

---

一般口演

## 一般口演15

### データベース・臨床研究

2017年11月21日(火) 16:30 ~ 18:00 E会場 (10F 会議室1003)

---

## [2-E-3-OP15-2] 研究者のためのインハウス臨床研究データベース GeneralDataBaseの開発

渡辺 浩（独立研究開発法人 国立長寿医療研究センター）

（開発の経緯と目的）当施設で進行中の研究基盤構築プロジェクトで今回「研究者のためのインハウス臨床研究データベース」ともいべき GDBを開発し導入した。一般に研究者や部門独自のデータ集積は個別のデータベースシステム（Excelやファイルメーカーなど）で保管されていることが多い。これらの中には実際には研究者同士で共有化した方が効率的なことも多いデータ項目がある事は明白である。（問題点）しかしながらこれらの共有を妨げているのは「一部の項目は共有可能だがデータをすべて渡すわけにはいかない」「手続きが面倒」「患者実名が入ってしまっており、渡しにくい」などの事情と思われた。（方法）そこで病院系のネットワーク内にデータインポートに特化したシンプルなデータ構造のDBを構築した。（GDB:GeneralDataBase）データ構造を細分化した結果、患者IDと測定項目名と測定結果と日時程度の情報だけを持つレコードを、リレーションのほぼ無いテーブルに格納することで充分であることがわかった。これにExcelやファイルメーカーなどの外部データベースからのインポートが簡便に出来るようなインターフェースを構築した。部門データベースからGDBへのデータコピーは部門側から起動させる仕様とし、データを吸い取るような仕様にはしなかった。これにより研究者側の裁量権を尊重する形となった。さらにこの病院系のGDBは日次処理で患者情報が匿名化され、研究所内の情報系キャンパスネットワーク内に複製される仕組みを作った。（考察）現在データとしては高齢者総合評価指数CGA等データが既に数多く格納されている。当施設のバイオバンクの検体に付随させる修飾情報としてまずは使う予定である。今後は当施設における重要な研究用データベースのプロトタイプとなると思われる。

# 研究者のためのインハウス臨床研究データベース

## - General Database の開発 -

渡辺浩\*1

\*1 国立研究開発法人 国立長寿医療研究センター

## In-house clinical database for researchers

Hiroshi Watanabe\*1

\*1 National Center for Geriatrics and Gerontology

For support of researchers' work, we have developed GDB General DataBase. The feature is simple data structure. (Patient ID, measurement item name, date and time, measurement result, unit) and versatile input interface. Also, by synchronizing this chunk with anonymization processing and putting it in the research network, GDB made easy to use not only for physicians but also for researchers. Even for users who do not have database creation know-how, we also introduced a Simple DataBase system that allows simple patient database creation. Although we have not reached this operation yet, if we continue to enrich the contents in the future, it seemed that the needs in the field were high.

Keywords: Database, Clinical research, Data structure, Anonymous

### 1. 背景

これまで 長寿医療研究センターでは「研究者のための支援基盤構築」を医療情報室主催でいくつかのを行なってきた。先の沖縄での医療情報学連合大会では「モジュール単位開発のメリットを活かした研究者支援基盤システム構築の報告」を発表した。これは個人情報が入った SS-MIX 標準化ストレージ<sup>1)</sup>を匿名化モジュールで匿名処理を行った「匿名化ストレージ(仮称)」を活用したもので、これにより研究者はここに電子カルテのデータを匿名化処理する必要がなくなり、施設全体のセキュリティリスクも軽減するというものであった。これに連動した今回 GDB プロジェクトを進めたので報告する。GDBとは general database であり、研究者のための汎用的な要素を持つデータベースシステムである。そもそもの経緯としては、当施設内の医師を含めた研究者らは「もともと症例データベースシステムなどは持っていないが、イチから作るのは面倒である」「セキュリティをしっかりとところで管理しないといけない」などで医療情報室へ支援を求める声が多かった。また既存でデータベースを持っていた場合も「電子カルテ内に蓄積しているが患者情報が入っているなどで研究利用がしにくい」「データを他研究者へ共有させたいが出ている項目と渡せないデータ項目が混在しており共有させにくい」などの「部門で孤立したデータベースソースが散在」していることもわかった。さらにこの頃、ナショナルセンターバイオバンク事業<sup>2)</sup>内で施設内研究者のための統合データベース構築、インハウスデータベースなるものの構築も求められていた。

### 2. 目的

求められるシステムの仕様を検討した。「個人情報のあるデータはセキュリティが保証されること」「様々な環境のデータベースソースを取り込むことができること」「データベース作成経験のないユーザでもシンプルな症例データベースが作成できること」「研究者が活用するデータは匿名化処理がされること」とした

### 3. 方法 システムの概要

基本的には病院ネットワーク内にファイルメーカー基盤のデ

ータベースを構築することから始めた。

#### 構造

様々なデータ構造からのデータを取り込む必要があることから、できる限りデータ構造をシンプルな必要項目のみとしてみた。想定上、データベースの項目を細分化していくと「患者(被検者)識別 ID」「測定日時」「測定項目名(コード)」「測定結果」「単位」に落とし込められると考えた。さらには、それぞれのデータ項目のオリジナルヘリンクできるように「元のデータベース識別子」もつけることにした。またこのデータベース自体のテーブルにはあまり複雑なリレーションを持たせないようにした。

