Chapter 12
Design of an IT System for Hepatocellular Carcinoma

Leonard Berliner and Heinz U. Lemke

Abstract During the development of an Information Technology System for Predictive, Preventive and Personalized Medicine (ITS-PM) for hepatocellular cancer (HCC) a wide number of variables or Information Entities (IEs) will be identified, and their relative value will be determined. These include factors reflecting: (1) clinical assessment of the patient including functional status, liver function, degree of cirrhosis, and comorbidities; (2) tumor biology, at a molecular, genetic, and anatomic level; (3) tumor burden and individual patient response; and, (4) medical and operative treatments and their outcomes. Beyond the development of database systems, our goals include the development of a realistic, plausible approach to the development of Digital Patient Models (DPMs) and Model Guided Therapy (MGT). These will be based on a complex of database and knowledge management systems capable of data storage, data mining, data analysis, and decision support. In this Chapter we have outlined the required structure and function of an ITS-PM that would be suitable for these tasks. The database structure, composed of three layers, has been described and sample entity-relationship diagrams populated from the clinical material described in Chaps. 3–11 have been presented. Methodologies are proposed that include Multi-Entity Bayesian Networks (MEBN), Reference Model for Open Distributed Processing (RM-ODP), and Service-Oriented Architecture (SOA), which can be considered a subset of RM-ODP. These can provide the comprehensive techniques and structures to be employed to successfully meet the requirements for an ITS-PM.

Keywords Personalized medicine · Hepatocellular carcinoma · Information technology · Model guided therapy · Information technology system for predictive,

L. Berliner (✉)
Weill Medical College of Cornell University, New York, USA
e-mail: LeonardB23@AOL.com

New York Methodist Hospital, Brooklyn, NY, USA

H. U. Lemke
Technical University of Berlin, Berlin, Germany
e-mail: hulemke@cars-int.org

University of Southern California, Los Angeles, USA

preventive and personalized medicine (ITS-PM) · Digital patient model · Patient-specific model · Model-based medical evidence · Bayesian network · Reference model for open distributed processing (RM-ODP) · Service-oriented architecture (SOA)

12.1 Introduction

In the preceding Chapters we have defined the current state of knowledge, as well as the limits of our knowledge, with respect to hepatocellular carcinoma (HCC). To advance Predictive, Preventive and Personalized Medicine (PPPM), we will be examining HCC in the final two chapters from a more integrated point of view, combining epidemiology, risk factors, infectious etiologies, pathology, microenvironment, biomarkers, screening and diagnostic technologies, and treatment modalities (single, combined, and/or sequential). In this Chapter we will be exploring the ways in which Information Technology (IT) may optimize our ability to manage patients with HCC in a multidisciplinary setting along with Model-Guided Therapy (MGT) as outlined in Chaps. 1 and 2. This will require the development of systems to provide unified access to general medical and patient-specific information for medical researchers and health care providers from different disciplines including hepatologists, gastroenterologists, medical and surgical oncologists, liver transplant teams, interventional radiologists, and radiation oncologists.

It is our assumption that the development of improved IT will promote an approach based on a global understanding of disease and treatment outcomes, rather than reliance primarily upon local availability and expertise. To this end, we need technologies and information systems to optimize the vast amount of information in various repositories by these health care providers and investigators from random controlled trials (RCTs), as well as other data sources.

With this in mind, we will begin to explore the daunting task of defining the IT specifications that would fulfill the requirements for an Information Technology System for Predictive, Preventive and Personalized Medicine (ITS-PM), using a model of HCC as a use-case. Ultimately, to handle the vast amount of available information, we will need to define and develop new types of database solutions and end-user applications. The database solutions should include certain features—easily accessible links to data sources and repositories, functionality that is well organized and easily expandable, the facility for queries that will promote probabilistic and statistically valid investigations, and, features to facilitate decision support and research.

Beyond the selection and development of database systems, the larger task is to find a way of using IT to pool, integrate, and correlate the following: (1) the clinical information relating to diagnosis and treatment of HCC; (2) the research data relating to epidemiology, virology, and pathology at the anatomic, molecular, and genetic levels; and, (3) the role of MGT and Patient-Specific Modeling. One of our goals is to propose a realistic, plausible approach to the development of Digital

Patient Models (DPMs), based on a complex of database and knowledge management systems capable of data storage, data mining, data analysis, and decision support. At this time, there is no system or collection of systems on the market that can accomplish these tasks. In this Chapter we will undertake a systematic approach to identify, analyze, and organize a combination of actual and/or potential software entities that could be assembled with the appropriate architecture to achieve these goals. At this time, the tools that are available to us include database management systems, physiologic models, web services, other mid-layer services, and a variety of tools to create appropriate end-user software and graphics applications. These components would be combined to form a subset of a much larger and more comprehensive Therapy and Imaging Model Management Systems (TIMMS) system as described in Chap. 2.

HCC has been selected as a "use-case" for the development of an ITS-PM. A tentative IT framework, composed of a variety of components, will be described that has the capability to integrate the following: the Patient-Specific Model (PSM) itself (that includes the complete medical description of any number of patients), and the various sources of medical information that may be local or remotely accessed through the Internet. It should be possible to view and access the proposed ITS-PM from multiple points of view, to extract different kinds of information and perform different kinds of tasks by medical practitioners, researchers, and epidemiologists. For example, user interface requirements for the medical oncologist versus the geneticist evaluating DNA sequences will be quite different. This not only reflects the different tasks, and therefore the different needs of the end-users, but also reflects that each end-user will have a somewhat different view of the DPM, itself. The complete collection of DPM databases can provide a view or representation of the patient as required for a variety of specific tasks, whether they are related to achieving improved treatment outcomes, enhanced patient safety, and/or for engaging in basic medical research (Fig. 12.1).

