

How to organise your data to meet your analysis needs?

Introduction	2
How to determine your information needs?	2
How to organise your data?	3
Getting the insights for your model	4
Data model design: what and how are you going to record data in your database?	6
Conceptual data modeling: entities, properties and relationships	7
Events centered data models based on the event standard format: who did that to whom?	8
Main entities: event, victim and perpetrator	9
The events standard in practice	11
Document repository data model	12
Track complaints data model	13
Monitoring judicial response data model	13
What's next once you have designed your data model?	14

Introduction

As Human Rights (HR) practitioners we manage large volumes of information from different sources, mostly cases about HR violations. Although we understand what information is relevant and how to collect it, it is not always easy to find the most effective way to organise it. While the advancement of technology invites us to overcome traditional data models, we are not familiar on how to take advantage of it.

This guide is intended to boost your team up at the time of building a database. No matter the platform you will use, as a first step we need to decide which data to collect and how to organise it. This will determine the kind of analysis that we will be able to perform and, thus, the insights that we will be able to deliver to our recipients.

How to determine your information needs?

The information needs assessment is the basis of a successful documentation work, as it determines the data we need and the best way to organise it. This process should be carried out at the beginning of the project, considering several elements, such as:

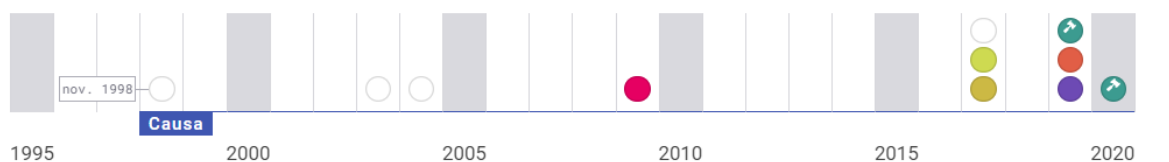
- The objectives of the documentation work: what is the data for? What is the core of your database? What do you want to capture? What questions should it answer?
- The type of data to be stored in the database: Is it quantitative or qualitative data? Is it numbers, texts, images, videos? Is there sensitive data?
- The kind of outputs that you want to produce: statistics, reports, graphic reports, storytelling, etc.
- The type of analysis you are going to perform in the database: counts, lists, case summaries, ad-hoc queries, etc.
- The users' needs: accessibility, automated functionalities, outputs as the ones mentioned above, etc.
- The kind and amount of data you are going to collect, paying attention to quality and quantity.
- The available resources: time, financial, technology, personnel, etc.

- The people involved in the different stages of the documentation process and the access permissions they will have.

In order to visualize your needs and better understand its implications for the database we strongly recommend you to list the questions you will ask the database and the graphs or visualizations you would like to obtain. Here we include some examples of different outputs you can get from your database:

Example 2: Timeline

This visualization allows an organized analysis of the most relevant milestones of the judicial case by time References.



White shows that the complaint is issued or presented in front of the court; pink represents admissibility reports; green for merits report; orange for hearings; purple for orders of the court; and green with the court hammer for judgements.

How to organise your data?

In the next section you will learn where to look to get the information you need to build your data model, which is the way you will organise your data in your database.

Then, we will explain the components of the data model and we will illustrate some of the purposes shared by most human rights organizations with different data model examples. Finally, we will expose some ideas that you should consider before implementing your model.

Getting the insights for your model

Designing your data structure is a main need. Try to summarize what you want to reflect in your database in one sentence. Making your design visual is a good way to reflect the specific purpose of your project.

Abstractly speaking, most of the things you do might fit into these or similar proposals:

- Capture facts: registering events by capturing all the elements that will help you to have a clear understanding of the fact. This general purpose could apply, for instance, to document HR violations and abuses through the *victim-act-perpetrator* model, summarized in the sentence *who did what to whom?* Of course, you can add as much complexity as you need to this model, but these are the three basic things (*victim*, *act* and *perpetrator*) you will need to describe in your database, as a starting point.
- Track processes: monitoring series of activities normally extended over time, with a chronological order or, at least, based on a model of action-result. Think, for instance, about the whole process through which laws are made, or the different stages of the complaint handling procedure. Most of these processes have something in common: they are guided by a well-defined workflow in which different paths are clearly identified.
- Check facts against a benchmark or HR standard: measuring values by comparing them with other values to show progression over time (from a baseline) or the degree of compliance with a standard taken as reference. Monitoring the progressive realization of human rights is a representative case of this objective.
- Improve document accessibility: although the work of HR documenters is always supported by documents, sometimes the document accessibility becomes an aim in itself. Facilitating the organization and retrieval of laws, caselaw, evidence, reports, media pieces or court resolutions contributes to other processes which ultimately aim to protect and promote HR.

