

SARS-CoV-2 感染リスクオントロジーの提案

A proposal for SARS-CoV-2 infection risk ontology

江上周作^{1*} 大向一輝² 山本泰智³
Shusaku Egami¹ Ikki Ohmukai² Yasunori Yamamoto³
伊藤真和吏⁴ 坂根昌一⁵ 網淳子⁶ 奥村貴史^{7,8}
Maori Ito⁴ Shoichi Sakane⁵ Junko Ami⁶ Takashi Okumura^{7,8}

¹ 産業技術総合研究所人工知能研究センター

¹ Artificial Intelligence Research Center, National Institute of Advanced Industry Science and Technology, Tokyo, Japan

² 東京大学大学院人文社会系研究科

² Graduate School of Humanities and Sociology, The University of Tokyo, Tokyo, Japan

³ 情報・システム研究機構 ライフサイエンス統合データベースセンター

³ Database Center for Life Science, Research Organization of Information and Systems, Chiba, Japan

⁴ PLOD info

⁵ シスコシステムズ合同会社イノベーションラボ

⁵ Innovation Labs, Cisco Systems, G.K.

⁶ WIDE Project

⁷ 北見工業大学

⁷ Kitami Institute of Technology, Hokkaido, Japan

⁸ 国立保健医療科学院

⁸ National Institute of Public Health, Japan

Abstract: 感染力の強い感染症に罹患した患者が生じた際、公衆衛生当局は、患者の健康状態や移動等の情報を開示し、社会に注意喚起することがある。著者らは、この患者の移動情報を構造化し、Patient Locational Open Data (PLOD) として公開し、患者との感染リスクの把握に応用することを提案してきた。この移動情報に加えて、移動先となる場所ごとの性質や感染リスクを高める要素について定性的な情報が記述されていれば、感染リスクの自動的な評価が可能となる。しかし、場所ごとにどのような感染リスクがあるかという知識は整理されていない。また、患者側の聞き取り調査で得られる移動に関する用語の統制もされていなかった。そこで本稿では、現在感染拡大している新型コロナウイルス感染症 (COVID-19) を対象として、場所に関する潜在的な感染リスクを定義したオントロジーを提案する。まず、これまでの PLOD の取り組みを紹介する。その上で、その機能をさらに強化する感染リスクオントロジーを示し、実運用に向けたシナリオの検討と考察を行う。

1 はじめに

結核や麻疹、新型インフルエンザ、新型コロナウイルス感染症に代表される感染力の強い感染症が発生した場合、公衆衛生当局は、濃厚接触者を特定して健康監視する必要がある。この接触者調査は、公衆衛生当

局による聞き取りにより行われる。しかし、公共交通機関や何らかの大規模施設を訪問した際には、全ての接触者を追跡することが困難となる。こうした不特定多数との接触が生じた際、当局は、患者の移動情報を公開し、接触可能性のある住民に保健所への連絡を呼びかける。しかし、住民の多くは、家族に免疫不全患者がいる等のケースを除いて、こうした当局からの情報提供に興味を払わない。報道機関に提供した情報は記事にはなるものの、それらの記事の拡散力は限定的

*連絡先：国立研究開発法人産業技術総合研究所人工知能研究センター
〒135-0064 東京都江東区青海 2-4-7
E-mail: s-egami@aist.go.jp

であり、有効な対策に繋がらない懸念があった。

このように、感染症への罹患リスクについて、公衆衛生当局は潜在的な接触者に効果的に情報提供することが出来ずにきた。もし、このプレスリリース自体をオープンデータとして公開することができれば、蔓延地域の住民は携帯電話に蓄積される自らの位置情報履歴との交差を計算し、患者との感染リスクを簡単に把握することが可能となる。そこで、我々は公開された患者の移動情報を対象として、患者位置情報に関するオープンデータを Resource Description Framework (RDF) 形式で公開するための手法を提案し [10], Patient Locational Open Data (PLOD) プロジェクトとして実用化に向けた活動をしている¹。

一方、新型コロナウイルス接触確認アプリ COCOA[16] に代表される、スマートフォン用いて得られる各種情報に基づいて感染リスクを通知するシステムには、電車での移動時の濃厚接触の誤検知や、間接的な接触を検知することができないなどの課題がある。こうした誤検知や検知漏れには、現状は手作業による対応が強いられている。公衆衛生当局におけるこの負担を軽減し、感染リスクを自動的に評価するためには、場所ごとの性質や感染リスクを高める要素について定性的な情報が必要となる。しかし、そもそも場所ごとにどのような感染リスクがあるかという知識は整理されておらず、聞き取り調査で得られる場所情報に関する用語の統制もされていない。

そこで、本研究では現在感染拡大している新型コロナウイルス感染症 (COVID-19) を対象として、場所に関する潜在的な感染リスクの関係を定義したオントロジーを提案する。特に、世間一般に周知されている 3つの密 [18] や Three Cs[13] などの標語に基づき、施設や公共交通機関などの場所と、感染防止のために避けるべき要素との関係性を把握できるようにする。具体的には、場所と人との関係性の記述にアフォーダンスの概念を導入し、店舗、施設、公共交通機関などの場所が、三密のどれに該当するか推論できるようなオントロジーの開発を試みた。これにより、個人の行動履歴と紐づけることで、その一連の行動が持つ感染リスクを自動的に導出することが可能になるものと期待される。

以下ではまず、2章で関連研究について説明し、3章で PLOD の取り組みについて述べる。4章において、提案する感染リスクオントロジーについて述べる。5章で考察を述べ、6章でむすびとする。

