械学習



を回す必要がございます。

1枚めくっていただきまして、「AI研究者から見た問題点（その1）」というところがありますが、これはいただいた資料である程度アドレスされているようでありますけれども、こうやってつくりましたウェブ由来の学習データを、現状は他の研究組織等と共有することができないようでありますということであります。現在は、ライセンスをするに際しましては、学習済みのモデルをライセンスにしております。一方で、諸外国では、同じ学習データ、これはウェブ由来のものも含まれますけれども、それに対して、多数の学習アルゴリズムを多数の組織が試して、高精度のものを特定する。これは常識的に行われているプラクティスでございまして、日本では残念ながらそれができないということがございます。

全ての著作物由来の学習データの共有を認めていただく必要はないかと思っておりますけれども、そもそも誰でも閲覧可能なウェブデータ由来の学習データの共有が認められないというのは、いかがなものかと感じております。

ただし、「機械学習目的」等であれば譲渡してもよいといった、強い制約をかけるほどリスクは要検討であると考えておりまして、といたしますのは、学習データを人が精査することによって有効な学習アルゴリズムを特定する。これは通常の研究のプロセスでありますので、人が見てはいけない的な制約が入るのは本末転棟ではないかと思っております。

「学習データは全て公開すべし」的な縛り、ここでそういった議論がなされているかどうか分かりませんが、それも長い目で見ると研究の阻害要因になるのではないかと思っております。といたしますのは、データの作成というのは非常に手間、コストがかかりまして、実は研究者としての出世を考えるだけでありましたら、他人がつくったデータで論文を書いているほうがよほど得策であるといった現状がございます。

1枚めくっていただきまして、それ以外にも、人手でラベルが付与された学習データに限らず、ウェブデータそのものに関しても、組織をまたがったデータの提供は現状は黒であると認識しております。

例えば、NICTでウェブ100億ページ以上を実際に蓄積しておるわけでありますけれども、これを大学等の公的機関やこれから検索エンジン等を開発したいとおっしゃるようなベンチャー等に提供することは現状ではかなわないと認識しております。

数百億ページのウェブページを収集するというのは、お金と時間が非常にかかりますので、ベンチャー等では到底無理かということになるかと思っておりますけれども、例えば、国の機関がそういったところにそういったデータを提供することで、こういった産業を活性化するというのも夢ではないのではないかと考えている次第であります。

もう一点お話ししたいことがあるのでありますが、1枚めくっていただきまして、対話をするロボットの絵が描いてございます。これを現状我々が目標にして開発を進めているものでありますが、要するに、例えば、自動車の中でこういったいわゆる雑談あるいはブレインストーミングに類するような対話ができるようなシステムを目指して開発を行っ

ているところでございます。

1枚めくっていただきまして、開発を開始しまして半年程度なのですが、例えば、「iPS細胞で臓器を作るんだって」ということを言いますと、iPS細胞でがんワクチン用細胞量産技術を開発するかも、あるいはノーベル賞をとったかもといった発話ができるようになっているわけでありますが、実はこれは情報源はウェブでございます。

めくっていただきまして、「AI研究者から見た問題点（その3）」という最後のスライドを見ていただきたいのですが、現状、ウェブを情報源としてロボットのレスポンスをつくっているわけでありますが、それは極めて短いものですので、著作権法的には問題がないかと考えておりますけれども、今後、こういったロボットが進化するにつれて、より長く複雑な表現、創作性が認められそうな表現を発話する必要が出てくるのではないかと考えています。

こういった場合に出典がもしあった場合には、逐一引用もしくは検索エンジンの体裁をとってURL等を表示しなければいけないのか。あるいは、対話ロボットの場合ですと、URLを逐一読み上げなければいけないのか。そういったところは、少し悩んでいるところでございます。何らかの方法で解決ができるといいなと思っている次第でございます。

最後に、対話ロボットが創作性を認められる表現をみずから生成したとして、その取り扱い等々についても多少頭を悩ませているところではありますが、この辺に關しましては、もう議論がかなりされているようですので、割愛をさせていただきます。

私からは以上になります。

○中村委員長 どうもありがとうございます。具体的な問題提起をいただきました。

次に、文部科学省研究振興局の榎本参事官から説明をいただきたいと思っております。よろしくお願いいたします。

○榎本参事官 お手元の文部科学省からの提出資料をごらんいただけますでしょうか。喜連川先生が委員でいらっしゃるところで御紹介するのは非常に恐縮なのですが、私から大きく3点で御紹介いたします。

1 ページ、1 点目、NIIにおけますSINET整備に関してでございます。先ほども内閣府の報告書がありましたけれども、オープンサイエンス等が関心になっている中で、そうしたことが物理的に可能になっているのも学術ネットワークができているからこそでございます。この学術ネットワークに關しましては、NIIによりましてはSINETとして長年整備されてきたところでございます。このSINETに關しまして、この1 ページでは大きく3つ並べているところでございます。左側がSINET本体の部分でございますけれども、本年度から100ギガで、国内、海外とも必要なネットワークの体系でつなげているところでございます。そうしたものとあわせて、2番、3番でございますけれども、共通基盤の整備ということで、データの保存、利活用ができる基盤の開発、そして、機関リポジトリ、図書館における検索システム等の対応がございます。

2 ページにまいりますと、先ほどございましたけれども、オープンサイエンスが第5期

の基本計画で大きな柱となっているところでございます。第4期までは、研究情報基盤の整備といった一般的な話でございましたが、第5期からはオープンサイエンスということで、コンセプトがこの従来からのオープンアクセスにおけます論文にとどまらず、データも含めたものとして拡大されているところでございます。先ほど関口先生からお話がありましたけれども、現在、JSTやNII、NICT、関係機関等と連携しながら各種データに関するDOIを初めといたします規格の整備、国際的な議論に参加しているところでございます。こうしたネットワークと関連するインフラの整備がある中で、オープンサイエンスに対応していこうというところでございます。

3 ページは、NIIの各種事業ですので、これも割愛いたします。

4 ページのところでは、NIIの活動の一環といたしまして、情報学研究データリポジトリという、これは多くの企業とNIIで締結をいたしまして、各種の民間企業におけますデータの研究者におけます利活用を進めていこうという枠組みとして、こうしたデータリポジトリの整備を進めているところでございます。こういったところが、まず、NIIの幅広い活動のうち的一端でございます。

5 ページにまいりますと、こちらは、今年度28年度から始めております理化学研究所におきますAIPセンター、革新知能統合研究センターでございます。こちらに関しては、5 ページの左下にございますけれども、政府の「人工知能技術戦略会議」のもと、総務省や経産省の関係機関等々も密接に連携しながら行っていくものでございます。本年度から日本橋にセンターを立ち上げて、資料真ん中左側の箱がございますけれども、汎用基盤技術、目的指向基盤技術、社会における人工知能研究グループ、こういった3つの大きなチームを設けて、基盤的な技術、それから、目的指向の研究等を中心として取り組んでいるところでございます。

6 ページでは、現時点での研究のチームリーダーの例を挙げているところでございますけれども、順次拡充しているところでございまして、6 ページは現時点で理研のウェブサイトで公表しているところでございます。

こうしたチーム体制を組んでいながらでございますけれども、基本的には汎用基盤技術ということで、この理論、基盤技術に関する汎用的な検討が中心になりますが、7 ページにまいりますと、各種の応用先と連携した形で、目的指向研究も幅広く進めていくとしているところでございます。左右2つに分けましたけれども、左側が日本がもともと強い分野をAIで強化していく。例えば、京都大学のCiRAとの連携、名古屋大学の未来材料・システム研究所との連携、また、右側のほうで、国内の必要性という観点で、ヘルスケア、防災・減災、インフラ、観光といった観点でも幅広いデータも応用先と連携しながら活用して取り組んでいくとしているところでございます。

最後、8 ページはJSTでございます。JSTにおきましても、幅広い研究者支援をしている中で、戦略的創造研究推進事業として大きくJST事業がこの8 ページで3つございます。チーム型のCREST、個人研究のさきがけ、若手を中心に支援いたしますACT-Iとなっていると

ころでございます。このCREST、さきがけ、ACT-Iとある中で、今度は横軸に見てまいりますと、ここに四角が全体で7個ございますけれども、幅広いテーマに関しまして、大学や研究機関等を支援しているところでございます。

この中で、上から2つ目の箱で、ビッグデータの応用領域ということで、北大の田中先生に領域総括をしていただいている件がございますけれども、これを1枚めぐりまして、9ページ、10ページでございます。

CREST「ビッグデータ応用」といわれる分野、ここは、気象、農業、論文データ等、幅広いビッグデータの応用先を想定した研究を進めているところでございますけれども、現在、このJSTで準備中でございますが、こうしたさまざまなデータに関して、それぞれの研究者における研究にとどまらず、ビッグデータ応用の体験型ポータル開発を進めているところでございます。9ページにイメージ図がございますけれども、アプリ領域として、ここで、例えば、医療、農業、地球環境、都市インフラとしていますが、こうしたものに関して、研究成果を公開していくという観点でこのデータを整備、活用、公開していくことも進めているところでございます。

10ページは、その機能のイメージということで、ここでは、例えば、北海道における除雪プロジェクト、都市インフラと関連いたしますけれども、そうした非常にこの地域におきまして除排雪に関して多額の費用と人力的な配置が必要となっているところを、こういったものをどんどん効率化していこうという観点で、多様なデータを収集しているところ、これを単独の保持とせず幅広く公開して、さまざまな観点で研究に活用してもらおうとしているものでございます。

以上でございます。

○中村委員長 どうもありがとうございました。

委員の皆さんから、意見ないし質問をいただきたいと思います。

発言を希望される方は、ネームプレートを立てていただければと思います。

国立国会図書館の川鍋副部長がお越しになりましたので、申しわけないのですが、早速ではございますが、説明をお願いしてよろしいでしょうか。もうちょっと待っても結構なのですが。

○川鍋副部長 遅くなりまして、恐縮でございます。

国立国会図書館電子情報部の川鍋でございます。

私どもの「国立国会図書館におけるデータの収集・保存・提供の取組」について、資料7-4でお時間を頂戴いたしまして、御説明いたします。

1枚おめぐりいただきまして、スライド番号1のところ、私どもで持っておりますデータのうち、デジタル化資料の画像データ、インターネット上の資料、国会会議録、また、書誌データ・典拠データについて御説明いたします。

スライド番号2のデジタル化資料（画像データ）の提供でございますが、左の円グラフの下にございますように、私どもは国立国会図書館デジタルコレクションというサイトに

おきまして、現在、デジタル化資料約262万点を提供しております。過去の補正予算等におきまして、大量のデジタル化が行われた時期もございまして、現在、その点数を提供しております。そのうち、絶版等で入手困難な149万点については、全国の公共・大学図書館等に送信をしております。また、著作権保護期間の満了が確認できたもの等につきましては、50万点をインターネット公開しております。パブリックドメイン資料の35万点については、当館への申し込み不要で、営利・非営利を問わず自由に利用することが可能でございます。下に、課題としまして、テキストデータについて書いております。現在、目次のみ作成して提供しておりますが、私どもがOCR処理してテキストデータを作成して、スニペットで表示して提供するためには、著作権者の許諾が必要ということで、なかなかテキスト化が進んでいないという課題がございます。

3 ページ目に、インターネット資料・オンライン資料について、現状と課題をまとめました。右上の「現状」にありますように、私どもでインターネット資料の収集保存事業ということで、平成14年からウェブサイトの収集・保存をしております。国等の機関のウェブサイトについては、制度的な収集として、民間のウェブサイトにつきましては、発信者の許諾を得て収集しております。ウェブに掲載されている文献相当の資料をオンライン資料と呼んでおりますが、それについては無償かつ技術的な制限、DRMのないものについて制度収集を約40万点行っております。私ども国立国会図書館の施設内での閲覧が可能でし、発信者の許諾が得られたものについては、インターネットでも公開しております。課題につきましては、今、申し上げましたように、インターネット公開には発信者の許諾が必要でございまして、また、これらの資料を人工知能の学習用データで利用可能な形で複製の提供をする場合には法制度の整備が必要ということを課題として挙げております。

スライド番号4、国会会議録（テキスト・画像データ）の提供についてでございます。衆議院・参議院の事務局と共同いたしまして、会議録のフルテキスト・データベースを構築しております。第1回国会以降の会議録のテキストデータ、また、画像データを、現在173万ページ提供しております。こちらについては、検索結果の複製、また、フルテキスト・データベースにおいて、検索用APIでの機械的な連携、営利目的は除きますが、そちらが可能となっております。

最後のスライドでございます。書誌データ、典拠データということで、私ども国立国会図書館で所蔵している資料がこういった資料であるかという情報、本でしたらタイトルとか著者名などといった書誌情報と申しますが、それをOPACまたは国立国会図書館サーチから検索できるほか、API、RSS配信等で、現在、1,170万件、こちらのデータを提供しております。また、典拠データと申しまして、この書誌にひもづく著者などの個人名、団体名、また、いわゆるキーワード（件名）、そういった言葉についてもデータを管理しております。提供しております。この典拠データに関しては、まさにこの人工知能の学習用データとして有用ではないかと考えております。現在、122万件のデータを提供しております。これらのデータはリンクトデータでも提供しております。課題につきましては、現在、私ど

もの書誌データ、典拠データ、非営利目的の場合には無償で提供しておりますが、さらなる利用の拡大に向けた取り組みを必要としております。

書誌データ等のメタデータ、これはかなり権利的にはCC0での提供が望まれるところですが、とにかく我が国の場合、なかなか書誌データ等のメタデータのCC0での提供が進んでおりません。私どもも、書誌データについて、国立国会図書館サーチでも提供しておりますが、分野横断的な統合ポータルとして、ジャパンサーチ（仮称）の構想もございます。こういったメタデータの提供の権利のところは、CC0というところが大変望ましいと思っておりますけれども、その辺が課題と考えております。

