

# Team 0 – D&A

## Project – Phase I

Aditya Nair (2023111029)

Shreyas Deb (2023113024)

Vishak Kashyap K (2023113012)

### 1. Introduction

In an interconnected world, the spread of pathogens can have a far-reaching impact, often with devastating consequences. The **Pathogen Database** is designed to simulate and track the dynamics of pathogen spread and mutation across the globe, inspired by the model in *Plague Inc.*. This database documents infection patterns, government responses, research developments, and transmission pathways, providing a detailed and realistic framework for understanding the spread and evolution of pathogens. With an emphasis on statistical tracking, public health impacts, and mutation monitoring, this database is an essential tool for studying pandemic response and preparedness.

### 2. Purpose of the Database

The purpose of the Pathogen Database is to provide an interactive model of disease spread and control by cataloguing essential data points on pathogen characteristics, infection rates, and the efficacy of responses across different regions. It serves as an analytical platform for visualising the impact of various diseases and understanding the factors that contribute to their spread. By allowing users to track pathogen mutations, evaluate quarantine efforts, and monitor research progress toward cures and vaccines, this database helps support research, policy-making, and simulation efforts related to pandemic preparedness.

### 3. Users of the Database

User Role	Description
Public Health Organizations	Agencies like the World Health Organization and the Centers for Disease Control can use this database to monitor disease characteristics, track global infection patterns, and release public health guidelines.
Research Institutions	Scientists and medical researchers can access real-time data on pathogen mutations and resistance to better inform their research on vaccines, treatment protocols, and cures.
Academic Institutions	Students and educators in public health, epidemiology, and related fields can utilise the database for simulations, assignments, and research purposes.
Government Bodies	Policy-makers and emergency response teams can use the data to model the outcomes of different response strategies, such as travel restrictions, quarantine measures, and vaccination campaigns.
General Public	Informed citizens may consult the database to understand the global impact of pathogens and public health responses, gaining insights into the importance of preparedness and coordinated action in times of health crises.

### 4. Applications of the Database

- **Infection Tracking:** Monitor infection and lethality rates in each country, helping users visualise how quickly a pathogen spreads and which areas are most affected. This data can guide resource allocation for public health responses, like sending medical supplies and personnel to critical regions.
- **Mutation Analysis:** Track mutations in real-time to assess changes in pathogen characteristics, such as resistance to drugs or increased transmissibility. This feature helps researchers and healthcare professionals adjust treatment protocols and anticipate future challenges in containment and treatment.

- **Response Planning:** Simulate government and public health responses, such as quarantine enforcement, vaccination programs, and social distancing measures, allowing users to evaluate the effectiveness of these strategies. This data helps improve decision-making and supports policy-makers in managing real-world public health crises.
- **Vaccine and Cure Development:** Follow the progress of research labs worldwide as they work toward vaccines and treatments, providing estimated timelines and funding needs. Users can gain insights into the challenges faced in developing cures and the impact of funding and resource allocation on research speed.
- **Transmission Pathways Mapping:** Map out the routes and methods by which pathogens spread between countries, such as international travel, animal migration, or waterborne transmission. This application helps users model and predict the next areas at risk, allowing preemptive measures to be taken.
- **Public Awareness and Media Insights:** Generate reports for public and media outlets to increase awareness of ongoing pathogen threats and responses. The database can present high-level insights to educate the public on health crises and reinforce the importance of timely action and prevention.

## 5. Database Requirements

### a. Assumptions

- When vaccine drive happens in the same country of origin of vaccine, the exporter and importer country are the same.

