

SIGNATE COVID-19 Challenge

Phase 1



SIGNATE inc.
Find the sign of changing times.

Copyright © 2020 SIGNATE, Inc. All Rights Reserved.

- ・国や自治体によって配信方針がバラバラ
- ・PDF、画像形式などの非構造データが多い
- ・一元的・網羅的に状況が迅速に把握できない



Phase1では、COVID-19罹患者に関する、
マシンリーダブルかつデータ分析可能な
最大規模のデータセットの構築を目指します

1. 情報を集めて転載する

国・自治体のHPから患者情報を項目別に転載する

2. 情報を整理する

1の中で、特定のカラムについて内容を整理する



1.情報を集めて転載する



1. 新規患者に注目

シートからどの患者情報を入力するかを決めます

2. 自治体のHPから転載

対応する自治体のHPから情報をコピペします

3. フォーラムに報告

作業内容をフォーラムにあげます



1. 新規患者に注目

SIGNATE

6

都道府県コード	症例番号	都道府県症例番号	発症日	確定日	公表日	受診都道府県	居住都道府県	居住市区町村	年代	性別	職業
14	1292	14-1292				神奈川県 ▾			▼	▼	▼
14	1293	14-1293				神奈川県 ▾			▼	▼	▼
14	1294	14-1294				神奈川県 ▾			▼	▼	▼
14	1295	14-1295				神奈川県 ▾			▼	▼	▼
14	1296	14-1296				神奈川県 ▾			▼	▼	▼
14	1297	14-1297				神奈川県 ▾			▼	▼	▼
14	1298	14-1298				神奈川県 ▾			▼	▼	▼
14	1299	14-1299				神奈川県 ▾			▼	▼	▼
14	1300	14-1300				神奈川県 ▾			▼	▼	▼
14	1301	14-1301				神奈川県 ▾			▼	▼	▼
14	1302	14-1302				神奈川県 ▾			▼	▼	▼
14	1303	14-1303				神奈川県 ▾			▼	▼	▼
14	1304	14-1304				神奈川県 ▾			▼	▼	▼

まず、シート【罹患者】を見ます。

1行に1患者の情報があります。「都道府県番号」は【都道府県マスタ】に一覧があります。

「症例番号」は各都道府県HPで公表されているX例目という数値にあたります。

都道府県番号と症例番号を組み合わせたものが「都道府県症例番号」です。

- 登録されている患者：オレンジ色が多い行があります。こちらが情報の追加対象です。
- 患者がシートにない場合：行を挿入して情報を入力します。
- 患者が間違っている場合：情報を訂正します。その旨をフォーラムに投稿ねがいます。

2. 自治体のページから転載

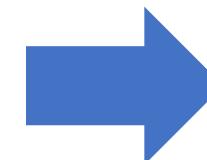
SIGNATE

7

情報発信元	URL
厚生労働省	https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000121431_00086.html
北海道	http://www.pref.hokkaido.lg.jp/hf/kth/kak/hasseijoukyou.htm
青森県	https://www.pref.aomori.lg.jp/welfare/health/wuhan-novel-coronavirus2020.html
宮城県	https://www.pref.miagi.jp/site/covid-19/02.html
秋田県	https://www.pref.akita.lg.jp/pages/archive/47957
山形県	https://www.pref.yamagata.jp/ou/bosai/020072/kochibou/coronavirus/coronavirus.html
福島県	https://www.pref.fukushima.lg.jp/sec/21045c/fukushima-hasseijoukyou.html
茨城県	https://www.pref.ibaraki.jp/1saigai/2019-ncov/index.html
栃木県	http://www.pref.tochigi.lg.jp/e04/welfare/hoken-eisei/kansen/hp/coronakensahasseijoukyou.html
群馬県	https://www.pref.gunma.jp/07/z87g_00016.html
埼玉県	https://www.pref.saitama.lg.jp/a0701/covid19/jokyo.html
千葉県	https://www.pref.chiba.lg.jp/shippei/press/2019/ncov-index.html

