民事判決情報の仮名処理の在り方等に関するＷＧ（第１回）

令和３年７月２１日（水）
１７：３０～１９：００
弁護士会館１６階来賓室

議　事　次　第

１　冒頭挨拶等

２　仮名処理の実情について

３　意見交換

４　次回以降の日程について

（配布資料）
１　民事判決情報の仮名処理の在り方等に関するＷＧ設置紙
２　民事判決情報の仮名処理の在り方等に関するＷＧ構成員及び出席者（第１回）
３　民事判決情報のオープンデータ化に向けた取りまとめ（2021 年 3 月 25 日）
４　自動仮名処理に係る実証実験結果（2021 年 1 月 26 日ＰＴ資料）
５　民事判決の仮名処理等の在り方に関するＷＧ・論点整理案
６　最高裁判所説明資料
７　第一法規株式会社説明資料
８　ウエストロー・ジャパン株式会社説明資料
９　株式会社ＬＩＣ説明資料

参考１　民事判決データベース化事業の在り方に関するＷＧ設置紙
参考２　民事判決オープンデータ化検討ＰＴ2020 年度活動報告記事（JLF NEWS 抜粋）

資料１

民事判決情報の仮名処理の在り方等に関するＷＧ

１　目的

「民事判決情報のオープンデータ化に向けた取りまとめ」を踏まえ，民事判決情報の適正な利活用に向けて，情報管理機関において実施すべき仮名処理の在り方等について更なる検討を加えること等を目的として，当ＷＧを設置する。

２　構成員

【座長】

山本和彦一橋大学教授

【メンバー・五十音順】

小塚荘一郎学習院大学教授
小町谷育子弁護士
菰田優日本弁護士連合会前事務総長
新堂明子法政大学教授
高須順一日弁連法務研究財団常務理事
中原太郎東京大学教授
町村泰貴成城大学教授
湯淺墾道明治大学教授
米村滋人東京大学教授

【オブザーバー】

内閣官房
法務省大臣官房司法法制部
法務省民事局
最高裁判所事務総局

【事務局＝日弁連法務研究財団】

大坪和敏日弁連法務研究財団事務局員
藤原靖夫日本弁護士連合会事務次長

３　その他

ＷＧは，令和４年２月を目途に，議論のとりまとめを目指す。
ＷＧの事務は，日弁連法務研究財団が，日本弁護士連合会の協力を得て行う。

資料２

民事判決情報の仮名処理の在り方等に関するＷＧ

構成員名簿及び出席者一覧（第１回）

２０２１年（令和３年）７月２１日

◎：ＷＧ座長

	構成員	出席者（第１回・7/21）
◎※	山本和彦（一橋大学教授）	山本和彦（一橋大学教授）
	小塚荘一郎（学習院大学教授）	小塚荘一郎（学習院大学教授）
※	小町谷育子（弁護士）	小町谷育子（弁護士）
※	菰田　優（日本弁護士連合会前事務総長）	菰田　優（日本弁護士連合会前事務総長）
	新堂明子（法政大学教授）	新堂明子（法政大学教授）
	高須順一（日弁連法務研究財団常務理事）	高須順一（日弁連法務研究財団常務理事）
※	中原太郎（東京大学教授）	中原太郎（東京大学教授）
※	町村泰貴（成城大学教授）	町村泰貴（成城大学教授）
※	湯淺墾道（明治大学教授）	湯淺墾道（明治大学教授）
※	米村滋人（東京大学教授）	米村滋人（東京大学教授）

	オブザーバー	出席者（第１回・7/21）
	内閣官房	坂本三郎（内閣審議官）
※	法務省	渡邊英夫（法務省司法法制部参事官） 脇村真治（法務省民事局参事官）
	最高裁判所事務総局	石井芳明（最高裁判所事務総局総務局第一課長）

事務局	出席者（第１回・7/21）
日弁連法務研究財団	大坪和敏（事務局員）
日本弁護士連合会	藤原靖夫（事務次長）

※　Web 会議システムによる出席者

資料３

民事判決情報のオープンデータ化に向けた取りまとめ

令和３年３月２５日

公益財団法人日弁連法務研究財団

第１　はじめに

21 世紀を迎え、法律問題は一段と複雑多様化、高度化、国際化の様相を強めている。このような現状に的確に対応するため、広く社会に開かれたシンクタンクの機能を有する総合的な研修・研究機関の創設・活動が望まれる中で、当財団は、開かれた研究機関として、各種研修・研究、法情報の提供を実践し、もって法の支配の確立に寄与することを目的として設立された。

後記のとおり、全ての民事判決情報をデータベース化し、社会全体の共有財産として活用する要望が高まっている現状において、裁判所や法務省といった公的機関だけではなく、民間においても広く法曹界及び関係団体の声を集め、適切に議論を進めることが必要であり、その議論の場を提供することは、当財団に課せられた重要な役割であると考える。そのような観点から、当財団は、「民事司法制度改革の課題の一つとして、民事裁判手続のＩＴ化に伴って、民事判決情報の活用拡充（判決情報のデータベース化を含む。）のニーズ・活用可能性が高まっていることを踏まえ、民事判決のデータ管理・活用を担う民間組織の立ち上げも視野に、民事判決データの管理及び利活用に当たり検討すべき課題・対応策について、幅広い観点から、実務的協議を行う。」ことを目的として、「民事判決のオープンデータ化検討ＰＴ」（以下「本ＰＴ」という。）を立ち上げた。本ＰＴの委員構成等は別紙のとおりであり、この分野に造詣の深い研究者や判例雑誌社・判例データベース会社等のほか、内閣官房、法務省、最高裁判所事務総局といった関係省庁等の参画を得て、令和２年３月２７日から本年３月２５日まで合計９回の会議が開催された。民事判決情報のオープンデータ化の実現に向けて、後記の Legalscape 社（以下「ＬＳ社」という。）が実施した実証実験の結果を踏まえつつ、幅広い観点から検討を加え、今後解決すべき課題の整理を行い、本取りまとめに至ったものである。

第２　ＡＩ・ＩＴ時代における民事判決情報の利活用に向けて

１　民事判決情報のデータベース化の意義

⑴　民事判決情報を国民・社会の全体で共有することの意義

ア　民事判決情報は、訴訟当事者間に生起した法的紛争についての裁判所の判断が示されているものであるが、その成果を広く国民や社会の全体で共有することにも重要な意義がある。古くは、平成１３年６月１２日に取りまとめられた司法制度改革審議会意見書において、「判例情報の提供により、裁判所による紛争解決の先例・基準を広く国民に示すことは、司法の国民に対する透明性を向上させ、説明責任を明確化するとい

うにとどまらず、紛争の予防・早期解決にも資するものである。」と指摘されていたところであり、また、最近では、民事司法制度改革推進に関する関係府省庁連絡会議が令和２年３月１０日に取りまとめた「民事司法制度改革の推進について」において、「民事判決情報は、国民にとって、紛争発生前には行動規範となるとともに、紛争発生後には当事者による紛争解決指針の一つともなり得るものであり、社会全体で共有・活用すべき重要な財産である。将来的に、ＡＩによる紛争解決手続のサポートの可能性があり、その活用が国家経済の活性化にもつながり得るものであることも踏まえると、現状、先例性の高い事件や社会的に関心の高い事件等の一部の事件に限定して一般に提供されている民事判決情報については、今後、より広く国民に提供されるべきである。」とされたところである。

イ　このように、民事判決情報は、紛争当事者だけでなく、国民や社会の全体で共有すべき公共財ともいうべき重要な資産であり、これをデータベース化した上で、広く国民や社会の利用に供することは、①司法の国民に対する透明性を向上させ、②国民に対して行動規範・紛争解決指針を示すとともに、③紛争解決手続に関するＡＩの開発等の研究を推進するための基盤ともなり得るものと考えられる。

⑵　利活用の現状と今後の展開可能性

ア　現在、全国各地の裁判所において年間約２０万件もの民事判決の言渡しがされているといわれているが、そのうち公開されているものは、先例性のあるものや重要性の高いもの等に限られ、その利活用の場面も、法曹関係者等による先例調査や研究者による研究など、ごく限られたものであったと思われる。

イ　もっとも、現在、民事訴訟分野のＩＴ化の実現に向けて、法制審議会の民事訴訟法（ＩＴ化関係）部会において民事訴訟法改正の調査審議が行われているところであり、このＩＴ化が実現すれば、判決書の原本そのものが電子化されたものとして作成されることが想定される。このような民事判決情報がデータベース化されて広く社会に提供されるのであれば、昨今のＩＴ・ＡＩ技術の著しい進展とも相まって、例えば、機械学習のためのデータとして活用され、法律専門家等を支援するＡＩ開発がされることで、法的サービスの一層の高度化が進んだり、ビッグデータとして活用され、産業や社会に新たな知見がもたらされることで、これまでにない仕組みやシステムの開発のきっかけとなったりするなど、民事判決情報の利活用の可能性は飛躍的に高まることが予想され、前記のような民事判決情報に期待される役割、機能がいかんなく発揮さ

れるものと思われる。

2　情報管理機関による統一的管理・提供の意義

⑴　実務の現状と課題

民事判決情報は、現在、裁判所のほか、判例雑誌社・判例データベース会社等の様々な機関によって、様々な媒体を通じて公開されているが、ほぼ例外なく、訴訟関係人のプライバシー等に配慮して一定の仮名化処理が施されている。もっとも、この仮名化の範囲や程度は、一様でなく、それぞれの機関において、独自の基準を策定した上で、人手と費用をかけて仮名化が行われているのであるが、これは社会全体としてみれば、いかにも非効率な状況にあるといえる。また、ＩＴ化実現後には、データ化された民事判決情報が大量に生成され蓄積されていくことが想定され、そのような大量の民事判決情報の利活用を念頭に置くのであれば、なおさらかかる現状を改善しなければならない。

⑵　情報管理機関による統一的管理・提供（本件スキーム）

上記⑴の課題は、民事判決情報の仮名化を一定の機関に集約して実施することで抜本的に解決することができるものと考えられる。

そこで、本検討会においては、別図のとおり、これまで各機関で行われていた仮名化処理を一定の機関において集約して行うことを念頭に置きつつ、この機関が、利活用機関のため、いわば民事判決情報の管理機関となって、裁判所から民事判決情報の提供を受け、効率的に一定の仮名化処理を施し、これをデータベース化して保管管理した上で、利活用機関に対して、オンデマンドに有償（効率化を前提とした仮名化処理費用相当額）で提供するというスキーム（以下「本件スキーム」という。）を試案として提示し、この試案をもとに論点を整理することとした。

本件スキームは、民事判決情報の利活用の在り方については、既存の判例雑誌社・判例データベース会社が今日まで果たしてきた社会的役割を前提として、利活用機関（ここでは仮に「機関」としているが研究者などの「個人」を排除するものではなく、その範囲は今後の検討課題である。）の活力とその創意工夫とに委ねつつ、民事判決情報の仮名化とその保管管理については、情報管理機関によって統一的に行われることで、これに要する社会的なコストを大幅に削減することが可能となるだけでなく、訴訟関係人のプライバシー等にも一層効果的に配慮することができるというメリットがある。

⑶　情報管理機関の適格性について

判決情報のオープンデータ化は、諸外国においては、国が国民に対する司法サービスとして実施している例も散見されているところであるが、

本件スキームにおいて、民事判決情報を仮名化等することは、民間の商業ベースを含めた利活用機関の活用の前提となる情報データの整備・提供を目的とするものであるから、司法機関の担う事務にはなじまないように思われ、民間主導の継続性ある枠組みが必要となると考えられる。

他方で、本件スキームにおいては、情報管理機関が、その立場を利用して、提供先の利活用機関を恣意的に選別したり、あるいは、仮名化等に要する費用の回収を超えて、多額の収益を上げたりするようなこととなるのであれば、本スキームの構想に至った経緯に照らしても、明らかに不相当といわざるを得ないであろう。したがって、情報管理機関は、本スキームの中核となるにふさわしい一定の公益性、中立性を有する機関である必要があるのでないかと考えられる。また、民事判決情報の仮名化に際しては、訴訟関係人のプライバシー等への配慮も不可欠であり、その情報の厳格な管理も求められることとなるから、法律的な専門性や知見のある機関が担うことが望ましいと考えられる。

こうした基本的視点を踏まえつつ、情報管理機関の在り方については、民事判決情報のデータベース化が、我が国の将来の民事司法のインフラの構築に資するものであることから、後記３⑵の裁判所や利活用機関との関係等も含めて、今後も検討を進めていくべき課題であるといえる。

３　本件スキームの実現可能性

⑴　ＡＩを活用した仮名化処理の技術的検証

ア　実証実験実施の意義

これまで、民事判決情報の仮名化は、人手による作業で行われてきたところであるが、前記のとおり年間約２０万件（ただし、対象とする民事判決の範囲については、後記第３の４のとおり、議論すべき問題であると考えられる。）ともいわれる民事判決情報を人手のみで仮名化することは、その作業量や費用等を踏まえるとおよそ現実的ではないものと考えられる。

そこで、本ＰＴにおいては、民事判決情報を効率的に仮名化するための方策を模索するため、ＡＩによる仮名化の技術的可能性を検証することとして、以下の実証実験を実施することとした。

イ　実証実験の概要とその成果

本ＰＴでは、ＬＳ社に委託して民事判決情報の機械処理による仮名化についての実証実験を行った。具体的には、本ＰＴの構成員でもある第一法規株式会社（以下「第一法規社」という。）とウエストロー・ジャパン株式会社の協力を得て両社から仮名化前後の民事判決情報などの提供を受け、そのうち第一法規社から提供を受けた約１６００件の判決情

