

地方自治体の持つオープンデータに関する 活用支援手法の提案

渡邊雄貴^{†1} 内田拓馬^{†1} 秋本泰良^{†1} 太田幸司^{†1} 梶克彦^{†1} 内藤克浩^{†1}
水野忠則^{†1} 菱田隆彰^{†1}

概要：近年、地方自治体は住民の住みよい街づくりや行政の「見える化」を進めるため、自身の所有する情報をオープンデータとして公開し始めている。しかし、複数の地方自治体が公開するオープンデータは、同種のデータでありながら含まれる項目名は統一されていないことや、含まれる情報の量や内容も不揃いであることがあり、データを実際に使用する際の障害となっている。また、自治体の持つ情報は主に所在や連絡先、機器の型番など自治体が対象を管理するために必要なものに限定されており、実際にそれらのデータを商用サービスなど他の目的のために使用する際には、必要な情報を全て満たすことが出来ない場合が多い。本研究では、自治体の公開しているオープンデータに着目し、複数の自治体が公開するデータを効率良く利用するために必要なデータ活用支援手法の提案を行う。具体的には、データ項目名の不統一な点や、含まれる項目の不揃いな点は、IPA が公開する共通語彙基盤[1]を用いた項目の統一、オントロジーを用いた必要なデータ項目の体系化を行うことでデータ内容を整理し検証できる仕組みを提案する。また、構築したオントロジーによって、公開しようとするデータに対して適切な内容を含んでいるかが確認できるチェックシステムの試作を行う。もう一つの問題点である実用上不足するデータ内容を補足する手段として、利用者の手によってオープンデータに関する追加情報を提供させることで、データ全体の補強を行う情報収集支援システムの設計と試作を行う。

1. はじめに

平成 25 年 6 月に、総務省より世界最先端 IT 国家想像宣言[2]が閣議決定された。この中でオープンデータを推進するために、ロードマップの策定、自由な二次利用可能を認める利用ルールの見直しを行うとともに、機械判読に適した国際基準データ形式でデータの公開が宣言された。それに伴い福井県鯖江市を始め各地方自治体が、住みよい街づくりや行政の「見える化」や国民参加を進めるため、AED やゴミ情報などの地方自治体が持っているデータの公開を進めている。

総務省は機械判読に適したデータ形式[3]での公開を推奨しているため、現在地方自治体によって公開されているオープンデータの多くは、CSV などの表形式のデータで公開されている。また、総務省からデータを適切な形式に変換し公開する為のツールを提供しているため、公開できるデータを作成する手段は用意されていると言える。しかし、オープンデータは実際に他者が利用できることが必要である。

地方自治体の持つ情報は、その自治体で管理している設備や施策に関する情報であり、地域ごとに管理している同じ種類の情報である。これらの情報は他社が活用することを考えた場合、できる限り同じ手法を用いてデータを取り扱えることが望ましい。しかし、現状では公開されているデータの項目名は自治体ごとに様々な名前がつけられており、利用時にはその判断を個別に行わなければならない。

また、公開するデータ項目は自治体ごとに定めており、使用したい情報が全ての自治体の情報から得られるわけではない。例えば、観光施設の電話番号や緯度、経度などは、スマートフォンのアプリケーションなどでは良く利用される情報だが、一部の自治体のオープンデータにはそれらが含まれていない場合があり、利用者側が用意する必要がある。

オープンデータの項目名の統一化やユーザビリティ、アクセシビリティに関するデータを体系化することの必要性をまとめた研究はいくつか行われている。庄司[4]は、「オープンデータの定義・目的・最新課題」の中で形式や語彙が揃っていないため使い勝手が悪いと述べている。また、「地方公共団体におけるオープンガバメントの推進に関する調査研究 [5]」ではデータ作成において指針となるものを明示しなければ職員による取り組みが失速してしまう可能性があるとして述べており、山根らは「オープンデータの RDF 化のための類似項目関係の推定 [6]」において既存の表形式のオープンデータは属性と値の組の単純な集合体であることを指摘し、オントロジーを活用し従属関係を明確化した上で属性間の意味構造を推定する研究を行った。

オープンデータを今後より良く活用するためには、公開するデータについて内容を整理し、同種のデータが複数存在する場合は、できる限り均質な内容とデータ形式で公開されるべきである。また、公開されたオープンデータとともに入手出来たほうが良い。足りない情報をうまく収集し公開できる仕組みが求められる。

