# Database Design

### Assessment One – Report

## Data Cleansing

The data provided had a few issues:

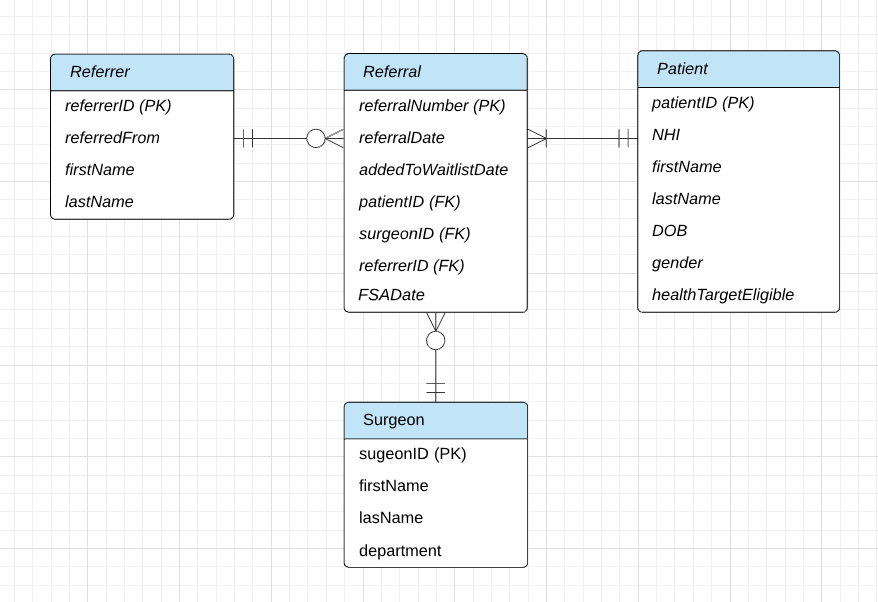
|  |  |
| --- | --- |
| Issue | Solution |
| Column of dates with no header | Ideally I would ask the client what the field represents. In this case I have left it out of the database entirely but have kept a copy of the data. |
| Duplicated NHI | NHI has duplicates and therefore doesn’t make it a good primary key (as well as security reasons), so I would inform the client about the duplicates but still enter the data in the meantime |
| Impossible dates | One where there was a DOB after a referral date – one man born in the 1700s. Both have been entered into the database. Client would be informed and provided a query to change it. |
| Missing FSA dates | The missing FSA dates will be left blank. I’m assuming these were emergency surgeries with no specialist appointment. |
| Referred from data / Referred by data slight inconsistencies | Some referrers had a lot of referrals from one place, but only one or two from another place. I’m assuming this is because there is some overlap within the ‘referred from’ field. For this reason, I changed the outliers to align with the majority. |

## Entities

I initially chose 3 entities, those being – Patient, Surgeon, Referral. However, once I went through the normalisation process, I discovered the transitive dependency between referred by and referred from, and created a new table for them call referrer,

## Attributes

Choosing the attributes for each entity was reasonable easy.



I decided to put department into the surgeon table as I am assuming that each surgeon specialises in a particular department. This is backed up by the data.

I also decided to split all the name attributes (surgeon name, patient name, referred by) and create separate fields for the first and last names of each person.

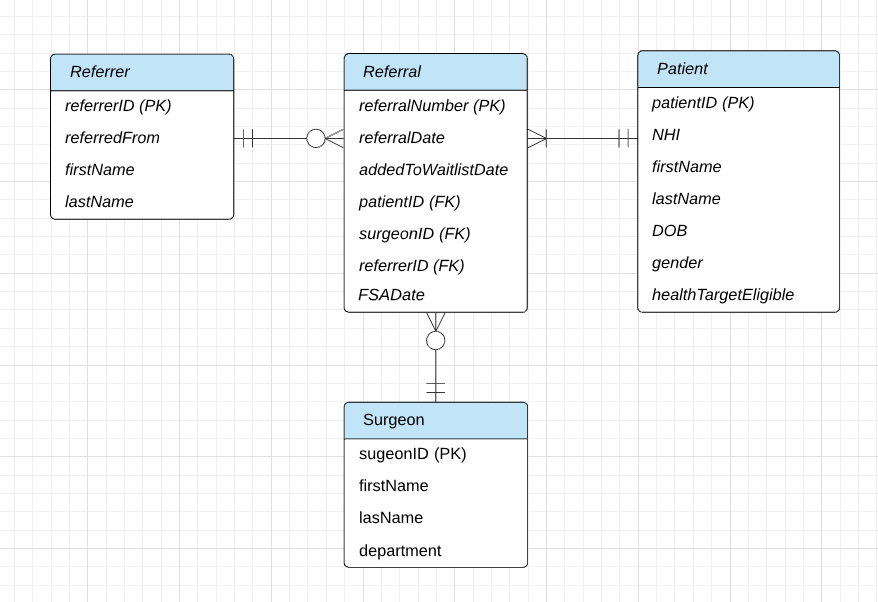
## Keys

There weren’t any attributes within the dataset that were suitable to be primary keys for any of the entities. This meant creating new attributes that would auto-increment when a new row is added. The added attributes were:

