Appendix for Illustration

(additional material and not part of the paper)

The aim of this appendix is to give the reviewer a glimpse on problems with large schemata. It is not the goal for a scientific paper to develop an application in detail. The paper uses elements of this schema as examples. The understanding of the examples however does not depend on this schema.

A Typical Large-Scale Schema Application: Public Health

Let us consider a typical large scale schema\(^1\). We omit a part of the application due to space limitations. We must, however, provide a substantial insight into the application domain. In particular we omit details about health care orders, health care shipments and delivery, health care claims, payment settlement, and health care referral. The diagram in Figure 1 depicts the main types, their subtypes (denoted by thick lines from subtype to the supertype), the associations among types, and uses already the extended ER model\(^2\) with its layering of relationship types on top of relationship types (e.g., HealthCareDelivery is build on top of HealthCareVisit).

Health care information systems store data on

- people and organisations that are concerned with patients, health care provider organisations, individual practitioners, insurance companies,
- relationships between parties such as patient relationships and practitioners relationships,
- types of services and goods available from the health care providers,
- types of agreements that exist between the various parties,
- records of health care services performed,
- claims submitted and the status of the claim,
- amounts directly owned from the patients as well as payments made by the patients,
- other supporting information such as accounting information to create the financial statements and human resource information to track personnel.

People and organisations in health care. Health care organisations need to track information about people and organisations with which they interact. Typical people involved into health care processes are patients, insured individuals, individual health care practitioners, administrators, provider staff support, and contact people such as those within an insurance company and in a pharmaceutical company.

They also need to track information about organisations involved in health care such as health care providers, employers and associated groups, insurance companies, health care networks, and health care associations.

There are some generic dimensions of people such as CONTACT, EMPLOYEE. A health care organisation may need to record various contacts within pharmaceutical companies, third party administration organisations, or insurance companies.

Various standard organisation rules include EMPLOYER, SUPPLIER, HOUSEHOLD, REGULATORY AGENCY, ORGANISATIONAL UNIT and INTERNAL ORGANISATION. ORGANISATION UNITS are subtyped into PARENT ORGANISATION, SUBSIDIARY, DIVISION, DEPARTMENT, and OTHER ORGANISATION. Depending on the purpose

---

\(^1\) The application example is a compilation of about two dozen of application schemata developed with the ID\(^2\) database development tool mainly in the 90ies in Arabian and Asian countries and the schemata discussed in [?, ?, ?]. The schema in this appendix is far more advanced than the SAP R/3 schema for health management. It is based on approaches developed in [? , ? , ? , ? , ?]. The schema has been used for an international response answering a call for proposals issued by the association of health care organisations in the Gulf area in 2004. The approach has been extended to [?].
Fig. 1. The Health Care Survey Grid with Main Types, their Main Sub-Types, and Relationship Types Without Integration of Attribute Pattern.
of the information system, we may represent them by role types, kind types, and special subtypes. The last opportunity is only chosen if there are essential and specific properties and thus attributes.

Typical health care industry organisations are Health Care Provider Organisation, Group, Network, Employer, Third Party Administrator, Insurance Provider, Payor, and Health Care Association. Providers may be subtyped into Institution, Health Care Practice, or others. An Institutional Provider represents organisations providing health facilities such as hospitals and psychiatric institutions. A Health Care Practice represents one or more individual health care practitioner who form a coalition to provide health care to patients. A Network is a collection of Health Care Provider Organisations that are linked together to provide services under certain guidelines established by the organisation that set up the network. A Group is a collection of individuals who are classified within an organisation to receive coverage through the organisation. Typical groups are Employer, Third Party Administrator, Insurance Provider, Payor (organisation that pays for the claims), and Health Care Association.

Insurance is a major constituent of health care. A person may be insured for health care. The Insured Party role captures information about people or organisations that have insurance. The Insured Organisation is the organisation that is insured and covers individuals for health care. It may also play the role of an Employer. Insured Individuals are important to track for proper insurance reimbursement. The Insured Contract Holder is the main party that is covered for the insurance. The Insured Dependent is a person being covered for a policy in addition to the insured contract holder.