遅くなった上で、恐縮でございます。

以上でございます。

○中村委員長 どうもありがとうございました。

御到着のところ、無理を申しあげまして、申しわけありませんでした。

委員の皆さんから、意見、コメント、質問などをお出しただければと思います。

まず、福井さん、どうぞ。

○福井委員 福井でございます。

ありがとうございます。

勉強になる御発表ばかりで、まだ消化ができないスピードのものも大分あったように思いますけれども、その中で、資料7-2、鳥澤先生にお伺いしたいです。実を言うと、私あたりにはこのぐらいのやわらかさがちょうどいい感じでありまして、非常にわかりやすかったのですけれども、その中で、問題点の1と記載いただきました。一方においては、学習用データを他の研究組織等との共有で悩みがあると書いていただいています。これは、想定されているのは、多数の団体や個人との間で、学習用データを相互に開放し合うような著作権の例外規定がない現状をおっしゃっているという理解でよろしいでしょうか。

○鳥澤センター長 不特定多数であるかはとわず、例えば、我々がシステムをライセンスしている企業さんなどでも、みずから学習データをさわりたい、さらにその性能向上を御自分たちでおやりになりたいという具体的なニーズがございます。

そのときに、要するに、不特定多数ではなくて1対1ということになるかと思っておりますけれども、そこでお渡しするのは、どうも我々の顧問弁護士にお聞きしますと、黒ではないかということで、ここをお願いしている次第でございます。

○福井委員 わかりました。

続いて、機械学習目的ならば相互に共有可能というのだとまだ制約としては強いということを書いていらっしゃると思いますが、それだとどんな目的がほかに記載としては考えられるでしょうか。

続けてお伺いしてしまうと、今度は逆に集めた学習用データを全て公開すべしだと、これも阻害要因であると考えていただいています。つまり、これは共有できるけれども、共有を義務づけられないと理解すればよろしいのでしょうか。

○鳥澤センター長 まず、1つ目の質問でありますけれども、その「機械学習目的」という言葉の意味かと思いますが、要するに、人が一切目に見ることなく機械学習機にデータを流し込むだけであればオーケーだという形での許諾だと仮にしますと、データを受け取った側は実際には研究ができないということになります。といたしますのは、機械学習機というのはよく誤解を受けるのですけれども、何かすごいプログラムがあって、何でもいからデータを流すと勝手に賢くなってくれるというイメージをお持ちの方は大勢いらっしゃるのですが、実はそうではなくて、機械学習にもいろいろな方法がありまして、それはデータであったり、あるいはアプリケーションであったりに合わせて選択をする必要が出てきます。その選択をする際には、学習データを精査する必要が出てきますので、広い意味では機械学習目的ということは可能かもしれませんが、そういった事情がございます。

もう一つ、2つ目の質問でありますけれども、ウェブ由来の学習データは全て公開すべしということが、実際、そういうことが議論されているかどうかは存じ上げないで書いておりますけれども、例えば、そういうことになると何が起きるかといいますと、要するに、他人が過去につくったデータを使って論文を書くだけであると。そういった研究者が恐らくアカデミアではふえてくるでしょう。といたしますのは、論文の数を稼ぐにはそれが一番効率がいいからであります。

一方で、学習データをつくるというのは非常にコストも時間もかかる作業でありますので、その学習データをある程度自分だけ使うという体制にして、自分がそのトピックに関しては論文が書けるといったことができないとすると、それだけコストをかけた意味がないということになりかねないのかなと思っています。

あとはそのビジネス界でも同じような現象が起きると思っております、要するに、どこかでお話があったかと思いますが、要するに、機械学習のアルゴリズム自体で余り差が出ないとすると、差をつけるとすればどういう学習データを保有しているかということで、ビジネス上の差異化を図ることになるかと思っておりますので、そこでもって全部オープンにすべしという話になってしまうと、本末転倒かと思っております。

○福井委員 確認のためのお尋ねだったので、共有を義務づけられないと理解しました。

最後に1点だけ、問題点のその3のところ、上のほうのポツですけれども、対話ロボットが創作性の認められそうな表現を発話する必要が出てくる。これはロボットが発話する内容に注目されたのだと思います。その下の小さいポツのところ、このときに出典があれば逐一引用等の体裁をとるとするとかなり大変だと書いていただいているが、これが今度は逆に学習している世の中のいろいろな文書に対する出典元の明示のことをおっしゃっているのだと思うのです。

○鳥澤センター長 すみません。ちょっと誤解があるかと思いますが、要するに、学習ということと検索というのは我々の観点では違うものになります。違いをここで説明するのはなかなか難しいのですけれども、学習といいますのは、ニューラルネットのように、も

もとの学習データの現状をとどめないような何かに落とし込んでしまうことであります。

一方で、ここでロボットが出典から何か発話の内容を出すときには、これはむしろウェブ等のテキストをストアしたデータベースから引っ張ってくるだけ、それはつまり検索ということになるかと思えますけれども、そういうことを意図してございます。

○福井委員 私のお尋ねは、この引用、出典等がある場合に、引用の体裁をとらなければいけないかというのは、例えば、出典はこれでしたよという出所の表示等を一々行うのでは、対話にならないだろうということをおっしゃっているわけですね。

○鳥澤センター長 そういうことです。

○福井委員 ということは、前提として、この対話ロボット発話の中には、ウェブの中などから引っ張ってきた、ためておいてもいいのですけれども、もとの文章が結構原型のまま発せられてしまうことを想定して、それだと出典元を表示しなければいけないのは大変だなということの御理解でよろしいでしょうか。

○鳥澤センター長 はい。

○福井委員 これはやはりかなり生じてきそうですか。それとも、この対話ロボットが相当高度化されていくと、もっと文章は消化されてしまって、出典はそんなに大きな、言ってみれば、もとのデータソースからそんなに大きなかたまりの文章しか引っ張ってこられないことが想定されていますか。

○鳥澤センター長 例えば、家庭にその対話ロボットが入るという話題はアメリカではよく話題になって実際に商品が出ていますけれども、例えば、有名なレシピサイトからとってきたレシピをそのままキッチンにいるお母さんにしゃべるということを考えると、これは恐らく創作性が認められると思いますので、それはできないということになりますね。そうすると、その大資本の企業が持っている自前のレシピサイトの情報だけしか提供できないのか。

そここのところはいろいろな考え方があって、無常件にその引用の要件を満たすことなくしゃべっていいというのは乱暴だと私も思いますけれども、その辺はどうなるのだろうか。

○福井委員 そうですね。そもそも引用に当たるかという大きなハードルがそこにはありそうですけれども、問題点はよくわかりました。ありがとうございました。

○瀬尾委員 短いプレゼンでしたが、大変大変勉強させていただきました。

幾つか順番にお話しさせていただきたいのですが、まず、目的は、何度も繰り返されているように、AIを日本の産業施策の中でどのようにプラスにしていくかというところで、今、検討を行っています。この鳥澤さんのお話の中で、こうしてほしいというところがあるのですけれども、現状で十分と思われていないから変えなければいけない部分があるということがまずわかりました。ただ今回、幾つかに分けて考えるべきではないかと思っています。というのは、学習データセットを組むために、データをクリーニングして組み立てることに実は非常にバリューがあるとすると、これについては、逆にオープン化をして



しまうと、そのバリューの価値が失われてしまうので、軽目か重目かは別にしてやや保護が必要な方向になる。クローズ側に入る。ただし、もともとの生データと呼ばれる部分に関しては、でき得る限り協力して共有してオープン化していくことが望ましいというお考えでしたね。

○鳥澤センター長 おっしゃるとおりです。

○瀬尾委員 私ももともとのデータをどう集めるかということに、多分、今のAI政策の基本がかかっていると思っています。今、いろいろな研究で、今日関口さんからご指摘ありましたが意外と生データはたくさんあるのだなと。例えば、先ほどNDLさんのウェブに関しての収集があったりしつつ、こちらの鳥澤さんのところでもウェブの収集があったりしているけれども、これが連携しているのかどうか。

つまり、どんどん生データを集めていくための施策をどういうふうにしたらいいかを専門的な現場の方々におっしゃっていただきたい。、例えば、ポータル的なストレージをつくって、ある程度のルールの中で、ともかく生データだけは共有させてくださいよと、そういう場をつくってほしいということが、私の理解ではそれが本当に必要だと思うのだけれども、そういうことをつくっていくことで大きく前進できるのかどうか、ご意見を伺いたい。今日のレクチャーをしていただいた現場の皆さんに、何が欲しくて、何があったら進むと思うのかをズバリ出していただいて、それがこれまでここで議論してきたこととどう合致するのか考えたいと思います。

私だってAIをそんなにわかっているわけではないのに、頭でっかちでこれがいいだろうなどと言ったら、全然現場と外れていたなどというのはよくある話ですから、リアリティーがあって、即効性がある、最も効率がいい部分をぜひ今日は皆さんからお伺いして、それを実現できる施策に落とし込むということ、今日はまとめなのでできたらいいなと思いました。

ただ、方向性としては、今言った、生データと学習済みセットのバリューの違いと、さらに生成物のバリューの違い、ただ、これについては、原著作物とのそごとかいろいろなことが出てくるということで、まずは生データの部分について、できれば皆さんに何がいいか、ぜひ教えていただきたい。それを生かせないとこの議論は余り意味がなくなってしまうと思いました。ぜひそこのところは、簡単にですけれども、皆さんからお伺いできればと思います。

○中村委員長 鳥澤さん、どうぞ。

○鳥澤センター長 どうもありがとうございます。

正直に申し上げまして、我々が生データというとウェブデータなのですけれども、それに関しましては、著作権法で他社に渡すことはまかりならぬという認識でおりましたので、どういうストレージでどういう共有のシステムでというのは考えていないというところが正直なところであります。ただ、著作権法でゴーが出せれば、例えば、そういう共有サーバーを立てるとか、あるいは何らかの方法でもってそのデータをトランスファーする方向

については、ぜひとも検討を進めたいと思っております。とにかくだめだということが出発点でありましたもので、すみません。そこまで考えが及んでおりません。

○中村委員長 ほかにいかがでしょうか。

関口さんでも、榎本さんでも。

○関口委員 ありがとうございます。

生データは、研究のほうの生データという観点であれば、今、それぞれの本当の分野分野によって全然ポリシーが違ってしまっていて、ここの地質の分野、材料とか、そういう物質の分野、本当の基礎物理などでもそれぞれ全然違うので、なかなかそこを統一的にやれという話に関しては、非常に過去にも難しかったことがあります。ですから、そういうのも踏まえて、今はそれぞれのそういうドメインドメインで、自分たちのポリシーでワークするものをきちんとやって、そういう境界領域のデータをどう扱うかというところを、ここがまだ余りきちんと議論ができていなくて、地質のデータと、例えば、防災のデータなどをどういう枠組みでやっていくのかというところできていない。

そこは何らかの相互運用できるようなクリアランスをとっていくということなのですが、それが、今、ぎりぎりできているのが、いわゆる政府系のオープンデータというところで、一定のポリシーでもって、ここはお互いに気にしなくていいですねとやっている。そういう意味では、非常に地道なプロセスでしか今はできていないという状況です。

○瀬尾委員 どう改善したらいいですか。制度の問題ではないですか。

○関口委員 そこが制度の問題ではないのではないかという気はしています。ですから、国内だけの問題ではないので、研究のコミュニティーは国際的なそれぞれ皆さんの活動をベースにされていますから、海外の動きと国内とのそういう整合性というのはなかなかとるのが難しいという認識で、今、動いているので、どうすればいいかということに関しては、まだちゃんとした回答は持っていません。

○中村委員長 ほかにいかがでしょうか。

どうぞ。

○川鍋副部長 国立国会図書館でございます。

やはり端的に申し上げれば、先ほど鳥澤先生がおっしゃったように、法制度の整備ということになるかと思えます。私どもで収集しましたインターネット資料、オンライン資料についても、現在は複製の提供の場合には許諾が必要ですので、国立国会図書館で収集したものについては権利制限でAIに提供できるという制度の整備が進めば、かなり違うかなと思えます。もちろんストレージ等の予算的な措置も課題としてありますけれども、やはり制度のところかなと思えます。

○中村委員長 よろしいでしょうか。

○瀬尾委員 今の制度の問題ではないという話があって、例えば、基本的に研究者の皆さんが自分のつくった生データ自体、収集した生データ自体も囲い込みたいと。余りオープンにしたいくないというベクトルが働いているのかどうか。それとも、それを出すような方

向にできるのかどうかというのが1点あるのと、もう一つ、ウェブにしても、例えば、検索エンジンの収集、クローリングに関しては、権利制限の中でできているわけですね。つまり、どれだけその権利者の権利を不当に害するか。先ほどみたいに、有名なレシピサイトを持ってきてべっと出して、それは幾らなんでもまずいと思います。もともとつくるインセンティブがなくなってしまうから。

だけれども、使える部分と使えない部分とあって、使える使い方と使えない使い方、つまり、権利者に対して不当に害するかどうかというところで、全部ではなくて、線が引けるはずだと思うのです。その部分についてできればより詳しく突き詰めて、先ほど川鍋さんもおっしゃったような話をきちんと実現化していかないと、やはりやめてしまうことになってしまうので、何とか考えなければいけないかなと思っています。

それと、今、先ほど関口さんがおっしゃった海外との連携、例えば、今、国会図書館さんが海外とのデータのやりとりをできていないのです。デジタルデータの送信とか。海外とコネクティングしていかないと日本はまたやはりガラパゴスになってしまうので、そういった現場の人たちからの海外とのバランス、それからウェブ収集とか、できないと言っても本当はもっと区別できる部分があるような区別、それから、中に囲い込もうとする力がもし働くのであれば、どこを出していくかをちゃんと区別する。とにかく早急にデータを集めないとならないというのが、AIにとってミッションなのです。そう私は思っている。だから、それについて皆さんのお知恵を少しずついただきながら、何とか施策に盛り込んでいくべきではないかと思っています。

ですので、今後ともそういう意見を現場の声というのをできるだけダイレクトにいただいていくことが必要かと思っておりますので、今後ともそれを続けていっていただければと思います。

問題提起で終わりました。

○榎本参事官 ありがとうございます。文部科学省です。

先ほど関口先生がおっしゃったように、なかなか研究者における研究データの整備は課題がある中というところがございます。先ほど御紹介したとおり、この政府としてオープンサイエンスを進めていくのだ、したがって、その研究成果、エビデンスに関しては抱え込まずに、もちろん非公開にすべきところはあるけれども、そこは認識した上で、基本的には成果を出していくのだという立てつけで、総論としては進めているところです。