2 関連研究

2.1 感染症対策と携帯位置情報

携帯電話の位置情報を公益活用する試みは、携帯電話が普及した 2000 年代の後半頃に起源がある [15]。もっともシンプルな形として、患者の追跡のために患者携帯の位置情報を利用するというものがあり、2015 年に韓国で生じた中東呼吸器症候群 (Middle East Respiratory Syndrome, MERS) 騒動において利用された報告がある [7]。しかしながら、携帯位置情報はプライバシー上極めてセンシティブであり、公衆衛生への活用においても倫理的な議論があった [6]。とりわけ、患者個々の位置情報を扱うことはハードルが高く、研究論文として報告される例も限定されてきた。

こうした状況は、2019 年より生じた COVID-19 のパンデミックによって、大きく変わりつつある。中国では、顔認証やセキュリティーカメラ、ソーシャルメディア管理を通常の人間による監視と組み合わせ、COVID-19 患者の接触者追跡を自動化したと言われている [8]。そのために、携帯電話の通話情報、位置情報など各種サービスの利用記録を企業から収集したとされる。中国においては、国内の交通・通信網を管理する企業の多くが国営であることから、迅速にデータを活用することが可能であった。韓国では、保健当局が、クレジットカードの使用記録、防犯カメラ映像、携帯電話の位置情報、公共交通用カード、入国管理記録などを精査することが可能であり、感染者や感染リスクのある人の移動履歴を突き止めている [5]。台湾でも、自宅隔離の対象者が携帯電話の位置情報を利用して追跡されるとされる [5]。こうしたサービスは、韓国における患者追跡の事例 [7] と筆者らによるわが国における患者との接触リスク通知サービスの提案 [11] を除いて、新型コロナウイルス感染症の大流行前には研究論文としての公開はほとんど知られていない。

2.2 COVID-19 とセマンティック技術

現在進行している COVID-19 のパンデミックにより、多方面で関連論文が急増しており、セマンティック Web 分野においても同様である。出版された文献から COVID-19 に関する知識を抽出してナレッジグラフを構築した研究 [2] が報告されているなど、文献を対象とした取り組みが多い。Kaggle では 200,000 を超える学術論文のデータセットを提供し、COVID-19 Open Research Dataset Challenge (CORD-19) を開催している²。このデータセットに基づいてナレッジグラフを

¹<https://www.plod.info/>

²<https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge>

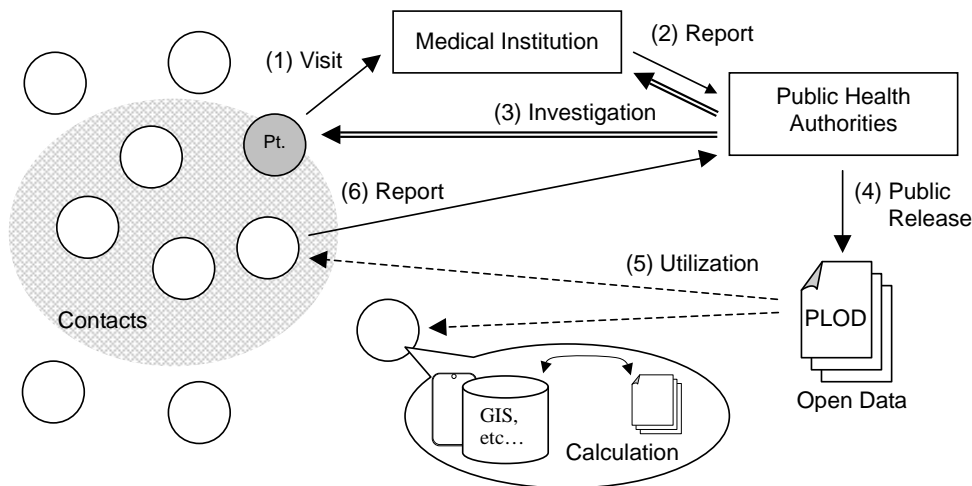


図 1: PLOD に基づく接触リスク計算の流れ

構築し、外部リソースとのリンク付けをした研究が報告されている [9][14].

また、生物医学の観点からはコロナウイルス感染症に関するオントロジーとして CIDO[4] が開発されており、COVID-19 に焦点を当てて CIDO を拡張したオントロジーも提案されている [1].

このように、計量書誌学的や生物医学的な分析の観点では、COVID-19 の課題に対するセマンティック技術の応用は進んでいるものの、感染リスク通知や追跡調査などの公衆衛生的な観点では、セマンティック技術の応用はそれほど進んでいない。我々の知る限り、施設や公共交通機関などの場所の種類に焦点を当て、用語の統制や感染リスクの推論を支援するオントロジーは提案されていない。

3 患者移動情報のオープンデータ化

公開しているプレスリリースを構造化したオープンデータとして公開することができれば、感染地域の住民は、携帯電話に蓄積される自らの位置情報履歴との交差を計算し、患者との感染リスクを簡単に把握することが可能となる。こうしたアプリケーションは、患者の訪問先が多いケースや、同時発生する患者数が多いケース等、すべての患者情報を検証していくことが困難となっていく場合において、とりわけ有効に機能することが期待される。そこで、我々はこれまでに、公開可能な情報として公開された患者の移動情報を対象として、患者の移動情報をオープンデータ化し活用するための手法について提案してきた (図 1)。

まず、(1) 患者が医療機関を受診し、医師が患者を診断する。次に、(2) 医師は、患者の発生を保健所に届け出る。さらに、(3) 保健所は、患者への聞き取り調査を実施し、(4) 患者発生のプレスリリースを公開する。我々は、このプレスリリースを、患者位置情報に関するオープンデータ (Patient Locational Open Data, PLOD) として、規定のフォーマットにおいて公開することを提案する。こうしたデータは、(5) 自由に利用することができ、患者移動のマッピングや個人にとっての患者との接触リスク推定に利用することができる。この推定においては、住民側は、自らの携帯電話を用いて記録されていく位置情報を利用する。そのうえで、接触が疑われる個人は、(6) 保健所に届け出て、その後の必要な指示を受けることになる。

