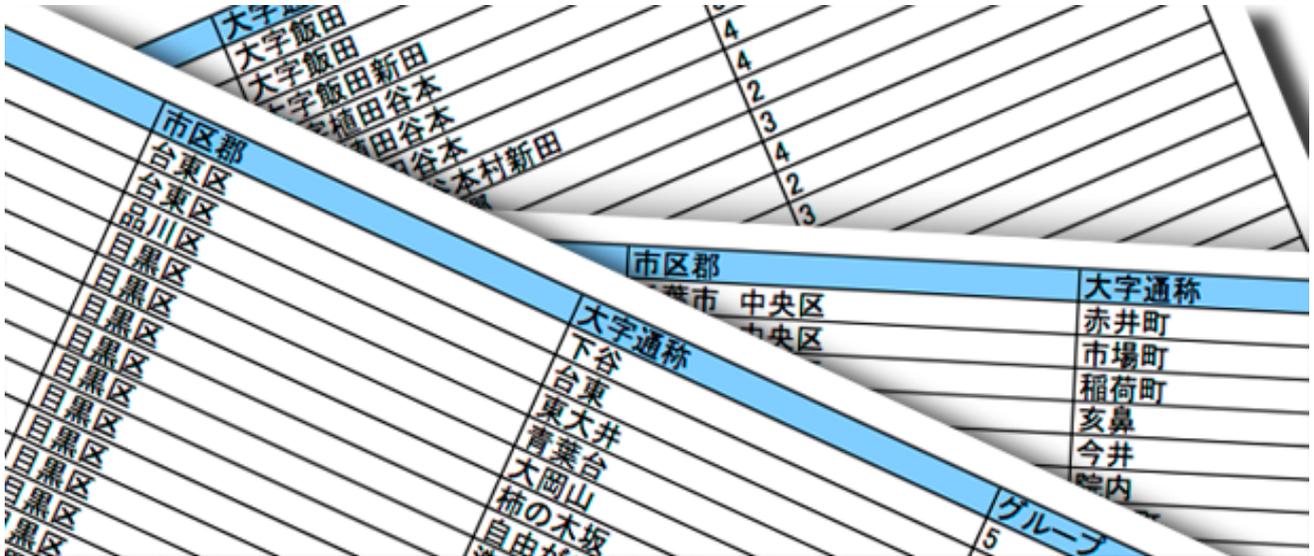


東日本大震災と情報、インターネット、Google

第9回：マシン リーダビリティの重要性



東日本大震災では情報サービスが大いに活用されたが、課題も浮かび上がってきた。その1つが「マシン リーダビリティ」(機械可読性)である。情報はコンピュータで処理しやすい形式で提供されることで、より効率的に、より広く活用されるようになるのだ。

2012年4月27日掲載

日本の情報化は進んでいるのか？

日本において、情報サービスは活用されているのだろうか？

このような質問をナンセンスだと思う人も多いだろう。

誰もが携帯電話を使ってコミュニケーションを取っているし、会社ではパソコンで文書を作り、表計算ソフトにデータを入力することが当たり前になっている。

ところが異なる観点から見ると、事情は違ってくる。日本生産性本部が発表した「労働生産性の国際比較 2011年版」によれば、日本の労働生産性は OECD 加盟 34 カ国中第 20 位。経済産業省は、2010 年に発表した「情報経済革新戦略」(<http://www.meti.go.jp/press/20100531005/20100531005-3.pdf>)の中で「我が国の労働生産性は 90 年代以降停滞しており、先進諸国中最低水準。IT を積極活用している諸外国と比べ、我が国の IT 投資は質も量も不足していることが一因」と指摘している。

この連載で紹介してきたように、東日本大震災では Google や Twitter などをはじめとする情報サービスが大いに活用された。しかし、当時の状況を細かく見ていくと、組織間での情報のやり取りに課題があることがわかってきた。

その課題が端的に現れたのが、東京電力の「計画停電」を巡る混乱ではないか。ここで見てきたのは、「マシン リーダビリティ」(機械可読性)を意識した情報提供の重要性である。

情報が錯綜した東京電力の計画停電

2011年3月11日(金)に発生した東日本大震災によって、東京電力、東北電力管内では、福島第1・

東日本大震災と情報、インターネット、Google

元データの再現性は非常に高い。元データを改ざんしにくいという特徴もあるため、出版・印刷分野での校正やデータの入稿、企業や官公庁での文書配布などに広く使われることになった。

数多くの長所を備える一方、PDF には短所もある。まず、PDF では文書としての見た目を優先しており、コンピュータでの自動処理についてはあまり考慮されていない。

例えば、見た目がまったく同じ複数の表が PDF 形式で提供されていたとする。見た目が同じなら、人間が読む分には何の支障もない。しかし、1 つの表は、ワープロソフトを使って罫線文字（「+」「-」などの文字）を使って体裁を整えているのかもしれない。空白文字を入れて、空きを調整しているのかもしれない。複数行にわたっている項目は、改行を入れているのか入れていないのかもわからない。PDF では見た目が同じでも、含まれているデータはまったく異なる構成になっていることがある。

また、PDF の仕様は複雑であり、閲覧・編集するためのソフトによっても挙動が変わってくる。一続きになっている段落をまとめてコピー＆ペーストしようとしても、必ずしも意図通りに文字を選択できるとは限らない。