報（テキストデータ）を教師データとして機械学習を行うことで、機械処理によって、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すべき対象語句を特定することができるか（「①対象語句の特定」）、及び当該対象語句について、どの程度の精度で適切な仮名語に置き換えられるか（「②語句属性の特定」）について実験をした[1]。

実証実験の結果としては、機械処理による仮名化の「①対象語句の特定」における再現率・精度は約９４％であり[2]、対象とされた民事判決情報のうち５割超のものについては、機械処理によって仮名化すべき対象語句として特定されたものが仮名後のデータにおいて仮名化されていた語句と完全に一致したという結果となった。また、機械処理による仮名化の「②語句属性の特定」における精度は約９８％であった。ＬＳ社の説明によれば、今後のカスタマイズ等により再現率・精度の一定の上昇が見込まれるものの、機械処理のみによる再現率・精度を１００％とすることは技術的には不可能であり、１００％を目指すためには、機械処理後に人手による確認・修正を行うことが不可欠であるとのことであった。もっとも、当初から全て人手で仮名化する場合と比較すれば、仮名化に要する時間を大幅に減縮することができる見込みであるし、人手での修正を支援するツールを開発すること等により、更なる時間の短縮を図ることも考えられるとのことであった。

また、ＬＳ社からは本スキームを実現するための実務体制の構築に要する費用の規定要因として、仮名処理関連費用として、自動仮名処理システムの開発・運用費用や人手による確認・修正に要する本件スキームに必ず必要となる費用のほか、その選択によって金額が大きく異なるデータベースの運用をクラウド上で行うかオンプレミスで行うかといった点、民事判決情報を提供するためのシステムや決済フローをどのようなものとするかといった点、ＩＴ化実現後の裁判所との連携をどのように行うかといった点等の今後の課題が示された。

⑵　本件スキームの体制構築に向けた今後の課題

---

[1] 「①対象語句の特定」及び「②語句属性の特定」は実証実験内での定義・呼称に従う。詳細は LS 社発表資料（https://www.jlf.or.jp/wp-content/uploads/2021/03/minjiodpt_siryou20210126.pdf）6 頁を参照

[2] 再現率（仮名すべき単語のうち、実際に仮名化できていた割合のこと）９４．５％、精度（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のこと）は９３．４％という結果であった。詳しいデータについては、ＬＳ社の発表資料の 14 頁を参照。

ア　体制構築の基本的指針

民事判決情報のデータベース化に向けたスキームの体制構築に際しては、利活用に至るまでの一連の過程をデジタル技術により一貫して効率的に利活用できるよう全体最適化されたシステムにしていくことが重要である[3]。本件スキームの体制構築の検討に当たっては、かかる視点を踏まえ、民事司法制度の担い手及び利用者にとっての効率的でより良い在り方を検討する必要がある。

イ　裁判所との連携の在り方

本件スキームを前提とした場合、情報管理機関に一定の適格性を求める必要があることは前記２⑶のとおりである。

もっとも、その適格性を具体的に検討するに当たっては、本件スキームを構想するに至った経緯や背景を踏まえつつ、民事判決情報を裁判所から提供を受ける根拠に立ち戻った検討が必要となろう。

ウ　利活用機関との連携の在り方

本件スキームを前提としたとしても、情報管理機関と利活用機関の連携の在り方については、様々な方法が考えられるところであるが、大きな方向性としては、①仮名化後の民事判決情報の提供先を限定することなく、広く一般に向けて、情報管理機関のホームページ等で直接公開する方法と、②提供先を一定の基準・要件を満たす特定の利活用機関（例えば、判例雑誌社・判例データベース会社・研究団体・研究機関・教育機関・官公庁などが考えられる。）に限定する方法とが考えられる。

この点につき、ＬＳ社見解によれば、公開規模に応じて実務体制及び公開システム等が大規模になるほど、当該体制・システムの構築に要する費用及び詳細検討に要する期間が大きくなるため、仮に①の方法をとった場合には、具体的にはサーバーの管理、公開システムや決済システムの構築・管理について、比較的長期かつ詳細なシステム検討及び一定の費用を要するのではないか、また仮に②の方法をとった場合には、比較的短期かつ小規模な実務体制検討・構築が可能なのではないか。加えて、①の方法を実現するプロセスとして、まず②の方法をとった上で、

---

[3] かかる視点は国が進める「デジタルトランスフォーメーション（ＤＸ）」（企業がビジネス環境の激しい変化に対応し、データとデジタル技術を活用して、顧客や社会のニーズを基に、製品やサービス、ビジネスモデルを変革するとともに、業務そのものや、組織、プロセス、企業文化・風土を変革し、競争上の優位性を確立すること（経済産業省「デジタルトランスフォーメーションを推進するためのガイドライン Ver.1.0」平成 30 年 12 月））にも通じるものである。

事後的に公開システムや決済システムの検討・構築を行う場合であっても、最終的なシステムの検討・開発・運用に係る費用及び追加で生じる検討期間はほぼ変わらないため、最終的に目指す方向性が①、②のいずれの方法であっても、足元で着手する方向性として②の方法を取ることも十分検討に値するのではないかとのことであった。ＬＳ社見解に加え、前記２⑶のとおり、情報管理機関が営利を目的とせず、仮名化等にかかる費用相当額を基本的には利活用機関から回収するという本件スキームを前提とすると、①の方法をとった場合に利活用機関の経済的負担が重くなってしまうことが懸念されるため、ひとまずは②の方向を志向することが現実的な方策ではないかと考えられる。

もっとも、②の方向を志向するとしても、提供先となる利活用機関が充足すべき基準・要件がどのようなものであるべきかといった点や、提供するに当たって利活用機関による民事判決情報の取扱いについてどのような行為規範を設けるべきかといった点については、今後も更に議論を深めていく必要があるものと考えられる。

エ　民事判決情報の提供を受け仮名化後にこれを利活用機関に提供するまでの業務フローの更なる検討

情報管理機関における民事判決情報の取得の在り方については、民事訴訟分野のＩＴ化の議論を注視しつつ、提供を受けるデータの形式等を含めた具体的なフローを更に検討していく必要があるものと思われる。

また、情報管理機関における民事判決情報の保管管理方法についても、仮名化前後の民事判決情報の保管方法（クラウドかオンプレミスか等）や、そのセキュリティ確保のための方策等についても検討する必要がある。

さらに、情報管理機関と利活用機関との連携の在り方についても、既存の利活用機関の実情やニーズを踏まえつつ、今後予想される利活用の有りようも視野に入れて、提供の方法、態様等につき更に議論を深める必要があるものと考えられる。

オ　情報管理機関が今後ＡＩのシステム開発を行う、更には人手による修正・確認業務など持続・安定的に情報管理を継続していくには、相当な規模の予算と人的資源を要することから、予算及び情報管理機関の在り方について検討していく必要がある。

第３　若干の論点整理

1　総論

本件スキームの実現に向けたシステム開発、体制の構築には、前記の諸課題を解決するための更なる検討を要するところであるが、情報管理機関が裁判所から民事判決情報の提供を受け、これに仮名化等の一定の処理を施した上で一元的に管理し、利活用機関の要望に沿ってこれを提供するに当たっては、現行法制との関係で、上記の検討と並行して以下の点を更に整理しておく必要があると思われる。

2　仮名化の要否、基準について

民事判決情報には、裁判の結論となる主文のほか、事実、理由、当事者及び法定代理人等が含まれており、事実の記載においては、請求を明らかにし、かつ、主文が正当であることを示すのに必要な主張を摘示しなければならないこととされている（民事訴訟法２５３条）。したがって、民事判決情報には、訴訟当事者の私生活上の秘密や営業秘密をはじめとする訴訟関係人に関する情報が含まれることが不可避であり、こうした情報の中には、訴訟関係人において、第三者に公開されることを望まないものも含まれることは想像に難くないところである。

これまでの実務においても、訴訟関係人のプライバシー等に配慮して、公表する裁判例については、一定の仮名化処理が施されるとともに、一定の類型の事件については公表しない取扱いがなされてきたものと思われる。情報管理機関においても、このような実務を踏まえて、提供を受けた民事判決情報については、これをそのまま利活用機関に提供するのではなく、一定の仮名化処理を施した上で提供することが想定される。もっとも、本件スキームでは、仮名化処理へのＡＩの活用が前提とされており、仮名化の基準をより一層明確化することが求められる。その仮名化の要否、範囲、程度については、事件類型に応じて、裁判の公開を定めた憲法８２条１項等の趣旨を踏まえた検討を要するものと考えられよう。

すなわち、憲法８２条１項は、裁判の対審及び判決を公開の法廷で行う旨を定めており、民事訴訟における当事者の攻撃防御とそれに対する裁判所の応答は、同条２項が定めるごく一部の対審を除いて、いずれも公開の法廷で行われることとなる。そして、そのプロセスが記録された訴訟記録については、口頭弁論の公開主義の趣旨を徹底する観点から、民事訴訟法９１条１項において、「何人も、裁判所書記官に対し、訴訟記録の閲覧を請求することができる。」と定められており、判決書の原本を含め、誰でも閲覧可能とされている。民事判決情報に訴訟関係人の情報が含まれ、これに配慮する必要があるとしても、民事判決情報は、このような公開のプロセスを経て生成されたものであり、誰でもアクセスすることができる情報であることに十

分留意する必要があろう。

他方で、上記の訴訟記録の閲覧等については、訴訟当事者の私生活上の秘密や営業秘密の保護の観点から、訴訟当事者の申立てに基づく秘密保護のための閲覧等の制限の規定（同法９２条）が設けられ、裁判の公開の要請との調整が図られており、上記の対審の公開禁止の規定と併せて、訴訟関係人の利害との調整を図る上で参考とすべきであろう。

もとより情報管理機関における民事判決情報の取扱いは、裁判制度における情報の取扱いそのものとは異なるものであるが、オンライン上で民事判決情報を公開することとなれば、訴訟当事者の私生活上の秘密等に及ぼす影響が裁判所でのみ判決閲覧を認める現行制度以上に大きくなるため、上記のような視点をも踏まえつつ、適切な利害調整を図るための検討を進めるべきである。

現在、民事訴訟分野のＩＴ化の実現に向けて、法制審議会の民事訴訟法（ＩＴ化関係）部会において民事訴訟法改正について調査審議が行われているところであり、ＩＴ化が実現した後の訴訟記録の閲覧等の在り方についても議論が進められている。今後は、こうした議論も注視しつつ、ＩＴ化実現後の訴訟記録の閲覧等の制度との関係も視野に入れた検討を進める必要があるものと思われる。

３　仮名化すべき情報、程度について

前記のとおり、民事判決情報の仮名化の要否、範囲及び程度については、事件類型によってもその内容は変わり得るとの指摘もあったところであり、更なる検討を要するが、これまでの実務を踏まえると、まずは、以下の点を中心に整理を進めるべきであろう。

⑴　類型的に仮名化すべき情報

ア　個人の氏名・住所、生年月日等

個人の氏名については、歴史上の人物やプライバシー権侵害を上回る公共の利益が認められる者の氏名、後述するビッグデータに関する議論等を除けば、仮名化すべき要請が最も高い個人情報であると考えられる。また、住所及び生年月日等の個人情報についても、仮名化することが相当な情報ではないかと思われる（もちろん、住所や生年月日については、どこまでの範囲で仮名化する必要があるかという議論は残る。）。

このほかにも、例えば、金融機関の口座番号や携帯電話番号といった類型的に仮名化することが適当な情報があるものと考えられるものの、個人の特定につながり得る情報について、どのような範囲のものをどの程度まで仮名化すべきと考えるか、明確な基準の策定が可能である

かどうか等も含めて、更に議論を深めていく必要があるものと考えられる。

イ　法人の名称

法人の名称については、実務の現状をみても、仮名化の有無は、具体的事案に応じて検討がされているものと考えられ、一律の基準を示すことは難しい状況にあるようにも考えられる。

この点については、法人については、プライバシー権等が観念されないことから、個人とは異なり、そもそも仮名化の必要はないのではないかとの意見があった。他方で、民事判決情報がインターネット上で公開されることによるレピュテーションリスクを気にするあまり、紛争解決手段として裁判手続を選択すること自体を躊躇させてしまう可能性があるのでないかとの指摘もされたが、このような指摘に対しては、国際的には、情報を秘匿するために、公開が予定される裁判でなく、仲裁その他ＡＤＲが活用されるなど、紛争当事者において事案に応じた紛争解決手段が選別されている実情にあり、上記のような指摘を過大に評価することで、かえって現在よりも利用可能な民事判決情報の情報量が減少してしまうことがあってはならないのでないかとの意見もあった。こうした点も踏まえつつ、引き続き検討していくべき課題であると思われる。

ウ　商品名等の固有名詞

商品名や地域名などの固有名詞については、個人（法人）名を仮名化しても、当該固有名詞の記載により、その個人（法人）が特定されてしまう可能性がある場合もあり、その仮名化の必要性が生ずる一方で、例えば、知的財産権関係訴訟については、仮名化すると公開の意義が損なわれてしまうとの指摘もあったところであり、こうした点も踏まえつつ、引き続き検討していくべき課題であると思われる。環境・公害関係訴訟における地域名も同様である。

⑵　特異な情報であるためや、訴訟手続外で公となっている情報と併せることにより、個人等が推知される情報

ア　実務の現状においては、公開されている民事判決情報の多くは、氏名等の個人情報の仮名化が行われているものの、個人等が推知される特異情報や、公となっている情報と併せることにより個人等が推知される情報についてまで仮名化は行われていないのではないかと思われる。

一方で、個人情報の保護に関する法律（以下「個人情報保護法」という。）との関係においては、「他の情報と容易に照合することができ、それにより特定の個人を識別することができることとなるもの」も「個人

