

## 知的財産戦略に資するオープンデータの活用

情報検索委員会  
第1小委員会\*

**抄 録** 世界各国でデータのオープン化推進の動きが活発化しており、米国特許商標庁等の知財庁においても審査情報等の特許データのオープン化が始まっている。しかし、特許データや非特許データのオープン化が進んできている一方で、これらオープンデータを活用した分析手法が広く普及しているとは言い難い。ユーザーが独自にオープンデータを活用するためのハードルはまだ高いと言わざるを得ない。本研究では、オープンデータの定義やライセンスに関する情報への理解を深めると共に、オープンデータの活用手法およびオープンデータを活用した分析事例の提供を通じ、オープンデータ活用のハードルを下げることで、会員企業によるオープンデータの活用を促進し、知的財産戦略の高度化に貢献することを目的とする。

### 目 次

1. はじめに
2. 本研究の目的
3. オープンデータ
  3. 1 オープンデータの定義
  3. 2 オープンデータ活用の利点
  3. 3 オープンデータのライセンス
4. オープンデータの提供状況
  4. 1 日本のオープンデータ
  4. 2 米国のオープンデータ
  4. 3 欧州のオープンデータ
5. オープンデータの知財分析への活用
  5. 1 有価証券報告書データの活用
  5. 2 ナショナルプロジェクト・データの活用
  5. 3 世界銀行データの活用
6. おわりに

### 1. はじめに

世界各国において経済の活性化や新事業の創出を目的としたデータのオープン化が推進されており、政府が主導したオープンデータ・ポータルサイト（データカタログサイト等とも称さ

れ、オープンデータを一覧したり検索したりすることができるWebサイト）が開設され活用の促進が図られている。

我が国でも、国若しくは地方公共団体又は独立行政法人若しくはその他の事業者によって提供される官民データの活用促進に関する施策を推進することで安心安全で快適な生活実現に寄与するための法律として、官民データ活用推進基本法が2016年12月14日に施行されており、経済の活性化や国際競争力の強化につながるものとしてデータの活用が期待されている。

オープン化されているデータは多岐に亘り、一般的な非特許のオープンデータの他にも、米国特許商標庁（USPTO）等の知財庁では審査情報等の特許データのオープン化が進められており、知的財産戦略（以下、「知財戦略」ともいう）策定のための分析等において広く活用が望まれている。

\* 2017年度 The First Subcommittee, Information Search Committee

## 2. 本研究の目的

特許データや非特許データのオープン化が進み様々なデータが入手できる仕組みが構築されてきている一方で、オープンデータをどのように知財分析へ活用すれば良いのかについて、入門的な解説や活用の事例はまだ少なく、オープンデータの活用に対するハードルの高さが現在の課題と捉えている。

これからの知財分析においては、提供されるサービスを利用するだけでなく、オープンデータ活用による独自のデータ分析を通じ、高い付加価値を生み出していくことが求められているのではないかと考えている。

そこで、本研究では具体的なオープンデータ活用の手法や事例を提供することで、会員企業におけるオープンデータ利用のハードルを下げ、知財戦略の高度化に貢献することを目的として、オープンデータの活用手法の開発や事例の検討に取り組むこととした。

## 3. オープンデータ

### 3.1 オープンデータの定義

オープンデータの定義については、様々な説明がなされているが、本稿では定義の一例として、総務省のWebサイトに記載の定義<sup>1)</sup>に基づき説明する。オープンデータとは、「機械判読に適したデータ形式で、二次利用が可能な利用ルールで公開されたデータ」であり「人手を多くかけずにデータの二次利用を可能とするもの」を示す。同じく総務省のWebサイト<sup>2)</sup>では、オープンデータの5つの段階とデータ形式として、図1のような整理が行われている。

図1の整理によるとオープンデータと言うためには少なくとも2段階の条件を満たす必要があり、望ましくは特定の商用ソフトウェアのデータ形式ではなく、XML形式やCSV形式といったオープンフォーマットであることが求められている。即ち、PDF形式や画像のデータ等、人が判読するためのデータを単に公開するだけではオープンデータとは言い難いことが示唆されている。

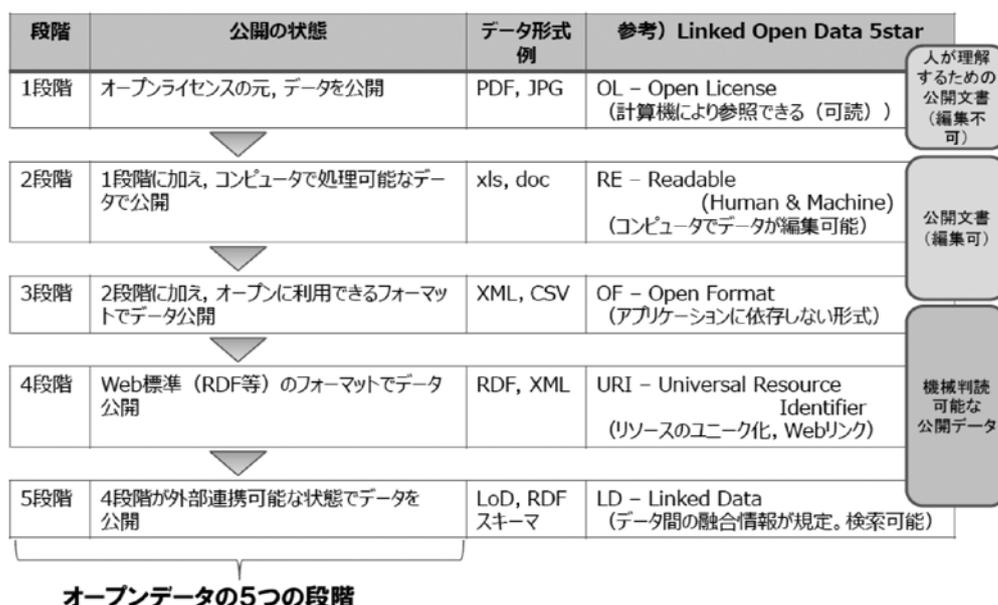


図1 オープンデータの5つの段階とデータ形式

また、利用条件の明確化のため、データを公開する際にはパブリック・ライセンスとすることが主流であるが、全てのオープンデータが必ずしも「誰もが自由に利用，再利用，再配布可能」なライセンスで提供されているわけではないことに注意が必要である。オープンデータのライセンスについては、3. 3オープンデータのライセンスで説明する。

### 3. 2 オープンデータ活用の利点

政府によるオープンデータの意義・目的として、電子行政オープンデータ戦略<sup>3)</sup>の中で、次の3点が挙げられている。

- 透明性・信頼性の向上  
公共データが二次利用可能な形で提供されることで、国民が政府の政策等に関する分析、判断を行うことが可能となり、行政の透明性が高まり、行政への国民からの信頼性を高めることができる。
- 国民参加・官民協働の推進  
官民の情報共有が図られることで、公共サービスや民間サービスの創出が促進され、我が国を取り巻く諸状況にも適切に対応することができる。
- 経済の活性化・行政の効率化  
公共データが二次利用可能な形で提供されることで、新ビジネスの創出や企業活動の効率化等が促され経済活性化が図られる。また、国や地方自治体における政策決定等においても、業務の効率化、高度化が図られる。

