

イベントレポート

目次

- 1 イベント概要
- 2 イベント目的
- 3 イベント内容
- 4 アンケート結果
- 5 成果
- 6 課題
- 7 課題

1 イベント概要

イベント名

Social Hack Day #56

開催日時

2023年12月16日（土）10:45～18：00

対象者：

- ・ オープンデータに興味がある人
- ・ 自治体職員
- ・ エンジニア

2 イベント目的

2022年度に報告した「[DataOpsを活用したオープンデータ利活用に関する報告書](#)」におけるオープンデータ利活用プロセスやツール等の有効性の確認を目的として本イベントを開催した。

3 イベント内容

今回のイベントでは、データ利用者が緊急避妊検索プロジェクトで使用している診療所データの可視化に活用することを前提としている。

データ利用者の観点で、データを有効に活用するにはどのような構成とすべきかをデータ提供者のプロセスである「データの収集蓄積」を通して議論し、このプロセスの有効性を確認する。

データの可視化に際し、以下の作業を実施した。

- ・ 最新の診療所データをPDFからExcel(xlsx)に変換
- ・ データ可視化の為、変換したデータをスプレッドシート上で修正
- ・ ChatGPTの有用性確認の為、ChatGPTによる修正したデータのレビュー及びデータの可視化

イベント内で使用したデータ、ツールについては以下に記載する。

使用したデータ

緊急避妊に係る取組について | 厚生労働省

※ご協力いただいたプロジェクト：[アフターピル検索](#)

使用したツール

- [PDFをExcelに変換](#)
PDFのデータをExcel(xlsx)に変換する為のツール
- [PDF Excel 変換](#)
PDFのデータをExcel(xlsx)に変換する為のツール(こちらの方が効率よく変換することが可能だった)
- [地方公共団体コード住所\(J-LIS\)](#)
郵便番号や住所から地方公共団体コードを確認する為のツール
- [BODIK ジオコーダー](#)
住所から緯度・経度の情報を取得するツール

タイムテーブル

11:00 本日の活動内容共有
11:30 プロジェクト開始、チーム内自己紹介
11:50 プロジェクト（DataOps）概要説明
12:00 当日の作業内容共有
12:05 質疑応答
12:10 休憩
13:10 再開
13:15 データ提供者の方からのコメント
13:30 CKAN拡張機能を使用した際の効果説明（デモを実施）
13:35 PDF→Excel変換作業開始
14:50 生成系AIレビュー体験
15:00 プロジェクト中間発表
15:15 写真撮影
15:20 シャッフル（概要説明、作業内容共有、質疑応答）
15:50 カルチャーの分野のオープンデータ
16:00 再シャッフル（意見交換、今後の作業方針）
16:30 ChatGPTを使用したデータの可視化
16:55 アンケート
17:00 最終成果報告
17:15 報告会終了

CKAN拡張機能のイメージ

CKAN拡張機能はDataOpsプロセスの課題抽出・計画・データの公開・取得・活用・フィードバックの要素を満たす機能を有している。
この機能の内、フィードバックのデモ実施イメージは以下。

ログイン保存

ckan

Topデータセット組織グループ利活用方法About

検索

組織 / C3Lab / 緊急避妊に係る診療が可能な産婦人科医療機関等一覧

緊急避妊に係る診療が可能な産婦人科医療機関等一覧(青森県).pdf / コメント

データセット

medical-institution

データリソース

緊急避妊に係る診療が可能な産婦人科医療機関等一覧(青森県).pdf

コメント数

1

コメント

評価☆☆☆☆☆

要望

送信

2023/12/15 17:06 要望☆☆☆☆☆

PDFでは少し使いにくいです。
excel形式かcsv形式で公開して頂けると嬉しいです。

生成系AIレビューのイメージ

ChatGPTを使用して、変換したデータのレビューを実施。
事前にオープンデータの項目定義を学習させ、その定義にもとづいたフィードバックを生成。
生成系AIレビュー体験の実施イメージは以下。