An alternative model could also be to include the Insured Organisation as a subtype of the Organisation Role and the Insured Person as a subtype of Person Role. Because there is probably more common information about the insurance information surrounding these parties, the decision to sub-type based on the Insurance Party has been preferred.

We observe in applications a number of standard relationships such as Organisation Contact Relationship, Supplier Relationship Employment, Organisation Rollup. The Family Dependency is a typical example of a Household Membership. Health care uses also some specific relationship types such as Patient Practitioner Relationship, Patient Provider Relationship, Practice Affiliation, and Provider Network. A doctor may act as a primary care provider (PCP). The Patient Provider Relationship identifies which patients are with which Health Care Provider Organisation. The Practice Affiliation type identifies which Individual Health Care Practitioners are associated with which Health Care Provider Organisation.

People and organisations may play a variety of roles. The Party Role type stores the kind of role somebody may play. The Party Relationship type links roles to relationships that exist between parties. A party has in a certain party role a party relationship with another party and its role.

Health care facilities such as Hospital, Office, Room, Clinic are used in health care. Therefore, we use the generic type Facility that combines these facilities. Additional subtypes of Facility are Medical Building, Ambulatory Surgery Center, and Floor. We might also include specific facilities such as Bed, which would be related to Room.

Patient, Individual Health Care Practitioner, Health Care Provider Organisation are typical examples of Party Role. The Party Qualification and Party Skill are associative types that maintain the competencies and background expertise for Party. To avoid redundancy we introduce the Skill Type and Qualification Type that allow to collect all related information into one type.

Health care is also based on Licence restrictions that are valid for certain states or a Geographic Boundary for which the licence may apply.

Specific patient information may be kept in types such as Medical Condition and Physical Characteristic.

- People and organisations that are concerned with patients, health care provider organisations, individual practitioners, insurance companies;
- Relationships between parties such as patient relationships and practitioners relationships;
- Types of services and goods available from the health care providers;
- Types of agreements that exist between the various parties;
- Records of health care services performed;
- Claims submitted and the status of the claim;
- Amounts directly owned from the patients as well as payments made by the patients;
- Other supporting information such as accounting information to create the financial statements and human resource information to track personnel.

Roles: patients, insured individuals, individual health care practitioners, administrators, provider staff support, contact people (insurance company, pharmaceutical company), ...
Other parties: Employer, Supplier, Household, Regulatory Agency, Organisational Unit (Parent Organisation, Subsidiary, Division, Department, and Other Organisation), Internal Organisation, ...

Generic dimensions: Contact, Employee, Party Qualification, Party Skill, Skill Type, Qualification Type


Generic associations: Party Role, Party Relationship, ...

Health care facilities: Hospital, Office, Room, Clinic, ...

Generic facility types: Facility (Medical Building, Ambulatory Surgery Center, Floor, ... Bed, Room), ...

Specific types: Licence within a Geographic Boundary, ...

Generic characterisations: Medical Condition, Physical Characteristic, ...

...

Health care products. Health care organisations still consider themselves as being service-oriented. They are service providers, they perform procedures, offer diagnoses, and help patients through their time and expertise. What about pharmaceuticals, supplies, and medical equipment they may offer? Therefore they offer at the same time goods and services. We thus may generalise services and goods to Products. To be more neutral we model them by by Health Care Offering.

It contains however typical product characteristics such as supplier product, inventory item storage, price components, cost components, and price components.

Reporting schemata. Reporting schemata are typically OLAP cubes. OLAP cubes can be represented by relationship types which components are OLAP dimensions and which attribute types are derived functions [?].

Cubes used for health applications analyse how successful the health care enterprises have been treating patients. These cubes need to support the following:

Financial analysis: Balance sheets and statement trends allow to determine trends on the income and the profitability over time, incident types, patient types, health care practitioner types etc.

Human resource analysis: Employees can be classified regarding age, gender, marital status, position and other demographic information.

Claims analysis: History of claims and settlements can be classified regarding service codes, types of diagnosis, episode types, geographic areas, dates, and payors. Trend analysis allows an insight what types of health care deliveries have been reimbursed and allows to predict what to expect regarding insurance receipts.

