

# データセットの情報提示及び形式自動検査によって 自治体のオープンデータ公開を支援するソフトウェアの評価

渡邊 輔<sup>†</sup> 陳 思楠<sup>†</sup> 佐伯 幸郎<sup>††</sup> 中村 匡秀<sup>†,††</sup>

<sup>†</sup> 神戸大学 〒657-8501 神戸市灘区六甲台町 1-1

<sup>††</sup> 高知工科大学 〒782-8502 高知県香美市土佐山田町宮ノ口 185

<sup>†††</sup> 理化学研究所・革新知能統合研究センター 〒103-0027 東京都中央区日本橋 1-4-1

E-mail: †tasuku@ws.cs.kobe-u.ac.jp, ††chensinan@gold.kobe-u.ac.jp, †††saiki.sachio@kochi-tech.ac.jp,  
†††masa-n@cmds.kobe-u.ac.jp

あらまし 政府や自治体は、保有データをオープンデータとして公開し、データ利用を促進することで都市に新たな価値を生みだすことを目指している。しかし、公開されているオープンデータは、データの形式やデータの内容が自治体によって異なっており、利活用が困難である。オープンデータとして公開されるデータは、機械判別できるファイル形式でかつ、データ構造が一定の標準に従った形式であることが理想であるが、現状では、自治体の職員が手作業でデータ形式の検証・修正を行っているため、標準への準拠が不十分である。本稿では、先行研究として実装した、データ形式を自動検査することによって、標準化されたオープンデータ作成のためのデータ検証・修正プロセスを支援するシステム「OpenDataValidator」を利用した、自治体職員によるオープンデータ検証・修正実験を行った。その結果から、提案システムを利用して実際の自治体のオープンデータ作成を容易化することができるかを検証した。

キーワード オープンデータ、地方自治体、データ標準化、データ検証

## Evaluation of software to support open data publication by municipalities through information presentation and automatic format checking of datasets

Tasuku WATANABE<sup>†</sup>, Sinan CHEN<sup>†</sup>, Sachio SAIKI<sup>††</sup>, and Masahide NAKAMURA<sup>†,††</sup>

<sup>†</sup> Kobe University, Rokkodai-cho 1-1, Nada-ku, Kobe, Hyogo, 657-8501 Japan

<sup>††</sup> Kochi University of Technology, 185 Tosayamadacho Miyakouchi, Kami, Kochi, 782-8502 Japan

<sup>†††</sup> Riken AIP, 1-4-1 Nihon-bashi, Chuo-ku, Tokyo, 103-0027 Japan

E-mail: †tasuku@ws.cs.kobe-u.ac.jp, ††chensinan@gold.kobe-u.ac.jp, †††saiki.sachio@kochi-tech.ac.jp,  
†††masa-n@cmds.kobe-u.ac.jp

**Abstract** The format and content of the open data released to the public differ from municipality to municipality, making it difficult to make use of the data. Ideally, data to be released as open data should be in a file format that can be machine-discriminated, and the data structure should be in a format that conforms to a certain standard. In this paper, we conducted an experiment of open data validation and correction by municipal employees using "OpenDataValidator," a system that supports the process of data validation and correction for the creation of standardized open data by automatically checking the data format, which was implemented as a previous study. Based on the results, we verified whether the proposed system can be used to facilitate the actual creation of open data by local governments.

**Key words** Open data, Local government, Data standardization, Data validation

### 1. はじめに

情報化社会が進行するなか、様々な場所に蓄えられたデータ

を活用することで社会を豊かにする方法が模索されており、国や地方自治体が保有するデータをオープンデータ [1] として公

開する動きが広まっている。オープンデータは、様々なステークホルダーによる利活用が期待され、実際にオープンデータを活用したサービスが提供されている。

しかし、公開されているオープンデータは、データの形式やデータの内容が自治体によって異なっており、利活用が困難であるという問題が存在する。オープンデータとして公開されるデータは、機械判別できるファイル形式でかつ、データ構造が一定の標準に従った形式で公開されることによって、十全に利活用を行うことができる。しかし、現状では、自治体の職員が手作業でデータ形式の検証・修正を行っているため、標準への準拠が不十分なデータがオープンデータとして公開されている。

そこで本研究グループでは、このような問題の解決を目指して、データ形式を自動検査することによって、標準化されたオープンデータ作成のためのデータ検証・修正プロセスを支援するシステムを提案し、その具体的な実装として、**OpenDataValidator**の実装を行っている。

