Relational Database of Treatment Planning System Information

ME Kantor[a*], G Starkschall[a], P Balter[a],
[a]MD Anderson Cancer Center, Houston, TX, 77030, USA

ABSTRACT

The purpose of the present work was to develop a relational database and associated applications to facilitate retrospective review of data present in radiation treatment plans. The data source was a commercial radiation treatment planning system (Pinnacle3, Philips Medical Systems, Milpitas CA), which is specifically characterized by an open data storage format and internal scripting capability. The database is an open-source, relational database (PostgreSQL, PostgreSQL Global Development Group, http://www.postgresql.org). The data is presented through a web interface in addition to being fully query-accessible using standard tools. A database schema was created to organize the large collection of parameters used to generate treatment plans as well as the parameters that characterized these plans. The system was implemented through a combination of the treatment planning system's internal scripting language and externally executed code. Data is exported in a way that is transparent to the user, through integration into an existing and routinely-used process. The system has been transparently incorporated into our radiation treatment planning workflow. The website-based database interface has allowed users with minimal training to extract information from the database.

1.   INTRODUCTION

Effective retrospective review of radiation treatments often requires gaining access to information related to a particular radiation treatment plan. Gaining access to such information typically includes the tasks of manually retrieving the treatment plan from storage, navigating its contents, and transcribing plan information into the analysis software. This process can be time-consuming and error-prone, especially when information needs to be extracted from a large number of treatment plans for clinical studies. Demand exists for removing these difficulties and streamlining the workflow by minimizing the necessary actions and user interaction. Operating directly on saved plan files is desirable, but often some of the information calculated during a treatment planning session is not preserved in the saved plan files and thus is not available for automated analysis. Such analysis would be useful for clinical studies and general research. In a large institution, thousands of treatment plans from many years of planning experience have been removed from treatment planning systems (TPS) and archived to low-tier storage. Accessing such data adds another level of difficulty to the process.

The entire procedure of identifying, locating, and downloading archived plan data is complex and repetitive, consuming much valuable personnel time. The large amount of data available in these thousands of treatment plans is a benefit to any retrospective study so an automated system to extract data and populate the database was desired.

The purpose of the present work is to describe a system that can store, organize, and present radiation treatment plan data. It was designed around a relational database, along with an abstraction layer to integrate it with a commercial radiation TPS (Pinnacle[3]: Philips Medical Systems, Milpitas, CA). With this system, a researcher possessing novice database and programming skills can quickly find and manipulate relevant information from a potentially abundant data source. When possible, the software components were created using cross-platform, non-proprietary software.

*mkantor@mdanderson.org; phone +1-713-792-0295; fax +1-713-563-2545; www.mdanderson.org

2.   METHOD

A. Radiation treatment planning system

The commercial radiation TPS used in this study facilitated this work because of two key features: (1) an accessible data format and (2) internal scripting capabilities.

The TPS operates on a Solaris (Sun Microsystems, Santa Clara CA) platform and stores plan data in a structured hierarchical format as a set of human-readable ASCII files. This arrangement greatly simplified the task of parsing and analyzing stored information. This format and its relation of data are easily understood by a physicist familiar with the planning process.

The TPS has an internal scripting interface with its own proprietary language and data structures. The data structures are organized in a multi-level object-oriented relationship, and the language largely consists of functions that can be applied to these objects and the object's properties. Through the use of this language, the TPS can be automated to perform any series of tasks within its feature set. The data structures contain data loaded from the stored files at the time the TPS is started, and temporary data that is only accessible while the TPS is open.

An example of this temporary data is the dose planned to specified regions of interest (ROI). This information is a primary target for retrieval and storage because it describes important plan evaluation data. The internal scripting language is able to access and output all of the sought-after statistics and underlying data, making them available for use outside of the TPS. When populating this database from stored plan files, these data must be recalculated.

B. Database

To accommodate unanticipated uses and applications of the data collected from the TPS, how it could be accessed, and how it could be manipulated, we searched for a database system that met the criteria of: compatibility with various operating systems, flexible access options, and ample documentation of features and design.