したがって、政府におきまして、JSPSやJSTのようなファンディング機関に関しては、今後、その研究成果に関してはデータも含めて公開をどう具体的に義務化させていくのかというのは、手順を踏む必要はございますけれども、公的に資金を得た研究については、データを公開する方向で進めていくという流れにもなっているところです。

あとは、関口先生がおっしゃったように、分野別にいろいろな状況がある中で、学問分野ごとのコミュニティーにどうより主体的にこの検討を進めてもらえるかということだと思っています。そういった分野ごとの検討がさらに進んでいくと、いろいろなデータを

格納する場をどうしようかという議論も、恐らくNIIやJSTにおいて具体的にできると思っています。

ですので、まず、物理的にストレージをどうしようかという整備もあるものの、研究者コミュニティでそのことを活発にさせていくことと両方相まって進めていくと、流れとしてはできていますので、あとは戦術、戦略をどう具体化させようかというフェーズには来ていると思っています。

○関口委員 ちょっと補足をさせていただいていいですか。

○中村委員長 どうぞ。

○関口委員 先生の御質問に対する補足なのでございますけれども、その研究者は、今、いろいろな環境の中でオープンにしていくというところには、割とポジティブに取り組もうとはしています。ただ、もちろんある種の自分としての貴重なデータに関しては、ある機関のエンバゴということは必要となっているという議論がベースになっているところ です。

ただ、一方で、オープンにすることは全く問題はないのですが、逆にデータをメンテナンスしてくれと言われるところで、研究者というのは、なかなかそこは自分の仕事ではない。それでは、そのいろいろなキュレーションを含めて、データのマネジメントを誰がどのようにやってくれるのかというところは、今、そういう意味ではふわっとしているというところで、どういうふうにデータを、機関リポジトリというものを、例えば、我々がきちんと提供するのか、それとも、NIIさんにある部分を提供していただくのか、各機関でやるのかというところがまだ余りクリアになっていないというのが、そういう意味では、制度としては現状整っていないところではないかと思えます。

○中村委員長 ありがとうございます。

森さん、どうぞ。

○森委員 ありがとうございます。大変勉強になりました。

1点申し上げたいのは、前回からの続きになりますけれども、権利制限ということをごどのように進めていくかということが明らかな課題であるということかと思えます。今の御発表で非常にその点は明確になったかと思うのですけれども、研究者の方が手をかけたデータを公開させられるのは困るというお話が出たけれども、一つの視点として、利用できるかどうかという知的財産権の話と、その公開状態になっているかどうかというのは別でして、それは区別して議論いただいたほうがいいかなと思えます。

オープンにしなければいけませんなどということは法制度としては非常に珍しくて、実は人から聞いた話をベースに翻訳ウェブサイトで調べたのですけれども、韓国に位置情報に関する法律というものがあまして、位置情報を何らかの立場によって管理している人、例えば、日本でいうと通信キャリアなどはそうなのかもしれませんが、そういった事業者は、ほかの事業者からどういうIDとの連携なのかわからないですけれども、一定のIDで位置情報の照会があった場合に教えてあげなくてはいけないみたいな規定、これは私が日本

に訳されたものを見て衝撃を受けたのです。私の経験としては真実ですが、その翻訳サイト等がどの程度信用できるのかわかりません。

ただ、そういう手元にあるデータを必ずオープンにしないみたいなことは、それは非常に法制度としてはイレギュラー中のイレギュラーですので、そのことを余り心配されることなく、権利制限のほう、公開されているもの、アクセスできるものを利用できるのかどうか、あるいはライセンスを受けたもののライセンスの効果の範囲がどこなのかということも議論していったほうが良いと思います。

以上です。

○中村委員長 上野さん、お願いします。

○上野委員 ありがとうございます。

国会図書館の川鍋さんに2点、鳥澤さんに1点お伺いしたいと思います。

まず、最初に、国会図書館に関しましては、最後に御指摘になった「④書誌データ、典拠データの提供」についてお伺いしたいと思います。先ほど、書誌データや典拠データは大変有用なデータであるにもかかわらず、現状では、非営利目的の場合に無償で提供しているのみだというお話がございました。ただ、書誌データも典拠データも著作権のない単なるメタデータではないかと思っておりますので、その提供ができないとしましても、それは著作権法上の問題ではないように思います。そういったしますと、現在、非営利目的に限って無償提供されているというのは、何か別の法令上の制約があるということでしょうか、あるいは、運用上、権利者との協議等に基づいて制約されているということでしょうか。そのあたりの具体的な原因についてお伺いしたいと思います。

2点目は、最初のほうで御指摘になった「①デジタル化資料（画像データ）の提供」についてであります。この点に関して、現状ですと、OCR処理をしてスニペット表示をすることができないという御指摘がありましたが、それはそのとおりかと思っております。

ただ、現在、著作権法の改正の検討をしている中で、いわゆる「所在提供サービス」を一定の範囲で適法にする権利制限規定を設ける方向で進められているところです。もしこの改正がうまくいったといたしますと、国会図書館にもこの規定は適用されて、国会図書館が、その所蔵資料の中にある言葉をキーワードとして検索できるようにするサービスを提供し、そこで当該キーワード前後をスニペット表示することが、著作権法上はできるようになるようにも思うわけですが、この点についてどのようにお考えでしょうか。

それから、鳥澤さんからは、先ほど、ウェブデータ由来の学習データの共有は認められるべきだとおっしゃる中で、たとえ認められたとしても機械学習目的などという制約が課されると問題であるという御指摘がございました。確かに、機械学習の過程で人間が関与することも考えられますので、機械学習目的というのが狭い意味の機械による学習という場合に限定されてしまうと問題だというのは、確かにそのとおりかと思っております。

ただ、現状の著作権法47条の7という規定は、世界でもこれだけ大きな範囲の制限規定はないのですけれども、あれにも一応「電子計算機による」という条件が入っております。

したがって、同条の規定によって情報解析が許されるといっても、電子計算機を使っていなければいけないということになります。ただ、一部でも人間がかかわったら、もうこの規定は適用されないかかというところではありませんで、「電子計算機による」解析が主であれば、人間がかかわってもよいと解釈されております。

そういう観点からいたしますと、ウェブ由来の学習用データの共有について、例えば、解析目的という限定が課されたとしても、特に大きな問題はないようにも思うわけですが、この点についてはどのようにお考えでしょうか。

以上です。

○川鍋副部長 国立国会図書館でございます。御質問ありがとうございます。

上野先生の御質問の1点目、書誌データ等の提供についての著作権法由来のところの障壁といいますか、私ども館内で議論しておりますのは、著作権法との関係で、私ども国会図書館でつくったデータを国有財産として提供することと無償ということの関係をどう考えるかというところで、今、館内でも議論の整理がついていないというところが、私どもの事情といいますか、そういうことではございます。

2点目の御質問のスライド1に関しての所在検索サービスとしての本文テキストの件でございます。今、お話がありましたように、文化庁様で法整備の検討をされていると認識しておりますので、その影響を私どもとしても注視しているところです。権利関係の法整備が進みましたら、私どものほうでのテキストデータを作って、それを検索利用、スニペット表示するというところは進んでいくと考えております。

○鳥澤センター長 鳥澤です。御質問どうもありがとうございます。

おっしゃるとおりで、これは法律の素人である私から見たときに、要するに、恐ろしいなキーワードがいっぱい並んでいますと、制約条件としてこれはだめなのではないかと勝手に思ってしまうわけですが、その辺御配慮いただけるとありがたいかなぐらいのつもりであります。

以上です。

○中村委員長 ありがとうございます。

今枝さん、飯田さん、お願いします。

○今枝委員 簡単な意見を1つと、多分これは委員の方にしたほうがいいかと思うのですが、非常に稚拙な質問が1つございます。

まず、意見として、今、研究目的のデータの話と商用目的のデータの話というのをもう少し整理したほうがいいような感じを受けました。つまり、著作権法でも例外の規定の中に幾つかそういったことが列挙されているわけですので、そこを分けないと多分整理がつかないのかなという気はしております。すみません。アバウトな意見です。

すみません。これは非常に稚拙な質問なのですが、今、ウェイバックマシンとかウェブアーカイブといわれているもの、これは違法なんでしょうか。

つまり、私は、実は侵害訴訟の中で3、4回、ウェイバックマシンのデータをもとに、

特許の裁判ですと公然実施ですとか文献公知までも主張して、最終的に日本ではまだ認められていないのですが、アメリカとかでは幾つか認められている判例もあると聞いております。つまり、証拠性があるということで、これは違法なものであった場合に証拠性の関係が見えないと思ひまして、何が言いたいかといいますと、デッドコピーを丸々ウェブからして違法性はないという状況はどういう状況なのかというのは、実はよくわかっていなかったということがあります。

もう一つ、これもまた非常に稚拙です。先ほど翻訳の話が出ました。ウェブのものをそのまま翻訳をして出力をする。このページを翻訳して出力をするというサービスが、うちはやっていませんけれども、ほかのところではやっているように見受けられます。これは一つ一つ当然ながら全てのウェブサイトから許諾を得るわけにはいかないのです、これは違法なのか、合法なのか。

それらが仮に合法だったとしたら、そこら辺の応用がきくような気がしております、非常に素人的な質問で申しわけないのですけれども、委員の中にも法律の素人がおりました申しわけないですが、どなたかお答えいただけると。

○中村委員長 この場には、裁判官はいませんが、文化庁と学者と弁護士はおられますので、どなたかが答えられると思います。

いかがでしょうか。

○奥郵委員 すみません。来たところなので前後がよくわかっていないので間違っているかもしれませんが、今、ちょっと伺いました点ですが、検索エンジンに関するキャッシュがアメリカでフェアユースになるかどうかという地裁レベルの裁判例があったとは思いません。その判決がデファクトとなってサービスされているということであって、かなり高いレベル、例えば、最高裁であるとかでオーケーであるということまでが出ているわけではありません。インターネット検索エンジンが始まったころに、キャッシュという機能があったときに、それがフェアユースになるのではないかという判例がたしか1つはあったと思います。

ただ、検索エンジンのキャッシュとウェイバックのキャッシュが同じなのと言われると、前者は検索のための提供ということで、トランスフォーマティブということが言えるのですけれども、そうではない単なるアーカイブとしてのものが大丈夫なのかというと、多分私の記憶では正面から争われて、イエス、ノーというのはアメリカでも出ていなかったのではないかと。すみません。最近では追いかけていないのですけれども、アメリカについてはそういう理解をしております。

以上です。

○上野委員 日本のことについてだけ補足いたしますけれども、現状の47条の6という、検索エンジンを著作権法上適法にするような明文の規定は、おそらく日本しか持っていないものだと思いますけれども、同条の規定においても、一応「送信可能化された情報」という条件がかかっておりますので、すでに消失してしまったウェブサイトについては、現

状において送信可能化されているものに当たりませんし、それを検索できるようにするサービスも、現在アクセスできるウェブサイトの所在を検索するものではないため、同条の規定は適用されないと考えられております。

ただ、立法論としては、ウェブアーカイブのようなサービスも一定の範囲で適法にする規定を設けることが検討されるべきだと私は思っております。なお、それ以前に、アメリカで行われているサービスだといえる場合は、そもそも日本法が適用されるものではないと理解できようかと思えます。

以上です。

○中村委員長 どうぞ。

○福井委員 証拠能力について補足いたしましょうか。

○今枝委員 ありがとうございます。

○福井委員 これはほかの方々に同意いただけるかわかりませんが、ウェイバックマシンそのものが仮に著作権侵害であるとしても、著作権侵害の行為によって収集されたからといって証拠能力が否定されるものではありません。そのもの自体が、真実、過去に存在したウェブページの内容であると信じられる理由があれば、証拠能力というのは認められようかと思えます。特に民事においてはこれは言えると思えます。

○今枝委員 ありがとうございます。

つまり、全くそれは応用がきかないということで理解いたしました。

○中村委員長 飯田さん、お願いします。

○飯田委員 企業から見たコメントをさせてもらいたいのですが、鳥澤先生から問題点として挙げていただいたその3のところは、大変そうだなと思うところがあって、有名なレシピサイトの例というのは多分すごく端的なものだと思うのですが、そうだけでなくAIを今後いろいろなものに搭載していきます。そうしたときに、AIが発明をしていく、そういうときに、権利侵害が起こったときに、このライアビリティはどうなのかというところは、やはり大きな問題ではないかと思えます。当然、メーカーとしては、お客さんのせいにするわけにはいきませんし、仮にメーカーがライアビリティをとるのだったら、相当そのAIには制限をかけていかなければいけなくなる。そうすると、AIの技術というものの進展が阻害されるのではないかと思えます。

なので、AI生成物にいろいろな権利を与えることを検討すると同時に、そのAIが生成したもののライアビリティをどうするのかというところを考えていかなければいけないのではないかと思えます。現行法のままでいけば当然規制になるので、それはメーカーとしても相当工数をかけてそのAIの生成を規制していかなければならないことになるかと思えます。それが1点です。

もう一つは、やはりデータの共有ですけれども、企業としては、競争・協調領域というものが当然あります。競争領域というところのデータは、基本的には外には出たくないと思うのが企業だと思うのです。あと、協調領域に関しても、協調するということは企業



同士の目的を持ってやっていますので、その目的を同じにしたグループ、そういう組織の中では提供はしていくのだろうと思います。

もう一つ、その競争でも協調でもない領域はあるのです。というのも、例えば、トヨタ自動車でしたら、今、車にいろいろなセンサーをつけています。そういうところにいろいろ吸い上げてくるのですけれども、これをいわゆる競争・協調領域でも使わない、いわゆる捨てるデータも間違いなくあるのです。例えば、車のワイパーが動く情報をとっていますと。これは何に使うかという、今のところトヨタでそんなものを使ってもしょうがないのです。

