

## 民事判決のオープンデータ化検討PT（第6回）

令和2年11月11日（水）  
13:00～14:30  
弁護士会館16階来賓室

### 議 事 次 第

- 1 冒頭挨拶等
- 2 実証実験の経過報告
- 3 今後の論点整理について
- 4 意見交換
- 5 次回以降の日程について

（配布資料）

- 1 民事判決のオープンデータ化検討PT構成員名簿及び出席者一覧（第6回）
- 2 株式会社 Legalscape 提出資料
- 3 論点整理に関する資料
- 4 町村泰貴教授提出資料

- 参考1 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）  
参考2 民事判決のオープンデータ化検討PT第7回以降の日程（案）

民事判決のオープンデータ化検討PT 構成員名簿及び出席者一覧（第6回）

2020年（令和2年）11月11日

◎：座長      ○：座長代理

	構成員	出席者（第6回・11/11）
◎	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	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	<ご欠席>
※	第一法規株式会社	田中英弥（代表取締役社長）
※		川原崎晶子（出版編集局編集第一部長）

	オブザーバー	出席者（第6回・11/11）
	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	出席者（第6回・11/11）
	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

# 機械による自動仮名処理に係る実証実験

－ 実験アプローチ解説・初期結果報告 －

2020年11月11日  
株式会社Legalscape

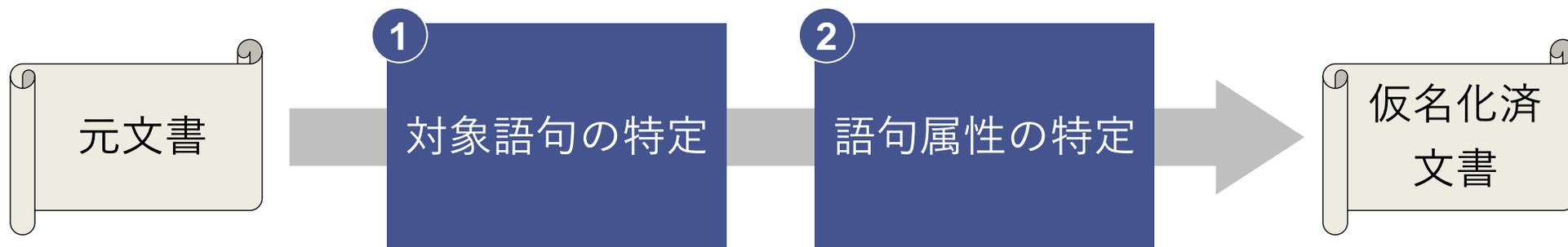
- 本実証実験（'20年10月-'21年1月予定）では、民事判決オープンデータ化検討に資することを目的とし、「機械による仮名処理の技術的可能性」を評価
  - 実験により、機械処理の精度、および誤り傾向を評価
  - 機械処理結果の誤りを人手により修正する場合の費用対効果を推定（人手修正については、機械処理結果を基にしたシステム補助による効率化も加味）
- 現時点で、機械処理の2つの段階「①対象語句の特定」「②語句属性の特定」のうち、①における機械処理精度を初期的に確認
  - 実験用学習データとして、別途受領した人手処理済データ約1,600件の内、まずは約800件<sup>1)</sup>を使用
  - データの学習により「①対象語句の特定」を実行する検証用システムを初期的に構築、精度を確認（学習モデルのチューニング、法律言語資源の追加、等は未実施）
  - 確認の結果、初期段階においても、約9割の精度は達成可能と判明
- 今後、最終報告に向け、機械処理全体の精度、および人手修正の費用対効果を検証・評価
  - ①の学習データの追加、学習モデルのチューニング、法律言語資源の追加、等による処理精度の向上余地を確認
  - ②も含めた全体での検証用システムを構築、処理精度を検証・評価
  - 上記の機械処理精度の評価を踏まえ、人手修正のアプローチを構想、コストを試算

本報告では、実験アプローチの全体像と現時点までの初期結果、今後の進め方についてご説明・ご報告

1. 機械学習の入力データとして、まずはデータ整形が不要なデータ（約800件）を使用。今後データ整形の上で入力データを増やす

1. 実験アプローチ
2. 初期結果
3. 今後の進め方

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（旧姓佐藤、以下山田）は東京都渋谷区にて...