オープンデータ項目定義書の確認

2023年12月16日

 匿名



20210303_resources_dataset_doc_02....
スプレッドシート

添付した資料はデジタル庁が提示しているオープンデータの項目定義書です。
開いて中身を確認し、理解できたら[はい]と教えてください。



ChatGPT

ファイルを開いて確認しました。これはデジタル庁が提示しているオープンデータの項目定義書で、複数のシートが含まれています。シートの内容は以下の通りです：

1. 表紙
2. 改訂履歴
3. データ項目定義書について
4. AED設置箇所一覧
5. 介護サービス事業所一覧
6. 医療機関一覧
7. 文化財一覧
8. 観光施設一覧
9. イベント一覧
10. 公衆無線LANアクセスポイント一覧
11. 公衆トイレ一覧
12. 消防水利施設一覧
13. 指定緊急避難場所一覧
14. 地域・年齢別人口
15. 公共施設一覧
16. 子育て施設一覧
17. オープンデータ一覧
18. 食品等営業許可・届出一覧
19. 学校給食献立情報メタデータ
20. 学校給食献立情報
21. 小中学校通学区域情報メタデータ
22. 小中学校通学区域情報
23. ファイル名命名規則
24. データ項目特記事項

はい、理解しました。👍

匿名

export.csv
スプレッドシート

[オープンデータの項目定義書]に基づいて、
添付の[export.csv]に対してフィードバックを生成してください。

ChatGPT

添付された「export.csv」ファイルを確認しました。このCSVファイルには以下のカラムが含まれています：

1. 都道府県
2. 全国地方公共団体コード(6桁)
3. 施設名
4. 郵便番号
5. 住所
6. 電話番号
7. ウェブサイトURL
8. 産科、婦人科又は産婦人科の標榜の有無
9. 医療機関における緊急避妊に係る対面診療への対応可能時間帯
10. 常時の緊急避妊薬の在庫の有無
11. 緯度
12. 経度

このデータは、主に医療機関の情報を含むもので、特に産科や婦人科のサービス、緊急避妊薬の在庫状況などに焦点を当てています。ファイルの最初の数行を見ると、滋賀県の医療機関の情報が含まれていることがわかります。

フィードバックを生成するために、データの整合性、完全性、そしてオープンデータの項目定義書に基づいた適切なフォーマットへの準拠を検討します。以下の点に注意してデータを確認することをお勧めします：

1. **データの完全性**：いくつかのカラム（例えば「全国地方公共団体コード(6桁)」）で欠損値が見られます。これらの値が必須であれば、欠損しているデータの補完が必要です。
2. **データの一貫性**：各カラムのデータが一貫したフォーマットで入力されているかを確認します。例えば、電話番号や郵便番号のフォーマットが一定であるか、ウェブサイトURLが適切な形式であるかなどです。
3. **項目定義の準拠**：オープンデータの項目定義書に基づいて、必要な項目が全て含まれているか、またそれらが適切に定義されているかを確認します。
4. **地理的データの精度**：緯度と経度のデータが正確かどうかを検証します。これは特に地図上での施設の位置特定に重要です。

これらの点に基づいて、データをさらに分析し、改善するための具体的なフィードバックを生成できます。 [↗]



匿名

具体的なフィードバックを生成していただきたいです。



ChatGPT

具体的なフィードバックを生成するために、添付された「export.csv」ファイルのデータを詳細に分析し、以下の点に注目して評価しました：

1. 全国地方公共団体コードの欠損：

- ・ 問題：サンプルデータでは「全国地方公共団体コード(6桁)」の列にデータが欠けています。
- ・ フィードバック：このコードは各施設を一意的に識別するため重要です。欠けているデータを追加するか、この欠損が意図的であればその理由を文書化することを推奨します。