3.1 患者移動情報のスキーマ

患者情報を PLOD として公開するうえで、プレスリリースに含まれる情報の構造化について提案する。現在、日本においてプレスリリースを通じて公開される感染に関する情報には、以下のような情報が自由記述されており、自動処理を行うことが難しい。

- 文書の発行者・連絡先情報等
- 患者自身の属性情報 (年齢帯や性別)
- 患者の罹患している感染症情報 (名称や感染力の程度)
- その他医学的背景情報
- 患者自体の大まかな臨床的経過

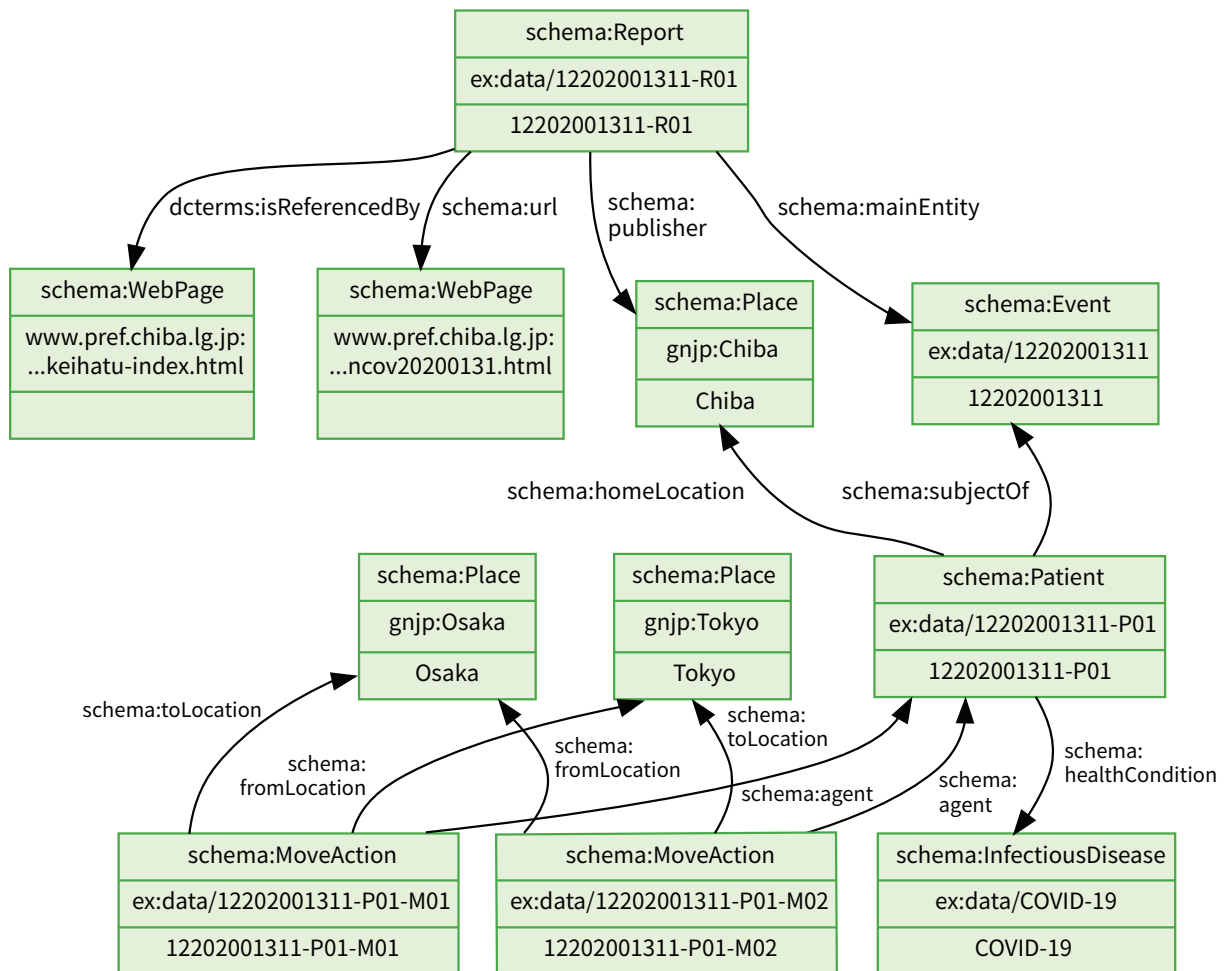


図 2: 患者移動情報の PLOD モデル

● 患者の移動情報

メタデータにはプレスリリースの発信者である政府機関、あるいは自治体の名称、公表日時が含まれる。患者情報には、患者情報にはデモグラフィック属性として年齢、性別、職種、居住地が含まれ、疾患情報として感染症名や感染力の程度等が含まれる。その他の医学的背景情報として、発熱日、確定日、臨床定期経過、既往歴など、今回の感染症に関わる患者の医学的情報が公開される。そのうえで、移動情報の概要や渡航歴が記載される。移動情報には日時あるいは日時の範囲、場所が含まれる。特定の場所に立ち寄っていないなどの否定的情報についてもここに記述される。このように、プレスリリースには、患者の状況が概観できる状況が含まれる。ただし、実際に提供される情報の質や量は、プレスリリースごとに大きく変わる。プレスリリースによっては、とある患者について詳述することもあるが、とある疾患に何名が罹患したかという患者数のみの公開に留まることもある。

患者の移動情報の多くは、下記の要素の組み合わせで表現される。i) 移動した時間、ないし、時間帯; ii-a) 患者が訪問した施設や住所; ないし ii-b) 利用した公共交通機関のリスト、である。ただし、時には、とある場所への訪問を否定する形の情報となっている場合もある。プレスリリースに含まれる患者の移動情報を PLOD モデルにて表現したものを図 2 に示す。この例では、東京から大阪に移動した後、大阪から東京に戻ったケースを示している。