### b. Strong Entity Types & Attributes

Entity Type	Attribute	Attribute Type	Data Type / Domain	Constraints
Pathogen	id	Key	Varchar	PRIMARY KEY
	scientific_name	Simple, Single-valued	Varchar	NOT NULL
	type	Simple, Single-valued	Enum (Virus, Bacteria, Protozoa, Fungus)	NOT NULL
	lethality_rate	Simple, Single-valued	Int (0 to 100)	NOT NULL
	transmission_method	Simple, Multi-valued	Enum (Airborne, Waterborne, Bloodborne, Droplet, Contact, Sexual)	NOT NULL
	predecessor	Simple, Single-Valued	A valid Pathogen id	Recursive Relationship, can be NULL
	incubation_period	Simple, Single-valued	No. of days, Int ( $\geq 0$ )	NOT NULL
	mutation_probability	Simple, single-valued	Int (0 to 100)	NOT NULL
	resistance	Composite, Multi-valued	<medium: Varchar, resistance_level: Int (0 to 100)>	(None)
Mutation (subclass of Pathogen)	id	Key	Varchar	PRIMARY KEY (Parent_Pathogen_ID should be prefix of ID)
	scientific_name	Simple, Single-valued	Varchar	NOT NULL
	type	Simple, Single-valued	Enum	NOT NULL
	lethality_rate	Simple, Single-valued	Int (0 to 100)	NOT NULL

Entity Type	Attribute	Attribute Type	Data Type / Domain	Constraints
	transmission_method	Simple, Multi-valued	Enum	NOT NULL
	predecessor	Simple, Single-valued	A valid pathogen id	Same as Predecessor of Pathogen with Pathogen(ID) = Mutation(Parent_Pathogen_ID); can be NULL
	incubation_period	Simple, Single-valued	No. of days, Int ( $\geq 0$ )	NOT NULL
	mutation_probability	Simple, Single-valued	Int (0 to 100)	NOT NULL
	resistance	Composite, Multi-valued	<medium: Varchar, resistance_level: Int (0 to 100)>	(None)
	parent_pathogen_id	Foreign Key, Single-valued	Varchar	NOT NULL, REFERENCES Pathogen(ID), does not exist in Mutation(ID)
	mutation_id	Derived attribute	Varchar	Derived from ID & Parent_Pathogen_ID
	date_discovered	Composite	<Year, Month, Date>	NOT NULL
Country	country_id	Key	Varchar	PRIMARY KEY
	name	Simple, Single-valued	Varchar	NOT NULL
	population	Simple, Single-valued	Int	NOT NULL
	health_quality_index	Simple, Single-valued	Int	(None)
	climate	Simple, Single-valued	Enum	NOT NULL
	borders_open	Simple, Single-valued	Boolean	(None)
Research Lab	lab_id	Key	Varchar	PRIMARY KEY
	name	Simple, Single-valued	Varchar	NOT NULL
	location	Foreign Key, Single-valued	Varchar	NOT NULL, REFERENCES Country(Country_ID)
	total_funding	Simple, Single-valued	Float	(None)
	projects	Derived attribute	Int	Derived by querying Research_Projects
Research Project	project_id	Key	Varchar	PRIMARY KEY
	total_funding	Derived attribute	Float	Derived from individual fundings from Participates_In reln.
	labs	Multi-valued, Foreign Key	Varchar	NOT NULL, REFERENCES Research_Lab(Lab_ID)
	pathogen_id	Foreign Key	Varchar	NOT NULL, REFERENCES Pathogen(ID)
	duration	Composite	<Start_Date, End_Date>	Start_Date cannot be NULL

Entity Type	Attribute	Attribute Type	Data Type / Domain	Constraints
	research_focus	Multi-valued	Varchar	(None)
	milestones	Multi-valued, Composite	<Milestone, Date_achieved>	(None)
	current_status	Simple, Single-valued	Varchar	(None)

