

データ主導型イノベーション

Data Matters
Value of Data Innovation

Google™



Google™

C O N T E N T S

Special Report
 ゲーグルの内外で起きている
 “データ主導型
 イノベーション”
 の価値

Interview
 国際大学 GLOCOM 主任研究員
 庄司昌彦氏に聞く
 自由な活用を前提に
 取り組むことで
 オープンデータの
 価値が引き出せる

Symposium
 「Innovation Nippon プロジェクト」シンポジウム・シリーズ
 第1回 オープンデータの
 イノベーション・ポテンシャル
 日本のイノベーションを
 けん引する資源
 オープンデータの
 活用に向けた方策を探る

Google Public Policy Blog
 [ゲーグルの公共政策ブログ]

S p e c i a l R e p o r t

ゲーグルの内外で起きている

“データ主導型 イノベーション”

ゲーグルのエンジニアは検索
 パターンの分析を繰り返
 し、サービスの品質向上に努めて
 います。スペルチェック機能がそ
 の代表例です。ユーザーが検索を
 行うとき、例えば「Beatls」と入
 カミスをして、検索結果には
 「Beatles」と表示されます。エ
 ンジニアリングチームはこの機能
 を実現するため、検索ログから一
 般的なスペルとタイプミスを割り
 出し、ユーザーが求める検索結果
 が確実に表示されるようにしてい
 ます。ただし、これはよく知られて
 いる話です。本特集では、あまり知
 られていない「データ主導型イノ
 ベーション」の数々を紹介します。

1998年、ゲーグルの共同創業者ラリー・
 ペイジとセルゲイ・ブリンは、「世界中の
 情報を整理し、世界中の人々がアクセス
 できて使えるようにすること」をスローガ
 ンに掲げ、独自の検索エンジンの開発を
 開始。今やゲーグルは世界中のユーザー
 を獲得し、誰もが知る存在となりました。
 そのゲーグル社内での多様なデータの利
 用法について尋ねられる場合も多いので
 すが、基本的にはサービスの品質向上、
 システムのセキュリティ維持、そして不正
 行為からユーザーを守ることに活用し
 ています。
 このように、多くのユーザーが利用す
 るサービスの品質を保つ上で「データ」や
 「情報」は不可欠な要素となっており、
 “データ主導型のイノベーション”も急速
 に増えています。ここでは、ゲーグルの内
 外で起こっているその事例を紹介します。

図1 インフルトレンド：インフルエンザの流行が予測される地域や期間を確認
 できるツール。



Value of Data Innovation

の価値

インフルトレンド

2008年11月にゲーグルがリリースした
 「インフルトレンド」は、インフルエンザ
 の流行が予測される地域や期間を確認で
 きるツールです。予測精度は米疾病対策
 センター（CDC）と比べても遜色なく、
 しかも医療機関から集めたデータを分析
 するCDCより2週間も早く情報を入手で
 きます（図1）。

インフルトレンドは、ゲーグル検索のロ
 グを利用して開発されています。数人のエ
 ンジニアが発想したアイデアに基づき、
 インフルエンザの流行に関する過去の
 データをCDCから入手。それを同期間
 の検索ログと照らし合わせたところ、一
 部の検索キーワードがインフルエンザの
 流行を予測する指標となることが分かり
 ました。

図2 ゲーグルマップ：ナビモードを使用すれば、目的地への道順と所要時間
 がリアルタイムで表示される。



ここで押さえておきたい重要なポイント
 があります。20世紀には、「ヒトゲノ
 ムとは何か」「ペニシリンで細菌増殖を抑
 制できる理由は」など、事実の解明に基
 づいて様々な問題が解決されました。た
 だ、21世紀に私たちが直面している問題はタ
 イプがまったく異なります。現在の社会
 が抱える問題は複雑で、それを解決する
 には前世紀に解明されたすべての事実を
 把握し、理解しておかなければなりません。

しかし、大量のデータを最先端のコン
 ピューティング・パワーで統合すれば、
 そうした問題を解決できることが判明し
 ました。検索データでインフルエンザの
 流行に対処するなど、これまで誰が想像
 できたでしょうか。また、このようなデー
 タはほかでも有効活用できます。

ゲーグルマップ

日常的に利用するゲーグルマップを考
 えてみましょう。ゲーグルでは、パソコ
 ンか携帯端末かを問わず、ユーザーの所
 在地に基づいて最適な情報を提供できま
 す。例えば、Android 端末でゲーグルマ
 ップのナビモードを使用すれば、目的地へ

の道順と所要時間がリアルタイムで表示
 されます。あるいは、現在地周辺のレス
 トランやバー、ガソリン・スタンド、
 ATMなどを見つけることも可能です。自
 転車の場合は、自転車専用レーンやサイ
 クリング・ロードの有無、道路の勾配な
 どに基づく自転車向けのルートを確認で
 きます。ゲーグルでは大量のデータ処理
 でこうした情報を提供しています（図2）。

ゲーグルマップの地図にレイヤを重ね
 るケースを考えてみると、複数のデータ
 を統合するメリットについて理解しやす
 くなります。例えば、リアルタイムの交
 通情報を重ねると、事故の発生地点や渋
 滞状況を確認し、ルートを変更すべきか
 どうかを的確に判断できます（図3）。

ゲーグルインサイト

ゲーグルのチーフ・エコノミストは、
 特に計量経済学の観点から検索キーワ
 ードの傾向を詳しく分析しています。これ
 により、景気後退と密接に関連している
 検索キーワードのグループが確認されま
 した。

グレーのバーは、当該語句の検索数が

図3 ゲーグルマップ：交通情報を重ねると、事故の発生地点や渋滞状況を
 確認し、ルートを変更すべきかどうかを的確に判断できる。

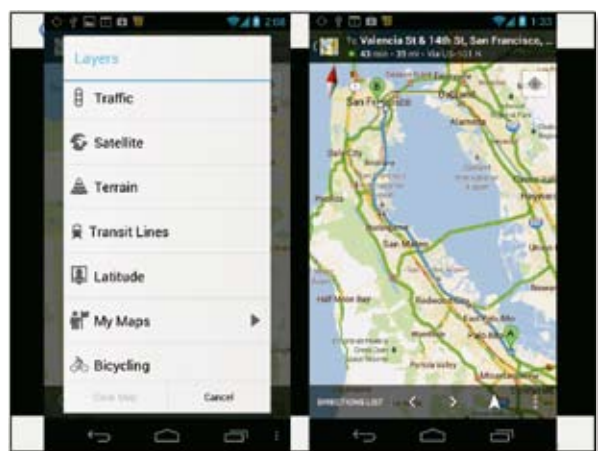
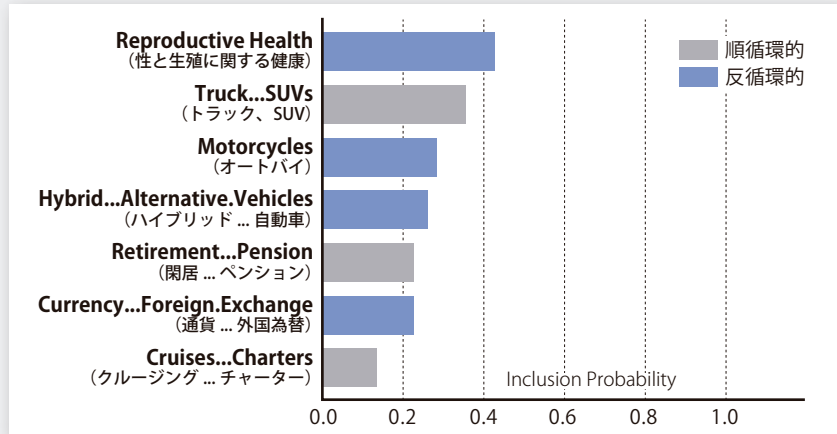