本稿では、OpenDataValidator が実際の自治体の職員によって利用された際のシステムの性能について評価するための実験を行い、その結果について考察を行った。

実験は兵庫県三田市の職員に協力を依頼し、実際にオープンデータ作成に関わる職員に、OpenDataValidator を活用した、オープンデータ検証・修正作業を行ってもらう形で実施した。

その結果、OpenDataValidator によって、標準に準拠したデータセットを作成することができることが確認された。また、実際のオープンデータ公開作業においても、提案システムは自治体職員が容易に利用することができ、データチェックにかかる時間を削減することができる可能性が示された。

## 2. 準備

### 2.1 オープンデータ

政府及び地方自治体はデジタル技術をまちづくりに取り入れ、都市機能の最適化や市民生活の質の向上を目指すスマートシティの取り組みを進めている。その取り組みの一つとしてオープンデータの取り組みがある。政府や自治体が保有する公共データを公開し、データ利用を促進することで都市に新たな価値を生みだすことを目的としており、各団体はオープンデータ基本指針 [2] に従ってデータ公開を行っている。オープンデータ基本指針によると、オープンデータは「営利、非営利目的問わずに二次利用可能」、「機械判別に適したデータ構造を持つ」、「無償利用可能」などの特性を満たすものであり、誰でもデータを活用できる。

オープンデータの公開形式については、機械判別に適した構造及びファイル形式で公開することを原則としており、オープンデータの達成度の評価指標として用いられている「5つ星」[3]の指標を参考に、より活用しやすい形式での公開に努めることとされている。

また、データの利活用を容易にするため、データの構造についても、一般的に利用されている、自治体標準オープンデータセットなどの標準形式に従ったデータ作成が推奨されている。

### 2.2 自治体標準オープンデータセット

自治体標準オープンデータセット [4] とは、オープンデータの公開とその利活用を促進することを目的として、政府として公開を推奨するデータと、公開するデータの作成にあたり、準拠すべきルールやフォーマット等を取りまとめたものである。

この標準に従って作成されたデータセットは、同じく標準に従った他のデータセットと容易に結合することができ、状況を選ばず、様々な場面で活用することができる。デジタル庁のホームページにて、各データセットのデータ項目定義書が公開されており、各自治体はこのデータ項目定義書に従ってデータを公開することが推奨されている。データ項目定義書では、データセットに必要なデータ項目の定義や、そのデータ項目のデータ形式・制約について定義されている。

全てのオープンデータがこのような標準形式に従って公開されることが理想ではあるが、標準形式に従ったデータを作成するには、専門知識と相応の時間が必要となり、自治体の職員の負担が大きい。オープンデータとして公開されるデータは、数十の列と数百の行で構成されていることが多く、そのデータの構造やデータの内容を目視で確認することは困難である。また、データの修正箇所を正しく把握し、修正するには専門知識が必要であり、自治体の規模によっては、人的リソースの不足や専門知識を持った職員の不足によって、データの修正が困難である場合がある。

### 2.3 先行研究:OpenDataValidator

本研究グループでは、このような課題の解決を目指して、標準化されたオープンデータ作成のためのデータ検証・修正プロセスを支援するシステムを提案し [5]、その具体的な実装として、OpenDataValidator の実装を行っている。

## 3. OpenDataValidator

### 3.1 システム概要

OpenDataValidator は、データの形式検査の自動化によって、オープンデータ検証・修正プロセスを支援する Web アプリケーションである。標準化されたオープンデータの構造に定義される、必要な項目名やその項目のデータ形式等の情報を DB に保存し、その情報をもとに入力データが標準に準拠しているかどうかを自動で判別する機能を提供する。また、データチェック機能は、一般的な自治体職員が利用することを想定し、高度な知識を必要としない仕組みを実現する。これらの要件を満たすため、OpenDataValidator は、図 1 に示すように、3つの機能を提供する。以下では、それぞれの機能について詳しく述べる。

#### 3.1.1 F1: データチェック関数管理機能

データチェック関数管理機能は、システム管理者（プログラムの知識を持つ技術者）が、データチェック関数をシステムに登録するための機能である。データチェック関数とは、入力となる列データに対して、その列が持つ各値が、求められるデータ形式に従っているかどうかを判定するための関数である。確認するデータ形式は、データ構造の標準形式の項目定義書に記