1点目は公共データ特有のオープンデータ活用の利点と言えるが、2点目および3点目については、公共データに限らず、オープンデータ一般に通じる利点とも言える。

知的財産分野においても、非特許データのオープン化が進むことで、特許データと組み合わせた分析をユーザー自身で行える幅が広がると思われる。例えば、有価証券報告書のデータを

容易に取り扱えるようになれば、特許データと非特許の企業データを組み合わせた分析を効率的に行うことができる。多種多様なデータのオープン化が進むことでこれまでの分析とは異なる切り口での分析から新たな知見を得られることが期待される。

### 3. 3 オープンデータのライセンス

オープンデータが著作権で保護されている場合がある。このような場合でも著作権処理を行うことなくデータの二次利用ができるよう、パブリック・ライセンスが取り入れられている。

パブリック・ライセンスは、著作権者が公衆に著作物の無断利用を認める範囲について意思表示を行うためのものである。パブリック・ライセンスの代表例はクリエイティブ・コモンズ・ライセンス<sup>4)</sup>（以下、CCライセンス）である。このライセンスには4種類の利用条件「表示 (BY)」、「継承 (SA)」、「非営利 (NC)」、「改変禁止 (ND)」を組み合わせた6種類がある（図2）。このライセンスにより利用者は、著作者が表示した利用条件さえ守れば、著作物を自由に二次利用することが可能となる。また著作者は著作権を保持したままデータを流通させることができる<sup>5)</sup>。

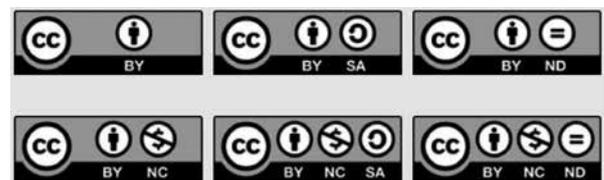


図2 CCライセンス一覧

前述のパブリック・ライセンスはCCライセンス以外にも種類があり、国が異なる場合はもとより、同じ国でもデータの公開者で異なるライセンスが適用される場合があるのでデータごとにライセンス情報を確認する必要がある。

各国のオープンデータのライセンスについて

説明する。

日本では、データのライセンス欄にBB-CYの表示があるものであって、公表者が著作権を有するものはCCライセンスの「表示4.0国際」により二次利用できる<sup>6)</sup>。ただしライセンス欄に「政府標準利用規約（第2.0版）」の表記があるものはこれに従う必要がある。

米国では、連邦政府が著作者となるデータについては、著作権で保護されずにパブリックドメインとなるため自由に二次利用できる。しかし、連邦政府以外の自治体が公開するデータについては各自治体の規約によるため注意が必要である。データが連邦政府かそれ以外のものかは、後述するDATA.GOVのデータ一覧に表示されるバナー（Federal等）により区別できる。

欧州では、データベース権およびPSI指令（European directive on the re-use of public sector information）に対する各国の方針の違いから、国ごとにライセンス形態が異なっている。例えば、英国およびフランスでは独自のライセンスを制定している。英国のOpen Government License（OGL）は、著作権だけでなくデータベース権の対象も二次利用できるようにしている。また、フランスのLicence Ouverteもデータベース権に対応しており、CCライセンスと互換性を持たせている。一方、ドイツではCCライセンスの「表示4.0国際」を採用している。

ライセンスをめぐる最近の動向に触れる。現在AI・機械学習のためのライセンス整備が行われ始めている。Linux Foundationは、機械学習に用いるトレーニングデータを扱うためのライセンスとして2017年10月にCommunity Data License Agreementを発表した<sup>7)</sup>。このような新しいライセンスの導入により今後ますますオープンデータの活用が促進されていくと予想される。

## 4. オープンデータの提供状況

### 4.1 日本のオープンデータ

日本では、経済産業省が2010年頃からデータのオープン化への取り組みを開始し、2013年には国内初のオープンデータ・ポータルサイトであるOpen Data METI<sup>8)</sup>を公開した。これにより経済産業省の保有データ170セットを公開し、他の府省庁に先駆けてノウハウ収集を実施した。その後、経済産業省は2015年に産業構造や人口動態などの官民ビッグデータを集約し、可視化するシステムとして、地域経済分析システム（RESAS）<sup>9)</sup>を公開、2016年には経済産業省版法人ポータルサイトを公開して実証実験を行った後に、利便性の向上を図ったシステムとして法人インフォメーション<sup>10)</sup>を2017年に公開している。

また、2014年には、内閣官房情報通信技術総合戦略室による企画・立案の下、総務省行政管理局が運用する、各府省庁が公開する公共データの横断的検索を可能としたオープンデータ・ポータルサイトDATA.GO.JP<sup>11)</sup>が公開されている。DATA.GO.JPでは、各データが省庁やデータの種別に応じて、タグ付けされており、そのタグからユーザーが欲しい情報を選択し、PDF形式、xls形式等の形式でダウンロードが可能となっている。

今後の計画として、政府は2016年に「オープンデータ2.0」として、2020年までを集中取組期間と定め、オープンデータの更なる深化を図るとしている。2017年に官民データ活用推進基本計画を策定し、8つの重点分野（電子行政、健康・医療・介護、観光、金融、農林水産、ものづくり、インフラ・防災・減災等、移動）を定め、官民データを活用して日本の持つ課題を解決するための施策を推進している。

本稿では、5. オープンデータの知財分析へ

の活用において、日本のオープンデータとして金融庁が所管するEDINET<sup>12)</sup>で公開されている有価証券報告書のデータ活用事例を報告する。

## 4. 2 米国のオープンデータ

米国では、オバマ大統領の8年間の在任期間中IT政策重視の路線を追求してきた。就任翌日、透明性とオープンガバメントに関する大統領覚書を発出し、この覚書を基盤として2009年5月に電子政府政策の中核をなす「オープンガバメント・イニシアチブ」を打ち出した。オープンガバメント・イニシアチブの目的は、政府の効率化や透明性向上だけでなく、政府の有する情報・データをインターネット上で誰もが利用しやすい形式で公開することにより、新たなビジネスの創出と経済成長につなげることである。これと同時に各政府機関によって公開されたデータへのアクセスを容易にする政府データポータルサイトDATA.GOV<sup>13)</sup>を開設し、同年12月には各政府機関に対してデータをオンライン上で公開することを義務付けるオープンガバメントに関する連邦指令を発出した<sup>14)</sup>。

2009年に開設されたDATA.GOVでは、政府機関、州、都市が保有する233,640のデータセットを公開している。(2018年3月時点)

各データの種別は地理空間データとそれ以外に分けられるが、地理空間データが全体の半数以上を占めている。また、それぞれHTML形式、PDF形式、XML形式、ZIP形式、TIFF形式、JSON形式等、多様なフォーマットで投稿されており、各データセットは農業、気候、消費財、環境、エネルギー等、14のトピックスや、各省庁、州、都市毎に分類されている。