阿部秀彦を始めとする Google の地図関連エンジニアたちは、電車運行情報などさまざまな災害関連情報を地図上に表示してユーザーに提供していた。計画停電についても、東京電力提供の PDF から住所やグループ番号を抽出して地図上にマッピングしようとしていたが、この作業にはかなり手こずったという。表が複数ページに分かれている場合や、項目が改行で区切られている場合など、複雑な条件を解析するスクリプト（プログラム）を書き、作業を進めていった。苦勞の末、16 日（水）には、Google マップ上で計画停電の地域を地図上で確認できる「停電マップ」が公開された。

計画停電の開始から数日後には、東京電力から PDF と共に、元データの Excel ファイルも提供されるようになったが、この処理も難物だった。一見データが整然と並んでいるようでも非表示の列があったり、ファイルによって項目名がいきなり増えるということもあった。表の形式も統一されておらず、地域によって専用のスクリプトを書く必要もあったという。



Google が公開した停電マップでは、ある地域がどのグループに属しているかを調べられた。

グラフのデータも公開することで、活用の幅が広がる

計画停電が続く中、3 月 22 日（火）に東京電力は電力使用状況グラフを公式サイト上で公開した。これは、東京電力管内における 1 時間ごとの電力使用実績を棒グラフで掲載したものである。前日の実績、そして前年同日の実績も折れ線グラフで表示されているという、これまでにない取り組みであった。

ただし、このグラフは画像データとして掲載されており、元の数値データが公開されていなかった。数値データとして用意されていれば、集計して特定期間の使用状況を分析するなど、活用の幅が広がるはずだ。データ活用を希望する企業からも同様の相談がなされていたこともあり、経済産業省 情報経済課の吉川徳明さんと

情報政策課の守谷学さんは東京電力に働きかけ、24日(木)から数値データもCSV形式(テキストをカンマで区切った形式)で提供されるようになった。CSVならコンピュータで読み取って容易に処理・加工することができる。

実は、CSV形式でデータが提供される以前に、グラフの画像データをコンピュータで処理して、そこから元の数値データを推測するプログラムを書いていた強者もいた。ただ、当然のことながら、情報元自身のデータが一番正確であり、処理もしやすい。

CSV形式でデータが公開されてから2日間ほどで、ネット上には50を超える電力状況データを利用したサービスが立ち上がった。汎用的なデータにすることで、活用範囲が格段に広まるという好例だろう。

こうしたデータ利用の取り組みを、より多くの企業の協力を得て加速するため、経済産業省商務情報政策局は3月30日(水)に「東北地方太平洋沖地震等に係る情報提供のデータ形式について(周知依頼)」と題した事務連絡を、社団法人日本経済団体連合会に対して行っている。

この文書の一部を引用しよう。

「データが直接htmlで記述されていたり、csv等比較的自動処理が容易な形式でデータが供給されていれば、インターネット上のさまざまなコンテンツやアプリケーションの制作者の方々に携帯電話でも閲覧できるようなアプリケーションの開発や、より使いやすいwebページの構築などを促すことができます。また、これによって、被災地はもとより、直接被災地以外の地域においても、提供情報の利用を促進することが期待できます。つきましては、円滑な情報提供を図る観点から、ホームページにおいて情報提供を行う場合には、極力PDF等自動処理がしにくいデータ形式のみによらず、htmlやcsv等の自動処理に適したデータ形式を併用したり、別途オープンな情報提供APIを整備するなど、データを提供する方法について、ご無理のない範囲で、特段の配慮をいただけますよう、貴会会員各社にご周知方お願い申し上げます。」

民間企業に対して、行政側からこうしたデータ交換の方針について依頼を行ったのは、異例である。

活用しやすいデータを公開するために

今回取り上げた課題について、整理しよう。

まず、情報化というのは、紙で行っていた作業を単純にコンピュータに置き換えるだけでは不十分だ。作成したデータを人間が読みやすくするのは別に、コンピュータにも処理しやすくすることで何倍にも活用の幅が広がる。

特に統計的なデータを公開するのであれば、きちんと構造化するのが望ましい。構造化といっても、必ずしも複雑な作業が必要になるわけではない。例えば、表計算ソフトを使ってデータを作成するのであれば、行と列を意識して、どのデータがどの項目を表しているのか、曖昧にならないようにする。空白文字や空行などで見栄えを整えるのはやめる、といった点に注意する。ファイル名やフォルダ名についても、自動処理が行いやすいように何らかの規則性を持たせるべきだろう。そして先にも述べたように、統計データの提供が目的なら、



東京電力の「でんき予報」からは、CSV形式でのデータダウンロードも可能になっている。

CSV などのシンプルなフォーマットも同時に提供すべきだ。

IT 活用というと大規模なシステムが話題に上ることが多いが、こうした小さな点に注意するだけで組織内／間のデータ再利用は格段に容易になる。これは、震災のような非常時に限ったことではない。

ただし、こうした取り組みは、個々人がやるだけでは不十分なも確かだ。企業や自治体全体に、データの再利用や自動処理することの意義を浸透させることが重要になってくる。

取材、執筆、編集：林信行 / 山路達也