|  |  |
| --- | --- |
| Entity | Primary Key |
| Referral | referralNumber |
| Patient | patientID |
| Surgeon | surgeonID |
| Referrer | referrerID |

I decided not to use NHI as a PK because there were some patients that had the same NHI as others.

## Connectivity & Relationships

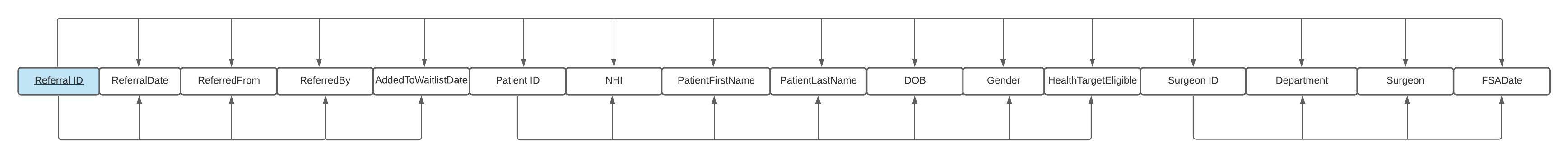


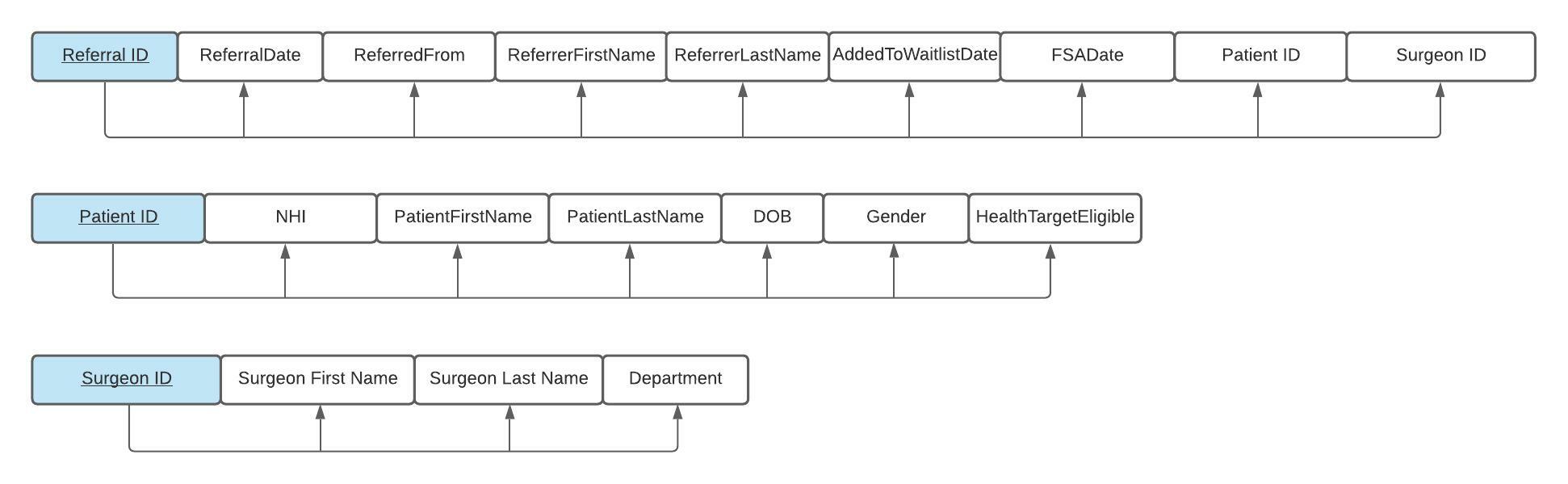
A single Referral can have one and only one patient, one and only one surgeon, one and only one referrer.

A Referrer can have zero to many referrals. The reason why I did zero to many (and not 1 to many) is because the referrers seem to all be doctors and may be entered into the database before even sending a referral. This is the same reasoning for why surgeon is a 0 to many relationship.