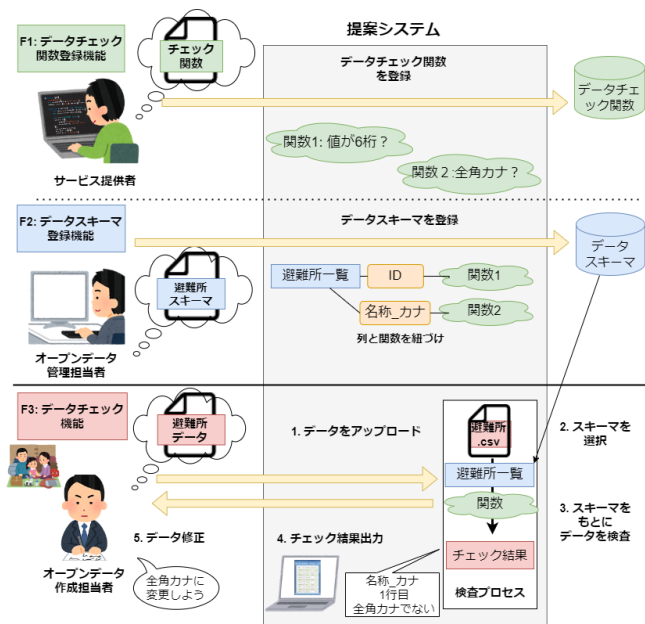


図 1 OpenDataValidator のシステム概要

載されている情報をもとに作成される。データチェック関数をシステムに登録する際には、システムから関数を自動で呼び出すための情報や登録する必要がある、関数名や、引数情報、関数詳細等の情報を登録する。

### 3.1.2 F2: データスキーマ管理機能

データスキーマ管理機能は、オープンデータ管理担当者（データ構造の標準形式を理解している人）が、オープンデータの標準形式をシステム上に定義し、システムに登録するための機能である。データスキーマとは、データセットの構造を定義したものであり、データセットに必要な列名の情報や、各列のデータが満たすべきデータ形式の情報を含む。本機能では、Web アプリの画面上で、データセットの詳細や、データセットの列名を入力し、各列に対して適用するデータチェック関数を選択することで、データスキーマをシステム上に定義する。

### 3.1.3 F3: データチェック実行機能

データチェック実行機能は、オープンデータ作成担当者（自治体のオープンデータ作成を担当する職員）が、データチェック関数管理機能とデータスキーマ管理機能で管理されている情報をもとに、データセットの形式を自動でチェックするための機能である。本機能では、Web アプリの画面上で、チェック対象のデータセットをアップロードし、検査に利用するデータスキーマを選択することで、データセットの形式を自動でチェックすることができる。検査結果は、図 2 に示すようにデータテーブル上に具体的な修正箇所を表示し、図 3 に示すようにその詳細をコンソール上に出力することで、データセットの修正箇所や修正方法を確認することができる。

## 4. 地方自治体職員による評価実験

### 4.1 実験の目的

OpenDataValidator は、自治体職員による、標準に準拠したオープンデータ公開を支援するためのシステムであるが、実

際の自治体職員が本システムを利用して、データのチェック作業を正しく行うことができるかどうかを検証する必要がある。そこで、兵庫県の三田市役所に協力を依頼し、実際にオープンデータ作成に関わる職員に、OpenDataValidator を活用した、オープンデータ作成作業を行ってもらった実験を実施した。本章では、以下のような RQ1, RQ2 をリサーチクエスションとして設定し、これらのリサーチクエスションを実験を行うことによって検証する。

**RQ1: 自治体職員が提案システムを利用して、標準に従ったデータセットを作成できるか**

提案システムを利用して、データセットの作成、データ形式の検査を行ってもらい、作成されたデータセットがどの程度自治体標準オープンデータセットに準拠できているかを検証する。

**RQ2: 提案システムは実際のオープンデータ公開作業にどの程度受け入れられるか**

提案システムを利用して、データセットの作成、データ形式の検査を行ってもらい、アンケートを通して、提案システムを利用したデータセット作成の実際の現場における活用可能性について検証する。

### 4.2 実験の概要

実験では、協力していただいた職員に、提案システムの利用状況をそれぞれ想定した 2 つの実験タスクを実施してもらい、その後、RQ2 を確かめるためのアンケート項目に回答してもらった。また、RQ1 を確かめるため、タスク終了後、それぞれのタスクで作成されたデータセットに対して検証を行った。以下では、各タスクの詳細と RQ の検証方法について詳しく述べる。

