# PROJECT DESCRIPTION: HotZone

In a previous semester's project, students developed a pilot application to assist the Centre for Health Protection (CHP) with data management and contact tracing for COVID-19 cases. Following that success, our project for this semester will tackle another problem for the CHP.

Background:

During Hong Kong's current SARS-CoV-2 epidemic, CHP's epidemiologists have focused on the histories of contacts between cases to better understand transmission of the virus. Similar research identified clusters of infections based on reported contacts and determined transmission chains within them. For example, 3 large clusters were discovered that were linked to: (a) four bars across Hong Kong, (b) a wedding and a preceding social event, and (c) a temple. The clusters and transmission chains within the clusters are shown below:

Adam, et al. Sept. 17, 2020
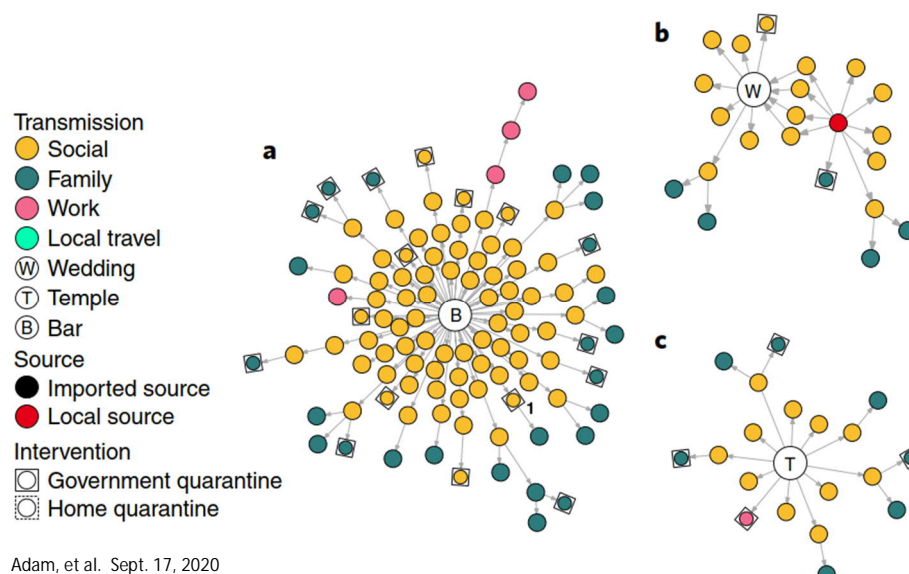
Identifying clusters based on contact tracing is extremely time consuming and may miss clusters associated with environmental transmission. Environmental transmission was important in, for example, the SARS-CoV epidemic in 2003 - many cases were found to be concentrated in certain locations such as Amoy Gardens in Ngau Tau Kok, where the spread of the virus was the result of bad drainage design rather than direct person-to-person contact.

CHP wants to identify clusters of cases linked by location to get early warning of possible hot-zones of infection. This will allow CHP to alert the public about such zones, and to take action to investigate and contain the spread of the virus in those zones. Currently CHP records and manages all data in Excel spreadsheets and has no means to cluster cases by locations visited. CHP needs an application, *HotZone*, to support management of data related to locations visited by known cases, to identify clusters, and to visualise the clustering results. CHP intends the application to become its standard tool for identifying hot-zones in the event of future epidemics of SARS-CoV-2 and other viral infections. CHP also wants to use the application to analyse data from previous viral epidemics to find clusters that were not previously identified.

CHP's priorities for the new application are: (a) to improve the quality and efficiency of location data recording, and (b) to perform clustering. Other features, including those supporting data entry for other case-related data, are felt to be straightforward and/or less urgently required.

Your individual project will take *HotZone* through Inception and the first iteration of development of Release 1. You will deploy your product increment at the end of the iteration. The objective of the individual project is for you to gain hands-on experience preparing typical project deliverables, demonstrate that you can obtain and use data from Hong Kong Government Geodata Store datasets, and acquire a working knowledge of the Django framework.

Subsequently, in your group project, your team will complete Release 1 in a single additional iteration and then develop and deploy Release 2 in a further iteration.

