



# オープン&ビッグデータ活用・地方創生推進機構

平成26年度第3回 技術委員会

## 評価版ツールの状況報告

2015.03.03

オープン&ビッグデータ活用・地方創生推進機構 事務局

# 評価版ツールの位置づけ

## ■ 今年度の検討項目の中での位置づけ

1. オープンデータガイドの精査 **第2回委員会にて報告**

2. 情報流通連携基盤 外部仕様書の精査

3. 評価版ツールの作成

◇情報流通連携基盤システム 外部仕様書の参照実装パッケージ

◇ボキャブラリ管理サイト

◇「オープンデータガイド」9.3節の「技術的指針」に関するチェックツール

◇メタデータ抽出支援ツール

**第3回委員会にて報告**

4. 次年度実施内容の検討

## ■ 今回報告するツールの扱い

▶ ツール自体は、オープンソースで公開予定。

▶ 次年度の検討テーマ案の1つである「ツール集の検討」に先立ち、昨年度の技術委員会や実証の中で要望されたツールを試作したもの。



# 1. 表形式データへの指針に 対するチェックツール

# 「表形式データへの指針に対するチェックツール」概要

## ■ 提供する機能

- ▶ 「オープンデータガイド」第9章に掲載されている、表形式データに対する以下の指針を満たしているか否かをチェックするツール
  - ◇人間による判断を支援するツール

グレード	指針	
1	指針1	1つのファイルは、1種類の表から構成されるべきである。
	指針2	ヘッダは、1行から構成されるべきである。
2	指針3	データでない情報を、レコードに含めないことが望ましい。
	指針4	全てのフィールドは、他のフィールドと結合されないことが望ましい。
	指針5	値がない場合を除き、フィールドを空白にしない（省略しない）ことが望ましい。
	指針6	年の値には、西暦表記を備えることが望ましい。
	指針8	利用している文字コードを明記することが望ましい。また、国際的に広く利用されている文字コードを利用することが望ましい。

# ツールによるチェック結果(1)

## ■ 指針1を満たさない場合

ファイルX

月	A市	B市	C市	D町
1	-4.5	-0.5	1.6	11.3
2	-6.8	-2.1	0.4	8.4
3	-2.4	1.9	3.8	13.5
4	0.2	3.4	6.5	17.3

月	A市	B市	C市	D町
1	230	58	377	103
2	169	43	422	122
3	144	54	322	144
4	232	102	145	133



### 表形式データに関する指針 チェックツール

gl1\_invalid.xlsx  
6.8 KIB

✓

出力結果の言語 日本語

デリミタ コンマ

チェック

結果  
(gl1\_invalid.xlsx)

⚠ 下記をご確認ください。

指針	関連セル番号
指針 1: 一つのファイルは、1種類の表から構成されるべきである。	A9, B9, C9, D9, E9
指針 3: データのない情報を、フィールドに含めないことが望ましい。	A9, B9, C9, D9, E9
指針 7: フィールドの単位が明記されていることが望ましい。	A3, B3, C3, D3, E3

Copyright © YRP Ubiquitous Networking Laboratory. All Rights Reserved.

# ツールによるチェック結果(2)

## ■ 指針2を満たさない場合

月	気温			
	A市	B市	C市	D町
1	-4.5	-0.5	1.6	11.3
2	-6.8	-2.1	0.4	8.4
3	-2.4	1.9	3.8	13.5
4	0.2	3.4	6.5	17.3



### 表形式データに関する指針 チェックツール

gl2\_invalid.xlsx ✔  
6.7 KiB

出力結果の言語 日本語

デリミタ コンマ

チェック

結果  
(gl2\_invalid.xlsx)

⚠ 下記をご確認ください。

指針	関連セル番号
指針 2: ヘッダは、1 行から構成されるべきである。	C2, B3
指針 5: 値がない場合を除き、フィールドを空白にしない(省略しない)ことが望ましい。	C2, D2, E2, A3
指針 7: フィールドの単位が明記されていることが望ましい。	A3, B3, C3, D3, E3

Copyright © YRP Ubiquitous Networking Laboratory. All Rights Reserved.