情報」に該当するものと整理されており（同法２条１項１号）、医療情報のデータベース化についてのルールを定めたいわゆる次世代医療基盤法においては、このような情報まで含めて匿名化を行うことが求められている。

イ　この点について、本ＰＴにおいては、現状の仮名化の程度（前記ア）で公開することに不法行為法上の違法があるとは考え難いとする意見や、これらの情報についてまで仮名化されてしまうと、内容が分からなくなってしまい、公開の意義が失われてしまうことを懸念する意見が多数出されたところである。また、個人等が推知される情報であるかどうかを判別することは、機械処理による場合はもちろんのこと、人手による場合であっても、多分に主観的な判断が求められることとなり、そのような対応は困難ではないかとも思われる。

ウ　そこで、個人等を推知させる情報の仮名化を要するかどうかについては、前記２で指摘したところも踏まえつつ、引き続き検討すべき事項と思われる。

⑶　営業秘密

営業秘密については、そのような情報が民事判決情報に含まれることがあったとしても、その情報の性質に照らして、訴訟当事者以外の者においてそれが営業秘密に該当するかどうかの判断を的確に行うこと自体が困難であると考えられるから、情報管理機関において、民事判決情報に含まれる営業秘密を漏れなく抽出して仮名化することは困難ではないかとも思われる。

この点については、民事訴訟法９２条１項２号において、営業秘密について閲覧等制限の申立ての手続が定められているのであるから、本来的には当該手続において対応すべき事柄であるとの意見が複数の委員から出されたところであり、このような意見も踏まえつつ、引き続き検討すべき事項と思われる。

⑷　利活用機関と仮名化の程度

本ＰＴにおいては、民事判決情報の仮名化については、利活用機関における利活用の方法に応じて仮名化の程度を変えることも検討すべきでないかとの指摘もあったところである。特に、利活用の目的がビッグデータとしてＡＩに読み込ませることにある場合には、そもそも仮名化をすること自体が無意味であり、そのことで所期の目的を達成することができず、かえって利活用を阻害することになりはしないか、むしろ、問題は目的外使用されるおそれにあり、そのような弊害に対応するための適切な仕組みを検討すべきでないかとの意見も出たところである。

この点については、仮名化処理をすることによるビッグデータとしての活用に対する弊害の大きさや、ビッグデータとして活用する機関からの情報漏洩等のリスクの大きさ、ビッグデータとして活用する場合には仮名化をすること自体が無意味といいきれるのか等を考慮していく必要があるものと思われ、その利活用の実態も踏まえつつ、今後も引き続き議論していくべき課題であるといえる。

⑸　事後的な是正の手段について

個人情報や営業秘密等について、画一的な基準により仮名化を実施したとしても、人手による確認のミス等によって、仮名漏れや仮名過多が生じる可能性があるし、基準においては仮名化の対象とならないものであっても、個別の事情によっては仮名化することが相当なものがあったり、逆に、基準においては仮名化の対象となるものであっても、当該民事判決情報を分析検討する観点からは仮名化によって所期の目的を達することができなくなってしまう場合もあり得るものと思われる。このような場合に対応するために、仮名化についての不服申立てのような制度を設け、事後的な是正の機会を確保すべきであるとの意見が出たところである。この点については、今後も引き続き議論していくべき課題であるといえる。

４　対象とすべき民事判決情報の範囲

⑴　いわゆる調書判決について

民事判決は年間２０万件程度の言渡しがされているものの、これらの中には、判決の言渡しが判決書の原本に基づかないでされる、いわゆる調書判決（民事訴訟法２５４条１項）も相当数含まれている（令和元年の第一審民事通常訴訟事件における地方裁判所における欠席判決は２万４７８０件、簡易裁判所における欠席判決は９万６８３１件[4]）。この調書判決は、その要件に照らしても、一般に、対席判決と比較して先例的価値に乏しいものが多いと考えられる。また、調書には、当事者及び法定代理人、主文のほか、請求及び理由の要旨が記載されることとなるが、この請求の記載については、実務上、調書に訴状の写しを添付する形式で記載されることが多い。したがって、このような判決情報までデータベース化するとなると、より多くの手間や費用を要することも想定されることからすれば、調書判決を対象範囲から除外するということも考えられる。そこで、ＩＴ化が実現した後の民事訴訟における判決書の在り方や利活用機関のニーズ等を踏まえつつ、更に検討することが考えられる。

⑵　データ化されていない民事判決について

[4] 数値は裁判所データブック２０２０による。

前記の民事訴訟法（ＩＴ化関係）部会において調査審議されているところであるが、民事訴訟のＩＴ化が実現されれば、判決もデータの形式で作成・保存されることになるのでないかと考えられることから、民事判決情報の情報管理機関を通じた利活用は、一定の仮名化処理を施すにせよ、現状の実務と比較して、容易で効率的なものになると考えられる。

一方、既に言渡しがされた民事判決の判決書の原本については、紙媒体の形式で保存されているところ、このような民事判決情報をデータベース化するためには、既にインターネット上で公開されている民事判決情報を利用する場合を除けば、紙媒体の形式で保存されている判決書を用いて、ＯＣＲ等の技術により、データに変換することから必要になることが見込まれる。これらを考慮すると、ＬＳ社の実証実験による試算以上に費用及び手間がかかることになるものと考えられる。

上記を踏まえると、まずは、判決書がデータの形式で作成・保存されることとなった場合における民事判決情報のデータベース化の在り方を検討することが相当であろう。

５　個人情報保護法との関係の整理

⑴　総論

本件スキームでは、情報管理機関は、裁判所から継続的に民事判決情報の提供を受け、これに仮名化処理を施した上で、利活用機関に提供することを想定していることから、その詳細設計如何では、情報管理機関において、仮名化処理の前後を問わず、大量の民事判決情報を保管管理することとなる。前記のとおり、民事判決情報には、訴訟関係人に関する情報が含まれることが不可避であることから、情報管理機関が保有する情報には、「個人情報」（個人情報保護法２条１項）が含まれ、事例によっては、「要配慮個人情報」（同条３項）が含まれるものも相当数存在することが予想される。したがって、今後、情報管理機関における管理及び提供の在り方を検討するに当たっては、個人情報保護法との関係を更に整理する必要があるものと思われ、差し当たり、現時点においては、以下の諸点を指摘しておきたい。

⑵　「個人情報取扱事業者」（「個人情報データベース等」）該当性

仮に、情報管理機関において提供を受けた民事判決情報やこれに仮名化処理を施したものをどれだけ集積しても、その集積したものが「個人情報データベース等」（同条４項）に該当しないというのであれば、情報管理機関は「個人情報取扱事業者」に該当しないということとなり（同条５項）、個人情報取扱事業者の義務等を定めた同法第四章の適用はないということになる。

この点については、情報管理機関において集積されたものが「特定の個人情報を電子計算機を用いて検索することができるように体系的に構成したもの」（同条４項１号）と整理されるのかどうかが一つの論点となるものと思われるが、情報管理機関における民事判決情報の取扱いや業務フローの在り方の検討とも深く関わるところであり、この点と併せて更に検討を深めていく必要があるものと思われる。

⑶　同法第四章の規定が適用される場合の規制について

仮に、情報管理機関が個人情報取扱事業者に該当するのであれば、情報管理機関による民事判決情報の取得、管理、適用については、個人情報の取得に関する制限の規定（同法１７条等）、第三者提供の制限の規定（同法２３条等）等の適用を受けることとなるから、これらの規定との関係をどのように整理するのかについて更に検討を深める必要がある。

なお、個人情報保護法については、令和２年に改正がされ、現在、更なる改正に向けた法律案が国会に提出されたところであり、いずれにしても、その動向も注視しつつ検討を進めていく必要があろう。

第４　おわりに

今後、民事判決情報のデータベース化の試みを進める際に検討すべき課題は、前記第２、第３のとおりであるが、大きく分けると①情報管理機関と利活用機関の関係の在り方や、仮名化にかかる費用の抑え方といった、民間主導の継続性ある枠組みを構築するために検討すべき課題（第２）と、②現行法制との関係やあるべき仮名化の程度といった法律的観点から検討すべき課題（第３）と、に分けられるものと考えられる。

したがって、令和３年度以降は、①及び②の課題をそれぞれ別の場で検討を進めていくことが相当であると考えられる一方で、相互に関連する部分もあるところから、必要に応じて情報交換等を行いつつ、今後も力強く民事判決情報のデータベース化の取組を前進させていくことが重要であると考える。

当財団は、その設立趣旨にのっとり、次年度以降もこの取組の中核となる機関として、本ＰＴで明らかとなった諸課題の解決に精力的に取り組み、民事判決情報のデータベース化の実現に向けて力強く前進させていく所存である。今回御協力頂いた関係者の方々に改めて感謝の念を述べさせて頂くとともに、民事判決情報のデータベース化の実現に向けて、今後もこの取組に御協力頂ければ幸いである。

2020. 3. 27 **別紙**

日弁連法務研究財団主催「民事判決のオープンデータ化検討ＰＴ」

1　目的

民事司法制度改革の課題の一つとして、民事裁判手続のＩＴ化に伴って、民事判決情報の活用拡充（判決情報のデータベース化を含む）のニーズ・活用可能性が高まっていることを踏まえ、民事判決データの管理・活用を担う民間組織の立ち上げも視野に、民事判決データの管理及び利活用に当たり検討すべき課題・対応策について、幅広い観点から、実務的協議を行う。

2　位置付け及び構成員

日弁連法務研究財団の活動の一環として、民事判決データの活用に向け、同財団が受け皿となる可能性も含め、検討を行う。構成員は、以下のとおり。

（座　　長）　菊地裕太郎日本弁護士連合会会長

（メンバー）　横溝高至日弁連法務研究財団専務理事
日本弁護士連合会
山本和彦一橋大学教授
町村泰貴成城大学教授
一般財団法人司法協会
一般財団法人法曹会
株式会社Legalscape
ウエストロー・ジャパン株式会社
株式会社有斐閣
第一法規株式会社

（オブザーバー）内閣官房
法務省
最高裁判所事務総局

（事務局）大坪和敏日弁連法務研究財団事務局員
藤原靖夫日本弁護士連合会事務次長

3　その他

・　ＰＴは、令和２年３月２７日に第一回会合を開催する。当面、半年間をメドに、議論のとりまとめを目指す。

・　ＰＴの事務は、日弁連法務研究財団が、日本弁護士連合会などの協力を得て行う。

別紙

民事判決のオープンデータ化検討ＰＴ　構成員名簿

２０２１年（令和３年）３月２５日現在

◎：座長　　○：座長代理

	構成員	氏名
◎	菊地裕太郎（日本弁護士連合会前会長）	菊地裕太郎（日本弁護士連合会前会長）
○	早稲田祐美子（日弁連法務研究財団専務理事）	早稲田祐美子（日弁連法務研究財団専務理事）
	日本弁護士連合会	渕上玲子（事務総長） 菰田　優（前事務総長）
	山本和彦（一橋大学教授）	山本和彦（一橋大学教授）
	町村泰貴（成城大学教授）	町村泰貴（成城大学教授）
	小塚荘一郎（学習院大学教授）	小塚荘一郎（学習院大学教授）
	一般財団法人司法協会	松本英司（複写事業部長） 川端素子（出版事業部長）
	一般財団法人法曹会	清水　孝（主事）
	株式会社 Legalscape	八木田樹（代表取締役・最高経営責任者） 津金澤佳亨（最高執行責任者） 城戸祐亮（取締役・最高技術責任者） 久本空海（ソフトウェアエンジニア）
	ウエストロー・ジャパン株式会社	中村裕哲（執行役員）
	株式会社有斐閣	高橋　均（常務取締役）
	第一法規株式会社	田中英弥（代表取締役社長） 川原﨑晶子（出版編集局編集第一部長）

オブザーバー	氏名
内閣官房	坂本三郎（内閣審議官）
法務省	金子　修（法務省司法法制部長） 大野晃宏（法務省民事局参事官）
最高裁判所事務総局	村田斉志（最高裁判所事務総局総務局長）