図4 検索キーワードから景気動向を分析



景気に対して「順循環的」に変動する可能性を示しており、好景気と結び付いた検索語句を確認できます。ブルーのバーは「反循環的」に変動する可能性を表しており、景気後退に関連する語句が分かります。「Reproductive Health(性と生殖に関する健康)」は景気の低迷期に検索される可能性が最も高く、支出や貯蓄に関連する「Truck(トラック)」「Pension(年金)」「Cruise(クルージング)」「Charter(チャーター)」や、ガソリン消費の節約、貨幣価値に関連する「Motorcycle(オートバイ)」「Vehicle(自動車)」「Exchange rate(為替レート)」など、一般的に上位にランクしそうな語句よりも高いことが分かります(図4)。

チーフ・エコノミストも最初はこの意外な結果に驚きましたが、景気と出産の関連性に関する米国ピュー研究所(米国のシンクタンク)の調査結果を見てその謎が解けました。2007年以降、米国では出生率が下がり続けていますが、この落ち込みは経済状況の悪化と連動しているようです。そこで、「Reproductive Health」と「Birth Control(避妊)」が検索される可能性をグラフ化し、消費マインドと比較してみました(図5)。

このデータからは、景気の低迷期における独特の行動パターンが浮き彫りになりました。景気が悪いと人は出産を望まなくなり、避妊の方法を検索するように

なるようです。

なお、同分析にはグーグルの無料ツール「グーグルトレンド」を使用しました。グーグルトレンドでは、特定の検索キーワードの動向把握や、特定の時期の人気キーワードを確認することができます。

経済効果

次に経済効果について見てみましょう。検索や広告などのデータ主導型ビジネスは、すでに多大な経済価値を生み出しています(図6)。

【データ】

- ・インタラクティブ広告：米国内で3000億ドル(約30兆円)規模の経済活動を創出。
- ・米国内のインターネット広告の市場規模は、同国GDPの2.1%に相当。
- ・インターネット検索がもたらす総合的

な経済価値7800億ドル(約77兆円)のうち、検索プロバイダーが占める割合は4%のみ。残りはユーザーや各分野の企業が占める。

- ・2010年、グーグルは640億ドル(約6.3兆円)規模の経済活動を創出(米国内の企業、ウェブサイト運営者、非営利団体が対象)。

調査結果はほかにもあります。米テキサス大学の2011年の調査によると、「フォーチュン1000」の平均的な企業でデータの有用性が10%アップすると、年間収益が20億1000万ドル(約1996億円)増加。データの入手可能性が10%アップすると、年間純利益が6567万ドル(約65億円)増加となっています(図7)。

【コンサルティング会社マッキンゼーの2011年の調査結果】

- ・米国の医療分野でデータ利用の効率性と有効性がアップすると、3000億ドル(約30兆円)の経済効果がもたらされる。
- ・EUの行政部門のデータを有効利用できれば、年間2500億ユーロ(約33兆円)の経済効果が生まれる。
- ・米国の小売業界でデータ主導型のイノベーションが進み、データに基づく意思決定が確立されれば、営業利益率の60%向上が見込める。
- ・位置情報を世界規模で有効活用できれば、それだけで消費者余剰の年間6000億ドル(約59兆円)増加が見込める。

図5 グーグルインサイトで「景気と出産の関連性」をグラフ化

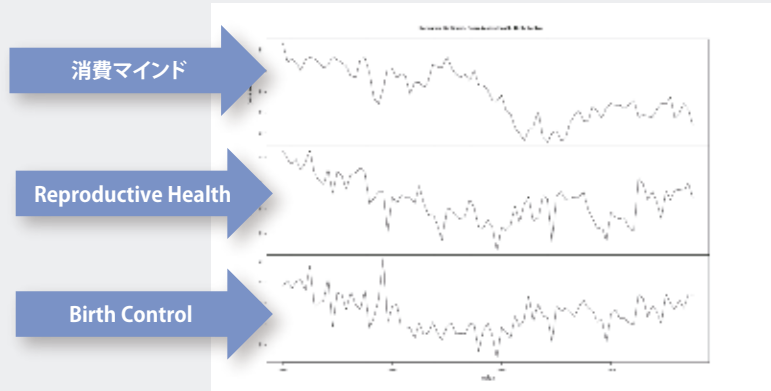


図6 データ主導型ビジネスによる経済効果

インタラクティブ広告	3000億ドル(約30兆円)規模の経済活動を創出
インターネット広告の市場規模	GDPの2.1%に相当
検索	総合的な経済価値は7800億ドル(約77兆円) 検索プロバイダーの割合は4%のみ
グーグル(2010年)	640億ドル(約6.3兆円)規模の経済活動を創出

オープンデータ

最近自治体や国家レベルでオープンデータの活用が進んでいます。データソースの一般公開でイノベーションを加速し、市民や国民の利益につながる新たな価値を生み出すのが狙いです。例えば、公共交通機関のデータに誰でもアクセスできるようにすれば、アプリ開発者による乗客向けの便利なアプリ作成が可能となり、公共交通機関の利用を促進できます。公共交通機関の利用者数が増えれば、運営主体である国や自治体の歳入増も見込めるでしょう。

米ハーバード・ケネディスクールの透明性政策プロジェクトでは、公共交通機関のデータのフォーマット、アクセス可能性と対応アプリ数の関係を調査しました。調査した5つの交通機関のうち、ポートランドのTriMetとボストンのMBTAがこのアプローチを積極的に実践しており、利用者数に対するアプリ数の割合が最も高くなっています。逆に、オープンデータの導入が最も遅れているワシントンのWMATAでは、2011年現在でわずか11個のアプリしか提供されていませんでした(図8)。

開かれた政府の推進

各国の政府も公共データの公開を推進。一般市民が自由にデータにアクセスし、

図7 その他の「データの可能性を引き出す」経済効果

データの有用性が10%アップ＝	年間20億1000万ドル(約1996億円)の増収
データの入手可能性が10%アップ＝	年間純利益が6567万ドル(約65億円)増加
データの可能性	・米国の医療分野で 3000億ドル(約30兆円) 規模の経済効果を創出 ・EUの行政部門で 2500億ユーロ(約33兆円) 規模の経済効果を創出 ・米国の小売業の営業利益率が 60% 向上
世界規模の位置情報:	消費者余剰の年間 6000億ドル(約59兆円) 増加が見込まれる