3.2 接触情報のスキーマ

患者や濃厚接触者への聞き取り調査時に、国立研究感染症研究所の提供する調査票³や北海道大学公衆衛生教室の提供する積極的疫学調査調査票楽々活用セット⁴な

³<https://www.niid.go.jp/niid/ja/diseases/ka/corona-virus/2019-ncov/2484-idsc/9357-2019-ncov-02.html>

⁴<http://publichealth.med.hokudai.ac.jp/covid19/covid-19-楽々集計セット>

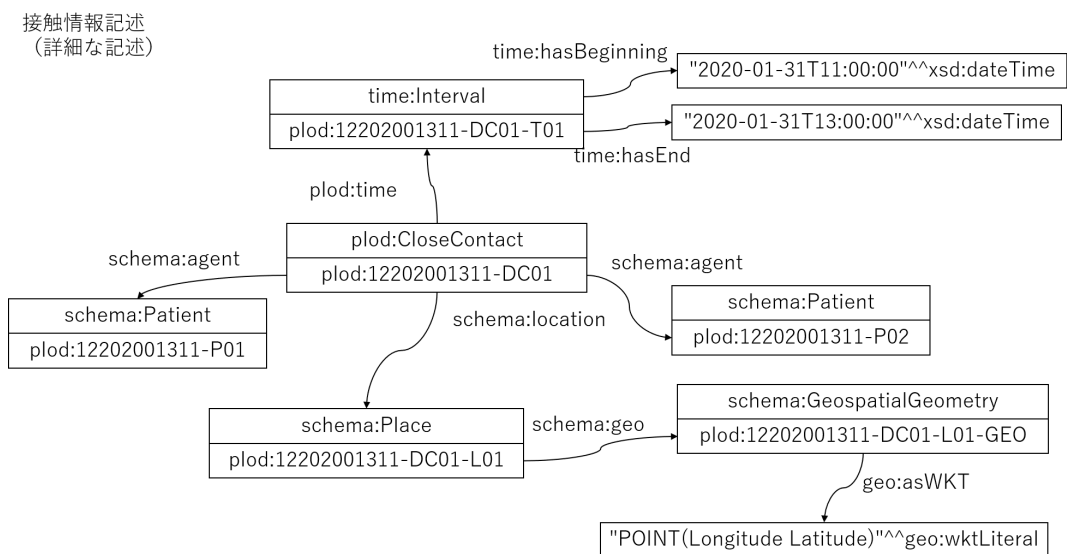


図 3: 接触情報のスキーマ

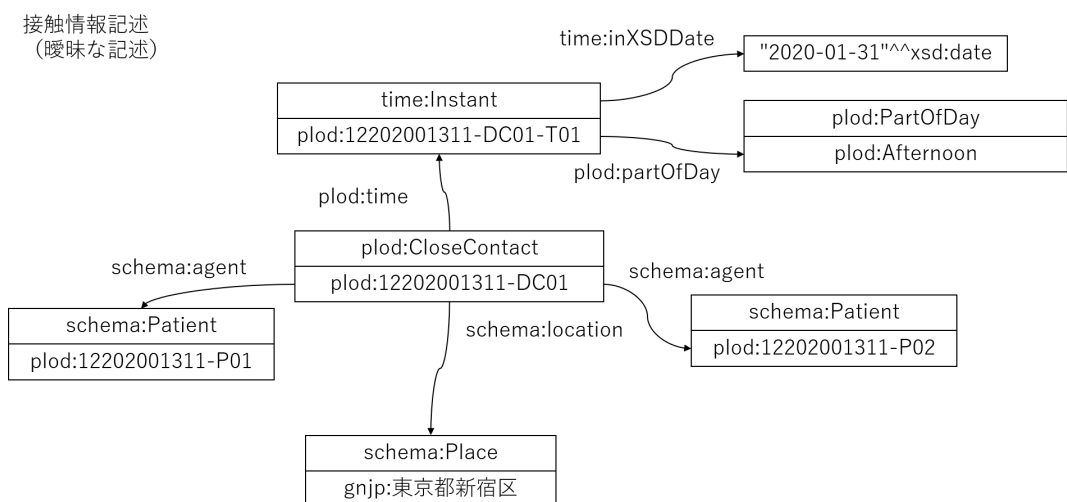


図 4: 曖昧な接触情報のスキーマ

どを利用することで、接触者氏名、接触場所、接触状況、日時情報を詳細に収集することが可能である。こうした直接接触情報のデータはプライバシーに関わるため公開することはできないが、個人が特定できないよう匿名化した上で構造化したデータは、追跡調査に欠かせないリソースである。

本研究では患者接触情報の記述に対応できるように PLOD のスキーマを拡張した。図 3 に接触情報部分のスキーマを示す。聞き取り調査においては接触時間や場所に関して曖昧にしか記憶していない場合は多く、絶対的な時空間情報の記述だけでなく抽象的で曖昧な情報の記述にも対応する必要がある。図 4 は曖昧な接触情報を記述した例である。このように、詳細度の高い接触情報と曖昧な接触情報の両方の記述に対応できる

ようスキーマを設計している。

4 感染リスクオントロジー

我々は、感染リスクの推論を次の三段階に分け、段階的な実現を図っている。本稿では、まず第一段階の推論を可能にするオントロジーの構築を目標とした。

- (1) 第一段階（標語レベル）：対象者のいる場所や行動が、3つの密などの標語で指定されたリスクのどれかに該当することを推論する
- (2) 第二段階（感染経路レベル）：飛沫感染、直接/間接接触感染、エアロゾル感染、媒介物感染などの具体的な感染経路に該当することを推論する

COVID-19 Risk Index

Risk levels for exposure vary based on four main factors:

- Enclosed space**
- Duration of interaction**
- Crowds**
Density of people + challenges for social distancing
- Forceful exhalation**
Sneezing, yelling, singing, and coughing

