

特許情報のオープンデータ化

——USPTOの取り組みを中心に——

栗 原 潔*

抄 録 再利用可能なデータを自由に流通させることで価値を得る「オープンデータ」の考え方に注目が集まっている。特許制度の目的のひとつが発明の公開にあることから当然に特許関連情報においてもオープンデータの思想はきわめて有効である。USPTO（米国特許商標庁、United States Patent and Trademark Office）は長きにわたりオープンデータ戦略を積極的に推進しており、単なる公報の公開に留まらず、審査経過情報や訴訟関連情報を含めた多様な情報を、リアルタイム性が高く、再利用が容易な形式で無償提供し、自由な利活用を推進している。これらのオープンデータを活用した革新的な民間企業の事例も出現しつつあり、特許関連データの付加価値をさらに高めている。これらのオープンデータ化の事例に学ぶべき点は多いと思われる。

目 次

1. はじめに
2. オープンデータの活用
 2. 1 オープンデータの価値
 2. 2 オープンデータ提供のポイント
 2. 3 特許情報とオープンデータ
3. USPTOのオープンデータ戦略
 3. 1 USPTOのIT戦略とオープンデータの位置づけ
 3. 2 USPTOのオープンデータの取り組み
4. USPTOオープンデータサービスの活用例
5. おわりに

1. はじめに

オープンデータとは「自由に使える再利用もでき、かつ誰でも再配布できるようなデータ」¹⁾と定義される。言うまでもなく、データには組織の戦略的資産として組織内に限定して管理すべきデータ（典型的には営業秘密）もあれば、公共財として広く公開することで万人が価値を得られるタイプのデータもある。後者にあたるデータがオープンデータである。ソフトウェア

（コンピュータ・プログラム）においても、企業独自のクローズドなソフトウェアと、社会的基盤として機能し得るオープンソースソフトウェアがあり、両者がそれぞれに価値を提供できるのと同様である。

近年になり、データが提供する社会的な価値に対する認識の一般化、および、大量データ処理テクノロジーの低廉化等の要因により、オープンデータへの注目が高まっている。典型的な動きは、オープンガバメントデータとも呼ばれる、政府等の公共機関が提供するオープンデータである。e-Gov（電子政府）の取組みとオープンデータは表裏一体の存在である。

米国政府は、オバマ政権時代よりオープンガバメントデータに力を入れ始めており、Data.govと呼ばれるオープンデータのポータルを運営している²⁾。日本政府も、IT総合戦略本部が平成25年に「電子行政オープンデータ推進のためのロードマップ」³⁾を定め、Data.govと同様のオープンデータポータルとしてData.go.jp等

* 弁理士 Kiyoshi KURIHARA

を開設している。

もちろん、オープンデータの提供は政府機関に限定されるものではない。アカデミアの領域では調査研究関連データの公開は以前から行われてきているし、民間企業がCSR（企業の社会的責任、Corporate Social Responsibility）の一環として、あるいは、企業イメージの向上のために自社のデータ資産をオープンデータとして公開することもあり得る。また、厳密な意味では、オープンデータとは呼べないが、有償でデータ資産（たとえば地図データ）を販売するビジネスもある。過去においては、このようなデータ販売は、特定の企業グループ向けに高額な料金で提供されることが多かったが、近年は、多数のユーザーに向けた比較的安価な月額料金またはアクセス数課金でデータ資産を提供しているケースもある。

2. オープンデータの活用

2. 1. オープンデータの価値

適切に収集・整理されたデータが大きな価値を生み出すことは論を俟たないが、それをオープン化することでさらにその価値を高めることができる。

たとえば、ある政府機関が、都市の渋滞に関する長期的な調査履歴データを保有していたとする。それを組織内の計画業務等に活用するだけではデータの価値を十分に発揮することはできない。

そのデータに基づいたレポートを作成し、ウェブ上で閲覧可能にすることは、オープンデータ化の第一歩である。データが組織内で限定的に使用されていた場合と比較して、データの提供する価値は増すが、まだ、その潜在的可能性が十分に発揮されているとは言いがたい。データの再利用がほとんど考慮されていないからである。

