

民事判決情報データベース化検討会

第5回会議議事録

- 第1 日時 令和5年2月22日(水) 自 午後3時
至 午後5時30分
- 第2 場所 オンライン開催
- 第3 議事
- 1 開会
 - 2 民事判決情報の利活用等について
 - 3 民事判決情報データベース運用上の課題について
 - 4 日本弁護士連合会による海外調査の追加報告
 - 5 次回以降の議事、日時等の説明
 - 6 閉会

議 事

山本座長：

それでは、定刻となりましたので、民事判決情報データベース化検討会第5回会議を開会したいと思います。

本日も御多用の中、御出席を賜りまして誠にありがとうございます。

なお、本日は所用のため、増田委員が午後3時15分頃から参加、楢松委員が午後4時頃から参加、また小町谷委員が午後5時頃退出の御予定と伺っております。

それでは、本日の審議に入ります前に、配布資料等について事務局から説明をお願いいたします。

事務局：

渡邊です。資料の確認をさせていただきます。資料1ですが、本日予定の有識者ヒアリング等の概要を記載したものになります。資料2から4は、本日お話を伺います皆様から御提出いただいた資料でございます。最後に会議用資料として次回以降の日程等について記載したものを配布しております。資料の確認は以上となります。

山本座長：

ありがとうございました。それでは早速、本日の議事に入りたいと思います。本日は何人かの方々からお話をお伺いすることとありますが、まずは報道機関の立場からみた民事判決情報の利活用の可能性等についてヒアリングを実施したいと思います。御協力をいただく有識者の御紹介等を事務局からお願いいたします。

事務局：

渡邊です。本日は読売新聞東京本社論説副委員長の富所浩介様にお越しいただきました。民事判決情報の利活用の可能性につきましては、これまでも有識者ヒアリングで示されてきたところがございますが、富所様には資料1の第1のとおり、報道機関の立場からみた民事判決情報の利活用の可能性についてお話しいただくとともに、今後の議論の参考として、報道機関における訴訟関係人の権利利益への配慮の現状についてお話しただきたいと考えております。御発表いただきました後、質疑応答にも御対応いただく予定でございますけれども、所用のため午後3時40分頃には御退出の予定と伺っておりますので、委員の皆様におかれましてはその点を含みおきのほどよろしくをお願いいたします。以上でございます。

山本座長：

ありがとうございました。それでは富所様よろしくをお願いいたします。

富所様（読売新聞）：

皆さんどうも初めまして。読売新聞社の富所と申します。本日はよろしく申し上げます。簡単に略歴を紹介させていただきたいと思うのですが、これまで主に司法や教育分野の取材を担当してまいりました。現在は論説委員会というところで、社説やコラ

ムの執筆等に携わっております。本日は民事判決データベースの活用や課題等について、裁判報道に関わってきた立場から意見を申し上げたいと思っております。よろしくお願いいたします。

私も記者になって30年以上になるものですから、裁判の取材も大分変わってきているのではないかと思います。先日現場の記者の方に状況を確認しました。そうしましたら、デジタル化がいろいろ進んだ面などはあるのですけれども、基本はそんなに変わっていないという印象です。まずはそうした現状からお伝えできればと思っております。

資料を見ていただきたいと思いますが、最初の1ですけれども、裁判の報道は大きく分けるとこうした4点に集約できるかと思っております。まず一つ目は、裁判そのものの取材です。実際に民事裁判の法廷に足を運んで、原告・被告双方の主張を聞いたり、訴訟の進行状況を確認したり、それから判決の内容を聞いたりというような作業をしています。今回関係してくるのは2番と3番の部分になると思うのですけれども、二つ目は、裁判所のデータベースや民間のデータベースなども使っていますので、こうしたものを活用しながら訴訟の意義や同様の訴訟の存在等を確認しています。3番目は作業としては一緒なのですけれども、こうしたデータベース以外に紙媒体の判例集であるとか論文等も読んで取材の参考にしているというところです。最後の4番目は、訴訟の当事者あるいは代理人弁護士取材ということになります。当事者・代理人の方が記者会見を開くケース等も結構ありますので、そこで判決の内容であったり、意義、それから判決に対する見解等を伺います。こうした取材を通じて記事が出来上がっていくというところになります。今申し上げた2番と3番の、データベースであるとか判例集ですが、これはもちろん全部載っているわけではありませんので、どちらにも該当がない裁判みたいなものがままあります。ですので、民事訴訟の全件の判決が閲覧可能になるデータベースができるということになると、こうした抜け落ちがなくなって取材上かなり助かることになるというふうに思っております。

では、どうしてこのデータベース機能というのが必要なのか、どのように使っているのかという点ですけれども、スライド2枚目になりますが、ざっくり言うとニュースの価値判断、訴訟のポイント整理、それから全国的な傾向の分析ということになると思います。

ニュースの価値判断についてはこの次のスライドを御覧いただければと思うのですが、判決の価値ということになると、事案に社会性があるということはもちろんなのですが、あと司法判断の希少性等も影響してきます。やはり判例になるような、初めての司法判断というのはニュース性が高いということになるのですが、この初めてかどうかを確認する作業というのが実は極めて難しく、非常に苦勞しています。過去の記事を①から③まで並べてみましたけれども、①は、ファスト映画の無断投稿をめぐる民事訴訟の判決の時にこういう記事が載っているのですが、これは見ていただければ分かるように「元の映画のあらすじを説明した動画で、投稿者に賠償が命じられたのは初めて」というように「初めて」と言い切っています。これは各方面に取材をして、過去に一度もこうした裁判がない

ということが確認でき、自信を持って言える場合にしかこういう表現をいたしません。ですので、これは相当取材を尽くさないとういう表現ができないということになります。それから②なのですが、これは初めての可能性が高いけれども絶対そうとも言い切れないという、もしかしたらどこかに隠れたものがあるかもしれないというときに「初めてとみられる」みたいな表現を使うこともあります。これはかなり確度が高いけれども言い切れないという場合ですね。あと、③として書いてあるのが「〇〇を認定する判決は異例という」というふうになっていますが、これは過去に同種の事例がぼつぼつ見受けられるけれどもかなり少ないという場合に、「異例という」、しかもここはクレジットとして「原告側によると」というように明示をしています。ですので、もしかしたら本当はもっといっぱいあるかもしれないけれども、原告側によるとこれは異例ですということで、我々も取材上異例だということを確認できているということになります。データベースはこのように使っています。

それから二つ目のスライドになりますが、訴訟のポイント整理です。これは実際に紙面に掲載された一覧表なのですが、例えばインターネット上の記述について削除とか表示停止を求めた裁判のケースの一覧表が左側になります。右側が、同性婚訴訟で各地裁の判決等がどうなっているかというものを並べたものです。ですので、これはやはり東京の地裁でこういう判断が出ています、大阪でこういう判断が出ていますというようなものを過去のデータベース等から調べて、そこから抜粋をしてこのような表を作成しているということになります。このほかにも、後で少し御紹介しますが、読売新聞社にも記事のデータベースが別途ありますので、こうした記事のデータベース等も活用しながら間違いがないように重ねてチェックをしているというところになります。ここで非常に難しいところは、左側の表なのですが、高裁という所の右側に色が暗く付いている所は、「最高裁に係属中」という記載があります。つまり、控訴、上告されて現在は最高裁に係属しているということをここで明示しているのですが、ここが、現在の裁判が控訴されているのか上告されているのかというのが、報道機関としてはリアルタイムで知りたいところなのです。ですので、新たに導入される民事判決のデータベースで、更新がいつの段階のものなのかということは、報道機関としては非常に重要なものになります。今日現在控訴されているのか、上告されているのか、それから判決がまだ出ていないのかというような、そういうところをやはりリアルタイムで確認できるようになると非常にありがたいということになります。これは少し技術的な問題もあると思いますので、もしかしたらなるのかもしれませんが。

それから、データベースを使う理由の最後のものとしては、全国的な傾向を知るということです。次のスライドになりますが、これは読売新聞社のデータベースの一部を抜粋したものです。これは過去の記事を地方版も含めて一気に検索できるのですが、その中に「コロナ 給付金 訴訟」というようにワードを入れてみました。そうしますと、結構話題になりました、誤って給付金を送ってしまった市町村の裁判が一番上に出てきて

いますし、それから真ん中辺りには「給付金不支給で事業者が国提訴」というような裁判も出てきます。それから最後の所に「性風俗に給付金 国側は争う姿勢」というようなものも出てきますし、上から2番目は刑事でありますけれども、経産省の元官僚によるコロナ給付金詐欺事件の裁判がありました、こういうようなことが一度に見られるわけです。こうやって裁判の状況を一覧することによって、全国のトレンドを知ることができます。しかも、地域も例えば大分とか大阪というふうに出てきますので、こうした裁判がどういう地域で集中的に起きているのかとか、全国で万遍なく起きているのかというようなことを確認することができることとなります。こうして、こういった検索が結果的に新しい記事にまたつながっていくというようなこともたくさんあります。

それから次のスライドになりますが、報道機関としては民事判決のデータベースのエンドユーザーという面もなくはないのですが、こうして情報を受け取って、それを加工して世の中に発信するという利活用機関としての役割もあります。むしろこちらの役割の方が圧倒的に大きいかもしれません。ですので、その場合のアクセスできる情報というのは、できる限り加工されていない状態のものが望ましいというふうに考えています。AIによって匿名化された情報、判決を仮に読んだとしても、ニュースの正しい価値を判断するのはかなり難しく、また必要な情報の取舍選択といった配慮も難しくなってしまうと思います。例えばですけれども、個人の極めて私的な内容を含んだ裁判の判決があっても、提訴自体や司法判断に社会性や公益性がある場合は、やはり報じることで世間の人に知っていただく必要性があるのだと思います。その場合、私的な部分に踏み込んだ判決の内容というのは、プライバシー侵害等に留意をしまして記事にする場合は適宜匿名化をしています。あるいは、裁判の対象になっているのが社会的責任の重い有名企業だったりする場合は、これは、その会社が訴えた、あるいは訴えられたということに意義があるという場合もありますので、実名で報道することもあります。中身を見て一つ一つどういった配慮が必要か、必要でないかということを報道機関は自らの責任において判断していますので、事前にそれが制約されてしまうと、価値の判断や配慮というものが難しくなってしまうということが言えると思います。それからあともう一つ、現実的な問題として例えば原告や被告がXやYで表示されて手元に届いた場合は、それが男性なのか女性なのか、若い人なのか高齢者なのかといったような属性も一切分からなくなってしまうということになります。そうすると、訴訟の全体像がつかめないばかりでなくて、新聞はXやYと記載するわけにもいかないのです、その場合に記述するのがかなり難しくなってしまうという問題があります。あと、集団訴訟のようなケースですと、例えばXの1から50とあった場合に、果たしてこのXの1とか2、10とか15といっぱい出てきたりすると、これはどの人なのかということ判別するのも難しくなります。こういったこともありまして、やはり匿名化が必要な場合でも最小限にとどめていただきたいというのが私どもの見解です。それからあともう一つは、裁判の報道は判決の言い渡しがあったその日にしなければあまり意味がなくなってしまうと思います。3日も4日も遅れて記事になるようでは、

情報の価値も薄れてしまいます。ですので、現在も判例集等にいろいろと添付されていますが、訴訟の概要や評釈といった付加的な情報も付いていたりすると非常にスムーズな取材と原稿の処理にもつながって、使い勝手が良くなるのではないかと考えています。それから、やはり控訴中、上告中といった情報にできるだけリアルタイムで接することができたら良いなということがあります。これは非常に手間の掛かる作業だと思いますし、技術的な問題もあると思いますので、もしかすると長期的な課題になるのかもしれない。あとは、こうした機能を使う場合に現実的な問題として費用がどのくらいかかるか、1件当たりの検索で料金が発生するのか、定額制なのかとか、この辺りも使う側からすると気になるのではないかと考えております。