本研究では、自治体の公開しているオープンデータに着目し、複数の自治体が公開するデータを効率よく利用するために必要なデータ活用支援手法の提案を行う。具体的

^{†1} 愛知工業大学
Aichi Institute of Technology

は、データ項目名の不統一な点や、含まれる項目の不揃いな点は、IPA が公開する共通語彙基盤を用いた項目の統一化、オントロジーを用いた必要なデータ項目の体系化を行うことでデータ内容を整理し検証できる仕組みを提案する。また、構築したオントロジーによって、公開しようとするデータに対して適切な内容を含んでいるかが確認できるチェックシステムの試作を行う。もう一つの問題点である実用上不足するデータ内容を補足する手段として、利用者の手によってオープンデータに関する追加情報を提供させることで、データ全体の補強を行う情報収集支援システムの設計と試作を行う。

2. オープンデータ分析

2.1 オープンデータの現状

現在、福井県鯖江市を始め各地方自治体がオープンデータの公開を始めている。また、公開されるデータは様々

であり、統計や人口や政府が調査したデータや AED、ゴミ情報、避難所の情報など、自治体が運用上必要な情報や自治体が整備し管理が必要な公共設備の情報が中心である。データ形式は CSV や XLS といった表形式のデータを公開しているところが多く、それらのデータを RDF 形式に変換したデータを公開している自治体も存在している。公開されているデータに含まれる項目は自治体ごとで独自で定められている。そのためデータに対する項目名も自治体ごとにさだめているため複数の自治体で同種のデータが公開されている場合、同じ内容を示す内容であってもその項目名は統一されていない。また、データの項目はさだめられていてもその内容が含まれていないものも存在し、用意されている全ての項目について実際のデータが揃っているわけではない。表 1 は静岡県島田市が公開している AED に関するデータである。データの項目は設置されている AED に関するデータである。データの項目は設置されている AED の画像を入手できる URL や、AED の危機に関する

表 1 静岡県島田市の AED データ

	A	B	C	D	E	F	G	H	I	J	K	L
1	名称	名称(英語)	設置場所	住所	緯度	経度	画像URL	パッドのタイプ	シリアルナンバー	モデルナンバー	AEDの有効期限	AEDのタイプ
2	島田市役所:福祉課カウンター		福祉課カウンター	島田市中央町	34.836439	138.176044						
3	島田市立看護専門学校:体育館		体育館	島田市野田10	34.847132	138.183893						
4	保健福祉センター:1階受付		1階受付	島田市中河町	34.840594	138.179057						
5	プラザおおり:1階事務所		1階事務所	島田市中央町	34.836577	138.174602						
6	島田市博物館:本館事務所		本館事務所	島田市河原1丁	34.831342	138.154388						
7	ローズアリーナ1階:1階総合受付		1階総合受付	島田市野田16	34.850162	138.17528						
8	ローズアリーナ2階:2階プール		2階プール受付	島田市野田16	34.850162	138.17528						
9	ローズアリーナ3階:3階弓道場		3階弓道場	島田市野田16	34.850162	138.17528						
10	島田球場:1階ロビー		1階ロビー	島田市横井4丁	34.824926	138.174597						