しかし、レポート作成の元になった生データを“再利用可能”なデータとして一般公開すれば、データの価値は飛躍的に増大する。たとえば、そのデータを活用して渋滞回避機能を備えたカーナビ・サービスやアプリを開発する企業が登場する可能性がある。多くの開発者が公正な競争を行うことで、様々な顧客の声に適切に応えられるサービスが提供され、消費者が選択の自由による利益を得られる可能性も高くなる。

さらに重要な点として、他のオープンデータ、たとえば、気象情報データと道路交通情報データを組み合わせることで、より正確な渋滞予測機能が提供できる可能性が生まれる。また、カーナビ・サービスに機能を付加した別のサービス（たとえば、タクシーと電車でどちらが早く目的地に着けるかを判定してくれるサービス）が開発される可能性も生まれる。データが相互に組み合わせられ、様々な者によって利活用されて価値を増していく、いわば“データの価値連鎖（バリューチェーン）”が実現できる。データの収集時には思いもつかなかった新たな活用法のイノベーションが生まれる可能性もある。同じデータであってもオープンデータ化することによって、特定の組織がデータを囲い込んでいた場合と比較して全体的な価値はるかに大きくなる。

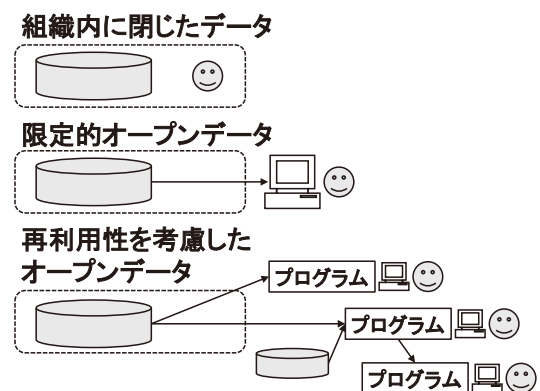


図1 オープンデータ概念

もし、上記の組織が公共機関であるならばデ

ータを国民に対する無料サービスとして提供することが好ましいが、私的企業であれば、他者によるデータの利用に対して課金を行うことも考えられる。こうすることで、価値が高いデータの収集や分析情報提供のインセンティブが生まれる。厳密には有料でのデータ公開サービスは「オープンデータ」の定義には合致しないとも考えられるが、言葉の厳密な定義の問題は別として、データの自由な流通と複数データ資産の統合によって新たな価値が生まれるという点に変わりはない。

2. 2 オープンデータ提供のポイント

本稿冒頭に挙げたオープンデータの定義「自由に使えて再利用もでき、かつ誰でも再配布できるようなデータ」からわかるように、真の意味のオープンデータを実現するためには、単にデータを公開するだけでは十分ではない。“再利用しやすい”こと、より正確に言えば、“コンピューター・プログラムによって再利用しやすい”こと（別の言い方をすれば、“機械可読性が高い”こと）が不可欠の要件である。

価値があるデータは様々な活用ができることでさらに価値を増す。画面に表示して読む、あるいは、紙に印刷して読むというのは、無数にあるデータ活用法のほんの一部に過ぎない。このほんの一部のみにフォーカスしてしまい、せっかくのデータの価値を十分に発揮できないようにしてしまう事例も見られる⁴⁾。データを再利用するために、印刷した書類の数字を手手で再入力するというやり方はきわめて不効率なやり方であり、そのような作業はコンピューターに任せなければならない。

データをコンピューター・プログラムによって再利用しやすくするには自己記述型の形式で提供することが重要である。自己記述型とはデータ項目そのものとそのデータ項目に関するデータ（メタデータ）が共に含まれている形式で

ある。現時点で、このような自己記述性を実現するための確立した標準としてはXML (eXtensible Markup Language) とJSON (ジェイソン, JavaScript Object Notation) がある。