事務局	氏名
日弁連法務研究財団	大坪和敏（事務局員）
日本弁護士連合会	藤原靖夫（事務次長）

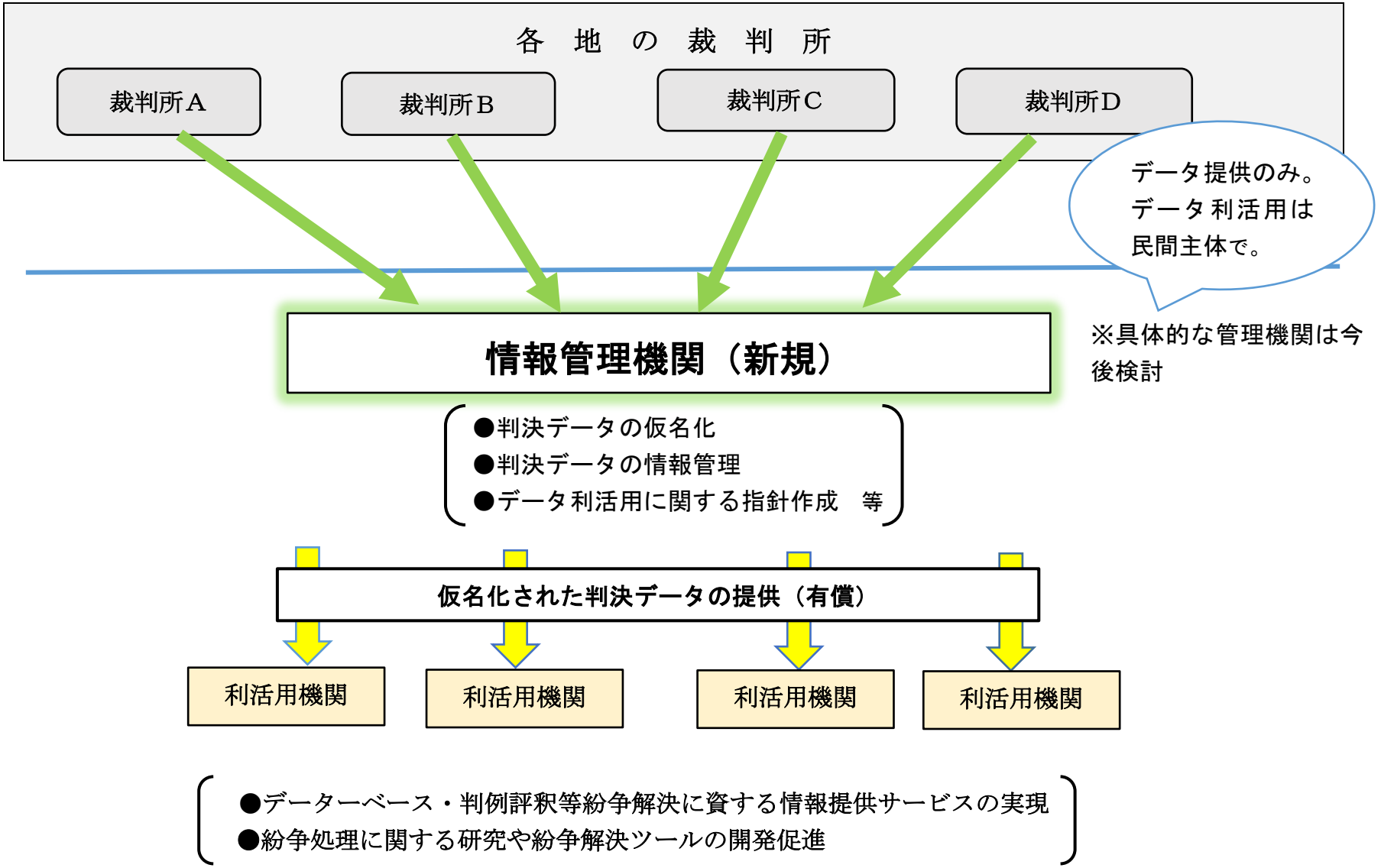
別紙

民事判決のオープンデータ化検討ＰＴ　開催状況

第１回	令和２年３月２７日	○ＰＴ設置の趣旨説明 ○政府における民事司法制度改革の議論について ○民事判例オープンデータ化の可能性と課題について ○判決情報のデータ処理，匿名化について
第２回	令和２年５月２５日	○有識者からのヒアリング ・医療データの匿名利活用の状況について ・民事判決の匿名化について ○諸外国の判決情報の匿名化状況について ○検討課題について
第３回	令和２年６月２９日	○有識者からのヒアリング ・判決データの利活用と個人情報保護法 ・他分野における匿名化加工・利活用の取組について ○諸外国の判決情報の匿名化状況について ○実証実験の準備状況について
第４回	令和２年７月２７日	○有識者からのヒアリング ・民事判決のオープンデータ化の意義と利活用の可能性について ○実証実験の準備状況について
第５回	令和２年９月２日	○有識者からのヒアリング ・情報公開の観点からみた民事判決のオープンデータ化とその課題 ○実証実験の進行について
第６回	令和２年１１月１１日	○実証実験の経過報告 ○今後の論点整理について
第７回	令和３年１月２６日	○実証実験の報告 ○今後の論点整理について
第８回	令和３年２月２５日	○取りまとめ（たたき台）について
第９回	令和３年３月２５日	○取りまとめについて ○次年度について

別図

民事判決情報のオープンデータ化（考えられるイメージ私案）



資料４

Confidential

LEGALSCAPE
WE EMPOWER LEGAL PROFESSIONALS

民事判決オープンデータ化検討**PT**

自動仮名処理に係る実証実験

－追加実験内容及び、最終結果の報告－

2021年1月26日

株式会社Legalscape

要旨

- 本実証実験（’**20**年**10**月**-’21**年**1**月実施）では、民事判決オープンデータ化検討に資することを目的とし、機械による仮名処理の技術的可能性を実験・評価した上で、仮名処理の実務体制構築の方向性オプション・主要検討項目のたたき台を整理した
  - 前回の実験経過報告（’20年11月PT時報告）では、機械処理の2つの段階「①対象語句の特定」「②語句属性の特定」のうち、①における機械処理性能を初期的に確認・報告した
  - 以降本日までに「①の性能向上の取組み」及び「②も含めた機械処理の実装・性能評価」を実施した

- 機械処理性能は、「①対象語句の特定」において再現率・精度ともに**94%**前後（箇所ベース）を達成し、「②語句属性の特定」精度は**98%**（属性グループベース[1)]）を達成した
  - とくに、①においては5割超の判例で100%の正解（機械よって完全に正しい仮名処理が行われた）となり、既存の日本語新聞や海外判決文等における自然言語処理性能と比較し、高い性能を達成
  - 今後、更なる研究開発等により一定の性能向上は不可能ではないと推察されるものの、他方で機械処理のみによる100％の性能実現は原理的に不可能であるため、実務上は人手による追加修正の要否・程度の検討が不可欠

- 人手による修正方針には複数のオプションがあるが、いずれにおいても費用・手間等は（今後継続的に本**PT**の目的・条件等に立ち返り確認する必要があるものの）実現に向けた詳細検討が可能な範囲である
  - 人手による確認・修正の方針のオプションには、大きく全文を確認・修正する方針か、機械出力の周辺のみを確認・修正することで、一定の誤りを許容する代わりに確認・修正の手間を削減する方針がある
  - また、いずれの方針においても人手による修正を効率化するツールの開発が可能であり、仮名化に要する時間を現状比で1/6～1/20程度に削減することが可能な想定
  - 今後の実務体制構築にあたっては、仮名処理システムの検討・開発・運用に加え、データ公開の規模や方式等を踏まえた全体の公開システム・実務オペレーション体制に係る検討を要する

1. 語句属性の特定精度の定義については本編p.11に記載

本資料の内容

1. 実証実験の概要（目的と位置づけ）
2. 実証実験の結果：自動仮名処理の技術的可能性
3. 実務体制構築の方向性オプション（議論用叩き台）

実証実験の目的

背景

民事判決のオープンデータ化を行う上で、仮名処理をどのように／どの程度、施すべきかが検討されている

- 特に、技術的な観点からは、判決の大量公開にあたって、一定の機械的な自動処理が前提とされる中で、機械的な処理の可能性、精度評価が必須
- 同時に、機械による自動処理が完全でないことも想定されるため、人手による追加修正の要否・経済性の評価・検討も必須

∨

目的

本実験は、機械処理精度の評価、及び、人手による修正の費用対効果を検証・推定することで、取るべき仮名処理手法／体制構築の検討に資することを目的とする

実証実験の流れ・位置付け

- 約1,600件の判例を対象に実験を行い、自動仮名処理の技術的な可能性を検証・評価
- 技術的可能性の評価を踏まえ、実務体制構築の検討材料を提示

仮名処理の実務体制を構築するまでの流れ

	1. 機械処理の精度水準を見定める	2. 人手修正の費用対効果を概算する	3. 実務体制の検討・構築を進める
検討論点	機械処理精度の水準は？ • 各誤りの質の違い、量は？ • どこまで向上可能か？	実務的観点から、どの程度、人手で修正すべきか？ • 修正する場合の作業量・費用は？ • 支援ツールでの効率化効果は？	左記の技術的可能性及び、本件趣旨／政策的観点を踏まえ、どのような実務体制を構築すべきか？
検討／実施アプローチ	実証実験 本実験の範囲 約1,600件の判例データを対象に実験 必要に応じて、実験規模・詳細度を上げる • データの追加 • 案件類型ごとの分析 （精度要求水準の異なる案件類型を設定・分類し、分析）		システム開発・体制構築 システムの仕様策定・開発 支援ツールの開発に着手 実務体制（事業モデル含め）を検討・構築 • 公開対象者・データ利用目的等による処理精度水準の設定の違いがあれば考慮
	本資料にて結果を報告		本資料にて検討の枠組・材料を提示

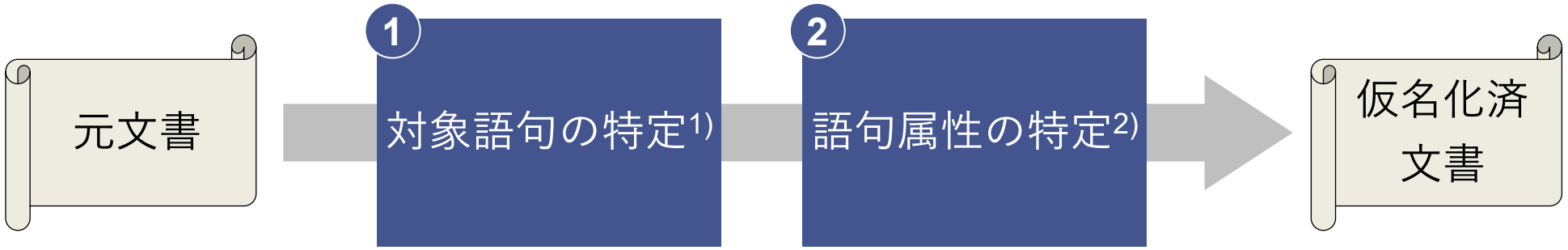
LEGALSCAPE
WE EMPOWER LEGAL PROFESSIONALS

本資料の内容

1. 実証実験の概要（目的と位置づけ）
2. 実証実験の結果：自動仮名処理の技術的可能性
   - ① 実験のアプローチ
   - ② 機械処理の精度水準
   - ③ 人手修正の費用対効果
3. 実務体制構築の方向性オプション（議論用叩き台）