### c. Weak Entity Types & Attributes

Entity Type	Attribute	Attribute Type	Data Type / Domain	Constraints
<b>Government Response</b>  (identified by Country & Pathogen)	response_id	Partial Key	Varchar	(None)
	country_id	Foreign Key	Varchar	NOT NULL, REFERENCES Country(Country_ID), identifying relationship
	pathogen_id	Foreign Key	Varchar	NOT NULL, REFERENCES Pathogen(ID), identifying relationship
	date_implemented	Simple, Single-valued	Date	NOT NULL
	response_type	Simple, Single-valued	Enum	NOT NULL (Quarantine, Travel Ban, Social Distancing, Vaccination Program, Humanitarian Programmes, etc.)
	response_severity	Simple, Single-valued	Enum	NOT NULL (Level of severity declared by government on an action - Low, Med, High, etc)
	effectiveness_rate	Simple, Single-valued	Int (0 to 100)	NOT NULL
<b>Vaccine</b>  (identified by Pathogen & Research Project)	vaccine_id	Partial Key	Varchar	(None)
	name	Simple, Single-valued	Varchar	NOT NULL
	pathogen_id	Foreign Key, Single-valued	Varchar	NOT NULL, REFERENCES Pathogen(ID), identifying relationship
	project_id	Foreign Key, Single-valued	Varchar	NOT NULL, REFERENCES Research_Project(Project_ID), identifying relationship
	date_discovered	Simple, Single-valued	Date	NOT NULL
	effectiveness	Simple, Single-valued	Int (%age from 0 to 100)	NOT NULL
	number_of_administrations	Simple, Single-valued	Int	NOT NULL
	number_of_doses	Simple, Single-valued	Int	NOT NULL
	distribution_cost	Simple, Single-valued	Float	(None)

	side_effects	Simple, Single-valued	Enum	(None, Mild, Severe, etc.)
--	--------------	--------------------------	------	----------------------------

#### d. Relationship Types

##### Degree 2 (Binary Relationships)

##### 1. Is\_A – Subclass-Superclass Relationship

- Subclass: Mutation
- Superclass: Pathogen
- Description: A Mutation Is\_A Pathogen (in our model).  
A mutation entry in the Mutation type can be considered to be its own organism, with its own properties, and can participate in all the same relationship types as any Pathogen. And hence, for convenience, Mutation was made a subclass of Pathogen, with additional attributes. But a mutation is scientifically the same species as the alpha pathogen, with a slightly different set of behaviours. And, therefore, a surrogate key ID was introduced as a common key attribute between the two entity types rather than the scientific name, which is technically unique.
- Deriving Mutation ID: We place an additional constraint on the ID attribute in Mutation, such that the Parent\_Pathogen\_ID of any Mutation entity is always a prefix string of its ID, so that the Mutation ID is straight-forward to derive.  
Ideally, it is preferred that (Parent\_Pathogen\_ID, Mutation\_ID) be the composite key of a Mutation. However, it was copying every single other attribute from Pathogen – just that its Pathogen ID was no longer unique/key. Hence, the workaround of introducing a surrogate key ID was made for convenience of defining relationships and we do not have to create a complex jugaad such as defining relationships twice or creating a union.

##### 2. Predecessor\_Of

- Participating Entity Types: Pathogen (Successor: Partial), Pathogen (Predecessor: Partial)
- Description: Pathogen is a predecessor of another Pathogen
- Cardinality Ratio: Predecessor : Successor :: 1 : M
- Participation Constraint: Predecessor – Partial (0, 1), Successor – Partial (0, M)
- Extra Attributes: Similarity\_Score (percentage similarity between two pathogens)

##### 3. Mutated\_From

- Participating Entities: Mutation, Pathogen
- Description: Pathogen is the pathogen species for a Mutation. A Pathogen that is also a Mutation cannot be the parent pathogen for another mutation, i.e., parent of Mutation has to be a Pathogen that is not in Mutation.
- Cardinality Ratio: N : 1
- Participation Constraints: Mutation – Total (1, 1), Pathogen – Partial (0, N)