A *Precision Surgery View* may be utilized to enhance surgical guidance for improved safety and efficiency; a *Surgical Workflow View* may be employed in the Operating Room to optimize the surgical process; a *Physiological View* would optimize the process of patient monitoring; a *Decision Support View* would provide assistance in the selection of best treatments; a *Biomarkers and Imaging View* could be employed to help gain a deeper understanding of disease fundamentals, e.g. oncology; and, a *Disease/Epidemiology View* may be utilized to pool large numbers of DPMs to gain insight into patient populations and epidemiology (Model-based Medical Evidence [MBME]).

A few points, from the Chaps. 3 through 11, will serve as reminders of the complexity of creating an ITS-PM for HCC: (1) the treatment spectrum for HCC extends from one extreme to the other, i.e. from transplantation of the entire liver to targeted therapy with Sorafenib at the molecular level; (2) HCC is often treated without tissue diagnosis, i.e. with radiologic and biochemical confirmation; (3) the understanding of the hepatic microenvironment and its relationship to HCC is evolving; and, (4) there are limitations in the RCTs comparing different minimally invasive treatments and/or their roles in down-staging of advanced cases. The science behind
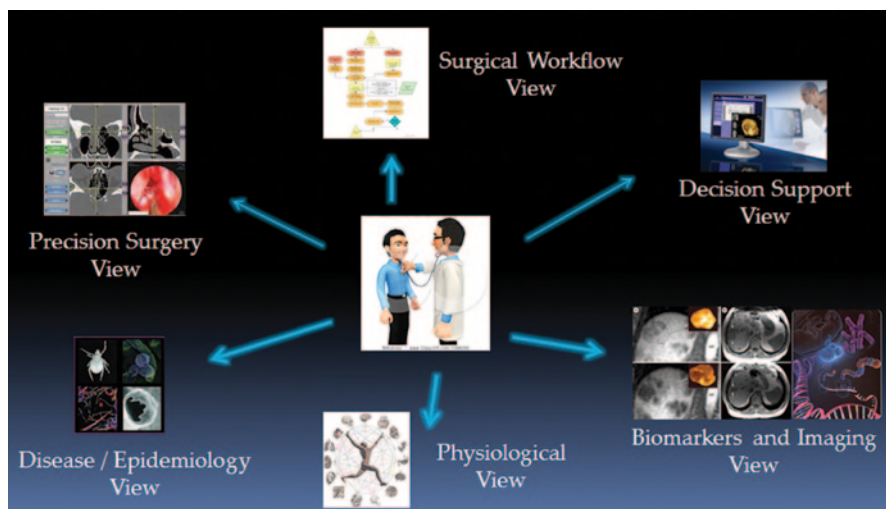
Fig. 12.1 The complete digital patient model provides a variety of views, or representations, of the patient depending on the specific tasks, requirements or areas of interest of the end-user

our treatment choices can be thought of as being in a state of evolution. There are differences of opinion, as well as newly emerging evidence, concerning many facets of HCC and its treatment. Therefore, the ITS-PM system under development must be sufficiently broad, sensitive, and flexible enough to help organize and make sense out of the widespread and disparate information available. It is hoped that the ITS-PM will help fill the gaps of our knowledge by incorporating and integrating new information into the existing fund of medical knowledge and help us make the best decisions for our patients, even when medical knowledge is incomplete. As in any medical decision support system, it is important to emphasize that the role of the ITS-PM is not to replace the physician in decision making, but rather to assist the decision making process, such as at a hospital's Tumor Board.

In summary, the development of the ITS-PM for HCC will provide a comprehensive system to identify and then determine the relative value of the wide number of variables: (1) factors reflecting clinical assessment of the patient including functional status, liver function, degree of cirrhosis, and comorbidities; (2) factors reflecting tumor biology at a molecular, genetic, and anatomic level; (3) factors reflecting tumor burden and individual patient response; and (4) factors reflecting medical and operative treatments and their outcomes. If this project is successful, it can serve as a prototype for IT solutions to assist in the diagnosis, research, and management of other cancers as well as non-malignant diseases.

Fig. 12.2 A graphic representation of the points of view utilized in the reference model for open-distributed processes. (Adapted from [1])

Open-Distributed Processing View Model
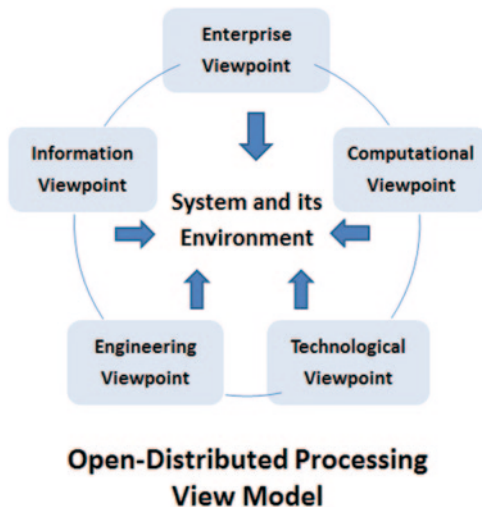
12.1.1   ITS-PM: Organization and Architecture

12.1.1.1   Requirements for an ITS-PM

The first task is to consider and define the requirements for an IT approach for PPPM with respect to HCC. It is probably best if we divide this task into broad categories, each of which will have its own focus, data types, tasks, and solutions.