機械処理の流れ

①対象語句の特定、②語句属性の特定の２つの機械処理を施す



原告山田太郎（以下「山田」という）は東京都渋谷区にて…

- *山田太郎*
- *山田*
- *渋谷区*

- 同一: *山田太郎* = *山田*
- 住所: *渋谷区*

*原告**X**（以下「**X**」という）は 東京都（以下略）にて…*

1. 言語処理分野の研究での「*固有表現抽出*」に該当
2. 言語処理分野の研究での「*関係抽出*」などに該当

LEGALSCAPE
WE EMPOWER LEGAL PROFESSIONALS

機械処理の流れ：❷語句属性の特定

置換語の選定には、グルーピング・タイプ分類・名付けの各段階における判定を要し、
各判定の精度が最終的な処理精度に影響を及ぼす

仮名処理対象の語句全体の中から、同じラベルを付けるべき語句同士をグループ分け

各グループに対して、人名／住所／車両番号／郵便番号、等のタイプを分類

各グループに対して、タイプに則した適切な仮名を付与

LEGALSCAPE
WE EMPOWER LEGAL PROFESSIONALS

処理誤りの意味合い

❶❷それぞれの機械処理における誤りは、異なる実務上の問題をもたらす

「❶対象語句の特定」での誤り

ケースI：仮名の「漏れ」

山中太郎（旧姓佐藤）は…

↓

X（旧姓佐藤）は…

個人名の漏洩
→ プライバシーリスク

ケースII：仮名の「過多」

山中の道で…

↓

Xの道で…

文意の不明／変化
→ 閲覧者の利用性・権利の問題

「❷語句属性の特定」での誤り

ケースIII：仮名記号の不一致

山中太郎（旧姓佐藤）は…

↓

X（旧姓Y）は…

文意の変化
→ 閲覧者の利用性・権利の問題

- プライバシーリスクの問題は、①の結果のみが影響
- 閲覧者の利用性／権利上の問題は、①②両方の結果が影響

誤りの種類と評価指標：❶対象語句の特定（1/2）

「仮名漏れ（偽陰性）」と「仮名過多（偽陽性）」という２種類の誤りがある

誤りの種類

機械出力 ＼ 正解	仮名対象	一般語句
仮名実施	A 正答	B 仮名過多 *（偽陽性）*
仮名せず	C 仮名漏れ *（偽陰性）*	D 正答

評価指標

$$\text{精度} = \frac{A}{A + B}$$

仮名実施のうち何割が正解か
（全く特定しなければ100%）

トレードオフの関係

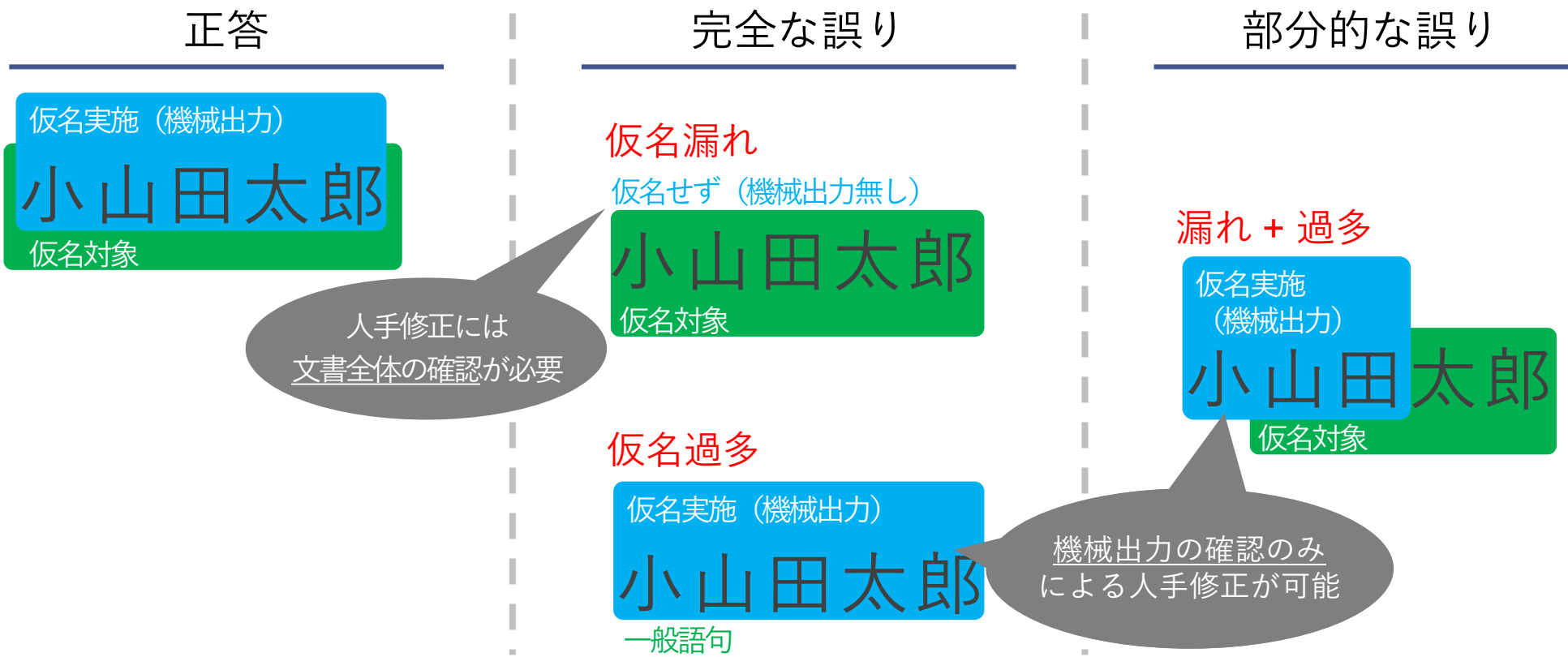
$$\text{再現率} = \frac{A}{A + C}$$

仮名対象のうち何割を仮名できたか
（全文字を特定すれば100%）

- プライバシーリスクには「仮名漏れ（偽陰性）」が影響
- 閲覧者の利用性／権利には「仮名過多（偽陽性）」が影響

誤りの種類と評価指標：❶対象語句の特定（2/2）

仮名漏れ／仮名過多には、さらに「完全な誤り」と「部分的な誤り」がある

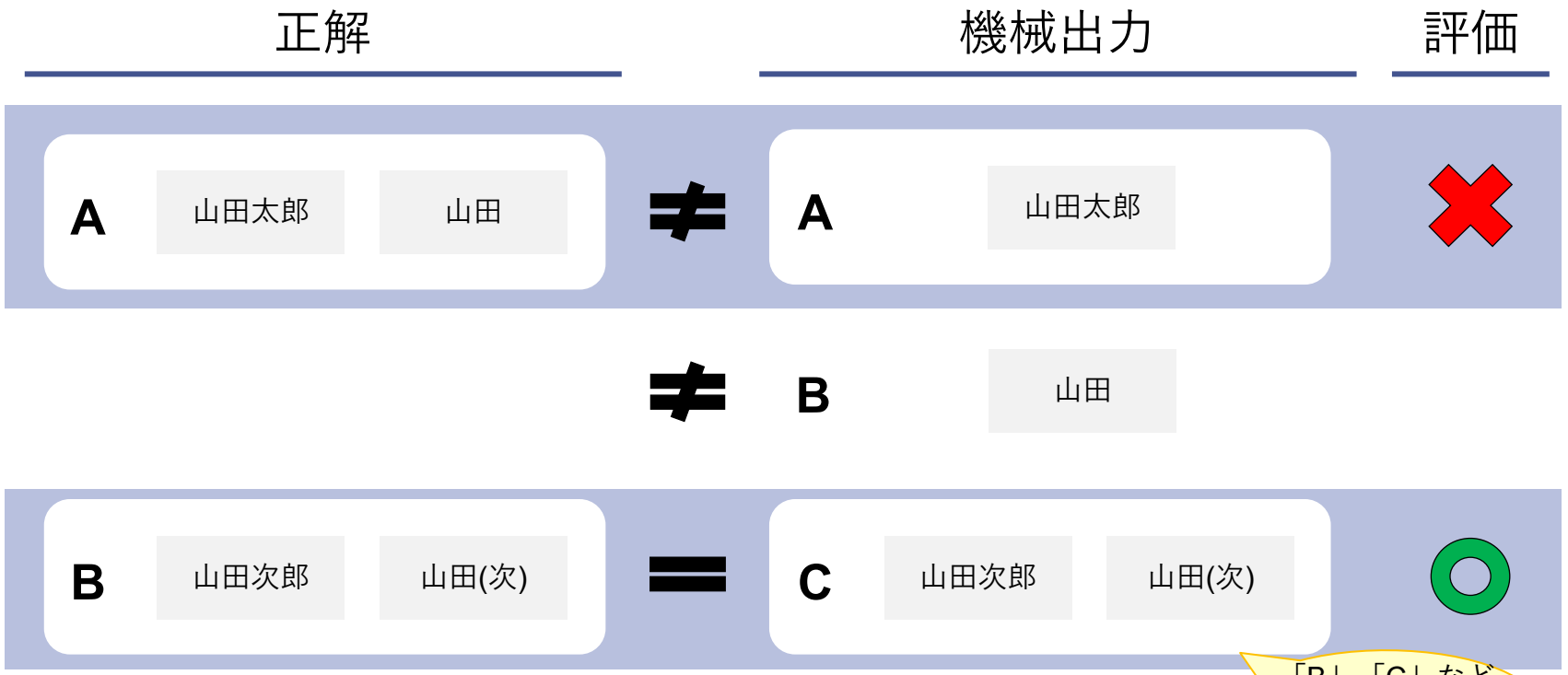


- 人手修正をする場合、修正対象を「完全な誤りのうちの仮名過多」と「部分的な誤り」に絞ることで、修正コストを下げることが可能

（＝文書全体を確認するのではなく、機械出力のみ確認することで修正の手間を減らす）

誤りの種類と評価指標：❷語句属性の特定

仮名対象語のグループの一致割合を「精度」として定量評価



精度
（グループ一致割合）

$$= \frac{\text{正解と一致する機械出力グループ数}}{\text{正解グループ数}} = \frac{1}{2} = 50\%$$

本資料の内容

1. 実証実験の概要（目的と位置づけ）
2. 実証実験の結果：自動仮名処理の技術的可能性
   - ① 実験のアプローチ
   - ② 機械処理の精度水準
   - ③ 人手修正の費用対効果
3. 実務体制構築の方向性オプション（議論用叩き台）

機械処理の精度水準

実証実験の
最終結果

技術詳細
（性能改善の取り組み）

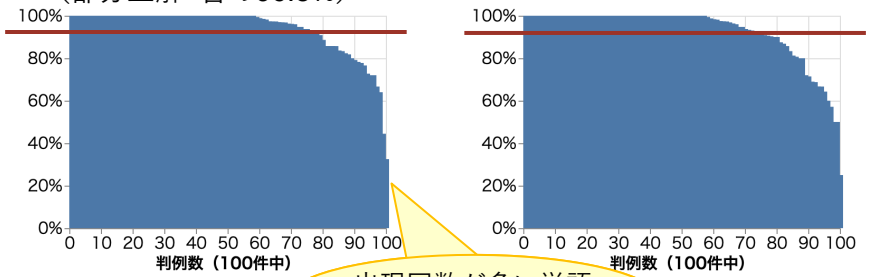
実証実験の最終結果：全体像

- 判決公開に向け、仮名処理実務に機械処理を採用／組込ことは、充分に検討に値する
- 本実験結果は、自然言語処理における1つの研究としても、一定の成果を実現できた

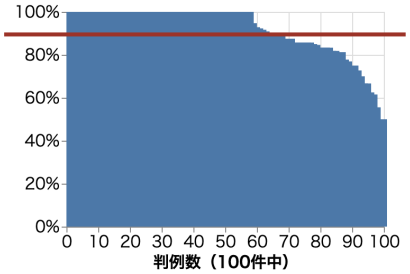
1 対象語句の特定

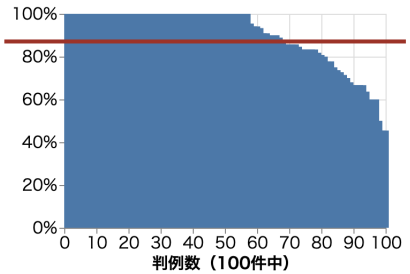
2 語句属性の特定

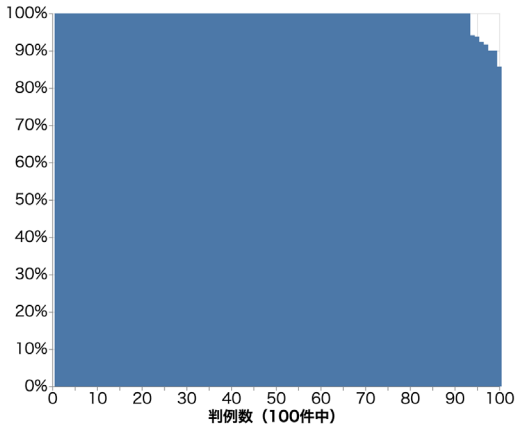
	再現率（仮名漏れ）	精度（仮名過多）	精度（グループ一致割合[2]）
仮名箇所数ベース（1単語でも複数回カウント）	94.5% （部分正解[1]含め96.8%）	93.4% （部分正解[1]含め96.4%）	98.0%
仮名単語種類ベース（1つの単語は1回カウント）	91.8% （部分正解[1]含め96.4%）	87.6% （部分正解[1]含め94.5%）	



出現回数が多い単語における誤りが比較的少ない







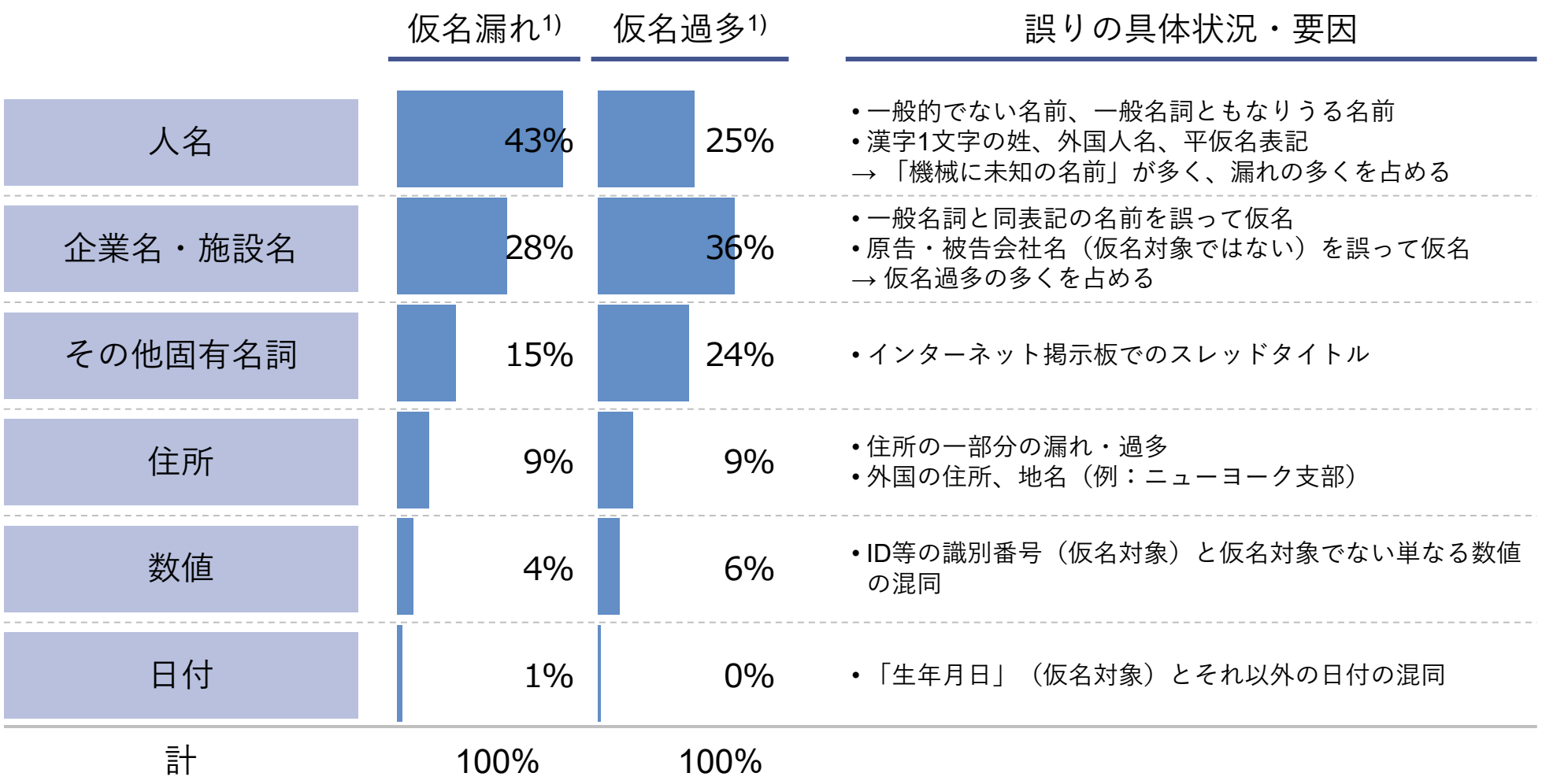
※正解の仮名箇所に対して、②を実施した際の精度

5割の判例は修正を必要とせず、必要とする場合でも、平均7箇所・3単語の修正に留まる

1. 「部分正解含め」た再現率／精度とは、p.10記載の「正答」及び「部分的な誤り」の両方を正解とした場合の再現率／精度を指す
2. グループ一致割合の定義は、p.11参照

実証実験の最終結果：詳細：❶対象語句の特定

- 誤り全体を仮名対象の語句タイプ別に分類
- 誤りの80%以上は、人名、企業・施設名、その他固有名詞の認識誤りによる



本実験にて各種精度改善（後頁にて詳述）を施した上でも、なお誤るケースは、
「未知の人名・ニックネーム」／「一般名だが固有名」等の機械的判断が難しいケース

1. 前頁箇所数ベースの誤りについて、分析

実証実験の最終結果：詳細：❶対象語句の特定

- 網羅の困難さから、仮名漏れが発生
- 逆に一般名詞との混同などにより、仮名過多も発生

人名の誤り

仮名漏れ
非定形の名

正解
ネットネームのtanaka@net

機械出力
ネットネームのtanaka@net

人名の誤り

仮名過多
一般語と混同

正解
被告は金を要求した

機械出力
被告は金を要求した

企業名の誤り

仮名漏れ
既知でない名

正解
被告は、すみれフラワーに ...