だけれども、ワイパーを使えば、これは雨が降っているとか、そういうデータに使えるのです。トヨタとしては、別にそんなものは全然オープンに提供していけばいいと思うのです。ただ、こういうデータは、いわゆるビジネスでは使わないので、そこにデータのメンテナンスであったり、先ほど御指摘があったように、データマネジメントという負担がかかれば、捨てるかどうかといったら多分捨てるというのがその企業の選択になってしまうのではないかと思います。

なので、企業の中にもそういう本当に提供していいデータというのはたくさんあると思うので、そういうところを、より提供しやすい、そういう構えというか、そういうマネジメントというものを国先導でやってもらえれば、よりそのデータの共有につながっていくのではないかと思います。

以上です。

○中村委員長 林さん。

○林委員 ありがとうございます。

すみません。時間の関係で、一言だけ申し上げます。

本日、この後、取りまとめの議論に入ると思うのですが、ここまでの議論の中で、データの保護対象についての話と、公開、非公開の話が少し交錯していたような感じがいたしました。また、鳥澤先生からのお話、疑問点を挙げられたような、現場でのお悩みというものもあったと思います。

この議論、今、挙げたそれぞれはある意味クライテリアの違う話ですので、今後、取りまとめるに当たっては、マトリックスのようにそれを整理して議論するとよろしいのではないかと感じました。

現場でのお悩みについては、経産省のデータ流通のプロジェクトでは、クローズドで御相談を受けてそれに専門家が回答したり、また今後、作られるガイドラインも役立つのではないかと思います。

それから、マトリックスのほうなのですけれども、対象物について、生データという言い方をしたり、有名なレシピサイトのようなウェブ由来のデータと言ったり、著作物由来のデータといったりしていますが、それぞれの範囲はベン図的にいうとだんだん狭まっていく関係にあるかと思います。対象範囲の違うものについて、公開、非公開の場合と組み

合わせると議論が非常にややこしいことになりますので、そこは切り分けて議論していったほうがよろしいかと思いました。

以上です。

○中村委員長 森さん、どうぞ。

○森委員 ありがとうございます。

1点だけ、細かいことですが、先ほど川鍋さんから国有財産法の話がありました。これは、政府のオープンデータのときに当然大々的に問題になったことでして、所管はたしか財務省だったのではないかと思いますけれども、その際に内閣官房で議論をいたしまして、基本的にはオープンデータとの関係では克服された問題です。私はどの条文をどう克服したかというのは忘れてしまったのですが、内閣官房にお尋ねいただければ、それはいけるということで大丈夫だと思います。

以上です。

○中村委員長 喜連川さん。

○喜連川委員 いろいろお話をお伺いさせていただいたのですが、全体として見ますと若干つまみ食いの見えてしまうのではないかという気がします。

例えば、関口委員のデータというのは産総研がフォルダーのデータなのです。でも、NICTが議論しているのは他人のデータなのです。あるいは、今日は余り議論がありませんでしたが、我々がAMEDなどとおつき合いをしていますと、医療系のデータは多分これと全然取り扱いが違うのです。

何を申し上げたいかといいますと、データのスペクトラムというのは、今日議論がなされているデータの空間よりもはるかに大きなところがあるというのは理解しなくては行かず、昔、グーグルが出たときは、ウェブのデータが大きいと彼らは主張したわけですが、現状、我々の側からすると極めて小さなフラクショナルでしかないと言っても過言ではないのではないかと思います。

世界でこの手のことを誰が一生懸命やっているのかという感覚からしますと、やはりイギリスがその歴史を持っているような気がしまして、たまたまですけれども、きのう、UKのデータ開発の人が来られていたのですが、相当人数をかけながら国家としてサポートしているのです。大学の中に張りつけてはいるのですが、サラリーは国から出すというぐらい非常に重いオペレーションをなされておられて、先ほどデータを出すのはいいのだけれどもその責任はとれないということもあったわけですが、全くそのとおりで、したがって、全てのデータをばかみたい丁寧にキュレーションしてみたって、何の役にも立たないわけですから、そういうめり張りも含めてどうするかというのは、データ依存の部分があります。

この辺は一番しっかり議論したのは学術会議ではないかという気がするのですが、今、このまとめからしますと、データを引っこ抜いて、中にソフトに食わせればしまいという時代の議論は、もう時代感からすると大体勝負は決まったなと。今度、我々が考える

のはもう一歩先なのです。つまり、データをビット、バイトで引っこ抜く口ということから、そうではなくて、ソフトウェアがそのデータの意味を理解しながら、つまり、データがアンダースタANDINGをしながら何に使えるかという、そういう新しいマーケット領域も含めながら考えていく時代になっているというのが世の中の広さ感かなという気がしています。

先ほどNICTの鳥澤先生が御指摘になりました研究者からデータを剥ぎ取ったら、記録がなくなってしまう。だから、これは論文も書く能力がない研究者がこうなるといけないのですけれども、そういう言い方というよりも、むしろデータをオーグメントするというのがこれからすごく重要になるのです。つまり、今のマシンラーニングといいますか、学習というのは、データを生で食わせている領域というのはもはや終わってしまっていて、なかなかレアで入らないものがどんどんアディショナルにつくりながら学習をさせていくというのがコンピテンシーになる。

そうしますと、それはその原作者の意図がそこに入っておりますので、したがって、何が言いたいかといいますと、そういう意味で、原データとそれがエクスパンドしたものは、もっともっと広いデータの世界というものをある程度整理してから議論をしたほうがよくて、ピンポイントでつまみ食いして、これで問題が解決すれば何となく心地がいいのかなというのではないのではないかとというのが、私の印象でございます。

失礼いたしました。

○瀬尾委員 いろいろ細かいことなのですが、先ほどあった飯田さんのお話でもみんなそうなのですが、企業さんのつくっているデータは、自分のために努力してつくっているデータですね。確かにデータは2つに分けるべきだと思っています。、ひとつは、いわゆる私有データ、例えば、ウェブでもそうだし、企業が持っているデータ、それに対して公的なデータ、研究機関がやったり国の機関でやってきたデータを分けないといけないと思っています。

特に著作権法の話が出て、今日、専門家の皆さんがたくさんいらっしゃるけれども、と、やはり不当に害するようなことがあってはいけないと思います。ウェブ上だから、AIの発展のためにみんな権利を捨てて犠牲になって使わせてくれという、あの鍋釜供出みたいな世界、それはやはりまずい。けれども、使い方によってはそればかりではないですね。サーチエンジンのときに、あれは誰も困らないのでアグリーされた。そういうことがあったら、著作権の制度を変えていくことについては考えていくべきだと思います。それはバランスの問題だから。ただ、個人の財産と社会の発展とのバランスなので、正直、これは時間がかかると思います。1年、2年、法改正をするのだとやはり時間がかかっていくと思う。

今、水田さんもいらっしゃるって、非常に文化庁さんは早くなさっていらっしゃるけれども、やはり2年とか3年かかっているのです。この誰かの所有しているデータをAIに使っていく部分については、やはりちょっと長いスパンで考えていくしかないでしょう。それ

で取り組んでいって、今の2つを切り分けていく。こういうビジョンをここで出すべきではないかと思う。これに対して、そうではないオフィシャルだったら早急にまとめていく施策をすること、これは所有権がないというか、そんなに主張しないのだから、国の制度として使うことができるわけですね。そこは切り分けて、ともかく使えるものを早急にふやしていくような施策と、それから長期的な視野を持ってやっていくということをしなければいけないのではないかと思います。そこは、まさに今、喜連川先生がおっしゃるように、違う意味かもしれないけれども、切り分けが必要だと思っています。

先ほど1つちらっと国会図書館さんが海外のことを言いましたけれども、日本のデータだけではなくて海外ともつながっていかないと、これはデータが非常に限定的になってしまう。今後、オープンな範囲を国内と国外にきちんと連携していく部分もつくっていかないとはいけません。

だから、前に、それこそ川鍋さんがおっしゃったけれども、日本の図書館資料は海外に送れないのですとか、海外はみんなやっているのですとかと聞いたけれども、何で送れないのですかと言ったら送れないのですと言われて終わったのですけれども、それはまずいでしょう。だから、そういう海外との連携についても国内体制を整えて、ともかくより短い時間で多く使える施策として組み合わせていくべきでしょう。、人の私有にかかわる部分のデータの利用や企業さんのデータ活用などはより長期的な施策として切り分けるしかないと思います。企業さんも多分、持っているデータを出すみたいな領域になると急にひゅっといなくなると思うのですが、一生懸命自分の会社でお金をかけてやったものがとられるかもしれないと思ったら、そこは消極的になるのも仕方ありません。だから、そのところについては、2つ議論を分けて、データセットを考えていくことが重要ななと思いました。

ただ、著作権が研究者の皆さんに委縮行為を生んでいる。これは一番最初、去年の会議で、喜連川さんも言われた、著作権はわからないのだけれども何だか使えないとかとおっしゃったのですけれども、そういう専門家でないとはわからなくて委縮効果を生むような状況というのは極めてまずいと思いますので、これについては、どういう形で解消するのかは別にしても取り組むべき課題だと感じました。

以上です。

○戸田委員 企業からのコメントということで、最後の報告書の骨子のところで言おうかと思っていたのですが、データの分類の議論されているようなので一言申し上げます。もう少しデータを分類したり、ユースケースを精緻に場合分けしたりしていく必要があると思います。

私から見ると、(1) 個人発生データというか、個人がかかわるデータ、(2) 自然発生データ、それから、(3) 企業が保有するデータ、(4) ある程度オープンなデータになるかもわかりませんが、行政が保有しているようなデータ、そういったデータの整理も有用かもしれません。

先ほど林先生が言われた2軸のマトリックスで整理するというのは、私も大賛成であります。今、申し上げたデータの種類とデータにかかわる主体を2軸にとって整理ができるのではないかと。横軸は、(1)データを発生させるのは誰なのか、(2)それを収集、蓄積するのは誰なのか、(3)それから、それをAI等で抽出して分析、加工していく人、(4)さらにそのデータの利活用をした便益を受ける人、こういった横軸とデータの種類の縦軸にとり、マトリックスで整理ができるのではないかと思います。

皆さんが議論されているように、特定の領域での議論と、全体をどうしていくのかの議論が混在している気がしますので、早急に手を打つべきところとそうではなく中長期的に議論すべきところを、明確に分けて検討したほうがいいのではないかと思います。

以上です。

○中村委員長 どうぞ、木全委員。

○木全委員 企業の話が出ましたので、私からも。

やはり同じようなところがあって、よく知財の特許などでオープン、クローズドという話があるわけなのですが、データのところもオープン、クローズドの部分がありまして、先ほどもお話がありましたように、差別化の源泉になっているようなところは出せない。逆に、一緒に組んでやったほうが利があるところは協調として組んでやりますので、そこは出さないあるいは出すということに対してはそれぞれインセンティブが働いて動いていますので、何かそここのところの境目を変えようと思うと、例えば、仲間がふえたので協調領域のほうをでかくしておいたほうが利があるからよりデータが出てくるとか、そういう周りの仕組みのほうも考えていく必要があるだろうなというのが1点でございます。

もう一点は、先ほどのデータの分類分けのことなのですが、やはり私も分類分けは賛成でして、どこかだけで議論をすると、そこで決めたことが全データに及ぶのです。そのときのデメリット側を持つデータもあるはずでして、そういう場合分けは必要なと。そのときに、ひょっとしたら、データのオーナーだとか、あるいは種類以外にもそのデータの性質みたいなものもあるかなという気もしております。多分同じ議論なのだろうと思うのですが、例えば、基本的な科学データのように、ずっと一回調べられたら普遍のものと、どこかのセンサーデータみたいに、そのデータをとっていることに意味があるのだけれども、その一瞬一瞬のデータが意味がないようなものもやはりデータなので、そのようなデータの性質に基づく分類も場合によってはあるのではないかと。

先ほど、企業的に余り価値がないのでとっているのだけれども捨てるだけみたいなものは、まさしくそういう、多分常にアップデートされていて、使い方によっては使えるのかもしれないのだけれども、メンテをするほどでもなく、そのまま垂れ流しになっているという類いのものなのではないかという気もしますので、そこをとることに意味があるものなのか、ためていくことによって価値が生まれていくものなのか、端からその価値を有しているものなのかというのは、切り分けもあるのではなかろうかと思えます。

○中村委員長 ありがとうございます。

どうぞ。

○福井委員 もう次に移らないといけない時間だとは思いますが、ちょっと最後の瞬間に、私は非常に議論が自分の中で混沌としてきてしまったので、一言、申し上げたいと思います。

最後の瞬間に、このデータはまずい、このデータはまずいという、ややいいところどりの議論が出て混沌としてきているのではないかという印象であります。

例えば、研究目的でのデータと商用目的でのデータは分けるべきではないか、公的なデータと民間が持っているデータは分けるべきではないのか。ユースケースに応じて使い分けるべきではないのか。それぞれのデータに特性があり、また、ユースケースごとに考え方を要するというのは当然のことです。それ大いに議論すべきことだと思います。

また、同時に、各企業なりが持っているデータを出さない自由があるということは、先ほど森委員もおっしゃったとおりです。これは当然であり、公開しないで抱え込んでおくことは構いません。公開を義務づけられるという法制度、この話は少なくとも本日は出ていないだろうと思います。

今、我々が少なくともかなり大きな議題として目の前に提示されているのは、47条の7という現行法の中には制限規定がありまして、ここではコンピューターでの解析目的での既存の著作物の利用はできると書いてあるのです。しかし、それは特定者間での共有ができないのではないか、また、公衆へのデータベースの提供はできないのではないかということが、大きな課題として、データ活用にとっては、今、提示されている。だから、この部分を変えていくかどうか、その必要があるのかどうかということについて、私はある方向性を出すことが、この会に少なくとも提示されているのかなと理解していました。その方向性を混沌とさせるような議論にもし聞こえてしまうと、私には異論のあるところです。

これは各企業がその持っているデータを公開すべきかどうかという話ではなくて、各企業さんなりあるいは各団体なりが持っているデータの中には、当然いろいろなデータがまざっているわけです。公的なデータ由来のものもあれば、民間データ由来のものもある。研究目的で最初は集められたものもあれば、商用目的で最初に集められたものもあるでしょう。それぞれのものがまざっている。それらの原データの中にある著作権に制限をきかせるかどうかということ、これが少なくとも今大きなテーマなのではないでしょうか。