これは私が申し上げるべきことではないかもしれませんが資料には書きませんでしたけれども、民事判決を幅広いエンドユーザーの目に触れる形にするという場合には、これとはまた別の議論が必要になるのだと思います。例えば閲覧の対象が私たちのような報道機関の場合もあるでしょうし、法科大学院や大学の研究や教育、学習等に使われる場合もあるでしょう。あるいは企業の法務の担当者が、自分の会社が抱えている紛争と同様のケースを調べて解決に役立てたいといったような場合などもあると思います。用途と目的は多岐にわたると思いますので、そうしたケースごとにどこまでマスキングをするのかしないのか、そうした範囲や必要性も変わってくるのではないかというふうに考えています。それで、あえて一つ言うのであれば、やはりこれがDVやストーカーといった加害者側に悪用されて、仮に元交際相手や元配偶者の居場所を突き止めるのに悪用されるようなことがあると、データベースの存在意義そのものが揺らぎかねないということがあると思います。仮に住所がマスキングされていたとしても、判決の認定内容等から類推できるということももしかしたらあるかもしれません。その辺りはやはり細かな配慮が必要になるのかもしれないし、現在のAIにどこまでそれができるのかどうかということも、私は、その点知見がありませんのではっきりとは分かりませんが、やはりきめ細かく対応するということが必要かと思っています。これは性犯罪や虐待等に係る裁判も同様かと思っています。最後になりますけれども、今申し上げたとおり、データを提供する相手、それから使い方に応じてマスキングの有無等々についてはきめ細かい切り分けをしていただければなというふうに考えております。私からは以上でございます。どうもありがとうございました。

山本座長：

富所様、ありがとうございました。それでは、15時40分頃まで御対応いただけるということですので、せっかくの機会でありますので、是非活発に御質問等いただければと思います。どなたからでも結構ですので、いつもどおり挙手機能で挙手いただければと思います。小塚委員お願いいたします。

小塚委員：

学習院大学の小塚です。どうもありがとうございました。特に利活用機関として利用さ

れるという御発言であるとか、やはり企業が当事者の場合は実名というのが社会的にも重要だというような御発言を大変重く受け止めました。それで、時間が限られているようですので二つ、簡潔に伺いたいと思います。一つ目の御質問は、現在ここで議論しているのは、ある年度から全件公開を始めますという議論だと思いますが、価値判断的なことをおっしゃったので、過去の分の公開ということをどうお考えになるか、とりわけ過去分は全件ではなくある時点から全件公開されているという、継ぎはぎの状態が一定期間続くことをどうお考えになるかというのを伺いたいと思います。

二つ目の御質問は、最後におっしゃったこととの関係で、新聞記者の方はジャーナリストとしての職業倫理のようなものがあると思いますが、同じ会社でも例えば週刊誌なども出しておられた時期が少なくともあったと思いますが、結構外注取材というのが世界にはありますね。マスコミ、マスメディアであっても外注取材などの場合の当事者の倫理等はどういうふうに考えていったらよいか、少しその辺りも御存じのことがありましたら教えていただけますでしょうか。

富所様（読売新聞）：

どうもありがとうございます。1番目の、過去の分の公開に関しては、これは私も非常に関心がありまして、本来的に申し上げると、データベースはやはり全部分かった方が当然ありがたいのですね。ですが、これは裁判資料のデジタル化の問題と密接に絡むと思うので、過去のものを今からデジタル処理してデータベースに入力するという作業は恐らく膨大な手順が必要になると思いますし、ある程度継ぎはぎになるのは現時点ではやむを得ないかというふうに思っています。ただ、先ほど申し上げたように過去に同じものがないかどうかという、無いことの証明はものすごく難しいので、本当は全部そろっていないと無いということも証明できないということになってしまうので、ある方がありがたいというのは実際のところであります。

それから外注なのですが、少なくとも裁判の取材に関しては、私どもは、自社のものについては、自分たちの自前の記者でやっておりまして、あとは外の方に傍聴記を書いてもらうことなどはあるのですけれども、掲載の責任は会社にあるので、載せる前に、例えば事実関係に誤りはないか、行き過ぎた表現はないかと、そういうところをチェックをしています。もちろん自社の掲載基準みたいなものもありますので、そういうものに照らして適切かどうかということはチェックしています。ただこれは、今御指摘のように、会社によってもかなり温度差があると思いますし、相当踏み込んで書くことによって世間のニーズに応えるというような、そういう考え方もあり得るのかもしれませんが、ですので、それはやはり個々の会社の考え方と責任において処理をするというのが報道の世界の原則かなというふうに思っています。以上です。

小塚委員：

承知しました。ありがとうございました。

山本座長：

ありがとうございました。それでは続きまして、板倉委員お願いいたします。

板倉委員：

ありがとうございます。四つあるのですが、一つは、利活用機関になることも考えられているということなのですが、これは個社で考えられているのか、もう少し広く団体として考えられているのか。その際に生データの方が良いという御意見はそうだろうと思うのですが、これは会議全体へのコメントになりますが、一部の利活用機関には加工の程度を変えるということがあり得るのかというのが論点かと思いました。

もう一つは、なるべくリアルタイムで欲しいということで、それは報道なのでそうだろうと思うのですが、恐らく裁判所である程度たまってから出すということになると思うのですが、どれくらいの頻度でしたら報道として耐え得るのか、一方で最高裁としてはどれくらいの頻度でしたら出せるのか、月1だと新聞は多分つらいかなというところだと思うのですね。週1でしたらそれがあり得るのかというところですね。

3番目は要望なのですが、新聞で報道される際に正確な判決年月日とか事件番号とかは書かれていないと思うのですけれども、こうやっていっぱいになると書いていただかないと読者の方も検証できないので書いてほしいなというのがありますが、それは書かないでしょうかという話です。

4点目はまた少しはみ出す論点かもしれませんが、結構報道の方は開廷表を御覧になっていて、私が関係している裁判とかでも「開廷表からなくなったのですが終わりましたか」とか聞かれて「言えません」とかいうふうに回答したことがあるのですが、開廷表も後々はネタとしては欲しいでしょうかというのが4番目です。以上よろしく申し上げます。

富所様（読売新聞）：

どうもありがとうございます。利活用機関として使うというのは、これまでの私の取材経験と今の現場の利用者から、恐らくそういう使い方になるだろうということで申し上げました。実際に多分使う段になって、では会社としてどのように使うのか、どの程度使うのかということは判断することになりますので、現時点では私見と言われれば私見ということになります。

それから、リアルタイムの問題なのですが、これは、その日のうちにデータベースに載るというのは現時点では私どもは考えておりません。そうなったら良いなということではありますが、ただ私どもの会社にもデータベースがありまして、支局も各都道府県にありますので、そこで大きな裁判等は記事になっていたりするのですね。それはリアルタイムで当日には見られるようになっていきます。ですので、それも使いながら、古い判決等についてはデータベースも活用しながら、という複合的な活用に恐らくはなっていくと思うので、これは私も無理を承知でといいますか、そうなったらうれしいなという意味で申し上げた次第であります。

それから、判決の年月日は、通常判決の記事は判決があった翌日の新聞、オンラインであれば当日にはもう出ますし、何日に〇〇地裁で判決が言い渡されたと記事には書いて

あるので、日付と地裁は分かるのですね。事件番号は、一般の読者にはなかなかなじみがなくて、それを報道の中に記載するというのは、もしかすると難しいかなというふうには思っております。

御指摘のとおり開廷表は、その日にどういう裁判があるのかというのは司法記者であれば大概の者は、もちろん全部ではありませんけれども、何か大事なものを見落としがないかというのをチェックするために見ておまして、確かにそういう情報もあつたら便利だなというふうには思いますが、例えばそれも全国の一般のエンドユーザー全部に開廷表としてぱっと見られるようになっていいものかどうかというところも、これもまた別の議論が必要かもしれないなというふうには少し思いました。私からは以上です。

板倉委員：

ありがとうございます。

山本座長：

ありがとうございました。残りの時間は10分弱なのですが、3人の方から手が挙がっていますので、少し時間を考慮しながら御質問いただければと思います。小町谷委員お願いします。

小町谷委員：

弁護士をしております小町谷と申します。判決等の情報というのは、基本的に司法記者クラブの方は判決の情報をそのまま入手でき、一般の方が入手できない情報を多数入手していると理解しています。そうすると、このデータベースを作ることは、司法記者の方がアクセスできる情報以上に、どこに最も価値があるとお考えでしょうか。

富所様（読売新聞）：

どうもありがとうございます。過去のデータベース等に載っておらず確認できなくて、過去の裁判も自社の記事しか残っていないというようなものもあるのです。その場合は、当該地裁がある支局なり本社なりに連絡をして「判決を見たいのだけれども、参考になるようなものはないですか」と言って取り寄せたりしているのです。ですので、こういう作業が不要になって一発で見られるようになってくると、取材・執筆をして読者にいち早く情報を届けるという意味では非常に効果が高いかなと思っております。ただ、そうすると欲しいなと思うのはやはり古い判決だったりするので、それがデジタル化以降というと、使い勝手が良くなってくるまでには、もしかしたらある程度の時間が必要になるのかなと思っております。以上です。

小町谷委員：

ありがとうございます。

山本座長：

ありがとうございました。それでは鹿島委員お願いいたします。

鹿島委員：

ありがとうございます。時間もないので手短にというところなのですが、本日は貴重な御報告ありがとうございました。司法書士の鹿島と申します。これまで簡裁の裁判例が記事になるということは少なかったように思うのですが、悪質商法とかの社会問題となるような消費者事件というのが、基本的に統計上だと平均被害額は70万から80万円というふうにも言われていますので、そのような意味でも個別事件として最初に簡裁に提訴されることは多いと思っているので、今後データベース化によって報道機関の方に目に留まる機会が増えるということは私たちが期待をしている部分ではございます。一方で判決の中には欠席判決も大変多いので、判決文で訴状引用がされる場面がありまして、裁判官の判断と読み誤ってしまうというような、情報の利用に関する懸念と申しますか、情報リテラシー上の懸念のようなところもあると思うのですが、先ほどの御報告で利活用機関としての参入も視野に入れるというところでしたが、私どもは、そのようなところが期待するところではあるのですけれども、ほかに何か利活用機関として想定される目的や方法等があれば教えていただければと思います。よろしく願いいたします。

富所様（読売新聞）：

どうもありがとうございます。今簡裁の話が出ましたけれども、やはり私たちも裁判の取材をする中で、目が行き届かないなという反省は常にあります。やはり裁判の量が膨大ですし、裁判の多くは離婚訴訟だったり建物の明け渡しだったりするわけで、もしかしたらその辺りに重要なものが隠れているのかもしれないけれども、やはりなかなか目が行き届かない。簡裁も同じような理由で目が行き届きません。ですので、今までの私どもの感覚からすると、実はこういう裁判があってこういう判決が出ていましたということ了他社に先駆けて報じることが報道機関としての意義のようなどころがあるのですけれども、それがデータベースで簡単に出てきてしまうようになると、現場の取材記者の頑張りというよりは、いかに早く検索するかというようなことが問われてくるようになります。そうすると、我々の仕事ぶりももしかしたら変わっていくのかもしれない。たくさんある情報をうまく分類をしたり、分析をしたりしながら、いわゆるデータジャーナリズムと言われるのですけれども、そういったような報道の方向に司法ももしかしたら今後は進んでいくのかもしれないというのは、今回のお話を伺ってとても感じているところです。