Reference Model for Open Distributed Processes and Service-Oriented Architecture

It is imperative that comprehensive and cohesive hardware and software architecture is provided for the ITS-PM to allow each section to function independently, while synchronized and in communication with each other section. Reference Model for Open Distributed Processing (RM-ODP) and Service-Oriented Architecture (SOA) (which may be considered a related subset of RM-ODP and is perhaps more widely known) are standards, methodologies, or approaches to enterprise system development that could help fulfill the necessary requirements.

   RM-ODP is an International Organization for Standardization (ISO) standard that gives a solid basis for describing and building widely distributed systems and applications in a systematic way. Emphasis is placed on the need to build such systems with evolution in mind by identifying the concerns of major stakeholders and then expressing the design as a series of linked viewpoints representing these concerns. Each stakeholder can then develop an appropriate view of the system with a minimum of interference from the others [1] (Fig. 12.2) (Table 12.1).

Table 12.1 Viewpoints utilized in the reference model for open-distributed processes. (Adapted from [1])

Viewpoints for reference model for open-distributed processes
The *enterprise viewpoint* focuses on the organizational situation in which the design activity is to take place. It concentrates on the objectives, business rules, and operational policies that need to be supported by the system being designed.
The *information viewpoint* concentrates on the modeling of the shared information manipulated within the enterprise of interest. By providing a common model that can be referenced from throughout a complete piece of design, we can ensure that the same interpretation of information is applied at all points.
The *computational viewpoint* is concerned with the development of the high-level design of the processes and applications supporting the enterprise activities. It uses the familiar tools for object-oriented software design, expressing its models in terms of objects with strong encapsulation boundaries, interacting at typed interfaces by performing a sequence of operations (or passing continuous streams of information).
The *engineering viewpoint* tackles the problem of diversity in infrastructure provision; it gives the prescriptions for supporting the necessary abstract computational interactions in a range of different situations. It thereby offers a way to avoid lock-in to specific platforms or infrastructure mechanisms.
The *technology viewpoint* is concerned with managing real-world constraints, such as restrictions on the hardware available to implement the system within budget, or the existing application platforms on which the applications must run.

Once the requirements and the approach to fulfill these requirements have been developed, reviewed, and approved by the overall team, the wide variety enterprise software components need to be created and assembled. SOA provides the infrastructure and organization required for both connectivity and interaction between a wide variety of programs and functions (services) that may be written in different software languages to provide proper and secure transactions. SOA does not imply a specific technology or creation of a single all-encompassing program. Rather, SOA is an architectural paradigm and discipline that may be used to build infrastructures enabling those with needs (consumers) and those with capabilities (providers) to interact via services across disparate domains of technology and ownership [2].

Implementation of a SOA will provide for user interfaces, messaging between users, storage of data, access to data and services, establishment of workflow processes, and system security. When properly conceived, SOA is sufficiently flexible to allow incremental development and implementation of the functionality required by the organization. While SOA is often associated with Web Services, it is important to understand that the services provided by SOA need not be web based. SOA is often associated with the streamlining of business practices; however, the organization, interchangeability, and flexibility of SOA can provide advantages for the scientific and medical community as well, that faces similar obstacles created by the wide variety of software and IT tools that are currently difficult to integrate. For the purposes of this article, the importance of SOA resides in its ability for the scientific and medical community to find a realistic methodology for creating a useable and secure system, composed of complex and disparate entities, including Electronic
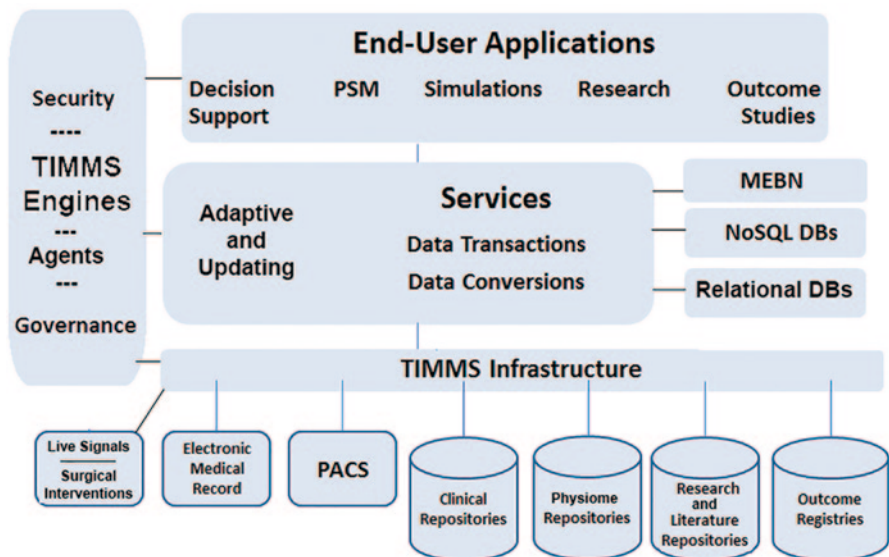
Fig. 12.3 A schematic for organization of an ITS-PHC. This diagram reorganizes many of the TIMMS components in a structure that will enable the secure interchange of information between data sources, database management systems, data analysis systems, and end-user applications. (Legend: *PSM* patient specific model; *TIMMS* therapy and imaging model management system; *PACS* picture archiving and communications system; *MEBN* multi-entity Bayesian network; *NoSQL* not only structured query language; *DBs* databases)

Medical Records, Hospital and Radiology Information Systems, research databases and repositories, as well as the database systems that will form the core of an ITS-PM.