これらデータセットの公開に加え、各々のオープンデータを活用したアプリサービスの情報開示(大気汚染の状況把握、健康的な食物をスーパー内で検索、医者検索、等)も行われており、オープンデータの具体的な活用に向けた取

り組みが促されていることが窺える。

本稿では、5. オープンデータの知財分析への活用において、米国のオープンデータとして、米国証券取引委員会(SEC)が所管するEDGAR<sup>15)</sup>で公開されている有価証券報告書に類する年次報告のデータ、および米国に本部を置く世界銀行が公開するデータの活用事例を報告する。

### (1) 米国特許商標庁のオープンデータ

米国特許商標庁(USPTO)のオープンデータ戦略については栗原氏による知財管理誌での論説<sup>16)</sup>にて述べられているので、本稿ではUSPTOのオープンデータ入手やその活用について述べる。

USPTOはOpen data and mobility<sup>17)</sup>というプラットフォームにてオープンデータを提供している(表1)。

Open data and mobilityでは、オープンデータを利用するための各種メニューが用意されており、データを可視化するためのツール、バルクデータのダウンロードサイトやアプリケーション・プログラム・インターフェース(API)が提供されている。

オープンデータの入手に関して以下に触れる。例えば、Bulk Data Productsというメニューは、オープンデータを含む様々な情報を扱うサイトになっている。この中のBulk Data Storage System (BDSS)というメニューでは、オープンデータカタログが用意されており、米国の特許や商標に関する出願情報や包袋情報の他、公報データや譲渡情報データなどが入手可能となっている。

Open Data Portalでは、登録公報や公開公報データを検索するためのAPIが、PTAB APIでは、USPTO審判部のデータにアクセスするためのAPIがそれぞれ提供されている。

Open data and mobilityでは、2017年にPAIR Bulk Dataが廃止され、同様のデータである、Patent Examination Data System(PEDS)<sup>18)</sup>が

追加された。PEDSは、USPTOが提供するバルクデータを入手するためのポータルサイトの的なメニューになっており、米国の出願情報や包装情報を、XML形式やJSON形式等でダウンロードすることが可能である。

表1 Open data and mobilityから入手可能なオープンデータの例

メニュー	サブメニュー	概要	データ形式
Patent Examination Data System (PEDS)		USPTOにおける出願や包装情報等のバルクデータをダウンロード可能。APIの利用も可能	XML, JSON
Bulk Data Products	Bulk Data Storage System (BDSS)	特許公報、公開公報、譲渡情報、包装情報、維持年金納付に関するデータ、その他商標に関する公報などが各種形式でダウンロード可能	PDF, TXT, TIFF, XML
Open Data Portal	Application Programming Interfaces	登録公報および公開公報データを検索するためのAPIを公開	
PTAB API (Beta)		PTABの審判情報等を検索するためのAPIを公開	

現在、PEDSは1981年出願以降の書誌情報や審査経過情報等を掲載しており、PEDSを利用することで、無料で検索し、必要な情報をダウンロードすることができる。Webインターフェースの他にもAPIが利用可能である。また、すべてのデータを一括でダウンロードすることも可能であり、ダウンロードしたデータを用いて独自のデータベースを構築することも可能である。

但し、現在のPEDSのデータは不完全であり、一部データの欠落がみられる。本件についてUSPTOに問い合わせを継続したが、本稿執筆時点においては有効な回答を得ることはできなかった。しかしながら、USPTOがオープンデータの整備に力を入れていることは紛れもなく、今後の整備状況によっては実務での活用に値するようになると思われる。

本研究では前述のAPIを活用し、Excel上で

米国特許の検索が可能となるマクロを試作した。USPTOが公開する特許情報がExcel上に直接取り込めるようになることで、既に利用している商用データベース等の検索結果と組み合わせ、新たな観点での分析のためのデータ補完を行うような活用を想定している。サンプルでは、Application Numberを入力してマクロを実行するとFirst Named Applicant等に加え、Group Art UnitやExaminer NameをPEDSからExcelに直接取り込むことができる(図3)。

なお、本ツールは新興国等における特許調査ツールとしての活用可能性も将来考えられることから、PEDSにアクセスするマクロのソースコードはクリエイティブ・コモンズのパブリックドメインライセンス(CC0)として当協会の会員専用のWebサイト<sup>19)</sup>で公開する。

<b>Application Number:</b>	13/799,098	←入力
<b>First Named Applicant:</b>	Apple Inc.	
<b>Filing or 371 (c) Date:</b>	2013/3/13	
<b>Status:</b>	Patented Case	
<b>Patent Number:</b>	9329685	
<b>Class / Subclass:</b>	340/665000	
<b>Group Art Unit:</b>	2682	←出力
<b>Examiner Name:</b>	WANG, JACK K	

図3 サンプルマクロの実行例

## (2) 米国食品医薬品局のオープンデータ

米国食品医薬品局(FDA)は、国民の公衆衛生の保護および向上を図ることを目的とする連邦政府機関である。医薬品、医療機器、食品などに関する規則の策定や承認を行っており、製薬メーカーや医療機器メーカーにとっては関わりが深い。FDAの提供するオープンデータについては、以下のデータベース等から利用可能である。

### 1) Drugs@FDA<sup>20)</sup>

Drugs@FDAは米国で承認された医薬品に関するデータベースであり、医薬品名、有効成分、申請番号で検索でき、承認情報や添付文書等を

参照することが可能である。

### 2) Orange Book<sup>21)</sup>

Orange Bookは先発医薬品と後発医薬品（ジェネリック医薬品）の生物学的同等性の判定を行い、その結果を掲載した発刊物であり、製薬業界を中心に広く活用されている。データベースでは、承認薬についてメーカー名、活性成分名、並びに特許情報が収録されている。特許番号で承認薬情報を検索することも可能である。

Orange Bookに収録されている承認薬はDrugs@FDAにも収録されているが、収録データに違いがある。例えばDrugs@FDAには特許情報は含まれていない。

### 3) openFDA<sup>22)</sup>

openFDAは2014年より提供されているデータベースで、医薬品、医療機器および食品に関して、有害事象報告、リコール情報、医薬品表示等に関するデータが提供されている。APIやCSV形式、JSON形式等が利用可能である。各データの検索または分析等で利用するためのリサーチツールも提供されている。リサーチツールの例としては、特定の薬品における有害事象報告数の推移を示すためのChange point analysis for adverse event reportsや、医薬品表示の検索が可能なDrug product labeling browser等がある。

これらFDAが提供するオープンデータは、医薬品メーカーや医療機器メーカーにとっては重要な情報源となっており、今後の知財分析への活用が期待される。

## 4.3 欧州のオープンデータ

欧州のオープンデータ政策は、2003年のPSI指令により、政府保有情報の再利用を可能にすることから始まった。その後、オープンデータ政策は英国を筆頭として緩やかに進められていたが、2009年から2011年にかけて続々と各国でオープンデータ・ポータルサイトの設立、ライ