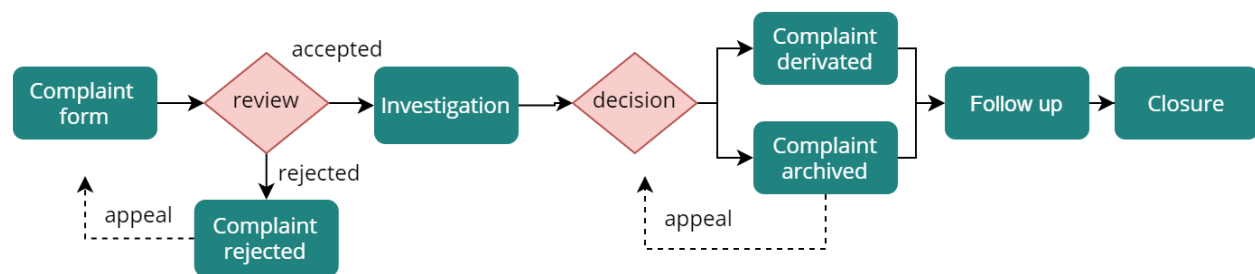
Certainly, there will be many cases that do not fit exactly these examples or that might add many complexities to them. However, it is always useful to reach a certain level of abstraction in order to discover the general patterns behind your activity.

One of the simplest ways to make your ideas clear is representing them visually. There are many free resources to help you in this visualization task, as diagrams.net, [Miro](https://miro.com) or [Mural](https://mural.com), between others.

While the data model (the conceptual representation of the database) is the most accurate technique to structure data, there are alternative helpful ways to visualize the tasks and procedures involved in data management that could prepare you to design your data model better. For example, imagine how data will flow in the situation you want to capture or how your team will gather, process and analyze information. Take a look at the examples below:

Example 1: Complaint tracking data workflow:

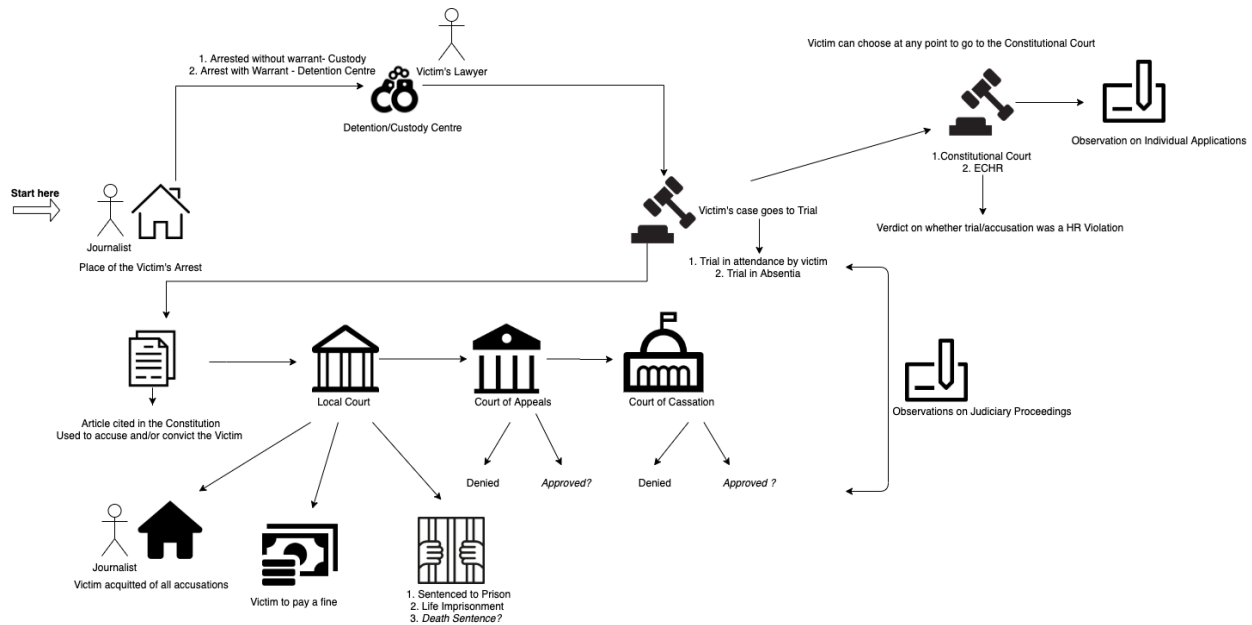
This is a simple schema that will help you to identify some components of your data model. As a test, think whether someone who knows nothing about your work would be able to understand your diagram.