鹿島委員：

ありがとうございます。

山本座長：

ありがとうございました。それでは町村委員お願いいたします。

町村委員：

本日はお話ありがとうございました。成城大学の町村と申します。今考えられているのは判決情報なのですよね。記者の方がこれまでよく利用されている訴訟記録等を追って

いくというようなことは、デジタル化されても残念ながら簡単にはならないし、データベースとしても提供されないというのが一応の前提です。私はそうあってはならないと思っていますが。そういうことで、今でも、見られる訴訟記録を記者の方は追って行って事件の内容を深く知るといふことをされると思うのですけれども、そういうときに仮名化された民事判決データというのは、過去に類似のものがあるかどうかを探すというのは分かりますが、当該事件を知る上でどれほどの意味があるのかというのが少し疑問に思うところなのです。仮に追っていた事件が判決として公開されたというときに、それが仮名化されたデータで出てきたとしても、それは既に知っていることなのではないかと思うので、そうするとやはり新たにデータとして出たものに注目をして、事件番号等から記録を追っていくというような使い方をするのではないかなと思うのですけれども、そうすると仮名化していても実名はそうやって取れますので、実名を取って復元した形での利用というのをメディアの方はされるのだろうなというふうに想像しているのですが、それで間違いないでしょうか。逆に言うと、それを禁じられては困るというふうに想像しているのですが、いかがなものでしょうか。

富所様（読売新聞）：

ありがとうございます。それが非常に難しい問題で、やはり記者として一つの裁判をしっかり追いたいというときに、訴状から準備書面から全部目を通して、最終的に判決も含めて全体像を評価するというのがあるべき姿なのだろうなと思います。もちろんどこまで時間的に追いきれるのかという問題はあるのですが、できる限りやれることはやった方が良く、そうあるべきだと私も思っています。ですが、それが匿名化されて出てきた情報は、今の話だと多分少し遅れて出てくる形になるのでしょうかから、今やっている裁判について後から匿名化されて出てくるということに関しては、そんなにそのデータベースを見て何か知見を得るといふようなことには多分ならないと思います。なので、今あるものは今あるものとしてしっかり追っていくという形になっていくのだろうなと思います。ただ、判例集とかを見ていても、こんな裁判があったのかというように、後から気づくこと、目に触れるということはあるのですね。ですので、それをきっかけにほかにも全国で同じようなことが起きているのではないかと、少し古めの判決から掘り下げていくようなプロセスもあると思います。ですので、原則はやはり実名を含めてマスキングされていないものにアクセスしたいというのはもちろんあるのですけれども、一部匿名化されているから、それが役に立たないということにはならないのかもしれませんが。本当にケースバイケースだと思います。あまり答えになっていないかもしれませんが。

山本座長：

ありがとうございました。それでは、予定された時間になりましたので、恐らくまだ御質問はあろうかと思いますが、この程度にさせていただければと思います。富所様には大変お忙しい中、御対応をいただきまして誠にありがとうございました。貴重なお話を我々の今後の審議に生かしていきたいというふうに存じます。ありがとうございました。

た。

富所様（読売新聞）：

ありがとうございました。

山本座長：

それでは、富所様にはこれで退出いただければと思います。引き続きまして、次の議題に移りたいと思います。事務局の方からお願いできますか。

事務局：

渡邊です。これまで委員の皆様からはデータベース化実現のために必要な民事判決情報の取得、仮名処理、管理提供のプロセスについて現状の想定やイメージを確認したいという御意見を頂いていたところでございます。この点につきましては、日弁連法務研究財団の民事判決のオープンデータ化検討PTにおいて、必要なシステムや体制の在り方に関する検討が重ねられてきたところでございます。そこで、本日はそのPT事務局の大坪和敏弁護士にお越しいただきまして、PTにおける実証実験の結果やその結果等を踏まえた議論、検討の状況について御紹介いただきたいと思います。御発表いただいた後、質疑応答にも御対応いただく予定でございます。なお本日は、この後の予定として杉村委員に海外調査の追加報告をお願いしておりますけれども、大坪弁護士にお話しいただく民事判決情報データベース運用上の課題につきましては、皆様の御関心がとても高いところかと存じますので、大坪弁護士には御質問が尽きるまで御対応をお願いしたいということで依頼しているところでございます。以上でございます。

山本座長：

ありがとうございました。それでは大坪弁護士からよろしく願いいたします。

大坪様：

日弁連法務研究財団の民事判決のオープンデータ化検討PTにおいて事務局をしております弁護士の和敏と申します。本日は発言の機会を頂きましてありがとうございます。法務研究財団における民事判決情報データベース化事業の検討状況についてお話しさせていただきます。

資料の3ページ目を見ていただきたいのですが、法務研究財団について最初に御紹介させていただきます。当初は財団法人として設立しておりましたが、財団法人日弁連法務研究財団は日弁連や公認会計士協会、税理士会、弁理士会、司法書士会等の関係団体の協力を得て1998年4月に設立されたものです。その後、公益認定を受けまして2010年10月に公益財団法人日弁連法務研究財団となっています。財団の活動内容としましては、主に会員を対象とした研修事業と研究事業を中心として、この資料に記載されているような事業を行っております。研究の事業につきましては、財団自体が直接研究するということがあまりなく、会員の方が取り組まれる研究テーマに対して助成するということが中心になっています。この資料の右側の一番下にあります法科大学院認証評

価事業のように、その時の状況によって個別の事業を行っているというものもあります。今回の民事判決情報のオープンデータ化の検討は、必ずしも財団の本来の業務とは言えないものではあるのですが、研究活動の一環あるいはその研究に寄与するものということで取り組んでいるものです。

資料4枚目の「民事判決のオープンデータ化検討PT」についてです。このPTは、元々は2020年3月に、民事司法制度改革推進に関する関係府省庁連絡会議で「民事司法制度改革の推進について」というものを取りまとめておりますけれども、その中で「法務省は、民事判決情報を広く国民に提供することについて、司法府の判断を尊重した上で、ニーズやあい路等につき必要な検討をする。また、最高裁判所においては、民事判決情報の提供も含め、法務省における上記検討に協力することが期待される。」とされています。これを受けまして、法務省、最高裁の御参画もありまして、財団の研究活動の一環としてこのPTを2020年3月に立ち上げました。このPTでは、民事判決情報のデータベース化に向けた課題の抽出や、AIを活用した機械による自動仮名化処理のための実証実験を中心として、有識者ヒアリング・意見交換を行っております。その中間的な取りまとめとして、2021年3月に、「民事判決情報のオープンデータ化に向けた取りまとめ」を発表しているところです。この取りまとめにおきましては、事業化の前提となる構想として、民事判決情報のオープンデータ化の考えられる私案というものを示しております。そこについては後ほど御紹介いたします。この取りまとめを受けまして、2021年7月に、PTの下に二つのWGが発足しました。一つは民事判決データベース化事業の在り方に関するWGで、こちらでは事業を行うに当たっての体制整備の在り方を検討することとして現在も検討を継続しております。もう一つは、民事判決情報の仮名処理の在り方等に関するWGで、座長は山本和彦先生にお願いしております。こちらでは、研究者の方々に多数御参加いただきまして、法的な観点からの検討を行っております。この検討の結果につきましては、昨年6月に、「民事判決情報の適正な利活用に向けた制度の在り方に関する提言」として公表しているところです。

次のページは、PTにおける私案について、若干整理してスライドにしたものになります。左側が現状になっておりますけれども、現状は、裁判所の民事判決につきましては、商用のデータベース会社や判例雑誌等の出版社が、それぞれ裁判所から独自に紙の判決文を入手してデータ化するとともに、各社それぞれの基準に従って仮名処理をした上で一般に提供されています。これとは別に、裁判所でも裁判所のホームページにおきまして一部の判決を公開しており、こちらでも裁判所が独自に仮名処理をしているというふうになっています。裁判所の方は詳しくは分かりませんが、データベース会社等や出版社は、それぞれ相当のお金をかけて仮名処理をしているというのが現状になっております。これに対しまして、財団のPTで検討しております私案というものの内容が資料の右側になります。こちらでは、裁判所から民事判決情報の提供を受ける機関として、真ん中の黄色い所ですが、情報管理機関というものを設けまして、その情報管理機関で

各データベース会社や出版社がこれまで人手と費用をかけて行っている仮名化の作業等を集約して行って、そこで仮名化されたデータを、緑色の所のデータベース会社や出版社に有償で提供するということを予定しています。右側の緑の部分については利活用機関というふうに我々は呼んでいるところです。

資料の6ページは「私案実現のために必要なシステム・体制」です。情報管理機関としては、主として次のような業務を行うということを予定しております。まず、裁判所から仮名処理前の民事判決情報を取得して、それを管理するということです。次に、取得した民事判決をAIによって仮名処理、そしてAIによって仮名処理したものを人手で修正・ダブルチェック、そして仮名処理した後の民事判決情報を利活用機関に提供、最後に、現在の紙の判決書やそれをテキスト化したデータにはない、付加的な情報についても従来の紙の判決書の情報に付加して裁判所から提供を受け、これについても利活用機関に提供するということを予定しているところです。この付加的な情報としましては、どのようなニーズがあるかについて、事業の在り方に関するワーキンググループに参加しているデータベース会社等にヒアリングをしたところ、判決のIDとか判決の言渡し日、裁判所のID、事件番号、審級関係、原審裁判所の判決のID、口頭弁論の終結の日、閲覧等の制限決定があった場合、更正決定があった場合等の情報について、付加的に提供いただきたいというニーズがあるということが分かっております。

次のページは、民事判決情報取得から仮名処理の流れを図にしております。真ん中の情報管理機関の黄色の部分について、細分化したものが次の8ページになっております。こちらを見ていただきたいと思います。仮名処理の中身について、システムを作るに当たって、作業を細分化したものがこのスライドになっています。仮名化するにはこれらのオレンジ色の箱のそれぞれの作業が必要になると考えられておまして、各作業の役割分担につきましては、最初の「システム」のところで機械仮名化まで終えた上で、「担当①」がチェック作業を行うなどします。作業内容・役割分担が変わるということもあります。これはダブルチェックを前提としておりますので、担当①と②の担当者がチェックをして最終的に提供するところに移るという仮名作業を予定しているところです。

次のスライドに移りまして、情報管理機関から利活用機関への提供の流れを図にしたものです。ダウンロードの条件として赤い字で書いてありますが、全件をダウンロードする場合と、例えば労働事件等といった特定の分野の判決を全件ダウンロードする、また欠席判決等の調書判決は要らないのでそれを除いた全件をダウンロードしたいというようなニーズがあり得るところです。提供後に更正決定や民事訴訟法92条の秘密の保護のための閲覧等の制限がされている場合に、その情報をどのように利活用機関に提供するかということも問題になるわけですが、それについては今後の検討課題ということになっております。情報管理機関で保管した判決については、基本的には新しいものに更新するということを予定しているところです。その更新したものを再度、更正決定等と一緒に個別に提供することになるということが今のところ考えられております。なお、地裁の判決

の場合には、提供後に控訴審・上告審等の判決が出たときには、それらの審級の結び付きが分かるように提供することになると思われます。提供済みの地裁の判決について、その審級の情報等その都度新しく更新されるものを提供するか、その方法についても基本的には更正決定等の情報と同様に考えられるところですが、こちらについても今後細部について詰めていくということになります。仮名化前のデータの保管ということも問題になっており、こちらについても、これまでのヒアリングの中で一定のニーズが考えられるところですが、利活用機関には今のところ提供しないということを想定しております。以上の、裁判所からの判決データを取得してから利活用機関へ提供する時間ですが、技術的・体制的には数日で可能というふうには考えられているところです。ただ、実際にはシステムの都合等もありますので、現実的な時間はもう少しかかるのではないかと見込んでいます。さらに、全体的なシステム構築にかかる初期の費用に関しては、今のところ概算で見積りを何社かから出しているところですが、平均すると1億5,000万円程度かかるのではないかと見積もっています。