出典:「Measuring the Business Impacts of Effective Data」テキサス大学(2011年)、「Big data: The next frontier for innovation, competition, and productivity」マッキンゼー社(2011年)

画期的な方法で活用できるようにしています。オープンデータを通じて、かつてないほど政府の透明化が進んでいます。では、国家レベルの取り組みをいくつか紹介しましょう。

ロンドンでは警察の犯罪マップが公共データとして市民に公開されており、データのマッシュアップ(統合)の好例と評価されています。犯罪マップは、犯罪に関するデータと地図情報を組み合わせたものです。

フランスでは、政府機関のEtalabが公共データのオープン化を推進しています。その成果の代表例として挙げられるのが、フランスのテクノロジー・ソリューション・プロバイダーAntidot社が公共

データを組み合わせて構築した、フランスの史跡情報サイト「Monuments Historiques」です。このページはフランス語で構成されていますが、グーグル翻訳を利用して読むことができました(図9)。

米国では、公共データのポータルサイト「Data.gov」で、膨大な量のデータがアプリ開発者向けに公開されています。

多様なデータを組み合わせる

多様なデータを組み合わせると、ユーザーが利用しなくなる魅力的なコンテンツとなり、提示した内容の理解も進みます。例えばグーグルのサービスの1つ「Politics & Elections(政治と選挙)」

図8 公共交通機関のデータとアクセス数、対応アプリ数の関係

出典: 透明性政策プロジェクト(2012)

都市	ポートランド	ボストン	シカゴ	ワシントン	ニューヨーク
運営組織					
平日の平均利用者数(2011年)	325,400	1,292,000	1,717,200	1,446,400	10,287,600
アプリ数(2011年)	44	47	22	11	58
比率(アプリ数/利用者数)	1/7,000	1/27,000	1/78,000	1/131,000	1/177,000
開発者との関係	強	強	中	弱	弱
マーケティングの積極性	低	低	中	低	高
データの公開時期	2006/2007	2009	2009/2010	2010	2011/2012

図9 フランスの史跡情報サイト「Monuments Historiques」政府機関のEtalabが公共データのオープン化を推進している。



は、次の大統領選に注目している米国のユーザー向けのサイトです。検索のトレンド、ニュースの内容やYouTube動画を基に、Googleが候補者ごとに作成した注目ワードのチャートを確認することができます(図10)。

位置情報とYouTubeのアップロード情報を組み合わせることにより、有権者から寄せられた動画を本人の居住地に関連づけて地図上に表示することも可能です。情報の発信元が分かれば、選挙の動向を把握しやすいでしょう(図11)。

「Politics & Elections」は、米国専用のサイトではありません。例えば、エジプト人がエジプト版のページにアクセスして現住所を入力すれば、最高選挙管

理委員会から投票方法に関する情報を引き出すことができます。

Google翻訳

データは将来的に、政策の諸問題の解決に役立つような革新的なツールも生み出します。Google翻訳もその種の好例といえるでしょう。元のウェブページと手作業で翻訳されたページの類似パターンを探り、利用する仕組みです。この処理を多数のサンプルページで実施することにより、検索サービスを多様な言語に対応させる機械翻訳エンジンを開発することができました(図12)。

ウェブページの翻訳に限らず、最近では母国語が違うユーザー同士のリアルタ

イム・チャットにも対応しています。Googleのブラウザ「Chrome」は既存のデータに基づき、ユーザーが開いたページが非母国語であると確認すると、それを通知し翻訳を提案する機能があります。Google翻訳のページに直接アクセスし、翻訳するテキストを手動で貼り付けることも可能です。様々な翻訳データの蓄積が進むほど、細かなニュアンスを読み取る精度が上がり、翻訳の質も向上します。これは機械翻訳の機能向上にデータが貢献している一例です。

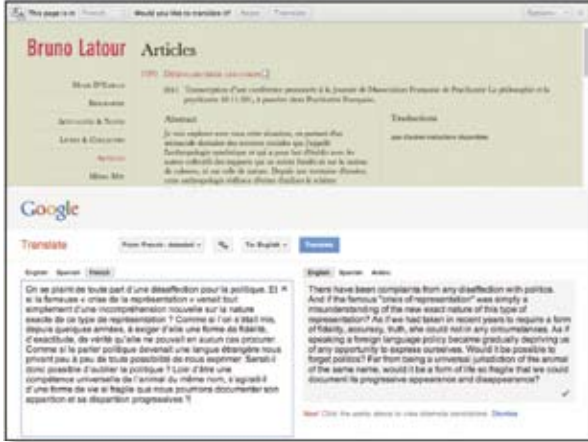
ウイルスなどの対抗手段

以上のように、データは前向きなイノベーションに利用できるのはもちろん、スパムやフィッシング、ウイルスなどの対抗手段として利用することもできます。Googleのセーフ・ブラウジング技術はその一例です(図13)。

Googleのセーフ・ブラウジング技術では数十億のURLを調査し、次のような危険なサイトを検出します。

- ・マルウェア・サイト：キーロガーなどのクライムウェアを、ユーザーのパソコンに強制インストールする
- ・フィッシング・サイト(銀行を装った偽サイトなど)：本物のサイトを装い、訪問者のユーザー名やパスワードを盗み取る

図12 Google翻訳：検索サービスを多様な言語に対応させた機械翻訳エンジン。



検出された不正サイトは、Googleの検索結果で危険なサイトであることが明示されます。検索結果に警告を表示したのは、主要な検索サイトではGoogleが初めてでした。

Googleのサービスのユーザーだけでなく、すべてのインターネット・ユーザーを保護することが目的ですので、「悪意のあるサイト」のデータはSafe Browsing APIを使用すれば、誰でも無料で利用できます。

例えば、AppleのSafariやFirefoxでもこのデータが利用されています。検索ログのデータを利用すると、ウェブスパムが紛れていない適切な検索結果を返すことができます。IPアドレスやクッキー情報などのログデータからは、検索の品質を測定する様々な指標を作成することが可能です(インデックスのサイズやカバレッジ、検索の「鮮度」、スパムなど)。