2. 電話番号のフォーマット：

- ・ 問題：電話番号のフォーマットが一定ではありません。
- ・ フィードバック：電話番号を標準化したフォーマット（例：国コードの有無、ハイフンの使用）で記載することを推奨します。

3. ウェブサイトURLのフォーマット：

- ・ 問題：いくつかのURLが不完全または誤っている可能性があります。
- ・ フィードバック：すべてのURLが正しいかどうかを確認し、必要に応じて修正します。また、URLは"http://"または"https://"で始まることを確認してください。

4. 緯度と経度の精度：

- ・ 問題：緯度と経度の値が正確であるかどうかは確認が必要です。
- ・ フィードバック：地図上で緯度と経度の値を確認し、施設の正確な位置を反映しているかどうかを検証します。

5. データの一貫性：

- ・ 問題：いくつかの項目で情報の記載方法が一貫していない可能性があります。
- ・ フィードバック：例えば、「緊急避妊薬の在庫の有無」の列では、在庫情報の記述方法を統一することを提案します。また、情報の一貫性を高めるために、特定のカテゴリーのデータ入力規則を設定すると良いでしょう。

これらのフィードバックは、データの品質を向上させ、利用者が情報をより簡単に理解し活用できるようにするために重要です。また、データの整合性と信頼性を確保するためにも必要です。

ChatGPTを使用したデータの可視化イメージ

ChatGPTを使用してデータの可視化を実施。

変換したデータをChatGPTに読み込ませ、可視化された地図データを生成。

また幅の調整や表示項目の追加指示への対応についても確認。

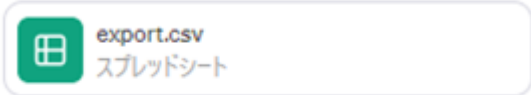
ChatGPTを使用したデータの可視化で実施した会話ログのイメージは以下。

地図上で可視化

2023年12月16日

① このチャットには、Advanced Data Analysisによって生成されたファイルや画像が含まれていますが、まだ共有チャットで表示されていません。

匿名



この「export.csv」を地図上に可視化してください。

ChatGPT

このCSVファイルには、「都道府県」、「施設名」、「住所」、そして各施設の「緯度」と「経度」の情報が含まれています。これらのデータを用いて、地図上に施設の位置を可視化することができます。次に、地図上にこれらの施設をプロットしてみましょう。[↗]

地図上に施設の位置を可視化しました。以下のリンクから地図をダウンロードし、ブラウザで表示することができます。

地図のダウンロード

この地図では、各施設がマーカーで表示されており、マーカーをクリックすることで施設名が表示されます。[↗]

匿名

popupの横幅を大きくしてください

ChatGPT

ポップアップの横幅を拡大し、地図を更新しました。以下のリンクから改良された地図をダウンロードして、ブラウザで表示することができます。

改良された地図のダウンロード

この更新された地図では、各施設のポップアップがより広いスペースで表示されます。[↗]

匿名

ついでに電話番号もpopupに出してほしいです

ChatGPT

ポップアップに施設名と電話番号を表示するように地図を更新しました。以下のリンクからこの新しい地図をダウンロードして、ブラウザで表示できます。

電話番号付き地図のダウンロード

この地図では、各マーカーのポップアップに施設名とその電話番号が表示されます。[↗]

DataOpsのプロセスがオープンデータ利活用促進につながるかの根拠を得る為、イベント参加者にアンケートを実施した。

5 成果

成果	関連するDataOpsプロセス							
	課題抽出	計画	データの収集蓄積	データの公開	データの取得	データの処理	データの活用	フィードバック
国の制度関連の取り組みを行っているプロジェクトでDataOpsの概念を組み込んでいきたいという意見が上がった								
自治体がオープンデータを提供するプロセスを体験することで、作業量等の観点から自治体職員が現状の業務に加えて、手作業でデータ加工を行うことは負荷の観点で困難であると感じた			○					