次に、10 ページ目の「仮名処理ツール・体制」です。こちらでは、PTで行いました実証実験の結果の概要を書かせていただいております。現在判例データベース会社に掲載されている地方裁判所の判決は、全体の数%ということになっているようですけれども、判例雑誌出版社や判例データベース会社は現在、基本的に手作業で仮名処理をしているということでした。これを年間20万件程度言い渡されている判決について、全て手作業で仮名処理しなければならないということになりますと、到底対応することが難しいと考えられておりますので、機械処理をする必要があるのではないかとということになります。財団のPTでは株式会社Legalscapeに委託しまして、AIを利用した機械処理で仮名処理ができるのか、仮名化前の民事判決情報について、どの程度の精度で仮名化ができるのかについて実証実験をしております。実証実験の詳細については時間の関係で省略させていただきますけれども、簡単に概要を御紹介しますと次のとおりでございます。まず、機械処理の流れとしては、大きく対象の語句の特定と、語句の属性の特定という2つのステップに分けて処理を行っております。対象の語句の特定は、対象となる文章の中から、例えば「山田太郎」といった仮名化する対象となる語句を特定するもので、言語処理分野の研究の中では固有表現抽出に当たります。次のステップの語句の属性の特定は、仮名化する対象の語句ごとに適切な仮名の記号に置き換える処理で、例えば「山田太郎」と別の箇所に出てくる「山田」という記載が同じ人であるという場合には同じ記号を付与し、その山田という人が別人であれば別の記号を付与するというような処理になります。これは言語処理分野の研究では関係抽出というものに当たります。実証実験の結果では、対象の語句の特定につきましては、仮名すべき単語のうち実際に仮名化できた割合を再現率と言っておりますが、再現率につきましては94.5%、つまり5.5%ぐらいは仮名が漏れているということになります。仮名処理を実施した単語のうち、仮名化すべきであった単語の割合・精度につきましては93.4%という結果でございます。この精度というのは、

つまり 6.6%が仮名し過ぎということになります。最後の語句の属性の特定という面では、仮名対象のグループの一致割合・精度は 98%という結果になっております。この実証実験の結果につきましては、今後カスタマイズすることによりさらに精度を向上させるということが可能であるものの、時間の経過とともに新しい個人の名前が出てくる等の変化をするということがありますので、機械処理だけで 100%仮名化を行うということは不可能であって、機械処理後に人手により確認・修正作業を行うことが不可欠だろうというふうに言われております。実証実験では、更に人手による確認・修正ツールの開発・運用実験を実施しております、それが次の 11 ページ目の実証実験の内容になっております。

この実験では、自動仮名化処理システムというものによって施された仮名処理を、人手で修正するのにかかる時間というものを検証しております。この実験では、第一法規株式会社様による仮名処理基準に基づいて仮名処理作業を行っているところです。この実証実験の作業では、2 人の担当者の方に修正作業とダブルチェック作業というものを行ってもらうことにしました。1 人の担当者の修正作業というのは、仮名処理システムによって仮名処理が施された判決文が画面に表示されまして、その表示されたものに対して適宜修正を加えて仮名処理を完璧にするという作業になっています。その後、別の担当者がダブルチェック作業を行うということになっています。ここでは、最初の担当者が実施した修正作業によって完璧に仮名処理がされた判決文が画面に表示されるので、それに対して適宜修正を加えて仮名処理を行い、完全なものにするということになっています。このように、ここで書かれている二つのシナリオでどのくらい時間がかかるかというものを調べております。一つ目の全文シナリオは、図の中の「1 全文確認修正」とされているものですが、これは仮名漏れも仮名し過ぎも排除するというものをゴールとしています。この判決文の全文を上から下まで確認して、完璧な仮名処理を実施するというものを考えています。二つ目がその下の方にある「2 機械出力のみ修正」になります。この機械出力のみの修正は、システムをもう少し信頼して、仮名漏れはないか、あっても極めて少ないということ想定しまして、仮名し過ぎを排除するというものをゴールとしています。判決文の中でシステムが仮名すべきと判断した箇所を一つ一つ全て確認して、高速な仮名処理を実施するという、このシナリオでは全文シナリオで実施した作業の一部を省略した形になっております。

次の 12 ページ目は、実証実験の結果と概要になります。文字数を考慮して修正した結果になりますが、かかった時間について見ますと、シナリオ 1 の全文確認修正では約 13 分かかっておりました。シナリオ 2 の機械出力のみの修正では約 4 分という結果になっております。ヒアリングの結果、現状では平均的に 1 件 30 から 60 分程度かかっているというところが、大きく効率化されていたということが分かりました。

最後の 13 ページ目です。ダブルチェックを前提として 1 件当たりの作業時間は十数分ということになりますので、欠席判決を含む約 20 万件を、毎日更新をする場合の人手での修正には、16 人程度の体制が必要となるのではないかと考えられます。年間のコスト

は約 4,400 万円かかるのではないかと見積もられております。資料の最後の方には実証実験に参加した方の参考の御意見が記載されています。簡単ですけれども私の発表は以上となります。ありがとうございました。

山本座長：

大坪弁護士、ありがとうございました。それでは質疑応答に移りたいと思います。先ほど事務局からありましたとおり、大坪弁護士には一応御質問が尽きるまで対応をお願いしているということでございますので、委員の皆様におかれましても時間を気にせず御質問、御議論をいただければと思います。それではどなたからでも結構ですので、御発言をお願いしたいと思います。湯淺委員お願いいたします。

湯淺委員：

明治大学の湯淺でございます。大坪弁護士、どうもありがとうございました。3点質問させていただきたいのですが、1点目は簡単な質問でございます、仮名のし過ぎ、仮名過多という御説明がございましたけれども、仮名のし過ぎというのは、例えば「湯淺は〇〇した」というときに「湯淺」の所だけを仮名すべきなら「湯淺は」の所まで仮名をしてしまったという意味なのか、仮名過多とはどういうことかということでございます。

2点目が、今日はコスト感も御紹介いただきまして大変貴重なお話であったかと思いますが、業者にある程度費用感を問い合わせたりされたときに、先生の今日御紹介いただきました費用感は、いわゆるセキュリティ対策を含めた部分なのか、それともそれは含んでいなくて、純粹にシステムのところだけなのかということ、もし可能でしたらお教えいただけますと幸いです。

3点目は、逆にプロジェクトを通じて、特に個人だけではなくて裁判所から提供される生の判決部分は非常に機微な情報が入った形で提供されるのかなと思っておりますが、そのセキュリティ対策をどうすべきかということにつきまして、何かPT全体あるいは先生の御感想として、「やはりここはこうしないとまずいのではないか」というところが何かございましたらお教えいただけますと幸いです。よろしくお願いいたします。

大坪様：

ありがとうございました。まず、仮名し過ぎの点は、湯淺先生がおっしゃったとおり、例としてはもう少し適切なものがあるかもしれませんが、本来すべきでない所を仮名し過ぎてしまったということになります。

2点目については、一応今回見積りを出すに当たっては、前提の条件がありますけれども、その前提の条件に従ってセキュリティについても考慮していただくということは予定しております。前提の条件というところは、例えば、まだ決まっていないところで、仮名化前の判決についても情報管理機関で一時保管することになるかもしれませんけれども、それを長く持つということになると、きちんとしたセキュリティが必要になってくるということが考えられるのですが、今のところそんなに長く持たなくてもいいのではな

いかとも考えられていて、仮名化後のこともそうですけれども、結局情報自体をどの程度の期間保管するかによってセキュリティも変わってくるのではないかと考えております。今のところは、大体 20 年ぐらい保管する前提でのセキュリティということで考えております。

3 点目の機微な情報については、これからの議論になるのかと思いますし、ざっくりとした回答になってしまいますけれども、一定の基準を設けてもやはり事案によっては言ってみれば匿名化しなければいけないというような情報もあるのだろうというふうに思っています。その点についてどうするかというところは、なかなか基準は決められないところではないかと思われて、そこにつきましては、結局 AI による機械化の仮名処理を 100% ではできないということになりますと、人手でやらざるを得ないところが残っておりますので、この人手のところでも多少柔軟な対応というのがやはり必要になってくるのだろうと個人的には考えております。お答えになっているでしょうか。

湯淺委員：

大変貴重なお答えを頂きましてどうもありがとうございました。

山本座長：

ありがとうございました。それでは米村委員お願いいたします。

米村委員：

本日は貴重なお話を頂戴いたしまして、大変勉強になりました。ありがとうございました。私からは、この実証実験の中身に関して、やや細かいところになりますが何点かお伺いしたいと思います。私が聞き漏らしたところを含んでおりましたら申し訳ございません。

まず、この選ばれた 150 件というのは、どういう基準で選ばれたのかというのが 1 点目でございます。

それから関連して、その 150 件の判決文の長さというのは、どの程度であったのかというのが 2 点目です。最後の結果の表示のところ、文字数を考慮して時間数を補正されたというような御説明があったように思いますが、具体的にどういう形で補正をされたのか、例えば一般的な判決文の平均文字数のようなものを計算されて、それによって補正したというようなことか、それとも単純に 1 文字当たりにかかる時間数を計算したというようなことなのか、その辺りの補正の仕方を少し教えていただければ大変ありがたく思います。

大坪様：

ありがとうございます。すみません。実証実験の詳細な中身については、私も頭に入れていないところがございます。正確にお答えできないかもしれませんが、できる範囲でお答えさせていただきます。まず 150 件の判決を選んだ基準というのは、基本的には既に判例データベース会社の方で仮名化して公表されているデータがありますの

で、それをランダムに選んでいただいたということになっています。特にこういう事案という指定はなく、現在の 150 件を平均的に取ってくださいということで依頼して、それで提供を受けたものを使っております。実証実験に当たっての文字数とか分量についても、その中で平均的なものを頂くことを前提としておりましたので、例えば原発訴訟のような何百ページもあるようなものについては、今回は頂いていないということになっております。事前のヒアリング等を踏まえまして、一つの判決の平均的な判決書の枚数が 10 枚程度というところだったと思いますので、そういうものを前提として、ざっくりと依頼をしているところで、実際に Legalscape の方で文字数の修正を何文字で調整したかというところについては資料を見れば分かるのかもしれませんが、私の方でそこまで頭には入れておりませんでした。

米村委員：

ありがとうございました。今のような質問をさせていただきましたのは、精度との関係で、機械処理が活用可能な事案に限定があるのではないかと考えたことが理由です。もちろん、こういった機械処理の精度が十分であれば、もう機械処理だけで良いということになりそうですが、機械処理というのは、件数がたくさんあるものほど精度が上がっていくというところがあるはずですので、件数の少ない事案類型については十分な処理ができず、仮名漏れあるいは仮名過多が生じやすくなるという可能性があるように思います。そういった、件数が少なく、しかも仮名漏れ等による修正の必要性がかなり大きいような事案に関しては、特に全文確認修正のような手法を併用していくこともあり得るように思いましたので、どの辺りの事案を念頭に置いてこの実証実験をされたのかということを知りたかったというのが基本的な趣旨でございます。

山本座長：

ありがとうございました。それでは続きまして、小塚委員お願いいたします。

小塚委員：

小塚です。今この大坪先生の資料の「出典」というところを見ましたら、私も参加していた時に御報告いただいた実証実験のままのお話だったということを理解いたしました。恐らく米村先生の御質問の細かいところも、この資料の 11 ページ目に書かれている出典を御覧いただけると出ていると思います。

私からは 2 点ですが、一つは、金額について、この資料に書かれていないところで複数社に見積りを取られたというこの金額に、最後のスライドに出ている作業員のコストというのがオンされてくるという理解なのか、それともこれは含んで見積りをしておられるということなのか。オンされてくるとなると、この書き方で大丈夫なのかと思っていきます。というのは、実験のときには既に今データベース会社で作業している方をお願いをしているわけですので、既に熟練した方であり、かつ私の理解に間違いがなければ恐らく従業員として雇われている方ではないかと思うのですね。仮にこの情報管理機関に集約

することになりますと、データベース会社では雇用をやめられると思いますし、そもそも誰が雇用するのだという問題が出てくるし、場合によってはそういう外注業者のようなものを作るのか派遣になるのかということも出てきますし、今作業しておられる方は、例えば10年後、20年後となってくれば御退職とかいうこともあり得るので、そうすると新規の方のトレーニングのようなことも必要になる。そういうことはこのコストには含まれていないと思いますので、その点を指摘するとともに、先ほどの見積りに含まれているのかどうかという点について御質問させていただきたいと思います。