Health care delivery outcome analysis: The outcome of health care deliveries can be analysed under various circumstances.

Health care episode outcome analysis: The outcome of different kinds of health care episodes is of specific interest depending on various circumstances.

The cube in Figure 2 accommodates the last need of analysis. It allows for instance to analyse the positive outcomes of health care treatment. Questions such as “How successful have health care treatments been in treating patients for particular health care episodes in particular regions and in dependence on treatment time?” can be answered. The fact type is represented as a relationship type and uses derived attribute types for the number of episodes, the number of health care visits, the average length of the episode, and the total charges associated with the episodes. The purpose of this relationship type is to provide measurements in order to analyse the success and lack of success for various health care episodes depending on eight dimensions.

The dimension OUTCOMES provides an explicit representation of the outcome of the episode. This dimension OUTCOMES is used as a separation dimension. The seven other dimensions EPISODES, DIAGNOSES, INCIDENTS, TIMES, PRACTITIONERS, PROVIDES, and PATIENTS allow the analysis of health care episodes to be viewed under different conditions. These dimensions are used in the cube for clustering and are supported by drill-down, roll-up, dice and slice operations. For instance, the PATIENT dimension supports a separation of outcomes depending on the main characteristics of a patient. Typically dimensions are hierarchically structured.
There are several issues to consider when deploying and using OLAP cubes within an application. What is the data volume that needs to be available in the cube? Which part of the cube must be materialised and which part of the cube may remain to be virtualised? Transformation processing time to create and to update the cube may go far beyond what can be tolerated. The data refresh frequency and the data transformation processing window size must be sufficient to meet the business needs. Therefore we need to determine the needs, the query issuing frequency and the response time behavior in such applications. Depending on the infrastructure, the cube may be built on one server or may be distributed within a network of servers. Therefore we need to consider the computation time within the distribution. Object-level security needs may increase transformation processing time and thus may make it harder to meet processing window schedules.