#### インポート I/F

外部にあるエクセルなどは 上記の形式に精製して CSV の形で取り込みを可能にした。本システムが ファイルメーカーを基盤に作っているので、外部のファイルメーカーデータベースであればこの GDB についているスクリプト(内包プログラム)のトリガーを起動させることもできる。そのため外部のデータベースが通常運用している中で、定期的に自動的にデータの転送なども行うことができるようにした。この際には部門データベースから GDB へのデータコピー転送は部門側から起動させる仕様とし、GDB がデータを吸い取るような仕様にはしなかった。これにより研究者側の裁量権を尊重する形となった。これらのシンプルなデータ構造と取り込みのインターフェースを公開することで外部からのデータの受け入れの可能性を広げた。

#### 保存・匿名化処理

以前の情報学会でも発表した、当施設内にある「匿名化モジュール」を今回も活用した。元々バイオバンクシステム用に導入したものだが、すなわち電子カルテネットワークと研究用のキャンパスネットワークの間にある DMZ ネットワーク内に患者の ID を匿名変換する機能に特化したサーバーをおき、これによってバイオバンク登録患者の匿名変換や標準化ストレージの匿名変換をさせたものである。今回の GDB も病院ネットワーク内には医療目的で使えるように実名を含めて管理している。これをバッチ処理で ID を匿名変換して同様の構造のものを匿名化 GDB として複製して置いた。

#### アウトプット

今回のデータベースにはデータの検索や多機能な出力仕様は作らなかった。データを貯めやすくしておき、研究活用の際には大まかに切り取ったデータを部門、ローカルなデータベースで加工編集してもらうことを目的としている

### 初心者用シンプルデータベース

GDB とともに作成経験のないユーザ向けの Simple Database SDB の開発にも取り組んだ。標準化ストレージからデータの取り込み機能を持ち、ある程度の患者属性情報は自動引用されるので ID を入れるだけで氏名や生年月日が引用されるようになっている。日々の診療の合間に独自入力項目を含めた情報を更新できるようになっている。また GDB への送り込み機能が DB にプリセットされている。

## 4. 結果

このシステムはまだプロトタイプなので実運用まで至っていないのだが、バイオバンクの情報部門でモデルケースとしてテスト運用中である。現在のデータ蓄積状況としては 13000 件の MMSE データ(ミニメンタルステート検査:認知症の診断に使われる質問形式セットのデータ)が貯められている。さらに VSRAD データ(MRI による脳の萎縮度を数値化したデータ)の写し込みも検討中の段階である。

## 5. 考察

元々は施設内で置き場の困ったデータや部門個別要望に対応するために検討していたものであったが、進むにつれて研究者のためのインハウスデータベースのプロトタイプと成り得る可能性のあるソリューションであることがわかった。現時点で現場を感じるメリットとはデータにアクセスしやすく使いやすいというものであった。

しかしながら同時に研究者からはこの GDB に入れられたデータの信用性はどこが担保するのか、と言った意見も聞かれた。上記で伝えたように本ソリューションは限定的な目的で当初は作成されていたため、ユーザが想定でき、ある程度善意の運用イメージがつけやすかったが、システム規模が拡張するに従い重要視すべき問題となった。このデータベースはあくまでも研究の糸口を見出すために利用するものと位置付けている。そのため前向き研究利用やバリデーションをしつかりと求められる研究はターゲットではない。しかしながら利用する立場を考慮しデータのトレーサビリティを担保するためリンク元のデータベース識別子を持たせるようにした。更には今後は根拠となる文献のリンクを乗せるなどの案も検討中である。このデータベースはコンテンツが増えればより魅力的になると思われる。まずは多くのユーザに周知使ってもらうことを検討している。

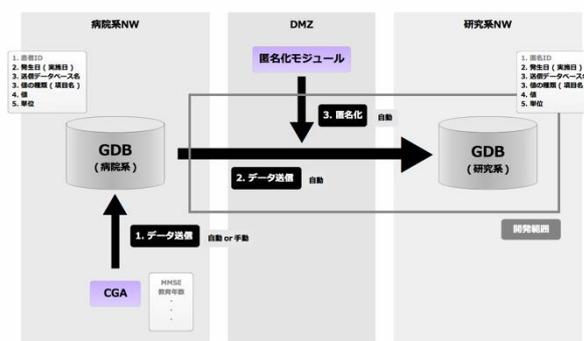


図1 システム概要



図2 シンプルなメニュー

0000004	2017/04/05 (水)	CGA	透視版	13
0000005	2017/04/06 (木)	CGA	透視版	14
0000006	2017/04/07 (金)	CGA	透視版	15
0000007	2017/04/08 (土)	CGA	透視版	16
0000008	2017/04/09 (日)	CGA	透視版	17

図3 シンプルなデータ構成

## 6. 参考文献

- 1) Kimura, M., Nakayasu, K., Ohshima, Y., Fujita, N., Nakashima, N., Jozaki, H., ... Tani, S. (2011). SS-MIX: a ministry project to promote standardized healthcare information exchange. *Methods of Information in Medicine*, 50(2), 131-9.
- 2) 加藤規弘, 鶏内雅司, 河野隆志, 後藤雄一, 斎藤美紀子, 穴戸稔聡, ... 松原洋一. (2015). わが国の保健医療システムにおけるゲノム医療の展望 国立高度専門医療研究センターの取り組みと提言(解説). *医療*, 69(8-9), 375-382.