センスの制定が行われ、急速に環境整備が進められた。今や30カ国以上が自国のポータルサイトを保有している。表2にその一部を紹介する。欧州全土でオープンデータへの関心が高いことが窺える。

欧州では、EUが提供するオープンデータ・ポータルサイトとして、EU Open Data PortalおよびEUROPEAN DATA PORTALがある。前者は、2012年に欧州委員会の文書の再利用を目的として開設された。後者は、2015年にPSI指令に基づき開設された。両者の違いとして収録データの違いが挙げられる。前者は、欧州機関が提供するデータを収録しているのに対し、後者は欧州各国のデータを公開することを主眼としているため、各国が公開したデータにアクセスでき、さらに前者のデータも収録している。データのカテゴリは科学、環境、経済、貿易、教育等と多岐に渡っており、フォーマットはデータごとに異なるがPDF形式、xls形式、XML形式等が見られた。ただし、ドイツやフランスのような非英語圏では、データ内容が英語以外の現地語であるケースが多く、利用する際は注意が必要である。

表2 欧州のオープンデータ・ポータルサイト

機関	Webサイト名
EU	EU Open Data Portal <sup>23)</sup>
EU	EUROPEAN DATA POTAL <sup>24)</sup>
英国	DATA.GOV.UK <sup>25)</sup>
フランス	data.gouv.fr <sup>26)</sup>
ドイツ	GOVDATA <sup>27)</sup>

欧州の中でもIndustry4.0の推進などで注目されているドイツの企業情報に着目して調査を行った。その結果、ドイツ企業の財務報告書や財務諸表を連邦ガゼットBundesanzeiger<sup>28)</sup>で閲覧できることが分かった。しかしながら、現時点では、EDINETやEDGARのように機械判読可

能なファイル形式でのデータ取得には対応していないことが判明した。また、企業情報に関しては、登記情報のWebサイトUNTERNEHMENS REGISTER<sup>29)</sup>において有償の場合があるものの、バランスシートや登記情報を確認することができた。しかし、これらのデータも機械判読可能な形式でのダウンロードはできないことが確認された。一方、ドイツの連邦金融監督庁(BaFin)<sup>30)</sup>には、過去1年以内の証券取引記録や目論見書のデータベースがあり、XML形式でのダウンロードが可能となっている。なお、欧州連合としては2020年より上場企業の財務諸表をEDINETやEDGARと同じXBRL形式<sup>31)</sup>で提出することが決定されている<sup>32)</sup>。欧州企業の財務情報分析の効率化が期待される。

本稿では、5. オープンデータの知財分析への活用において、欧州のオープンデータとして公開されている、ナショナルプロジェクトに関するデータの活用事例を報告する。

## 5. オープンデータの知財分析への活用

オープンデータとして提供されているデータを、特許情報と組み合わせることで新たな観点からの特許分析へつなげるための取り組み事例を報告する。本稿では分析結果よりも活用方法の提案に主眼を置くため、各分析事例における結論の評価を割愛することに理解を願う。

### 5.1 有価証券報告書データの活用

代表的な企業情報である有価証券報告書に注目し、オープンデータとしての提供状況、活用の可能性等について検討した結果を報告する。

#### (1) 有価証券報告書の情報について

日本では金融庁がEDINET、米国では米国証券取引委員会がEDGARというシステムで有価証券報告書の情報をオープンデータとして提供している。EDINET、EDGARともデータ形式

として財務情報の二次利用を容易とするために開発されたXBRL形式というデータフォーマットを採用している。これによりEDINET、EDGAR上の情報は、容易に表計算ソフトなどに取り込んで分析や加工を行うことができる。

有価証券報告書は公平かつ適時に開示し投資家を保護することを主目的にオープン化が進んでおり、政府のオープンデータ政策に先行していると思われる。たとえばデータフォーマットXBRL形式の普及推進団体(XBRL Japan<sup>33)</sup>)は、日米の有価証券報告書で採用されていること等を理由に、XBRL形式を政府データのオープン化における標準的なデータフォーマットの1つとして提案している<sup>34)</sup>。

#### (2) 有価証券報告書データ活用ツール

非商用利用では無償利用が可能なソフトウェアやAPIを使い、Excel上で有価証券報告書のオープンデータを活用できるツールを試作した。

試作したツールで利用しているソフトウェアやAPIの商用利用のためには必要なライセンスを提供元から取得しなければならない。また、非商用利用であっても提供元の利用規約等に沿った利用が必要である点に注意を願う。

##### 1) EDINET活用ツール

一般社団法人XBRL JapanがWebサイトで紹介している「有報キャッチャーEXCELツール」<sup>35)</sup>を利用し、Excel上で該当セルに企業名(出願人名)を入力すると、対応するセルにその企業の過去数年の売上高、純利益、研究開発費などの財務情報をEDINETから自動ダウンロードして表示する機能(図4)を実現した。

この機能は、基本的には以下のA)~C)をExcel上に記述することで実現した。

- A) 予めEDINETからEDINETコードリスト<sup>36)</sup>を入手し、企業名、証券コードおよび決算日に対応付けた対応表(表3)を作成。
- B) 上記ツールにより提供される関数(式1)

A	B	G	L	P
企業名 (出願 人名)	売上: 16年4 -17年	利益: 6年4月 17年3月	研究開発 6年4月 17年3月	:百万円 12年4月 13年3月
A社	514,	54,11	33,74	
B社	359,	11,60	11,51	
C社	249,	40,63	17,01	
D社	166,	9,14	6,46	
E社	121,	8,77	2,58	3,333
F社	72,0	2,54	3,49	
G社	55,5	1,19	1,45	

企業名を入力すると  
該当する企業の過去の売上高/純利益/研究開発費をEDINETからダウンロード、Excelの該当セルに出力/表示

図4 EDINET試作ツール動作例

にて、該当企業・該当年度に対応する有価証券情報があるURLを入手。引数となる securityCode (証券コード), baseDate (決算日) は、A) の対応表を参照して入力。

C) 上記ツールにより提供される関数 (式2) にて、B) で入手した有価証券情報のURL から売上高などの財務情報を入手。

- (ア) url  
B) の手順で入手したurl
- (イ) elementName  
該当する要素名 (NetSales等)  
baseDate (決算日)  
A) の対応表から記載
- (ウ) periodStart (期間開始日)  
決算日から1年前の日付

UCXEdinetXBRLInstanceLatestASR (security Code, baseDate) … (式1)

UCXFact (url, elementName, baseDate, period Start) … (式2)

表3 対応表の例

企業名	証券コード	決算日
A社	137X	2017/5/31
B社	775X	2017/5/31
C社	67XX	2017/3/31
D社	232X	2017/3/31
E社	...	2017/6/30
F社	...	2017/6/30
G社	...	2017/3/31