自治体のHP
から患者の詳細情報を確認

罹患者の公表情報例（北海道157例目）の公表情報から一部抜粋	
【患者（道発表157例目）の概要】	
年 齢	50代
性 別	男性
国 籍	日本
居住地	旭川市
職 業	医師
症状・経過	<p>【経過】 2月26日(水) 倦怠感、頭重感、咽頭の違和感が出現（咳、発熱なし） 3月4日(水) 市内の医療機関A（勤務先病院）を受診、インフルエンザ検査陰性、胸部CT検査で肺炎像は認められず、以降、自宅療養していた。 3月16日(月) 本人から旭川市保健所に相談 3月17日(火) 新型コロナウイルス検査を実施 3月18日(水) 検査の結果、陽性と確認 保健所の調整のもと、市内の医療機関Bに入院</p> <p>【現在の状況】 発熱はなく、頭痛がある。</p>
行動歴	<p>海外渡航歴なし。 2月26日(水)の症状出現後、3月4日(水)午前9時まで、土日を除き通常勤務。 3月4日(水)以降は自宅療養とし、外出はしていない。 3月18日(水) 保健所の調整のもと、市内の医療機関Bに入院。 勤務先の職員に体調不良者はいない。接触患者等の体調については、調査中。</p>
※ 旭川市でこれまで取り扱った検査数（3月19日午前9時時点） 総検査数：164検体（127人）、うち市内在住者の陽性（10人）	



対応する項目を記入して
いきましょう。
【公表日】HPに公開され
た日付です。必ず入力願
います。

データの中には選択式のものがあります。選択項目がない場合は、
フォーラムなどでSIGNATEまで連絡ねがいます。

SIGNATE COVID-19 Case Dataset 【罹患者】

都道府県コード	症例番号	都道府県症例番号	発症日	確定日	公表日
1	157	1-157	2020/02/26	2020/03/18	2020/03/19
～					情報源
～					備考
http://www.pref.hokkaido.lg.jp/hf/kth/kak/k					非公表

3. フォーラムに報告

SIGNATE

8

The screenshot shows the SIGNATE platform's navigation bar with links for Competition, Quest, and Career. Below the navigation bar, there are tabs for Competitions, Rankings, Stats, and other support pages like 'コンペへの参加方法' (How to participate in a competition), '機能一覧' (List of features), 'よくある質問' (FAQ), and 'お問い合わせ' (Contact us). A red 'New' badge is visible above the 'Forum' section. The 'Forum' section contains two buttons: '+ ディスカッションを作成' (Create a discussion) and 'フォーラム機能について' (About forum functions). A post by 'SIGNATE' is shown, reading: '貢献いただいた方はこちらに返信願います！ (2nd)' (Please reply here if you contributed! (2nd)). The post was made 2 hours ago.

フォーラム「貢献いただいた方はこちらに返信願います！ (2nd)」にアクセスして作業内容を投稿ねがいます。

作業していて判断に迷うときは、フォーラム「作業の進め方で不明な点はこちらに返信願います！」に投稿ねがいます。

- ・入力したい患者がシートにないときは？
- ・一部だけ未入力が残っている患者を見つ
けたときは？



質問回答1

SIGNATE

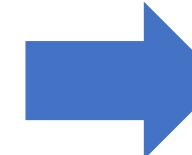
10

質問：入力したい患者がシートにないときは？

回答：シートの適当な位置に行を挿入して、情報を入力してください。

例：埼玉県の1001例目の患者をシートに入力したいとき

都道府県コード	症例番号	都道府県症例番号	発症日	確定日	公表日	受診都道府県
11	999	11-999	2020-05-15	2020-05-21	2020-05-22	埼玉県 ▾
11	1000	11-1000			2020-05-23	埼玉県 ▾
12	1	12-1	2020/01/20	2020/01/31	2020/01/31	千葉県 ▾
12	2	12-2	2020/02/01	2020/01/30	2020/01/30	千葉県 ▾



都道府県コード	症例番号	都道府県症例番号	発症日	確定日	公表日	受診都道府県
11	999	11-999	2020-05-15	2020-05-21	2020-05-22	埼玉県 ▾
11	1000	11-1000			2020-05-23	埼玉県 ▾
11	1001	11-1001				千葉県 ▾
12	1	12-1	2020/01/20	2020/01/31	2020/01/31	千葉県 ▾
12	2	12-2	2020/02/01	2020/01/30	2020/01/30	千葉県 ▾

埼玉県の1000例目（都道府県症例番号 11-1000）と千葉県の1例目（都道府県症例番号 12-1）の間に新しい行を挿入してください。

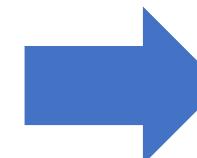
その上で、先ずは「都道府県コード」、「症例番号」、「都道府県症例番号」を入力してから、その他の情報を転記してください。

質問：一部だけ未入力が残っている患者を見つけたときは？

回答：自治体等のHPを確認して未入力分の内容を埋めてください。

例：発症日が空欄のままであった場合

都道府県コード	症例番号	都道府県症例番号	発症日	確定日	公表日	受診都道府県
11 850	11-850		2020-04-29	2020-04-29	埼玉県	▼