## ツールによるチェック結果(3)

### ■ 指針1～8を満たす場合

月	A市の気温 [°C]	B市の気温 [°C]	C市の気温 [°C]	D町の気温 [°C]
1	-4.5	-0.5	1.6	11.3
2	-6.8	-2.1	0.4	8.4
3	-2.4	1.9	3.8	13.5
4	0.2	3.4	6.5	17.3



衣形式データに関する指針 チェックツール

gl8\_valid.csv ✓  
0.1 KiB

出力結果の言語

デリミタ

結果  
(gl8\_valid.csv)  
✓ このファイルは指針1～8に準拠しています

Copyright © YRP Ubiquitous Networking Laboratory. All Rights Reserved.



## 2. メタデータ抽出ツール



# 「メタデータ抽出ツール」概要

## ■ 主要な機能

- ▶ webに公開されているオープンデータをデータカタログに登録する際に、公開されているオープンデータのファイルからメタデータを自動抽出するツール

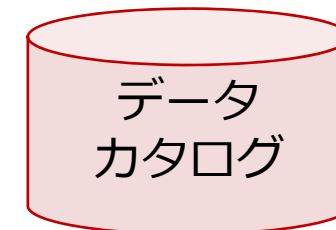
ファイル名	作成日	ファイルサイズ	説明
xx.csv	2015-02-15	12,345	xxに関する統計データ
xx.xlsx	2015-02-15	88,776	xxに関する統計データ
yy.csv	2015-02-17	23,456	yyに関する統計データ
yy.xlsx	2015-02-17	99,887	yyに関する統計データ

(2) メタデータの作成

(1) メタデータを抽出

web上のオープンデータ  
 xxに関する統計データ  
 ([CSV形式](#))([EXCEL形式](#))  
 yyに関する統計データ  
 ([CSV形式](#))([EXCEL形式](#))  
 ...

(3) データの登録



# 「メタデータ抽出ツール」概要

---

## ■ 抽出できるファイル形式

- ▶ Word
- ▶ Excel
- ▶ PowerPoint
- ▶ PDF
- ▶ HTML

## ■ 抽出できるメタデータ

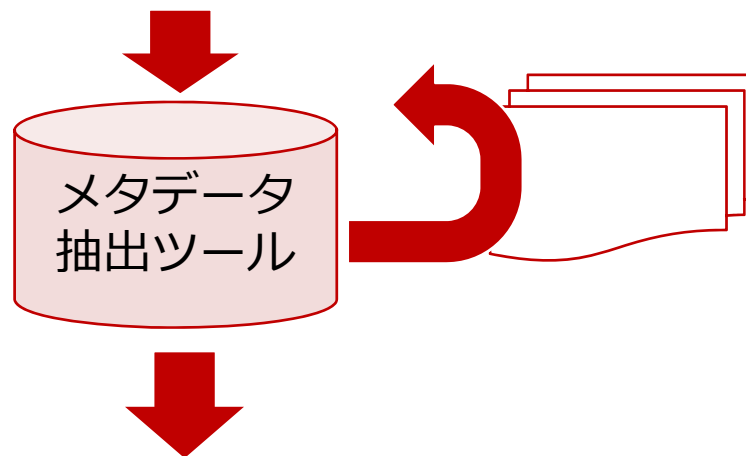
- ▶ ファイル名
- ▶ ファイルサイズ (バイト単位)
- ▶ ファイルタイプ
- ▶ 作成日時
- ▶ 最終更新日時
- ▶ キーワード
- ▶ タイトル
- ▶ 作成者
- ▶ ファイルの説明 (HTMLであれば“meta description”タブに記載された文字列)
- ▶ ファイル中にある「指定された用語」

# 「メタデータ抽出ツール」概要

## ■ 動作概要

入力 (Excel+ 設定ファイル+ Formによる入力)