前述したツール (有報キャッチャーEXCEL ツール) のダウンロードファイルには他にも具体的な使用例が含まれているので、応用にあたっての参考となる。

また、財務情報と別途入手した特許情報とを利用したグラフを自動作成することも可能である。(図5)

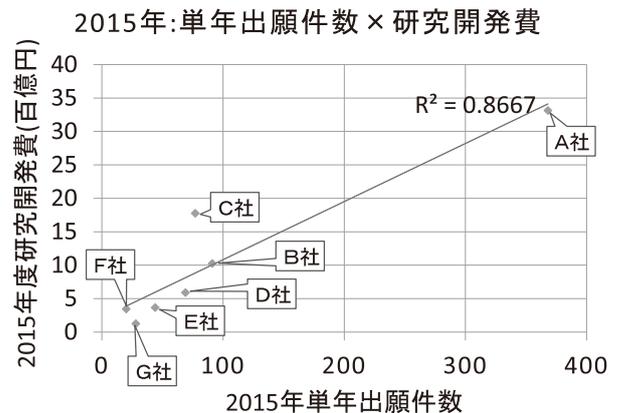


図5 特許情報と財務情報の相関図 (例)

## 2) EDGAR活用ツール

投資情報配信企業が提供するEDGARデータのWebサービスAPI<sup>37)</sup> を利用し、Excel上で調査対象企業に対応するティッカーシンボル<sup>38)</sup> を入力することで、該当する企業の過去の売上高などの財務情報をEDGARからダウンロードすることができるツールを試作した。

まず、前述のAPIを利用するための利用登録が必要である (無料)。Webサービス提供元に

アクセスし、利用登録<sup>39)</sup>を行う。利用登録の手順については提供元のWebサイトを参照願う。

利用登録、サインイン、および設定を行うことで、Application keyを入手することができる。Application keyはAPIの実行のために必要な識別子である。サンプルで提供する関数例を利用する際には、それぞれ取得したApplication keyを使用することが必要である。Application keyの入手手順は提供元のWebサイトを参照願う。なお、その際に求められる入力項目「Web Site」は入力必須であるが内容は自由である。

APIによって得られるEDGARの情報は多岐にわたり、詳細はサービス提供元のWebサイト<sup>40)</sup>を参照願う。サンプルとして、下記にExcel上にEDGARから、(式3)の【Tickersymbol】に対応する企業の2013年度～2017年度のTotal revenue, Net Income, Operating Profit, Research Development Expense, Selling, General & Administrative Expense等をXML形式で取り込む関数を提供する(式3)。なお、(式3)の【Application key】には前述したApplication keyを記述する。

(式3)の関数で取り込まれたXML形式のデータはそのままでは取り扱いが煩雑である。ExcelにはXML形式から必要な情報を取り出す関数として、FILTERXML関数が提供されている。サンプルとして、取り込んだXML形式のデータが格納されているセルの番地を【XML】に記入すると2013年度のTotal revenueに対するデータを取り出す使用例を提供する(式4)。

なお、Total revenue以外のフィールドについては、式4での@field='totalrevenue'に対応するフィールド名を記述する。詳細については提供元のWebサイトおよび取り込んだXML形式のデータの内容を参照願う。また、row [1]の [ ]の中の数字がEDGARから取り込んだデータの年度に対応しており、サンプルでは2013年から取り込んでいるので、最初の2013年

度が1に対応している。以降の年度は2, 3と増やすことで取り込むことができる。試作ツールの動作例を図6に示す。

WEBSERVICE (“http://edgaronline.api.mashery.com/v2/corefinancials/ann.xml?fields=IncomeStatementConsolidated&primarysymbols=【Tickersymbol】&fiscalperiod=2013q1~2017q4&sortby=fiscalperiod+asc&appkey=【Application key】”) … (式3)

FILTERXML (【XML】 , “//response/result/rows/row[1]/values/value[@field='totalrevenue']”) … (式4)

Apple	FY13	FY14	FY15
Total revenue	\$233,715,000,000	\$170,910,000,000	\$233,715,000,000
Net Income	\$39,510,000,000	\$37,037,000,000	\$53,394,000,000
Operating profit	\$52,503,000,000	\$48,999,000,000	\$71,230,000,000
Research Development expense	\$6,041,000,000	\$4,475,000,000	\$8,067,000,000
Selling General Administrative expenses	\$11,993,000,000	\$10,830,000,000	\$14,329,000,000

図6 EDGAR試作ツールの動作例

## 5.2 ナショナルプロジェクト・データの活用

近年、AI, IoT, ビッグデータ等を始めとするデジタルテクノロジーの普及に伴い、例えば、タクシー業界に対するUberや民泊業界に対するAirbnb等のベンチャー企業による、既存産業への新たなイノベーションおよび破壊をもたらす新規参入が目立ち始めている。これらビジネス環境の変化は、同一業界において従来から活動している企業にとって無視できない脅威となっているため、注視し続けなければならない。

しかし、これらベンチャー企業は特許出願の件数がまだ少ないケースが多い。そのため単に

同一業界内の特許出願件数に注目し、分析してしまうと、従来から同一業界で活動していた企業の出願件数に埋もれてしまう。そのため、マクロ観点でこれら注視すべきベンチャー企業の出現を早期に予見することが難しい。

これに加え、業界・技術分野によっては各企業のオープン&クローズ戦略に基づき、重要技術であってもあえて出願せず自社の重要ノウハウとして秘匿する方針がある。そのため前述と同様に、単に特許出願件数や各特許データから読み取れる情報を基軸とした分析を行ってしまうと、対象業界・技術分野の研究開発動向や競合企業の把握が十分に判別できないケースも存在する。このような昨今の状況から、特許出願件数の動向に依拠した調査だけでは自社ビジネスの業界内における競合企業の把握が一層難しくなっている。

本稿では、特許データの分析だけでは埋もれてしまう企業の把握可能性を検討するために、ナショナルプロジェクトに関するデータ（ナショナルプロジェクト・データ）を活用した。ナショナルプロジェクトを受託する企業であるからには、当該ナショナルプロジェクトに寄与するアドバンテージ（技術、人材、コネクション、等）が期待されている企業であると考えられる。

そこで、ナショナルプロジェクト・データを用い、特定の技術分野におけるナショナルプロジェクトに参加している企業を抽出し、特許データ分析からは得られない競合把握に活用できないか検討した。

検討にあたり、本稿では、EU Open Data Portalのオープンデータから、Horizon 2020<sup>41)</sup>の参加者企業のデータ<sup>42)</sup>を調査した。

プロジェクト5GExおよびSESAMEにおける参加企業の累積出願件数の分布は図7および図8の通りである。

5GExではREDZINC SERVICES LIMITED, BERLIN INSTITUTE FOR SOFTWARE

DEFINED NETWORKS等、SESAMEではVIRTUAL OPEN SYSTEMS, ATHONET SRL等を始めとし、両プロジェクトともに特許出願件数が10件以下の参加企業がおよそ半数を占めていることが分かった。