機械出力
被告は、すみれフラワーに ...

企業名の誤り

仮名過多
一般語と混同

正解
被告は、国立病院に ...

機械出力
被告は、国立病院に ...

実証実験の最終結果：詳細：❶ 対象語句の特定

- 自由度の高い固有名詞や数値は網羅が困難で仮名漏れが発生
- 住所の誤りはほとんどの場合において部分的な仮名過多・漏れ

固有名詞の誤り

仮名漏れ
自由記述名

正解
ネット掲示板「東京グルメ案内」

機械出力
ネット掲示板「東京グルメ案内」

住所の誤り

仮名過多
部分的なミス

正解
東京都北区北部地域

機械出力
東京都北区北部地域

数値の誤り

仮名漏れ
未知の形式

正解
ログインID: 12345

機械出力
ログインID: 12345

日付の誤り

仮名漏れ
生年月日

正解
被告は1月1日に息子Aを出産

機械出力
被告は1月1日に息子Aを出産

実証実験の最終結果：詳細：❷ 語句属性の特定

ルールベースによるシステムで100%に近い精度を達成

機械的アプローチ

- 古典的なルールベースによる手法のみ
  - チューニングしやすく再現性が高い
- 採用したルール
  - 「○○（以下○○という。）」などの明示的な記述を利用
  - 「フルネーム」と「姓」「名」の対応づけ
  - ひらがな・カタカナ表記の対応づけ（山田＝ヤマダ）

機械処理の結果

- 正解率：98.0 %

  ※1　「正解の仮名語グループ」（仮名記号）のうち「機械処理で正しく出力したグループ」の数の割合。ステップ①は完璧と仮定し独立に評価

  ※2　旧姓・通称の区別は今回評価対象外とした

- 誤り分析
  - 明示的でない略称
    - 「東京タナカ病院」 → 「タナカ病院」
  - 一意に特定できない人名
    - 「田中太郎」「田中次郎」が登場する判例での「田中」
  - 同表記だが区別される名前
    - 「田中が所有する田中ビル」→「Aが所有するBビル」
  - その他、明示的でない別表記
    - ニックネーム「田中直樹」→「ナオ」
    - インターネット上の名前「田中」→「TANAKA@PC」

- ステップ②の誤りは人間が読む際には容易に発見できるため、利便性・情報公開性の毀損は限定的と思料
- ステップ①の誤りが後続の②に伝播することを踏まえても、①の方がより重要

実証実験の
最終結果

技術詳細
（性能改善の取り組み）

技術詳細：性能改善の取り組み

- 主な改善アプローチを網羅的に検討・採用し、性能改善を達成
- 大規模統計モデルBERT及び個別ルールにより、仮名漏れの大幅削減に成功（仮名過多も減少）

	採用手法	再現率（仮名漏れ）	精度（仮名過多）
前回PT時点	spaCy	85.2%	92.4%
最終結果	BERT+ルール	94.5%	93.4%

各手法での改善幅	ⅰ spaCy+法律辞書	+ 0.3%	(0.6%)
	ⅱ BERT	+ 8.8%	+ 0.8%
	ⅱ BERT+生データ	+ 7.7%	+ 0.6%
	ⅲ BERT＋ルール	+ 9.3%	+ 1.0%

一通りの改善手法を実施し、

最も改善幅の大きい**BERT+**ルールによる改善手法を最終結果として採用した

更なる改善余地は小さいため、次のステップでは、仮名処理実務としての要求性能／人手修正方針検討に進むことが望ましいと思料。次頁以降、各改善手法の詳細を解説

手法：spaCy＋法律用語辞書の追加

- 法律用語辞書の用語の出現が少なく、性能への影響なし
- 辞書追加による更なる性能向上は、企業名や稀な人名を入れる必要

追加辞書の詳細

- 法律用語辞書として2種類の言語資源を受領
  - 有斐閣様『法律用語辞典』
  - 商事法務様『会社法コンメンタール事項索引データ』
- これらのデータから、既存の機械処理用辞書に未収録の1.2万語を整備し利用
- 今回対象とする判例データの全文約900万語中、0.9%のみが該当する法律用語であった
  - （例）「弁論の全趣旨」、「不法行為」、「訴訟代理人」

追加結果

- 性能への影響はなし
  - 追加された語彙は専門用語・学術語も多く、判例本文中の出現頻度は小さい上、仮名対象と識別が困難な語がほとんど含まれていない
  - 例えば「訴訟代理人」を「訴訟/代理人」と2単語と解釈していたとしても、結果への影響はなし

→ むしろ人名辞書・企業名辞書等のさらなる追加が有効

尚、本実験では、**BERT**の学習に利用した**Wikipedia125**万記事に含まれる固有名詞をカバーしており、更なる固有名詞の充実のためには、特殊・レアな人名等の辞書が必要

手法：大規模な統計モデル（BERT）＋生データ

判例データ以外の大規模言語データを活用した統計モデル（BERT）の導入により、
特に仮名漏れが大幅に改善

大規模統計モデル（BERT）の概要と結果

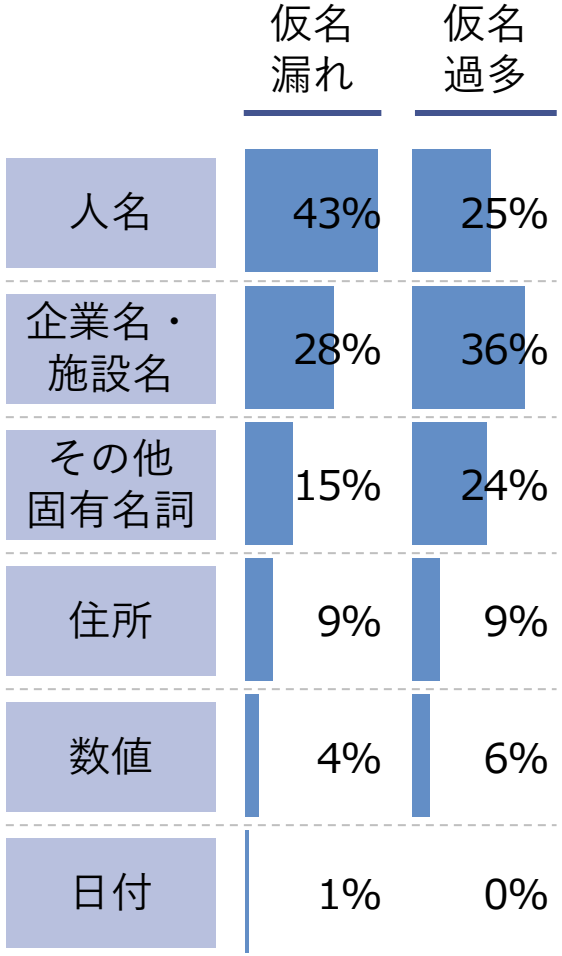
- 大規模なデータにより事前に学習した超大規模の統計モデルを活用
  - Google社による「BERT」
  - 今回は、日本語Wikipedia全記事（125万記事）を利用して学習
- 大規模なため一般的なPCでは利用できず、特殊な計算機環境が必要
- 前回PT時に比べ、特に仮名漏れが大幅に改善
  85.2% → 94.0%

生データ（ 第一法規様判例データ）の活用

- Wikipediaに加えて、第一法規様から受領した2.5万件の仮名処理後判例データを大規模統計モデルの学習に活用
  - 仮名処理後のみであり、処理前との対応はない
  - 約300万行、1.0GBのテキスト
- 大幅な性能向上は確認できなかった
  - 他の研究事例と比較して小規模（参考: Google社 - 33億語のテキスト）
  - 仮名処理後のデータであり固有名詞情報が存在しないため有効性が小さいと想定

手法：大規模な統計モデル（BERT）＋ルールの追加

- 一部分のみ、ルールにより誤り修正が可能
- ほとんどの誤りは固有名詞であり、ルール対応は困難



ルール設定による改善アプローチと結果

判例の構造情報活用

- 「当事者」節を直接ルール解析し、明示されている原告・被告名を把握

→ **1.7%**の仮名単語の改善

モデル出力仮名単語の活用

- 統計モデルがある単語を仮名処理していたら、全文書中で文脈によらずその語を仮名

→ 漏れが**2.5%**改善、一方で過多が**3.5%**増加

文字パターンの活用

- 日付、電話番号等をパターンにより抽出

→ 日付パターンでは著しく過多増加 **+27%**

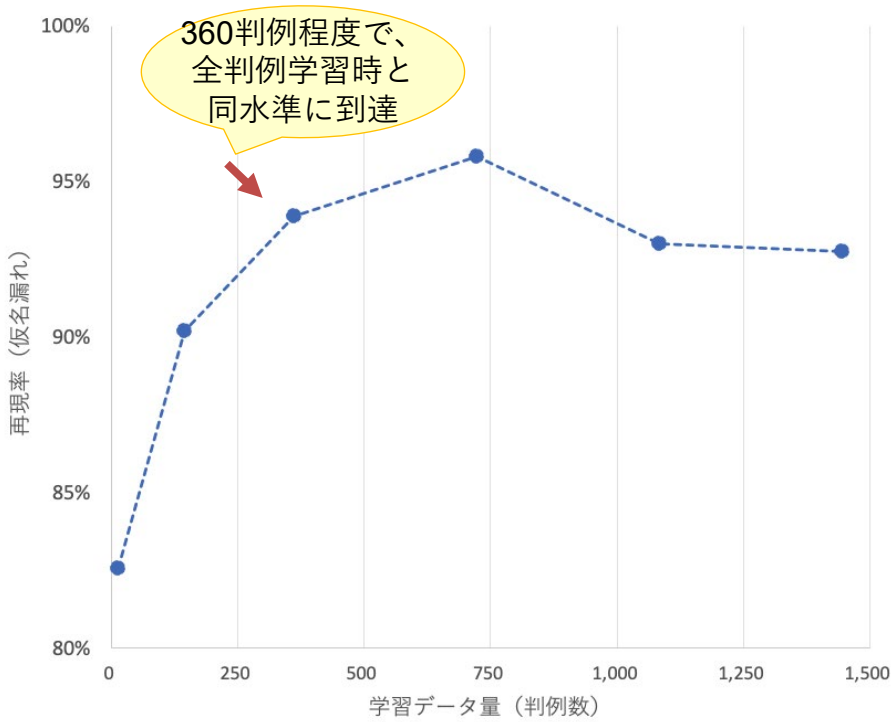
→ 他の有効なパターンでは 仮名漏れが**0.1%**減少

上記を総合しての最大の改善量
漏れ: 仮名単語 2.0%（仮名箇所 0.5%）
過多: 仮名単語 0.3%（仮名箇所 0.2%）

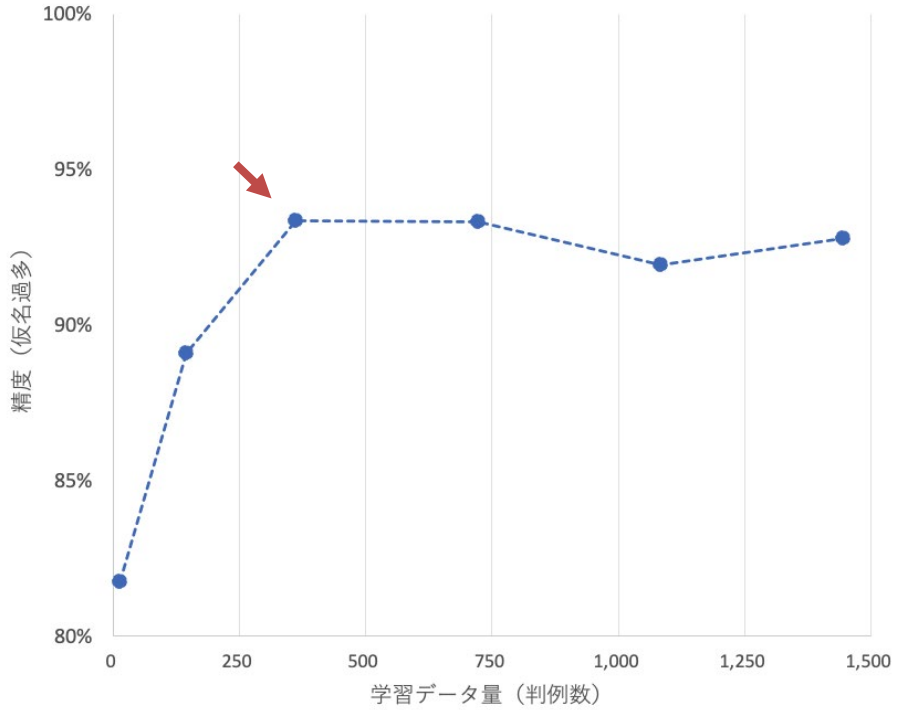
参考）学習データの追加による性能改善余地

現時点で入手可能性の比較的高い範囲の判例数（最大でも数万程度想定）の範囲においては、更なる判例データの追加による性能向上は限定的な可能性が高いと思料

再現率（仮名漏れ）の学習データ量別変化



精度（仮名過多）の学習データ量別変化



360判例（現在利用できるデータの25%）でも全判例利用による学習と同程度の性能
（追加するデータの中身によっては、逆に性能が低下する可能性も有り）

LEGALSCAPE
WE EMPOWER LEGAL PROFESSIONALS

本資料の内容

1. 実証実験の概要（目的と位置づけ）
2. 実証実験の結果：自動仮名処理の技術的可能性
   - ① 実験のアプローチ
   - ② 機械処理の精度水準
   - ③ 人手修正の費用対効果
3. 実務体制構築の方向性オプション（議論用叩き台）