成果	関連するDataOpsプロセス						
	課題抽出	計画	データの収集蓄積	データの公開	データの取得	データの処理	データの活用
オープンデータ提供者にとって生成系AIによるレビューが有効な手段であることを実感できた			○				
CKAN拡張機能（ckanext-feedback）はフィードバックをトリガーとしてDataOpsプロセスの課題抽出・計画・データの公開・取得・活用の要素を満たす機能を有している為、それをイベント参加者が実際に体験することでDataOpsプロセスにおける善循環サイクルへの理解を促進できた	○	○		○	○		○
博物館では所蔵品のデータの提供を積極的に実施している為、カルチャーの領域でも活用してみたいという意見が上がった							
ChatGPTを使ってデータ構造についてレビューを実施したところ生成系AIレビューのイメージのようなアイデアを検討してくれた			○				
ChatGPTによる可視化は細かい要求に対しても適切にデータを生成してくれる為、データ利用者にとって活用できるツールであることを実感した							○

この中でも特に重要な成果は以下である。

- 自治体がオープンデータを提供するプロセスを体験することで、作業量等の観点から自治体職員が現状の業務に加えて、手作業でデータ加工を行うことは負荷の観点で困難であると感じた

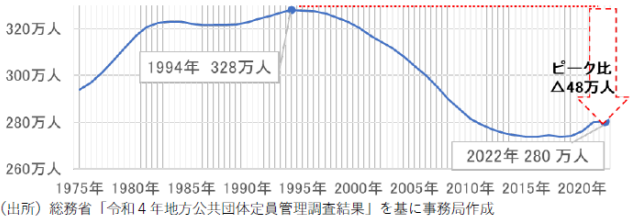
データ提供者に負担がかかることで、データ公開に対する抵抗感が強くなる可能性がある。
その為、データ提供者に負担をかけずにデータを加工する仕組みを今後検討する必要がある。

なお、以下内閣官房が公開している資料^[^1]では、DX・情報関係業務の担当職員数は、3名以下の自治体が54.6%(950団体/1,741団体)、1名以下の自治体が16.9%(295団体/1,741団体)と報告されている。過半数が3名以下であるため、自治体職員の負担は大きいと考えられる。

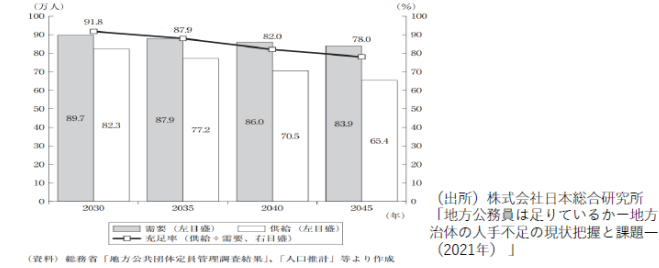
公共サービスの持続可能性への懸念 ①地方公共団体の職員数

○ 地方公共団体の職員数は、業務の効率化等によりピーク時から減少
○ デジタル化による更なる業務効率化も、3名以下の自治体が55%(1名以下の自治体は17%)であり、課題