図7 5GEx参加企業の累積出願件数分布

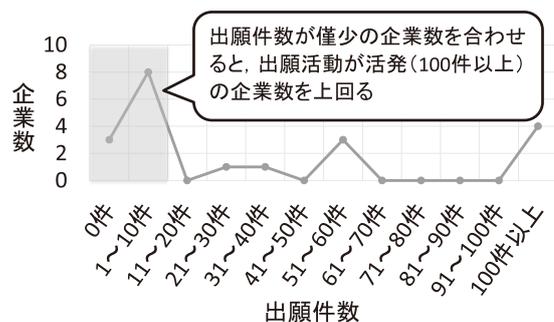


図8 SESAME参加企業の累積出願件数分布

前述のとおり、出願件数としては皆無もしくは僅少であってもナショナルプロジェクトへの参加企業であれば何らかのアドバンテージを保有しているはずであり、その技術分野によっては、自社にとって将来の競合企業となる可能性が考えられる。しかし、例えば出願の有無や出願件数のランキングによるベンチマーク等のような、特許データに偏った分析、判断を行ってしまうと、前述のような出願件数が皆無もしくは僅少である一方、ビジネス観点からは注視すべき将来の競合企業を見落としてしまいうることが分かる。

以上の結果より、オープンデータを活用することで、特許データ分析上の注目すべき企業だ

けでなく、技術・事業上の注目すべき企業の手がかりをつかめることが分かった。

また、ナショナルプロジェクト・データについて、日本においては、オープンデータ・ポータルサイトDATA.GO.JPから「経済産業省の補助金等の交付決定に関する情報」のデータを取得できるが、収録年が平成22年度～平成24年度であり、かつHTML形式で格納されているため、分析するうえで扱いにくい点が見受けられた。一方、経済産業省の同データの格納先<sup>43)</sup>へアクセスすると、上記に加え、平成25年度～平成29年度のデータをxls形式で取得することが出来る。なお、国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）からの公募情報は、NEDOのWebサイト<sup>44)</sup>から確認することが出来たが、いずれのオープンデータ・ポータルサイトにも格納されていなかった。

これら日本におけるナショナルプロジェクト・データの開示状況は、欧州と比べるとまだ整備の途上である様子が見受けられた。ナショナルプロジェクト・データは、特許データからでは読み取れない企業のピックアップなどに役立てられることから、今後の整備に期待したい。

前述の通り、これからの知財戦略には、特許データからだけでは抽出、判別しづらい企業を、これらオープンデータを活用することによる新たな観点から見出し、調査対象に加えた分析を行った提言も求められているのではないだろうか。知財戦略において、量での議論が中心であった従前よりも質を重視する流れの中では、量的には少数であっても注目すべき特許をいかに見出し、分析するかが今後重要になってくるものと思われる。具体的には、件数は少ないが質の高い特許を所持する中小規模の協業相手の探索や、外国の有力ローカル企業の探索などでも役立つと考える。これらをもとに有力な協業相手が探索できれば、自社特許のライセンスによる事業拡大のサポートや、共同研究による新た

な知財獲得など自社の知財戦略に組み込む検討ができる。

### 5.3 世界銀行データの活用

世界銀行が各国の様々なデータをオープンデータとして公開していることを受け、特許データとを組み合わせることによる新たな観点からの特許分析について検討した。

世界銀行のWebサイト<sup>45)</sup>上の「データ一覧」を選択することで、オープン・データ・イニシアティブというオープンデータ・ポータルサイトにアクセスできる。そのサイトで情報を取得したい国名を選択したうえでCSV形式、XML形式、xls形式などのデータ形式でデータを一括取得することができる。データ項目によってはデータが格納されていない場合もあるが、古いものでは、1960年からデータが格納されている。なお、取得可能なデータには、広く活用されている国内総生産（GDP）以外にも知財分析に活用が考えられそうなデータとして例えば表4のようなものが公開されている。

表4 世界銀行で公開されているデータ（抜粋）

ハイテク製品輸出（輸出製品に占める割合）
総人口
人口増加率（年率）
1人当たり国内総生産（名目米ドル）
インフレ率、GDPデフレーター（年率）
軍事費（対GDP比）
中央政府債務総額（対GDP比）
研究開発費（対GDP比）
二酸化炭素排出量（1人当たりメトリックトン）
エネルギー消費量（1人当たり石油キロ換算）
電力消費量（1人当たりキロワット時、kWh）
上場企業の時価総額（対GDP比）
サービス等、付加価値（対GDP比）
工業、付加価値（対GDP比）
農業、付加価値（対GDP比）

今回は世界銀行で公開されているデータのうち、GDPおよび二酸化炭素（CO2）排出量と特許行政年次報告書で公開されている諸国の特許出願件数の推移とを組み合わせることで、新たな出願国となる可能性がある国を予測できるかを、過去の歴史と比較して検証した。

CO2排出量やGDPは市場となりうる国の基盤整備に応じて増加する値である。一方、特許出願件数は市場としての基盤が確立した国（地域）に対する市場としての期待値であると考え、①産業の発展に応じCO2排出量が増加し、②発展した産業によってGDPが上昇し、③産業保護のため特許出願が増加するとの仮説を立て、CO2排出量およびGDPの変動から、特許出願件数の増加の予測を試みた。

仮説検証のため、既に市場が成熟している日本と、近年経済成長が著しいことが明らかな中国について、前述の①②③の関係を確認した。

本研究では（国民1人あたりの）CO2排出量とGDPとのデータについて、長期的傾向をみるために、100%を基準とした前年からの変化率の後方10年の移動平均値を指標として用いて分析を行った。また、日本と中国の特許出願件数は特許庁発行の特許行政年次報告書のデータを利用した。

日本では、図9を参照すると、1990年頃まで一旦停滞していたCO2排出量の増加が再び始まり、1992年頃にはGDPの増加率の上昇傾向がみられた。その後1995年頃から停滞していた特許出願件数の増加が確認された。

同様に、中国では、図10を参照すると、2002年頃まではなだらかな増加をしていたCO2排出量が、急増するのに伴い、2005年頃からGDPの成長率が増加したことが確認された。その後2009年頃からの中国における特許出願件数の爆発的な増加は記憶にも新しい。特許出願件数の増加には経済成長の他にも、当該国の知財政策が強い影響をもたらすため、法改正等の背景情