The database system chosen was PostgreSQL (Global Development Group, http://www.postgresql.org). This software fully satisfied the requirements named above and offered other appealing features such as native support for *n*-dimensional arrays as table column types, making it well-suited for storing any associated pixel or tabular data directly in a table cell.

*mkantor@mdanderson.org; phone +1-713-792-0295; fax +1-713-563-2545 ; www.mdanderson.org
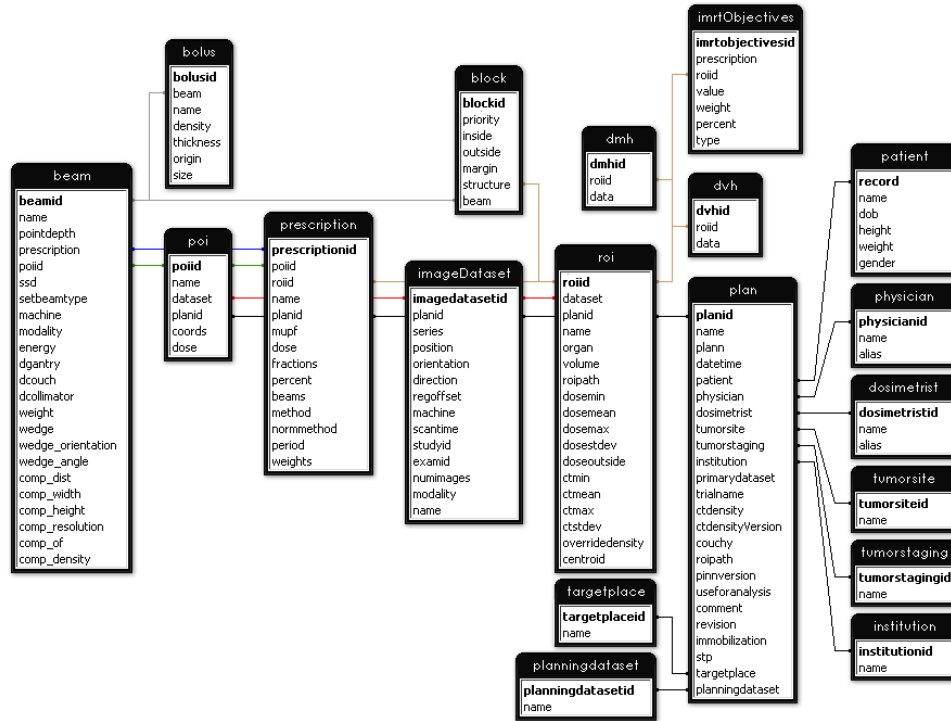
Figure 1. Complete schema of the database. This structure shows the child-parent relationships of the tables, where each table represents a type of object in the plan.

Each treatment plan contains a large amount of information, but the database is not intended to represent the plan entirely. A subset of important information was identified to define the characteristics of each plan. This selection was used to define the database's organizational structure, or schema (Fig. 1). The relation of objects in the TPS is soundly designed and has stayed consistent over many major releases. We chose to follow its general structure in the design of our tables; for example, there are tables called "plan," "prescriptions," and "beams;" each object in "beams" references the object in "prescriptions" to which it belongs, which in turn references an object in "plan".[1] Database tables can also reference real-world objects such as patients and physicians. The table "plan" holds descriptive information from the TPS, and acts as an anchor that is referenced by most other tables. This design makes database searches efficient because most of the fields desirable in a search are placed into the "plan" table, reducing the need to join many tables.

C. Software

Not all needed functionality was available in the TPS scripting language, such as the ability to parse ASCII data and interact with a networked database server. These functions were facilitated by the creation of several additional programs. These included (1) a file parser capable of converting structured TPS files into hierarchical objects, (2) an abstraction function to take these raw objects as input and convert them to processed objects more suited for database input, and (3) a general function that accepts processed objects as input and queries the database. The Python programming language (Python Software Foundation, http://www.python.org/psf) was chosen to implement this functionality due to its ease of use, portability, highly object-oriented design, and our development expertise with it. Custom input parsers and abstraction functions can be made for other treatment planning data sources, allowing the general query functions to stay isolated from frequent modifications.