#### 4.2.1 Task1. 三田市指定緊急避難場所一覧データセットの形式チェック

このタスクでは、現在公開しているオープンデータセットを自治体標準オープンデータセットに準拠させたいという利用状況を想定した作業を行った。具体的には、現在三田市オープンデータカタログサイト [6] に公開されている「三田市指定避難所一覧」データセットに対して、提案されたシステム上で、自治体標準オープンデータセットである「03. 指定緊急避難場所一覧」に準拠しているかどうかを確認した。その後、その結果をもとに、「三田市指定避難所一覧」データセットを「03. 指定緊急避難場所一覧」データセットに準拠した形に修正する作業を実施した。

#### 4.2.2 Task2. 三田市オープンデータ一覧データセットの作成と形式チェック

このタスクでは、これから作成する新たなオープンデータセットを自治体標準オープンデータセットに準拠させたいという利用状況を想定した作業を行った。具体的には、現在三田市のオープンデータカタログサイトでは公開できていない、「三田市オープンデータ一覧」データセットを新たに作成し、そのデータセットに対して、提案されたシステム上で、自治体標準オープンデータセットである「06. オープンデータ一覧」に準拠

データチェック実行画面

作成データセット選択      データアップロード      データチェック実行

データチェックを実行し、修正点を確認してください

CSVをアップロード      ニードデータチェックを実行      コンソール出力リセット

| 行名 | ID | 名称                        | 名称_カナ         | 所在地_連絡先記                     | 住所             | 方面 | 緯度              | 経度         |
|----|----|---------------------------|---------------|------------------------------|----------------|----|-----------------|------------|
| 1  | 1  | 三田小学校                     | サンダシヨウガッコウ    | 兵庫県三田市豊新町2番20号               | 豊新町2番20号       |    | 34.8873475135   | 220.20051  |
| 2  | 2  | 八幡中学校                     | ハチケイチュウガッコウ   | 兵庫県三田市八幡町120番50地             | 八幡町120番50地     |    | 34.876268343135 | 225.959023 |
| 3  | 3  | さくら市民センター                 | サクラシミンセンタ     | 兵庫県三田市三田町2番14号               | 三田町2番14号       |    | 34.88268040135  | 225.929643 |
| 4  | 4  | 城山公園                      | シロヤマコウエン      | 兵庫県三田市三田131-1番地              | 三田131-1番地      |    | 34.8926079135   | 229.12122  |
| 5  | 5  | 三輪小学校                     | ミワノガッコウ       | 兵庫県三田市三輪1-12-11              | 三輪1-12-11      |    | 34.89089151135  | 229.9717   |
| 6  | 6  | 北谷小学校                     | キタヤノガッコウ      | 兵庫県三田市北谷町53番地                | 北谷町53番地        |    | 34.9037844135   | 221.1413   |
| 7  | 7  | 志手小学校                     | シテハラシヨウガッコウ   | 兵庫県三田市志手881番地                | 志手881番地        |    | 34.91612627135  | 229.1894   |
| 8  | 8  | 上野台中学校                    | ウエノダイチュウガッコウ  | 兵庫県三田市志手東1145番地              | 志手東1145番地      |    | 34.914033       | 135.236989 |
| 9  | 9  | 広野小学校                     | ヒロノガッコウ       | 兵庫県三田市上野沢29番地                | 上野沢29番地        |    | 34.92507783135  | 230.7795   |
| 10 | 10 | 広野市民センター                  | ヒロノシミンセンタ     | 兵庫県三田市上野沢29番地1               | 上野沢29番地1       |    | 34.92710714135  | 189.6909   |
| 11 | 11 | 母子山の総合館                   | モクシヤマノミナミセン   | 兵庫県三田市母子495番地                | 母子495番地        |    | 35.02207176135  | 232.2191   |
| 12 | 12 | 小野田公園                     | オノダコウエン       | 兵庫県三田市小野1740                 | 小野1740         |    | 34.95626572135  | 233.7834   |
| 13 | 13 | 有馬富士共生センター                | アリマフジキョウセイセンタ | 兵庫県三田市志手960番地                | 志手960番地        |    | 34.933465       | 135.230272 |
| 14 | 14 | 豊平小学校                     | タカヒラシヨウガッコウ   | 兵庫県三田市下豊172番地                | 下豊172番地        |    | 34.95067881     | 135.271354 |
| 15 | 15 | 高平ふるさと交流センターカワヒラファクトリージョセ | カワヒラファクトリージョセ | 兵庫県三田市高平200番地                | 高平200番地        |    | 34.9777091135   | 234.0302   |
| 16 | 16 | 都市民センター                   | アイシミンセンタ      | 兵庫県三田市大川1307番地4              | 大川1307番地44     |    | 34.94900469135  | 138.9598   |
| 17 | 17 | 阪中学校                      | アイシユウガッコウ     | 兵庫県三田市大川1307番地3              | 大川1307番地36     |    | 34.95052164135  | 137.7482   |
| 18 | 18 | つつがの小学校                   | ツツガノガッコウ      | 兵庫県三田市つつがの1番地7番つがの丘南7丁目19番地1 | つつがの丘南7丁目19番地1 |    | 34.9456677135   | 230.7277   |
| 19 | 19 | 阪小学校                      | アイシユウガッコウ     | 兵庫県三田市西郷477-1                | 西郷477-1        |    | 34.96191384135  | 143.2728   |
| 20 | 20 | 本志小学校                     | ホンシヨウガッコウ     | 兵庫県三田市東本志1810番地              | 東本志1810番地      |    | 34.96137778135  | 185.6074   |
| 21 | 21 | 豊南中学校                     | ナガカキユウガッコウ    | 兵庫県三田市豊南649番地                | 豊南649番地        |    | 34.944602179135 | 278.6544   |
| 22 | 22 | ふれあいと創造の園                 | フレアイトソウゾウエン   | 兵庫県三田市西41129番地1              | 西41129番地1      |    | 34.94609526135  | 167.9904   |
| 23 | 23 | 阪南小学校                     | ムコウガッコウ       | 兵庫県三田市東本志1810番地              | 東本志1810番地      |    | 34.95847398135  | 207.9108   |
| 24 | 24 | 狭間小学校                     | ハバヤマガッコウ      | 兵庫県三田市狭間が丘4丁目14番             | 狭間が丘4丁目14番     |    | 34.890269       | 135.221214 |
| 25 | 25 | 狭間中学校                     | ハバヤマユウガッコウ    | 兵庫県三田市狭間が丘4丁目1番              | 狭間が丘4丁目1番      |    | 34.87993498135  | 202.0298   |
| 26 | 26 | 富士小学校                     | フジシヨウガッコウ     | 兵庫県三田市富士が丘1丁目124             | 富士が丘1丁目124     |    | 34.88967778     | 135.194309 |
| 27 | 27 | 富士中学校                     | フジチュウガッコウ     | 兵庫県三田市富士が丘3丁目724             | 富士が丘3丁目724     |    | 34.88869414135  | 189.1226   |