たとえば、以下のようなXML形式データの一部を例にとると、「新佐 一郎」は審査官名であることが（人間にとってもコンピューター・プログラムにとっても）容易、かつ、正確にわかる。テキストやPDF形式でデータを公開する場合と比較して、コンピューター・プログラムによって自動的に集約や統計的処理を行うことがはるかに容易になる。

```
<Examiner><Name>新佐 一郎</Name>  
</Examiner>
```

オープンデータの公開方法は、バルクデータ提供（ファイル転送）型とAPI（Application Programming Interface）提供型とに大きく分けられる。大量データを利用者の元に置いて自由な分析を行いたい場合には前者が、リアルタイム性が高い少量データをプログラムから適宜アクセスして使用する場合には後者が望ましいため、オープンデータ提供者側としては両方の形態を可能にすべきである。また、前述のとおり、いずれの場合でも公開されるデータ形式はXMLやJSONなどの自己記述型の形式を使用すべきである。

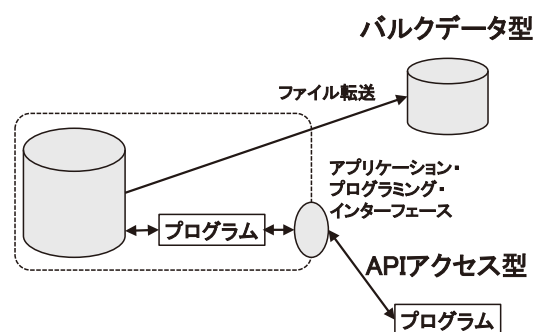


図2 オープンデータの提供方式

バルクデータ型で提供されるデータは、ある程度サマリーしたデータと詳細データに分けられるが、両方を提供するべきである。頻繁に使われる形式のサマリー・データを提供することは、ネットワークとサーバの負担を軽減する上で好ましいが、データ収集時には想定されなかった用途も含めて、多様な切り口での分析を可能にするためには、詳細データへのアクセスを可能にしておく必要がある。

2. 3 特許情報とオープンデータ

言うまでもなく、特許関連情報もオープンデータの重要なカテゴリーのひとつである。そもそも、特許制度の重要な目的のひとつに発明の公開がある。発明の一定期間の独占というインセンティブを与えることでその代償として発明の公開を強制して秘匿化を抑制し、重複開発を防ぐことが特許制度の重要な目的のひとつである。

特許関連情報としては、公報データ（公開公報および特許公報）も重要だが、審査経過関連データの重要性も無視できない。後の例で示すように、審査経過データを分析することで、特許実務家、および、特許庁自身にとっての重要な知見が得られる。

前述のとおり、公開とは単に紙ベースの情報を画面で閲覧可能にすることに留まるわけではない。画面表示あるいは印刷して読むということはデータの活用法のほんの一例に過ぎない。特許関連情報を公共財として再利用がしやすい形でオープンデータとして提供することは、オープンデータの中でもとりわけ価値が高い形態のひとつと考えられる。

3. USPTOのオープンデータ戦略

3. 1 USPTOのIT戦略とオープンデータの位置づけ

オープンデータとしての特許情報活用の参考

になる事例として、USPTOのオープンデータ戦略について見ていきたい。

USPTOのIT戦略文書「STRATEGIC INFORMATION TECHNOLOGY PLAN FOR FY 2015-2018」⁵⁾によれば、USPTOは審査事務手続きの効率性向上やレガシー・システムの現代化に加えて、「情報配布システムと透明性」を重要なIT戦略の項目として挙げており、特許関連データのオープンデータ化にフォーカスが当たっていることがわかる。

全体的に見て、USPTOのIT戦略は民間企業的な先進性とスピード感を感じさせるものである。GEエアロスペースやAOL等、民間での経験豊富な人物であるジョン・オウエンス氏をCIO（最高情報責任者）として採用している点にもよるのかもしれない。

3. 2 USPTOのオープンデータの取り組み

以下に、USPTOのオープンデータの具体的な取り組みの例を見ていこう。これらのデータはUSPTOのウェブサイト⁶⁾から、特別な申込み手続等なしに誰でも利用できる。なお、一般的な公報データそのものの公開についてはここでは触れていない。