The example shows a basic complaint management workflow. The complaint form is submitted by the victim and, then, reviewed by the HR organization, which decides to investigate it or not. Once the investigation is ready, the complaint can be referred to another institution (a legal instance, for example) or archived. This hypothetical example allows you to identify the type of things you will need to record in your database if you want to track complaints, as we will see later in this document.

Example 2: Journalists' arrests tracking

Here is another workflow that reflects how journalists' detentions occur in a certain place. The workflow helps to identify certain elements that later will be part of the data model.



Data model design: what and how are you going to record data in your database?

Although the term data model is commonly used by software developers to describe data structure in computational coding, the conceptual approach to data modelling stresses the importance of design, planification, and anticipating needs or probable issues to solve.

The data model is an abstraction that describes how data is organised in your database. Designing the structure of our data system requires understanding the type of objects you are going to describe and how are they related to each other. Addressing the data model design consciously as a first stage will save you time and effort. We strongly invite you to illustrate the data structure before starting to record or migrate your data.

The data model is conditioned by: analysis needs, type of data, volume of data, frequency of use and updates, relationships between data and the type of tool used to store and process data, among others.

Conceptual data modeling: entities, properties and relationships

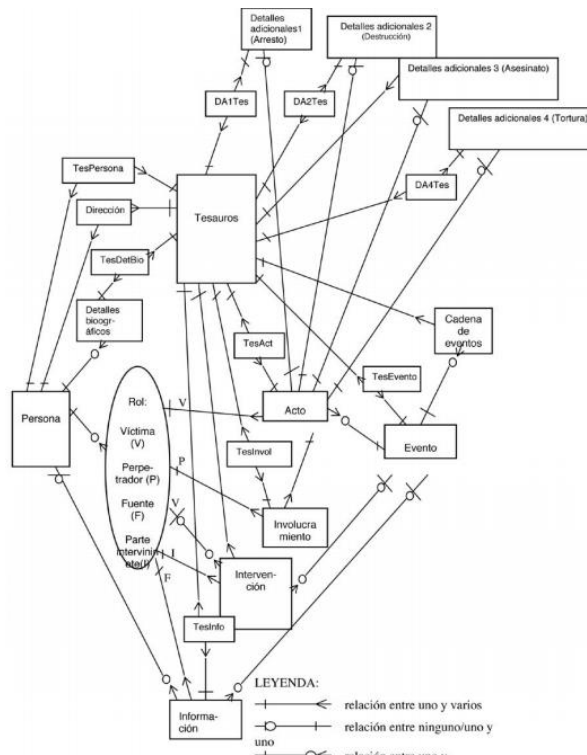
Conceptually speaking, the data model is a structured view of your data founded on entities, properties (or attributes) and relationships.

Entities are objects described in the database by means of properties. An entity could be a person, an object, a document or any item we want to record data about. In the human rights field, we normally describe victims, HR violation events, laws, acts, courts, etc. This notion is similar to the idea of unit of analysis in social research methodology.

Properties are descriptive attributes that are assigned to entities. There are various types of properties, such as text, numeric, select one from a list, select multiple from a list, date, time, etc. For instance, a victim (entity) is described through a name (text), an age (number), a date of birth (date), etc. This notion is familiar and could be articulated to the concept of variable, in statistics.