コンソール

データセット情報

データチェック関数情報

Warning

["メッセージ":"データセットの定義に従っていない(列名/市町村名)がカラムに含まれてい  
ます。列名の修正をお願いします"]

Warning

["メッセージ":"必須でない列(全国地方公共団体  
コード)がカラムに含まれていないためチェックをス  
キップしました"]

Success

列「ID」に 関数「半角英数字が埋める関  
数」を適用しました

Failed

列「ID」に 関数「値の桁数を確かめる関数」  
を適用しました

Success

列「ID」に 関数「列内で値が重複していない  
ことを確かめる関数」を適用しました

Failed

列「名称\_カナ」に 関数「全角カタカナが埋め  
られる関数」を適用しました

図 2 データチェック出力結果例

列「ID」に 関数「値の桁数を確かめる関数」  
を適用しました

```
{
  "関数詳細": "各値の桁数が 12 桁に一致するかを確認しま  
す。",
  "チェック結果": "一致しない値が検出されました"  
一致しない値とその行番号: {
    1: "1",
    2: "2",
    3: "3",
    4: "4",
    5: "5",
    6: "6",
    7: "7",
    8: "8",
    9: "9",
    10: "10",
    11: "11",
    12: "12",
    13: "13",
    14: "14",
    15: "15",
    16: "16"
  }
}
```

図 3 データチェック内容詳細例

しているかどうかを確認した。その後、その結果をもとに、「三  
田市オープンデータ一覧」データセットを「06. オープンデータ  
一覧」データセットに準拠した形に修正する作業を実施した。

#### 4.2.3 RQ の検証方法

RQ1 を確かめるための方法として、各タスクによって作成さ  
れたデータセットに対して、自治体標準オープンデータセット  
の項目定義書を参照しながら、手動で検査を行い、誤っていた  
箇所をカウントすることによって、データ全体のどの程度の割  
合が標準に準拠しているかどうかを判定する方法を利用する。  
また、RQ2 を確かめるための方法として、タスクを実施しても  
らった後に表 1 に示すアンケート項目に回答してもらった。そ  
れぞれのアンケート項目は、ソフトウェアの品質に関する国際  
規格である SQuaRE [7] の利用時の品質を考慮して作成した。  
各アンケート項目は 4 段階（1. そう思う、2. どちらかといえ  
ばそう思う、3. どちらかといえばそう思わない、4. そう思わな  
い）で評価してもらい、その評価をつけた理由についてもあわ  
せて質問した。また、Q6 に関しては、システムの評価点及び  
改善点について自由記述で回答してもらった。

