

データベース管理の実際と問題点

山口県立大学看護学部
中澤 港 <minato@ypu.jp>

データベースの必要性

- 大規模なデータなら使い捨ては非現実的
- 作図ソフトや統計処理ソフトを使いわけるために、それぞれに必要な形式のデータを出力できるような管理方式が必要
- 公開可能なデータベースが蓄積されていけば、メタアナリシスに役立つ可能性もある

データ管理方式

- スタンドアロンのパソコンで, 1つのソフトを使って
入力・編集・保存
 - 利点: 手軽
 - 欠点: 分業できない。ソフト依存性あり
- スタンドアロンのパソコンで, 入力用のフロントエンドとデータ管理のバックエンドを両方作動させて入力・編集・保存
 - 利点: データが壊れにくい。入力しやすい
 - 欠点: フロントエンドの設計が面倒。分業しにくい
- バックエンドをサーバで, フロントエンドをクライアント(パソコン)で動作させる
 - 利点: 分業・協業しやすい。入力しやすい
 - 欠点: ネットワーク環境でないと作業できない

データファイルのいろいろ

- テキストデータ
 - 利点: 可搬性が高い。クラッシュしても復旧しやすい
 - 欠点: 見通しはよくない。複雑な構造はもたせにくい。検索に時間がかかる(ただしインデックスをつけるソフトを使えば高速化できる)
- 表計算ソフトのデータ
 - 利点: 小さいデータの場合は見通しがいい。
 - 欠点: 複雑な構造はもたせにくい。可搬性が低い。検索に時間がかかる。クラッシュに弱い
- データベースソフトのデータ
 - 利点: 検索が速い。複雑な構造ももたせられる。いいソフトならクラッシュに強い
 - 欠点: DBMS 作動下でしか中身が見えない

現状の問題点

- かなり大きなデータでも表計算ソフトで入力されている場合が多い(とくに保健・看護系)
- データの質のチェックが不十分な場合が多い(誤入力はいうまでもないが、本来は再コーディングが必要な場合でもされないままの場合がある)
- データベースの構造が予め十分に練られて設計されていないため、あとで保守するのが難しいことがある。
 - コンピュータ専門家でないとはDBMS の設計が難しい
 - コンピュータ専門家は、必ずしも疫学や保健・医療の専門家ではない

将来の方向性

- 必要ならばいつでもデータの質のチェックに立ち返れるようにするため、個票と照合できる状態で、DBMS で管理しておくべき。ハードウェア的なバックアップも定期的に必要
- 同時に、定期的に、個人情報と連結不能な状態にしたテキストデータ(詳細にデータ構造などをドキュメントとして記述しておくことも必要)を出力しておくべき
 - 他の研究者が再分析やメタアナリシスに利用できる
 - 万が一クラッシュしてもテキストからデータを再構成できる利点