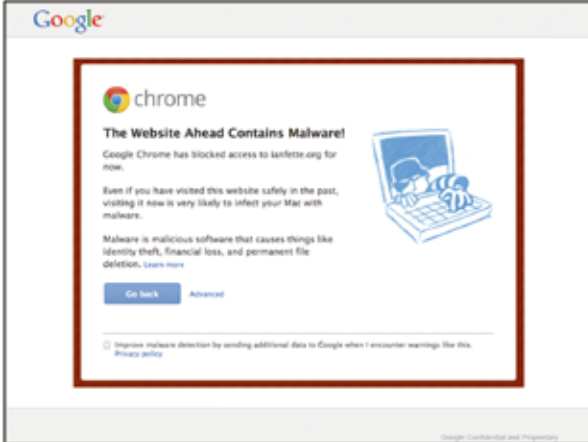
新しい指標の作成では、ログデータを

見直し、過去の検索やその結果を参考にすることが重要です。Googleでは検索ログを辿って時間をさかのぼり、数カ月前の検索でどのような結果が出ているか、しっかりと確認しています。新種のスパムを正確に検出する指標を作成する際には、スパムの成功事例のトラッキングを開始するだけでなく、ログデータを見直し、過去数カ月または過去数年間に同タイプのスパムにどのように対処してきたかも確認するようにしています(図14)。

こうした手法を正当なユーザーの検索だけに適用し、ボットによる不正な検索などを対象から除外する上では、IPアドレスとクッキー情報が重要になります。例えば、同じ検索を何度も繰り返すボットがある場合は、ユーザーにスパムが表示される頻度を測定する前にIPアドレスやクッキー情報を利用して、そうした検索を除外しておく必要があります。このように、関連性が高く適切な検索結果をユーザーに表示するには、ログデータの活用が欠かせません。

収集したデータは、ユーザーの保護にも役立っています。例えば、広告主の余計な出費につながり、広告ネットワークの信頼性を損ねる不正なクリックは、コンピュータのモデリング技術を利用して検出しています。IPアドレスは、1人のユーザーを表す場合もあれば、1000台のパソコンで構成されたボットネットを表す

図13 Googleのセーフ・ブラウジング技術は数十億のURLを調査し、危険なサイトを検出する。



場合もあります。

データ主導型時代の課題

以上の例が示すように、現在の様々な問題を解決する上でデータの活用や分析の重要性が高まっていることがお分かりいただけると思います。今後、私たちが直面するのは、多様な種類の膨大なデータを把握していなければ解決できない問題です。そうしたデータが、ユーザーの行動をトリガーとするものが含まれることが多いことも論をまたないでしょう。

しかし、新しい問題といっても、地域開発や犯罪撲滅、公害対策などの従来の取り組みと大きな差異はありません。プライバシーや著作権、表現の自由に関しては、関連する政策担当者が振興と課題克服の両面から検討する必要があります。何よりも、データ・イノベーションを促進し、誰もがそのメリットを享受できる環境を整えていくことが鍵となるのです(図15)。

図10 Googleの「Politics & Elections(政治と選挙)」候補者ごとに作成した注目ワードのチャートを確認できる。



図11 「Politics & Elections(政治と選挙)」は、有権者から寄せられた動画を本人の居住地に関連づけて地図上に表示することもできる。

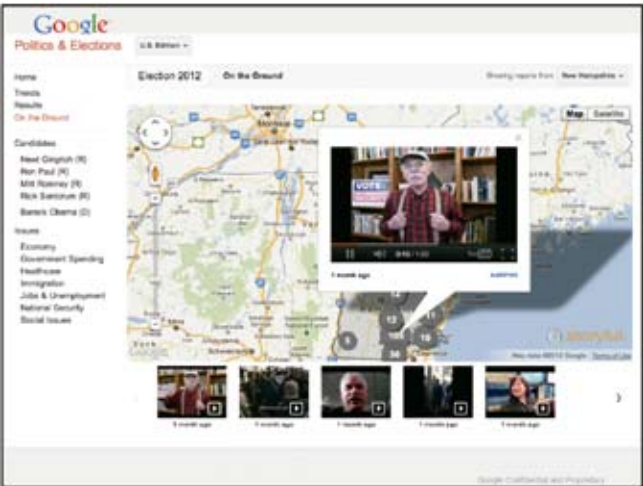


図14 スパム/詐欺行為への対策

事前対策
フィルタ 自動アルゴリズムですべてのクリックを分析し、不正なクリックをリアルタイムで排除する
オフライン分析 自動アルゴリズムと手動解析
事後対策
調査 クリックの品質保持チームが手作業で実施
例：不正なクリックの検出

図15 データ主導型時代の課題

- ◎ インフラのサポート
- ◎ プライバシーとイノベーションのバランス
- ◎ 経済的利益の追求
- ◎ データ・アクセスを阻害する障壁の打破
- ◎ 所有権に関する問題の解決

国際大学 GLOCOM 主任研究員 庄司昌彦氏に聞く

自由な活用を前提に取り組むことで オープンデータの価値が引き出せる



庄司昌彦

政府や公的機関が保有する様々なデータを再利用可能な形で公開する「オープンデータ」。社会的、経済的に大きな価値を生むと期待されており、国内でも近年急速に関心が高まっています。しかし、効果的な活用のためには、政府、自治体、企業、市民の間で理解を促進しつつ、複数の課題を乗り越える必要があります。オープンデータ活用に積極的な海外事情にも詳しい国際大学GLOCOMの庄司昌彦氏に、オープンデータの可能性や活用に向けた課題などを伺いました。

——政府の新たなIT戦略「世界最先端IT国家創造宣言」でオープンデータの推進が盛り込まれるなど、活用に向けた取り組みが進んでいます。

庄司 私が2010年にIT戦略本部のタスクフォース委員になった頃は、海外での先進的な取り組みを話題にしていた程度でした。このたび、政府がオープンデータ推進の具体的な戦略を示すなど、急速に関心度が高まっています。ただ、まだ途に就いたばかりというのが現状ではないでしょうか。

そもそもオープンデータの定義に誤解があると思います。オープンデータとは、単にデータを公開することではありません。ここで言う“オープン”とは利用条件が広く開かれていることを指しています。ネット

上で公開していても、無断で利用できなかったり、利用が制限されたりしてはオープンデータとはいえません。また、紙やPDFファイルなど、利用しにくい形式で公開されている状態もあり。そう考えると、まだまだ使えるデータは少ないといえるでしょう。政府が先導してオープンデータを推進することには、大きな意義があります。

——オープンデータによってどのような効果が期待されているのでしょうか。

庄司 政府や公共機関が保有するデータを活用することで、期待される効果は3つあります。その1つ目は、政治の透明

日本にはオープンデータではないが公的情報を利用した世界最先端レベルのサービスも存在する

◎ 道路交通

主体：地方自治体・国土省・警察・高速道路会社ほか
→ 道路交通情報センター → 民間企業/メディア
情報：道路渋滞、工事、事故、通行状況 → 到着時間予測、事故対策

◎ 鉄道・バス

主体：鉄道バス会社・子会社 → 民間企業
情報：時刻表、運行情報 → ナビサービス

◎ 気象

主体：気象庁 → 気象業務支援センター → 民間企業/メディア
情報：気象予報、警報、過去データ、緊急地震速報
→ 高度な気象情報サービス、防災・災害対応



性を高めること。国の予算を誰がどの地域でいくら使っているのかなどを、データに基づいて市民が考え、議論できるという効果があります。

2つ目は、官民協働の実現です。行政サービスに民間企業や市民の力を借りるためのデータとして活用するのです。英国の「FixMyStreet」というサイトでは地域内の道路や公共物などが破損している場合、市民が投稿して解決につなげるサイトとして活用されています。