p.3再掲）実証実験の流れ・位置付け

- 約1,600件の判例を対象に実験を行い、自動仮名処理の技術的な可能性を検証・評価
- 技術的可能性の評価を踏まえ、実務体制構築の検討材料を提示

仮名処理の実務体制を構築するまでの流れ

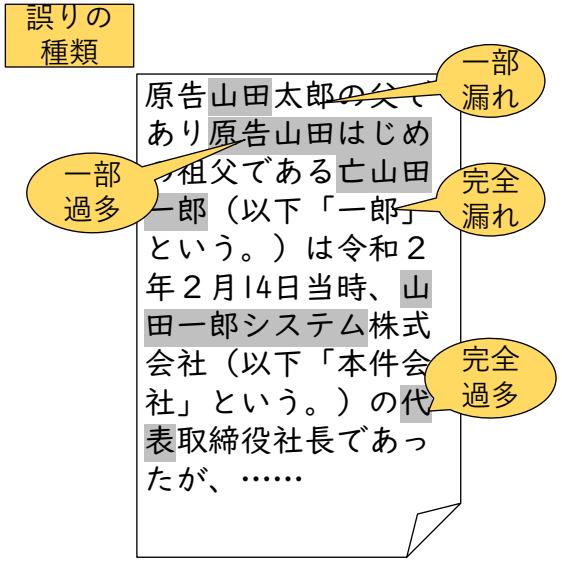
	1. 機械処理の精度水準を見定める	2. 人手修正の費用対効果を概算する	3. 実務体制の検討・構築を進める
検討論点	機械処理精度の水準は？ • 各誤りの質の違い、量は？ • どこまで向上可能か？	実務的観点から、どの程度、人手で修正すべきか？ • 修正する場合の作業量・費用は？ • 支援ツールでの効率化効果は？	左記の技術的可能性及び、本件趣旨／政策的観点を踏まえ、どのような実務体制を構築すべきか？
検討／実施アプローチ	**実証実験**		**システム開発・体制構築**
	本実験の範囲 約1,600件の判例データを対象に実験		システムの仕様策定・開発 支援ツールの開発に着手 実務体制（事業モデル含め）を検討・構築 • 公開対象者・データ利用目的等による処理精度水準の設定の違いがあれば考慮
	必要に応じて、実験規模・詳細度を上げる • データの追加 • 案件類型ごとの分析 （精度要求水準の異なる案件類型を設定・分類し、分析）		
	本資料にて結果を報告		本資料にて検討の枠組・材料を提示

LEGALSCAPE
WE EMPOWER LEGAL PROFESSIONALS

人手修正のオプション

- まず、人手修正の前段の機械処理において、漏れ／しすぎのトレードオフを調整する
- その上で、大枠の修正のオプションは、「判決文を全部見る」か「機械出力だけ見る」かの2つ

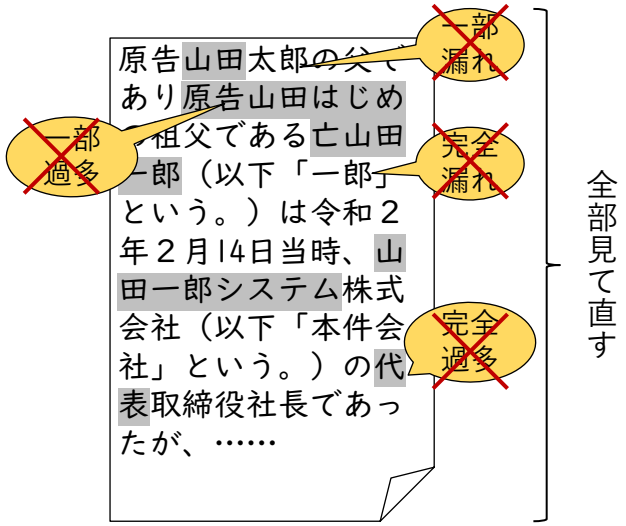
機械のトレードオフの調整



“漏れ”と”しすぎ”のトレードオフの間でバランスを取る

- ”全部黒塗り” ⇔ ”何もしない”

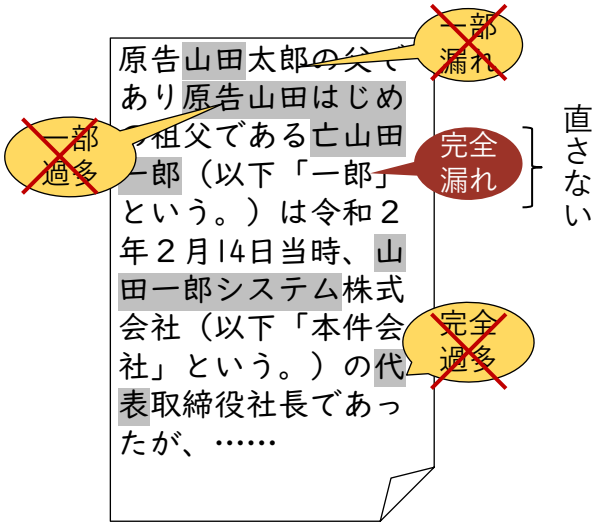
1 判決文を全部見て修正（全部修正）



完全な仮名を行うために、判決文全体を確認

- 全部見るが、修正ツールによるアシスト・効率化は可能（同一語句の同時選択／似た語句のサジェスト、等）

2 機械出力だけ見て修正（一部修正）



一部の「漏れ」を許容し、機械が判定した部分の周辺のみを確認

- 機械が判定した部分から、確認対象をさらに絞り込むことも可（人名／地名／番号、等）
- 機械で「多めに仮名」すれば、「漏れ」割合を一定低減可能

オプションの評価

- 仮名処理に係る費用は大きく分けて、人手処理の人件費と機械処理システムの費用
- 人手処理については、最大で（全文確認修正の場合）、2,500万円前後のイメージ

仮に10万件を想定した場合

				❶全文確認修正	❷機械出力のみ修正	3. 人手修正なし	参考
費用イメージ	仮名処理に係る費用	人手修正	1件あたり費用[1)]（時間）	250円（作業時間10分） ※従来作業時間が60分かかる判決を想定	75円（作業時間3分）	–	機械処理・支援を利用せず、人手で仮名処理した場合は、1件約1,500円程度かかる想定 ※仮名作業時間は約60分を想定
			10万件の場合	2,500万円／年 ※単純計算による概算	750万円／年 ※単純計算による概算	–	
		機械処理システム（開発費＋運営費）		（システム費用の考え方／方針検討については、次章記載） • 機械処理AIの開発・改善・アップデート • 作業ツール／システムの開発・保守			

人手修正作業の人員体制イメージ

年間10万件の判例を、1日ごとに処理・公開する場合は、最大でも**9**人体制で人手修正を施すイメージ（修正作業量を削減しない場合）

- 1日判例公開数＝約417件（10万件÷240日）※営業日ベース
- 9人体制であれば、1日約432件を処理可能（1人が1日で修正可能な判例数＝48件）

公開を1日ごとに行わない場合（例えば半年に一回）は、都度短期集中的に作業を実施

案件ごとの❶／❷の組合せや機械処理のチューニング等の個々の改善項目含め、
より詳細なオプションは存在するも、次章にて大枠の方向性を定めた上での詳細検討が望ましいと思料

1. 人手作業人員の時給を1,500円と仮置きし、試算

本資料の内容

1. 実証実験の概要（目的と位置づけ）
2. 実証実験の結果：自動仮名処理の技術的可能性
3. 実務体制構築の方向性オプション（議論用叩き台）

p.3再掲）実証実験の流れ・位置付け

- 約1,600件の判例を対象に実験を行い、自動仮名処理の技術的な可能性を検証・評価
- 技術的可能性の評価を踏まえ、実務体制構築の検討材料を提示

仮名処理の実務体制を構築するまでの流れ

	1. 機械処理の精度水準を見定める	2. 人手修正の費用対効果を概算する	3. 実務体制の検討・構築を進める
検討論点	機械処理精度の水準は？ • 各誤りの質の違い、量は？ • どこまで向上可能か？	実務的観点から、どの程度、人手で修正すべきか？ • 修正する場合の作業量・費用は？ • 支援ツールでの効率化効果は？	左記の技術的可能性及び、本件趣旨／政策的観点を踏まえ、どのような実務体制を構築すべきか？
検討／実施アプローチ	実証実験		システム開発・体制構築
	本実験の範囲 約1,600件の判例データを対象に実験		システムの仕様策定・開発 支援ツールの開発に着手 実務体制（事業モデル含め）を検討・構築 • 公開対象者・データ利用目的等による処理精度水準の設定の違いがあれば考慮
	必要に応じて、実験規模・詳細度を上げる • データの追加 • 案件類型ごとの分析 （精度要求水準の異なる案件類型を設定・分類し、分析）		

本資料にて検討の枠組・材料を提示

実務体制構築に向けた検討のステップ（議論用）

技術的可能性及び、本件趣旨／政策的観点を踏まえ、どのような実務体制を構築すべきか？

① 議論の出発点として、「民事判決オープンデータ化」の対象範囲・方式において、
どのようなゴール設定が、選択肢としてあり得るか？

- 全判決公開／一部判決公開、一般公開／特定事業者向け、等

② 各ゴール設定ごとに、必要となるシステム開発費用、準備・調整期間、等の
「取組／検討の“大きさ”」はどのように異なるか？

- 同時に、それぞれにおける残論点／詳細化すべき項目はどう変わるか？

本資料では、①②に関し、たたき台を記載

③ 取組／検討の“大きさ”を認識しつつ、
本件オープンデータ化の趣旨・政策的位置づけに立ち返った際、
現時点でゴールとして見定め、詳細検討を進めるべき方向性はどれか？

- 今回、重視すべき観点／政策としての評価基準は何か？
- 評価基準を踏まえ、選択・詳細化・磨くべき、ゴール／方向性はどれか？

④ 方向性踏まえ、今後詳細検討すべき事項及び検討ステップ・スケジュールは？

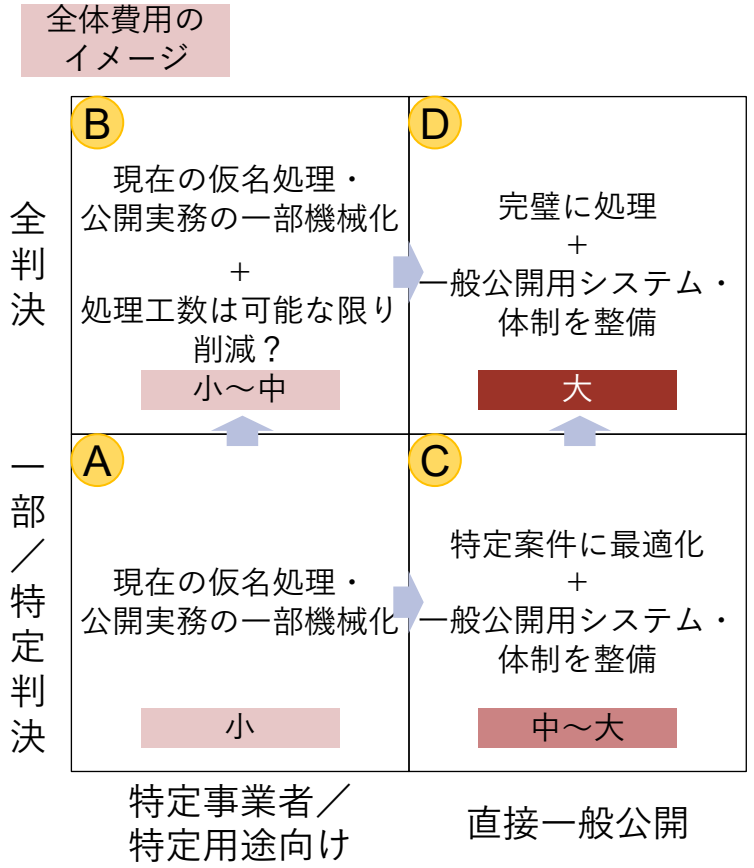
叩き台を元に、今後の方針を議論

本資料では、プロジェクト全体のシステム費用／準備・調整期間や、
今後の詳細検討項目が変わる分岐点（①＋②）を明確化し、議論のたたき台とする

①②実務体制構築の方向性オプション（議論用叩き台）

- 従前のPTでの議論を踏まえ、公開対象範囲・方式の組合せとして4つの方向性を仮置き
- 選ぶ方向性次第で、プロジェクト全体の費用／調整期間、検討項目が変わる想定