2点目はセキュリティに関して、作業の方がどういう環境で作業をするかということもセキュリティの問題に入ると私は思います。最近ではカフェなどでよくパソコンを使ってオンライン会議をしている方もいらっしゃるし、ああいう会社はセキュリティのポリシーはどうなっているのだろうとすごく気になっているのですけれども、恐らくこの作業はそういうことはしていただいている作業だと思いますので、そうすると作業場所の確保とか、あるいはテレワークはできないことにしてしまうのかとか、そういうことによってもコストがまた増えそうな気がするのですが、その辺りについて、WGの方でその後何か議論していることがありましたら教えてください。よろしく申し上げます。

大坪様：

ありがとうございます。まず1点目の金額のところですが、1億5,000万円というのは、作業員の方の費用、ランニングコストとは別になります。最初のシステム開発のところで1億5,000万円ぐらいと見積もっているところです。そのランニングコストにつきましては、この委託の仕方をどうするかというところは必ずしも決まっていないところですのでございまして、情報管理機関でそういう仮名作業をする方を雇うということは今のところ考えていないのではないかと思います。だからどうするかというと、例えばその業者の方が仮名処理をするシステム開発や人手による修正の部分も含めて委託を受けるということもあり得るかなと思います。ここをどうするかというところは、実際今のやられている方々のお力も借りる必要があると思うのですけれども、御意見も踏まえてできるだけ費用を抑える形で考える必要があるかと思っています。本当は大まかに言いますと、できるだけ機械である程度のものでできれば、比較的熟練した方でなくてもできるようになるといいかとは思っています。つまり、大学生のアルバイトとかでもできるのが本当は望ましかったのかもしれないのですけれども、いきなりそういうことはできないでしょうから、その辺は、将来はそういう形で費用を抑えられるようにAIの方の精度も上げていく必要があるのかなというふうに思っております。

2点目のセキュリティに関して御指摘の点につきましては、問題としてはありますけれども、それについてどう対応するかというところは、費用のところとも関わってきますが、どういう形で対応するかというところは全く決まっておらず、まだ検討はしていませんでした。その点についても今後十分に配慮しなければいけないということが分かりました。ありがとうございます。

山本座長：

ありがとうございました。それでは異委員お願いいたします。

異委員：

東京大学の巽と申します。本日はありがとうございました。私から2点ございまして、一つは仮名化のし過ぎというところに関して、もう少しお伺いしたいと思います。先ほどの湯淺先生の御質問は、例えばですけれども、判決文中に「巽智彦は、〇〇した」というときに「巽智彦」だけXにすればいいところ「巽智彦は、」までXにしてしまって日本語が通らなくなっているというときに、元の文章に戻すということを問題にされていたと思うのですが、そういうことにとどまっているという御主旨なのか。もしくは、資料の13ページには、実証実験参加者の意見の中に「推知情報をどこまで仮名化するかで人手の作業量が増減する」という御意見が載っていますけれども、例えば推知情報ですと、形式的な基準ののっとなって仮名化・匿名化をしたのですが、判決文全体を読んだときに、このレベルで仮名化をしてしまうと、要するに判決の中身が分からなくなってしまうとか、そこまでやらなくてもこの事案類型だったら推知はされないでしょうというような、より実質的な判断を作業員の方がされて、元に戻すという作業までされたのか。仮名化し過ぎというときに、形式的に日本語として通らないものを直したというレベルなのか、もう少し実質的な判断を人の手で加えたということなのかというところを追加で教えていただければというのが1点目です。

2点目は、仮名化そのものの話ではないのですが、7ページのイメージ図と、その直前の6ページで付加情報の提供ということをおっしゃっていたときに、付加情報の内容として、審級情報ですとか様々なものがニーズとしてはあるとおっしゃっていて、そういうものは判決文自体ではなく、別の情報として出てくるということになっていると思うのですが、こうした情報の付加は今回の実証実験ではひとまず対象外で、判決文本体の仮名化のところだけやったということでもいいのかというのが2点目の質問です。

2点目に関してもう少し申し上げますと、審級情報ですとか、または並行している別の請求や訴訟の情報とか、裁判所の側で判決文情報とまとめて情報管理機関に提供できそうなものはそうするというスキームを念頭に置かれていたのか、それとも、判決文データと付加情報が別々に提供されて、情報管理機関の側でいちいちそれを突き合わせるというスキームを想定されていたのか。付加情報は今回の実証実験の対象ではないのだと思うのですが、財団PTの議論の中でその取扱方について何らかの議論がありましたら、参考までに少し教えていただきたいというのが2点目に付随した質問です。

大坪様：

ありがとうございます。まず1点目なのですが、結論として、推知情報については、今回の実証実験では考慮しておりませんので、特にそれについて仮名処理をするということはやっておりません。ですので、先ほどの仮名のし過ぎというのは、文章が通らなくな

ることあるのかもしれませんが、形式的なところが中心になっております。

2点目ですけれど、これは、いわゆるメタ情報といわれるものを付加していただいた方が利便性が高まるということで、こちらにつきましては今後、最高裁との協議によって変わってくるところかと思えます。利活用機関の方で、こういう情報も付加情報として追加してほしいという要望を踏まえて最高裁に申し入れて、それについて御検討いただくということを予定しているところです。現時点で確実にこれは頂けるというようなところまではお話はできないのですけれども、例えば審級情報であれば実現可能性は高いということになっております。その審級の情報について、どういう形で提供を受けることになるのかということについても、裁判所のシステムにも関わるところなのかもしれませんが、まだ決まっていないところなので、できるだけつながりがはっきり分かるように一体にしてもらうのが良いのかもしれませんが、その辺はまだ詰めたところまでは検討ができてないところです。

異委員：

ありがとうございます。よく分かりました。

山本座長：

ありがとうございます。それではタニグチ委員お願いいたします。

タニグチ委員：

大坪先生、本日は貴重なお話を聞かせていただいて、どうもありがとうございました。日本電気のタニグチと申します。2点質問させていただけたらと思っております。1点目は、今回の実証実験の御説明の中で、機械で処理をして、その後は人の手でダブルチェックする体制を組まれるということだったのですけれども、先ほど仮名化のし過ぎの状況があるというお話がありましたが、やはり機械でやっても人の手でやられたとしても、何らかの仮名化漏れのようなリスクというのものもあるのかなというふうには考えているのです。先ほどの富所様のお話の中でも、仮名化しなければいけない情報が漏れることによって、このデータベースの存在意義そのものが問われてしまうような、重篤な事案も発生し得るのではないかというようなことも示唆されていたと思うのですけれども、財団PTの議論の中で、この辺りの仮名化しなければいけないところについての、何かの漏れがあったときのリスクというののはどのように評価されているのかというのを聞かせていただきたいというのが1点目の質問でございます。

もう1点目は、データベースの事業の在り方そのものについての中で御検討されていたのかもしれないですけれども、議論の中で、今回裁判所から一括して情報管理機関にデータが送られて、そこで仮名処理するということが想定されていると思うのですが、この運用が始まった場合に、今裁判所で行われているような、個別のそういった情報で公開するというような形での情報提供がなくなっていくということを想定されているのか、それともそういった場合の情報提供もある一定レベルでは行われていて、言ってみれば情

報管理機関で仮名処理をしたデータとそうでないデータ・情報というのが併存していくような形というのがあり得るというふうに想定をされて議論されていたのか、その辺りについて分かる範囲で教えていただければと思います。どうぞよろしくお願いいたします。

大坪様：

ありがとうございます。まず1点目ですけれども、仮名漏れなり仮名に問題があったときのリスクについては大分議論をしております。むしろ情報管理機関なり利活用機関としては、一定の基準に従って仮名処理をしたものについては、それが仮にプライバシー侵害に当たるようなことを当事者の方から言われても責任を問われないようにできないかという方法で考えていまして、そもそも大前提となりますのは、裁判の公開の下で判決自体は誰でも閲覧できるということになっておりますので、そういう前提からすると、名前や住所については仮名するのですけれども、そういうものを仮名した上で更にプライバシー侵害という形で損害賠償請求を求められても、それについてはやはり責任を問われるべきではないのではないかと思います。責任を問われることになると、利活用機関としても困るということと言われています。ただ、その点についてこれまでの検討では、法律で免責規定を設けるなどの手当てをするというのは難しいということなので、実際には個別の判断で損害賠償請求が立つかどうかというのを考えましようというような流れには今のところなっています。いずれにしても、逆の立場からするとプライバシーを侵害された側として、その点については何か対応を求める必要があるのかもしれないのですけれども、今のところは損害が発生したときに損害賠償という一般論で対応するというのが一つです。

さらに、必ずしも財団PTの中で議論したわけではないのですけれども、情報管理機関なり利活用機関で提供した情報について、ある程度何らかの権利侵害があるというようなことが分かった段階で、その権利者の方から不服申し立てをしてもらって、それについて早めに、例えば判決言渡しがした段階で、判決のこういう情報については公にしないでくれというような申入れもあり得るのかなと思います。それについて、そもそもそういったものを認めるかどうかということも議論はしたのですけれども、少なくとも個別の申入れの機会を設けることによって検討することで個別には対応することも可能かと思っております。例えば性犯罪等の特殊な事例について一般の基準ではやはり対応できないケースも少なからずあるだろうと思われまますので、そういうものについては、やはり個別に対応せざるを得ないと思います。議論の中では、そもそもそういう性犯罪等の特殊な事件については、情報管理機関なり利活用機関に、裁判所の方から情報を提供すること自体もやめるべきではないかという御意見もあったのですけれども、それはそもそも公開されている判決情報ということで、比較的少数な意見ということもあって、そういう個別なものについてはもっと別に法律の中で、例えば訴訟記録の閲覧を制限するとか、性犯罪なりDVの犯罪に関して名前とか住所を秘匿する制度というのが今回民事訴訟法改正で設け

られておりますので、そういうもので対応していくということを考えて、それらで対応できないものについてはまずは一旦情報管理機関なりでオープンにするという前提で作業しましょうということになっております。ですので、必ずしも完全ではないのですけれども、引き続き仮名漏れのときのリスクをどう除いていくかというのは検討せざるを得ないところではあるのですが、一応今のところの議論はそんな状況でございます。

タニグチ委員：

ありがとうございました。

山本座長：

ありがとうございました。それでは板倉委員お願いいたします。

板倉委員：

ありがとうございます。基本的には実作業する委託先の事業者にもまると委託したいということで、情報管理機関側としては特段パーマネントでは置かないという想定だったと思うのですが、本当にそれで大丈夫なのかなと思ひまして。今の異議申立ての問題を考えると、エスカレーションがないということはあり得ないと思うのです。その場合に御質問したいのは、本当にゼロの想定なのか、情報管理機関側で一人なのか二人なのか、異議申立てのコストをどのようにお考えなのか、今回の 4,300 万円にはそれは全く入っていないと思いますので、そこがどれくらいなのかということと、あとは一応相見積りを取っていただいたということなのですが、恐らくこれは最初注目されるので、どういふように委託先を選定したのか、入札等の透明性が問われると思いますが、その点は何か議論等がありましたでしょうかという、以上 2、3 点の質問であります。

大坪様：

まず、情報管理機関の体制に関しては、これはどこが情報管理機関になるか決まっていないところで勝手に議論しているところではあるのですけれども、責任者としては、弁護士を 1 人は、常駐するかどうかはともかくとして、何らかの管理責任者として置くということを想定しております。その管理責任者の費用については、確かにまだ計算はしていないところで、これがプラスになってくる可能性はあります。さらに、多分情報管理機関としても、先ほどは細かくは説明していなかったのですが、最終チェックのところがあるので、この最終チェックのところについては、資料 8 ページで「担当①」「担当②」がチェックして、「最終チェック」というところがありますけれども、これはどの程度までチェックするかというところは必ずしも決まっていないところですが、情報管理機関の作業担当者の方で、形式的なところを最後は見ることになる、そんなに細かく見ることは難しいかもしれませんが、ということも予定しております。そのところの費用についても、必ずしも今のところは計算できていないところになっております。