A process was created to facilitate automated import of archived treatment plans. The input parser was re-used, since the files are from the same TPS for which it was originally designed. The abstraction function was created as a MATLAB

---

[1] It is important to differentiate between the object identified as "plan" in this database and the DICOM-RT object RT-PLAN.

*mkantor@mdanderson.org; phone +1-713-792-0295; fax +1-713-563-2545 ; www.mdanderson.org

program (The MathWorks, Inc., Natick, MA) that output data to the general query function. The MATLAB program calculates dose-volume histograms (DVHs) and various statistics by transforming the dose grid into the coordinate system of the ROI outlines, and uses these outlines to isolate the relevant voxels. Great care was taken to fine-tune the algorithms so that they precisely agreed with the TPS results for all calculated data; specialized linear-interpolation algorithms and single-precision floating point versions of built-in MATLAB functions are a few examples of what was necessary to reach numeric agreement with quantities calculated using the TPS.

D. Export of data during a planning session

- Current Trial
  - Name
  - Revision numbers
- Regions of Interest (ROI)
  - ROI name
  - Volume (cm^3)
  - Percentage of volume outside the dose grid
  - Reference to CT dataset on which ROI has been delineated
  - CT numbers within ROI: maximum, minimum, mean, standard deviation
  - Dose values (in cGy) within ROI: maximum, minimum, mean, standard deviation
  - Full set of vertices for every polygon defining the ROI on each axial plane
- Points of Interest (POI)
  - POI name
  - Coordinates of the POI
  - Dose values (in cGy) at the POI
- Dose-Volume Histograms (DVH)
  - Identification of parent ROI
  - Volume of each dose bin, using a 10-cGy bin width

Table 1. Information exported by the treatment planning system internal scripts, the majority of which is only available in system memory during a treatment planning session

*mkantor@mdanderson.org; phone +1-713-792-0295; fax +1-713-563-2545 ; www.mdanderson.org

Figure 2. Graphical user interface (GUI) presented to the treatment planner allowing additional plan-related details to be added during export. If an option is not present in this GUI, it may be added immediately using the control in the far-right column.

At the conclusion of the treatment planning process, our planners execute a series of TPS scripts that assist with the documentation tasks required for each treatment plan. The data export process was integrated into these scripts, and performs its tasks transparently to the planner (Table 1). If desired, the planner can use the GUI of the documentation software to turn off the export feature, or add additional information which is relevant to the plan but not contained in the TPS (Fig. 2).

E. Web-based interface

We developed a website interface for this database to facilitate data extraction and queries for retrospective reviews. This website enabled us to provide access to the database in a user-friendly format from any location within our intranet. Using a website as the primary database interface allows the integration of convenient server-side technologies such as on-the-fly vector and bitmap graphic generation, file format conversion, and user authentication. With the availability of resources and previous development experience as a decision-guide, the PHP: Hypertext Processor language (PHP) (The PHP Group, http://www.php.org) was chosen for implementation of the server-side code, and the Apache web server (The Apache Software Foundation, http://www.apache.org) was chosen to host the website. The database software offers connection options to a large variety of server-side languages and server software, so these choices had no effect on the core purpose of the work, and are merely one of many ways to access and present the data within.

*mkantor@mdanderson.org; phone +1-713-792-0295; fax +1-713-563-2545 ; www.mdanderson.org

Figure 3. A table showing the most recent plans exported into the database, with options for filtering and searching all exported plans. The displayed information is sufficient for a basic search and overview, with each row showing a specific patient's most recently exported plan. In the data structure of the commercial treatment planning system, each patient may have multiple plans and each plan may contain multiple trials, so specifics are shown for a particular export. The physician and dosimetrist are set uniquely per plan.

Figure 4. The header of a webpage document containing detailed information on exported plans. The "Versions" section allows the user to toggle between different plans for the selected patient, and to set a specific version as active, indicating that specific version as representative of the patient in broad data mining. Adobe Acrobat (pdf) files associated with this plan are listed as attachments accessible through hyperlinks.

