Tracing patient PLOD by mobile phones
Mitigation of epidemic risks based on patient locational open data

Ikki Ohmukai
The University of Tokyo, Japan
i2k@l.u-tokyo.ac.jp

Yasunori Yamamoto
Research Organization of Information and Systems, Japan
yy@dbcls.rois.ac.jp

Maori Ito
PLOD info, Japan
maorinphone@gmail.com

Takashi Okumura
National Institute of Public Health, Japan
Kitami Institute of Technology, Japan
tokumura@mail.kitami-it.ac.jp

Abstract—**In the cases when public health authorities confirm patients with highly contagious diseases, they release summaries about patient locations and travel information to alert the public. Due to privacy concerns, these releases do not include the detailed data, and typically comprise the information about commercial facilities and public transportation used by the patients. This paper proposes a framework to represent the information in Resource Description Framework (RDF). The open data would then be used by residents for automated estimation of the potential risks of contacts, combined with the location information stored in their mobile phones.**

Index Terms—**public health preparedness, infectious diseases, location, linked open data**

I. INTRODUCTION

In the cases when public health authorities confirm patients with highly contagious diseases, such as measles, tuberculosis, and the novel coronavirus, they have to identify close contacts of infected patients to monitor their health conditions for a certain period [1]. This *contact tracing* is mostly performed manually by officials. However, concerning the usage of public transportation and visiting public facilities, the officials are unable to trace all possible contacts.

To mitigate the potential risks associated with such situations, public health authorities release the reports on the travel information of the infected patients, expecting potential contacts to report themselves. Nevertheless, due to privacy concerns, authorities cannot provide the detailed information about the patients, and therefore, these releases typically comprise the data related to commercial facilities and public transportation used by the patients. Moreover, public health authorities are not capable of efficiently alerting the potential contacts about the risk associated with the infectious agents, who are mostly unconcerned.

We consider that efficient sharing of the data on the patient location information can be used to bridge this gap by personalizing the alerts for each resident through the automated estimation of potential intersections between the resident and the infected patients based on the location history stored in their own mobile phones. Such approach can be helpful, particularly, when patients have visited various places, and when the number of patients has increases to a certain

threshold exceeding the capacity of ordinary people to process the trace information corresponding to all patients.

The proposed open data framework for tracing the patient information is illustrated in Figure 1. First, a patient visits a medical institution and a physician confirms the infection case (1). Secondly, the physician reports to a regional public health center (2). Third, the officials perform the necessary investigation (3) and publicly alert the information by an announcement (4). In the proposed scheme, this announcement is released as the open data established in a form of the Patient Locational Open Data (PLOD). These data can be publicly used to perform mapping of patient traces and estimate the risks of potential contacts for individuals (5). To perform this estimation, residents can employ the location history information on their mobile phones. A suspicious contact is supposed to be reported to the public health authority with the purpose of obtaining further instructions (6).
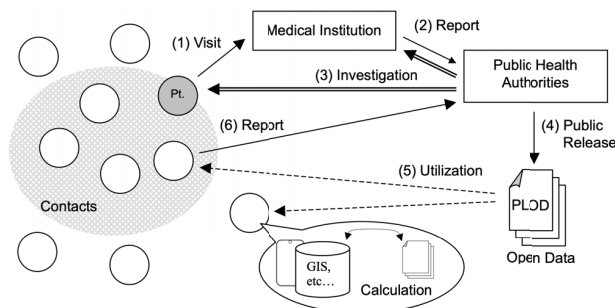

Fig. 1. Overview of the proposed approach

II. RELATED WORKS

Utilization of the mobile location information for the public purposes was initiated in the late 2000s, when mobile phones became commonly available across the world [2]. It was deemed as a straightforward approach to trace patient by analyzing the location information provided by mobile devices. However, the trace information is considered as the privacy data, and there are ethical concerns even considering public health applications [3].