業者の選定につきましては、現在、相見積りを取っているところですが、これからその選定について再度 RFP を出した上で、実際にまた提案を出してもらい、その中で

ういうふうを選ぶかは決まっていらないのですけれども、恐らく今のところ、財団としては、財団の WG なり PT の中で承認を得るというプロセスになっていくかなというふうに思っているところです。若干気になるのが、当初は実証実験について、Legalscape 社というところをお願いしていたわけですが、実証実験をしているので、そのまま AI の仮名処理のところについては Legalscape に委託することが考えられています。仮名処理のシステム開発については、Legalscape が一歩先んじているのかなというところがあります。そもそも実証実験のときに Legalscape に頼んだことについての公平性のようなところも問題にはなり得たので、何社かに同じような実証実験をやってもらうということも考えられたところなのですが、それぞれやり方が違うこともあり、比較しようがないのではないかということで断念した経緯があります。その点、AI による仮名処理のシステム開発について、今後どういう形で入札的なことをやっていくかというのは課題としてあるところです。

板倉委員：

ありがとうございます。これは、マイナンバーも頻繁に入札とかで問題になるのですが、永久にやらなければいけない仕組みで、原則は元が税金なので毎年入札だということになるのでしょうかけれども、ノウハウが全然たまらない形で毎年業者が変わるのも困るので、本当は情報管理機関の方できちんとノウハウがたまる形が良いと思うのです。妙案があるわけではないのですけれども、簡単に引継ぎとかもできないなと思って、難しいなと思った次第です。

山本座長：

ありがとうございました。それでは中原委員お願いいたします。

中原委員：

中原です。貴重なお話をどうもありがとうございました。私も仮名化のところではなくて、付加情報のところについて伺いたいと思うのですけれども、利活用機関というのが現状想定されるようなデータベース会社であるとか出版社以外にも、様々な主体というのが想定され得るということで、先ほどの富所さんのお話の中でもそういった面が出てきたわけでありまして。そのときに情報管理機関からどのようなデータを得られるのかということは非常に重要で、実際問題その付加情報というようなものの有用性というのが、利活用機関となるのか、それともエンドユーザーになるのかの判断を分けるような気がします。

先ほどの異委員に対するお答えを伺っていたときに、基本的に裁判所と協議をするというふうにおっしゃっていたのですが、その具体的なことなのではけれども、情報管理機関として、自分たちで何か付加情報を付けたらというようなことは想定されるのか、そのような場合にどれだけのコストや時間がかかるのか、更に言えば、情報管理機関としてどこまでのことを担えるのか。利活用機関からのダウンロード要求のところでも様々なも

のが想定されるということだったのですけれども、そこである分野の事件についての付加情報をそのような形で付ければそのようなダウンロード要求も可能になるというような形で、その情報管理機関が出してくるデータによって使い勝手というのは変わってくると思うのですが、どこまでのことを情報管理機関として対応するのかということについて伺わせていただければと思います。以上です。

大坪様：

ありがとうございます。情報管理機関としましては、まず取組の当初の前提としては、既存のデータベース会社なり出版社の方の事業を圧迫しないということを前提としておりましたので、情報管理機関の方で裁判所から頂いたデータに何らかの付加的な情報を付けるのは、できるだけしないということを考えております。付加情報を付けるようなシステムを作ることにする費用面がかなりのものになるのではないかとこのように見込まれましたので、最低限のところでは付加情報としては考えていないところになります。最低限のところとしては、ある程度裁判所の方からも、事件類型的なところを大きくは分けていただけることを予定しておりますが、何らかの検索語によって、この検索語に該当する判決が選べるような形でのデータベース化までは、情報管理機関としては難しいのではないかとこのように考えている次第です。お答えになったでしょうか。

中原委員：

はい。どうもありがとうございました。

山本座長：

ありがとうございました。それでは米村委員お願いいたします。

米村委員：

度々すみません。先ほど実証実験の中身についてお尋ねしましたが、出典元の PDF ファイルを見ると詳細が分かるということをお塚委員から御示唆いただきましたので、そちらを拝見しまして追加のお伺いをさせていただきます。最初にお話を伺った際には、何故こちらの実証実験では、かかった時間とか人手の点だけを分析されたのか、そもそもの仮名化のクオリティなどはどうだったのかということの比較があってもよかったのに、何故それをされなかったのかということをお少し疑問に感じたのですが、出典元の PDF ファイルを拝見しますと、2020 年度の実証実験で、機械処理のクオリティについては既に検証済みで、それを受けて人手や実施体制の構築に向けた実証実験というものを今回されて、本日は後者の実証実験の御説明を頂いたということだったのかなというように理解いたしましたが、その理解でよろしいかというのが 1 点目でございます。

その上で、今日の何件かの御質問は、機械処理のクオリティに関わるものだったように思いますけれども、結局一般の方々が気にされるのは、この機械処理の仮名化で本当にきちんとした仮名化処理ができるのかということだろうと思います。仮名過多の問題には十分に対応できるけれども仮名漏れが防げないのだとすると、それはそれで非常に大

きな問題なのではないかと考える人がかなり多いようにも思いますので、少なくとも有意な仮名漏れはないのだということをお示しいただくことが大変重要になってくるように思うのです。その点を2020年度の実証実験で検証されたのかということをお示しいただければと思います。

大坪様：

ありがとうございます。2020年度の実証実験の結果についてもお話ししたかと存じますが、95%前後の精度という結果が出ていて、いずれにしても100%にならないということが分かっておりましたので、仮名化のクオリティについては、これ以上追求してもあまり意味はないのではないかとこのように考えておりました。その上で、人手でやらなければいけないということでしたので、人手で行う費用がどのくらいになるかということも2度目の実証実験でやったわけです。検討の中では、ある程度機械に頼る形も考えてはいたのですが、結論的には全件一通り人手で見ざるを得ないというふうに考えておりましたので、そういう意味で、仮名漏れについては、人手で最終的には防ぐしかないのではないかとこのように、あまりクオリティについて追求しなかったと、ほかの方がどうか分かりませんが、私の方では思っています。いずれこの技術が進歩すれば精度は上がっていくだろうというざっくりとした感覚はあると思います。

米村委員：

やはり、事案類型とか、どういうタイプの判決なのかということによって、機械にとっても人間にとっても、仮名化のしやすいものとしにくいものがあるはずなのですね。実証実験というからには、どういうタイプの事件を何件選んで、比較の結果、どのようところに仮名漏れが起こっているという結果が出てきたのかが非常に重要で、例えば貸金返還請求訴訟とか過払金請求訴訟のような、件数が非常に多い事件での精度が95%だったというのなら、ほかの事案はもっと低いのではないかとこのことにもなりますし、ほかの類型、例えば損害賠償請求訴訟などではいろいろな事案が雑多に入っていますので、そういうものを全部合わせて95%だったということだとすると、かなり定型化された、件数が多くてはっきり確立されているような事件類型であれば、もっと精度は高く、機械処理のみでも実用に耐える可能性が出てくるように思います。しかし、そういう細かいデータがないと、この問題は検討できないのではないのでしょうか。全件当然にこの仕組みを導入することにはなりそうもないですし、あるいは全件必ず人手で確認しなければならぬということも、さすがに実務的にもたないのではないかと思います。したがって、どの範囲であればこの仕組みがうまくいきそうなのかということを実証実験で確かめていただくのが本来は望ましいと思いますし、そのことが分かるデータの出し方をさせていただきたいというのが私の希望です。

大坪様：

貴重な御意見をありがとうございました。実証実験の対象となる事件類型を選ぶとき

に、比較的偏りないような形で広く事件類型は選ぶようにはしたのですが、米村先生がおっしゃるような形での深い追求はしていなかったもので、その点は今後こちらでも検討をさせていただきたいと思います。

山本座長：

ありがとうございました。それでは小塚委員お願いいたします。

小塚委員：

小塚でございます。すみません、先ほど私の質問は既に終わっているのですが、板倉先生が質問された時に再審査の話をして、大坪先生の方から弁護士が最低 1 人そこにいて対応するのだという御回答があったのですが、私の記憶している PT か WG の議論では、審査委員会のようなものがあって弁護士や研究者等の有識者の方たちが最終的に決めるということで、もちろん有識者委員会は無給でやりますと言えばコストには影響しないわけですが、いずれにしてももう少し重い仕組みを予測していたのではないかと思います、私の記憶が違うのかどうか確認させてください。

大坪様：

ありがとうございます。小塚先生がおっしゃるとおり、判断に迷うケースについては財団側の管理責任者だけでは対応できず、外部の何らかの有識者の組織、調査委員会的なもので判断していただくということを想定してはいます。

小塚委員：

そうですね。ですので、まず事前のところ、最終的に作業員が判断できなければ、弁護士の先生お一人の対応なのかもしれない。そういう部分があるのと、事後の修正というところで、事後の修正も過多・過少両方あるということだったと思いますけれども、そこはもう少し仕組み的なものがあるという理解ですね。そこは記録にとどめたいと思いました。ありがとうございました。

大坪様：

ありがとうございます。

山本座長：

ありがとうございました。それでは鹿島委員お願いいたします。

鹿島委員：

本日は貴重な御報告を頂きましてありがとうございます。司法書士の鹿島と申します。先ほどの費用感のようなところの話の中で具体的な数字がいくつか出てきたところなのですが、1 億いくらというのがインシヤルコストとしてということでもよろしいのでしょうか。となると、ランニングコストの部分が、仮名処理の部分が 4,400 万円プラス最初のコストとしてその値段ということになると理解をしているので、ここに示されている、先ほどおっしゃっていた開発の部分の数字に関しても、基本的には仮名処理の部分は

AI のツールの部分だけということによろしいのでしょうか。

大坪様：

いえ、そこにはAI のツールと裁判所からもらったものを最終的に提供するまでの、全体のシステムの開発にかかる費用も含んでいます。

鹿島委員：

承知しました。この4,400万円の部分に関しては、あくまで仮名処理の部分のみにかかるコストということで間違いはないということですか。

大坪様：

はい。

鹿島委員：

承知しました。この提供を受け続けることに関するコスト感というところは、試算はされているのでしょうか。

大坪様：

システム開発の中に含んでおりますので、裁判所から自動的に、毎日か週に何回かその日に出された判決の提供を受ける形のシステムを開発することになります。裁判所のシステムとどうつながるかという課題はあるのですけれども、システムの中でやるということを考えています。それについてはシステム開発の中の費用に含まれています。

鹿島委員：

では、以後はかからないということでしょうか。

大坪様：

はい。

鹿島委員：

承知しました。ありがとうございます。

山本座長：

ありがとうございました。それでは板倉委員お願いいたします。

板倉委員：

ありがとうございます。今の鹿島先生のお話と私の先ほどのお話の追加のようなところなのですが、AI のプログラムそのものの著作権は絶対にこの情報管理機関もしくは国の方で持たないといけません。今回は元から持っているところをお願いしたのでいいのだという話でしたけれども、著作権は絶対にくれませんので、AI の開発コストが必ずイニシャルでかかってきます。このAI を、毎年若しくは何年かで変わるかもしれませんが、委託先にライセンスして使うというような形でやらないと、AI を持ってないとなると全く駄目ですし、ライセンスを受けて使うというのも安定性の観点からあり得ないと思いますので、AI をゼロから学習させて開発するコストを乗せておかなければいけないと思