##### 4. Infection

- Participating Entity Types: Pathogen (Partial), Country (Partial)
- Description: Country being infected by a Pathogen.
- Cardinality Ratio: M : N
- Participation Constraints: Pathogen – Partial (0, N), Country – Partial (0, M)
- Extra Attributes: First\_Case\_Date (Date), Total\_Infected (Int), Total\_Deaths (Int)

##### 5. Participates\_In

- Participating Entity Types: Research Lab, Research Project
- Description: A Research Lab conducts/participates in various Research Projects.
- Cardinality Ratio: M : N
- Participation Constraints: Research Lab – Partial (0, N), Research Project – Total (1, M)
- Extra Attributes: Budget\_Allocation (Float – fund allocated by the lab)

##### 6. Located\_In

- Participating Entity Types: Country, Research Lab
- Description: Research Lab is located in Country

- Cardinality Ratio: 1 : N
- Participation Constraints: Country - Partial (0, N), Research Lab - Total (1, 1)

### Degree 3 (Ternary Relationships)

1. Vaccine\_Development (Identifying Relationship for Vaccine)
  - Participating Entity Types: Research Project, Vaccine, Pathogen
  - Description: A specific Vaccine is developed as part of a Research Project targeting a particular Pathogen.
  - Cardinality Ratio: 1 : M : N
  - Participation Constraints: Research Project - Partial (0, M \* N), Vaccine - Total (1, N), Pathogen - Partial (0, M)
2. Country\_Response\_To\_Pathogen (Identifying Relationship for Government Response)
  - Participating Entity Types: Country, Government Response, Pathogen
  - Description: Represents the Government Response by a Country to a particular Pathogen.
  - Cardinality Ratio: 1 : M : N
  - Participation Constraints: Country - Partial (0, M), Government Response - Total (1, 1), Pathogen - Partial (0, N)
3. Research\_Focus
  - Participating Entity Types: Research Lab, Research Project, Pathogen
  - Description: A Research Lab has a particular research focus within a Research Project on a given Pathogen.
  - Cardinality Ratio: M : N : N
  - Participation Constraints: Research Lab - Partial (0, N), Research Project - Total (1, M), Pathogen - Partial (0, N)

### Degree 4 (Quaternary Relationships)

1. Global\_Vaccine\_Distribution
  - Participating Entity Types: Country(Importer Country and Exporter Country), Government Response, Vaccine
  - Description: Importer Country imports Vaccine from Exporter Country as part of Government Response(only those Government Responses take part which are of type 'Vaccination Drive')
  - Extra Attributes: Doses\_Distributed (Int - number of doses sent to each country), Distribution\_Cost\_Per\_Country (Float - cost per country), Coverage\_Percentage (Int - percentage of population covered by vaccine in each country)
  - Cardinality Ratio: 1 : M : N : N
  - Participation Constraints: Country(Importer) - Partial, Country(Exporter) - Partial, Government Response - Partial, Vaccine - Partial

## 6. Functional Requirements

### a. Modifications (CRUD Operations)

- *Creation:* Functions like -
  - addNewPathogen(...): Creates a new Pathogen entity. Checks for integrity constraints for Predecessor foreign key. Additionally, Predecessor cannot have the same ID as the entity being inserted. Example - INSERT into PATHOGEN<1, 'samplus pathogenus', Virus, -50, Contact, NULL, 5, 10, 40> (This is an INVALID Insert due to the negative value of the lethality\_rate which should be between 0 and 100)  
INSERT into PATHOGEN<1, 'samplus pathogenus', Virus, 50, Contact, NULL, 5, 10, 40> (This is a VALID insert)
  - addNewMutation(PathogenID, MutationID, ...): Creates a new Mutation entity, with ID as concatenation of arguments PathogenID•MutationID, Parent\_Pathogen\_ID as PathogenID. PathogenID will be checked for referential integrity against Pathogen table. The rest of the attributes undergo identical constraint checks as addNewPathogen().