Low

Walking outdoors
With or without pets

Running or biking
Alone or with another person

Staying at home
Alone or with members of your household

Outdoor picnic or porch dining
With non-household people and physical distancing

Picking up takeout food, coffee, or groceries from stores
Risk: Potential crowding

Low / Medium

Playing "distanced" sports outside
Ex. Tennis or golf

Grocery shopping
Risk: Indoor, close contact, potential clustering of people, high-touch surfaces

Retail shopping



Medium

Visiting hospital emergency department
Risk: Indoor, potential clustering of people

Medical office visit
Risk: Indoor, close contact, potential clustering of people, high-touch surfaces

Dentist appointment
Risk: Indoor, close contact, patient not wearing a mask

Taking a taxi or a ride-sharing service
Risk: Depending on frequency of cleaning, duration of ride, and number of passengers

Outdoor restaurant dining
Risk: Close contact, potential challenges to wear a mask during eating

Museum
Risk: Indoor, close contact, potential clustering of people

Medium / High

Exercising at a gym
Risk: Indoor, close contact/potential clustering of people, high-touch surfaces, difficult to wear a mask, high respiratory rate

Hair/nail salon and barbershops
Risk: Prolonged close contact, difficult to wear a mask

Working in an office
Risk: Indoor, high-touch surfaces, prolonged close contact/potential clustering of people

Indoor restaurant or coffee shop
Risk: Indoor, prolonged close contact/potential clustering of people, high-touch surfaces, while eating and drinking

High

Bars and nightclubs
Risk: Enclosed space, prolonged close contact/potential clustering of people, high respiratory rate, yelling/projection of voice

Indoor party
Risk: Indoor, prolonged close contact/potential clustering of people
Additional risks: alcohol (loss of awareness), shared beverages (coughing)

Playing contact sports
Football, basketball, soccer, etc.
Risk: Prolonged close contact/potential clustering of people, high respiratory rate, unable to wear a mask

Air travel
Risk: Enclosed space, prolonged close contact/potential clustering of people, high-touch surfaces, yelling/projection of voice

Public transportation Subway or bus
Risk: Enclosed space, prolonged close contact/potential clustering of people, and high-touch surfaces

Religious services
Risk: Enclosed space, prolonged close contact/potential clustering of people, high-touch surfaces, yelling/projection of voice

Concert
Risk: Enclosed space, prolonged close contact/potential clustering of people, high-touch surfaces, yelling/projection of voice

Movie theater or live theater
Risk: Enclosed space, prolonged close contact/potential clustering of people, high-touch surfaces

Watching sports
Risk: Prolonged close contact/potential clustering of people, high-touch surfaces, yelling/projection of voice, enclosed spaces (if indoor)

REOPEN INTELLIGENTLY. REOPEN SAFELY.

図 5: COVID-19 Risk Index[3]

| | A | B | C | D | E | F | G | H | I | J | K | L |
|----|----------------------------------|---------------|----------------|--------|-------------------------|---------------|--------------------------------|---------------------|-----------------------|---------------------------------|-----------------------------|---|
| 1 | | | | | | | | | | | | |
| | | Risk | enclosed space | indoor | prolonged close contact | close contact | potential clustering of people | high-touch surfaces | high respiratory rate | unable/difficult to wear a mask | yelling/projection of voice | |
| 5 | Religious services | High | 1 | | 1 | | 1 | 1 | | | | |
| 6 | Watching sports | High | (if indoor) | | 1 | | 1 | 1 | | | | |
| 7 | Indoor party | High | | 1 | 1 | | 1 | 1 | | | | |
| 8 | Air travel | High | 1 | | 1 | | 1 | 1 | | | | |
| 9 | Concert | High | 1 | | 1 | | 1 | 1 | | | | |
| 10 | theater or live theater | High | 1 | | 1 | | 1 | 1 | | | | |
| 11 | Exercising at a gym | Medium / High | | | 1 | | 1 | 1 | 1 | 1 | 1 | |
| 12 | Hair/nail salon and barbershops | Medium / High | | | | 1 | | | | | 1 | |
| 13 | Working in an office | Medium / High | | | 1 | 1 | | 1 | 1 | | | |
| 14 | Indoor restaurant or coffee shop | Medium / High | | | 1 | 1 | | 1 | 1 | | | |
| 15 | Visiting | Medium / High | | | 1 | 1 | | 1 | 1 | | | |

図 6: COVID-19 Risk Index を整理したスプレッドシートの一部

- 第三段階（専門知識レベル）：感染症医学や空間疫学の専門知識やエビデンスとともにリスクを推論する。

場所に関する感染リスクの要素についてまとめた標語としては、厚生労働省の3つの密[18]やWHOのThree Cs[13]が代表的である。また、COVID-19 RECOVERY CONSULTING, LLCは感染リスク指標として図5のチャートを提案している[3]。その他、各所で感染リスクをまとめた指標が提案されている。本稿では前述の指標[3]を参考にオントロジーを試作した。

4.1 オントロジーの試作

本研究では次の設計方針でオントロジー構築した。以降、実際の構築手順とともに説明する。

- 場所と行動の関係をアフォーダンス的に記述する
- チャート内で使用されている語彙を可能な限り3つの密における“密閉”、“密集”、“密接”に対応付ける

まず、図5のチャートから必要な情報を抽出し、リスクのある行動を行名、その行動や場所に関する性質を列名とし、行動ごとにどのような感染リスクがあるか集計できるようにスプレッドシート(図6)上で情報を整理した。整理したスプレッドシートを確認し、デザイン上の問題で記載が不足していると考えられるデータを一部補っている。

次に、整形した表の行名から施設名などの場所の語彙を、列名から行動や場所の性質を表す語彙を抽出し、それぞれクラスとして定義する。例えば、行名から“Bar”や“Gym”といった施設の種類を抽出し、これをクラスとして定義する。