いました。これは絶対にこちらで著作権を持っていないと、永久に動かすシステムですし、永久にそれを学習させていかないといけない話ですので、そこがもし抜けているとしたら、AI の開発コストはライセンスでは駄目で確実にこちらで持つ必要があります。なので、もう一回やるのだということになるかもしれませんが、既にあるので事実上複製の分はくれるということでしたらいいかもしれませんが、しかしながらそこは絶対に、ほかの知財や権利の処理も含めてイニシャルで乗ってきてかなりかかると思いますので、そこは見込まないといけないなと思いました。以上です。

大坪様：

ありがとうございます。その前提で今後交渉をしたいと思います。今までそういう形で話はしていたのですけれども、若干曖昧なところもあったのかもしれませんが。

山本座長：

ありがとうございました。それでは宍戸委員お願いいたします。

宍戸委員：

宍戸でございます。非常に多くの御指摘があって、それをまとめるとか確認するということもできないのですけれども、お話を伺っていて懸念がいくつかあります。その懸念との関係で言うと、一つは、仮名化についてのアルゴリズム、ソフトウェアを開発してやっていくということが、ある種の標準的な、少なくとも今回実証事業としてやっていただいたものは分かるのですが、本当は長い判決文であるとか、要注意な判決とそうでないものを最初に仕分けするとかいった作業が実は必要なのではないかと思います。そこは裁判所様の方でフラグを立てていただくとかいったことも当然あると思うのですけれども、管理機関の方で、例えば判決をまず1回目にばっと読ませて、これはかなり丁寧にやらなければいけないものですか、これは典型的に処理できるものですかといったものをはじくようなことを人間がやるのか、そこ自体AI がやるのか、そういった手間の部分、最初のスクリーニングが実は必要なのではないかとこのことを思いました。というのがまず1点目の意見でございます。

2点目は、ある種のランニングコストのような、今想定されているあまり難し過ぎない事件などから積算すると、20 万件の場合これくらいだろうということは今回計算していただいたわけですが、それ以外にも様々なコストがあるということが分かってきているわけですね。管理コストのようなものもありますし、AI の開発コスト、あるいはAI の改善のようなコストもあるだろうということもあります。また裁判所からデータを頂くという場合の、そこが専用線なのかとか、どういう形でもらうのかといったコスト、それから実際に手を動かして仮名化をしてくださる人を、どういう場所で、どういうセキュリティ体制の下でやっていただくのか。そこが上がれば上がるほど当然コストもかさんでいくわけでありまして。それからもう一つあるのは、最初のオリジナルの判決データもらって仮名化した判決データを作る、データベースを作ることによって、その後もオリ

ジナルのものをずっと持ち続けるのか、あるいは作業中のデータについて、しっかり削除するのか、取っておくのかといったことによってもコストはかなり変わってくると思います。ということで、今回の真ん中の、仮名化に伴うランニングコストの部分は計算していただいたわけですが、それ自体精査が必要だという、先ほど来出てきている論点とは別に、これ以外にどんなコストがあり得るだろうか、それについてどれだけの水準のものを要求することによってどれだけコストがかかってくるのだろうかということの洗い直しがどこかで必要であり、それは法務省でやっていただくのか、または財団様にお願いするのか、やはりその作業が必要かなというふうに思いました。

3点目は、仮名化はとりあえずしたということですが、推知情報をどうするかという問題が残っているだろうと思います。推知情報についても、やはり危ないものはある程度管理機関の方で作業していただくというスキームにするのか、当然そうなりますと、もう1枚法律家の方が入って1件1件丁寧にチェックするといったことが必要になるかもしれないと思います。最初にある種の前さばきをやることによって、その部分が軽減されるという部分もあるかもしれませんが、そういったことが起き得ると思います。そうではなく、もう推知情報については見ないで、後は利活用機関に渡してそちらで作業していただき、そこで問題があった場合は利活用機関が責任を負っていただくという考え方も当然あり得るわけですね。ここら辺の見合いが一つ考えどころかなと思います。それぞれに伴うリスク、コスト、メリットを整理した方が良いかなと思います。

最後に4点目、仮名化のAIないし著作権をしっかりと押さえておくということと同時に、仮名化の作業を実際に人間がAIを使って、その後も人間が見るということについてのしっかりとしたマニュアルとか、人が入れ替わってもできるようにする研修の体制、またそれに伴うリクルーティングやその費用もしっかり計算しておくということがサステイナブルな仕組みを考える上で必要だなということを改めて感じたところです。

以上、非常に貴重な御報告を頂いて、更にそれに基づいて多くの構成員からいろいろな御指摘がありましたので、私の方で考えたことを申し上げたところです。ありがとうございます。

大坪様：

大変貴重な御意見ありがとうございました。

山本座長：

ありがとうございました。それでは事務局からお願いいたします。

事務局：

渡邊です。先ほど仮名処理のクオリティのことが話題になりましたので、若干事務局からも補足をさせていただきます。大坪弁護士御提出の資料で主に言及されていたのは、PTで実施された2回の実証実験のうちの2回目の方ということになります。先ほど話題になりました仮名処理のクオリティのことですが、これは1回目の実証実験で検証された

ものでした。本日の資料では、その点について詳細に言及されておりませんでしたので、少し皆様に御懸念が生じてしまったのかなと思いついて伺っておりました。この件は、おつて大坪弁護士と協議させていただき、次回以降資料として追加させていただけたらと思つます。

私の手元にその資料がございますので簡単に御紹介しておきますと、1回目の実証実験では、機械処理の精度が現在の技術水準でどの程度のものかというところを検証したわけですが、そのために用意された裁判例は1,600件で、偏りのないものということで選別されたものであります。これを用いて実験が行われたのですが、機械処理の精度だけでなく、先ほどから問題になっている仮名漏れ・仮名過多でございますけれども、どういったものに漏れや過多が生じやすかつたのかということの分析もしていただいております。例えば仮名漏れで申し上げますと、一番多かつたのが人名でございます。続いて企業名、施設名、こういったところに漏れが生じやすいという結果が出ております。仮名過多につきましては、これもやはり企業名、施設名が多くございまして、続いて人名が多いという結果になっております。その上で、どういった類型の事案にこういった漏れや過多が生じやすいのかというところは更に分析を進めていく必要があろうかと思つますが、詳細については実証実験の報告書に記載がございますので、次回までに提出させていただければと思つております。以上でございます。

山本座長：

ありがとうございます。資料の追加を頂けるということですので、よろしくお願ひいたします。それでは湯浅委員お願ひいたします。

湯浅委員：

湯浅でございます。先ほど来、各委員の先生方からコストとAIの著作権の問題がございましたので、若干追加的に申し上げさせていただきますと、セキュリティの部分は正直申しまして、どの程度のセキュリティリスクを想定して、それに対してどういうシステムにするかということによって全然変わってきますので、非常に強固に24時間365日のネットワークの監視をすつとか、裁判所とこの機関との間の判決文のやり取りのところも専用線を引くとか、そのレベルになってくつとそこだけで年間数千万というところになってくつと思つます。あるいは、加工前の生データを20年程度保存する予定というお話でございましたが、ここをオンラインのストレージにするのか、インターネットと切り離した媒体に物理的に強固に保存するのかによつても変わってきますので、これは日弁連にお願いすべきなのか法務省にお願いすべきなのか、どの段階でどの程度のセキュリティレベルとリスクを許容するかというレベルをある程度考えないと、費用の試算もしようがないのかなという気がいたしました。

2点目は、板倉委員が御指摘の問題は、正におっしゃるとおりではあると思つます。他方で、今自治体を中心に、急速にAIの利活用が行政サービスにおいて進んでいるわけですが、総務省が自治体向けに出しているガイドラインにおきましても、必ずしも人

工知能あるいはASP型やSaaS型のサービスについて、知的財産権等の権利を自治体側が完全に持つようにということは言っておりません。権利の在り方には留意する必要があるというところでとどまっています。それは、恐らく従来のシステムと同じように、人工知能やASP型、SaaS型の知的財産権を地方自治体側に持たせると、ものすごく巨額の費用になるからなのだと思います。他方で、板倉委員の御指摘のように、継続的、安定的なサービスを提供しようと思ったら永久にやるのだからという御指摘がありました。そのとおりではあるのですが、こちら側で著作権を持つとすると、恐らくその部分の費用の回収を、いわばエコシステムとしてこの判決データベースのオープン化のところで、年間の利用料ベースとかで回収していくということは困難だと思うのです。したがって、板倉委員が御指摘のように、仮名化等を行う部分のAI等の権利を完全に持つという形でこの制度を考えていくのであれば、その部分の費用は別に切り出さないと、民事判決をオープン化するというコスト試算が全く変わってきてしまうのではないかと懸念しております。もちろん、板倉委員がおっしゃるように、権利として持つのは望ましいというのは重々理解してはおりますが、コスト面と全体のエコシステムとの兼ね合いということをご考慮する必要があるのではないかとございます。

山本座長：

ありがとうございました。ほかにはよろしいでしょうか。長時間にわたりまして大変有益な御議論をいただけたかと思えます。今日の大坪先生からの御報告でかなり具体的なプロセスあるいはコストを含めたイメージというのが共有できたのではないかと思います。そしてそれは、今後の我々の制度の議論の前提として非常に有益だったということかと思えます。逆に、我々がこれから議論していくこの制度をどうするかというのが、また今のお話のコスト等に対しても影響を与えていく面が当然出てくるのだろうというふうに思いますので、相互にまた情報を共有しながら議論を進めていく必要があるのかなと思えました。いずれにしろ大変有益な御意見を多々頂きましてありがとうございました。大坪弁護士にも長時間にわたりお付き合いいただきましてありがとうございました。

大坪様：

ありがとうございました。引き続きよろしく申し上げます。

山本座長：

それでは、今日はもう一つ、海外調査の追加報告という点がございます。事務局の方からお願いいたします。

事務局：

渡邊です。第3回会議ですけれども、日本弁護士連合会から海外調査の御報告を頂いたところでございますが、本日は追加の御報告があるということで杉村委員に御準備いただいております。御発表いただきまして、その後質疑応答に御対応いただきたいと考えております。以上です。

山本座長：

それでは杉村委員お願いいたします。

杉村委員：

よろしくお願いいたします。それでは始めさせていただきます。2回目の調査報告になります。

スライド2を御覧ください。調査対象の国および地域は、今回は4か国について御報告させていただきます。フランス、ドイツ、英国、イングランド及びウェールズですね、アメリカ合衆国については、連邦及びカルフォルニア州をピックアップしています。

スライド3を御覧ください。概要なのですが、前回の御報告と同様に、この四つの国に関しては、各国の政府機関や弁護士会等に日弁連の方から書面で照会を出しました。ですが、きちんと書面で回答が返ってきたのが残念ながらドイツだけになっておりまして、ドイツは連邦弁護士会からきちんと御回答を頂くことができました。フランスに関しては、日弁連の会員で、フランスの弁護士資格を持って今フランスで仕事をされている方が国際室の연구원としていらっしやいまして、宮田さんとおっしゃいます。宮田さんに調査の依頼をいたしました。あと、イギリス・アメリカに関しましては、国際室の嘱託が公になっている資料であったり、自分の留学経験であったり、現地にいる弁護士に対しての聞き取り調査等をして、調査をしたという結果になっております。少し不十分な点もあるかと思いますがお許してください。

スライド4を御覧ください。前回と同じような表になりますので、御説明を簡単にさせていただきますと、まず民事判決を公開しているかどうかの点に関しましては、いずれも公開はしているということでございます。ただし、裁判所のウェブサイトにおける全件の公開なのかどうかというところは、いずれの国の回答も調査も、全件をしているというわけではないという回答となっております。後ほど詳しく見ていきます。仮名処理については、全くしていないという所はなくて、フランスとドイツは基本的に行っているというような回答が返ってきていますが、アメリカとイギリスについても一部の事件は仮名処理がされているものがあるということになります。裁判所から全判決を取得する機関や団体があるかどうかに関しましては、全判決をとるところはそもそもの国にもないという形ではありました。ただし、こちらに載せさせていただいているような固有名詞が出ているものは、国として、行政機関であったりオフィシャルのものとして、かなり判決情報を取得して公開するということをやっている団体がございます。こちらも後ほど詳しく説明をさせていただきます。そういうものを通じて言いますと、裁判所からの取得費用については、無償であるとか、無償と推測されるものがあつたということになります。ただし、ドイツの弁護士会からの回答としては、行政機関による運営サイトは無償で裁判所から取得しているけれども、民間の会社がどうしているかというのは弁護士会では分からないという回答となっております。また、判決を取得して公開している機関や団体の利用に関しては、いずれも一般国民が利用可能ということで、利用に資格制限はないという