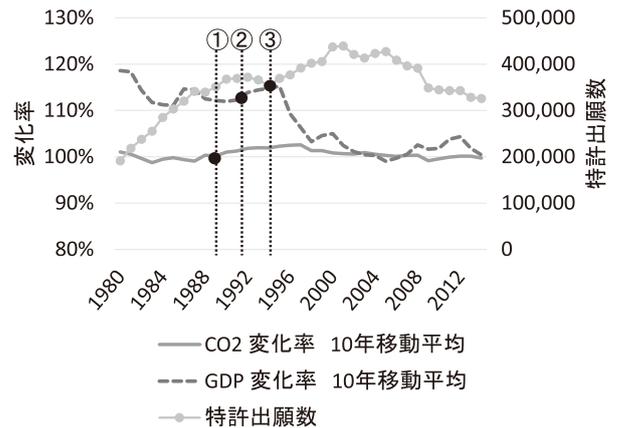


図9 日本のデータ推移

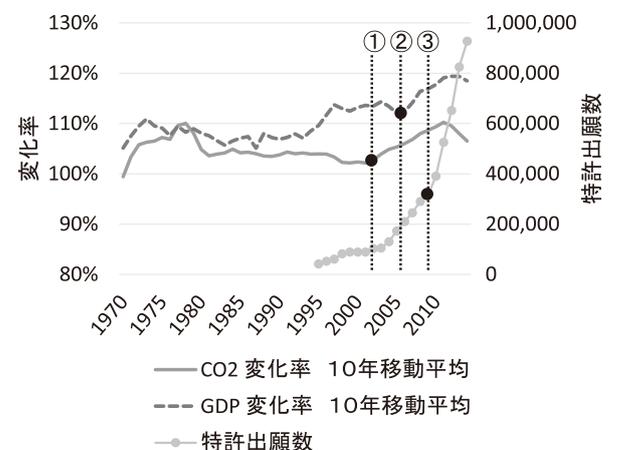


図10 中国のデータ推移

報の収集も重要である。

日本および中国の事例から、本稿での仮説はある程度支持されうると推定し、サンプル事例として、今後の経済発展が期待されるアフリカ諸国から、ナイジェリアとケニアの出願国としての立ち上がりの予測を試みた。

ナイジェリアでは、図11を参照すると、1990年代で減少が続いていたCO2排出量変化率の移動平均値が、2000年頃から100%以上に転じ、2002年頃からGDP変化率の移動平均値も前年比で10%以上増加していることが分かる。よってナイジェリアでは、特許出願国としての素地を備えつつあるのではないかと推測される。またナイジェリアの2014年の1人当たりの名目GDPは中国の2008年頃に相当することから、今後の

知財制度や司法制度の発達に応じて出願国として検討すべき国であると考えられると思われる。

ケニアでは、図12を参照すると、1990年代後半にCO2排出量の増加がみられたものの2000年以降は減少が続いていた。2009年頃からCO2の排出量が増加に転じたものの、GDPは、なだらかな成長を継続し、急激な動きは見られない。また、ケニアの2014年の1人当たりの名目GDPは中国の2003年頃に相当することから、出願国として期待するにはまだ時間を要するのではないと思われる。

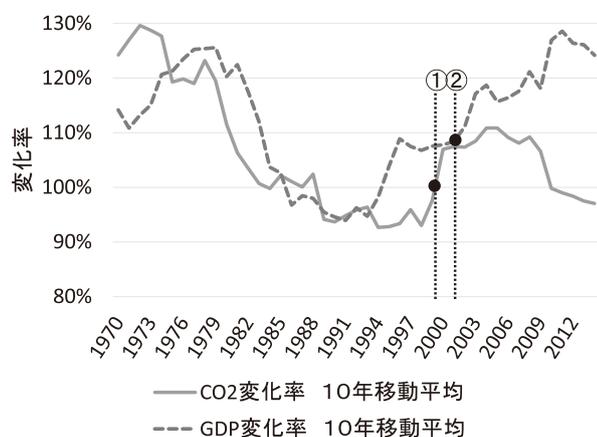


図11 ナイジェリアのデータ推移

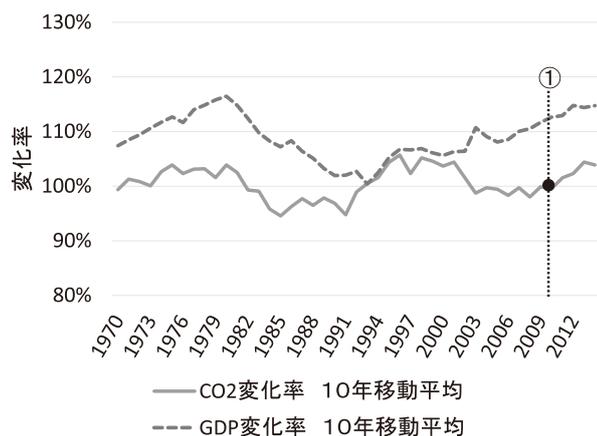


図12 ケニアのデータ推移

今回報告した世界銀行が提供する各国のオープンデータは取得が容易であり、長期間収録されているデータが多いことから、特許データと非特許データとを組み合わせたマクロ分析のハードルを下げることに貢献できると考えている。

本稿での分析では、前述したデータの組み合わせによって、出願国の可能性を検討したが、現在では先進国を中心にCO2削減技術が開発され、CO2とGDPの関係に変化が生じており<sup>46)</sup>、過去のような結果にならない可能性があると思われる。分析としては十分な検証を行うことはできていないと認識しているが、オープンデータを活用した新たな視点での分析によって、競合に先んじた知財投資を行うためのツールとしての可能性を示すことができた。

## 6. おわりに

本研究を通じ、知財分析に活用できる非特許オープンデータが数多く存在していることが分かった。また、本稿では、オープンデータを活用した特許分析の手法の提案等の具体的な活用事例の報告に取り組んだ。

本研究で紹介した、オープンデータ活用の具体的な手法および事例によって、会員企業等におけるオープンデータ活用のハードルが下がる効果を期待している。知的財産部門は、ビジネスチャンスの探索にも貢献できる素養を備えた組織であると考えている。オープンデータを活用した新たな切り口での知財分析によって、我が国企業における知的財産戦略の高度化に貢献できることを切に願う。

本稿は、2017年度情報検索委員会第1小委員会第2ワーキンググループのメンバーである杉本敏彦（日本電気特許技術情報センター、副委員長）、金子浩之（富士通）、生井丈夫（テルモ）、西仁実（日鉄住金総研）、村上耕二郎（リコー）、森田昂樹（トヨタテクニカルディベロプメント）の執筆によるものである。

