

Data Pipeline Project

HEALTH INSURANCE RECORDS – XML MODELLING

SAFA EL AZRAK

Modeling of the XML database of a health insurance company

In this project, I considered that a health insurance company has Eletronic medical records for all its employees.

The chosen model is represented below:

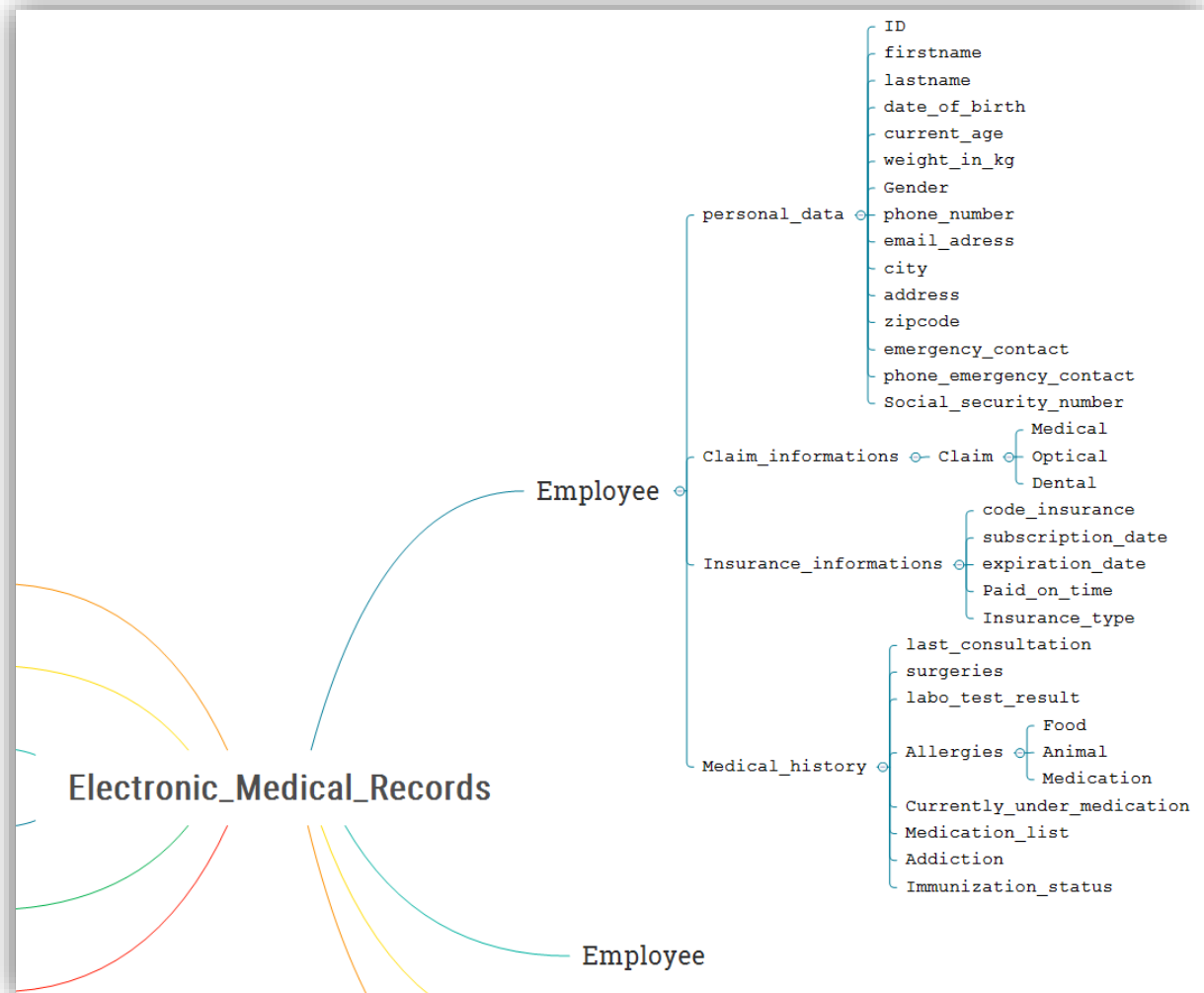


Figure 1 : Representation of the XML modelling

XML File

My root element is “Eletronic_Medical_Records” that is a list of Employees : a sequence having all the elements “Employee”. The number of Employees in the records is unbounded. In my example, I randomly worked on 10 Employees.