列名から抽出する語彙は一部変換が必要になり、ここでアフォーダンスの概念を導入する。アフォーダンスとは「特定の有機体(群)が特定の環境内に生息している時、その環境の中の特定の対象(群)・事象(群)が、その特定の対象(群)・事象(群)との関係で特定の有機体(群)に対して提供する行為の可能性(opportunities)[17]と定義されている。李らによると、「環境に置かれ

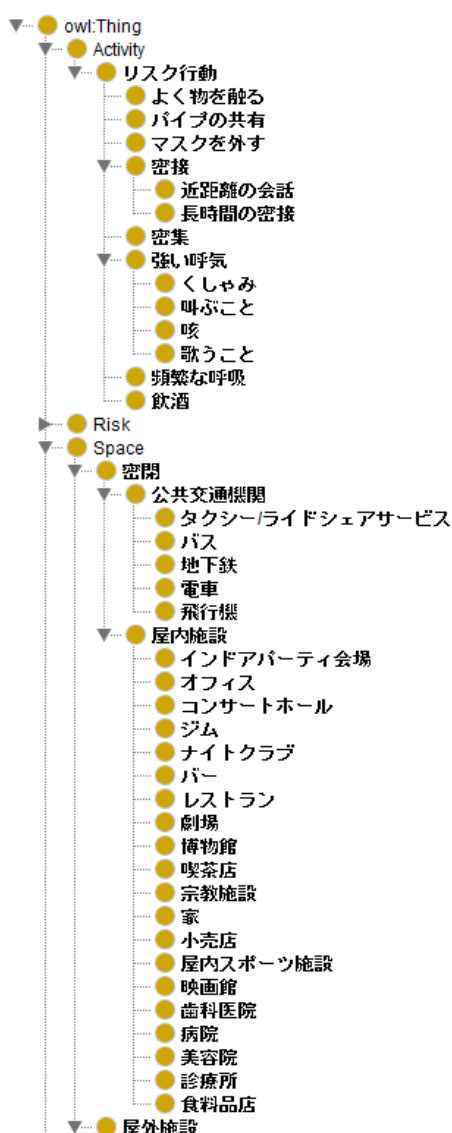


図 7: 感染リスクオントロジーのクラス階層図の一部

た認知体が環境との相互作用によって、環境上で実行できる行為の可能性を与える情報あるいは可能性そのものがアフォードンスである」[20]とされている。本研究においては「店、施設、公共交通機関等の場所が、人間に対して潜在的にどのような感染リスク行動を可能にするのか」という関係性を定義する。

例えば、スプレッドシートの行名となっている“Exercising at a gym (ジムで運動する)”から“ジム”を抽出し、屋内施設クラスのサブクラスとして作成する。列名となっている“unable/difficult to wear a mask (マスクを着用することが困難)”という性質は“Remove a mask (マスクを外す)”という行動に変換し、リスク行動クラスのサブクラスとして作成する。そして、“ジム”が“マスクを外す”という行動をアフォードするという関係性を、ジムクラスが持つプロパティ制約と

して次の記述論理式の通りに定義する。

$$Gym \sqsubseteq \exists \text{afford}.\{\text{remove_a_mask_instance}\} \quad (1)$$

これは、Gym クラスのインスタンスについて、afford プロパティの値の少なくとも一つの値は remove_a_mask_instance であることを意味する。remove_a_mask_instance は Remove_a_mask クラスの唯一のインスタンスである。このように、アフォードンスの概念を導入して場所と行動の関係性について定義する。

図 7 に提案するオントロジーのクラス階層図を示す。場所に関するクラスは Space クラスを最上位としており、行動に関しては Activity クラスを最上位としている。Activity クラスのサブクラスとして感染リスク行動クラスを作成している。提案するオントロジーでは、将来的に感染リスクを「3つの密」の指標で集計することを想定している。そのため、“密閉”を Space クラスのサブクラスに、“密集”、“密接”をリスク行動のサブクラスとして定義している。屋内施設や公共交通機関は密閉クラスのサブクラスと定義し、それらの場所が密集、密接やその他の感染リスク行動をアフォードするという知識を定義している。構築した Web Ontology Language (OWL) ファイルを公開している⁵。

4.2 推論シナリオの検討

「Aさんが2020年4月1日12:00から13:00までレストランBで食事した」という情報が取得できたときに、この情報が三密のどれに該当するか推論するというシナリオを想定する。実際、患者や濃厚接触者への聞き取り調査において3.1節で述べた調査票を使用することで、このような詳細な情報は取得可能である。また、非感染者が感染リスクを把握するため、このようなデータを何らかのアプリから入力するという事例も考えられる。この情報は次の要素を持つイベントデータである。

- 人物：Aさん
- 時間：2020年4月1日12:00から2020年4月1日13:00
- 空間：レストランB
- 行動：食事する

このイベントデータを、提案した感染リスクオントロジーに基づいてナレッジグラフ化すると、図8のように表現することができる。

尚、事前に飲食店名や店舗形態のリストを関係機関から取得できているという想定である。レストランBは(屋内)レストランであるため、“前提として”密閉空間であり、密集行動、長時間の密接やマスクを外す