○中村委員長 どうぞ。

○井内局長 集約の方向に向かおうと思った会で拡散してしまった感じがありますので、少し整理をしたいと思います。

いろいろなデータ、質的なデータ、公的なデータ、あるいはデータの種類によっても、あるいはオーナー、フォルダーによってもいろいろなものがあるというのが大前提の中で、最初にスコープをそれなりに私どもなりに絞って、特に産業界、企業にとっていろいろなビジネスの選択肢が広がるようなやり方をするためにどうしたらいいか、まさに今、福井先生がおっしゃったような観点で議論していただいていたという話なのですけれども、今

日はそこに公的なデータが入りましたので、さらにそこに入っている著作権の問題は入りましましたので、これは新しい文脈の話が入ってきております。

でも、最終的にまとめるときには、もちろんデータ全体の中で我々はこのスコープのところを議論しているのだというのはいち少し明確にすべきだと、それはおっしゃるとおりだと思いますが、この場で、今回のラウンドで全体を扱うということは多分難しいと思いますので、かつ、それは政府部内のほかの場でもいろいろ議論されていることがありますので、それは取り入れてやりたいと思っております。

著作権との関係は、後ほどまた著作権法の今の見直しの状況等がございますので、その中でも多分議論を続けていただければいいのかなという感じがいたします。47条の7の問題は確かにあろうかと思えますけれども、それはむしろデータの活用、利活用の選択肢をふやす方向でどうしていくのかというところをぜひ御確認いただければと思います。

○中村委員長 喜連川さん、何か。

○喜連川委員 今の議論と少しずれるのかもしれませんが、国際法に關しましての御指摘もいろいろあろうかと思うのですが、ことしの1月の年初に、ホワイトハウスはグローバル・シェアリング・オブ・リサーチデータというメッセージを出しております、若干問題なのは、トランプの就任式の前だったので、何がどうなるのかよくわからないという大きな不安があるのですけれども、そこまでグローバル化というのは来ておまして、先ほど関口委員からRDAのお話がありましたけれども、去年は、3月に東京でRDAを公開させたということもありまして、そこそこそういう意味の流れはできているというのが現状だと理解しています。

ただ、この委員会が一体何を目的にするのかということは、大分前にもいろいろ議論したかと思うのですけれども、どっちに法律を持っていったほうが日本の国益になるのかというところは理解した上で次の法施策をしましょうということだったと理解しています。

RDAをやっている、全然入らない国というのはいっぱいあります。したがって、世界で一番研究費を使っている国が入らない可能性があります。その中で日本がどういう戦略をとるかということも、一方でしっかりと考えておく必要があります、なかなかこの問題というのはそう簡単ではないというところを、大人の解釈がある段階では入るといことは共通認識としてとっておく必要があるのだろうとは思っております。

それから、民間が持っているデータというお話がありましたけれども、ここは何か議論が混沌としていると思うのですが、公のデータというのは、要するに、公的資金、つまり、税金でつくったデータとそうでないものというところで線をきちんと引いておくということが一番重要であります。

したがって、先ほどの鳥澤先生の場合は若干ややこしいのです。公的なサーバーで、人のつくったものをとってくる。だから、半分ぐらい自分のもので半分ぐらい他人のものみたいになってしまっているという、でも、こういうややこしい例はいっぱいあるので類型化する必要があるわけですけれども、公的資金でつくったデータは、原則はオープンにし

ましようというのがG8、G7の合意でありまして、したがって、完全にグローバルだという理解でいいのではないかと思っています。

ここでの議論は、実はそれよりも、今、データ利活用が余り進んでいない民をどうステイミュレートするかということのほうが国家としては大きいのではないかと思っております。そういうときに、前回、福井先生から契約法ということでおっしゃっていただいたのですけれども、それを十分小さな企業が理解できるのかという、テンプレートがあればうれしいのではないかという気がするの1つと、業界によって、データのシェアリングが非常に清廉潔白になされているところと、非常にデータを守り、絶対に出さないというところと物すごく温度差が違っています。この辺は現実を見据えながら一体どうあるべきかというのを、政府としてこういう形がということをおっしゃっていただくことは、私は相当価値があると思いますし、先ほどのワイパーの例というのは古臭い例なので全然おもしろくないのですけれども、少なくともソサエティアルベネフィットといえますか、社会に対して価値を与えるもののデータを出そうという、ある種のCSR的な部分は官としても推奨するような、そういういい枠組みがあると非常にうまく回るのではないかと思った次第です。

以上でございます。

○中村委員長 どうぞ。

○戸田委員 福井先生に誤解があるといけないと思って発言させていただきますと、喜連川先生がおっしゃったように、産業界といっても業種によってデータに関しては考え方が違うのは事実です。少なくとも私は、著作権法の文脈でいくと、柔軟な権利制限規定というのは入れてほしいという思っている立場でありまして、それが前提で議論をしてきたつもりです。

日立のような会社は、製造現場も持っていますから、データを生み出すほうもたくさんあるのですけれども、会社全体の方向としては、自分の会社の中にあるデータだけではなくて、お客様とか、社会にあるデータを利活用してビジネスを推進していこう、という立場でございます。それがSociety5.0を推進していくスタンスだと思いますので、今日の議論というのは少し混乱があったかもしれませんが、産業界としては、データ利活用を推進していこうというスタンスには変わりありません。

○中村委員長 どうもありがとうございました。

この議論で45分ぐらい押しまして、今日はあと2つテーマがあるのですが、少なくとも残りの次のテーマは非常に重要なテーマなので、それはきちんとこなしたいと思います。

次に進みたいと思います。

AI生成物とAI学習用データの著作権関連の議論、ここまでの議論でもかなり関連する議論はございましたけれども、そこに入っていきたいと思います。まず、事務局で前の会議で各委員から出していただいた意見を踏まえて、AI生成物に関する資料を準備いただきましたし、特許庁からも資料の準備をいただきました。また、AI学習用データの著作権関連



について文化庁から資料を準備していただいていますので、それぞれ順に説明をいただければと思います。

まず、事務局から。

○岸本参事官 再び資料2をごらんいただきまして、5ページ以降を簡単に御紹介させていただきたいと思います。AI生成物に関する御意見なのですが、総論のところの3つ目でございます。技術が日々進歩していて、欧米等でも議論が余り進んでいない。後ほど欧州調査報告もさせていただきますけれども、こういう現状で、特別なルールや法律を定めるよりは慎重な議論が必要であるという御意見がありました。また、AI生成物を用いたサービスに関しましては、ビジネス関連発明であれば特許が認められる。また、創作に関しましては、人間が創作にかかわっていれば、道具を使ったという評価で対応可能であるという御意見がございました。

6ページ目でございます。AI生成物が問題となる可能性ですけれども、上から2つ目あるいは4つ目あたりですけれども、他人の著作物がそのまま出力されたのであれば、依拠性ありと考えるよいのではないかと。あるいは、アクセスがあれば依拠性ありということではないかという御意見がある一方で、留意点のところですが、インターネット掲載を根拠にアクセスがあるとされてしまうと、独自創作の抗弁が機能しなくなって表現の自由空間が狭まってしまうのではないかとという御意見もございました。

7ページ目でございます。AI生成物の権利濫用の問題についてですけれども、著作物に関しましては、AIがつくったものに関して人間が創作したと僭称されてしまうという問題があるという御指摘、発明のところの2つ目ですけれども、マテリアルゲノムのようなものについて、パテントトロール的につくられてしまうという問題がある。一方で、著作権を実際に主張できる状況にはならないのではないかと、あるいは市場原理に任せるという手もあるのではないかとという御意見がございました。

次に、資料4をごらんいただきたいと思います。

2ページ目でございますけれども、かつて著作権審議会の第9小委員会で、これは平成5年に出されたものですけれども、報告書の中でコンピューターによる創作物について、人の「創作意図」及び「創作的寄与」があれば、人がコンピューターを道具として使用して創作したと評価ができると整理されたことを踏まえまして、AI生成物の分類を試みております。

上段がAIを道具として利用して人が創作するイメージでございます。左端のほうに学習済みモデル作成者がおりますけれども、この作成者が学習済みモデルをつくるまでの過程において、最終的なAI生成物の表現についての創作的な創作意図を用い、あるいは創作的な寄与をしている場合が考えられるのではないかとということでございます。

また、右側のほうに「利用者」と書いてありますけれども、この利用者が学習済みモデルに何らかの創作意図に基づく入力を行い、その結果、出てきた生成物の表現に創作的寄与があったと認められる場合も考えられるところでございます。

さらにその先ですけれども、その生成物に対して少しアレンジを加えるとか、あるいは修正を行うという行為をすれば、それもまた成果である創作物に対して、創作意図、創作的寄与があったと評価される可能性もあるわけでごさいます、こちらに書いてある、この上の段に書いてあるAI生成物と、その先に矢印でつながっている創作物が、人がAIを道具として利用した創作物と言えるのではないかと考えております。

具体的なイメージが湧かないかと思ひまして、3ページに「AI生成物の事例」ということで事例を挙げておりますけれども、例えば、その上のほうに書いてありますAIがいろいろなジャンルの曲をたくさん学習して、ビートルズ風の楽曲を出し、それを人が編曲した。そこに創作的な寄与があったという例が該当するのではないかと考えております。

2ページ目に戻っていただきまして、下の段でごさいます。こちらのほう、その学習済みモデルの作成者というのが、そのモデルの作成の過程において、最終的なアウトプットであるAI生成物の表現に関して、創作意図も創作的寄与もあったとは認められない場合ということでごさいます。右側の利用者も学習済みモデルに対して、簡単な、例えば、キーを1つ押すという操作しか行っていないような場合には、出てきたAI生成物というのが、たとえ外形上著作物と認められるような創作的な表現であったとしても、AIが自律的に生成した「AI創作物」であるということで、現行法の著作物とは言えないのではないかと整理でごさいます。

必ずしもここに該当しないかもしれないのですけれども、3ページの下の方のところを挙げております、利用者が写真とスタイル画像を選択すると、スタイル画像の画風に加工した写真が出力されるという画像の自動生成の例が、比較的人の関与の程度が少ないAI生成物と言える例なのではないかと考えております。

4ページ目を飛ばしまして、5ページ目でごさいます。先ほど創作意図と創作的寄与があれば、AIを道具として用いて人が創作したものと言えるのではないかという御説明をしたのですけれども、どの程度のものがあればそう言えるのかという点について、一つの御参考になるのではないかと考えております。

人がコンピューター創作物を「道具」として用いて著作物を創作したと認められるための要件ということで、先ほど申し上げた第9小委員会でもまとめられた報告書でごさいます。下の①のところをごらんいただきますと、創作意図というものに関しまして、具体的な結果物の態様についてあらかじめ確定的な意図を有することまでは要求されず、当初の段階では「コンピューターを使用して自らの個性の表れとみられる何らかの表現を有する結果物を作る」という程度の意図があれば足りるものとしております。

②のところ、創作的寄与につきましては、個々の事例に応じて判断するとしつつ、例えば、機械翻訳におけるアート編集、作曲における多数の計画らが選択・修正等を行うことにより、最終的なみずからの創造的個性に最も適合するものを作成していく一連の過程などが、創作的寄与であるとしております。

6ページ目、7ページ目がAIに関して残された論点について整理されたものでございま

す。1つ目は、先ほど福井先生からも御指摘がございましたけれども、AI学習用データの作成者と学習主体が異なる場合の問題でございます。共通の情報解析という目的のもとで、特定当事者間で学習用データの譲渡あるいはそのウェブやクラウド上にアップする行為による提示が行われるわけですが、それが現行著作権法の47条の7の適用によって適法と考えることができるのかという問題でございます。

譲渡につきましては、公衆譲渡に関する権利制限規定が47条の10というところがございますが、そちらにはこのようなもの場合についての規定はございませんので、公衆譲渡はできないけれども、公衆でないものへの譲渡はできるという解釈ができるのですが、この「公衆」ではないとされる範囲につきましては、下記の裁判例では、かなり限定的に解釈をされておりますので、特定当事者間でのデータのやりとりが「特定かつ少数」の者への譲渡と裁判例に照らして特定少数者への譲渡と認められる必要があるというのが一つでございます。

2つ目の矢印でございますが、学習とデータをウェブ上にアップする行為につきましては、「公衆」ではなくて特定当事者に向けた送信であれば可能と、つまり、ウェブやクラウド上に置いてあったとしても、特定少数の者がアクセス可能な状態であれば、公衆送信権侵害等はないと解されるということでございます。このような点に留意して、提示を行う必要があるということでございます。

6ページ目の下の●、枠で囲っているところでございますが、2つ目の論点として、AI生成物に対する人間の創作的寄与の程度についてでございます。先ほど第9小委員会の報告について簡単に御説明をしましたが、これをAIに照らして考えますと、学習済みモデルの作成者と利用者のそれぞれにどの程度の創作的寄与があれば人間がAIを道具として利用してつくった創作物として認められるのかということでございます。学習済みモデルの作成者というのは、利用者がどのようなデータを入力し、それによってどのようなAI生成物が出力されるかというのは、あらかじめ把握しておりませんので、そのような中で、AI生成物に対して創作的寄与が認められる場合というのは一体どういう場合が考えられるのかということがまず一つはございます。それから、利用者についても、どのようなデータを選択し、入力すれば創作的な寄与があったと評価できるのかということが問題になってくると考えております。

7ページ目でございますが、3つ目の論点。これまで余り御議論いただいてこなかった点ですが、AI生成物が問題となる、侵害物になってしまう可能性についてでございます。まず、学習用データに含まれていた著作物がAI生成物としてそのまま出力される場合、この場合に論ずるべき点としては、依拠性と責任主体という2つだろうと考えております。

依拠性につきましては、先ほどのこれまで御意見の御紹介の中でも、学習用データに含まれる著作物がそのまま出力された場合には依拠性があるという整理としてはどうかという御意見がありましたけれども、そのように考えることができるかどうか。