#### 4.3 実験の手順

実験は、兵庫県三田市の三田市役所にて、オープンデータ公  
開に関わる職員 3 名に協力していただき、実施した。

実験に協力していただいた職員のスキルレベルについては、

参加者 A は、オープンデータ公開に関わる業務の経験はない  
が、Excel のスキルは高く、自治体標準オープンデータセット  
について十分理解し、定義に従ったデータセットを作成できる  
というレベルである。参加者 B は、オープンデータ公開に関わ  
る業務を 2 年間行っており、調べながら自力で Excel の操作を  
行うことができる。また、自治体標準オープンデータセットに  
ついては、概要は理解しているというレベルである。参加者 C  
は、オープンデータ公開に関わる業務を行っており、Excel の  
操作にはまだ不慣れであるが、他人の指導があれば作業が行え  
るレベルである。また、自治体標準オープンデータセットにつ  
いては、概要は理解しているというレベルである。

実験は事前準備、タスク 1、タスク 2 の順番で行った。以下に  
各プロセスにて行ったことを詳しく述べる。

##### 4.3.1 事前準備

ここでは、タスク 1、タスク 2 の形式チェックに利用するため  
の情報を自治体標準オープンデータセットのデータ項目定義書  
から抽出し、データチェック関数として python の関数として  
実装、データセットの他の情報とともにシステムに登録した。

まず、その定義書の情報からチェックすべき形式を抜き出し、  
データチェック関数として実装した。その後、システムの「デー  
タチェック関数登録機能」を利用してシステムに登録した。

次に、「スキーマ登録機能」を利用して、03. 指定緊急避難場

表 1 タスク実施後のアンケート項目

| No. | Sentence                                      |
|-----|-----------------------------------------------|
| Q1  | 実際にオープンデータを作成する際にこのシステムは有<br>用だと思いますか？        |
| Q2  | このシステムを利用することで、データチェックに必要<br>な人員は削減できると思いますか？ |
| Q3  | このシステムを利用することで、データチェックにかか<br>る時間は減ると思いますか？    |
| Q4  | 一般的な自治体職員がこのシステムを利用してデータ<br>チェックを行えると思いますか？   |
| Q5  | システムは想定した通りに機能しましたか？                          |
| Q6  | 良かった点・ご意見・改善点等があれば記述してくださ<br>い                |

所一覧データセット, 06. オープンデータ一覧データセットに関するデータスキーマをシステムに登録した。これによって、各データセットに必要な列名の情報や、各列に対して行うチェックの情報がシステムに登録され、自動で標準形式に準拠しているかどうか検査できるようになった。

#### 4.4 タスクの作業手順

タスク 1, タスク 2 では共通して、以下のステップに従って参加者に作業を行ってもらった。

Step1. チェック元データを用意する

Step2. システムにアクセスする

Step3. データチェック用スキーマを選択する

Step4. システムにチェック元データをアップロードする

Step5. システムでデータチェックを実行する

Step6. チェック結果をもとにデータを修正する

以下で、各ステップの詳細について述べる。

Step1 では、これから提案システムでチェックを実行する対象となるファイルを準備した。タスク 1 では、三田市オープンデータカタログサイトにて公開されている、三田市指定避難所一覧というデータセットをサイトからダウンロードした。タスク 2 では、三田市が保有する未公開のオープンデータ一覧データを用意していただき、作成中のオープンデータとしてチェックを実行する対象とした。

Step2 では、提案システムの URL を Web ブラウザ上に入力し、その後システムへのログイン処理を行った。

Step3 では、提案システム上で「データチェック」機能にアクセスし、データチェックのために利用するデータを選択した。

Step4 では、提案システム上でチェックを実行する対象のファイルをアップロードした。

Step5 では、提案システム上でデータチェックを実行し、出力結果の確認を行った。出力結果画面の例は図 2 のようになっている。画面の左側にはチェック対象のデータが表示され、チェックの結果、修正が必要と判定されたデータは背景が赤く表示される。画面の右側には、チェックの内容の詳細が出力され、どの列に対して、どの関数が適用され、結果はどのようなであったのかを確認することができる。チェック内容の詳細の出力例は図 3 のようになっている。