⁵<https://github.com/PL0D-info/PL0D>

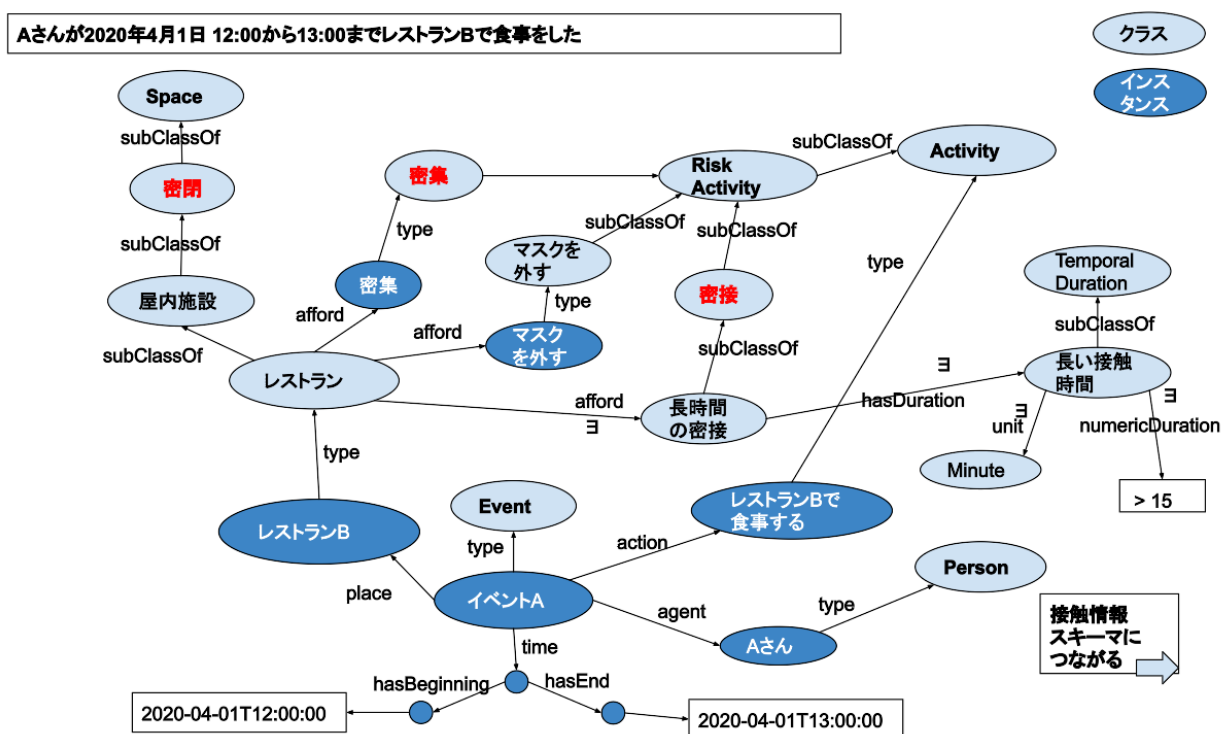


図 8: シナリオデータのナレッジグラフ化

行動の可能性を人に与える。したがって、このイベントは“前提知識に基づく”三密のすべてに該当する感染リスクの高いイベントであることが把握できる。しかし、実際には店舗における換気実施の有無や、食事の時間、密接行動の有無などの個別の状況を考慮する必要がある。例えば、提案した感染リスクオントロジーでは“長時間の密接”は“長い接触時間”を持つと定義しており、また、便宜的に、長い接触時間とは15分以上であると定義している。一方で、Aさんの食事時間は2020年4月1日12:00から2020年4月1日13:00の1時間である。したがって、このファクトデータをオントロジーに当てはめると、Aさんのイベントは長い接触時間を持つ長時間の密接に該当する可能性があると推論できる。このように、個別の状況を考慮するための知識を事前に定義しておくことで、前提知識に基づく一様な推論だけでなく、設定条件に基づく柔軟な推論が可能になるだろう。

5 考察

5.1 患者移動情報のオープンデータ化の考察

現時点でのスキーマ仕様は、実アプリによる検証を経ていないドラフト水準のものであり、実運用までにはいくつかのステップが存在する。本稿でも言及した

ように、プレスリリースにはさまざまな種類の情報が記載されており、時折、現行のPLODにて表現することが困難なイレギュラーな情報が含まれることにある。そのために、今後、実際の活用を通じて、仕様自体をブラッシュアップしていく必要がある。一方で、プレスリリース側自体も、標準化に向けて改善されていく必要があるだろう。こうしたPLODとプレスリリース双方の発展に向けては、現行のPLOD仕様の表現力が限定されているとしても、まずは稼動する実験サービスを実現することが重要と考えられる。そうした実験サービスの活用を通じて、仕様とアプリケーションの双方が改善され、現段階ではクリアとなっていない様々な実用上のニーズへと応えていくことが可能となる。

5.2 感染リスクオントロジーの考察

本研究では、まず標語レベルでの感染リスクの推論に向けてオントロジーを試作したが、実運用までにはいくつかのステップが存在する。4.2節で例示した推論を幅広く可能にするために、個別の状況記述に関する語彙を洗い出して統制しておく必要がある。一方で、実データにはイレギュラーな情報や未知語が混在していることが想定されるため、患者移動情報と同様に実際の活用を通じて、仕様自体をブラッシュアップしていく、拡張性・柔軟性を付与する必要がある。

また、聞き取り調査やアプリケーションから得られる情報をオントロジーのクラスと対応付けるステップが必要になる。調査票の時点で項目が細分化されていると、必要な情報（インスタンス）を抽出して該当するクラスに分類することが容易である。しかしながら、項目の過度な細分化は現場の負荷上昇につながるため、聞き取り調査では文章による記述が多くなることも想定される。したがって、自然言語処理技術を用いたキーワード抽出やクラス-インスタンス間のマッチング処理が必要である。

本研究では、厚労省が掲示している3つの密と、COVID-19 RECoVERY CONSULTING, LLCのCOVID-19 Risk Indexチャートを参考にオントロジーを試作したが、他に、内閣官房が掲示している「5つの場面 [19]」やINFORM Covid-19 Risk Index[12]など、異なる標語や指標が存在する。さらにそれらの指標ごとに感染リスクの定義が異なっている。特に、国や地域により施設環境、人口密度や行動文化は異なるため、感染リスクの定義のローカライズが必要であると考えている。

6 おわりに

本稿では、我々のこれまでの取り組みとして、感染症患者の移動情報に関するプレスリリース情報を、RDFを用いてオープンデータ化する枠組みを紹介した。また、接触情報記述に対応するためスキーマを拡張したことを報告した。さらに、場所に関する感染リスクの推論を可能にするオントロジーを試作し、実運用に向けたシナリオの検討と考察を行った。