都道府県コード	症例番号	都道府県症例番号	発症日	確定日	公表日	受診都道府県
11 850	11-850		非公表	2020-04-29	2020-04-29	埼玉県

自治体などのHPを確認して、未入力の項目に関する情報を参照してください。

確認ができたら、その内容をシートに記入してください。なお、対象の項目が情報源を参照しても確認できなかった場合は、上の例のように非公表と記入してください。

一部の項目の入力だけの場合でも、作業結果をフォーラムに投稿願います。



2.情報を探査する



1. 症状

典型的な症状を把握するために整理します

2. 関係性

クラスターを把握するために整理します



1. 症状

SIGNATE

14

A	B	C	D	E	F	G
都道府県コード	症例番号	都道府県症例番号	症状・経過	日付	症状	確認済フラグ
17 9	17-9	17-9	3月22日(日曜日)鼻閉、味覚・嗅覚障害 3月24日(火曜日)県外の医療機関を受診 3月27日(金曜日)帰国者・接触者相談センターに相談 帰国者・接触者外来を受診 37度台の発熱 3月28日(土曜日)石川県保健環境センターにおいて、PCR検査を実施したところ、陽性と判明し、入院	-	-	-
				2020/03/22	icd=R448;鼻閉	▼
				2020/03/22	icd=R431;嗅覚障害	▼
				2020/03/22	icd=R432;味覚障害	▼
				2020/03/24	受診	▼
				2020/03/27	icd=R509;発熱	▼
				2020/03/27	受診	▼
				2020/03/28	入院	▼
17 10	17-10	17-10	3月28日（土曜日）37度台の発熱 3月30日（月曜日）帰国者・接触者相談センターに相談。帰国者・接触者外来を受診、肺炎と診断。石川県保健環境センターにおいて、PCR検査を実施したところ、陽性と判明	-	-	-

シート【症状】を見ます。A~D列には【罹患者】から転記した情報があり、こちらは編集不要です。そこから1行下がったE、F列には、各種症状、および、**受診、入院、退院**の情報を機械的に抽出した結果をトランザクション形式で記載しています。

- 抽出されているE、F列の内容が正しければ、その行の「確認済フラグ」に1を入力してください。
- 抽出されているE、F列の内容が誤っていれば、その行を修正してください。
- 症状、および、**受診、入院、退院**のいずれかで抽出が不足しれいてば、行を追加して追記してください。

- ・症状と合わせて記載されている icd とは何か？
- ・修正が必要なのは例えばどのようなケースがあるか？
- ・症状を追記する場合はどのように入力する？



質問：症状と合わせて記載されている icd とは何か？

回答：症状を特定するためのコードです

都道府県コード	症例番号	都道府県症例番号	症状・経過	日付	症状
7 36	7-36		4月 1日（水）味覚、嗅覚異常あり 4月 5日（日）倦怠感、咳あり。その後、8日（水）まで継続。 4月 7日（火）医療機関受診 4月 9日（水）医療機関受診 4月 10日（金）新型コロナウイルスの検査結果、陽性と判明	-	-
				2020/04/01	icd=R431;嗅覚異常
				2020/04/01	icd=R432;味覚異常
				2020/04/05	icd=R53;倦怠感

比較分析などへの活用も考慮して、症状については標準規格として規定された国際疾病分類第10版※1におけるコード（ICD10コード）も併記しました。

「症状・経過」から機械的に症状を抽出する際には、表記ゆれを加味しながら集計するために万病辞書※2を活用しております。

※1 国際疾病分類第10版, <http://www.byomei.org/icd10/index.html>

※2 万病辞書, <http://sociocom.jp/~data/2018-manbyo/index.html>

質問回答2

SIGNATE

17

質問：修正が必要なのは例えばどのようなケースがあるか？

回答：「症状以外のものが抽出」、「日付情報が不十分」などが主です

例：検査項目が症状として抽出されてしまっている

1	都道府 県コード	症例番 号	都道府県 症例番号	症状・経過	日付	症状	確認済フラグ
16985	13 14	13-14		<p>2020/02/02:咳症状あり</p> <p>2020/02/05:発熱があり、都内医療機関Aを受診</p> <p>2020/02/10:都内医療機関Bを受診。同日都外に出張(新幹線)</p> <p>2020/02/12:症状が改善せず、都内医療機関Cを受診、医療機関Dを紹介され、そのまま入院(この間、インフルエンザ等の検査を実施していたが、特定の感染症と診断されず)</p> <p>2020/02/14:東京都健康安全研究センターで検査実施</p> <p>2020/02/15:陽性判明</p>	-	-	-
16986					2020/02/02	icd=R05;咳	▼
16987					2020/02/05	icd=R509;発熱	▼
16988					2020/02/05	受診	▼
16989					2020/02/10	受診	▼
16990					2020/02/12	icd=J111;インフルエンザ	▼
16991					2020/02/12	受診	▼