(1) PAIR Bulk Data

米国特許の実務を行った方であれば、PAIR (Patent Application Information Retrieval) をご存じだろう。米国特許出願の審査経過情報(包袋情報)の検索・閲覧システムである。USPTOのサイトから、この情報をXMLまたはJSON形式で一括ダウンロードすることが可能になっている。全件ダウンロードだけでなく、たとえば、出願人、審査官名、日付等により条件を指定して一部のデータをダウンロードすることもできる。

これにより、特定の出願の情報を画面で閲覧するだけではなく、様々な統計的分析が可能に

なる。たとえば、後述の事例に挙げたような審査官別の統計情報は、単に画面上でデータを閲覧するだけでは得られない重要な知見である。

(2) Patents View

40年にわたる米国特許データを可視化できるウェブサイト（および、アプリケーション）である。地域別の特許出願状況、特許出願間の引用関係が非常に見やすい対話型の魅力的ビジュアライゼーションで提供される。紙面ではとうてい再現できないのでウェブ画面上⁷⁾で試してみることをお勧めしたい。



図3 Patents Viewの画面

前述のオープンデータのセオリーどおり、Patents ViewはAPIも提供しているので、他のデータやアプリケーションと容易に組み合わせることができる。たとえば、大学の教育レベルと特許出願件数の地域別相関関係、特許出願件数と地域経済指標との相関関係を可視化すること等も可能であり、興味深い洞察が得られる可能性があるだろう。

(3) PTAB API

PTAB (Patent Trial and Appeal) (審判部)のデータにアクセスするためのAPIが提供されている。データは日次で更新されている。現時点では、ベータ版での提供であるが、ウェブサイトには既にコメントが寄せられており、開発者

の意見を積極的に取り入れていることが窺える。

(4) Global Dossier

世界五大特許庁（米国，日本，EU，中国，韓国）の出願・審査情報を横断的に閲覧できるサービスである。複数国をまたがって同一特許ファミリーに属する出願を閲覧できる点に特に価値がある。日本のJ-PlatPat（特許情報プラットフォーム）が提供している「ワン・ポータル・ドシエ」のデータベースと同内容だが、USPTOのサービスでは入力履歴が保存されるなどユーザー・インターフェース上は多少の違いがある。もちろん、ヘルプ情報が日本語で表示されること等、J-PlatPatの方が日本の一般的ユーザーにとってはより使いやすいであろう。

(5) Assignments Data

1980年からの米国特許の譲渡状況を記録したデータベースである。USPTOのウェブサイト提供される譲渡データベースの検索機能はあいまい検索ができないなど機能が限定的だが、APIが提供されているため、利用者側で必要に応じてより複雑な検索機能を実現することも可能であろう。

(6) Patent Litigation Docket Reports Data

米国では、PACER (Public Access to Court Electronics Records) と呼ばれるシステムを通じて、原則的にすべての連邦裁判書類が（係争中のものも含めて）閲覧可能となっている。PACERのアクセスにはユーザー登録（クレジットカード登録）とシステム維持料としての料金支払いが必要であるが、USPTOではPACERで提供されるデータから、1963年～2015年の74,623件の特許訴訟関連データを抽出し、バルクデータとして無料ダウンロード可能にしている。

データは、Stataという統計ソフトウェアの

形式に加えて、一般的なcsv（カンマ区切り）形式でも提供されているのでExcelでも処理可能であり、単純な分析であれば手軽に行うことができる。たとえば、各年にある企業が原告又は被告となった特許訴訟の件数をカウントするなどの処理はExcelの標準的操作だけでも実行可能である。

(7) Patent Claims Research Dataset

特許公開公報および特許公報に自然言語処理のアルゴリズムを適用することで得られた様々なクレーム関連の統計情報がバルクデータとしてダウンロード可能になっている。もともとは、特許の権利範囲に関する研究論文作成のために収集されたデータであるが、他の研究者にとっての重複作業を防ぐためにオープンデータ化されている。

