

# 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 #58

開催日時

2024年2月17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、[前回のイベント](#)で議論した、データ提供者とデータ利用者の間を仲介する二次データ提供者の作業プロセスにフォーカスを当て、そのプロセスの一部を実際に実施した。

二次データ提供者の作業プロセスは既存のDataOpsプロセスには存在しない為、以下を想定。

1. 一次データ提供者(データ提供者：自治体職員など)が公開したデータを取得する（二次データ提供者によるデータ取得プロセス）
2. 取得したデータを機械判読可能なデータへ加工する（二次データ提供者によるデータ処理プロセス）
3. データ利用者が利用できる形で提供・公開を行う（二次データ提供者によるデータ公開プロセス）
4. 一連のプロセスをCIなどで自動化する（二次データ提供者による自動化プロセス）

今回はこのうち、1、2、4の作業を実施した。

実際の作業内容としては以下の通り。

- ・ 厚労省が公開している「緊急避妊薬医療機関」のPDFデータ一括取得（dimによるデータ更新管理含む）

- 取得したPDFデータのCSV変換（フォーマットが異なるPDFについても考慮）
- PDF取得からCSV変換までの GitHub Actions による自動化

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

- [Robocorp](#)  
RPAフレームワーク、本イベントではPDFからテキストを抽出する用途で使用
- [tabula](#)  
PDFデータをCSVやExcelデータに変換するライブラリ
- [pandas](#)  
Pythonのデータ解析ライブラリ
- [dim](#)  
データインストールマネージャー  
プロジェクト内のオープン データを管理する

## タイムテーブル

11:00 SocialHackDay概要説明  
11:05 全体自己紹介  
11:10 持ち込みプロジェクト全体説明  
11:15 プロジェクト毎に分室  
11:20 DataOpsの詳細説明  
11:30 質疑・意見  
11:35 自己紹介  
11:40 厚労省公開の医療機関等一覧データのpdf\_to\_csv自動変換について  
11:55 休憩  
13:00 再開  
13:00 Robocorpの検証  
13:30 tabulaの検証  
13:55 リンク切れの問題について  
14:00 ~ 15:00 以下の残課題に関して分担作業

### 【残課題】

- PDF to CSV
  - ヘッダーがうまく表示されていない箇所の修正
  - ヘッダー毎に分かれてしまっている部分の修正
- リンク切れ
  - dimの機能でデータ更新管理(GitHub Actionsの作成)
  - 別プロジェクトのGoogleColabで試す

15:00 中間報告  
15:10 シャッフル  
15:15 シャッフル（質疑応答、意見）  
15:30 後半の開発タイム開始  
15:30～16:50 残課題に関して分担作業  
16:50 アンケート  
17:00 進捗・成果報告  
17:15 終了