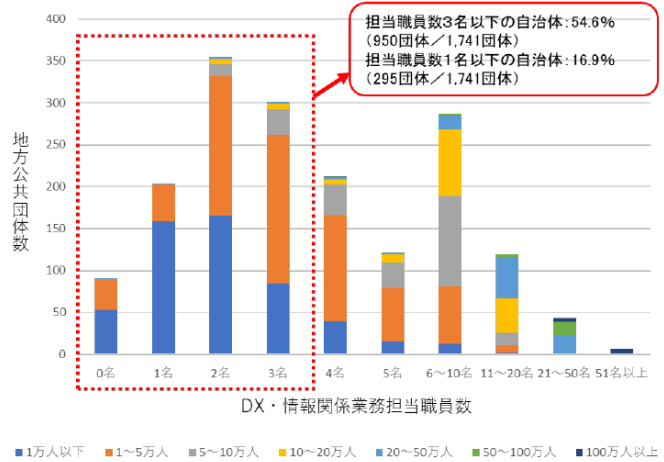
《地方公共団体（都道府県・市町村）職員総数の推移》



《地方公務員不足の将来推計（市町村・普通会計）》



《（人口規模別）DX・情報関係業務担当職員数》



6 課題

イベントの中であがった課題を以下の表にまとめた。

課題	関連するDataOpsプロセス						
	課題抽出	計画	データの収集蓄積	データの公開	データの取得	データの処理	データの活用
オープンデータがPDFで公開されている為、データ利用者は扱いづらい			○	○		○	○
自動化して効率的にデータの整形を行いたい			○				
自治体側がファイルに変更を加えた際、その変更がデータ利用時に大きな障害になってしまう場合がある為（想定外の項目が追加されている等）、相互のコミュニケーションが必要だと感じた			○				○

課題	関連するDataOpsプロセス						
	課題抽出	計画	データの収集蓄積	データの公開	データの取得	データの処理	データの活用
<p>以下の理由から初めにフォーマットを用意しておく必要性を実感した</p> <ul style="list-style-type: none"> ● 地域ごとにフォーマットが異なる ● 半角全角、記号などの統一が行われていない ● 自治体がオープンデータに加えた工夫（「市」の行を追加してグルーピングしている）で、かえって扱いにくいデータ形式になってしまっている 			○			○	○
<p>大元のデータが更新されると、リンクが無効になってしまう問題が発生したため、データ更新の際、何かアナウンスがあるべきだと感じた</p>				○	○		
<p>文化資源の領域のオープンデータも利活用したいという意見があったが、オープンデータの認知度や法整備などの課題がある</p>		○					
<p>ChatGPTが出してくれたデータ構造は細かく、確かに正しいが項目が大量になってしまうためデータ作成のコストがかかる</p>			○				
<p>CKANにオープンデータを公開するにあたりバージョンを意識する必要があるが、有識者のサポートが必要になる可能性が高い</p>				○			
<p>ChatGPTによる可視化はデータ利用者にとって活用できるツールではあるが、指示文のちょっとしたニュアンスの違いで結果が変わってしまうこともある為、必ず結果が確かなものかを確認する必要がある</p>							○

この中でも特に重要な課題は以下である。

- オープンデータがPDFで公開されている為、データ利用者は扱いづらい
- 以下の理由から初めにフォーマットを用意しておく必要性を実感した
 - 地域ごとにフォーマットが異なる
 - 半角全角、記号などの統一が行われていない
 - 自治体がオープンデータに加えた工夫で、かえって扱いにくいデータ形式になってしまっている

データ利用者にとって利活用しやすいデータ(※1)を公開することが利活用の促進につながると思われる。
その為、データ構造やフォーマットの統一を意識してデータを加工する仕組みを今後検討する必要がある。

※1 利活用しやすいデータとは、機械判読(※2)可能なデータを指す。

※2 機械判読とはコンピュータプログラムが自動的にデータを加工、編集等ができること

7 まとめ

今回のイベントでは、データ利用者の観点で、データを有効に活用するにはどのような構成とすべきかについて議論した。

データ構造やフォーマットの統一を意識してデータを加工することは必要であるが、その為のデータ提供者の作業コストは高くなることが確認された。

今後の課題としては、データ提供者に負担をかけずに、データ利用者が利活用しやすいデータに加工する仕組みを検討する必要がある。

この仕組みができることによって、利活用の促進につながることが期待できる。

[^1]: 内閣官房デジタル行財政改革会議事務局, “「国・地方を通じたDXの推進」に係る討議資料,” 2 2024. [オンライン]. Available: https://www.cas.go.jp/jp/seisaku/digital_gyozaiikaikaku/taiwa6/240227taiwa1.pdf. [アクセス日: 29 03 2024].