At the same time, the research works on the statistical utilization of the mobile location information have gained
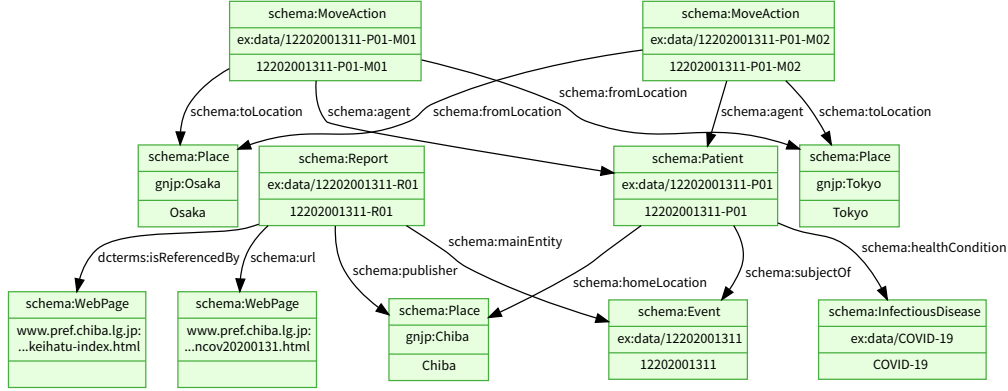
Fig. 2. RDF model of PLOD

wide popularity, as mobile phones produced a vast amount of the location information relevant for the improvement of population health [4]. In this regard, the statistical utilization of the mobile location information to control infectious diseases has been considered as a promising approach [5]. In the related studies, researchers basically used the anonymized or statistically processed data.

In 2020, the situation has drastically changed due to the pandemic of the new coronavirus infectious disease (COVID-19). In China, the government authority have deployed the facial recognition technology, security cameras, and social media analysis combined with the manual surveillance to automate the contact tracing with regard to COVID-19 patients [6]. In South Korea, public health authorities have monitored the credit card history, security camera footage, location information obtained from mobile devices, public transportation cards, immigration records, etc., to track individuals exposed to the risks of possible infection [7]. In Taiwan, the subjects of the home quarantine have been monitored by the authority, reportedly, through analyzing mobile phone signals [7]. Such applications of processing the location information with the purpose of crisis management of infectious diseases have been rarely published as research articles before the novel coronavirus pandemic in 2020, except for the cases of patient tracing in South Korea [8] and a proposal on a risk assessment service of potential infections in Japan [9].

The challenge associated with the knowledge representation of the information about infected patients and their location data lies in the fact that such approach should represent both the qualitative information corresponding to attributes and conditions of patients and quantitative information on the location, such as coordinates, distance, and range. To address the problem, the World Wide Web Consortium (W3C) proposed a framework that can consistently handle qualitative and quantitative information in terms of geospatial knowledge called GeoSPARQL [10].

III. DESIGN

In this section, we propose a structured representation of the press releases appropriate to publish the information as PLOD.

The press releases representing the current situation about infected patients usually comprise the following information formulated in natural language, which is not suitable for automated processing.

- Issuer and contact information of the document;
- Patient's attributes (age and gender);
- Name of the disease (infected agents) and infectivity;
- Other medical background;
- Clinical course of the patient;
- Travel information.

The metadata include the name of the government agency or local government, as a publisher of the releases, date and time of the publication. The patient information includes demographic attributes such as age, gender, occupation, and place of residence, as well as the information on the disease, such as the infectious agents, and the degree of infectivity. Other medical background and clinical course include the date of onset, confirmed date, and past medical history. Outline of the travel history might be included here as well. The releases contain the information to provide an overview on the patient situation; however, the amount of the information included in a particular release may vary. It may describe a single case or represent only the number of patients infected by a certain disease.

The travel information of patients is mostly composed of a combination of the following items: i) Travel time or range; ii-a) Name of the facility, or address visited by the patient; or ii-b) The list of public transportation used. In some cases, it may come in a negation form due to denying a visit of a specific place. Figure 2 represents an example of a press release that outlines the two travels of a patient: from Tokyo to Osaka, and from Osaka to Tokyo, as described in the PLOD model.