## PDFデータ一括取得のイメージ

dimを使用した厚労省公開の医療機関等一覧データのPDF一括取得を実施。

PDFデータ一括取得のイメージは以下。

```
● $ sudo dim install -P https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000186912_00002.html -e ".pdf" -n "shinryoujo"
Installed to ./data_files/shinryoujo_1/000876235.pdf
Installed to ./data_files/shinryoujo_2/001095355.pdf
Installed to ./data_files/shinryoujo_3/000838800.pdf
Installed to ./data_files/shinryoujo_4/000835130.pdf
Installed to ./data_files/shinryoujo_5/001202747.pdf
Installed to ./data_files/shinryoujo_6/000901552.pdf
Installed to ./data_files/shinryoujo_7/000624110.pdf
Installed to ./data_files/shinryoujo_8/000997937.pdf
Installed to ./data_files/shinryoujo_9/001065009.pdf
Installed to ./data_files/shinryoujo_10/000619186.pdf
Installed to ./data_files/shinryoujo_11/001078067.pdf
Installed to ./data_files/shinryoujo_12/000838802.pdf
Installed to ./data_files/shinryoujo_13/000791218.pdf
Installed to ./data_files/shinryoujo_14/001176052.pdf
Installed to ./data_files/shinryoujo_15/000620924.pdf
Installed to ./data_files/shinryoujo_16/001017232.pdf
Installed to ./data_files/shinryoujo_17/001036451.pdf
Installed to ./data_files/shinryoujo_18/001126005.pdf
Installed to ./data_files/shinryoujo_19/000605600.pdf
Installed to ./data_files/shinryoujo_20/000863647.pdf
Installed to ./data_files/shinryoujo_21/001176053.pdf
Installed to ./data_files/shinryoujo_22/001095356.pdf
Installed to ./data_files/shinryoujo_23/001119260.pdf
Installed to ./data_files/shinryoujo_24/000841406.pdf
Installed to ./data_files/shinryoujo_25/000608002.pdf
Installed to ./data_files/shinryoujo_26/000752223.pdf
Installed to ./data_files/shinryoujo_27/000909780.pdf
Installed to ./data_files/shinryoujo_28/000587703.pdf
Installed to ./data_files/shinryoujo_29/000946363.pdf
Installed to ./data_files/shinryoujo_30/000985736.pdf
Installed to ./data_files/shinryoujo_31/000659670.pdf
Installed to ./data_files/shinryoujo_32/001011512.pdf
Installed to ./data_files/shinryoujo_33/000588946.pdf
Installed to ./data_files/shinryoujo_34/000605683.pdf
Installed to ./data_files/shinryoujo_35/000605688.pdf
Installed to ./data_files/shinryoujo_36/000605713.pdf
Installed to ./data_files/shinryoujo_37/000619187.pdf
Installed to ./data_files/shinryoujo_38/001036448.pdf
Installed to ./data_files/shinryoujo_39/001094323.pdf
Installed to ./data_files/shinryoujo_40/000589242.pdf
Installed to ./data_files/shinryoujo_41/000985733.pdf
Installed to ./data_files/shinryoujo_42/000841831.pdf
Installed to ./data_files/shinryoujo_43/000605681.pdf
Installed to ./data_files/shinryoujo_44/000850751.pdf
Installed to ./data_files/shinryoujo_45/000591524.pdf
Installed to ./data_files/shinryoujo_46/000849991.pdf
Installed to ./data_files/shinryoujo_47/000638156.pdf
Installed to ./data_files/shinryoujo_48/000610144.pdf
Installed to ./data_files/shinryoujo_49/001203117.pdf
```

## PDFからCSV変換時のイメージ

tabulaとpandasを使用したPDF→CSV変換を実施。

PDFからCSV変換と、変換されたCSVのイメージは以下。

\$ python3 main.py

PREFECTURE\_NUMBER 1 北海道  
PREFECTURE\_NUMBER 2 青森県  
PREFECTURE\_NUMBER 3 岩手県  
PREFECTURE\_NUMBER 4 宮城県  
PREFECTURE\_NUMBER 5 秋田県  
PREFECTURE\_NUMBER 6 山形県  
PREFECTURE\_NUMBER 7 福島県  
PREFECTURE\_NUMBER 8 茨城県  
PREFECTURE\_NUMBER 9 栃木県  
PREFECTURE\_NUMBER 10 群馬県  
PREFECTURE\_NUMBER 11 埼玉県  
PREFECTURE\_NUMBER 12 千葉県  
PREFECTURE\_NUMBER 13 東京都  
PREFECTURE\_NUMBER 14 神奈川県  
PREFECTURE\_NUMBER 15 新潟県  
PREFECTURE\_NUMBER 16 富山県  
PREFECTURE\_NUMBER 17 石川県  
PREFECTURE\_NUMBER 18 福井県  
PREFECTURE\_NUMBER 19 山梨県  
PREFECTURE\_NUMBER 20 長野県  
PREFECTURE\_NUMBER 21 岐阜県  
PREFECTURE\_NUMBER 22 静岡県  
PREFECTURE\_NUMBER 23 愛知県  
PREFECTURE\_NUMBER 24 三重県  
PREFECTURE\_NUMBER 25 滋賀県  
PREFECTURE\_NUMBER 26 京都府  
PREFECTURE\_NUMBER 27 大阪府  
PREFECTURE\_NUMBER 28 兵庫県  
PREFECTURE\_NUMBER 29 奈良県  
PREFECTURE\_NUMBER 30 和歌山県  
PREFECTURE\_NUMBER 31 鳥取県  
PREFECTURE\_NUMBER 32 島根県  
PREFECTURE\_NUMBER 33 岡山県  
PREFECTURE\_NUMBER 34 広島県  
PREFECTURE\_NUMBER 35 山口県  
PREFECTURE\_NUMBER 36 徳島県  
PREFECTURE\_NUMBER 37 香川県  
PREFECTURE\_NUMBER 38 愛媛県  
PREFECTURE\_NUMBER 39 高知県  
PREFECTURE\_NUMBER 40 福岡県  
PREFECTURE\_NUMBER 41 佐賀県  
PREFECTURE\_NUMBER 42 長崎県  
PREFECTURE\_NUMBER 43 熊本県  
PREFECTURE\_NUMBER 44 大分県  
PREFECTURE\_NUMBER 45 宮崎県  
PREFECTURE\_NUMBER 46 鹿児島県  
PREFECTURE\_NUMBER 47 沖縄県

