




DandD プロジェクト

慶應大・理工 柴田 里程



企画セッション 「データとその属性情報 記述」

- ◆ 数理統計学
- ◆ 応用統計（理論を応用する？）
- ◆ データそのものに対する一般論
 - データの的確な把握
 - データの組織化
 - データの属性記述

データに対する視点の投影



- ◆ SPMS (丹後俊郎, 刈谷丈治, 1980—1981)
 - Basic Data と Variable Data
 - 個体 × 時間 × 変量
 - 医学研究データ

- ◆ NetCDF (Unidata)
 - 緯度 × 経度 × 時間 × 変量
 - 解析に必要な属性を付与
 - 気象データの国際的な交換

- ◆ DDI
 - 個体 × 変量
 - 解析に必要な属性を付与
 - 社会科学データの検索



◆ DandD (Data and Description)

- さまざまな組織化の可能性
- 複数の絡み合ったデータセット
- 特定の研究分野を想定しない汎用性
- 必要な属性, 背景情報の一般的な規定
- ネットワーク上に散在する様々なデータ
- 特定のソフトウェアに依存しない汎用性
- データに対する視点の反映
- データの変容
 - クリーニング
 - 抽出
 - 解析結果
 - モデル
 - ユーティライゼーション



DandD プロジェクト

- ◆ D&D (1987)
 - 「データ解析のための電子ジャーナル」
- ◆ DandD (1998-2000)
 - XML
 - 「データリテラシー」, 2001, 共立出版
- ◆ DandDII (2001-2003)
 - インターデータベース
- ◆ DandDIII (2003-)
 - データの変容を DandD インスタンスのネットワークで記述



DandD プロジェクトの2つの側面

◆ 理論

- 視点を反映した構造化
- 属性の洗い出しとその特徴付け
- クリーニングと正規化
- データモデル

◆ 開発

- サーバ・クライアントシステム
- データベースなどへのアクセス
- 解析ソフトウェアとの連携



DandD プロジェクトの要点

- ◆ データ本体と、その構造と属性記述の分離
 - データベース, データファイル, . . .
 - DandD インスタンス (XML 文書)
- ◆ ソフトウェア, プラットフォームからの独立性
- ◆ クライアント・サーバシステム
- ◆ ネットワークの活用
- ◆ 多国語対応
- ◆ データの変容をデータ本体に手を加えることなく
DandD インスタンスで記述



一例 (途中で形式の変わったFTPサイト)

- ◆ 年月日 → 年、月、日
- ◆ 固定形式, 浮動形式
- ◆ DandD インスタンスでの記述
 - DatabaseCode, Code
 - 水準の表現の変換



- ◆ データ解析のためのデータとその記述の一体化



- ◆ データの多様性を吸収し、その複雑な関連性を、データ本体に手を加えることなく表現する手段(エージェント)

DandD の普及

◆ 研究

- データの多様性と、視点の投影
- その蓄積と体系化

◆ 応用

- 共同作業の能率化
- データの洪水に対する戦い
- 適切なモデルの構築とユーティライゼーション

<http://www.stat.math.keio.ac.jp/DandDII>

<http://www.stat.math.keio.ac.jp/DandDIII>