なお、本稿で紹介するツール、サービスまたは公開するソースコードを利用した際に生じる不具合、不利益などに関して当協会および筆者は一切の責任を負わない。

## 注 記

- 1) 総務省 オープンデータ戦略の推進  
[http://www.soumu.go.jp/menu\\_seisaku/ictseisaku/icriyou/opendata/](http://www.soumu.go.jp/menu_seisaku/ictseisaku/icriyou/opendata/)
- 2) 総務省 オープンデータの5つの段階とデータ形式  
[http://warp.dandl.go.jp/info:ndljp/pid/11109954/www.soumu.go.jp/menu\\_seisaku/ictseisaku/icriyou/opendata/opendata01.html](http://warp.dandl.go.jp/info:ndljp/pid/11109954/www.soumu.go.jp/menu_seisaku/ictseisaku/icriyou/opendata/opendata01.html)
- 3) 電子行政オープンデータ戦略  
[https://www.kantei.go.jp/jp/singi/it2/pdf/120704\\_siryoushi.pdf](https://www.kantei.go.jp/jp/singi/it2/pdf/120704_siryoushi.pdf)
- 4) クリエイティブ・コモンズ・ライセンスとは  
<https://creativecommons.jp/licenses/>
- 5) 情報の科学と技術 65巻 12号, pp.509~514 (2015)
- 6) DATA.GO.JP 利用規約  
<http://www.data.go.jp/terms-of-use/terms-of-use/>
- 7) Community Data License Agreement  
<https://cdla.io/>  
[https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho\\_hyoka\\_kikaku/2018/sangyou/dai5/sankou6.pdf](https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2018/sangyou/dai5/sankou6.pdf)
- 8) Open Data METIは一定の役割を終えたとして2017年3月末をもって閉鎖された。
- 9) 地域経済分析システム (RESAS)  
<https://resas.go.jp/>
- 10) 法人インフォメーション  
<https://hojin-info.go.jp/hojin/TopPage>
- 11) DATA.GO.JP  
<http://www.data.go.jp/>
- 12) EDINET  
<http://disclosure.edinet-fsa.go.jp/>
- 13) DATA.GOV  
<https://www.data.gov/>
- 14) JETROニューヨークだより 2016年12月  
<https://www.ipa.go.jp/files/000056628.pdf>
- 15) EDGAR  
<https://www.sec.gov/edgar.shtml>
- 16) 栗原潔, 知財管理 Vol.67, No.6, pp.841~848 (2017)
- 17) Open data and mobility  
<https://www.uspto.gov/learning-and-resources/open-data-and-mobility>
- 18) Patent Examination Data System  
<https://ped.uspto.gov/peds/>  
(注) 使用ブラウザによっては表示に時間を要する場合があります。
- 19) 「知財管理」付録ページ (会員専用)  
<http://www.jipa.or.jp/kaiin/kikansi/chizai-kanri/furoku.html>
- 20) Drugs@FDA  
<https://www.accessdata.fda.gov/scripts/cder/daf/>
- 21) Orange Book  
<https://www.accessdata.fda.gov/scripts/cder/ob/>
- 22) OpenFDA  
<https://open.fda.gov/>
- 23) EU Open Data Portal  
<http://data.europa.eu/euodp/en/data/>
- 24) EUROPEAN DATA POTENTIAL  
<https://www.europeandataportal.eu/en/>
- 25) DATA.GOV.UK  
<https://data.gov.uk/>
- 26) data.gouv.fr  
<https://www.data.gouv.fr/en/>
- 27) GOVDATA  
<https://www.govdata.de/>
- 28) Bundesanzeiger  
<https://www.bundesanzeiger.de/>
- 29) UNTERNEHMENS REGISTER  
<https://www.unternehmensregister.de/ureg/>
- 30) ドイツの連邦金融監督庁 (BaFin)  
[https://www.bafin.de/EN/Homepage/homepage\\_node.html](https://www.bafin.de/EN/Homepage/homepage_node.html)
- 31) XBRL Japan Inc. XBRLとは  
[https://www.xbrl.or.jp/modules/pico1/index.php?content\\_id=7](https://www.xbrl.or.jp/modules/pico1/index.php?content_id=7)
- 32) 浅野圭子, デジタルレポーティングの潮流~世界的な潮流とわが国企業の挑戦の方向性~, PwC's View 第9号 (July 2017)  
<https://www.pwc.com/jp/ja/japan-knowledge/>

- pwcs-view/pdf/pwcs-view201707-05.pdf
- 33) XBRL Japan  
<https://www.xbrl.or.jp/>
- 34) XBRL Japanからの電子行政オープンデータ構想  
に対する提案 (2013)  
[https://www.xbrl.or.jp/uploads/fckeditor/  
event/09Panel04XBRLJapanSympo14.pdf](https://www.xbrl.or.jp/uploads/fckeditor/event/09Panel04XBRLJapanSympo14.pdf)
- 35) 有報キャッチャーEXCELツール  
<http://tecaweb.net/archives/290>
- 36) EDINET ダウンロード EDINETタクソノミ  
及びコードリスト  
[https://disclosure.edinet-fsa.go.jp/E0\\_1EW/  
BLMainController.jsp?uji.verb=W1E62071Init  
Display&uji.bean=ee.bean.W1E62071.EEW1E62  
071Bean&TID=W1E62071&PID=currentPage&  
SESSIONKEY=1506480722775&kbn=2&ken=53  
&res=53&idx=0&start=1&end=53&spfl=1&spf  
2=1&spf5=1&psr=1&pid=0&row=100&str=&fl  
g=&lgKbn=2&pkbn=0&skbn=1&dskb=&askb=  
&dflg=0&iflg=0&preId=1](https://disclosure.edinet-fsa.go.jp/E0_1EW/<br/>BLMainController.jsp?uji.verb=W1E62071Init<br/>Display&uji.bean=ee.bean.W1E62071.EEW1E62<br/>071Bean&TID=W1E62071&PID=currentPage&<br/>SESSIONKEY=1506480722775&kbn=2&ken=53<br/>&res=53&idx=0&start=1&end=53&spfl=1&spf<br/>2=1&spf5=1&psr=1&pid=0&row=100&str=&fl<br/>g=&lgKbn=2&pkbn=0&skbn=1&dskb=&askb=<br/>&dflg=0&iflg=0&preId=1)
- 37) EDGAR Online  
<http://developer.edgar-online.com/docs>
- 38) ティッカーシンボルの検索ページ  
<https://www.nasdaq.com/symbol/>
- 39) EDGAR Online Register  
[http://developer.edgar-online.com/member/  
register](http://developer.edgar-online.com/member/<br/>register)
- 40) EDGAR Online DataFied API Version 1  
<http://developer.edgar-online.com/docs/v1>
- 41) 全欧州規模で実施される、最大規模の研究・革  
新的開発を促進するためのフレームワーク。  
2014年からの7年間にわたり、約800億€の公的  
資金が提供されている。
- 42) CORDIS-EU research projects under Horizon  
2020 (2014-2020)  
[https://data.europa.eu/euodp/data/dataset/  
cordisH2020projects](https://data.europa.eu/euodp/data/dataset/<br/>cordisH2020projects)
- 43) 経済産業省 予算執行に関する情報開示 補助  
金等の交付決定に関する情報  
[http://www.meti.go.jp/information\\_2/  
publicoffer/index\\_result\\_info.html](http://www.meti.go.jp/information_2/<br/>publicoffer/index_result_info.html)
- 44) NEDO 公募情報 (一覧)  
<http://www.nedo.go.jp/search/?type=koubo>
- 45) World Bank Open Data  
<https://data.worldbank.org/>
- 46) IEA finds CO2 emissions flat for third straight  
year even as global economy grew in 2016  
[https://www.iea.org/newsroom/news/2017/  
march/iea-finds-co2-emissions-flat-for-third-  
straight-year-even-as-global-economy-grew.  
html](https://www.iea.org/newsroom/news/2017/<br/>march/iea-finds-co2-emissions-flat-for-third-<br/>straight-year-even-as-global-economy-grew.<br/>html)
- (URLの最終参照日は、全て2018年6月13日)
- (原稿受領日 2018年6月20日)