Figure 5. Detailed information on all regions of interest from a selected plan. The website can present associated data in various formats using the provided form controls.

*mkantor@mdanderson.org; phone +1-713-792-0295; fax +1-713-563-2545 ; www.mdanderson.org

The website interface formed a foundation on which future applications were built and contains demonstrations of potential uses. A plan summary page shows a compact list of the newest plans and facilitates searches and filtering by major categories (Fig. 3). The plan viewer shows the information available for a specific plan, and contains controls that can be used to track and compare multiple versions of the same plan in cases of repeated exports (Fig. 4). The viewer can display DVHs for user selectable ROIs (Figs. 5 & 6). The DVHs can also be exported to other formats (currently Excel/XML, CSV, and plain text) for further analysis.

3.   RESULTS

This system has been in clinical use for the past two years. During this time we have captured the plan data from nearly 5,800 plans exported by our dosimetrists, and over 700 plans imported from archived files. These plans spanned 4 major versions of our TPS (Pinnacle^3 versions 6 through 9). Data has been retrieved to support several retrospective reviews; full and tabular DVH data has primarily been sought by researchers who have used this system.

The process of exporting plan data from the TPS to the database is transparent, and never omitted since it is integrated into the treatment plan documentation process.
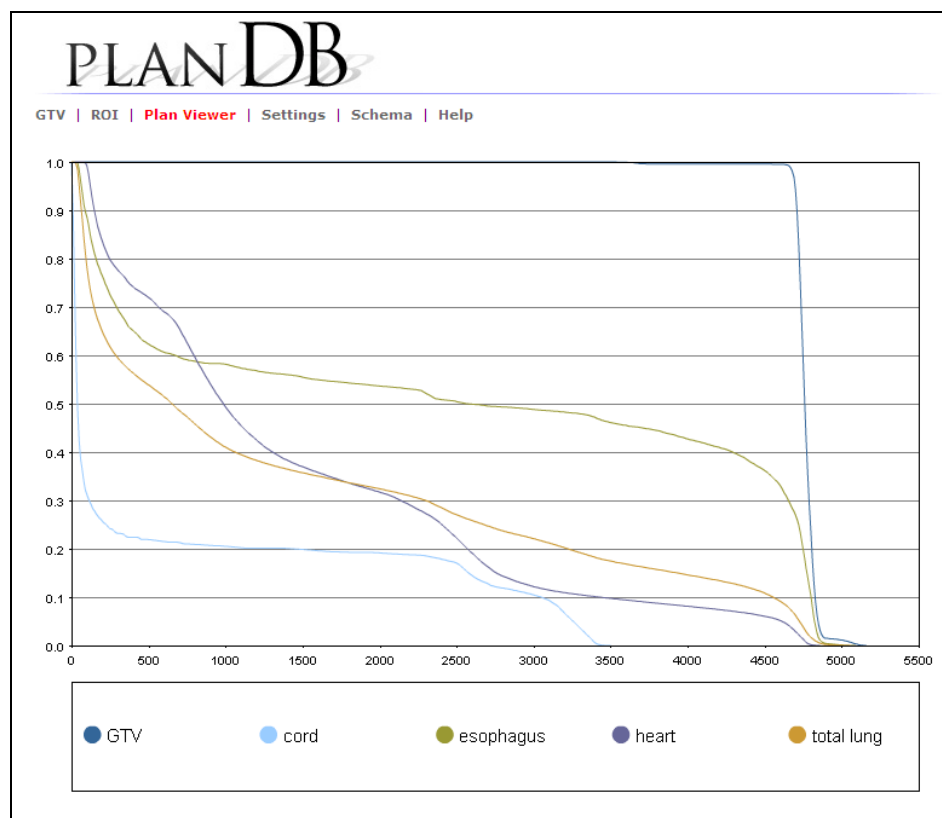

Figure 6. The website server uses the database to provide data points, colors, and names, which the client-based Adobe Flash Player uses to render this vector graphic.