	column 1	column 2	column 3	column 4	column 5
1	施設名	郵便番号	住所	電話番号	産科、婦人科又は産婦人科の標榜の有無
2	江別市立病院	067-8585	北海道江別市若草町6番地	011-382-5151	有
3	医療法人社団ポリウイメンズクリニック	067-0031	北海道江別市元町21番地の9	011-391-7577	有
4	医療法人社団みらいエナレディースクリニック	061-3209	北海道石狩市花川南9条1丁目86番地2	0133-72-8688	有
5	市立千歳市民病院	066-8550	北海道千歳市北光2丁目1番1号	0123-24-3000	有
6	マミーズクリニックちとせ	066-0038	北海道千歳市信濃2丁目1-13	0123-27-4103	有
7	医療法人社団さくら女性クリニック	066-0062	北海道千歳市千代田町7丁目1789-3千歳ステーションプラザ5F	0123-27-1323	有
8	医療法人社団産婦人科・小児科クリニックリブ	061-1441	北海道恵庭市住吉町1丁目8-10	0123-32-0181	有
9	みよしレディースクリニック	061-1133	北海道北広島市栄町1丁目5-3北広島ビル2F	011-373-8500	有
10	木古内町国民健康保険病院	049-0422	北海道木古内町字本町710	01392-2-2079	有
11	八雲総合病院	049-3197	北海道二海郡八雲町東雲町50番地	0137-63-2185	有
12	北海道立江差病院	043-0022	北海道樺山郡江差町伏木戸町484番地	0139-52-0036	有
13	JA北海道厚生連俱知安厚生病院	044-0004	北海道虻田郡俱知安町北4条東1丁目2番地	0136-22-1141	有
14	寿都町立寿都診療所	048-0406	北海道寿都町遊馬町72番地2	0136-62-2411	有
15	医療法人社団青見沢レディースクリニック	068-0028	北海道岩見沢市8条西19丁目3-1	0126-23-2222	有
16	砂川市立病院	073-0196	北海道砂川市西4条北3丁目1-1	0125-54-2131	有
17	医療法人社団たきかわ産科婦人科クリニック	073-0021	北海道滝川市本町2丁目5番18号	0125-23-3039	有
18	名寄市立総合病院	096-8511	北海道名寄市西7条南8丁目1番地	01654-3-3101	有
19	社会福祉法人北海道社会事業協会富良野病院	076-8765	北海道富良野市住吉町1-30	0167-23-2181	有
20	留萌市立病院	077-8511	北海道留萌市東雲町2丁目16番地	0164-49-1011	有
21	市立稚内病院	097-8555	北海道稚内市中央4丁目11番6号	0162-23-2771	有
22	利尻島国保中央病院	097-0401	北海道利尻郡利尻町富形字緑町11番地	0163-84-2626	有
23	網走厚生病院	093-0076	北海道網走市北6条西1丁目9番地	0152-43-3157	有
24	金川医院	093-0012	北海道網走市南2条西2丁目10番地	0152-43-2884	有
25	やまかわウイメンズクリニック	090-0804	北海道北見市桜町5丁目8番地	0157-66-0011	有
26	産科・婦人科北見レディースクリニック	090-0020	北海道北見市大通東2丁目11番地	0157-31-0303	有
27	医療法人社団公和会中村記念愛成病院	090-0051	北海道北見市高栄東町4丁目20番1号	0157-24-8131	有
28	広域紋別病院	094-8709	北海道紋別市落石町1-10-37	0158-24-3111	有
29	社会医療法人製鉄記念室蘭病院	050-0076	北海道室蘭市知別町1丁目45番地	0143-44-4650	有
30	医療法人緑風会石田内科骨傷科	049-5615	北海道洞爺湖町本町195	0142-76-3838	有
31	社会医療法人母恋日鋼記念病院	051-8501	北海道室蘭市新富町1-5-13	0143-24-1331	有
32	医療法人社団のぞみ会レディースクリニックぬまのはた	059-1304	北海道苫小牧市北栄町2-27-7	0144-53-0303	有
33	王子総合病院	053-8506	北海道苫小牧市若草町3丁目4-8	0144-32-8111	有
34	医療法人社団活仁会桜木ファミークリニック	053-0832	北海道苫小牧市桜木町2丁目25-1	0144-71-2351	有
35	新いもろが町市民病院	056-0004	北海道新いもろが町新町内崎町4丁目4番1号	0146-42-0181	有