- Other functions work similarly and check for integrity constraints, Names include - `addNewCountry()`, `addNewResearchLab()`, `addNewResearchProject()`, `addNewGovResponse()` and `addNewVaccine()`.
- Read:** Functions to read the instances of the different entities present in the database. Example - `readPathogen()`
- Update:** Functions to update the entities or relationships as follows -
  - `updateVaccine(...attribute to update...)`: Updates the Vaccine entity in cases such as 'number\_of\_administrations' increasing or new 'side\_effects' coming up or 'effectiveness\_rate' changing as more people take the vaccine. Example - `UPDATE Vaccine WHERE vaccine_id = 10 and pathogen_id = 1 and project_id = 3 CHANGE side_effects = side_effects + {mild fever}`
  - Other functions include - `updateResearchProject()`, `updateResearchLab()`, `updateGovResponse()`, etc,
- Delete:** Functions to delete entities or relationships:
  - `deleteMutation()` - If a study on a mutation was incorrect and was actually a new Pathogen, we perform deletion and a subsequent insertion is done in Pathogen. Example - `DELETE Mutation WHERE mutation_id = 5 and pathogen_id = 2`

## b. Retrieval Operations

### I. Selection/Projection/Aggregation/Search

- Selection:** Example operation - Selecting all the Government\_Response tuples corresponding to a particular pathogen. Query - `SELECT * FROM Government_Response WHERE pathogen_id = 1;`
- Projection:** Example operation - Selecting the name of all the Countries where the borders are open. Query - `SELECT name FROM Country WHERE borders_open = True;`
- Aggregation:** Example operation - Selecting the best performing vaccine for a given pathogen. Query - `SELECT MAX(effectiveness), vaccine_id FROM Vaccine WHERE pathogen_id = 1;`
- Search:** Example operation - Searching for a pathogen of certain subspecies by querying for 'likeness' of the `scientific_name`. The given query searches all pathogens which are of the virus family 'betacoronavirus'. Query - `SELECT * FROM Vaccine WHERE scientific_name LIKE "%betacoronavirus%";`

### II. Analysis Reports

- This report identifies the number of countries that have taken high-severity government responses against highly lethal pathogens. By showing the extent of strict measures globally, it helps policymakers evaluate the aggressiveness of responses based on pathogen lethality. Query -

```
SELECT COUNT(DISTINCT Government_Response.country_id) AS
High_Severity_Countries
FROM Government_Response
JOIN Pathogen ON Government_Response.pathogen_id = Pathogen.id
WHERE Government_Response.response_severity = 'High' AND Pathogen.lethality_rate >
75;
```

- This report calculates the total funding directed towards research projects focusing on airborne pathogens. Such a summary is essential for understanding resource allocation toward highly transmissible diseases, helping organisations prioritise funding for pathogens that pose the greatest spread risk.

```
SELECT SUM(Research_Project.total_funding) AS Total_Funding
FROM Research_Project
JOIN Pathogen ON Research_Project.pathogen_id = Pathogen.id
```

WHERE Pathogen.transmission\_method = 'Airborne';

## 7. Summary

The **Pathogen Database** provides a comprehensive model for tracking and analysing the spread and mutation of pathogens globally, inspired by *Plague Inc.* Through this database, users can monitor infection patterns, government responses, research developments, and transmission pathways, creating an essential tool for pandemic preparedness and public health policy. It includes diverse user roles, such as public health organisations, research institutions, government bodies, and the general public, each benefiting from its various applications.

The database supports a wide range of functionalities, including CRUD operations for data management and advanced retrieval operations like selection, projection, aggregation, and search. Analytical capabilities include generating detailed reports that provide insights into government responses to high-lethality pathogens and funding distribution for research projects on airborne pathogens, for example. By facilitating complex relationships among entities like countries, research labs, pathogens, and mutations, this database enables a realistic simulation environment for disease control and mitigation. The Pathogen Database thus stands as a valuable asset in supporting research, informed decision-making, and global health initiatives.