2つ目の責任主体についてですけれども、仮に依拠性と類似性が認められて、AI生成物の出力が権利侵害に該当するとなった場合、誰が権利侵害を行った主体と評価されるのか。利用者なのか、学習済みモデルの作成者なのかという問題でございます。利用者につきましては、単に出力しただけということであれば、著作権法の30条1項の私的複製に該当して権利侵害とはならないとも考えられるのですけれども、一方、おもしろいものがあったということで、SNSなどにアップしてしまいますと、公衆送信権侵害ということになるということでございます。学習済みモデルの作成者についてですけれども、少なくとも侵害行為を幫助したということで、出力はしていませんので、直接の侵害主体にはならないとしても、幫助したということで共同不法行為の責任を負う可能性がございますし、これまでの判例を見ますと、複製等の主体として差しとめ請求の対象となる可能性というのでも否定できないと考えられるところでございます。このような現状が学習済みモデルの作成行為というものを委縮させてしまう懸念があるのではないかとこの点について、御意見をいただきたいと考えております。

最後の論点ですけれども、AI創作物を人間の創作物として僭称する問題でございます。現行法上、AI創作物は保護されておりませんし、仮に今後保護の仕組みを考えると、保護の程度が人間の創作物に対するものよりも弱い場合、AI創作物を大量に生成して、人間の創作物と僭称して権利主張を行うトロールの人が出てくるのではないかとこの点をどう考えるかということでございます。

これらの論点について、本委員会として何らかの検討の方向性を出すのであれば、どのようにお考えになるかということについて御意見をいただきたいと考えております。

長くなりましたが、以上でございます。

○中村委員長 ありがとうございます。

続いて、特許庁の仁科企画調査官から、特許庁の取組状況について、報告をいただきたいと思っております。

○仁科企画調査官 座長、御指名をありがとうございます。

お手元に資料9があるかと思っておりますけれども、特許庁の取組状況につきまして、そちらで御説明をさせていただきたいと思っております。

スライド1に記載の検討会につきましては、昨年10月に開催しました、第1回の本検討委員会のほうで、開催の趣旨ですとかテーマにつきましては御紹介済みでございます。スライド2と3に、その趣旨ですとかテーマ、委員等について御紹介しておりますので、今日は説明を省略させていただきます。また、この検討会でございますけれども、議論を非公開で行っておりますことを補足させていただきます。

引き続きまして、スライド4をごらんください。検討会の開催状況でございますが、左側から始まりまして、第5回、昨年末でございますけれども、こちらで主要な論点及び検討の方向性につきまして、まとめさせていただいております。こちらにつきましては、後ほどスライド5と6を御参照いただきます。現在、第7回まで終了しておりまして、次回、

第8回が3月6日に開催の予定でございます。当日は、本日のテーマと関係いたしますAI創作物あるいは3Dプリンティングのデータに関しまして、その取り扱いについて議論していただく予定でございます。特許庁では、本件に関しまして委託調査研究を行っておりまして、その経過も踏まえまして議論していただく予定でございます。

スライド5と6に、先ほど御紹介しました12月末の第5回検討会で示されました主要な論点と検討の方向性をまとめさせていただいております。こちらはお時間が今日はございませんので、後ほど御参照いただく形をお願いいたします。

スライド7に移りまして、今後の検討でございますけれども、先ほど御紹介しましたとおり、3月の検討会でAI創作物に関する保護を論点としまして議論いただく予定でございます。こちらでは、AIの活用によりまして、創作に対する人間の関与の形態が変化しているという問題意識のもと、発明の保護ですとか発明者はどのように考えればよいかということにつきまして、先ほど申し上げました本年度実施の調査研究の内容を踏まえまして議論していただく予定でございます。こちらの検討会は、先ほど開催状況をごらんいただきましたとおり、4月上旬ぐらいに取りまとめを予定してございます。それ以降は、また別途審議会のほうで審議をする予定でございます。スライド7の右側でございます特許制度小委員会ですとか、あるいはその下位組織の基準に関するワーキンググループで御検討いただく予定でございます。

スライド8から、AIを活用した創作、ビジネスモデル、3Dプリンティング用データ、データ構造に関する検討状況につきまして、こちら途中経過でございますけれども、御報告させていただきます。

スライド9までめくっていただきますと、AIを活用した創作に関する論点でございますが、発明は御案内のとおり自然法則を利用した技術的思想の創作のうち高度なものと定義されておりまして、AIを活用した創作につきましては、この要件に該当するものも当然ございますので、「発明」として特許法上保護され得ると認識しているところでございます。

スライド10をごらんいただきますと、AI創作物が権利化されるまでには、スライド10に記載の図に示されておりますとおり、いろいろな人間が関与しています。こういったAIの活用によりまして、人間の創作への関与の形態がどんどん変化しているところを踏まえまして、今後、さまざまな人が関与してくる、AIを活用した発明の保護とか、その中における発明者を検討していく上での論点を整理するために、先ほど御紹介しました調査研究を行っております。AIに関しましては、AIが先ほど御紹介されましたような自律的な発明ができるかということに関しまして技術的な可能性も考慮しながら、今後、取るべき対応について、先ほど御紹介しました検討会で議論を行っていただく予定でございます。

スライド11は、AIを活用した創作に関する論点の続きでございますが、こちらの検討委員会の第3回の資料にございましたとおり、一般的には、学習済みモデルというのはAIプログラムとパラメータとの両者を合わせたものという形で定義されております。その定義に従う限りにおきましては、学習済みモデルは特許法上のプログラムに該当するため、こ

れは特許法で保護され得るという理解でございます。一方で、どのような学習モデルが特許法上の発明に該当するかについて、審査の判断手法を示すべきではないかという御意見もございますので、これにつきましては、スライド15で御紹介させていただき取り組みで判断手法を示すことを試みております。

スライド12でございますが、こちら第3回の本検討委員会で委員の方から御指摘がありました。ビジネスモデル特許に関する論点でございます。これにつきましては、新規性とか進歩性を有する学習済みモデルを、国際的な情勢も踏まえまして、適切に特許法で保護する必要があるという認識のもと、後ほど御紹介します判断事例の蓄積を行っていきたいと考えております。

スライド13の、3Dプリンティングのデータに関する論点につきましても、調査研究を行わせていただいております。スライド13に書いてございますとおり、3Dデータといいますが、単に座標データの集まりであります3Dスキャンデータから、その3Dプリンターを動かすための印刷経路の情報までを含みます加工データまで、いろいろな3Dデータがあることが調査研究の結果、わかっております。また、実際に3Dプリンターを動かすためのデータをつくるためには、非常に人の手を介す必要があるという指摘もあるところでございます。こういったところも踏まえまして、このスライドの下にございますとおり、3Dデータの取り扱いに関する論点としまして、産業財産権を有する物品が3Dデータを介して複製された場合に、これをどう保護するのか。特許法では、プログラム等を電気通信回線を通じて提供する行為につきまして、間接侵害が成立するとしておりますけれども、要は、この3Dデータが間接侵害を構成する物、プログラム等に該当するか否かというところの論点整理をさせていただいているところでございます。

スライド14を飛ばしまして、スライド15に移らせていただきます。こちらが、今、特許庁で検討を進めているものでございまして、データあるいはデータ構造、学習済みモデルにつきまして、特許権として保護されるか否かの予見性を高める観点から、審査手法を示すべきだという御指摘を受けておりますので、ちょうど今日の本検討委員会の裏番組で開催しておりますけれども、審査基準専門委員会ワーキンググループという検討体におきまして、今、申し上げましたようなデータとかデータ構造、学習済みモデルに関する発明の該当性を検討しております。そちらのワーキンググループで方向性を御承認いただければ、審査基準における事例として追加し、その充実化を図っていく方向で考えております。

特許庁からの説明は、以上でございます。

○中村委員長 ありがとうございます。

次に、文化庁の水田課長から、AI学習用データの著作権関連について説明をいただきたいと思っております。よろしく申し上げます。

○水田課長 文化庁の著作権課長をしております。水田と申します。どうぞよろしく御願いいたします。

資料8をごらんいただきたいと思っております。柔軟性のある権利制限規定の検討状況という

こととさせていただきます。

1 ページをごらんください。背景ということとさせていただきますが、文化庁の文化審議会著作権分科会におきましては、これまでも随時新しい時代に対応した制度のあり方について検討を行ってきたわけですが、さらなる変化に直面している中で、この中でも上野先生、奥邨先生にも入っていただきまして、平成27年にワーキングチームというのを審議会の法制・基本問題小委員会のもとにワーキングチームを置きまして、本年2月、実は先週24日の金曜日でございますが、中間まとめを取りまとめたところでございます。検討の背景にありますのは、この下の参考にあります2つの計画など、さまざまところで御指摘いただいているのは御承知の通りでございます。

2 ページをごらんください。

この中間まとめの概要でございますが、検討の手法といたしまして、著作物の利用のニーズを把握いたしました。文化庁においてニーズ募集を行いまして、そこで出てきましたニーズというのを①から⑥までの6つの類型に分類いたしました。これが優先的に検討すべきものということで捉えております。これらのニーズを権利制限の対象とすることの可否とあわせて、それらを含めました将来生じ得るニーズにも柔軟に対応できるような権利制限規定のあり方を検討してきたわけでございます。

②でございますが、この柔軟性の効果と影響につきまして、これを専門的に分析するために、社会調査や文献調査等の調査研究を実施して得られたデータをもとに検討を行っております。2つ目の○ですが、その結果としまして、利用の状況・場面を特定しない一般的・包括的な権利制限規定を設けることにつきましては、割合は小さいわけですが、訴訟リスクをとることに積極的な企業等については、「公正な利用」の促進効果が一定程度期待できる。その一方で、大半の企業や団体については、「公正な利用」の促進効果はそれほど期待できないという評価がございました。アンケートの結果、日本の企業については、法令遵守意識が高く、訴訟のリスクを冒して新規の事業を展開しがたいという結果が出てまいりました。一方で、現在の著作権に関する企業または個人の理解度からいたしますと、誤解に基づく侵害など「不公正な利用」を助長する可能性が高まるといったことも示唆されたわけでございます。

こういったことを踏まえまして、我が国における「柔軟性のある権利制限規定」の整備につきましては、一定の明確性と柔軟性、この適切なバランスを備えた複数の規定の組み合わせによる「多層的」な対応を行うことが適当という結論を得たところでございます。

3 ページをごらんいただけますでしょうか。

基本的な考え方を図示したものでございます。これは、横軸が権利者に及ぶ不利益、縦軸が社会的意義・公益性等ということで示しております。したがって、下のほうにありますグレーの三角形の部分は、これは社会的意義だとか公益性等と不利益を考えますと、権利制限をすべきではない、権利者の利益を不当に害する領域という部分でございます。こういった形の中で3層に分けて整理してはどうかという提言でございます。

一番左の第1層でございますが、著作物の本来的利用に該当せず、権利者の利益を通常害さないと評価できる行為類型です。本来的利用というのは、音楽を聞いて楽しむとか、絵を見て楽しむといった形での、いわゆる享受するような行為類型でございます。下の点線で囲んでありますような条文が現在の著作権法の中の条文では該当するであろうということでございます。内訳は、この後に御説明いたします。こういった類型については、行為類型を適切な範囲で抽象的に類型化を行い、柔軟性の高い規定を整備すべきということでございます。AIの深層学習などはここの中に入ってくるという理解でございます。

真ん中の第2層でございますが、これは著作物の本来的利用には該当せず、権利者に及び得る不利益が軽微な行為類型でございます。現行の規定ですと、インターネットの情報検索の規定がこちらに当たっております。また、いただいておりますニーズの中では、所在検索サービスとか情報解析サービスといったものがここに含まれてまいります。これは、1層と違ひまして、一部著作物そのものを結果として外に提供する行為が含まれてくるわけでございます。その場合の軽微の度合いというところが課題になってくるわけでございますが、これにつきましては、著作物の本来的な市場に影響を与えない、あるいは質的・量的に見て軽微だといった要件が必要となってくるのではないかといた言をいただいております。こういった部分につきましては、著作物の利用目的等によって大きくりに範囲を確定して、相当程度柔軟性のある規定を整備するという提言でございます。

第3層につきましては、公益的政策実現のために著作物の利用の促進が期待される行為類型ということでございます。下にありますように、教育とか障害者関係、報道関係ということですが、公益性が非常に高い部分ではございますが、その際、検討するに当たっては、非常に著作物全体を使う場合がございます。こういったところについては、利用目的ごとに公益性や権利者の利益等の調整に関する政治的判断が必要ということで、権利制限の範囲を画定した上で、それぞれの範囲ごとに適切な柔軟性を備えた規定を整備すべきだということで、同じ柔軟性の中でもこういった3層に分けて、それぞれ柔軟性の程度は違ってくるだろうという方向でございます。

次の4ページをごらんいただきますと、特に細かい点も関係してまいります。

第1層、著作物の本来的な理由には該当せず、権利者の利益を通常害さないと評価できる行為類型についてということで、少しまとめております。最初に、この考え方でございますけれども、以下の①、②、③のような行為は、著作権法の目的に照らせば、権利者の利益を通常害さないものと評価できるということでございます。通常害さないということです。ここの中にありますように、不当に害する場合もあり得るということでございますが、まず、この1つ目ですけれども、著作物の表現の知覚を伴わない利用行為ということで、情報通信設備のバックエンドで行われる著作物の蓄積等でございますが、赤で書いてありますように、こちらが特にAIの深層学習などにかかわる部分かと見ております。②としましては、著作物の表現の知覚を伴うが、利用目的・態様に照らして当該著作物の表現の享受に向けられたものと評価できない行為、例えば、これは技術開発の試験の用に



供するための著作物の利用といった例がございます。3つ目が、著作物の知覚を伴うけれども、情報処理や情報通信の円滑化・効率化等のために行われる利用行為であって、独立した経済的な重要性を有さないものということで、キャッシングですとかミラーリングといったものが例として挙げられております。今、申し上げましたような行為、この第1層に当たる行為類型は可能な限り幅広く権利制限の対象となるように抽象的に類型化を行いまして、柔軟性の高い権利制限を整備することが適当という結論をいただいております。