(8) Cancer Moonshot Patent Data

“Cancer Moonshot”とは米国政府が推進する癌の研究支援プロジェクトである。このデータは、全公報データから癌研究に関連する特許公報と公開公報のみを抽出したデータである。単純な特許分類による絞り込みだけではなく、審査情報も含めたデータ分析により複数の関連分野（医薬、診断方法、治療機器、遺伝子等）をまたがって、癌治療に関連する約27万件の公報を幅広く抽出している。

癌研究を行う組織にとっては、関連する特許文献の洗い出しは必然的に行わなければならない作業であろう。この作業をUSPTOが事前に行ってくれていることで、文献調査の手間を大幅に削減でき、より付加価値の高い研究活動に多くの資源を割けるようになるであろう。重複作業を防ぐという点で社会的な価値は大きい。まさに公共財としてのオープンデータが価値を提供している典型例と言えよう。

(9) Historical Patent Data Files

IPCやCPCによる特許分類は科学研究目的には必ずしも適切ではないことから、NBER（全米経済研究所、National Bureau of Economic Research）では独自の分類法が使用されている。この分類は、特許公報にしか適用されていなかったが、確率対応アルゴリズムの適用により公開公報等にも適用したものがバルクデータとしてダウンロード可能になっている。

(10) Data Visualization Center

個別のデータではなく、USPTOの審査状況の巨視的な管理指標を一覧するためのダッシュボード群であり、Patents Dashboard（特許）、Trademarks Dashboard（商標）、Office of Policy and International Affairs Dashboard（政策と国際情勢）、Patent Trial and Appeal Board Dashboard（特許審判）、Trademark Trial and Appeal Board Dashboard（商標審判）が提供されている。

Patents Dashboardを例にとると、最初のOAまでの平均期間、査定までの平均期間、未審査の出願バックログ数、RCEのバックログ数などの指標がメーターを模したグラフィック等により表示されている。データの更新はリアルタイムではなく、四半期単位程度である。

通常であれば、年次レベルで、かつ、紙ベースで公開される指標をできるだけリアルタイム性が高い形でビジュアルに提供する点がポイントである。このようなダッシュボード（コックピットとも呼ばれる）は、一般企業において、経営陣が業務パフォーマンスをリアルタイムで監視するためによく使われるテクノロジーである。USPTOも、あたかも民間企業のように、業務パフォーマンスを管理し、それを顧客（すなわち、企業知財部や特許事務所）にわかりやすい形で提供しようとしていることがわかる。

(11) Patent Statistics Reports and Data Products

ウェブ上でUSPTOの業務の年次サマリーレポートを様々な形式で作成するための機能である。あくまでもサマリー情報であって、詳細な分析機能が提供されているわけではないが、利用者がパラメーター（年度、技術分野等）を入力することで、カスタマイズされたレポートを作成できる点に特色がある。一般に、日本の政府機関における統計情報の公開は、印刷形式の出来合いのレポートを表示・印刷可能にただけであることが多いことから、このような情報公開のやり方には見習うべき点があるだろう。

(12) Open Data Portal

上記のオープンデータを含む様々な情報をひとつのポータル上で提供する取り組みである。前述のData.govやData.go.jpの事例からも明らかのように、利用者が、様々なオープンデータやアクセスAPIの中から自分の目的に合致したものを効率的に探し出せるようにするための共通窓口としてのポータルサイトの構築は重要である。現時点では、ベータ版での提供である。

4. USPTOオープンデータサービスの活用例

企業の知財部門や研究開発部門、そして、特許事務所等の組織は、USPTOが提供する上記の特許関連オープンデータを活用して様々な分析を行うことができる。もちろん、社内でプログラムを作成したり、スプレッドシートや統計ソフトを操作したりすることで様々な分析を行うだけでも大きな価値が提供されるが、より高度な分析を行い、その結果を顧客に有償で販売する付加価値サービスをビジネスとする企業も存在する。

この企業は、USPTOが提供するオープンデータを活用した様々な分析情報やビジュアルリー

ゼーション・ツールを、基本的には特許事務所向けの有料会員制サービスとして提供している⁸⁾。このサービスを利用した場合の登録査定率は利用していない場合の平均値を上回っており、登録までの期間も統計的に有意に短縮化できているとのことである⁹⁾。

