

# 安全データ移行

データ移行は、既存システムを廃止し新しいシステムを稼働させる際に、単純にデータをコピー／移動する作業だと軽視されがちです。しかし、実際のデータ移行は非常に複雑であり、業務プロセスに対する深い理解と技術的な実装経験が必要とします。

データ移行は、医薬品安全性システムの新規導入やアップグレードプログラムにおいて、最も重要かつ困難な工程の一つです。将来の症例処理を見据え、レガシーデータを円滑に移行するためには、厳格な規制遵守とデータに対する高い慎重さが求められます。

製薬企業およびCRO（医薬品開発業務受託機関）は、安全性システム導入ライフサイクルの中で、最も時間と重要性を要する活動であるデータ移行に、数百万ドルを投じています。業界の経験では、データ移行プロジェクトの60%以上が、スケジュール超過または予算超過に陥っているとされています。本書では、こうした超過が発生する要因を明らかにし、品質コスト

（Cost of Quality : CoQ）を削減する方法を検討します。

## なぜデータ移行が必要か

- 業務プロセスの改善
- 新システムの導入
- 規制要件の変更
- 市場統合や製品買収

## 仕組み

データ移行は、まずソースシステムの分析および詳細なデータプロファイリングから開始され、プロジェクト初期段階でデータ品質の問題を洗い出します。次に、機能要件を反映したソース-ターゲット間のマッピングを作成します。ETLツールまたはカスタムプログラムを用いて、ソースデータを抽出し、機能要件に基づいたデータ変換を行います。この詳細な分析により、業務ユーザーはデータ例外を特定し、本番ロード前にデータ修正を実施することが可能になります。最後に、移行後データの妥当性を確認するためのデータ検証が実施されます。

## 課題

### プロジェクトスコーピング

一般的に、データは十分な注意を払われず、データ移行のスコーピングは後回しにされがちです。その結果、プロジェクト後半でデータ品質問題が顕在化し、スケジュールや予算に深刻な影響を与えます。

### データプロファイリング

本来はプロジェクト初期に発見されるべきデータ品質問題が、後工程で判明するケースが多くあります。過去プロジェクトのプロファイリングクエリを流用する傾向があり、誤解を招く結果となる場合があります。

### 異なる安全性システムに関する知識不足

安全性データ移行には、複数の安全性システムに関する技術的理解と業務ドメイン知識が必要です。これらのスキルは専門性が高く、希少です。

### データ検証

レガシーデータのサンプリング検証は、品質チームにとって大きな課題です。適切なサンプルは、レガシーシステム内のすべてのデータシナリオを代表する必要があります。

### 方法論

体系的な方法論が欠如していると、業務部門とプロジェクトチームに大きな負担がかかります。業務チームと開発チームがサイロ化し、重要な意思決定の共有が困難になります。

### 症例数（データ量）

レガシーシステムには膨大なデータが蓄積されています。この増大し続けるデータ量は、性能ベンチマーク、IT計画、ハードウェアサイジングの面で課題となります。

## ソリューション

### 計画

データ移行プロジェクトは非常に複雑であるため、移行すべきデータを「必要かつ実現可能な範囲」に限定する必要があります。スコープを明確化した後、プロジェクトのスケジュールおよび予算を設定または再調整します。

## デザイン（設計）

直接マッピング、コードリストマッピング、カスタム業務ルールなど、すべてのデータ変換ルールを含むマッピング仕様書を作成します。要件、システムアーキテクチャ、データ量を考慮した設計を行います。

## 開発

マッピング仕様に基づき、ETL（抽出・変換・ロード）処理を開発し、各ルールを検証します。複数回のテスト実行により、早期にエラーを特定し、重要な判断を行うことができます。

## 検証

業務ユーザーは、帳票比較、症例処理、フォローアップ処理、自動テスト結果を通じてデータを検証します。

## 実行

ETLプログラムにより、ソースデータを抽出し、マッピング仕様に従って変換し、ターゲットシステムへロードします。

データ品質メトリクスは、計画・分析・設計・開発・検証の各フェーズで定期的に作成します。これらのメトリクスは、業務ユーザーにとっての可視性を高め、品質改善の進捗を示します。

## ツールおよびアプローチ

市場には多くの標準ETLツールが存在しますが、データ移行は一度きりの活動であるため、コスト面からカスタムプログラムが選択されることも多くあります。通常、ソースデータの規模や複雑性に応じて、以下の3つのアプローチが採用されます。

### テーブル単位

- テーブル単位でデータをロード
- 実行時間が速い
- 開発期間が短い
- 大規模移行（100万症例超）に適している
- 症例単位でのエラー追跡が困難
- 複雑な変換レイヤーが必要
- 一時的にターゲットシステム構造の変更が必要な場合あり

## プロファイル（分析）

データプロファイリングを実施し、データ品質問題を特定します。その上で、手作業またはプログラムによるデータクレンジングについて、適切な判断を行います。プロファイリングは、どのデータを移行すべきかを業務ユーザーが判断する上で有効であり、再作業やコード修正のコスト削減につながります。

## 症例単位

- 症例単位でデータをロード
- 症例レベルでのエラー追跡が可能
- 変換レイヤーが比較的単純
- ターゲットシステム構造への影響なし
- データ品質が低いソースに適している
- 中小規模移行（最大100万症例）に適している
- 実行時間が比較的遅い
- 開発期間が長い

## ハイブリッド（E2B+カスタムSQL）

- E2Bファイルを使用して移行
- E2Bによりシステム間のデータ仕様が標準化
- 規制関連データの大部分をカバー
- 約30%のデータ項目はE2Bでカバーされず、カスタムプログラムが必要

## まとめ

データ移行を成功させるためには、十分な注意と準備が不可欠です。適切な計画とデータプロファイリングが行われない場合、プロジェクトが予算超過・期間超過、あるいは失敗に終わるリスクが高まります。成功の鍵は以下の3点です。

- **把握**：移行前に想定外のシナリオを特定する
- **改善**：移行前または移行中にデータ品質問題を是正する
- **監視**：データ品質メトリクスを定期的に作成・共有する

Vitrana™が、お客様のニーズに適したデータ移行戦略の策定をどのように支援できるかについては、Eメール（[sales.jp@vitrana.com](mailto:sales.jp@vitrana.com)）または電話（080-7946-1343）までお問い合わせください。