To model the information, unique and persistent identifiers (PIDs) are required. In the proposed model, each entity has a persistent ID. An event entity is automatically generated based on the press release ID so that all elements are connected through the event entity, even when it is not explicitly mentioned in the press release. Patient IDs are constructed based on the event ID adding a unique number of a the patient.

```
1  <12202001311-P01> a schema:Patient ;
2    schema:healthCondition <COVID-19> ;
3    plod:dateConfirmed "2020-01-31" ;
4    foaf:age "20s" ;
5    schema:gender "Female" ;
6    schema:homeLocation gnjp:Chiba .
7
8  <12202001311-P01-M01> a schema:MoveAction;
9    schema:agent <12202001311-P01> ;
10   schema:startTime "2020-01-16" ;
11   schema:endTime "2020-01-16" ;
12   schema:fromLocation gnjp:Tokyo ;
13   schema:toLocation gnjp:Osaka ;
14   schema:instrument "Airplane" .
```

Fig. 3. Example of PLOD (excerpt)

To identify the travel information, their IDs are generated by adding an incremental number in the travel history associated with a patient ID. The PLOD URIs are determined uniformly from these IDs by using the HTTP scheme (https://), domain name (plod.info), and path information (/data/) as a prefix.

Most of the properties corresponding to each entity employ Schema.org. Missing attributes in Schema.org are originally defined with the prefix "ex:". Figure 3 represents the raw data of the PLOD model using the turtle format.

IV. ANALYSIS

A. Analysis of press releases for infectious diseases

To verify the appropriateness of the patient information described in the PLOD model, we collected the press releases corresponding to infectious disease patients published by the local governments in Japan. The survey results included the press releases of 189 measles cases and 39 COVID-19 ones starting from 2018 through 2020, in total, 228 cases published by the Ministry of Health, Labor, and Welfare, as well as prefectures and cities. We then categorized the documents according to the way used to describe the case trace information.

Among the 228 releases, 71 described the trace in the independent sections, while 9 releases mentioned the trace in the bodies of the announcements. Another 80 releases included the trace information using generic terms and avoided providing the specific location. Then, 16 releases notified that there were no risks of public exposure by the cases, and 45 releases did not provide any detailed information of the case. In sum, among of 228 releases, 224 contained the case descriptions and 80 included the trace information. The format of the press releases and the contents were not uniform: some of them included the names of specific places, and others did not present the information sufficiently detailed to trace the patients.

The survey is incomplete in the sense that the analysis has been conducted only based on the releases found by using search engines, and there are releases that can be found only in archives of the local governments. Nevertheless, it is highly probable that the press releases corresponding to the patients infected by contagious agents do not have a single standardized format. In this regard, announcements by public health authorities in English-speaking countries are deemed to have similar characteristics, concluding from the twelve announcements we found by using search engines. The tentative conclusion can be summarized as follows: the information provided by public health authorities is not standardized, and it is considered meaningful to propose a framework for the unified representation of the actual information about infected patients.

B. Evaluation of the proposed model

Next, we evaluate the expressive power of the proposed PLOD model based on the press releases that contain the detailed trace information about patients. For the evaluation, we analyzed 39 press releases issued in the period from Jan 16, 2020 to Feb 11, 2020 and arranged the available information according to the PLOD model. The metadata of the press release itself, the patient information, and the patient trace information were successfully expressed for all cases by using PIDs issued for each entity considered as a Subject. However, we confirmed that the information corresponding to Objects became empty for the cases in which the detailed data were not provided in the corresponding release. Among the considered 39 press releases, 23 included the patient information, and 8 out of those 23 represented the trace information of patients, including date and location.

To estimate the contact risk using the PLOD model, it is essential to include the origin of a trace or a destination, preferably, together with the detailed information about the visited places. To obtain the detailed description of the trace information, we extracted the steps from each case report, and 21 steps were identified on the basis of the 8 releases. Among these steps, 10 included the origin or destination data with varying granularity: 8 were described at the prefecture level, 1 as "oversea", and only 1 contained the facility-level specific information that could be converted into the coordinates (latitude and longitude).

We identified the four patterns in the trace information that could not be expressed appropriately in PLOD. First, the releases included a negation of a specific area, such as "overseas". Second, several included cases of the profession information, such as "a bus guide" and "a driver", indicated the possibility of trips but not did not specify it in a manner sufficient to trace the case. Figure 4 illustrates a case corresponding to this category. Third, there were the releases that indicated traces in generic terms, for example, "returning home with a child", which could not be converted into the coordinates. Lastly, several releases referred to the other press releases that could contain the trace information, such as "same as past case A". It is hardly possible to convert the first and the second cases into the coordinates, whereas the third and the fourth cases might be processed to obtain the required data. For example, we may infer that the "home" is supposed to be in the city where the press release has been issued, and, similarly, the reference in the fourth case might be appropriately processed. Both cases can be organized as the best fit for the SPARQL-

```
1 <29202001281-P01-M01> a schema:MoveAction
    .
2     schema:agent <29202001281-P01> ;
3     schema:startTime "2020-01-08" ;
4     schema:endTime "2020-01-11" ;
5     schema:fromLocation "" ;
6     schema:toLocation "" ;
7     schema:instrument "Bus"  .
```

Fig. 4. Example of a failure case

based queries that can infer the information even if it is not explicitly stated.