It is beyond the scope of this article to provide a complete RM-ODP enterprise proposal with detailed SOA schema. However, we will try to explore and define the overall objectives and processes (enterprise viewpoint), the requirements relating to data types and data exchange (information viewpoint), and the software categories (computational viewpoint). (In some cases, specific software components, categories or products may be mentioned. However, at this stage of development this is done for illustrative purposes only to indicate the feasibility of a required technology or process. Architectural detail, as well as specific hardware and software selection and development, would be determined much later in the project.) A simplified schematic for the organization of an ITS-PM is presented in Fig. 12.3.

Data Exchange

Provision needs to be made for the exchange of data and interchange of data types between the various forms of databases that will be accessed, processed, and analyzed by the proposed ITS-PM. The vast amounts of data that are available may reside within Electronic Medical Records, Hospital and Radiology Information

Systems, research databases and repositories, in the form of relational databases, multi-dimensional databases, or newer NoSQL databases that may be of several types. The data types utilized within the ITS-PM may include strings, numbers, Boolean functions, images, text files, and XML documents. Much of this information is already in a format that can be utilized for data analysis. However, many entries into the medical record are not in a format that can be readily assimilated and analyzed in an automated IT system. Efforts have been made to create structured reports in Radiology, such as Digital Imaging and Communications in Medicine Structured Report (DICOM SR) [3] and RadLex [4], in which data is stored in retrievable format, such as XML and JSON. It ultimately may be required that full implementation of the ITS-PM may require extensive use of Structured Reports, in an as yet to be defined format.

Database Systems

A wide variety of database systems are currently available and in widespread use. They may be found in hospital information systems, throughout business and internet enterprises, government organizations, and personal computer programs. The most commonly employed databases today, relational databases (RDBs), are based on relational database management systems (RDBMS), in which data are stored in tables that are linked by designated relationships. Data are most commonly extracted from these databases by Structured Query Language (SQL) queries.

A new class of database systems recently has been developed and is known as NoSQL ("Not only SQL"). These systems do not rely primarily on tables, and therefore generally do not use SQL for data manipulation. These databases differ from RDBs in the great speed with which they can handle and sort through large volumes of information and relationships, thereby enabling systems such as Google and Facebook. NoSQL databases may be designed to store records (e.g. key-value stores), to store documents (e.g. XML documents), and/or to store data, whose relations are well depicted and utilized with graphs and graph theory.

The proposed ITS-PM will most likely need to be able to make use of several types of database management systems, in both core programs and data repositories. Thus, the ITS-PM will be well-equipped for different functionalities.

Model Creation

The content and the organization of the ITS-PM should be flexible enough to allow manipulation of the information required for constructing a variety of models to support MGT. This may include, but would not be limited to, models of the patient, i.e. the DPM. It could also include the ability to create models of disease processes, models of patient populations, as well as models of genetic, physiologic, and molecular processes.

The design and structure of the DPM will be discussed in greater detail below.

Clinical Decision Support

The ITS-PM should provide a variety of functions, including data-mining and data-analysis to detect correlations, and ultimately, to reveal and elucidate causal relationships between the patient, the disease processes, and exogenous factors. Through these functions, it is proposed that the ITS-PM will assist in: (1) the understanding of diseases in individuals and populations; (2) basic medical research; and, (3) clinical decision support.

It is important that safeguards be established to ensure that objectivity and strict statistical methodology be employed to prevent erroneous conclusions to be drawn from rapidly accumulating data (e.g. "correlation does not imply causation"). This is especially true in medicine, in which decisions often are made with the best available information (i.e. incomplete knowledge).

12.1.2 Clinical and Research Components

12.1.2.1 Defining the Requirements of the ITS-PM

In this Chapter, the objectives of the ITS-PM will be explored with an emphasis on defining the major processes that will be required, as well as their categories and components (enterprise viewpoint). As these processes are brought into focus, the specific required data types and the requirements relating to data exchange will be enumerated (information viewpoint). The major mid-level software functions and end-user applications will be discussed (computational viewpoint). At this stage of development the following will be considered: (1) the DPM (relational and NoSQL database management systems); (2) clinical decision support including predictive and simulation functions with a Multi-entity Bayesian Network (MEBN) [5, 6]; (3) access to medical research databases; and, (4) modules for outcomes studies for the development of disease models, relating to individuals and populations, as well as for the evaluation of treatment protocols and technologies.