- 山田太郎
- 佐藤
- 山田
- 渋谷区

本報告までに実施

- 原告: 山田太郎
- 旧姓: 山田 → 佐藤
- 同一: 山田太郎 = 山田
- 住所: 渋谷区

今後実施

原告X（旧姓X、以下X）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

# 処理誤りの意味合い

2つ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

## 「①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  
→ 閲覧者の利用性低下

## 「②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  
→ 閲覧者の利用性低下

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

# ① 対象語句の特定： 誤りの種類と評価指標 (1/2)

「仮名漏れ（偽陰性）」と「仮名過多（偽陽性）」という2種類の誤りが存在

## 誤りの種類

		正解	
		仮名対象	一般語句
機械出力	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%)

$$\text{再現率} = \frac{A}{A + C}$$

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%)

トレードオフ  
の関係

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

# 1 対象語句の特定： 誤りの種類と評価指標 (2/2)

1. 実験アプローチ
2. 初期結果
3. 今後の進め方

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」が存在

## 正答

仮名実施 (機械出力)

小山田太郎

仮名対象

人手修正には  
文書全体の確認が必要

## 完全な誤り

仮名漏れ

仮名せず (機械出力無し)

小山田太郎

仮名対象

仮名過多

仮名実施 (機械出力)

小山田太郎

一般語句

機械出力の確認のみ  
による人手修正が可能

## 部分的な誤り

漏れ + 過多

仮名実施  
(機械出力)

小山田太郎

仮名対象

- 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事が可能

(= 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)

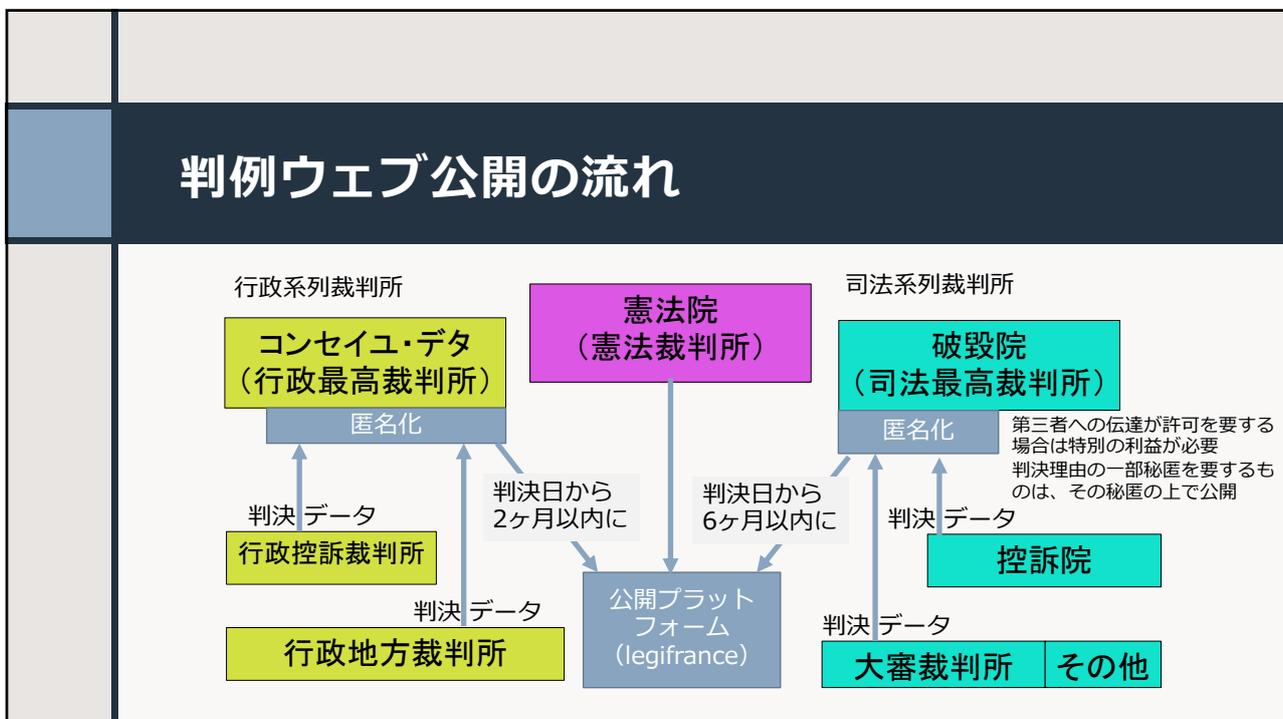
## 民事判決情報データベース化に向けた論点項目（たたき台）

令和2年11月11日

- 1 民事判決情報データベース化の意義
  - データベース化の目的、意義、利活用の可能性
  - 情報管理機関による一元的管理・提供の意義
  - 利活用機関との関係、提供のあり方
  