To make it is feasible to develop *HotZone* as a course project, many aspects of the application and its data have been grossly simplified. Don't be concerned by that.

Summary of information collected in initial interviews:

Some initial interviews have been conducted with representative stakeholders and are summarized here.

a) CHP Needs

Maintain records of each confirmed case of a viral infection of interest. Any member of CHP staff registered with *HotZone* may access and/or modify such records. Categories of data maintained for each case will correspond to those currently maintained manually in spreadsheets (see sample). Thus, CHP staff must be able to record and maintain for each case:

- Case Number (unique for a particular Infecting Virus),
- Patient,
- Infecting Virus,
- Date (on which the case was confirmed),
- Local or Imported.

Note: As we have seen for COVID-19, a patient diagnosed with the disease can recover, be confirmed as virus-free, and then later be re-infected by the same virus. CHP gives a different case number to each episode of infection. Thus, an individual person could be the patient for multiple cases, including multiple cases involving the same Infecting Virus.

- Data maintained for each patient:

  - Name,
  - Identity Document Number,
  - Date of Birth.

- Data maintained for each Infecting Virus:
  - Name,
  - Common Name (of associated disease),
  - Max. Infectious Period (days that a case with this virus could have been a transmission risk prior to being confirmed as infected and isolated. Example: for SARS-CoV-2 the period is 14 days).

- For each case, maintain records of locations visited by the patient during the Infectious Period. Again, the data categories will correspond to those currently maintained manually. Thus the application must record, for each confirmed case:

  - Name of Location Visited,
  - Address (if listed in Geodata dataset),
  - HK1980 Grid Coordinates,
  - Dates between which the case's patient was present at the location,
  - Category of the location with respect to the patient (currently one of Residence, Workplace, Visit).

- Query the Hong Kong GeoData Store using a location name supplied by the user and obtain location data for locations not already known to *HotZone*.

- Cluster cases based on geographical proximity of locations visited by patients within some specified time period of one another. Allow the user to specify the values of the clustering criteria (proximity and period).

- Visualise clustering results by superimposing cases, clusters and related data on Imagery Map or Topographic Map tiles obtained from the Hong Kong GeoData Store.

- Control access to *HotZone* through user authentication based on username and password.

- For each member of CHP staff registered with *HotZone*, maintain:

  - username,
  - password,
  - CHP staff number (7-digit string),
  - first name,
  - last name,
  - email address.

b) CHP Staff User Needs

- After being registered as a user by admin and allocated a username and password, be able to log in and access *HotZone* functions.

- Be able to change password.

- In the case of a forgotten password, be able to request to reset it (by, for example, having a token/link for password reset sent to their email account).

c) CHP Admin Staff

- Admin staff feel their needs are already met by Django Admin and they do not require any dedicated UI beyond that. That is, they can accomplish user registration and, where necessary, add and manage data directly through Django Admin.

- Have users' requests to change password or to reset lost passwords handled by the application without requiring Admin action.

d) Sample of case data

The following table shows an example of data currently maintained (poorly) for a case and the locations visited by the case's patient. Data are recorded only in English:

Case Number	Date Confirmed	Local or Imported
5094	29-09-2020	Local

Patient Name	Identity Document Number	Date of Birth
Chai, Wun Ching	D899336(6)	12-08-1974

Virus Name	Disease	Max. Infectious Period (days)
SARS-CoV-2	COVID-19	14

Location	Address	X Coord	Y Coord	Date From	Date To	Category
Hang Yue Building	334-350 Des Voeux Road West	832271	816470	15-09-2020	29-09-2020	Residence
St. Clare's Primary School	---	832498	816065	15-09-2020	29-09-2020	Workplace
Newman House	35-45 Johnston Road	835656	815343	19-09-2020	19-09-2020	Visit
Good Cheer Villa	16 Stanley Mound Road	839960	809236	20-09-2020	20-09-2020	Visit
Hau Fook Building	4-8 Yat Fu Lane | 562-570 Queen's Road West	831818	816382	27-09-2020	27-09-2020	Visit
Good Cheer Villa	16 Stanley Mound Road	839960	809236	27-09-2020	27-09-2020	Visit

e) Additional details for Location data *(Location Visited, Address, X Coord, Y Coord):*