Relationships allow entities to be connected to each other. They define how entities interact. For example, one act can have various victims, one perpetrator can be involved in various acts, etc. Relationships are very important in the data model since they bring the data structure “to life”. By this expression we mean: to support dynamic escenarios and flows, avoid or reduce data duplication, to circumscribe certain properties to a specific context (since out of that context certain data could generate confusions or mistakes).

In the next section, we provide simple visualizations of data models, based in real cases. Since the data model could get as complex as the image below, we decided to illustrate them in a more pedagogical way, if easier we could think of these approximations as standard formats.



For communication purposes, the data models you will see are illustrated in a simple way, and we intentionally avoided details that could interfere either with your understanding or our partners' privacy.

Events centered data models based on the event standard format: *who did that to whom?*

The model is meant to register events and provide clear understanding about them by describing three basic entities and the relationships between them: events, victims and perpetrators. The model helps to systematically record incidents enhancing consistency and enabling analysis, such as discovering trends and also patterns in large amounts of data.

The events standards format [that you can read in this document](#) was first launched in 1993 by the HURIDOCs' Task Force. It was designed to capture all the relevant information about events which might constitute human rights violations (normally for civil and political

rights). However, it can also be adapted to cover other aspects of human rights, including the progressive realization of economic, social and cultural rights. This model can be implemented through many different tools depending on the volume of information or the kind of analysis you want to perform, but the basic principles remain the same.

Following, we describe the basis of this model. Starting from the basic entities, it is easy to add new ones to complete the model and make it able to meet our requirements. We understand that each organization has its own particular needs, so the model can and should be customized in order to fit them.

Main entities: event, victim and perpetrator

The **event** constitutes the basic unit, the center of the model, and it encompasses one or various related acts taking place sequentially or at the same time. For human rights documenters, the scope of work is commonly composed of acts of commission and omission, which potentially constitute rights violations. Usually, the event is described through a set of core properties, as the title, the geographical location, the initial and final date, the description, the rights affected, the violation typology, the supporting evidence, etc.

The act, on the other hand, is the entity connected with victims and alleged perpetrators. Aside from this information, the act also has a specific date and location, and should include detailed information about the method of violence and its attribution.

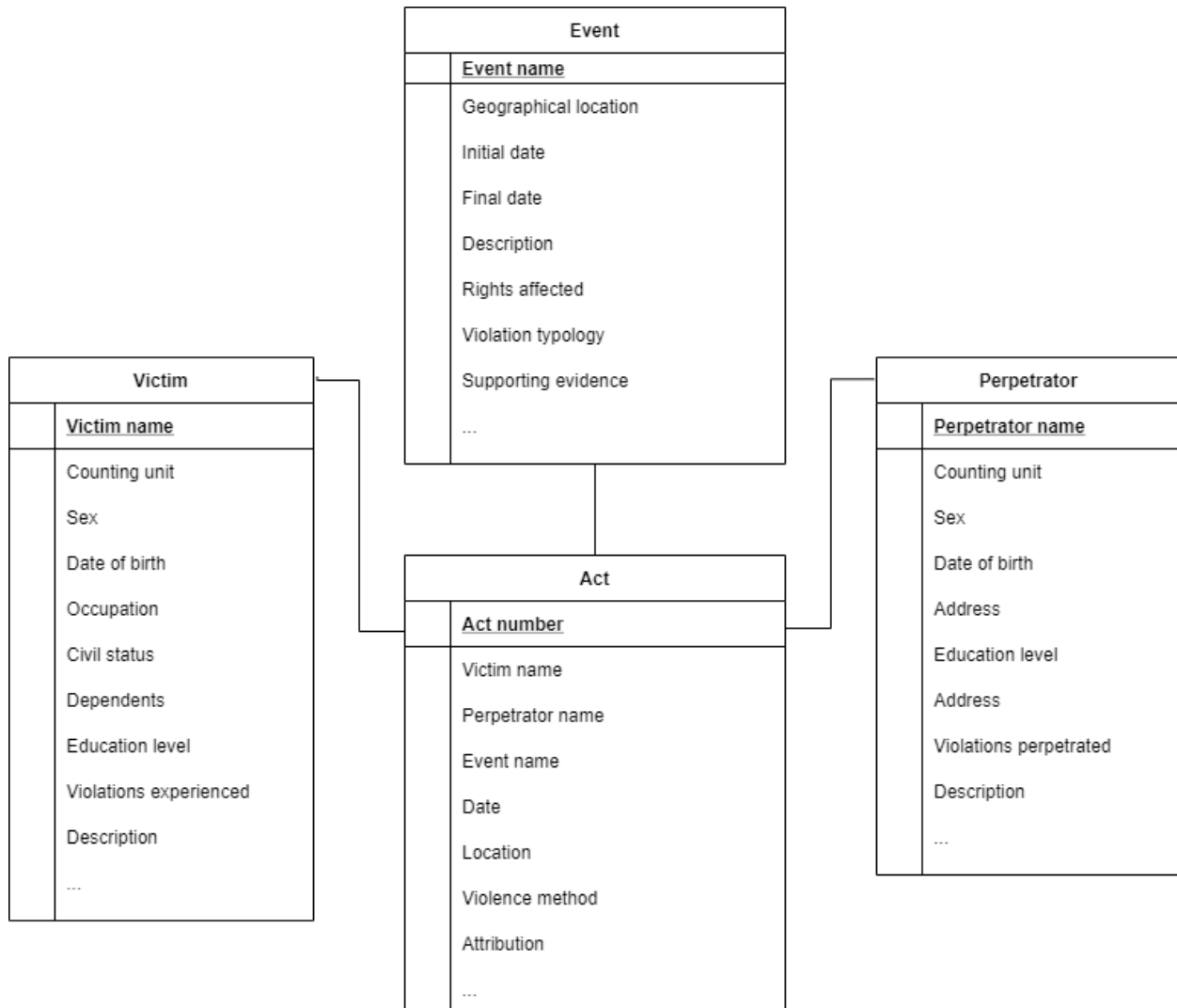
People might get involved in events through different roles, mainly as victims or perpetrators. People can also act as sources or intervening parties in the event, as witnesses or complainants.