- 2 現行法制との関係について
  - 個人情報保護法との関係
    - ・ 「個人情報取扱事業者」（「個人情報データベース等」）該当性
    - ・ 適用除外事由（同法76条1項3号）該当性
    - ・ 適用される場合の規制と対応策
  - 裁判の公開との関係
  - 訴訟記録の閲覧等、秘密保護のための閲覧等の制限に係る規律との関係
  - 訴訟関係人のプライバシー等の法的権利・利益との関係
  
- 3 匿名化すべき情報、程度について
  - 典型的に匿名化すべき情報
    - ・ 個人の氏名・住所、生年月日
    - ・ 法人の名称
    - ・ 商品名等の固有名詞
  - 特異な情報であるため個人等が推知される情報
  - 訴訟手続外で公となっている情報と併せると個人等が推知される情報
  - 営業秘密
  - 利活用機関と匿名化の程度
    - ・ 特に、ビックデータとしての利用と匿名化との関係
  - 匿名化作業の合理化・効率化のための方策
  
- 4 その他検討すべき論点
  - 情報管理機関のあり方
  - 対象とする民事判決情報の範囲

	<h1>フランスにおける 判決オープンデータ化</h1> <p>Loïc Cadiet報告書を踏まえた 法・政令の整理（追加） 町村泰貴</p>

1



2

## 匿名化の対象と程度・行政系列の裁判所

- ✓ 当事者または第三者の自然人の氏名は当然に秘匿措置
- ✓ 当事者、第三者、裁判官、検察官、書記官を特定する要素（住所、電話番号など識別子）
- ✓ それらの開示がそれらの者自身または周辺の者の安全、プライバシーの侵害となりうる場合に秘匿される。その判断は当事者と第三者について裁判を下した裁判所が行い、裁判官、検察官、書記官の場合はコンセイユ・データ争訟部部长または行政控訴裁判所所長が行う。
- ✓ 裁判官、検察官、書記官を特定するデータ
- ✓ プロファイルによる判決等の予測に再利用することはできない。
- ✓ 以上の規定は行政裁判法典L.10条、R.741-14条による

3

## 匿名化の対象と程度・司法系列の裁判所

- ✓ 当事者または第三者の自然人の氏名は当然に秘匿措置
  - ✓ 当事者、第三者、裁判官、検察官、書記官を特定する要素
  - ✓ それらの開示がそれらの者自身または周辺の者の安全、プライバシーの侵害となりうる場合に秘匿される。その判断は、当事者、第三者の場合は裁判を下した裁判官が、それ以外は裁判所所長が行う。
  - ✓ 裁判官、検察官、書記官を特定するデータ
  - ✓ プロファイルによる判決等の予測に再利用することはできない。
- 以上は司法組織法典L.111-13条、R.111-12条による

4

	<h2>匿名化に対する不服申立て</h2>
	<p>関係人はすべて、いつでも、行政系列ではコンセイユ・デタの司法官に対し、司法系列では破毀院の司法官に対し、秘匿措置の実施または解除を求めることができる。</p> <p>ただし、濫用的と判断される申立ては認められない。</p> <p>以上の規定は行政裁判法典R.741-15条、司法組織法典R.111-13条による</p>

5

	<h2>GDPRの下での個人情報保護との関係</h2> <p>Cadiet教授の報告書n° 42-63</p>
	<p><b>GDPR以前において、裁判例の公開が個人情報保護の観点で適法であること</b></p> <p>公益のために許される（CNIL=仏版個人情報保護委員会2001年11月29日通達）</p> <p><b>GDPRの下で、再識別化が不可能という意味での匿名化は困難</b></p> <p>判決要旨の公開しか許されないとすると、判決の完全性を害し、裁判を受ける権利を害することにもなる。</p> <p>法は行政情報に匿名化を求めているのに対して裁判情報には仮名化と再識別のリスクを予め考慮するよう求めるのみで、差をつけている。裁判情報の再識別化を困難にする措置は合理的範囲内で行えばよい。このことはGDPR86条が公文書へのアクセス権と個人情報保護との調和の上で開示を認めていることに合致する。</p> <p>結局、判決情報の公開は仮名化と再識別化が困難となるような条件の下で、可能と解される。</p> <p>判決情報の再利用は、要配慮個人情報に相当する種類の情報についても公文書の利用を可能とすべきことや表現の自由との関係で、許容される。ただしその種類の個人情報再利用を禁止する国内法は改正を要する。</p>

6

# 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)