- *HotZone* location descriptors are those used in the Hong Kong Government Geodata Store datasets. *HotZone* will interface with the Geodata Location Search API to allow CHP staff to search for and obtain these data directly. Details of the API are here:

    `https://geodata.gov.hk/gs/locationSearchAPI`

- As seen in the samples, some locations do not have an Address recorded in the datasets. For example, the JSON-formatted data obtained by querying Location Search with the string *St.+Clare%27s+Primary+School* is:

```
[{"addressZH":"","nameZH":"聖嘉勒小學","x":832498,"y":816065,"nameEN":"St. Clare's Primary School","addressEN":""}]
```

- *HotZone* will not maintain duplicate records of any Location. A new Location record is added only if it does not already exist in the system. Thus, if several cases' patients have visited *Newman House*, the visits will all link to its single location record.

f) Additional details for Visit data *(Date From, Date To, Category):*

- A visit represents any significant time spent by a patient at some Location. For most visits, the *Date From* and *Date To* will be the same indicating that the visit took place during a single day. Multiple Locations might be visited on the same day, as shown for *Hau Fook Building* and *Good Cheer Villa*.

- Repeat visits to the same Location on different days are recorded separately, as can be seen for *Good Cheer Villa*. An exception is made for Residences and Workplaces where the patient could be expected to spend considerable time over multiple days. In those cases, *Date From* will record the date of first visit and *Date To* will record the date of last visit during the period of interest.

g) Order of development to satisfy CHP's required release schedule

- Release 1 of *HotZone* will support the retrieval of location data to create the records of locations visited for a case. It will be developed in two iterations. The first iteration will handle the simple case where there is a single GeoData location that satisfies the search criteria and the location is not already known to *HotZone*. The second iteration will handle the situation where GeoData Location Search returns multiple locations and the user is required to select one of them. It will also handle the situation where there is a matching location already known to *HotZone*. In addition, the second iteration will implement basic authentication.

- Release 2 will add clustering.

- Release 3 will add cluster visualisation and all remaining features.

h) General assumptions

You may:

- assume that when an infection is confirmed, the patient is immediately isolated and, after that date, there will be no further visits related to that case;
- assume that users act in a disciplined way. For instance, CHP staff will never create duplicate records for the same case;
- assume that when a user updates data, there is no need to reflect that change immediately to other users who, for example, may be viewing the data;
- ignore the possibility of system or network failure and downtime.

i) Technical issues

- *HotZone* will be implemented in Python on Django 3.1 (any release: 3.1.1, 3.1.2, etc.);
- There is no constraint on choice of OS;
- You will develop locally, running on the Django default development server. You may use the SQLite DBMS if you wish, however at the end of each iteration in both the individual and group projects, you must switch to use PostgreSQL as DBMS and deploy to a cloud PaaS. You are, of course, free to install PostgreSQL locally and use it throughout development. You will be guided through the processes of switching to PostgreSQL and deploying;

- Again for simplicity during development, you may use the default development server with Django defaults to serve static assets if you need them (CSS, JavaScript, etc.). For production you'll use a more suitable approach to serve such assets and will be guided through that process too;

- You may assume you are free to use any of Django's built-in resources and third-party applications to implement *HotZone*. In fact, you are encouraged to do so. *djangopackages.org* and *PyPI* are good starting points when looking for third-party solutions – there are solutions for pretty much everything. At the start of the group project, we'll offer a tutorial covering the library and the method you must use for clustering.

j) Timeline

- Individual Project (Inception + Sprint 1 [Release 1, Iteration 1]):
  Begins: Thurs, Oct 1.
  Ends: Wed, Oct 28.

- Group Project (Sprint 2 [Release 1, Iteration 2] + Sprint 3 [Release 2]):
  Begins: Mon, Nov 2.
  Ends: Sun, Nov 29. (Release 1 delivered Sun, Nov 15.)