そして、3つ目が経済効果です。オープンデータによって新たなビジネスが生まれることや、先に述べたような官民協働の取り組みが結果として社会全体でのコスト削減につながることで。日本政府が最も期待しているのはこの部分ではないでしょうか。

——活用に向けた課題は何でしょう。データ分析力などの技術的な面でしょうか。

庄司 必ずしもそうではありません。技術的なことより、データ活用に向けた制度づくりや、民間企業、市民一人ひとりのマインドの持ち方だと思っています。民間の意欲をかき立て、創意工夫を引き出すような、制度面の整備や活用に向け

た運動の広がりが重要です。

例えば、自治体が地域住民に向けて子育てのためのガイドブックを作るとします。そのためには、幼稚園や保育園の情報だけでなく、病院や市役所、相談窓口、各種支援制度など様々な情報が収集されます。しかし、これだけでは、収集された情報はガイドブックを作るために使われるにとどまります。

一方、こうした情報をオープンデータとして公開するとどうでしょう。民間の人たちが情報を活用して、外国人向けに新たなメディアを作ったり、便利なアプリケーションが開発されたりと、さらに良い仕掛けが生まれる可能性があります。今はまだそのような発想が広く一般に浸透していないのです。

——政府や自治体などは保有する情報を公開することに抵抗感を持つことも

ありそうですね。

庄司 確かにその通りです。これまでの情報公開は、要求された情報を開示するというスタンスでした。しかし、これからは公開することを前提に情報の管理のしかた、収集のしかたを考えるべきです。それがG8のオープンデータ憲章でもうたわれている5つの原則の1つである「Open Data by Default」なのです。

他方、民間や一般市民など、データを活用する側は、データを組み合わせたり分析したり、新しい形で表現したりするなど工夫して、自分たちでビジネスや課題解決に役立てていくことが求められます。皆で知恵を出し合って、より良い社会を作ることが、地域、自治体の強みになります。そのためにも、政府や公共機関は量的にも質的にも十分なデータを提供すべきです。

——オープンデータの位置づけは「皆が知恵を出し合う土台」ともいえそうですね。

庄司 まさにそうですね。だからこそ、誰でも自由に使える形で提供しなければなりません。使い方を事前に想定して、使い方自体を限定するような提供方法では意味がない。政府のデータを公開することで、広く活用される知識の土台にしていくことが重要なのです。

周囲をどれだけ巻き込めるか
進化のプロセスがポイントになる

オープンデータ活用に向けた情報提供側の課題

利用しにくいデータ

- ◎ データとしての再利用が困難な複雑さ
- ◎ スキャン画像
- ◎ 語彙、形式の不統一

本当に「All Rights Reserved」なのか

- ◎ 必要性のない営利利用禁止
- ◎ 必要性のない学術利用限定

無駄な業務の見直し

- ◎ 紙でしか保存しない
- ◎ デジタルで作成したものを紙提出(その後手入力)している

法的根拠の整理

- ◎ 公開するよう定められている・ただし「書面」と書いてある
- ◎ 公開が禁じられている
- ◎ 公開の可否が定められていない
- ◎ 「目的外利用」の可否

所在・管理の状態

- ◎ 把握していない／探せない
- ◎ 死蔵されている

マインド

- ◎ 誤解・批判を避けたい

米国のオープンデータ活用事例



⑤ 米国の「The Climate Corporation」は、国立気象サービス(National Weather Service)から気象データ、農務省から収穫量と土壌データを得て、新しい農家向け保険商品Total Weather Insuranceを開発しました。

⑥ 米国の「Open311」は、緊急性の低いSOSを市民が行政に発信するサイト。

——海外の事例はどうでしょう。日本との違いを教えてください。

庄司 1つは、大規模なオープンデータの活用です。米国の保険会社である「The Climate Corporation」の農家向け保険では、国内250万カ所の気象データ、過去60年分の農作物収穫量データ、そして1500億カ所の土壌データの組み合わせにより、地域や作物ごとの被害発生率を予測して保険商品を販売しています。

こうした大規模で複雑なオープンデータの活用は海外でもあまり例がありません。しかしインパクトはこれだけではありません。オープンデータは、一部の専門家が活用するだけでなく、デスクトップで扱えるようなスモールデータから誰もがアイデア次第でたくさんのイノベーションを起こしていくことができるのです。例えば、米国の「Open311」は、緊急性の低いSOSを市民が行政に発信するサイトです。公共の施策に市民の参加が前提となっているのです。もちろん、日本

でも地域SNSのような発想は以前からあったのですが、欧米の方が活用や運用のプロセスがうまくいっています。日本の場合は、もっと長期的な視点から周りを巻き込むプロセスや官民協働のあり方を考える必要もあるでしょう。

——透明性という面でも欧米の方が進んでいるのでしょうか。

庄司 欧米に限りませんが、諸外国のオープンデータの議論では、税金の使われ方についての透明性や説明責任が日本以上に重視されています。政府の契約先や支出額なども細かなデータを積極的に公開しています。

日本でも「税金はどこへ行った?」というサイトを開発・普及する活動が立ち上がっており、オープン・ナレッジ・ファウンデーション・ジャパン(OKFJ)でも支援しています。これは自分の納めた税金の行方を項目ごとに分かりやすく伝えるために、工夫して単純化しています。横浜市をスタートに、すでに全国でも50

地域の中での対話を通して納得できる使われ方を追求したい

以上の地域で、自治体が公開しているデータを利用してサイトが作られています。民間や市民の政治意識の高まりとともに普及していった事例です。

OKFJは、英国発のグループであるオープン・ナレッジ・ファウンデーション(OKF)の日本グループです。OKFは2004年、ルーファス・ポロックという研究者でありアクティビストでもある人などが立ち上げ、現在は世界各国に拠点があります。主な活動内容は、公共データをはじめとした様々な情報を共有し、活用するためのツールやコミュニティづくりです。日本のグループは有志が集まって行動していたのですが、2012年に正式に発足しました。

——今後、我々はオープンデータとどのように向き合っていけばよいのでしょうか。

庄司 オバマ大統領は政権獲得後すぐにオープンガバメントの考え方を提示しました。そこで示された「透明」「参加」そして「協働」という3つの原則は、段階的に達成されていくものだと考えています。透明性が高まることで、意見を述べたり手を動かしたり参加する人々が増えてきます。そうした個人の参加が進んでいけば官民の組織の壁を越えた本格的な協働も実現していくと思います。

日本版オープンデータ活用事例



オープン・ナレッジ・ファウンデーション・ジャパン(OKFJ)を中心に制作されたサイト「税金はどこへ行った?」。原型となった「Where Does My Money Go?」は英国に本部のあるオープン・ナレッジ・ファウンデーションが開発。