## 4 アンケート結果

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。

### アンケート結果

## 5 成果

イベントの中で得られた成果を以下の表にまとめた。

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	フィードバック
tabulaとpandasを用いたPDF→CSV変換で、改行コードを置換することで変換することに成功した						○		
dimの機能を活用して、PDFを取得することに成功した					○			
CIによって定期的（1日1回）にPDFを一括ダウンロードし、CSVに自動変換するツールが完成した					○	○		

この中でも特に重要な成果は以下である。

- CIによって定期的（1日1回）にPDFを一括ダウンロードし、CSVに自動変換するツールが完成した
- 二次データ提供者によるデータの取得（PDF）から機械判読可能なデータ（CSV）に変換するまでの処理を自動化することに成功した。[\(opendata-pdf-to-csv\)](#)

後述の「課題」にも記載されている通り、いくつか課題が残っている為、今後も検討を進める必要がある。

## 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
オープンデータとして公開されているPDFが定期的に更新され、リンク切れが発生してしまう → dimによるデータ更新管理で対応					○		
リンク切れの判別は、PDFをダウンロードして確認するか、サイトの最終更新で確認するしかない → dimによるデータ更新管理で対応					○		
dimによるPDF取得時に2ファイルほど関係のないファイルが混ざっている					○		
PDFを一括ダウンロードして、csvのような機械可読性の高いものに自動変換したい → イベント内で作成完了					○	○	
Robocorpを使用したPDF→CSV変換ではPDF内のテーブルが認識されずテキストとして出力されてしまう						○	
tabulaとpandasを用いたPDF→CSV変換を行う際、以下の課題を確認 ● 変換後、出力されたCSVがPDFのページ毎にわかれてしまう → イベント内で対応済 ● PDFのヘッダーが毎回記述されてわかりにくい						○	
各都道府県によってフォーマットの差異がある（1ページ目だけにヘッダーが記載されている、独自の列があるなど）						○	

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
データが公開されているページにPDFの追加や順番の入れ替わり等が発生するとPDFの更新取得処理に失敗する可能性がある → CIが落ちることで検知可能					○		
今回のPDF→CSV変換処理は、幅広い分野のオープンデータに対しても検証を進めた方がいい						○	

この中でも特に重要な課題は以下である。

- 今回のPDF→CSV変換処理は、幅広い分野のオープンデータに対しても検証を進めた方がいい

今回のイベントでは、厚労省が公開している「緊急避妊薬医療機関」のPDFデータにて検証を進めたが、他のオープンデータに関しても同様の問題が発生する可能性がある為、引き続き幅広いオープンデータの検証を進める必要がある。

## 7 まとめ

今回のイベントでは、データ提供者とデータ利用者の間を仲介する二次データ提供者の作業プロセスにフォーカスを当て、PDFデータの取得からPDF→CSVへの変換処理までの自動化に成功した。

動くものは完成したものの、[6 課題](#)にも記載している問題もある為、今後も検討を進めていく必要がある。