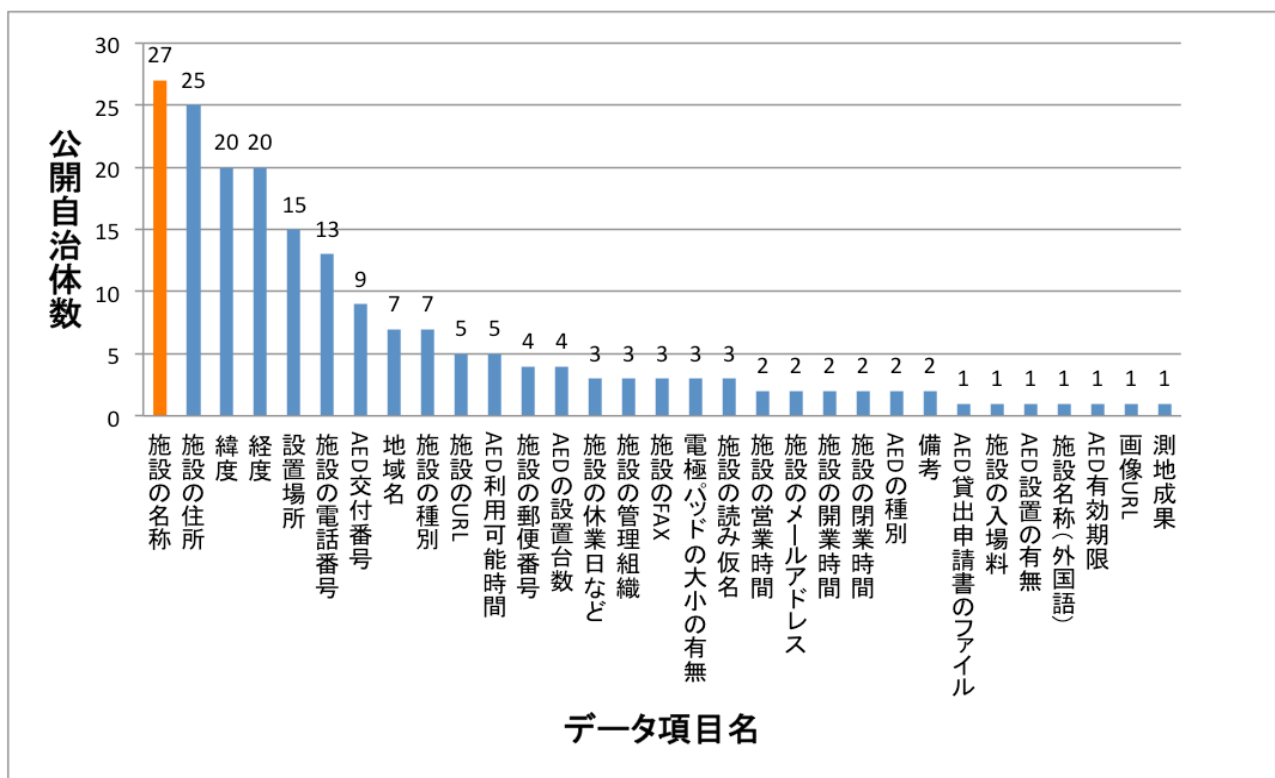


図 1 AED データ頻出項目グラフ

情報を含むための項目があるが、実際に公開されているデータには含まれていない。現状では、必要だと思われる項目と実際に自治体が用意できる情報には開きがあることがわかる。図1、図2は、すでに公開されているAEDおよび観光情報について、公開しているデータの項目とその項目を公開している自治体数を示したグラフである。施設名や緯度、経度または住所は多くの自治体が公開していることがわかるが、その他の多くの情報は限られた自治体しか公開しておらず、公開すべき項目が適切に選定されていないことも推定できる。

2.2 考察

現状では情報を公開することが優先されるため、各自治体が公開データに含む内容を十分吟味できないことが要因として考えられる。また、実際に公開する内容を推定する際に必要な条件が統一されていないため、それぞれの自治体の手探りで内容を定めていることも問題である。

実際に公開されたデータは、企業や一般市民が入手して利用することになるが、項目名の不一致や項目内容の不足があった場合、それらを修正し補足するために大きなコストをかけることになる。利用に不便なデータは、例え膨大なデータ量を持っていたとしても、整理にかかるコストが見合わなければ使われることはないだろう。今後オープンデータが積極的に利用されるようにするには、公開されるデータはできる限り項目名や項目の内容が整った均質なデ

ータを最初から提示するべきであり、自治体で用意できない情報は、何らかの形で別途入手し元のオープンデータとともに簡便に入手できるよう集約できる仕組みが必要である。

本稿では、これらの問題に対して二つの手法を提案する。一つ目は、データ作成者が適切な項目を確認する手法である。具体的には、既に公開されているデータの項目をオントロジーによって体系化し、オントロジーと今後公開しようとするデータ内の項目と比較を行うことで、作成者が項目の確認を行う。もう一つは、データ内で内容が用意できない項目について、データを補足する手法である。具体的には、補足してほしいデータに関係するゲームを一般市民に行ってもらい、そのゲームの活動によって必要な情報を提供してもらうことで、データの補足を行う。

3. 公開データ確認システム

本章では、公開されるオープンデータの内容を均質化するため、既存のデータ項目を用いた確認システムを構築する。データを確認するために、

- (1) 既存のオープンデータからオントロジーを作成してデータ項目の体系化
- (2) データ項目について共通語彙基盤を用いて統一化を行う。また、それらの情報を自動的に処理するデータ確認システムの試作を行う。