The DPM: Information Entities, PSM Template and MEBN

It is essential that the DPM should have the capacity to contain and organize information of any type that may be medically relevant. It will be required that these attributes will ultimately be organized into structures that can be utilized in a MEBN. At this time, for a DPM to achieve the wide range of functions that have been described, it would appear that the database structure be divided into three functional components or layers. These layers, which are more descriptive than physical, would include: (1) an Attribute Layer for data storage that would be best served with RDBs; (2) a Probabilistic Layer for data analysis and decision support that may be best served with MEBNs and graph theory; and, (3) an Action Layer that
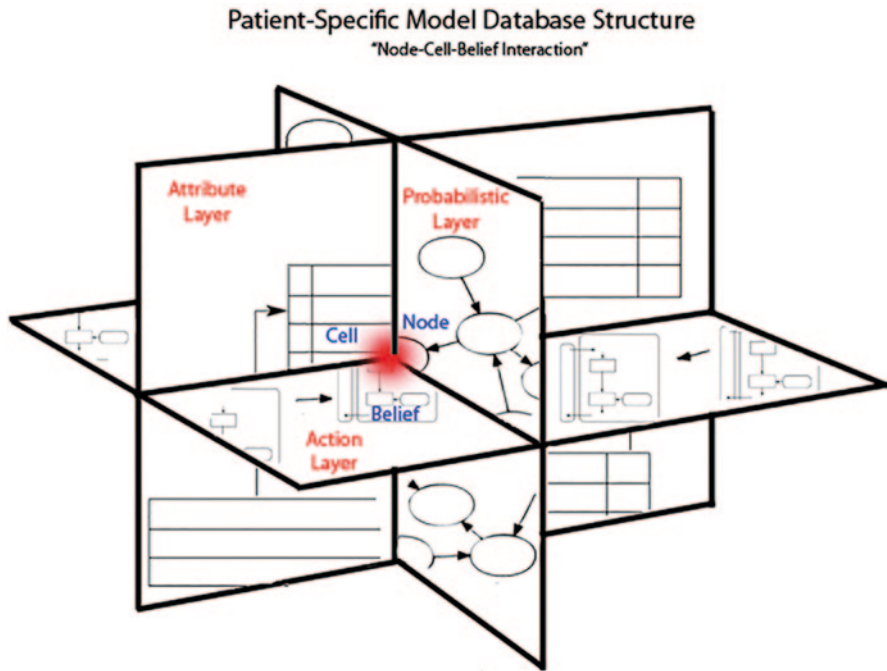
Fig. 12.4 The database structure of the generic PSM may be defined as three converging layers that allow the generic PSM to perform the many tasks assigned to it. Any given data point may contain the value of a patient attribute or Information entity that is associated with a certain probability distribution with respect to a clinical inquiry, and may be acted upon in decision support processes

would actively update databases and perform statistical analyses at specified times. (Fig. 12.4) Therefore, the database structure of a generic PSM may be defined as three converging layers that allow the PSM to perform the many tasks assigned to it. Any given data point may contain the value of a patient attribute or Information Entity (IE) that is associated with a certain probability distribution with respect to a clinical inquiry, and may be acted upon in decision support processes.

The first task in constructing a generic PSM is to organize the patient-specific information according to a generalized hierarchy of attributes or IEs, extending from most general to specific, as outlined in Chap. 2. From these IEs, an Entity-Relationship Diagram (ERD) may be designed, (Figs. 12.5a and b) from which a RDB may be constructed, as part of the Attribute Layer of the generic PSM. This RDB that will be populated with data from the many sources previously illustrated in Fig. 12.3 (the schematic for organization of an ITS-PM) in accordance with the organization described in the generic PSM template (also defined in Chap. 2), will serve as the reservoir of clinical data, biomarkers, images, and physiologic signals that will be utilized by the PSM. Information is accessed from RDBs by means of SQL queries.

Clinical Decision Support

Clinical decision support functions will reside predominantly within the Probabilistic Layer of the generic PSM. These functions will be available for evaluation and reorganization within categories of risk, diagnosis, prognosis, and treatment response for the purposes of clinical decision support. As envisioned in the proposed ITS-PM, the IEs stored within the RDB (including patient attributes, biomarkers, clinical data, and imaging data) will have greatly enhanced value in decision support systems when incorporated into MEBN and graph database systems.

As discussed in Chap. 2, in medicine, we must be able to reason in the presence of incomplete data and knowledge. There may be uncertainty regarding the existence of relationships among pieces of medical information, the strength of those relationships, and, constraints governing those relationships, such as, cause and effect. Bayesian inference and probability are logically coherent and provide tools and methodology to combine expert knowledge with statistical data, to represent cause-and-effect relationships, to learn from observations, to prevent over-fitting, and, to provide clear and understandable semantics. The ITS-PM will be able to make use of the Bayesian Belief Network or Model that is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG) (Fig. 12.6). In a DAG,