機械的な抽出にて誤って記載されてしまった不要な行（上記の例であれば16990行目）を、行ごと削除してください。

質問回答2 つづき

SIGNATE

18

質問：修正が必要なのは例えばどのようなケースがあるか？

回答：「症状以外のものが抽出」、「日付情報が不十分」などが主です

例：文脈から日付情報をYYYY/MM/DD形式に修正することができる

1	都道府県コード	症例番号	都道府県症例番号	症状・経過	日付	症状	確認済万ケ
6146		4 35	4-35	3月31日 発熱（37.0°C）。 4月1日 解熱。 2日 倦怠感出現。 4日 発熱（37.0°C）。福島県内の医療機関を受診。薬を処方された。 夜間に発熱（38.1°C）出現。 5日 解熱。 6日 嗅覚障害出現。 7日 コールセンターを通じて帰国者・接触者相談センターに相談。 8日 味覚障害出現。帰国者・接触者外来を受診。 9日 検査の結果、陽性が判明。入院先の感染症指定医療機関等を調整中。	-	-	-
6147					2020/03/31	icd=R509;発熱	
6148					2日	icd=R53;倦怠感	
6149					4日	icd=R509;発熱	

「症状・経過」において、年や月の情報が直接記載されていなくても、文脈から特定できる場合、「日付」カラムの値をYYYY/MM/DD形式に修正してください。

（上記の例であれば、2日→2020/04/02、3日→2020/04/04）

質問：症状を追記する場合はどのように入力する？

回答：【病名マスタ】を確認しながらICD10コードも合わせて入力してください

例：症状の抽出もれを補うとき

1	都道府県コード	症例番号	都道府県症例番号	症状・経過	日付	症状	確認済フラグ
5899		4 4	4-4	3月25日 発熱(39°C)、 だるさ 、咳 3月28日 症状は落ち着いたが医療機関Aを受診 3月28日 3月20日に患者A(4-3)と飲食をともにしていたことから、患者Aの相談を受けた帰国者・接触者相談センターから連絡 3月29日 帰国者・接触者外来を受診、検査の結果、陽性が判明	-	-	-
5900					2020/03/25	icd=R509;発熱	▼
5901					2020/03/25	icd=R05;咳	▼
5902					2020/03/28	受診	▼

【病名マスタ】を参照し、「出現形」において“だるさ”を検索。

出現形	標準病名	ICDコード
だるさ	全身倦怠感	icd=R53

「標準病名」と「ICDコード」を確認して、その内容を元に新規行を挿入しつつ入力してください。
(上記の例であれば、5900行目の下に行を挿入して、日付：2020/03/25、症状：icd=R53;全身倦怠感 を入力)

シート【罹患者】の「濃厚接触者状況」や、報道されたクラスター情報などを見ます。
確認できた患者同士の接触関係を、シート【罹患者関係】に整理します。

【罹患者関係】では、1行に2名の患者同士の関係情報を入力します。

- ・まだ登録されていない関係：行を挿入して関係情報を入力します。
- ・関係が間違っている場合：情報を訂正します。その旨をフォーラムに投稿ねがいます。

都道府県症例番号1	都道府県症例番号2	接触日	関係性	場所	情報源
1-218	1-238	非公表	同僚	▼ 北海道:札幌市 札幌呼吸器科病院	http://www.pref.hokkaido.lg.jp
1-218	1-249	非公表	医療従事者と患者	▼ 北海道:札幌市 札幌呼吸器科病院	http://www.pref.hokkaido.lg.jp
1-218	1-250	非公表	医療従事者と患者	▼ 北海道:札幌市 札幌呼吸器科病院	http://www.pref.hokkaido.lg.jp
1-218	1-251	非公表	医療従事者と患者	▼ 北海道:札幌市 札幌呼吸器科病院	http://www.pref.hokkaido.lg.jp
1-218	1-252	非公表	医療従事者と患者	▼ 北海道:札幌市 札幌呼吸器科病院	http://www.pref.hokkaido.lg.jp

「場所」についてはSIGNATEにて名寄せなどの修正を行う場合があります。
(名寄せした結果は【場所マスタ】にて管理)

ご協力ありがとうございます！