These entities are usually described with biographic information about the individuals or groups (families, organizations, communities, social groups, etc.) and are linked to acts.

The perpetrator is the individual or group who commits the act (or the one who fails to meet the obligations to protect and fulfill rights), it could be either a person (frequently a state official) or an institution like, a specific state office or the government.

Allowing disaggregation of data and giving space to be specific through the data model will be an asset, since you will be always able to aggregate data later, while the viceversa mode

(unpacking data) will no longer be possible. This is an example of how the simplified forms would look like:



The events standard in practice

In order to show the kind of results that we can get with this model, here we present some examples based on the standard:

Death Penalty Database, by Justice Project Pakistan (<https://data.jpp.org.pk/en>)

This database includes details as the number of prisoners on death row in Pakistan, the number of executions, trial details of the prisoners and other information regarding the death penalty in the country. It allows to filter prisoners according to certain attributes such as gender, age at the time of the alleged crime or history of mental illness, to show empirical evidence on how Pakistan uses the death penalty.

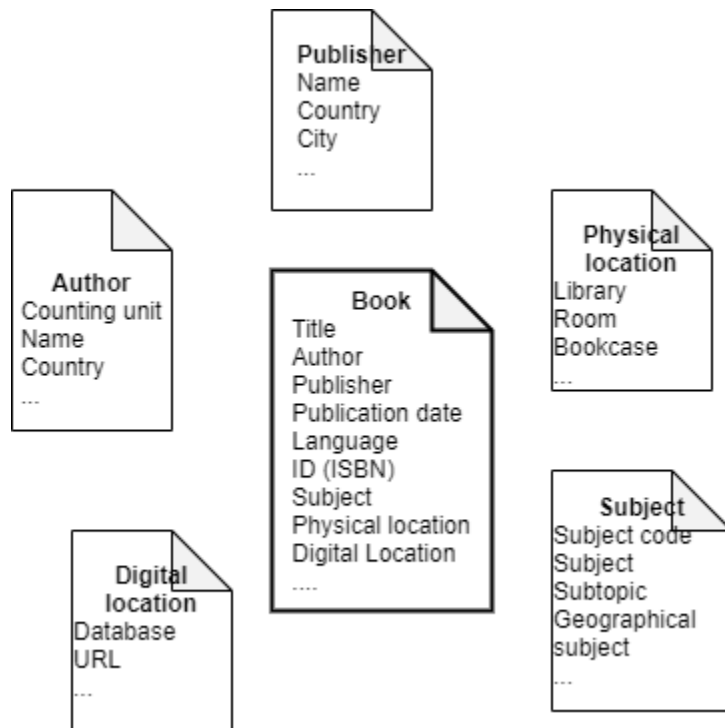
Cartography of violence against female journalists in Mexico, by CIMAC (Comunicación e Información de la Mujer Asociación Civil) (<https://cartografia-cimac.uwazi.io/es/library/map>)