ことでした。利用料については、民間の有料サービスがアメリカではすごく発達しているわけなのですが、こちらに名前を挙げさせていただいたような機関は、いずれも無償であったり、無償と推測されるということです。法令上の根拠は、ある国とない国があるということです。仮名処理については、裁判所から取得する段階で、裁判所の方で仮名処理がされているというのが基本的なもののごようでした。ということで少し詳しく見ていきたいと思えます。

スライド5を御覧ください。先ほどの、裁判所からオンラインで開示されているものが全部か一部かについては、基本的にどこも一部であるということになっておりまして、フランスについてはその詳細が分かりませんでした。ドイツからの回答については、裁判所の決定総数のうちのごく一部だということだったのですが、何かカテゴリー分けとかがあるわけではなく、裁判所や審級によってまちまちというような回答でございました。イギリスについては一部でありまして、最高裁と貴族院に関しては、サイトは見つかったものの、他の裁判所では見つからないものもあったというところでした。アメリカに関しては、連邦裁判所なのかそれぞれの州の裁判所なのかということによっても大分変わってきたりするようです。連邦最高裁判所については、まず1754年～2012年の判例というのは公式の判例集に全て掲載がされていて、それがウェブで掲載されているという形になっていて、それ以外のものについても、他の民間サイトなどで調査可能という旨の記載があるようでして、連邦最高裁判所の判例については、基本的には全て出ているような状況とのことでした。それ以外の連邦裁判所の判例については、アメリカの連邦政府のウェブサイトで一部が公開されているということです。州に関しては、カリフォルニア州を調べてくださったのですが、カリフォルニア州の最高裁の判例というのはウェブサイトで公開されていて、公式判例集への掲載というのは全件ということなのですが、公式判例集の全件掲載のものが全件オンライン化になっているかは確認ができなかったということです。あとは州の控訴審の裁判所等はウェブサイトを持っていて、一部公開されているような状況ということで、かなり地域によってもまちまちということが分かりました。

スライド6を御覧ください。次が仮名化処理に関して、です。フランスについては法令等もございませんし、デクレというものできちんといろいろなことが決められていて、判決に機密又は個人データが含まれている場合には、それは匿名化するというのが基本ということです。あとは、どういった情報かということで具体的に分かったものとしては、政府の審議の秘密であったり、国防又は国家安全保障以外に、プライバシーの保護であったり医療上の秘密、営業上の秘密、自然人に対する評価、価値判断、人の行動を明らかにする場合などがございまして、名前という以外にもかなり広くいろいろ伏せるという作業がされているようです。ドイツに関しましては、データ保護規制というもので、機微情報が公表される決定から除外されているということになっております。そのため、基本的に判決が公開される時に原告・被告という語が用いられていて、名前ではないというような回答が戻ってまいりました。あと、イギリスについても仮名化処理が行われているとい

うことだったのですが、詳細が分かりませんでした。ただ、これも詳しい先生がいらっしやったら後でフォローしていただければと思いますが、イギリスでは、個人情報の保護義務というのが裁判所には及ばないとされているのだというような記述に触れました。ただ、その更に先に行った、裁判所から情報をもらった人は個人情報保護義務が課せられているという話が載っておりまして、その辺りの個人情報の法制度の関係は、イギリスでは特殊なのかもしれません。アメリカに関しましては、皆さんの御想像どおりかもしれませんが、多くは仮名処理がされずに公開されているということなのですけれども、見ていくと仮名処理がされている裁判例も見つかります。現地の弁護士にヒアリングをしてもらったところ、少なくとも州レベルでは、多くの州で非公開についてのルールがあるようだということでした。こちらに関しては後で詳しく御説明いたします。

次が「審級や事件類型を問わず、全ての民事判決情報を裁判所から取得する機関・団体はあるか」というスライドになります。先ほどの表であったように、全てをごそつともらっている団体は、いずれの国もないということではありましたが、いくつか先ほどの表で名前を出させていただいた団体がありますので、説明させていただきます。フランスは Legifrance というものがございまして、こちらはフランス政府が開設したデータベースのウェブサイトになっております。裁判所ではなく、行政サービスの一環としてのウェブサイトだということなのですが、こちらに法令情報、官報などといったものがデジタルで載っているものになります。こちらに判決の情報というのも集めて公開するような仕組みが進んでおります。ドイツにつきましては、連邦法務省と地方法務局が共同して運営している司法ポータルサイトというものがございまして、こちらの司法ポータルサイトの方にはできる限り多くの判決を集めて公開するというようなことになっております。イギリスは、前の海外報告で LII のお話をさせていただいたかと思いますが、イギリスについても LII 運動というのが活発であったようでして、こちらについてはかなり判決情報の公開について公的な役割を担われているということです。ただ、これとは別に、2022 年に国の機関の方で、判決情報の開示というのが本格的に開始されたそうです。こちらはまだ始まったばかりですので、全件を出すには至っていないということなのですけれども、こちらの方の動きも出ているようで、判決情報の公開というのが進んでいるということです。続いてアメリカについては、無料で利用できるオフィシャルなものとしては、例えばハーバード・ロースクールが、ロースクールの図書館の蔵書である公の判例集をスキャンして無料公開するプロジェクト等を進行させているようでして、手元にあるものをどんどん開示していくという動きをされているそうです。

続いては、「裁判所から民事判決情報を取得する機関・団体がある場合の名称は何か」ということで、先ほどのものとは別の、民間的なものを載せてございます。フランスについては、民間の業者がたくさんおりまして、LexisNexis のようなものとか、あとは法律系のスタートアップ企業等もかなりあって、オープンデータとして公開されている判例情報を利用しているそうです。ドイツについては、先ほどの司法ポータルサイト以外にも、

無料のデータベース等があったり、有料のデータベースとして有名なものがあったりということで、他の団体もあるということです。すみません、イギリスは先ほどと同じことを書いてしまっているのですけれども、アメリカに関しては、有料の Westlaw とか LexisNexis が有名かと思うのですが、こちらにも全件掲載しているかということに関しては、全ての裁判所全件という情報は見当たらなかったということです。あと、トムソン・ロイターの方が判例の無料公開をやっていたらいいということでした。

スライド9を御覧ください。全体的な御報告は以上なのですが、調べている中で特にピックアップした方が良いかと思うものがございました。まず一つ目は、ドイツにおける匿名化処理についてです。先ほどドイツについては、「データ保護規制により、機微情報は公表される決定から除外されていて、原告、被告という語が用いられています」という御報告をさせていただいたのですが、これがどういう仕組みでやっているかについても弁護士会の方から回答を頂いております。そちらについては、「判決の公表方法に関する統一的な義務や規制はない」ということなのですが、逆に匿名化等については連邦州によって異なっていて、同じ地方の裁判所によっても実務運用が異なっているということもあって、裁判所によって随分違うようです。しかも、決定を公表するかどうか、またどのような形式で公表するかは裁判官自身が決定していて、裁判官が決めている点があるという回答がありました。

スライド10を御覧ください。もう少し具体的にどういったものが消されているのかということがこちらのスライドにございます。関係当事者のフルネームが記載されていることはほとんどないという回答でした。その他の詳細についても、多くがチェックされ、通常は匿名化されているということで、例えば、ニーダーザクセン州では、実名のほかに、法人名であったり、法人の代表者名、地名、生年月日、行政当局のファイル番号、労働組合または非営利団体の場所、通りの名称、団体名まで匿名化されているそうです。実際の匿名化作業を誰がやっていますかということに関しては、裁判官が自らやることもあるそうなのですが、多くの場合は裁判所の職員がやっているということです。図書館スタッフや報道担当官がこの作業を行う裁判所もあり、ヘッセン地方の裁判所では、集中的に匿名化を行う文書センターというものを設けていて、そちらで実際の作業をされているということでした。何を匿名化するかというのは、今見ていただいたようになりに違いますが、やはり保護に値する個人の権利次第なのだとすることを高等地方裁判所の報道スポークスマンは強調されているようでして、原則としては、当事者の氏名とか名称、当事者が誰であるかを導き出すことができるような推知情報のようなものも同様に匿名化されるということが原則のようです。

スライド11を御覧ください。続いてアメリカについては、先ほど申しましたように匿名化処理されずに公開されているというのが基本ではあるものの、一部匿名化処理がされています。ユタ州の弁護士にヒアリングした結果、州レベルでは多くの州で、一定の種類的事件については判例を公開しないというルールがあります。また、未成年者の氏名は、

プライバシー保護のためにイニシャルにされます。あと家事事件は非公開とされているというような回答があったそうです。

スライド12を御覧ください。実際に判例検索サイトなどで検索しますと、一方の氏名が“Jane Doe”や“John Doe”というように、日本でいうところの甲野太郎さんみたいな感じですかね、そういうふうになっているものはたくさんあるということです。ただ、調べていただいた方の御指摘としては、そもそもアメリカの訴訟制度では、プライバシー保護・当事者への危害防止ということで、「仮名での手続進行の申立て」というものが裁判の段階であります。ですので、判決が出る前の段階で、仮名で裁判ができるということになっているので、判決になった段階でもう仮名化処理といたしますか、本人の名前が判決書に載っていないという手続があるということで、そうするとそれが公開されたときに、公開に当たって仮名化処理したのか、元々の裁判で仮名になっていたのかというところの区別がつかないので、仮名で公開されているものの中には、そもそもの手続の中で仮名化をされているものもあるだろうという御指摘がありました。こちらは、日本でもちょうど2月20日から、氏名の秘匿の制度がきちんとスタートいたしまして、先ほど富所様から御指摘があったようなDV被害者だったり性犯罪の被害者だったり、相手方の当事者に名前とか住所を知られたくない方は、そちらがそもそも裁判手続の中で秘匿されるというケースもありますので、仮名化処理する以前に仮名になっている事件というのものもあるということも念頭に置いておいた方が良いのかなというふうに気付きました。私からは以上です。

山本座長：

ありがとうございました。大変詳細に御報告を頂いたかと思えます。恐らく皆さんからいろいろ御質問等あろうかと思うのですが、既にほぼ予定されている時間に達してしまっていますので、御質問がおありの委員は、法務省事務局を経由して杉村先生の方に質問を投げさせていただいて、お答えいただけるものは杉村先生の方でお答えいただき、それを委員全員で共有するという形にしたいと思えますがいかがでしょうか。法務省の方はそれでよろしいでしょうか。

事務局：

渡邊です。座長御提案のように進めさせていただきたいと存じます。

山本座長：

では杉村先生、恐縮ですけれどもそのような形で質問させていただければと思いますので、よろしく願いいたします。

杉村委員：

はい。お手数をお掛けいたしますが、よろしく願いいたします。

山本座長：

ありがとうございました。それでは、本日の審議はこの程度というふうにさせていただきます。

ければと思います。事務局の方から、今後の日程等について御説明いただけますでしょうか。

事務局：

渡邊です。次回以降の会議の予定については、会議用資料として配布いたしました資料のとおりとなります。議事の詳細は追って事務局から連絡を差し上げたいと思います。以上です。

山本座長：

ありがとうございました。本日も大変長時間にわたりまして熱心な議論をいただけたかと思えます。引き続き審議、御協力のほどどうかよろしく願いいたします。それでは本日はこれにて閉会としたいと思います。ありがとうございました。