Step6 では、Step5 で検出された修正が必要なデータに対して Excel 上で修正を行った。Step5 で確認したチェック内容の情報や、データ項目定義書の情報をもとに、標準オープンデータセットの形式に従うようにデータを修正した。

#### 4.5 実験結果

表 2, 表 3 にタスク 1, タスク 2 における、元データに存在した、標準に準拠せず修正が必要だった列名やデータの値の形式の数と、参加者が実際に修正を行った数を示す。さらに、表 4 に実験後に行ったアンケートの結果を示す。1 が「そう思う」、2 が「どちらかといえばそう思う」、3 が「どちらかといえ

表 2 Task1 における修正項目

| 修正項目     | All | A | B | C |
|----------|-----|---|---|---|
| 必須項目追加   | 2   | 2 | 2 | 2 |
| 必須項目形式修正 | 2   | 2 | 1 | 2 |
| 推奨項目名修正  | 4   | 4 | 4 | 4 |
| 推奨項目形式修正 | 3   | 3 | 3 | 3 |

表 3 Task2 における修正項目

| 修正項目     | All | A | B  | C |
|----------|-----|---|----|---|
| 必須項目追加   | 7   | 7 | 7  | 7 |
| 必須項目形式修正 | 2   | 1 | 1  | 1 |
| 推奨項目名修正  | 10  | 1 | 10 | 0 |
| 推奨項目形式修正 | 0   | 0 | 0  | 0 |

表 4 実験後のアンケート結果

| No. | 被験者 A | 被験者 B | 被験者 C |
|-----|-------|-------|-------|
| Q1  | 2     | 1     | 1     |
| Q2  | 2     | 1     | 1     |
| Q3  | 3     | 2     | 2     |
| Q4  | 2     | 1     | 2     |
| Q5  | 2     | 2     | 1     |

う思わない」、4 が「そう思わない」を表す。

## 5. 考 察

### 5.1 リサーチクエスションの検証

4.5 で得られた結果から、4.1 で述べたリサーチクエスションを検証する。

#### 5.1.1 RQ1: 自治体職員が提案システムを利用して、標準に従ったデータセットを作成できるか

表 2, 表 3 の各タスクにおける参加者が修正を行えたデータの数をみると、タスク 1 では全ての参加者が、必須データ項目を追加し、そのデータ形式を修正することができた。また、必須でないデータ項目についても、元データの列名を変更し、データ形式を修正することができた。また、タスク 2 では、全ての参加者が、必須データ項目を追加し、そのデータ形式を修正することができた。しかし、全参加者において、「ファイル形式」という項目の修正を行うことができなかった。この原因として考えられるのは、複数のファイル形式でデータを公開している際は、データを別々のセルで作成するというルールをシステム上でうまく伝えることができなかったということである。また、必須でないデータ項目については、参加者 B を除いて、元データの列名を正しく変更することができなかった。この原因として考えられるのは、システムが提示する情報が入力データの中で定義書に定義されていない列を明示するといったもののみであったため、参加者がシステムが提示する情報を理解し、データをどのように修正すればよいかわからなかったということである。

これらの結果から、自治体標準オープンデータセットに定められる最低限のデータ項目については、提案システムを利用することで、自治体職員が標準に従ったデータセットを作成する

ことができると言える。また、必須でないデータ項目の修正については、参加者のシステムの理解度によって、修正ができるかどうか異なることがわかった。

### 5.1.2 RQ2: 提案システムは実際のオープンデータ公開作業にどの程度受け入れられるか

表4のアンケートにおいて、Q1,Q2では、2名の参加者が「そう思う」、1名の参加者が「どちらかといえばそう思う」と回答し、その理由について、「エラー部を示してくれて編集が容易と感じたため」「目視でのデータ形式チェックは実際困難であるから」などの意見が得られた。

その一方で、Q3では、2名の参加者が「どちらかといえばそう思う」、1名の参加者が「どちらかといえばそう思わない」と回答した。その理由として、「本市の規模ではチェックを複数人で行っていないが、大きな自治体であれば人員の削減につながる可能性はある」などの意見が得られた。

Q4では、1名の参加者が「そう思う」、2名の参加者が「どちらかといえばそう思う」と回答した。その理由として、「データチェックは簡単に行えると思う。チェック後のデータ修正は一定のスキルが必要となるが、ある程度トレーニングすれば対応できると思う」や「操作は難しくないので、標準オープンデータセットや定義を理解できれば、操作できると感じた」などの意見が得られた。