現在、我々は自治体との調整を進めており、生に近いデータの提供を受けて、PLODスキーマや感染リスクオントロジーの実環境での検証を目指している。オープン化に関しては匿名化プロセスが必要であり、他の公開データとの組み合わせでも個人が特定できないよう、細心の注意を払う必要がある。まずは内部利用にフォーカスし、オントロジーやナレッジグラフなどのセマンティック技術による情報統合と推論処理により、保健所や関連機関に集中する負荷を軽減することを直近の課題としたい。

謝辞

本研究は、AMEDの課題番号JP20he0622042の支援を受けた。

参考文献

- [1] Shane Babcock, Lindsay G Cowell, John Beverley, and Barry Smith. The infectious disease ontology in the age of covid-19. 2020.
- [2] Daniel Daniel Domingo-Fernández, Shounak Baksi, Bruce Schultz, Yojana Gadiya, Reagon Karki, Tamara Raschka, Christian Ebeling, Martin Hofmann-Apitius, and Alpha Tom Kodamullil. COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology. *Bioinformatics*, 2020.
- [3] Dr. Ezekiel J. Emanuel, James P. Philips, and Saskia Popescu. COVID-19 Activity Risk Index. <https://www.covid19reopen.com/resources/covid-19-daily-activity-risk-index>.
- [4] Yongqun He, Hong Yu, Edison Ong, Yang Wang, Yingtong Liu, Anthony Huffman, Hsinhui Huang, John Beverley, Junguk Hur, Xiaolin Yang, et al. Cido, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis. *Scientific Data*, Vol. 7, No. 1, pp. 1–5, 2020.
- [5] Eun-Young Jeong. South Korea Tracks Virus Patients' Travels—and Publishes Them Online. <https://jp.wsj.com/articles/SB12291155354026644516304586207690851702666>.
- [6] Kerina Helen Jones, Helen Daniels, Sharon Heys, and David Vincent Ford. Toward an ethically founded framework for the use of mobile phone call detail records in health research. *JMIR mHealth and uHealth*, Vol. 7, No. 3, p. e11969, 2019.
- [7] KH Kim, TE Tandi, Jae Wook Choi, JM Moon, and MS Kim. Middle east respiratory syndrome coronavirus (mers-cov) outbreak in south korea, 2015: epidemiology, characteristics and public health implications. *Journal of Hospital Infection*, Vol. 95, No. 2, pp. 207–213, 2017.
- [8] Liza Lin. China Marshals Its Surveillance Powers Against Coronavirus. <https://www.wsj.com/articles/china-marshals-the-power-of-its-surveillance-state-in-fight-against-coronavirus-11580831633>.

- [9] Franck Michel, Fabien Gandon, Valentin Ah-Kane, Anna Bobasheva, Elena Cabrio, Olivier Corby, Raphaël Gazzotti, Alain Giboin, Santiago Marro, Tobias Mayer, et al. Covid-on-the-web: Knowledge graph and services to advance covid-19 research. In *International Semantic Web Conference*, pp. 294–310. Springer, 2020.
- [10] Ikki Ohmukai, Yasunori Yamamoto, Maori Ito, and Takashi Okumura. Tracing patient plod by mobile phones: Mitigation of epidemic risks based on patient locational open data. In *Proc. of the 29th IEEE International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), Web2Touch Track*, p. to appear. IEEE, 2020.
- [11] Takashi Okumura. Tracing infectious agents with mobile location information: A simple and effective countermeasure against epidemic risks. In *2019 IEEE Global Humanitarian Technology Conference (GHTC 2019)*. IEEE, 2019.
- [12] Karmen Poljansek, Luca Vernaccini, and Montserrat Marin Ferrer. INFORM Covid-19 Risk Index. <https://ec.europa.eu/jrc/en/publication/annual-reports/inform-covid-19-risk-index>.
- [13] World Health Organization Western Pacific Region. Avoid the Three Cs. <https://www.facebook.com/whowpro/posts/avoid-the-three-cs-be-aware-of-different-levels-of-risk-in-different-settings-co/1587901598047596/>.
- [14] Bram Steenwinckel, Gilles Vandewiele, Ilja Rausch, Pieter Heyvaert, Ruben Taelman, Pieter Colpaert, Pieter Simoens, Anastasia Dimou, Filip De Turck, and Femke Ongenaes. Facilitating the analysis of covid-19 literature through a knowledge graph. In *International Semantic Web Conference*, pp. 344–357. Springer, 2020.
- [15] Changhong Yang, Jun Yang, Xiangshu Luo, and Peng Gong. Use of mobile phones in an emergency reporting system for infectious disease surveillance after the sichuan earthquake in china. *Bulletin of the World Health Organization*, Vol. 87, pp. 619–623, 2009.
- [16] 厚生労働省. 新型コロナウイルス接触確認アプリ (COCOA) COVID-19 Contact-Confirming Application. https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/cocoa_00138.html.
- [17] 佐々木正人, 鈴木健太郎, 三嶋博之, 篠原香織, 半谷実香, 李銘義, 公文誠, 足立紀彦. 複雑系の科学と現代思想アフォーダンス複雑系の科学と現代思想アフォーダンス, 1997. *BME*, Vol. 12, No. 7, pp. 57–68, 1998.
- [18] 首相官邸. 3つの密を避けるための手引き. <https://www.kantei.go.jp/jp/content/000062771.pdf>.
- [19] 内閣官房新型コロナウイルス感染症対策推進室. 感染リスクが高まる「5つの場面」. <https://corona.go.jp/proposal/>.
- [20] 李銘義, 公文誠, 足立紀彦. 強化学習ロボットによるアフォーダンスの利用. *人工知能学会論文誌*, Vol. 16, No. 1, pp. 94–101, 2001.