Fig. 12.5 a A portion of a simplified entity-relationship diagram for a relational database is shown displaying the 1st, 2nd, and 3rd order information entities (as defined in Chap. 2) of a generic patient-specific model, and a 4th order entity: Hepatocellular Carcinoma. (Lege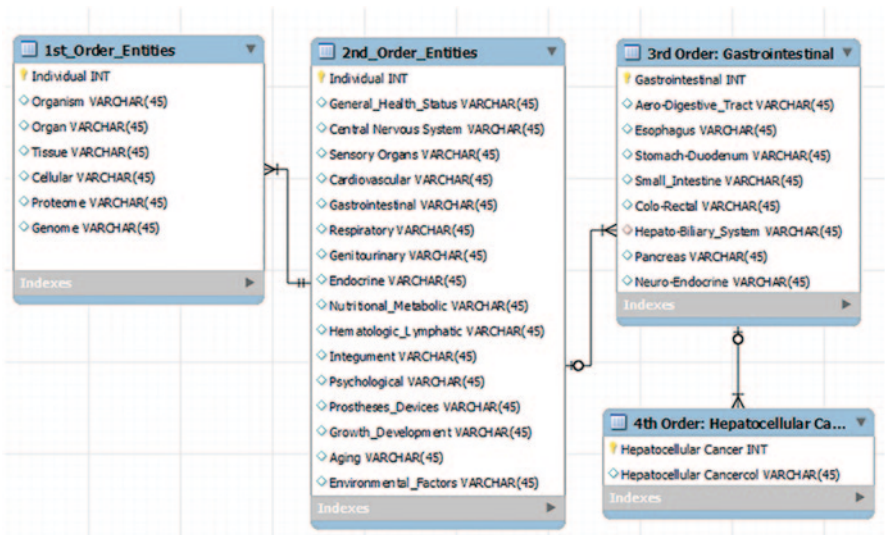nd: *INT* integer; *VARCHAR* includes text [characters, numbers, punctuation]). **b** A portion of a simplified entity-relationship diagram for a relational database is shown displaying the 5th order information entities relating to Hepatocellular Carcinoma
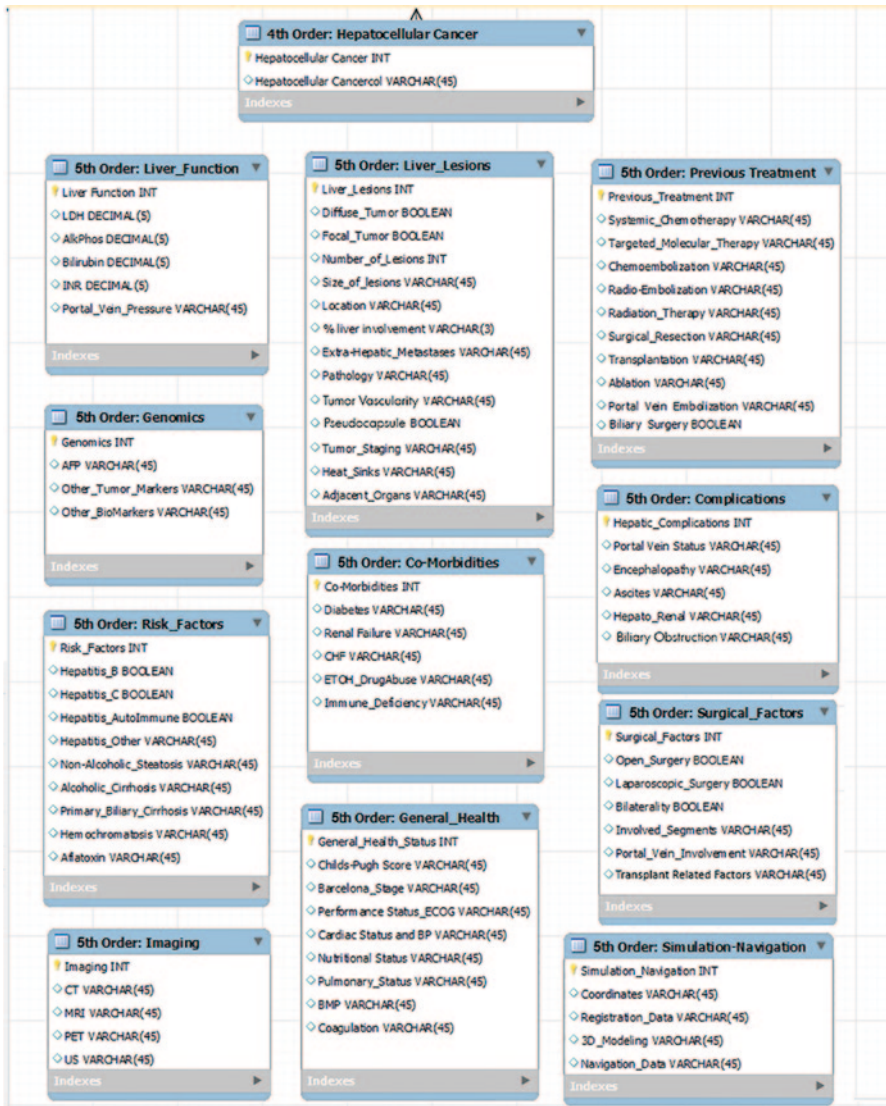
4th Order: Hepatocellular Cancer
- Hepatocellular Cancer INT
- Hepatocellular Cancercol VARCHAR(45)
- Indexes

5th Order: Liver_Function
- Liver Function INT
- LDH DECIMAL (5)
- AlkPhos DECIMAL(5)
- Bilirubin DECIMAL(5)
- INR DECIMAL (5)
- Portal_Vein_Pressure VARCHAR(45)
- Indexes

5th Order: Genomics
- Genomics INT
- AFP VARCHAR(45)
- Other_Tumor_Markers VARCHAR(45)
- Other_BioMarkers VARCHAR(45)
- Indexes

5th Order: Risk_Factors
- Risk_Factors INT
- Hepatitis_B BOOLEAN
- Hepatitis_C BOOLEAN
- Hepatitis_AutoImmune BOOLEAN
- Hepatitis_Other VARCHAR(45)
- Non-Alcoholic_Steatosis VARCHAR(45)
- Alcoholic_Cirrhosis VARCHAR(45)
- Primary_Biliary_Cirrhosis VARCHAR(45)
- Hemochromatosis VARCHAR(45)
- Aflatoxin VARCHAR(45)
- Indexes

5th Order: Imaging
- Imaging INT
- CT VARCHAR(45)
- MRI VARCHAR(45)
- PET VARCHAR(45)
- US VARCHAR(45)
- Indexes

5th Order: Liver_Lesions
- Liver_Lesions INT
- Diffuse_Tumor BOOLEAN
- Focal_Tumor BOOLEAN
- Number_of_Lesions INT
- Size_of_lesions VARCHAR(45)
- Location VARCHAR(45)
- % liver involvement VARCHAR(3)
- Extra-Hepatic_Metastases VARCHAR(45)
- Pathology VARCHAR(45)
- Tumor_Vascularity VARCHAR(45)
- Pseudocapsule BOOLEAN
- Tumor_Staging VARCHAR(45)
- Heat_Sinks VARCHAR(45)
- Adjacent_Organs VARCHAR(45)
- Indexes

5th Order: Co-Morbidities
- Co-Morbidities INT
- Diabetes VARCHAR(45)
- Renal Failure VARCHAR(45)
- CHF VARCHAR(45)
- ETOH_DrugAbuse VARCHAR(45)
- Immune_Deficiency VARCHAR(45)
- Indexes