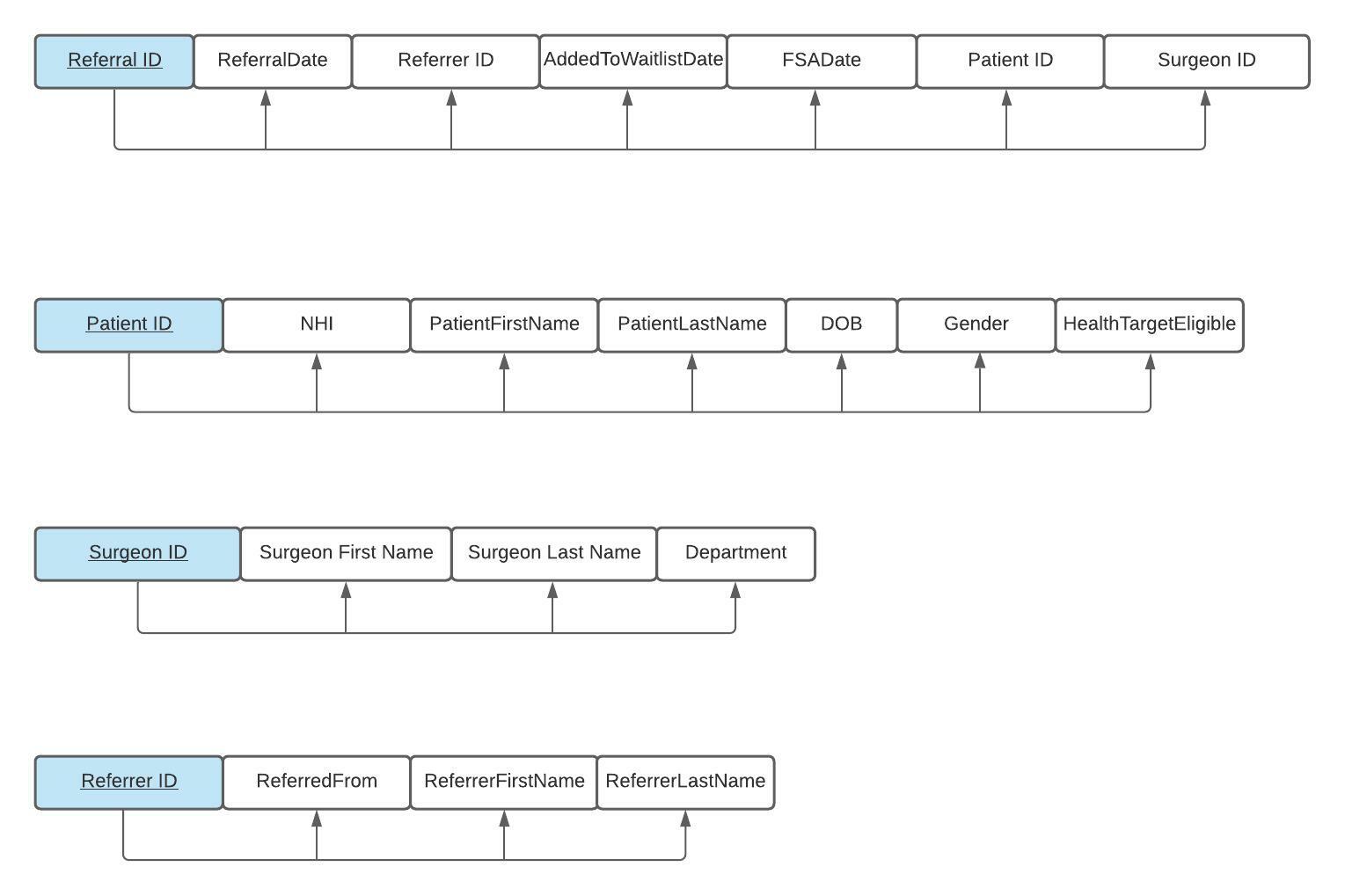
A Patient can have 1 to many referrals. The reason I have drawn ‘1 to many’ (rather than ‘0 to many’) is because a patient will likely be added to the database at the same time as the referral.

## Normalisation

To create The First Normal Form (1NF) I removed the attributes that had duplicated data from another field. This was the Year/Month field. I also removed the multivalued attributes into separate fields (firstName, lastName).



For the 2nd Normal Form, the partial dependencies within the First Normal form were removed and added into their own tables.

For the Third Normal Form, the transitive dependencies were removed the Referrer table was added.

## Bridging Tables

The Referrer table is the bridging table of all the other tables. This is to prevent the other entities from having a many to many relationships between each other and simplifies the database.

## Version Control

<https://github.com/swemuel/db-Ass-1/tree/main>