References

Generic Entity Types

separation of concern
depth normalisation
componentisation
view towers instead of redundant storage
<table>
<thead>
<tr>
<th>Entity Name</th>
<th>Entity Type</th>
<th>Entity Attribute Name</th>
<th>Entity Attribute Is Required</th>
<th>Entity Attribute Is FK</th>
<th>Entity Attribute Is FK</th>
<th>Entity Attribute Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>CERTIFICATION REQUIREMENT</td>
<td>Dependent</td>
<td>HEALTH CARE OFFERING ID</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ID</td>
</tr>
<tr>
<td>CERTIFICATION REQUIREMENT</td>
<td>Dependent</td>
<td>CERTIFICATION TYPE ID</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ID</td>
</tr>
<tr>
<td>CERTIFICATION STATUS</td>
<td>Dependent</td>
<td>STATUS TYPE ID</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ID</td>
</tr>
<tr>
<td>CERTIFICATION STATUS</td>
<td>Dependent</td>
<td>HEALTH CARE DELIVERY ID</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ID</td>
</tr>
<tr>
<td>CERTIFICATION STATUS TYPE</td>
<td>Dependent</td>
<td>STATUS DATE</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Date</td>
</tr>
<tr>
<td>CERTIFICATION TYPE</td>
<td>Independent</td>
<td>CERTIFICATION TYPE ID</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>ID</td>
</tr>
<tr>
<td>CERTIFICATION TYPE</td>
<td>Independent</td>
<td>DESCRIPTION</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Description</td>
</tr>
<tr>
<td>DELIVERY OUTCOME</td>
<td>Dependent</td>
<td>HEALTH CARE DELIVERY ID</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ID</td>
</tr>
<tr>
<td>DELIVERY OUTCOME</td>
<td>Dependent</td>
<td>DELIVERY OUTCOME ID</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>ID</td>
</tr>
<tr>
<td>DELIVERY OUTCOME</td>
<td>Dependent</td>
<td>OUTCOME TYPE ID</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>ID</td>
</tr>
<tr>
<td>DELIVERY OUTCOME</td>
<td>Dependent</td>
<td>OUTCOME EXPLANATION</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Long varchar</td>
</tr>
<tr>
<td>DIAGNOSIS</td>
<td>Independent</td>
<td>DIAGNOSIS ID</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>ID</td>
</tr>
<tr>
<td>DIAGNOSIS</td>
<td>Independent</td>
<td>HEALTH CARE EPISODE ID</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>ID</td>
</tr>
<tr>
<td>DIAGNOSIS</td>
<td>Independent</td>
<td>DIAGNOSIS TYPE ID</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>ID</td>
</tr>
<tr>
<td>DIAGNOSIS</td>
<td>Independent</td>
<td>DIAGNOSIS DATE</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Date</td>
</tr>
<tr>
<td>DIAGNOSIS TREATMENT</td>
<td>Dependent</td>
<td>DIAGNOSIS ID</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ID</td>
</tr>
<tr>
<td>DIAGNOSIS TREATMENT</td>
<td>Dependent</td>
<td>HEALTH CARE DELIVERY ID</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ID</td>
</tr>
<tr>
<td>DIAGNOSIS TYPE</td>
<td>Independent</td>
<td>DIAGNOSIS TYPE ID</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>ID</td>
</tr>
<tr>
<td>DIAGNOSIS TYPE</td>
<td>Independent</td>
<td>DESCRIPTION</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Description</td>
</tr>
<tr>
<td>DIAGNOSIS TYPE</td>
<td>Independent</td>
<td>ABBREVIATION</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Short varchar</td>
</tr>
<tr>
<td>DME DELIVERY</td>
<td>Dependent</td>
<td>HEALTH CARE DELIVERY ID</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ID</td>
</tr>
<tr>
<td>DRUG ADMINISTRATION</td>
<td>Dependent</td>
<td>HEALTH CARE DELIVERY ID</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ID</td>
</tr>
<tr>
<td>EPISODE OUTCOME</td>
<td>Dependent</td>
<td>HEALTH CARE EPISODE ID</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ID</td>
</tr>
<tr>
<td>EPISODE OUTCOME</td>
<td>Dependent</td>
<td>EPISODE OUTCOME ID</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>ID</td>
</tr>
<tr>
<td>EPISODE OUTCOME</td>
<td>Dependent</td>
<td>OUTCOME TYPE ID</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>ID</td>
</tr>
<tr>
<td>EPISODE OUTCOME</td>
<td>Dependent</td>
<td>OUTCOME EXPLANATION</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Long varchar</td>
</tr>
<tr>
<td>EXAMINATION</td>
<td>Dependent</td>
<td>HEALTH CARE DELIVERY ID</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ID</td>
</tr>
</tbody>
</table>
Fig. 4. Entity and Relationship Types (cutout)
The Person and Organization Dimension  People are characterized by their names. The structure of names is rather complex. A person may have many different roles at the same time. The type Person is used for storing information independent of jobs or roles. Since skills are common to several people we factor out the skills and education profiles. Due to the complexity of the identification of people we add an artificial identification ID to the type Person.

People and organizations are similar in various aspects. They act as parties in contracts, in work activities etc. Thus, we use the type Party in order to merge the common characteristics. If we decide to model organizations and people in separate fashion the number of relationship types associating people or organizations to their roles have to be doubled.

Since a person and an organization may be involved in a number of roles we categorize roles of parties by adding the component Role_Type to the type Party. Typical values in \text{dom(Role\_Type\_Description)} are “customer”, “supplier” and “producer”. Properties such as Credit_Rating may vary depending on the role the party is playing. Therefore this attribute is added to Party instead of adding the attribute to Person or Organization.

Parties may be in various relations. These relations vary a lot. We can represent any kind of relations in a separate relationship type. We decide however to use the characterization of relations by Kind. Priorities may range from “low” to “high”. Since relations vary over the lifespan of party objects we can also use the type Status for defining the current value. Typical values in \text{dom(Status\_Description)} are “active”, “inactive”.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{herm-diagram.png}
\caption{HERM Diagram of Person and Organization Pattern}
\end{figure}

In Figure 5 the pattern on people and organizations is represented. The pattern can be extended by more specific modelling of skills and education [AAB+98].