AI開発に関する部分を抜粋しておりますので、そちらについても御説明いたします。2ポツのところでございますが、この第1層に属する行為類型のうち相当程度のものは近年の累次の法改正によって既に権利制限の対象となっていると考えられるわけですが、技術の進展に伴いまして、現行規定に定める利用行為に類するものであるものの、現行規定の対象範囲から外れるおそれのある行為が新たに生じてきているという指摘がなされているということで、記述をいただいております。これは、現行規定の要件の一部に立法時に把握されたニーズの内容や技術仕様を前提として設定されたものがあるために、既存の規定と同様の趣旨が妥当する行為であるにもかかわらず、その行為が当該要件のために権利制限の対象から外れてしまうといった事態が生じたということでございます。この※にありますように、先ほど来出ております第47条の7につきましては、「統計的」という要件がAIによる深層学習に対応できていないのではないかとといった指摘や、複数の主体が協業で情報解析用データベースの作成と情報解析とを分担して行う場合に、権利制限が適用されないとの疑義があるといった指摘があります。これらの行為については、権利者の利益を害するものではないことから、権利制限の対象となるべき行為である旨の意見が示されたということで、なお、後者の指摘については現行法の解釈によっても対応可能であるとの意見もあったということが中間まとめに記述されております。

「この点」ということで最後でございますが、権利者の対価回収の機会を損なわない著作物の利用行為については、著作権法の目的に照らせば権利者の利益を通常害さないと考えられることから、現行規定と同様の趣旨が妥当する行為であれば、同様に権利制限の対象とすることが適当であるということで、この第1層に係る制度整備に当たり、現行規定についても、必要以上に個別具体的な形または特定の技術に偏った形で厳格な要件がなされていないか、それを確認した上で、技術的中立性にも配慮しながら必要な手当てを行うべきであると考えているということでまとめていただいております。

先週金曜日にこの中間まとめがまとまりまして、ちょうど本日から、このまとめにつきましてはパブリックコメントにかけております。3月29日までの予定でパブリックコメントをかけまして、その後、それらを踏まえまして最終的にまとめていきたいと考えております。

以上でございます。

○中村委員長 ありがとうございます。

それでは、委員から意見、コメント、質問をいただきたいと思っております。いかがでしょうか

か。

奥郵委員。

○奥郵委員 まず、1点目は特許庁さんに質問が1つと、2つ目は事務局に質問ですけれども、まず、1点目、特許庁さんの資料の11ページのところで、学習済みモデルはプログラム等に該当するため、産業財産権法上で保護され得ることになっています。これは確認だけなのですけれども、プログラムを抜きにした裸のパラメータだけはここでの検討の対象になっていないというか、ここで言う結論の「保護され得る」の中に入っていないという理解でよろしいでしょうか。

○仁科企画調査官 御質問ありがとうございます。

先生の御指摘のとおりでございまして、裸のパラメータといいますか、パラメータ自体につきましては、特許法の保護の対象にはなっていないとは考えております。

○奥郵委員 ありがとうございます。

あとは事務局のほうに質問なのですけれども、まず、スライド6ページのところで、下のほうに「AI生成物に対する人間の創作的寄与の程度について」との項目があります質問は、これは何のために聞いているのですかということです。ここでの問題意識は、すなわち、私は第1類型、第2類型と言っていますけれども、人間が道具としてAIを使ってつくった著作物と、そうではなくてAIが完全に自分でつくったものと、第1類型と第2類型と分けた場合に、どちらが第1類型になるのですか、第2類型になるのですかという議論のためにここで問題提起をされているのでしょうか。それとも、誰が著作者になるのかということの含みで聞いているのでしょうかということです。

というのは、今、第9小委員会の資料を引っ張っておられるのは、第1類型ですか、第2類型ですかということだけのための分析の部分だけが引かれています。第9小委員会では、第1類型の場合に誰が創作者になりますかと、そこから先もまだあります。それについての答えはまた第9小委員会として出しているのですけれども、そこは引用されていないのです。なので、どこまでを対象として議論されているかによっても、ここの考え方は変わってくると思いますので、そこは事務局の今回の狙いがどこなのかなというのがありません。

2点目の御質問は、7ページの「学習用データに含まれる著作物が」と書いてあるのですが、これは何をイメージしているのでしょうか。すなわち、学習用データの中の著作物がそのままの形でAIに収蔵されている場合をイメージしているのでしょうか。というのは、これはいろいろな場合があると思います。学習用データがまさに要素にばらばらに分解されて、人工知能中には、著作物としては残っていない。著作物性のある状態では残っていない。何らかの数字の特徴量だけがある。もちろんそれを著作物として見るのかどうかという議論があるわけです。そうではなしに、例えば、長い文章が3フレーズ、4フレーズぐらいにぶった切られて存在するのではこれは大分議論が違うのですけれども、これは何をイメージしてどういう場合を議論しようということなのか、それとも、その辺はまさに

これから分けて考えてほしいということなのか、その辺によっても大分検討が違ってくると思いますので、お尋ねした次第です。

以上です。

○岸本参事官 まず、1点目の何のために2点目の論点を議論しているのかということなのですが、先ほどの2ページ目のところで「AI生成物」の分類イメージということで、第1類型、第2類型を提出させていただいたのですが、第1類型のAIを道具として人が創作するイメージのところでも、結局、創作意図、創作的な寄与が具体的にあったと言えるような具体例というのが、先ほども御紹介しましたビートルズ風の楽曲を出力したようなところでして、最終的に人が明白に手を加えているような例しか思いつかないわけですし、その学習済みモデル作成者と利用者というのが、学習済みモデルもつくる過程あるいは学習済みモデルに入力を行う段階でどういう創作的な寄与をすれば、その後の編集とか一切なしに人が創作したと言える場合があるのかという点を、事務局としてもイメージがつかめなかったものですから、こういう論点も提示させていただいております。

2つ目の点ですけれども、学習用データに含まれる著作物というのは、そのままの著作物として独立して評価できるようなものが入力というか、学習されるというイメージで書いております。

○奥邮委員 違います。学習されるのではなしに、学習されたデータがどうやって保持されているかということについては、どういうイメージですか。そのまま残っているのですか。それとも、ばらばらになって著作物としては残っていないのですか。そこについては分けて考えていますかという議論です。

○岸本参事官 そのままの形で、AI生成物の中の一つの表現として出てきているというイメージです。

○奥邮委員 すみません。出てきているものではなくて、学習用データをどうやって持っていますかというのももう一個重要なのですけれども、そこは考えていないということですね。

○岸本参事官 はい。

○奥邮委員 わかりました。結構です。

○中村委員長 瀬尾さん、どうぞ。

○瀬尾委員 権利者団体にいろいろ顔を出しているのですが、今日言うと怒られるかもしれないのですが、基本的にこれはこの前からずっと、どうやって人が関与していて、どのようなツールであるのか、それともそのままAIがつくったのかということについて議論がいろいろあるかと思います。ただこの前から余りそれが関係ないのではないかという気がしてきました。つまり、AIがつくろうが、人がつくろうが、出てきたものが特定の著作物に対して非常に酷似しているもしくはそのままであったら著作権侵害になるし、もし人がAIでつくったものを自分の作品だと言った場合、その関与の度合いによってその著作物を分けることは多分不可能になってくるだろうということから、これまでの著作権ルール

をそのまま当てはめて検討すれば済むのではないかという気がしてきました。

ただ1つ言えることは、先ほどの議論にあった、どういう著作物をデータとして取り込めるかどうかは議論が別です。ただ、出てきたものに関しては、機械がつくろうが、人がつくろうが、基本的にAIがつくったと言えれば保護はなくなりますけれども、僭称した場合、もしこれは俺がつくったのだと言ったら言ったでいいではないか。それに対して警戒する必要はなくて、つまり、自分がコンピューターを使ってほとんど何もしないでできたものでも、すごくコンピューターを駆使して使ったものにしても、例えば、私たちが写真をつくるときにフォトショップを駆使して使ったものにしても、その関与の度合いは余り関係なくなってくるから、そこを議論するのは無駄ではないかという意味です。

もう一つ、例えば、前にあった議論で、これは私が言ったのだと思うのだけれども、どんどんコンピューターがいつぱいつくったものを人がつくったと言って強力な著作権で保護してしまうというのはどうなのかというのですけれども、よく考えると、そういう著作物がちまたにありふれると、価格は強烈に下落します。その中で、要するに、自分が著作物であると偽る意味も、そういうインセンティブがなくなるくらい落ちていく。だから、そういった市場の原則によって、どんどん無理して著作物であると自分が偽る必要もなくなってくると思いました。

もう一つは、今の著作物の利用例からして、どんどん利用がふえると、1対1の対価ではなくて、いわゆるまとまって月課金とかサブスクリプションモデルみたいな形になっていくとすると、利用の普及とでき上がったもののクオリティーのバランスに任せればよいことで、特別な法制度の手当ては余り必要ないのではないかという気がしました。

ただし、これによって私が現実に考えたのは、そうすると、ある程度どんどんクオリティーの高い作品が世の中にどんどん出てきてしまうので、現在のクリエイターの中で、よほどバリューのあるものをしないと食えなくなるというのはあります。

ただ、これは制度の問題ではなくて、技術の進歩によって食えなくなる商売はたくさんありますので、これについてはクリエイター側がAIに負けないものをつくっていく以外に手がないし、これによって進歩をとめることは好ましくないと思います。現在のAIシステムによってできた著作物というのは、たとえ関与の度合いがないものを自分の著作物だとして言った場合にしても、そうではなく本当にそれを駆使して使った場合にしても、もし本人が著作物だと言った場合には、著作物として認めざるを得ないのかなと。

AI学習セットをつくるためにも選択とか工夫が大変必要だとすると、そもそもAIをつくっていく時点でもそういう工夫が働くとすると、余りその点を考えているのは無駄なのではないかというのが、最近考えていることです。ですので、放置しておいたほうが発展を妨げないし、クリエイターの自助努力も高めるのではないかと思います。成果物に関しては、前にお話ししたことと違うことを思っているので、申し上げます。

○中村委員長 上野委員、お願いします。

○上野委員 質問というよりは、3点コメントをさせていただきます。

1つ目に、今回の討議用資料におきましては、AI生成物のイメージが2つに分けられております。すなわち、「AIを道具として利用した創作」と「AI創作物」というものであります。これは区別として重要なものだと思います。その上で、この資料では、区別のメルクマールといたしまして、かつての第9小委員会の考え方をヒントにして、「創作意図」と「創作的寄与」という観点を持ち出されて、これを基準に区別されているものと理解いたしました。

ただ、今回の資料では、その「創作意図」と「創作的寄与」が、学習済みモデルの利用者のみならず、学習済みモデルの作成者のところでも問題にされており、これが特徴的なように思います。これに対して、第9小委員会の報告書では、コンピューターシステムを使用する者に意図と寄与があるかどうかを問題にしているように思いますので、この図でいいますと右側といいたしでしょうか、学習済みモデルの利用者のところで「創作意図」と「創作的寄与」があるかどうかを問題にしているように思います。

そうすると、例えば、先ほど御紹介があったように、AIがビートルズ風の曲をつくったという実例がありまして、あの場合、確かに、編曲や作詞の部分は人間による新たな創作が行われているようですので、そこは確かに人が創作したと評価できるわけですが、楽曲の部分に関しましては、スタイルと曲の長さしか指定していないということのようですから、そうだとすれば、討議用資料にいう「簡単な指示」だけで楽曲が生成されたといえるように思います。そうすると、楽曲の部分は「AI創作物」と評価してよいのではないかと私は思います。

2点目は、「公衆」に関してであります。今回の資料でもスライド6におきまして、「特定当事者間での学習用データの譲渡」が公衆への譲渡に当たってしまうのではないかとということが論点になっております。ただ、難しいのは、この「特定」という言葉の意味であります。確かに、「特定」という言葉からいたしますと、一定の決まった企業グループの関係にあれば「特定」に当たるように聞こえるわけですが、著作権法上の「特定」という文言は、この言葉が持っている本来の意味を失っているように私は思います。実際、スライド6でも、「『特定』について『人的な結合関係が存在することを意味する』と狭く解釈する裁判例もある」とありますけれども、このような解釈はここに挙げられた裁判例だけが述べているというより、むしろ一般的な解釈というべきものです。例えば、加戸守行の逐条解説（六訂新版）におきましても、「『特定』といえますのは、行為者との間に個人的な結合関係があるものを指します」（73頁）と書かれています。これによれば、「特定」については、資料にいう「狭く」読む解釈が一般的なものというべきではないかと思えます。

3点目は、先ほど奥邨先生からも御指摘がありましたけれども、最後のスライドでありまして、「学習用データに含まれる著作物がそのまま出力された場合には依拠性があると考えられないか」という点です。これは、確かに先ほどの御指摘にもありましたように、もとの著作物がそのまま学習済みモデルの中に入っている場合と、そのままではなく、抽

象化されてあるいは断片として著作物性のない形で学習済みモデルの中に入っている場合の両方があると思うのですが、前者については依拠性があることは争えないと思います。他方、後者については議論があり得るところかと思えますけれども、私自身はこの場合にも基本的に依拠性は肯定されるのではないかと考えております。

以上です。

○中村委員長 福井さん。

○福井委員 時間はないのですが、余りにおもしろくなってしまったので、少し発言させてください。

事務局資料4の3のところで、瀬尾さんはAI生成物について著作物性を独自に人間の生成物と区別して議論するメリットに疑問を呈していきまして、私はこれはあると思うのです。従来から申し上げていることですが、知的財産権の付与というのは、そのような独占権を付与する必要があること、そして、その付与による社会的なデメリットが付与のメリットをよりも下回っていること、メリットがデメリットを超えていること、こういうことを超えて初めて許容されるものだと思います。

その点でいうと、瀬尾委員がおっしゃったみたいに、コンテンツが大量にあらわれて価格下落が生ずるといふならば、通常は必要性というのは低下しております。それに対して、デメリットは何かという、それほど大量のものに刑事罰まで生ずるような知的財産権が与えられてしまえば、新たな人々の活動というのは委縮する可能性がある。さらに、これをトロールというふうに使われるおそれもここでは指摘されているところであります。