The database includes information about gender based violence suffered by female journalists in Mexico. Regarding the event, it records different types of violations that you can filter by time or place. You can also filter by characteristics of the victim and the aggressor.

Freedom of Religion or Belief in North Korea database, by Korea Future Initiative (<https://koreafuture.uwazi.io/>)

This database allows users to access and organise evidence of religious freedom violations in North Korea. The database contains relational information that links the victims and the violations to the responsible perpetrators, locations, and organisations.

Document repository data model



This example shows how books are commonly described in libraries or bookshops. It is meant to facilitate adding new books and, above all, to facilitate retrieval by readers. Usually, you try to find books by title, author or subject (those are the most common access points). Many of the properties used to describe books belong to other entities: this is mostly how relationships are created between elements. One of the book's main properties is the author's name and, at the same time, every author entity has additional properties such as counting unit, nationality, etc. It is very common to see this model nourished with more types of documents, for instance multimedia files, maps, magazines, etc.

[The Cyrilla Database](#) gathers information about digital rights laws around the world in multiple languages and allows to retrieve resources by type (legislation, cases, analysis, etc.) and filter them by country and keyword. It is a great example on how to describe laws to facilitate its retrieval.

The repository model can easily be combined with other models. For example, if we go further with the library example, we could easily design something very similar to the events standard. Think, for instance, that the books described above are located in a library that loans these materials to library members. In order to do so, library members would constitute another entity with properties such as id, sex, age, email, telephone and educational level. The other main entity would be the loan itself, described with properties such as the loan period, the permission to take the material to home or to read at the library, etc.

This library loan structure could be summarized in the sentence: *What do we loan to who?*, similar to the event standard. A similar structure is used in many shops to register customers, orders and products: *Who ordered what?* Or *What do we sell to who?*

Track complaints data model

Complaint reception and management is a recurring task for national and international institutions in the human rights field. Data models for tracking complaints are essentially based in the stages of the complaint handling process.

This [case study covers the development of an information system to handle victim's complaints at a human rights organization](#). It shows how information needs were identified for a specific project and how they were converted into technical requirements for the database, including the data structure reflected in the data model.

Monitoring judicial response data model

Similar to the previous case, monitoring judicial response demands a model that covers the different stages of the judicial process and also includes a comprehensive description of relevant entities such as judicial instances.

[The SUMMA database](#), created by the Center for Justice and International Law and [explained in detail in this case study](#), constitutes a good example of how information should be registered and structured to facilitate its retrieval by users and how to provide aggregated information regarding cases.

What's next once you have designed your data model?

We have explained how to get the insights for your model and we have provided a few models that you can take as a starting point for your project. However, each project is different and you will need to adapt the model to fit your needs adding new entities, describing them with new properties or creating additional connections between them. The adequate data model is the one that allows you to record and organise the data you need to get the answers you are looking for.

To check that, test your model by adding some typical cases such as:

- The victim, event and perpetrator entities if you are documenting violations
- The stages of the complaint handling process if you are tracking complaints
- Different types of resources if your model is focused on accessibility.

Your data model should be able to capture various casuistry, so try recording rare cases that you can remember from your previous experience. Be sure you spend enough time designing and testing your model, as making amendments later would be more difficult.

Once you are done with testing, it is time to migrate your data. To do so, you can follow the basic steps mentioned [in this guide](#). Be ready to add new properties to describe your entities and new relationships to avoid redundancies between entities.

Even if your data model works for your current purpose, it is possible that your information needs regarding types of analysis, outcomes or data tipologies change from time to time. If that happens, you will need to update your data model to make it fit your new needs.