公共のデータをオープンにすることによりリスクはあります。いくつかのデータを掛け合わせることで、個人が特定できてしまう可能性もあるかもしれません。しかし、一方ではハザードマップのように、以前は公開することがタブー視されていたものが、今では多くの人から求められるように状況は変化しています。

——日本でも東日本大震災を機にオープンデータの捉え方が変わりました。

庄司 そうですね。例えば、放射線量のデータが公開されたことが挙げられます。多くの人が、生のデータが見たい、そして自分で判断したいと考えるきっかけになったと思います。数値に基づいて安全性の議論ができるようになったというのは大きな前進です。

電力の需給についても国民の見方が大きく変わりました。電力会社は電力に関するデータをグラフではなく、Excelで利用できるCSV形式で提供するようになりました。電力データのような、公共性が高く、注目されているデータが、使える形で提供されるようになったことは大きなインパクトがありました。オープンデータの活用事例としては、この生データを活用したアプリケーションが多数生まれました。

このような具体的目に見える事例は、今後に大きな意味をもたらします。オープンデータがどう活用できるのか、どんなメリットがあるのかを皆が実感できたは

ずです。この体験は、地域や社会、さらにはビジネスの世界で、次のイノベーションにつながるものだと思いますね。——活用が広まることで、社会の常識自体も変わると。
庄司 オープンデータの活用は、長期的な視点で取り組むべき活動です。意識が変わるまでには課題は多く道のりは遠いかもしれませんが、資源のない日本にとって、情報が重要な資源であることに間違いはないといえます。

データを公開することで市民の地域に対する関心が高まり、新しい活用方法が生まれてきます。そこには思いがけない使われ方や、地域社会への新しい発見があるかもしれません。

例えば、OKFの大きなイベントとして、「インターナショナル・オープンデータ・デイ」があります。これは世界の各都市で、一斉にオープンデータに関するイベントを開催します。時間を区切って、グループごとに集中的にソフトウェア制作を行う“ハッカソン”を実施します。

米国や英国ではトップダウンとボトムアップが咬み合ってオープンデータの活用が広がっています。日本も国の政策は動き出しましたから、地域の中で対話しながら、最も広く納得できる使い方を模索していきたいと考えています。

東日本大震災をきっかけに開発されたアプリケーションの例



東京電力 電力使用状況
Chrome extension
Yoshiki Takeoka



Sinsai.info



風@福島原発
石野正剛



NERV専用臨時電力
供給監視装置集団
@Molokheiya

東日本大震災の際に、政府のデータ公開と国民による活用が活発化した。経済産業省・LASDECがPDFではなくHTMLやCSVでの情報提供を要請。政府や東京電力などのデータを活用した自発的アプリ開発が国民の情報入手に貢献し、救援活動や復興に役立つアプリやサービスの開発、イベント開催が活性化した。

「Innovation Nipponプロジェクト」シンポジウム・シリーズ 第1回 オープンデータのイノベーション・ポテンシャル Review

日本のイノベーションをけん引する資源 オープンデータの活用に向けた方策を探る



グーグル株式会社と国際大学GLOCOMは、日本におけるICTを活用したプロ・イノベーション政策の推進を後押しすることを目指し、共同で「Innovation Nipponプロジェクト」を立ち上げました。同プロジェクトでは、有識者による政策課題への提言を取りまとめる研究活動、日本各地のイノベーション先進事例の紹介、政策担当者やオピニオンリーダーを招いて政策課題について議論するシンポジウムなどの開催を推進。9月12日に第1回「Innovation Nipponプロジェクト」シンポジウムが開催され、オープンデータの活用によるイノベーションの可能性について、様々な立場からの意見を踏まえた活発な議論が行われました。

2012年7月に電子行政オープンデータ戦略で決められたデータポータル試行版のローンチが近づく中、オープンデータへの取り組みは活発化しています。第1回「Innovation Nipponプロジェクト」シンポジウムでは、こうした動きを受けて、日本のオープンデータへの取り組みの現状と今後の課題について、有識者による講演が行われました。ここでは当日のプログラムに沿って概略を紹介します。



渡辺智暁氏
Tomoaki Watanabe
モデレーター、
国際大学GLOCOM
主幹研究員

同シンポジウムのモデレーターを務める国際大学GLOCOM主幹研究員でオープン・ナレッジ・ファウンデーション・ジャパン共同創設者、コモンズフィア常務理事の渡辺智暁氏は冒頭で「オープンデータは世界中で活動が広がっています。日本では何ができるのか議論を深めたい」とシンポジウムの趣旨を説明しました。

本格化するオープンデータへの取り組み

シンポジウムは、「オープンデータのフロンティア：イノベーションの課題は何か？」と題した、国際大学GLOCOM主幹研究員の庄司昌彦氏の基調講演で始まりました。

庄司氏は最初に「地域のリソースを有効活用して、課題を解決できる社会を作るには、オープンデータの活用が重要」と指摘。さらに、ここで言うオープンデータとは『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』であり、「単に公開されているだけでなく、広く利用条件が開かれているもの」であることを確認しました。

続いて庄司氏は、英国や米国で先行しているオープンデータの活用による透明性ある政府「オープンガバメント」について触れ、オープンデータの活用が国際的にも広がっていることを解説。2013年のG8サミットで作成された「オープンデータ憲章」を紹介しました。

「オープンデータの活用にあたって日本国内での転機となったのは、東日本大震災。例えば、放射線の安全性について元のデータを見た上での議論が広がり、電力の需給についてはデータがオープンになりました。さらにはデータを利用したアプリケーションが次々と生まれたのです」と説明。



庄司昌彦氏
Masahiko Shoji
国際大学GLOCOM
主幹研究員

また、日本のオープンデータの状況について「道路交通や鉄道・バス、そして気象といった分野では、世界最先端レベルの高度な展開が行われている。こうした基盤を生かすことが重要」と日本の強みを分析。このように最先端の分野から日本のオープンデータの活用の道が広がっていくことを強調しました。

さらに庄司氏は、自民党の安倍政権が早々に打ち出した「世界最先端IT国家創造宣言」の中で、データ活用の重要性が強調されていることを紹介し、電子行政のオープンデータ推進のためのロードマップを提示。データポータルの試行がまもなく始まることを背景に、「オープンデータへの取り組みが本格化するタイミングにきている」と改めて説明しました。

日本のイノベーションの課題として、「日本のオープンデータ活用にオリジナリティは本当にあるのか」という点を指摘。「世界に広がるためには、地域で生まれた動きを、スピード感を持って展開することが不可欠」と語りました。