5th Order: General_Health
- General_Health_Status INT
- Childs-Pugh Score VARCHAR(45)
- Barcelona_Stage VARCHAR(45)
- Performance Status_ECOG VARCHAR(45)
- Cardiac Status and BP VARCHAR(45)
- Nutritional Status VARCHAR(45)
- Pulmonary_Status VARCHAR(45)
- BMP VARCHAR(45)
- Coagulation VARCHAR(45)
- Indexes

5th Order: Previous Treatment
- Previous_Treatment INT
- Systemic_Chemotherapy VARCHAR(45)
- Targeted_Molecular_Therapy VARCHAR(45)
- Chemoembolization VARCHAR(45)
- Radio-Embolization VARCHAR(45)
- Radiation_Therapy VARCHAR(45)
- Surgical_Resection VARCHAR(45)
- Transplantation VARCHAR(45)
- Ablation VARCHAR(45)
- Portal Vein_Embolization VARCHAR(45)
- Biliary_Surgery BOOLEAN
- Indexes

5th Order: Complications
- Hepatic_Complications INT
- Portal Vein Status VARCHAR(45)
- Encephalopathy VARCHAR(45)
- Ascites VARCHAR(45)
- Hepato_Renal VARCHAR(45)
- Biliary Obstruction VARCHAR(45)
- Indexes

5th Order: Surgical_Factors
- Surgical_Factors INT
- Open_Surgery BOOLEAN
- Laparoscopic_Surgery BOOLEAN
- Bilaterality BOOLEAN
- Involved_Segments VARCHAR(45)
- Portal_Vein_Involvement VARCHAR(45)
- Transplant Related Factors VARCHAR(45)
- Indexes

5th Order: Simulation-Navigation
- Simulation_Navigation INT
- Coordinates VARCHAR(45)
- Registration_Data VARCHAR(45)
- 3D_Modeling VARCHAR(45)
- Navigation_Data VARCHAR(45)
- Indexes

Fig. 12.5 (continued)

each node (numbered circle) presents the attributes of each random variable or IE, while each edge (arrow) indicates the conditional dependency.

Bayesian networks are used for evidential reasoning or explanation. For example, a Bayesian network can be used to represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases.

As described in Chap. 2, building on the basic Bayesian Network, a MEBN is a logic system that integrates first-order logic with Bayesian probability theory [7]. A
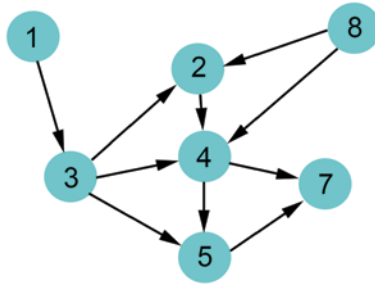
Fig. 12.6 Directed acyclic graph (DAG). Each node (*numbered circle*) presents the attributes of each random variable or information entity, while each edge (*arrow*) indicates the conditional dependency. For a multi-entity Bayesian network, the edges of the DAG also provide validated probability distributions, beyond conditional dependencies

MEBN can provide a descriptive and functional framework for quantifiable medical IEs. The nodes of the DAG in a MEBN contain the attributes of each random variable or IE, as supplied by the relational database systems, while the edges provide validated probability distributions, beyond conditional dependencies (Fig. 12.6). Thus, the MEBN can mathematically provide predictive capabilities and the ability to determine cause and effect relationships, over and above the descriptive, expandable, and correlational capabilities of a simple Bayesian network. Accordingly, the value of any given entity within the RDB systems can be enhanced by determining a relative value (probability distribution) for each factor within the appropriate contexts.

To create an effective clinical decision support system for HCC, utilizing a MEBN, the IEs identified as 5th order relating to HCC, will need to be assembled into MFrags and MTheories as outlined in Chap. 2. Initially, conditional relationships between IEs and their probability distributions will be determined by medical experts utilizing the best available evidence-based medicine. One critically important factor must be understood—it is the nature of Bayesian Networks to increase in accuracy as the system is tested and more information is added, according to Pearl's Bi-directional Belief Updating Algorithm [7]. The MFrags will be assembled to form graphs, e.g. Situation Specific Bayesian Networks (SSBN) to evaluate hypothetical conditions. Support for decision constructs in MEBN will be provided via Multi-Entity Decision Graphs (MEDG). As in any decision support system, the MEBN system will require ongoing updating and validation.