K	L	M	N	O	P
(リソース)タイトル	URL	(リソース)説明	データ形式	ファイルサイズ	最終更新日
title	url	description	format	size	last_modified_date
自由記述	自由記述	自由記述	択一選択	自由記述 (半角数字)	YYYY-MM-DD形式 (半角)



抽出対象ファイル群

出力 (Excel)

K	L	M	N	O	P
(リソース)タイトル	URL	(リソース)説明	データ形式	ファイルサイズ	最終更新日
title	url	description	format	size	last_modified_date
自由記述	自由記述	自由記述	択一選択	自由記述 (半角数字)	YYYY-MM-DD形式 (半角)
高度情報通信ネットワーク社会の形成に関する平成27年度概算要求について (HTML)	http://www.kantei.go.jp/jp/singi/it2/ot_hers/yosan/27siryou_1b.html	高度情報通信ネットワーク社会の形成に関する平成27年度概算要求について [HTML・PDF]	html	13573	
高度情報通信ネットワーク社会の形成に関する平成27年度概算要求について	http://www.kantei.go.jp/jp/singi/it2/ot_hers/yosan/27siryou_1b.pdf	高度情報通信ネットワーク社会の形成に関する平成27年度概算要求について	pdf	90310	2014-10-30

# 画面遷移

メタデータ抽出ツール

抽出対象のURLが記載された、登録用Excelファイルを指定してください。

プロパティの設定

登録用Excelファイルをドロップしてください。

Copyright © YRP Ubiquitous Networking Laboratory. All Rights Reserved.

## (1) 入力Excelファイルを指定

メタデータ抽出ツール

抽出するファイルを選択してください。

戻る 次へ

すべて ON すべて OFF

リンク先: <http://www.kantei.go.jp/>  
タイトル: トップ

リンク先: <http://jp.singi/index.html>  
タイトル: 会議等一覧

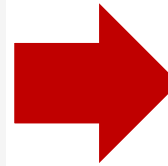
リンク先: <http://jp.singi/it2/index.html>  
タイトル: 高度情報通信ネットワーク社会推進戦略本部（IT総合戦略本部）

高度情報通信ネットワーク社会の形成に関する平成27年度概算要求について [HTML・PDF]  
リンク先: [http://jp.singi/it2/others/yosan/27siryou\\_1b.html](http://jp.singi/it2/others/yosan/27siryou_1b.html)  
タイトル: 高度情報通信ネットワーク社会の形成に関する平成27年度概算要求について(HTML)

高度情報通信ネットワーク社会の形成に関する平成27年度概算要求について [HTML・PDF]  
リンク先: [http://jp.singi/it2/others/yosan/27siryou\\_1b.pdf](http://jp.singi/it2/others/yosan/27siryou_1b.pdf)  
タイトル: 高度情報通信ネットワーク社会の形成に関する平成27年度概算要求について(PDF)

高度情報通信ネットワーク社会の形成に関する平成26年度予算について（府省庁別） [HTML・PDF]  
リンク先: <http://jp.singi/it2/other/yosan/26yosan.html>  
タイトル: 高度情報通信ネットワーク社会の形成に関する平成26年度予算について(府省庁別)

## (3) 抽出対象のファイルを指定



メタデータ抽出ツール

固定値と抽出元URLを選択してください。  
登録用Excelファイル: test.xlsx

戻る 次へ

抽出元URL: その他  
<http://www.kantei.go.jp/jp/singi/it2/others/yosan.html>

状態:

(データセット)タイトル:

(データセット)説明:

公表組織名: 内閣官房

連絡先:

作成者:

リリース日:

公開ウェブページ:

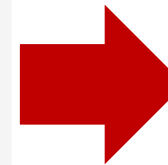
(リソース)タイトル:

URL:

(リソース)説明:

データ形式:

## (2) 固定のメタデータを設定



メタデータ抽出ツール

抽出処理

戻る

抽出完了しました。

処理されたファイル: test20150226\_13\_34\_44\_891.xlsx

Copyright © YRP Ubiquitous Networking Laboratory. All Rights Reserved.

## (4) 抽出完了