必要性がより低下し、逆にデメリットのほうも無視できないとなれば、ここはやはりAI生成物の著作物性を独自に議論するメリットは大いにあるだろうと思うわけでありまして。その点において、私は人間の寄与などの要素によってこれを区別しようという第9小委の方向性は基本的に支持できるものだと思っています。そして、事務局資料4の3にある具体例を出して、これによって、人の寄与がどの程度だったら、AI生成物とは言えないのかな、著作物に当たると、どの程度だったら当たらないと言えるのかという整理は、私は大変よろしいように思います。

上野先生の御指摘は全くごもっともではありますが、上のほうのビートルズ風の曲に関しては、人間の寄与もその後あったという点において、これは著作物とする結論には賛成です。さらに下のほう、人の関与の程度が少ないと考えられる、いわゆるマッシュアップの事例は、やはりこの程度ではなかなか認めがたいのではないかと感じるわけでありまして。

もう一点だけ申し上げます。事務局資料4の7で、依拠性ということが議論されています。学習用データに含まれる著作物がそのまま出力された場合には、依拠性があると。もしこれがそのまま出力されたときにしか依拠性がないという意味合いであるならば、私は反対であります。本来は、何がそのまま出力されているかというのは類似性で議論すべき問題であって、依拠性というのはそこに至る過程で原著作物に一体どのぐらい寄りかかっているかという論点であります。その意味で言うと、従来の判例の基準どおり、元の著

作物にアクセスがあって、それを知っていれば、依拠はあったと考えてもいいのではないかと思うわけであります。なぜか。

人間であってもある作品を知ったときに、それが自分の頭の中でどうプロセスされて最終的に成果物としてあらわれてきているかというのは、ブラックボックスでよくわからないわけであります。AIもこのところよくわからない。だから、そのところに法の議論は余り立ち入るべきではないと思うわけであります。もとの作品を知っていれば、アクセスがあれば、依拠という点ではもう解決したと考えて議論を単純化してしまってもいいのではないか。そして、その後で事務局が整理されている責任論というところにこの議論を集約させて、議論をシンプルにしていくほうがいいのではないかと思うわけであります。

以上でした。

○中村委員長 ありがとうございます。

皆さんにいただいた時間がここで来ましたが、少し延長をさせていただいて、発言をしたいという方はここで立てていただいた方にしたいと思います。

瀬尾さん、奥邨さん、今枝さん、戸田さんの順でお願いします。どうぞ。

○瀬尾委員 ごめんなさい。誤解があるかもしれないので、一言だけ。何もAI著作物に全部著作権を付与したほうが良いとは全く思っていません。AI著作物は著作権が基本的にはないものであるけれども、それをそういうふう言い立てる人が出てくることについて排除できないだろうと思われるので、その部分についての議論というのは余り効果がないのではないかと申し上げただけなので、決して積極的に著作権を付与したほうが良いとは全く言っていません。

○中村委員長 奥邨委員、どうぞ。

○奥邨委員 1点だけ、先ほど事務局が質問した件との関係で一言だけ補足します。

せっかく福井先生が整理したほうが良いのにとおっしゃったのと逆方向に行くのであれなのですけれども、基本的な考え方として、論理的に詰めていくと、従来の依拠性という言葉が持っていた広さから言えば、こうなるのだろうなどは思います。ただ、この解釈をしてしまうと、AIというのをいわばコピー機みたいなものと捉えることになりますので、本当にそれでいいのかなというのは場面場面でいろいろ考えていく必要が本来はあるのではないかと考えています。

あと2つあって、1つは、リバースエンジニアリング等々との整合性、説明性がつくのかどうかということも整理をする必要があるだろうと思います。アイデアは使っているけれども、表現は使っていない。でも、結果的に同じものになったときにどうするかということどうまく説明がつくのかどうか。

3点目は、先ほど人間の頭の中はブラックボックスだからというお話がありましたが、まさに私はそのとおりだと思っていて、従来の依拠性の議論というのは、人間というのはよくわからないのでそういう議論をしていたのですけれども、AIというのは、もちろんディープラーニングに難しいところがあったとしても、しょせんは機械ですから、人間より

は見える化が可能だということになると思うのです。とすると、ある程度、本当に表現は使っていないのです、アイデアしか使っていないのですということを技術的に証明できた場合も、従来型のアクセス云々という議論をしていいのかどうか、それと同じなのかなと、私も技術的なことも含めてまだ整理はついていないのですけれども、気にはなっておりますので、先ほどそういう質問をさせていただいたということです。ただ、まだ難しい問題ですし、いろいろな御議論があるでしょうから、別に私のこの意見でいかないとだめということを行っているわけでは全くありませんので、先生方の御意見を聞きながら整理していきたいと思っています。

以上です。

○中村委員長 今枝委員、お願いします。

○今枝委員 私は、先ほどの特許庁さんのお答えが何となく気になったので、1つだけ。

学習済みモデルは、AIプログラム+パラメータで、パラメータオンリーのことでは保護しないのですよねというのをイエスと答えられました。私もそのとおりだと思います。

ただ、この場合の学習済みモデルというのは、そもそもパラメータと分けることができないで、一体として既に一つのプログラムとして構成されていて、定義されるかもしれないけれども、結局は分けて保護するとか、分離したら、AIプログラム、つまり、カーネルになってしまうということなので、意味はない話になるのではないかと思いますのですけれども、そういう理解でいいですか、特許庁さん。ここは分けることはできないという理解でいるのですけれども。

○仁科企画調査官 技術的に意味のある形で分けられるか分けられないかということは、正直正確にはわからないところもあるのですが、私から提示しましたスライド11をごらんいただきまして、ニューラルネットワークの図を中央に描かせていただいておりますけれども、ここに書いてあります図でいいますと、「AIプログラム+パラメータ」というところの「パラメータ」に相当するものは、このニューラルネットワークの図に重みづけ係数 $W_1$ 、 $W_2$ 、 $W_3$ と書いてあるものになっています。この数字だけを切り出せるか切り出せないかと言われれば切り出せるかと思うのですけれども、ただ、この数字だけを切り出したとしても、ニューラルネットワークの構造と対にならないと意味がないので、その意味では、切り出せないというように思っています。

○今枝委員 多分、奥邨先生の質問のほうが、ここのパラメータだけを保護すべきかすべきでないかという意味で聞かれていたと思うのです。

そもそもパラメータを切り出すというのが、要するに、例えば、ディープラーニングをした後に、そのパラメータを切り出す行為をする理由もなければ、必要性もなければ、何もないと思うのです。定義としてはわかるのですが、パラメータをここで出す理由すらないのではないかと考えているのです。

○仁科企画調査官 一方で、こちらの検討委員会の過去の議論の中でも、このパラメータ自体を保護したらどうかという論点が挙がったと認識しておりますので、論点の一つであ



るのかなという理解でおります。

○今枝委員 わかりました。

○中村委員長 それでは、戸田委員、お願いします。

○戸田委員 私も、特許の件に関して、2つほどコメントをさせていただきます。

1つは、学習済みモデルをデータ構造で保護できるのではないかという点と、ビジネス関連発明の活用という点でございます。私が知る限り、この2つは、日本特許法特有といえますか、日本の特許法の保護が進んでいる分野だと思います。ですから、国際的な動向を踏まえて、これを御検討いただきたいというのが言いたいことであります。こういった権利は、正直申し上げまして、侵害立証が必ずしも容易ではなく、かつ、権利も安定しないものではないかと思えます。審査基準とかガイドラインを策定されるというのは賛成なのですけれども、特許出願を推奨してどんどん権利化していけばバラ色の未来があるということにはならないのではないかと考えています。国際的な動向も踏まえまして、ガイドラインの作成とか、皆様に向けてのアナウンスをやっていただきたいと思えます。

以上です。

○中村委員長 どうもありがとうございました。

時間が来てしまっているのですけれども、次回には取りまとめをしなければいけません。

そこで、恐縮なのですが、残りの一つに報告書の方向性骨子というものと、一つ的前提としてヨーロッパの調査というものをさせていただきましたので、その説明だけをちょっとさせていただいて、次回に持ち越すことにさせていただければと思います。

事務局から説明をお願いします。

○岸本参事官 資料5をごらんいただきたいと思えます。

これまでの御議論を踏まえまして、事務局として骨子案ということでまとめたものでございます。

初めに、基本的視点の後に、データ利活用促進のための知財制度のあり方ということで、現状と課題、論点としては、今日も参考資料4ということでおつけしておりますけれども、第5回の会合で論点整理としてお示しした論点、3ポツの方向性のところにつきましても、基本的には第5回の会合でお示しした考え方をベースにまとめていくということ、本日の御議論も加味しながらまとめていくということを考えております。

2ページ目、3ページ目は、AIの作成・利活用促進のための知財制度のあり方ということで、同様に第2回、第3回会合を踏まえまして、第3回の会合で本日参考資料4でおつけしておりますけれども、こちらでお示しした論点整理でお示しした考え方をもとに記述していくことを考えております。方向性についても、本日の議論をしっかりと踏まえまして、まとめていきたいと考えております。

簡単ですが、以上です。

○中村委員長 ありがとうございます。

ヨーロッパの調査のほうも、松村さんからお願いできますか。

○松村参事官補佐 知財事務局の松村です。

本検討委員会において、海外の議論の動向を調査すべきとの御意見を頂戴いたしまして、また、欧州委員会が本年1月に「欧州のデータエコノミーの構築」と題する書面を公表し、その問題意識はデータの自由流通を実現することによって社会・経済の発展を図るべく、データをめぐるさまざまな課題について利害関係で対話を行っていくことという趣旨の紙を出しまして、まさに我々の議論の方向性と同様のものでもあると思いましたので、こちらにヒアリングの意見交換を行いにまいりました。また、渡部委員長、上野委員の御紹介によりまして、マックスプランク研究所及びミュンヘン大学有識者4名の方と、このデータ、AIに関する議論の状況の意見交換を行ってまいりました。

まさにこのデータに関しては、このペーパーでも、まず、機械から得られる生データで、特に知財権の対象になっていないものについて適切なアクセスを確保し、利活用を促進するためにどのような方法が考えられるのかというところで、データのガイドラインですとか、データの契約のモデルをつくるのか、あとはデータ生成権みたいなものを付与してみるとか、さまざまなメニューが与えられていて、まさにこれは我々がそのデータの議論の際に提示したようなメニューと同様のものをごさしまして、問題意識は共通しているというのを確認してまいりました。

ただ、これから彼らはまずはヒアリングを行い、意見募集を4月まで行っていて、日本の企業にも積極的に意見提出をしてほしいということをおっしゃっておりますが、その意見募集を踏まえて、事例の調査を行い、これから方向性を決めていくという状況だと伺いました。なので、具体的に今後どうしていくかというよりも、これからまさに議論を詰めていくのだというところを伺ってまいりました。

特記すべきところは1つで、データベースに関するSui generis rightというものは、まさに今、見直しの段階に入っていることがわかりました。まさにIoT時代において、96年につくられたデータベースに関するSui generis rightがまさに機能するのかどうか。機能しないのであれば、ある種、改修を含めて議論をしていくというところで、ことしそういう議論の見直しを開始すると申しておりました。

また、有識者の御意見は資料6の右側に書いてございますけれども、データに関してはビジネスモデルが確立されていない中で知財権を付与するというのが、データの流通を阻害するので避けたほうがいいのかと、営業の秘密や契約での対応も可能であるみたいなことが言われていました。

また、Sui generis rightについても、なかなかデータベースを保護するものだと規定されているのだけれども、データとデータベースの区別が難しく、何が侵害になっているかわからず使いにくかったという評価をされていました。価値あるデータの不正取得について、欧州の国では、一般不正競争防止法に関して一般規定を設けているところもございまして、そういうものは日本にはないということを説明すると、行為規制アプローチというのも少しは考慮してもいいのではないかという御意見をいただきました。

AIに関しては、特に知財に関する議論はほとんど行われていないというところを確認いたしましたので、今後、我々の議論も含めて海外に発信しながら、引き続き情報交換を行っていったらと思っております。

簡単ですが、以上になります。

○中村委員長 どうもありがとうございました。

今日は、非常に多くの関係者の方にオールスターでお越しいただきまして、突っ込んだ議論をしていただきました。

最初のデータオープン化については、法制度をどうするかあるいはインセンティブをどう見るのか、海外との関係をどうするのかという御議論があった中で、今、ここで出せるアウトプットは何なのか、引き続き議論していくことは何なのかということを決めなければいけない。

もう一つのAI生成物についても、制度として今回どう捉えるのかということが、もう一度議論が戻ったという面もあるように感じていますけれども、そろそろこれも決める時期に来ております。

アウトプットは非常に悩ましいところではありますし、先ほど事務局に説明をいただいた報告書骨子の案も、今日の議論を踏まえるとこの時点で直さなければいけないところもありそうな気配でして、今回の結論に向けて、皆さんとの調整を深めたいと思います。

議論はあと1回で決着がつくかなと思っていたのですが、どうも難しそうだなというのが、今日の結論かなと思います。井内局長、いかがでしょうか。閉会の御挨拶を。

○井内局長 本当に、非常に貴重な意見ばかりいただきました。ありがとうございました。

私も100%の自信はもうなくなっていて、いろいろ文案はあるかなと思いますけれども、最大限努力したいと思います。

今日は非常に大きなスコープの御議論を特に前半でいただきましたので、それをどう反映させていくのか。そして、短期、中期に論点を分けて、何とか整理を試みたいと思います。

それから、先ほど欧州の報告がございましたけれども、欧州で聞いても、アメリカでは、政策当局、政策論としては、データについてもAIについても特段聞こえてこないということのございますし、個別にいろいろ米系企業の方に伺っても、どうもそのような感じのございますので、アメリカについては、学者の方々の議論はともかくとして、政策当局として特に何も無いのかなと理解をしております。

いずれにいたしましても、次回に向けまして、また個別にもいろいろと御意見を伺えると思いますので、よろしく願いいたします。

○中村委員長 事務局から、今後の予定をお願いします。

○岸本参事官 次回会合ですけれども、3月13日、月曜日、10時から12時までの予定となっております。よろしく願いいたします。

○中村委員長 そうということで、以上で閉会といたします。

どうもありがとうございました。長時間、本当に失礼いたしました。