The Address Dimension here: demonstration of pattern construction by separation of concern, injection of quality dimensions, overlaying with utilisation, introducing a log dimension

Handout mit Schema aus Vorlesung Informationsdienste 3 anstatt der anderen Schemata aus Preprint 09/2000

Addresses are one of the main component used for characterization of parties. The address pattern is based on the shows the structuring of the address dimension. Addresses are used in various occasions:

\footnote{Depending on the cultural environment people may have one or several first names, one or several last names, a suffix in the last name carrying information on the mothers maiden name or on the father’s name, a prefix carrying information on the first son if there is any etc. We are aware of these specific naming but we do not want to make the type too complex.}
Addresses are used for network contacts. These contacts are separated into email or phone or facsimile. Contacts have some properties in common. Others are specific to the form.

Addresses are used for mailing purposes. The form is dependent on the geographical area, on the communication forms used in the area. Geographical addresses follow the description of geographic boundaries.

Since we use addresses for the business we want to log the utilization. This utilization depends on the business process. In the codesign approach data are used in dialogue steps on the basis of provided media objects. In some cases, we archive the utilization for the sake of later retrieval.

The address pattern pictured in Figure 7 can be extended in various ways depending on the business processes:

- The relationship to parties may depend on the business task. In some cases one address is used, in others a different choice must be made.
- The contact time depends on time zones. Thus, we might add various characteristics applicable to time zones and areas.
- Streets may belong to various transport zones. Cities may have a number of districts. A street may belong to one or more districts.
- Search may be supported by algorithms such as SoundEx. In some cases we might be interested in explicit representation of an applicable ontology and of phonetic search.
- Countries may have a number of properties which are of interest as well, e.g. rules of business, rules of taxing, statistics, membership in international organizations, rules of business transactions, rules of pricing and accounting and rules of contacting people.
- Streets may have a long name and a number of applicable short names or abbreviations\(^3\). If we are interested in support for transportation then the geographical relation of streets is of interest. Bulk names are also applicable to city names.

\(^3\) By using such bulk names errors encountered by DATRAS in the German Phone Directory will be decreased from about 15% to about 4%.
Fig. 7. HERM Diagram of the Address-Party Association
Relational Star Schema for Person

CREATE TABLE person (  
    person_id NUMBER NOT NULL,  
    first_name VARCHAR2(20) NULL,  
    mi VARCHAR2(20) NULL,  
    gender VARCHAR2(20) NULL,  
    date_of_birth DATE NULL,  
    date_of_death DATE NULL  
);  
ALTER TABLE person  
    ADD ( PRIMARY KEY (person_id) ) ;

CREATE TABLE person_language (  
    person_language_seq_id NUMBER NOT NULL,  
    person_id NUMBER NOT NULL,  
    language_code VARCHAR2(20) NULL  
);  
ALTER TABLE person_language  
    ADD ( PRIMARY KEY (person_language_seq_id, person_id) ) ;

CREATE TABLE person_address (  
    person_address_seq_id NUMBER NOT NULL,  
    person_id NUMBER NOT NULL,  
    address_type VARCHAR2(20) NULL,  
    address_line_1 VARCHAR2(20) NULL,  
    address_line_2 VARCHAR2(20) NULL,  
    city VARCHAR2(20) NULL,  
    state VARCHAR2(20) NULL,  
    zip VARCHAR2(20) NULL  
);  
ALTER TABLE person_address  
    ADD ( PRIMARY KEY (person_address_seq_id, person_id) ) ;

CREATE TABLE person_phone (  
    person_phone_seq_id NUMBER NOT NULL,  
    person_id NUMBER NOT NULL,  
    phone_type VARCHAR2(20) NULL,  
    area_code NUMBER NULL,  
    phone_number NUMBER NULL,  
    extension NUMBER NULL  
);  
ALTER TABLE person_phone  
    ADD ( PRIMARY KEY (person_phone_seq_id, person_id) ) ;

ALTER TABLE person_language  
    ADD ( FOREIGN KEY (person_id)  
          REFERENCES person ) ;

ALTER TABLE person_address  
    ADD ( FOREIGN KEY (person_id)  
          REFERENCES person ) ;

ALTER TABLE person_phone  
    ADD ( FOREIGN KEY (person_id)  
          REFERENCES person ) ;