V. DISCUSSION

An indispensable part of the proposed approach is to develop the mobile applications that would use the open data to raise customized alerts for users. Mobile phones typically have a built-in application programming interfaces (API) for location services, which responds the position data of the mobile phone. Accordingly, mobile applications may periodically call the API to record their positions, which can be later used to estimate the intersections with the patients mentioned in the open data. Once the application finds that a patient enters in a predefined range corresponding to the phone, it may alert the owner of the phone.

The proposed application can be implemented in several ways. It could be developed as an independent application or as an embedded module in an application that uses a location service, such as applications for crisis management provided by local governments. As the proposed framework is based on the RDF, users or third parties can easily add the information relevant according to their needs. Organizing a hackathon is an approach to develop a variety of applications suitable for various usage options. As the public health information concerning infectious diseases may attract attention of a considerable number of people, the proposed application can be widely installed, once the application becomes available for the public use.

At the moment, the proposed framework is at a draft stage, and there could be steps required to implement before the public use becomes possible. As mentioned above, public releases have a variety of formats, and they often contain the irregular data, which are hardly possible to express in the current version of PLOD. To address the problem, continuous improvement for the specification is necessary, however, the formats of press releases need to be also standardized and updated. To facilitate the process of creating a unified data representation, it would be preferable to launch the service, even with the limited quantity in terms of releases and the limited quality in terms of expressive power. Then, during the actual use, the specification and application would evolve to meet the real needs.

VI. CONCLUSION

In this study, we have proposed a framework to represent information press releases focused on the current situation on infected patients in a form of the open data using RDF. To the best of our knowledge, this is the first attempt of applying the RDF and open data technology to mitigation of epidemic risks. Due to the ongoing pandemic caused by COVID-19, we consider that the proposed framework has a potential to be used worldwide.

The future research work will include the following three steps. First, we plan to collect the public releases, convert them into the open RDF data format, and distribute them aiming to facilitate the development of applications. This could serve as a demonstration for stakeholders to evaluate the utility of the open public health data concept. Second, the proposed framework needs to be supported by the authorities, such as the ministry of health or a national institution, in a centralized manner. Third, local governments are expected to initiate providing the data in the linked open data format directly in a decentralized manner.

In the field of epidemiology, public health fields have been using information technologies mostly for the purposes of statistics and prediction; however, they have rarely considered using such methods for prevention of diseases. Moreover, the concept of open data in the medical field has not been considered widely, and there exist only a limited number of publication dedicated to this subject. We envisage that the proposed approach may serve as a foundation for development of knowledge processing applications in public health, such as alerting about hospital overload under pandemic conditions.

REFERENCES

[1] S. A. Rasmussen and R. A. Goodman, *The CDC Field Epidemiology Manual*. Oxford University Press, 2018.

[2] C. Yang, J. Yang, X. Luo, and P. 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.

[3] K. H. Jones, H. Daniels, S. Heys, and D. V. 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.

[4] B. Ouedraogo, J. Gaudart, and J.-C. Dufour, "How does the cellular phone help in epidemiological surveillance? a review of the scientific literature," *Informatics for Health and Social Care*, vol. 44, no. 1, pp. 12–30, 2019.

[5] J. Chirombo, P. Diggle, D. Terlouw, and J. Read, "A review of models of human mobility for predicting infectious disease spread," *Modelling spatial processes of infectious diseases*, p. 60, 2018.

[6] L. 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.

[7] E.-Y. Jeong, "South Korea Tracks Virus Patients' Travels—and Publishes Them Online," https://jp.wsj.com/articles/SB12291155354026644516304586207690851702666.

[8] K. Kim, T. Tandi, J. W. Choi, J. Moon, and M. 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.

[9] T. 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.

[10] R. Battle and D. Kolas, "Geosparql: enabling a geospatial semantic web," *Semantic Web Journal*, vol. 3, no. 4, pp. 355–370, 2011.