*mkantor@mdanderson.org; phone +1-713-792-0295; fax +1-713-563-2545 ; www.mdanderson.org

planDB ID	PTV D_{95}	PTV D_{2cc}	Total Lung Mean	Esophagus $D_{1/3}$	Cord D_{2cc}
2055	74.855	76.903	1078.9	21.629	45.902
1739	74.755	76.433	1638.1	9.471	38.023
2069	66.348	70.689	1992.4	64.306	46.088
2053	73.479	79.361	1908.1	43.44	38.17
1983	72.06	79.487	1989.4	62.787	44.931
2000	42.022	53.062	1814.8	45.966	35.622
2052	73.782	79.696	2144.3	24.326	38.147
2085	61.098	66.81	1995.1	57.187	48.123
1999	73.518	79.975	1752.5	73.597	46.928
2054	58.831	64.465	2087.5	48.582	42.467
2168	39.303	44.099	1613.5	40.727	42.967
2056	74.515	79.011	1814.4	56.362	42.679
1967	68.196	74.268	2010.1	18.422	44.165
1961	74.477	78.281	1715.2	75.137	41.847
2043	48.698	56.846	2020.1	51.392	47.236

Table 2. An example of website output using the database source. Extracted from the database are all plans belonging to a specific institution. Requested statistics are automatically calculated, such as D_{95} (dose delivered to at least 95% of the ROI), D_{2cc} (minimum dose to the hottest 2 cm^3 of the ROI), mean dose, $D_{1/3}$ (dose delivered to at least 33.3% of the ROI). The column "planDB ID" is an internal reference to the database record and allows the website to provide contextual hyperlinks to the plan record directly from the table.

The importation from the archived files is highly automated and can be completed in batch mode, with a processing time of 6-10 minutes per treatment plan. In all aspects, the system's functionality and workflow met the design goals.

A unique feature of this system, compared to previous retrospective review tools, is its ability to update population statistics in real time. For example, a query was written to display specific dose levels to several structures for all patients associated with a specific "institution" as defined by the TPS (Table 2). This task is accomplished quickly using the website interface and the database. This particular example allows us to track the progress of a project over time and more importantly, it demonstrates the abilities of the system for real time monitoring of trends in single or multiple variables.

Another immediate application for this database is in the support of clinical trials. These trials often require reporting of various points along the DVH curve for various critical structures, e.g., V20 of the lung (the fraction of volume receiving 20 Gy or more). Web pages have been written for various protocols to display the relevant parameters, allowing data managers to quickly fill out the protocol datasheets.

4. DISCUSSION

Treatment plans contain a large amount of information in a digital format and thus available for placement into a database. Our database schema was designed around dosimetric reviews. In the future, the database can be expanded through creation of new tables and fields. The structure of this database allows this expansion without impacting existing applications since all fields and tables are explicitly referenced. The database software provides interface methods for dozens of programming languages and frameworks. Cooperation with other programmers and integration with their projects is simplified by this versatility.

*mkantor@mdanderson.org; phone +1-713-792-0295; fax +1-713-563-2545 ; www.mdanderson.org

Convenient options for placing data into the system are available from both inside and outside a treatment planning session. The website-based database interface makes this data available for speedy filtering and analysis.

Within this system, once a function or algorithm has been designed to retrieve and analyze a set of data based on its own custom criteria, it will continue to serve its purpose without the need for maintenance or adjustments. The output can be presented in almost any file format directly from a website interface. From this general description, several potential uses are immediately apparent:

1. Tracking of changes in structures for studies involving re-planning on weekly image sets.
2. Examining target dose coverage in the context of other factors, on a scale that would be a hindrance without the use of a database resource.
3. Analyzing statistical patterns by simultaneously monitoring plan parameters, such as beam angles, energies, etc., and the resulting tumor coverage or normal tissue sparing using a large set of input data.