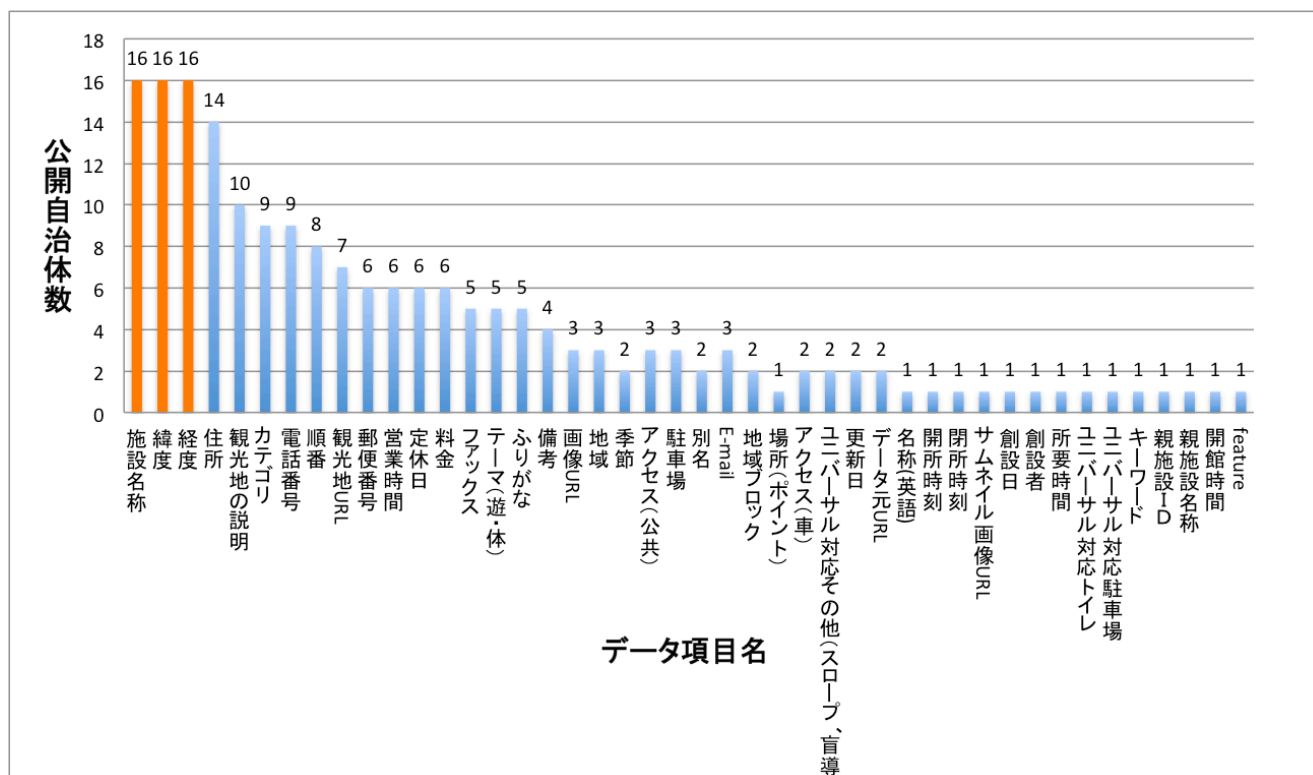


図2 観光情報の頻出項目グラフ

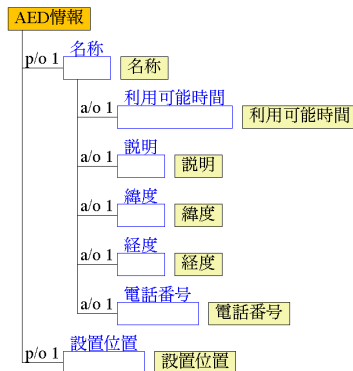


図3 AEDデータのオントロジー

3.1 項目の体系化と統一化の手順

まず、公開予定のオープンデータに必要なデータ項目を選定するために、既存のオープンデータの調査を行い、公開されているデータ項目数を集計する。多数の自治体が公開しているデータ項目を選定し、オントロジーの概念として用いる。今回我々はAED情報の必要項目の選定及びオントロジーによる体系化を行う。選定した項目を以下に示す。また製作したオントロジーは図3に示す。