4つの方向性オプション（議論用叩き台）



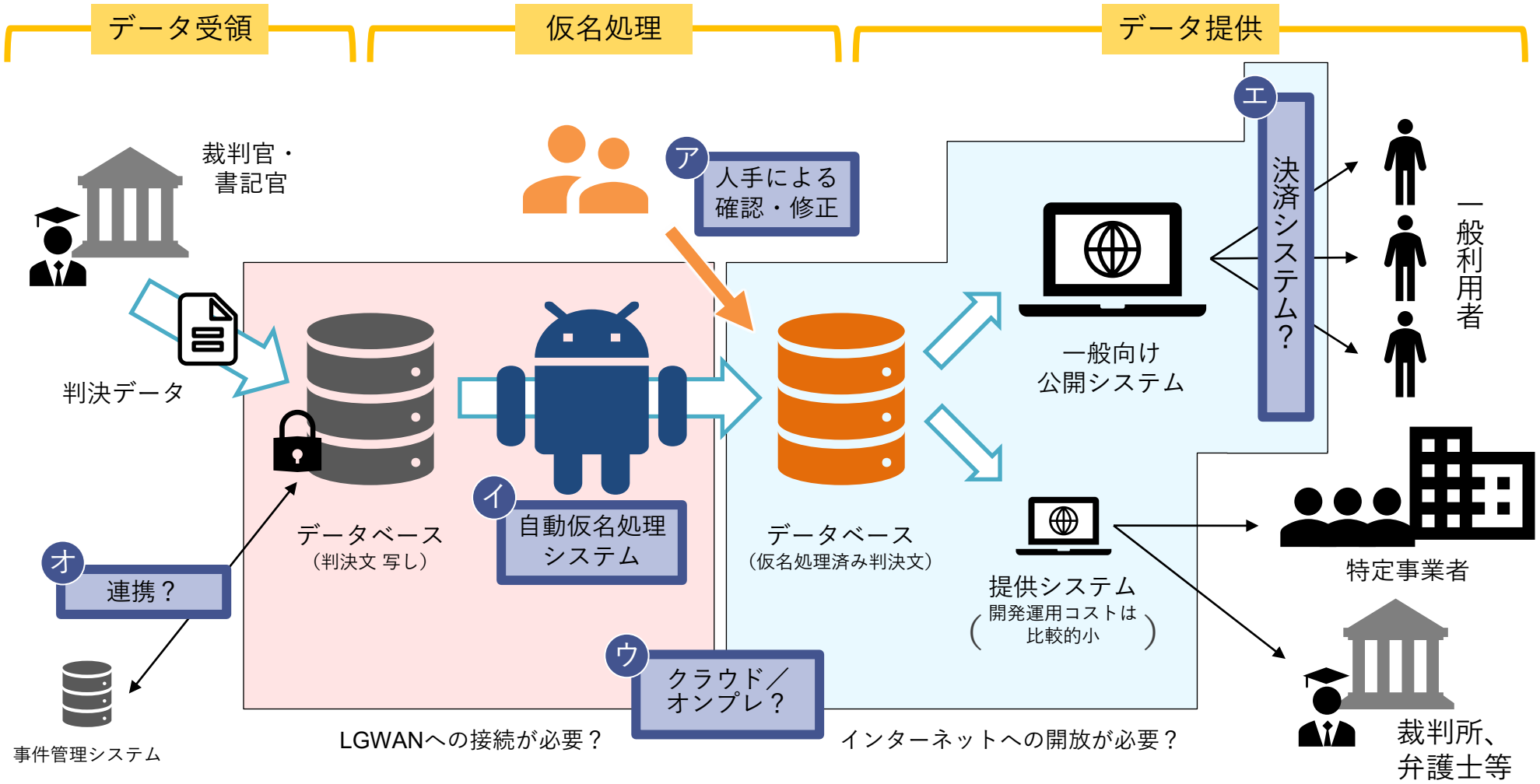
各方向性における取組のイメージ

	全体費用のイメージ	次に着手すべき検討項目イメージ
A • 現在、各判例データサービス事業者様にて行われている現行の仮名処理実務を機械処理により効率化／移管	小	• 持続的な座組の検討 • 仮名処理水準は現状と同水準が前提
B • 基本は現在の処理実務を効率化の方針 • 加えて、全件公開にあたり、処理工数削減余地を重点的に検討	小～中	• 持続的な座組の検討 • 特定案件／公開目的次第での、処理精度低減（修正コスト削減）余地の確認
C • 一般ユーザー向けのシステム開発、実務体制（問合せ等）構築の検討が必要 • 特定案件に鑑みた処理工数削減余地を検討	中～大	• 公開目的・範囲の整理 • その上で、必要処理精度／システム要件を詳細化
D • 一般ユーザー向けのシステム開発、実務体制（問合せ等）構築の検討が必要	大	• システムの範囲・仕様の検討 • 検討会議体・当事者の整理・調整

※費用の大まかな内訳・大きさ、変動要因については、次頁に記載

②システム費用の規定要因のイメージ（1/2）

ありうるシステム全体像を考えた際、費用を規定する分岐点は、仮名処理関連費用(㋐㋑)の他は、㋒クラウドか否か、㋓決済システムの有無、㋔他システムとの連携要否



②システム費用の規定要因のイメージ（2/2）

- いずれの方向性においても、システムのカバー範囲・仕様次第で、費用は大きく増減
- 公開目的／ゴールの詳細化と、システム要件・費用の具体化を両輪で進める必要

費用を規定する主な要因		A 一部判決→特定	B 全件→特定	C 一部判決→一般	D 全件→一般
仮名処理	ア 人手	処理工数は、「処理件数（全件／一部）」×「処理精度」が規定			
＋	イ 処理システム・ツール	機械処理AI＋作業ツール／システム			
データ公開システム	ウ クラウド／オンプレの選択	• オンプレの場合は、物理的なハードウェアの調達・設置・接続に加え、システム保守に係る人件費を要する • クラウドの場合は、サーバー容量・性能のみに応じて費用が増減 • セキュリティポリシー、等に応じて選択			
	エ 利用者情報の管理・決済システム等の有無	（データ提供事業者の数や、契約の流動性等次第も、基本は不要なイメージ）		• 個人情報・決済情報の管理、決済の実行システムの開発 • 提供対象（利用者）／提供方法、等の座組次第で要否が決まる	
＋					
裁判のIT化全体との連携	オ 事件管理システム等	（裁判のIT化は、長期的・大規模であり、本件オープンデータ化にて特定事業者・目的向けにデータ公開する場合は、取組対象として劣後するイメージ）		• 一般公開にあたり、利用者目線での「裁判システム」全体の整合性・連続性が必要等故、IT化検討との協働・協調が必要な可能性	

システム関連金額イメージ			
	初期	数千万円規模～（仮名処理費用がメイン）	～数十億円規模（公開用システムがメイン）
	運営・保守		

詳細な見積もり（初期／運営・保守費用の切り分け含め）のためには、より具体的な要件に基づく仕様検討が必要

LEGALSCAPE
WE EMPOWER LEGAL PROFESSIONALS

③④今後の詳細検討事項と進め方イメージ（案）

本年度PTにおいて体制構築の方向性を定め、次年度以降、座組・事業形式／システム要件・仕様含め、詳細化していく進め方が一案

	検討項目	時期イメージ（叩き台）
体制構築の方向性（現時点）を定める	1. 実務体制構築の方向性オプション（Ⓐ～Ⓓ）の内、どれを（優先的に）詳細検討するかを確定 2. 今後、詳細検討に入るために、どのような会議体・スケジュールを設定するかを検討 ➢ 協働すべき関連PT、当事者、等	本年度中（本PTにおける結論の位置付け）
方向性・ゴール設定の詳細化・明確化	1. 公開目的／ゴールの詳細化・明確化と、システム要件・費用の具体化を両輪で進める ➢ 費用が大幅に変動し得るシステム要件（前頁記載）については優先して有無・内容を検討 2. 公開機関の座組・事業形式を検討・確定する	来年度以降～
実際に、体制構築／システム開発に着手	• 具体化された実務体制方針、システム要件をもとに、予算確定・確保、システム開発を開始	

LEGALSCAPE
WE EMPOWER LEGAL PROFESSIONALS

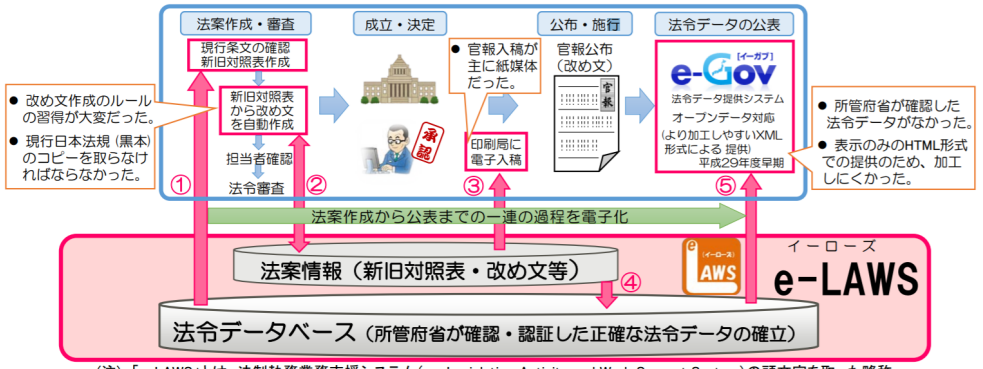
参考）総務省における事例

総務省・法制執務業務支援システムe-LAWSにて、採用されている法令データの作成・編集のためのシステム／インターフェースを参考にできる可能性

総務省公開資料[1]より：e-LAWSの概要

参考 法制執務業務支援システム（e-LAWS）の概要

- 「霞が関で働く女性有志」の提言を踏まえ、業務省力化・平準化の観点から、ICTを活用し、法案等関係資料の作成支援等を行うシステムの開発を行うことを決定（「国家公務員の女性活躍とワークライフバランス推進のための取組指針」（平成26年10月女性職員活躍・ワークライフバランス推進協議会決定）
- 総務省行政管理局において、一連の法案等作成業務を支援する「e-LAWS」を開発し、平成28年10月３日から各府省において本格運用を開始
- ①所管府省が確認・認証した正確な法令データを確立し、法令原本として活用できるデータベースを行政及び国民等へ提供、②新旧対照表から改め文を自動作成するなど、法案担当者の負担を大きく軽減



（注）「e-LAWS」とは、法制執務業務支援システム（e－Legislative Activity and Work Support System）の頭文字を取った略称

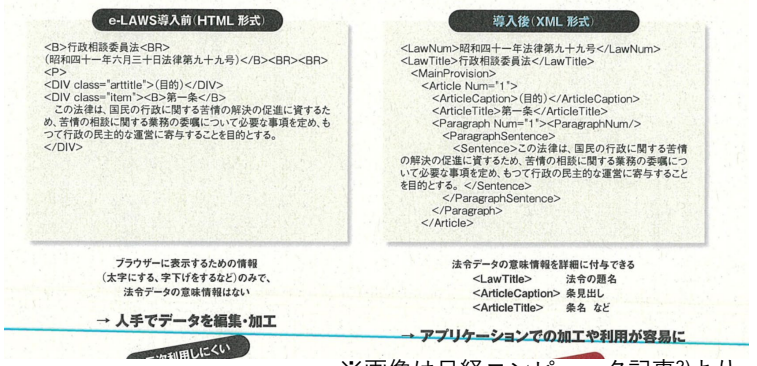
e-LAWS e-Legislative Activity and Work Support System

総務大臣談話

“e-LAWSの法改正支援ツールで、法令案の立案や参考資料作成の作業を飛躍的に省力化・効率化する”

法案作成・編集システム・インターフェースのイメージ[2]

法令データシステムのデータ形式を
機械可読性の低いHTMLから高いXMLへ変更



※画像は日経コンピュータ記事[2]より

法令作成者は、Wordのような直感的な編集ソフト／インターフェースによりXMLを編集

- 作成者は、XMLタグが直接見えない形式で法令データを作成・編集可能
- 編集結果をもとに（改正前後の法令の差分から）自動的に一部改正法令（改め文）を生成

→本件においても、**Word**等で書く際に、ハイライトやフォント変更のような動作で仮名対象を指定し、仮名化部分がタグ付されたデータを生成できる可能性

1. 総務省報道資料「法制執務業務支援システム（e-LAWS）の本格運用開始に際しての総務大臣談話」https://www.soumu.go.jp/main_content/000441588.pdf
2. 日経コンピュータ　2017年8月31日号 pp.44-45「紙の聖域に切り込む　戦後初、行政・司法のデジタル改革」

End of Document

資料５

民事判決の仮名処理等の在り方に関するＷＧ・論点整理案（仮）

【裁判所】　➡　【情報管理機関】　➡　【利活用機関】

Ｑ）提供根拠の整理　　Ｑ）提供根拠の整理

前→仮名化処理→後

仮名化基準の策定　※機械処理可能な程度の具体化を目指す

Ｑ）仮名化の要否

- 整理の視点　対立利益の調整（裁判の公開と訴訟関係人の権利利益）
  訴訟記録閲覧制度との関係の整理
  個人情報保護法の適用の整理

Ｑ）基準の策定

- 整理の視点　同上
- 対象裁判の整理（時点・種別・図表の仮名化）※事業化ＰＴでの整理による
- 判決類型による整理（公開の当否、基準変更の要否）
- 提供先による整理（基準変更の要否）
- **一般的仮名基準の策定**　※特に法人

Ｑ）適正化を図るための規律の要否　※事後的是正を図るための方策を含む。