While the MEBN will provide a complex system for answering specific questions relating to the management of patients with HCC, there are other tools available for the Probabilistic Layer of the PSM database system. In recent years, high-performance NoSQL databases have been used to find relationships between entities in very large networks, often with billions of objects. These database systems have been utilized for seeking information (e.g. Google), with vast social networks (e.g. Facebook), and to catalogue and find relationships in genetics research. These systems are known for their rapid answering time for complex queries, i.e.—traversals.
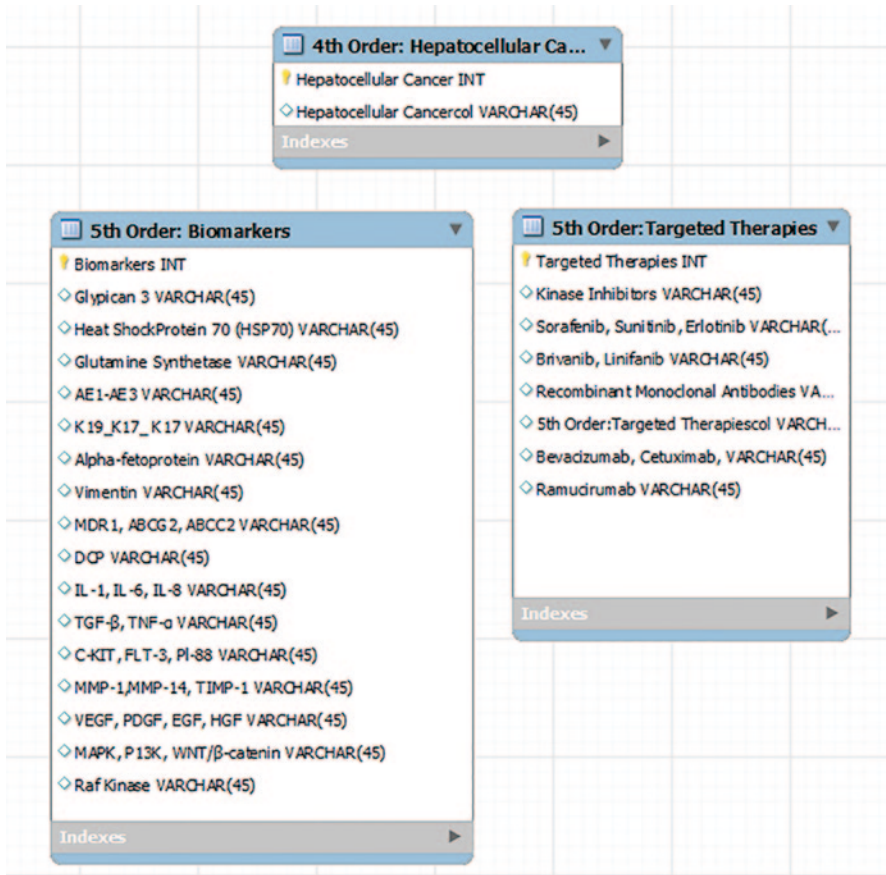
Fig. 12.7 A portion of a simplified entity-relationship diagram for a relational database that may be linked to a graph database for research in biomarker and targeted therapies is shown displaying the 5th order information entities relating to biomarkers and targeted therapies for HCC

One form of NoSQL database, the Graph Databases, may be especially useful by incorporating the IEs of the PSM template. Graph Databases can provide persistent storage for large volumes of data (nodes), to display relationships between entities implicit in the model (edges), to allow a unified view for multiple sources, and, are sufficiently flexible to manage unknown or dynamic schemas. Most importantly, Graph Databases can facilitate analysis of the connected information in network-like structures.

The ability of Graph Databases to find relationships within vast amounts of data will help provide a link between the domain of genetics and biomarkers research with the PSM. Figure 12.7 shows a portion of a simplified Entity-Relationship Diagram for a relational database that may be linked to a Graph database for research in biomarker and targeted therapies. Fifth order IEs relating to biomarkers and targeted therapies for HCC are displayed.

Action Layer

The third layer of the generic PSM database structure can be considered an Action Layer that will be designed to perform many of the tasks that will be required to update the PSM databases. The tasks performed as part of this action layer include data processing that will be required to ensure the increasing accuracy of the MEBN as indicated in Pearl's Theorem or Algorithm. These tasks may be accomplished by means of triggered sub-programs, and may include updating lab values in a graph database, recalculating probabilities in MEBNs, extracting data from structured reports such as imaging studies, and extracting data from the wide variety of local and remote repositories (e.g. genetic data). The system could be used locally at a clinical liver cancer center to monitor patient assessments, treatments, and outcomes.

This process will be facilitated when links can be established to provide access to medical research databases, as well as to established treatment registries through the TIMMS infrastructure as shown in Fig. 2.1, Chap. 2.

Conclusion

In this Chapter we have outlined the required structure and function of an ITS-PM that would be suitable to establish a use-case utilizing HCC within the context of the PSM and MGT. The database structure, composed of three layers, has been described and sample entity-relationship diagrams populated from the clinical material described in Chaps. 3–11 have been presented.

RM-ODP and SOA can provide the comprehensive methodologies to be employed to successfully meet the requirements for such an elaborate system.

In the concluding Chapter, the proposed benefits of this ITS-PM will be presented in the form of expert recommendations and outlook for PPPM and HCC.

References

1. Linington PF, Milosevic Z, Tanaka A, Vallecillo A (2012) Building enterprise systems with ODP: an introduction to open distributed processing. CRC Press, USA
2. Nickul D, Reitman L, Ward J, Wilber J (2007) Service Oriented Architecture (SOA) and Specialized Messaging Patterns. http://www.adobe.com/enterprise/pdfs/Services_Oriented_Architecture_from_Adobe.pdf. Accessed 6 Jan 2009
3. Hussein R, Engelmann U, Schroeter A, Meinzer H-P (2004) DICOM structured reporting: Part 1. Overview and characteristics. Radiographics 24:891–896
4. Radiological Society of North America: RadLex (2012) http://rsna.org/RadLex.aspx. Accessed 9 Dec 2012
5. Lemke HU, Berliner L (2010) Personalized medicine and model-guided therapy. In: Niederlag W, Lemke HU, Rienhoff O (eds) Personalisierte Medizin und Informationstechnologie, vol 15. Health Academy, Dresden pp 39–48
6. Lemke HU, Berliner L (2010) Personalized medicine and patient-specific modelling. In: Niederlag W, Lemke HU, Golubnitschaja O, Rienhoff O (eds) Personalisierte Medizin, vol 14. Health Academy, Dresden pp 155–164
7. Laskey KB (2008) MEBN: a language for first-order bayesian knowledge bases. Artif Intell 172(2–3):140–178