Many other prior projects involved the analysis of TPS data, but none have attempted to retrieve and store a broad set of these data in a general way. Some have focused on specific treatment sites, and others have been implemented using proprietary data structures, which limit their versatility. For example, Carolan et al.[1] conducted a large study by first isolating appropriate patients from a radiation planning database system, gathering related variables from other data sources, and applying a multivariate regression analysis to the resulting collection of data. In such a scenario it would be cumbersome to add more variables after the initial survey. In another example, Ehrgott and Winz[2] developed an interactive support system to aid in multicriteria optimization for radiation treatment planning by searching through a database of treatment plans. Similar applications in multiobjective treatment planning were also used by Craft et al.[3] and Thieke et al.[4] Retrospective analysis of care-type studies[5] have also made use of treatment planning databases. What makes the present system unique is that data are placed into a general-purpose database, where simple query adjustments can ease and accelerate the data collection for such a study.

The majority of software used in this work was developed on open source systems (Python, PostgreSQL, PHP, and Apache). Open-source has many advantages over proprietary systems including a large community of developers and users who quickly identify and fix errors, as well as contribute work toward the enhancement of feature sets. Also it does not require the purchase of any software licenses, making it highly portable in an academic environment. Finally, the abundant documentation eased development and familiarization with previously unencountered features. Full source code for that software is publicly available (for example, PostgreSQL source: http://www.postgresql.org/ftp/source/), easing any unforeseen compatibility issues if a major modification is necessary.

Two exceptions to the use of open-source criteria were the use of MATLAB and the commercial TPS. The choice of TPS is irrelevant to the database itself, since the concepts of regions, points, demographics, and DVH distributions are general to any type of radiation treatment planning, abstraction layers can be built around other systems as needed. We have initiated development of a DICOM-based abstraction layer that would facilitate the inclusion of more varied data sources. The choice of MATLAB was driven by its high level of functionality and its ubiquity in academic environments. Although several open-source options have emerged as alternatives to MATLAB, a major commitment would have been necessary to translate the mature MATLAB code and many associated functions to these platforms.

5. CONCLUSION

This system has been in use for 3 years and has helped to collect data for retrospect reviews and is an integral part of clinical trials on a program project grant. This system has democratized access to the treatment planning data allowing any researcher to quickly retrieve detailed planning data and import it into many analysis platforms. The choice of auto-population as part of the printing process, web-access, and the ability to build project specific reports have all shown to be valuable features. This is an example how the intrinsic digital nature of TPS data can be leveraged to facilitate ongoing research in radiation therapy.

*mkantor@mdanderson.org; phone +1-713-792-0295; fax +1-713-563-2545 ; www.mdanderson.org

REFERENCES

1. Carolan H, Sun AY, Bezjak A, Yi Q-L, Payne D, Kane G, Waldron J, Leighl N, Feld R, Burkes R, Keshavjee S, Shepherd F. Does the incidence and outcome of brain metastases in locally advanced non-small-cell lung cancer justify prophylactic cranial irradiation or early detection? Lung Cancer. 2005;49:109-115.
2. Ehrgott M, Winz I. Interactive decision support in radiation therapy treatment planning. OR Spectrum 2008; 30: 311-329, 2008.
3. Craft D, Halabi T, Shih HA, Bortfeld T. An approach for practical multiobjective IMRT treatment planning. Int J Radiation Oncol Biol Phys 2007; 69:1600-1607, 2007.
4. Thieke C, Küfer K-H, Monz M, Scherrer A, Alonso F, Oelfke U, Huber PE, Debus J, Bortfeld T. A new concept for interactive radiotherapy planning with multicriteria optimization: First clinical evaluation. Radiotherapy and Oncology 2007; 85:292-289, 2007.
5. Vordermark D, Marold D, Wirth S, Keilholz L, Pfändner K, Schimpke T, Thiel H-J, Willner J, Ziegler K, Guckenberger M, Flentje M. Patterns of care in the radiotherapy of prostate cancer in Northern Bavaria 1998-2000. Strahlentherapie und Onkologie 2007; 183:314-320, 2007.

*mkantor@mdanderson.org; phone +1-713-792-0295; fax +1-713-563-2545 ; www.mdanderson.org