公報や審査情報に基づいた分析データの提供を行う企業は日米共に多く存在するが、このサービスは情報の詳細度の点で特色があると思われる。たとえば、単なる平均的な審査期間やOA回数といった全般的な情報だけでなく、審査官ごとの詳細な情報が提供されている。



図4 審査官別情報の画面例

たとえば、出願の審査を担当する個別の審査官の平均OA回数がわかればその後に発生する費用や審査期間の目安が付けやすくなるであろう。また、個別の審査官に応じて中間処理の戦略を調整するなどの応用も考えられる。たとえば、ある審査官のRCE（継続審査要求）後の特許査定率が同一技術分野全体の平均よりはるかに低いのであれば、拒絶査定された場合にRCEは行わず、審判を請求した方が得策である等の判断を行うことができる。

もちろん、元となるデータは、USPTOから無償で提供されているので、誰でも同様の分析を行うことができる。組織内にリソースがある

のならば、自前で分析を行おうとする特許事務所や企業の知財部門があるかもしれないし、魅力的なサービスを提供できると考える企業が市場で公正な競争を行っていくこともあり得る。

さらに、他のオープンデータを組み合わせることで、データの価値をさらに高めようとする者も出てくるであろう。たとえば、審査関係のデータと訴訟関係のデータを組み合わせて分析することで、訴訟に勝つための出願戦略をアドバイスするサービスなども考えられる。このような分析を自社で行い独自の差別化要素とするか、公共の目的のためにオープンデータとして開放するか、有償付加価値データビジネスとして提供するかは自由である。いずれにせよ、前述の「データの価値連鎖」が実現され、データによって提供される価値が大幅に拡大されている点には変わりはない。これは、仮にUSPTOが公報データのPDF形式での提供といった限定的な公開しか行っていないのであれば実現できなかった価値である。

5. おわりに

組織の様々な意思決定においてデータが重要な役割を果たすことは論を俟たない。公共財としてのオープンデータの価値、そして、特許関連情報のオープンデータ化の価値についても同様である。

日本国内でも多くの特許情報分析会社が存在し、特許庁提供の公報等をベースに様々なデータ分析サービスを提供し、大きな価値をもたらしている。しかし、データ分析の価値は、元となるデータの再利用可能性、詳細度、範囲、そして、鮮度（リアルタイム性）に大きく依存する。その点では、USPTOの広範かつ積極的な

オープンデータ政策は特許関連データの一層の価値向上に貢献し、米国の特許制度自体の価値も高めていると思われる。また、訴訟関連情報のオープンデータ化についても、日米間で大きな差があると思われる（日米に法制度の違いはあるが、裁判情報を公開するという理念においては相違があるべきとは思われない）。司法関連のデータも含めた、日本の特許関連情報のオープンデータ政策について建設的な声を上げていくことが重要と思われる。

注 記

- 1) Open Knowledge Internationalのウェブサイト
<http://opendatahandbook.org/guide/ja/>
 - 2) <https://www.data.gov/>
 - 3) <http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryoushi.pdf>
 - 4) そのような事例のひとつとして、印刷した際の見栄えのみを追求し、データの再利用を考慮していない「神エクセル」とも呼ばれるスプレッドシートがある。（たとえば、「ネ申 Excel」問題 奥村晴彦（三重大学）
<http://oku.edu.mie-u.ac.jp/~okumura/SSS2013.pdf>
 - 5) <https://www.uspto.gov/sites/default/files/documents/strategicplan2015.pdf>
 - 6) USPTO Open data and mobility :
<https://www.uspto.gov/learning-and-resources/open-data-and-mobility>
 - 7) USPTO Patents View :
<http://www.patentsview.org>
 - 8) オープンデータの活用により明確なメリットが得られる可能性を示すためのものであり、サービスを宣伝・推奨する意図ではない。
 - 9) <https://www.juristat.com/>
- (URL参照日は全て2017年3月12日)

(原稿受領日 2017年2月17日)