これらの結果から、提案システムは、実際のオープンデータ公開作業において、容易に操作することができ、実際に、データ形式のチェックを行うことができるということがわかった。また、提案システムによるデータチェック後のデータ修正については、一定のスキルが必要となるが、トレーニングすれば対応できるということがわかった。

### 5.2 OpenDataValidator の利点

今回の実験で得られた意見をもとに、OpenDataValidator の利点について考察する。

アンケートを通して、「簡単にデータチェックができ、修正すべきセルが赤色で表示されるのが良かった」や「操作としては難しくないので」という意見が得られた。これらの意見から、OpenDataValidator は、容易に操作することができ、専門知識がない自治体職員でもデータ形式のチェックを行うことができるという利点があると言える。

また、表2、表3の修正を行えたデータの数から、OpenDataValidator は、おおむね高い精度で自治体職員によるデータ修正を支援することができるという利点があると言える。

### 5.3 OpenDataValidator の課題

今回の実験で得られた意見をもとに、OpenDataValidator の課題について考察する。

アンケートを通して、「データチェックは簡単に行えるが、データ修正は一定のスキルが必要となる」という意見が得られた。これらの意見から、より容易にデータ修正項目を理解することができるような情報提示が必要であると言える。

また、「実運用するためには、任意修正やデータが入っていないセルに対するエラー表示のあり方など、UI の観点で改善すべ

きとこころは残る」や「エラー表示をコンソールでクリック時にデータセルに移動するなどの機能があれば良い」、「必須列の検査は一度にまとめてしてほしい、何度もファイルの取り込みを行うのは手間に感じる」という意見が得られた。これらの意見から、より使いやすいUIを実現するために、データチェックの手順を明確に提示し、ユーザがシステムを用いて作業をより明確化することや、修正項目の情報により簡単にアクセスできるような機能が必要であると言える。

## 6. ま と め

本稿では、自治体職員による標準化されたオープンデータの公開を支援するためのシステムである、OpenDataValidator の評価実験を行った。兵庫県三田市の職員の協力のもと、実際のオープンデータ公開作業において、実装したシステムを用いることで、標準化されたオープンデータを作成することができるかを検証した。その結果、システムを用いることで、容易にデータの修正箇所を特定し、標準形式に従った形に修正を行うことができることが確認された。

今後の展望として、システムが提示するデータ修正のための情報を充実させ、より容易にデータ修正作業を行うことができるようにすることが挙げられる。また、オープンデータの作成作業を支援することができれば、オープンデータ公開プロセス全体を支援することができ、高品質なオープンデータの公開をより容易化できると考えられる。

謝辞 本研究の一部は JSPS 科研費 JP19H01138, JP20H05706, JP20H04014, JP20K11059, JP22H03699, JP19K02973, 若手研究 23K17006 の助成を受けて行われている。

本実験に多大なるご協力を賜りました、岩崎謙二様、吉川肇様、和田光平様をはじめとする兵庫県三田市の職員の皆様に心より感謝申し上げます。

## 文 献

- [1] 大向一輝, “日本におけるオープンデータの進展と展望,” 情報管理, vol.56, no.7, pp.440–447, 2013.
- [2] “オープンデータ基本指針 | 政府 cio ポータル,” <https://cio.go.jp/node/2357>. (Accessed on 2 February 2024).
- [3] “5つ星オープンデータ,” <https://5stardata.info/ja/>. (Accessed on 2 February 2024).
- [4] “自治体標準オープンデータセット (正式版) | デジタル庁,” [https://www.digital.go.jp/resources/open\\_data/municipal-standard-data-set-test/](https://www.digital.go.jp/resources/open_data/municipal-standard-data-set-test/). (Accessed on 2 February 2024).
- [5] 渡邊 輔, 陳 思楠, 佐伯幸郎, 中村匡秀, “地方自治体のオープンデータ公開プロセスを支援するアプリケーションの検討,” 電子情報通信学会技術研究報告, 第123巻, pp.019–025, Aug. 2023. 東京, 機械振興会館.
- [6] “三田市オープンデータカタログサイト,” <https://odcs.bodik.jp/282197/>. (Accessed on 2 February 2024).
- [7] “Iso/iec 25000 square series systems and software engineering — systems and software quality requirements and evaluation (square),” <https://committee.iso.org/sites/jtc1sc7/home/projects/flagship-standards/iso-25000-square-series.html>. (Accessed on 2 February 2024).