- ・ 名称
- ・ 利用可能時間
- ・ 説明
- ・ 緯度
- ・ 経度
- ・ 電話番号
- ・ 設置位置

図1を元に公開している自治体が半数を超えた項目は必要性が高い内容であると判断し、オントロジーに用いた。ものから選定を行った。また、二番目に多い住所の項目は、今回アプリ製作をする際座標データの方が重要であると考え採用しなかった。また、施設名の項目が多いことから、地方自治体の管理する施設は開館閉館時間が有り、利用できない場合があるため、説明や、利用可能時間の項目も必要であると判断した。

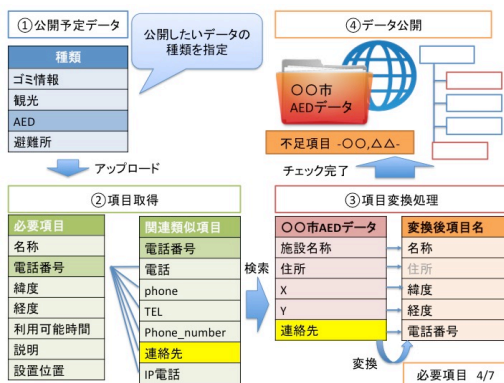


図4 システムの処理課程

次に、データ項目を統一化するため、共通語彙基盤を用いる。データ項目は各地方自治体によって様々であるため、予想される語彙を登録したデータベースを作成し、データ項目の差し替えを行う。

3.2 確認システムの概要

本システムは、地方自治体で公開する予定のデータをアップロードし、項目の有無の確認を行い、共通語彙基盤で項目を統一化した後公開する為のシステムである。手順を次の図4に示す。

- ①公開したいデータをアップロードしどの種類のデータを公開するかを選択する。
- ②内部で所持している、必要項目と関連類似項目を元にアップロードされたデータの項目名をチェックする。
- ③チェックして見つかった項目に関しては、共通語彙基盤の項目名に変換し、本システムに保存する。
- ④チェック及び変換の結果についてオントロジーを用いその関係性をユーザに表示し正しい内容と判断できたらデータを公開する。

本システムを用いることでどの項目が必要なのか、またどのデータが足りないのかを表示し、また項目名を統一化することで機械判読に適したデータ形式で変換した際に、オープンデータの自治体間でのデータ相互運用に繋がると考えられる。

3.3 データ解析の結果及び考察

地方自治体が公開しているAEDデータのCSVファイルを新規に公開しようとするデータを想定して、本システムにアップロードした結果を図5に示す。

今回我々は必要項目として7項目を提案した。データを検証すると、殆どの自治体が4項目以上のデータを公開し

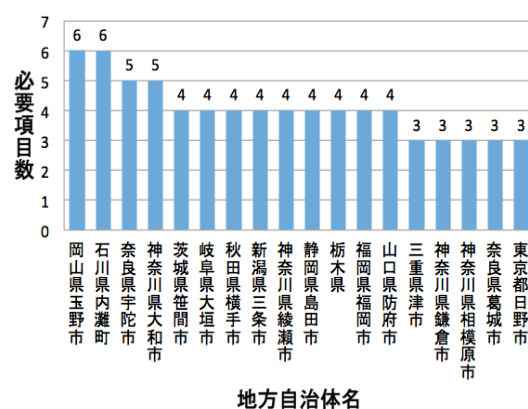


図5 CSVデータの必要項目チェック結果

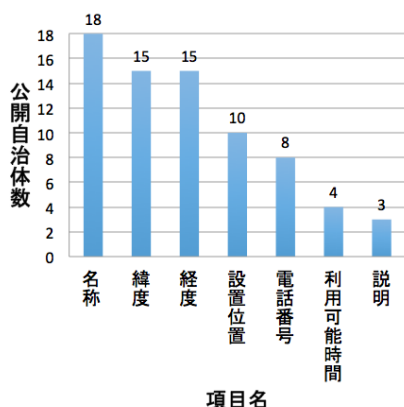


図6 自治体の公開項目数合計

ており、また図6より殆どの自治体が名称、緯度、経度の項目を記載しているため、データを公開する際の指標となるのではないかと考える。

4. 利活用結果及び考察

本章では、自治体では準備が難しいデータについて、一般市民などの利用者に対し自発的にデータ収集、提供に協力を促すデータ補足システムについて述べる。利用者に自発的な行動を促すため、ゲーミフィケーションを用いる。