推進には「Universal Language」が重要

基調講演に続いて、シンポジウムの登壇者からショートプレゼンテーションが行われました。

自民党所属の衆議院議員である橋本岳

氏が強調したのは、共通の価値観を持つことの重要性です。橋本氏は、欧州原子核研究機構を訪問した際に同機構の研究者たちが口にしていた「Science is Universal Language!」という言葉を紹介。「研究者は世界の各国から集まってきます。彼らの文化や思想などはそれぞれ異なりますが、『科学の追究』という同じ目的があるからこそ、コミュニケーションを深めることができるのです」とその理由を述べました。

また橋本氏は、政府が「世界最先端IT国家創造宣言」を提唱し、6月14日に閣議決定したことに触れ、「『日本再興戦略』と同じ日に決定したことには大きな意味がある。成長戦略の柱としてITを成長エンジンとして活用していこうということなのです」とその意義を説明。そのためには情報流通基盤の整備が必要であり、オープンデータの利活用についてのルールづくりも進められていることを紹介しました。



橋本岳氏
Gaku Hashimoto
衆議院議員
自民党

橋本氏は「今回の政府の宣言は官邸主導で決められ、担当大臣を設置するほど“本気度”が高く、省庁横断の体制が強化されている特徴があります。動き出したばかりですが、オープンデータならではの『Universal Language』が問われているのではないだろうか」と今後の展開を熱く語りました。

利活用のフェーズから 出口戦略を考える

千葉市でCIO補佐監を務める三木浩平氏が解説したのは、オープンデータの出口戦略である利活用のフェーズの重要性です。気象、交通などの各分野のデータの利活用にあたり、「オープンかクローズか」

といった開示の軸と、「データ量がビッグかスモールか」といった量の軸の2軸からなる4象限で整理し、それぞれについて利活用のアプローチを紹介しました。



三木浩平氏
Kohei Miki
千葉市CIO補佐監

例えば、スモールでオープンな領域での出口になるのは市民協働型の事業とした上で、「施設、公共サービス、イベント、観光などの情報が、街づくりや街おこし、ボランティアなどに使われることになる」と説明。一方、ビッグでオープンな領域では、地質や人口、気象などの統計的なデータがあり、企業がこうしたデータを利用して新しいビジネスやサービスなどの情報応用ビジネスを創り出していくことの重要性を述べました。

さらに、ビッグでクローズな領域は、医療や福祉、税、教育といった個人情報を中心であり、「公共団体が主体となってデータを管理しつつ、社会的な課題を解決するための事業を展開することになるでしょう」と解説。

実際に千葉市が取り組んでいる「ちば市民協働レポート実証実験(ちばレボ)」は、市民がスマートフォンで課題をレポートし、役所が処理を切り分ける市民協働型の事業。イノベーションの推進には、行政が情報を開示し、市民が主体的に解決にあたるということが重要なポイントになることが示されました。

農家向けの クラウドサービスを展開

民間企業の立場からオープンデータの活用について語ったのは、ウォーターセル株式会社の代表取締役の長井啓友氏です。同社では、クラウドを活用した農業支援システム「アグリノート」を展開しています。これは、航空写真マップと圃場

日時 2013年9月12日(木)
場所 国際大学グローバル・コミュニケーション・センター
主催 Innovation Nipponプロジェクト
後援 オープン・ナレッジ・ファウンデーション・ジャパン
登壇者
橋本岳(衆議院議員 自民党)、三木浩平(千葉市CIO補佐監)、長井啓友(ウォーターセル株式会社代表取締役)、藤井宏一郎(グーグル株式会社執行役員、公共政策部長)、庄司昌彦(国際大学GLOCOM主幹研究員)、渡辺智暁(モデレーター、国際大学GLOCOM主幹研究員)

をリンクさせ、利用者はパソコンやスマートフォンから作業内容を簡単に入力できるシステムです。作物の品質維持や管理が手軽に行える上、利用者同士で情報交換ができるほか、後継者育成にも役立つことが評価されています。特徴として、政府が提供している農業や肥料に関するデータベースも検索できるような仕組みが構築されています。



長井啓友氏
Hirotomo Nagai
ウォーターセル株式会社
代表取締役

「農業や肥料の使用に関しては厳しい制限があります。そこで、データとして取り込んで提供することで、確認しながら記録できるようにしているのです」(長井氏)。農業従事者の負担を減らしながら、ノウハウを蓄積していく目的で開発されたアグリノートの価値を高めるために、オープンデータを積極的に利用していることを強調しました。さらに長井氏は、今後必要となるデータとして「農作物の生育や病気のデータ」を挙げました。「農業で新たな農法にチャレンジする機会は1年に1回しか試すチャンスがないだけに、改革にも慎重になりがち。例えば、大学や研究室にある農作物のデータが開示されて利用できるようになれば、適切な農法にもチャレンジしやすくなるはず」。また、「各農家がいつ、どのような作業をしたかというデータが大量に蓄積されるようになれば、それをオープンデータとして提供していくことも考えたい」とアグリノート自体のオープンデータ化についても前向きに語りました。

「Innovation Nipponプロジェクト」とは

「Innovation Nipponプロジェクト」は、グーグル株式会社と国際大学GLOCOMIによる共同プロジェクトです。日本におけるICT活用のイノベーション推進を目的に、「プロ・イノベーション」の立場から、既存の制度、ガイドラインなどの法制度、産業振興の方向性、規制緩和のあり方、ビジネス慣行などに対して、世界の例を参照しつつ議論・研究を行い、適宜関係機関の政策企画・判断に役立つアウトプットを取りまとめます。
設立 © 2013年7月1日 発起人 © 国際大学グローバル・コミュニケーション・センター(GLOCOM)、グーグル株式会社
URL <http://www.innovation-nippon.jp/>

ガバメントデータの戦略的な活用を

グーグル株式会社の執行役員公共政策部長である藤井宏一郎のテーマは「ガバメントデータのマネタイズ」です。藤井は「ガバメントデータをマネタイズすることは挑戦的な課題といえます」と難しさを強調しながら、アプローチとして、ガバメントデータをビジネスに利用するG(政府) to B(企業)と、利用者に利便性を提供するG(政府) to C(消費者)の2つに分けて解説しました。



藤井宏一郎
Koichiro Fujii
グーグル株式会社
執行役員、
公共政策部長

G to Bには、製品やサービスの開発にデータを利用するプロダクト分野、その日の天候で仕入れを判断するといったオペレーション分野、売り上げ予測や業界動向をビジネスに役立てるマーケティング分野の3つの分野があるとした上で「収益増加または費用低減によりマネタイズ可能です」と藤井は指摘しました。

また、G to Cとしては、「ユーザーの判断を助ける」「図鑑のように情報そのものを消費する」あるいは「データにメディアが意見を加えて販売するデータジャーナリズム」などの方向が考えられるとし、ポイントとなるのは組み合わせによる価値の創出と指摘しました。例えば、観光ルートと宿泊施設、図書館の貸し出し履歴と書店の在庫情報など、異なる情報を組み合わせることで、新たな価値が生み出されると話しました。