For each Employee we have four other elements (groups of information) : their personal data, their claim information, their insurance information and their medical history.

1- The element “personal_data” contains general information about the employee. Beside the mandatory information to identify the person like the name, the date of birth, we have those to contact them and some information about the Age and the Weight (that could give us a little hint about the medical profile). If someone is overweighted or is getting old, he most likely to be consulting doctors and in need of the insurance reimbursement for example.

- 2- The “**Claim_information**” is to let the company know which fees the Employee has. We can count the claims per employee and categorize them into :
- Medical Claims
 - Optical Claims
 - Dental Claims

This data will allow the company to know which persons have the most claims (and the category of it).

- 3- The “**Insurance_information**” is about the policy insurance of each employee. Through this, the company can know when the employee subscribed for the insurance, the expiration date of it and even the type of the policy. Here I chose two types of insurances : Premium and Normal. Premium might for example be taking in charge the staying in a hospital for the first 3 days.
- 4- Finally, the element “**medical_history**” is all about getting a medical profile of the employee. Now again, it’s to predict which ones will need the most the insurance reimbursement.

Advantages and disadvantages of these choices for the processing of the data

The company is having electronic medical records for all of the employees. I chose to organize the data inside each element “Employee” and inside each element “Employee” we can access all information we want : personal data, claim information, medical history or insurance information. This means that any employee of the company can see a everything on each employee and have an overview of its file.

But in this case, if we say that for example, there are people managing the financial aspects of EMRs and others like specialist managing the medical part of the patients : the model I chose will lead everyone to consult the data in the same place and having access to the same information.

Maybe the specialist don’t need any information other that the medical history, they’re not interested in their insurance type or financial status. Other employees may be interested in financial aspect and not in the medical one. Another model of this data might be to have different elements like medical information or financial information and inside them we find the employees for example.

XML Schema

After writing the Xml file to model our data, we have to validate it against a schema. This consists in defining the different data types and the number of occurrences (maximum and minimum) an element can have.

XSL Transformations

I imagined four possible transformations for my model answering four different questions.

- **Retrieving Employees key information**

This scenario is to have an overview of the Employees, their gender and their age.

Employee Name	Employee ID	Security Number	Gender	Age
James Butt	1	9 5 6 7 10 11 8 12 4 2	Male	52
Josephine Darakjy	2	12 8 5 1 2 7 10 3 4 9	Female	51
Art Venere	3	6 7 10 11 3 8 2 1 9 4	Female	50
Sage Wieser	9	10 9 6 11 1 8 7 2 4 3	Male	24

Figure 2 : sample of the output of Transformation 1

- Retrieving Employees with their claim information

This scenario allows us to retrieve the employees with their total of claims (to see which ones demand a lot of claims and have big medical fees, probably to switch them to a more expensive insurance type for example). We can also display the category of their claims to have more details about the amount of each one.

Employee Name	Amount Medical claim	Amount Optical claim	Amount Dental claim	Total Amount Claim	Number of Claims
James Butt	0	56	80	136	3
Josephine Darakjy	0	567	980	1547	1
Art Venere	230	0	0	230	3
Sage Wieser	0	0	468	468	2

Figure 3 : sample of the output of Transformation 2

- Retrieving Employees with their Medical history (Health profile)

The “health profile” of an employee can also help the company predict if he has high or low medical expenses (dues to allergies, surgeries, currently under medication which indicates that a constant pharmaceutical claim for example). Maybe someone with high fees can be switched to a higher type of insurance.

Employee Name	Age	If had Surgeries	Is under Medication	Food Allergie	Animal Allergie	Medicine Allergie
James Butt	52	false	false	true	false	false
Josephine Darakjy	51	false	false	false	false	true
Art Venere	50	true	true	false	false	false
Sage Wieser	24	false	false	true	false	false

Figure 4 : sample of the output of Transformation 3

- Retrieving Employees with their Health insurance information

This scenario is useful to detect Employees whose insurance contract is ending soon to anticipate the renewal of the policy, the type of their insurance (to see if they can upgrade some employees) and the payment profile (is the Employee paying his contributions on time or not).

Employee ID	Employee Name	Insurance type	Insurance Expiration date	Paying on time
1	James Butt	Normal	2020-06-21	true
9	Sage Wieser	Normal	2021-07-29	true
6	Simona Morasca	Premium	2021-08-11	true
4	Lenna Paprocki	Normal	2021-08-29	true

Figure 5 : sample of the output of Transformation 4