4.1 データ補足システム

我々は本システムを利用し AED を検索するための「AEDMap-Hands-」というスマートフォンアプリを製作する。本アプリは、オープンデータの AED 情報を元に検索し、近くの AED を検索する。見つかった地点にどれくらいの時間で到着するのかを競わせるゲームである。現状の AED は整備が進んでいるが、必ずしも適切に利用されているとは言い難い。原因としては、AED に対する認知度の低さが考えられるため、利用者に対して認知度向上の重要性を解くことで、アプリの利用を促す。ゲームを行ったユーザは探した AED の写真を撮影し、ゲームのゴールとして写真をアプリから投稿する。この画像データは公開されている AED 情報の補足データとして公開することで、オープンデータの内容が充実することになる。

システムの概要を図7に示す。自治体で公開した AED の情報は、本アプリの AED 情報として利用し、ゲームを行って得られた写真はデータベースに集約され、オープンデータの一部として利用可能になる。

4.2 AEDMap-Hands-の動作概要

データは地方自治体が公開しているデータを用い最小限のデータ項目であると考えた名称、緯度、経度のデータを用い実際にアプリで活用できるかを検証した。

図8より、ダウンロードした AED のデータで近くの

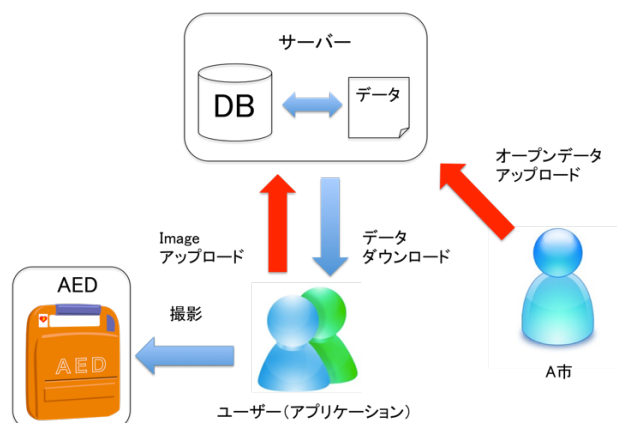


図7 AEDMap-Hans-利用時のシステム図

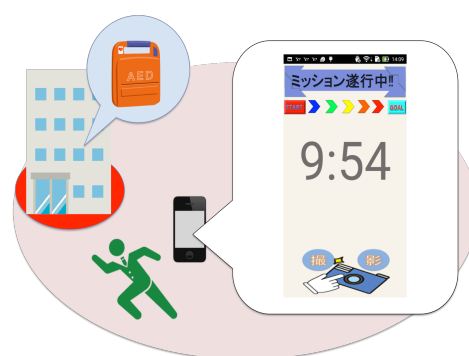


図8 ゲーム実行中画面



図9 AED 到着時の写真撮影時の画面

AED を検索しゲームを開始した。図9より、ゴール時に AED の画像とユーザ名を本システムに保存し、ゲームを終了する。

5. まとめ

本研究では、地方自治体の持つオープンデータに着目し、データの活用を活性化するための手法について検討を行い、公開されるデータを均質な内容とするための確認手法と、利用する際に不足する情報を利用者に補填させる手法の提案を行い、それぞれ試作システムを構築した。本システムを利用することで、オープンデータ内の項目名や項目の内

容が揃ったデータを入力することが可能となる。また、利用者からの情報を元にオープンデータの内容が補強されることで、更に便利なデータとして活用される可能性を示すことが出来た。オープンデータの利用活性化の一助となることが期待できる。

参考文献

- [1] 共通語彙基盤
(<http://goikiban.ipa.go.jp/node756>), アクセス日:2015/11/08
- [2] 世界最先端 IT 国家創造宣言
(<http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou1.pdf>), アクセス日:2015/11/08
- [3] オープンデータと言えるための条件
(http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/opendata/opendata01.html), アクセス日:2015/11/08
- [4] 庄司昌彦:オープンデータの定義・目的・最新の課題, 機関誌「智場」119号, p04-p15(2014)
- [5] 地方公共団体におけるオープンガバメントの推進に関する調査研究
(<https://www.kantei.go.jp/jp/singi/it2/densi/jichibukai/dai2/siryou2bessi.pdf>), アクセス日:2015/11/08
- [6] 山根昇平, 鵜飼孝典:オープンデータの RDF 化のための類似項目間関係の推定, 第 77 回全国大会講演論文集, Vol2015, No1, p481-p483