「ガバメントデータは、消費社会のデータと連携させることで、新しいライフスタイルの提案に転用することができるかもしれません。そこにも大きな可能性があると思います」と藤井。民間と行政が連携してお互いのデータを組み合わせて価値を生み出すことが、1つの鍵になりそうです。

蓄積されたデータが新たな価値を生む

その後の意見交換では、モデレーターを務めた渡辺氏がスモールデータの重要性を話題に取り上げました。橋本氏は「市民のニーズに基づくアプリケーションづくりは自治体が得意とするところ。国はデータ活用の枠組みを示したい」と話し、三木氏も「小さな取り組みや活動を大きく育てていくこと」の重要性を強調しました。また、長井氏は「アグリノートに蓄積されたデータを地域で共有して活用する方法も模索したい」と語りました。

庄司氏は、琵琶湖の水位を淡々と伝えるツイッターを例に挙げ、「そのデータの持つ意味や価値は人によって様々。琵琶湖の水位はにおいなどの生活環境の変化に結びついた情報なのだそうです」とデータの持つ未知の可能性を示唆。渡辺氏は、データを連携させることが価値の源泉であることを述べ、「例えば、政策の形成過程のデータも開示される必要があるのでは」と問題を提起しました。

データ主導型イノベーションによるさらなる価値創造には、多くの企業やユーザーがオープンなデータにアクセスでき、自由に活用できるための環境整備が必要であることが明らかとなったシンポジウムとなりました。

Google Public Policy Blog [グーグルの公共政策ブログ]

グーグルでは、公共政策に関する世界の様々な活動をブログで発信しています。ここではその一部を翻訳してご紹介します。

<http://googlepublicpolicy.blogspot.jp/>

◎ グーグル・パテントの対応範囲が拡大 Broadening Google Patents

<http://googlepublicpolicy.blogspot.jp/2013/09/broadening-google-patents.html>

昨年、先行技術検索機能の「Prior Art Finder」を追加し、欧州特許庁(EPO)の特許文獻も検索できるようになったGoogle Patentsですが、今回、新たに4つの特許機関(中国、ドイツ、カナダ、世

界知的所有権機関)の文獻も検索できるようになりました。こうした文獻の多くは特許出願時に先行技術の資料として利用できますので、米国のみならず、世界各国で特許の質の向上にお役立てください。

◎ レポート：グーグルの海賊行為対策 Report: How Google fights piracy

<http://googlepublicpolicy.blogspot.jp/2013/09/report-how-google-fights-piracy.html>

インターネットで公開される音楽や動画、文章、ソフトウェアは増加を続けており、作成者や作成場所も様々です。あらゆる創作活動はインターネットの普及で様変わりしましたが、オンラインでの著作権侵害は相変わらず問題となっています。グーグルのレポート「グーグ

ルの海賊行為対策」は、オンラインでの海賊行為対策としてグーグルが導入したプログラム、ポリシー、およびテクノロジーの概要をまとめたものです。

◎ 透明性向上を求める請願書 A petition for greater transparency

<http://googlepublicpolicy.blogspot.jp/2013/09/a-petition-for-greater-transparency.html>

グーグルは外国情報活動監視裁判所に請願書修正版を提出しました。この請願書は、今年7月にIT企業各社と市民権擁護団体が米国の議会と大統領に送った書簡の内容を反映したものです。外国情報監視法702条に基づき安全保障を目的としてグーグルが受けたあらゆる情報開示要請について、詳しい内容を公表することを許可するよう求めています。また、重要な公益に関わる問題であることから、

非公開ではなく公開の場での公聴会開催も併せて要請しました。また、本日は業界各社や事業者団体と共に、大統領直属機関である「機密情報と通信技術に関する見直しグループ」と面会する予定です。「安全保障を目的とした情報開示要請に関し、政府がその内容の公表を禁止すると、民主主義の根幹である基本的な自由が損なわれてしまう」という主張を繰り返し訴えます。

◎ オープンソースのサービスで利用できる特許を拡大 More patents in the service of open source

<http://googlepublicpolicy.blogspot.jp/2013/08/more-patents-in-service-of-open-source.html>

情報システムのイノベーションはオープンソース・ソフトウェアの登場で加速し、より優れた製品やサービスが低コストで提供されるようになりました。しかし、オープンソース・ソフトウェアの存在が大きくなるに従い、特許攻撃の件数も増えているのが実情です。グーグルは3月、一部の特許についてオープンソース・ソフトウェ

アの開発者やディストリビューター、ユーザーを提訴しないことを宣言する「Open Patent Non-Assertion (OPN) Pledge」を発表しました(ただし、相手から最初に提訴された場合はこの限りではありません)。この宣言の目的は、特許の競争的かつ防衛的な使用を促進し、オープンソースのイノベーションを支えることにあります。

◎ メンタルヘルス・サービスのオンラインでの認知度向上 Increasing Online Awareness of Mental Health Services

<http://googlepublicpolicy.blogspot.jp/2013/07/increasing-online-awareness-of-mental.html>

7月17日に米国で開催されるメンタルヘルス業界の全国大会で、グーグル、ツイッター、フェイスブックの3社が主催して、関連団体向けの講習会を開催します。テーマは「ソーシャルメディア・プラットフォームを利用して認知度を上げる方法」と「オープン・ディスカッションの

新たな手段を構築する方法」で、規模の大小を問わずあらゆる関連団体が参加できます。プレゼンテーションでは、団体専用アカウントの設定、フォロワー・コミュニティの構築、奉仕活動の実施、マルチメディアの統合や成果の測定などを取り上げます。

◎ 海賊・偽造行為の防止策で広告ネットワーク業界が合意 Ad Networks Agree on Industry Best Practices to Combat Piracy and Counterfeiting

<http://googlepublicpolicy.blogspot.jp/2013/07/ad-networks-agree-on-industry-best.html>

グーグルは、ホワイトハウスの知的財産執行調整官(IPEC)、インタラクティブ広告協会(IAB)、および主要な広告ネットワーク事業者と、「海賊行為と偽造行為を防止する広告ネットワークの自主的ベストプラクティスおよびガイドライン」を取りまとめ、署名しました。署名した広告ネットワーク事業者はこのベストプラクティスに基づき、偽造

品の販売を中心に構成されたウェブサイトや著作権を侵害しているウェブサイトの広告プログラムへの参加を禁止するポリシーを掲げ、維持することになります。業界全体でこのベストプラクティスを実践すれば、違法サイトの収益源を絶つことができ、インターネットの健全性を維持してイノベーションを促進することができます。

◎ ウェブの活用が米国に17%増の経済効果をもたらす America's businesses are growing. The web is helping

<http://googlepublicpolicy.blogspot.jp/2013/06/americas-businesses-are-growing-web-is.html>

2012年、グーグルの検索ツールと広告ツールは、米国内190万以上の企業・団体(広告主、サイト運営者や非営利団体)に前年比17%増

となる940億ドル(約9兆2556億円)規模の経済効果をもたらしました。州別の統計と地方で活躍する企業のウェブ活用事例もご覧ください。