

第4章：データ品質管理：移行成功の鍵

SAP データ移行プロジェクトにおける最大の落とし穴は、システムの機能や技術的な移行プロセスではなく、**データの品質**にあります。古いレガシーシステムから持ち込まれる不完全なデータや不整合なデータは、新しい SAP 環境の整合性を崩し、業務停止や分析の誤りを引き起こす根本原因となります。本章では、データ移行を成功に導くための核心的な要素であるデータ品質管理に焦点を当て、その戦略と具体的な実践手法を詳細に解説します。データ移行は、データ品質を一挙に改善し、企業のデータ資産価値を高める**絶好の機会**と捉えるべきです。

4.1 データプロファイリングと品質問題の特定

データ品質管理の第一歩は、「**現状を知ること**」です。旧システム内に存在するデータの真の状態を科学的に分析し、潜在的な問題を特定する作業が**データプロファイリング**です。これは、移行プロジェクトのリスクと工数を正確に見積もるための、最も重要な基礎作業となります。

4.1.1 データプロファイリングの定義と目的

データプロファイリングとは、移行対象となる旧システム内のデータを詳細に分析し、その**構造、内容、品質、および関係性**に関する統計情報やインサイトを取得するプロセスです。

主な目的：

1. **リスクと工数の見積もり**：データ品質の悪さ（欠損、重複、不整合など）の度合いを数値化し、それに基づいて**データクレンジング**や**データ検証**に必要な工数と期間を現実的に見積もる。
2. **変換ルールの明確化**：データの分布や値の多様性を把握することで、ソース（旧システム）からターゲット（SAP）へのマッピングに必要な**変換ロジック**や**標準化ルール**を具体的に定義する。
3. **データ駆動型の意思決定**：主観や推測ではなく、**客観的なデータ**に基づいて移行対象範囲の妥当性や、データ品質改善への投資の必要性を経営層に説明する根拠とする。

4.1.2 プロファイリングの具体的な手法

データプロファイリングは、専門のツール（SAP Data Services、サードパーティ製 DQ ツールなど）を活用して実施されます。主な分析項目は以下の通りです。

1. 構造分析（Structure Analysis）

- データ型の検証**：各項目のデータ型（文字列、数値、日付など）が正しく

定義されているか、格納されている値がその型と一致しているかを確認します。

- **事例：**品目コードの項目が数値型であるにもかかわらず、一部に「ABC-123」のようなハイフンを含む文字列が混入している場合、SAP へのロード時にエラーとなります。

- **必須項目の欠損率：**SAP で必須となる項目（例：販売組織、会社コード、税コードなど）について、旧データでの欠損率を算出します。欠損率が高いほど、後のクレンジング工数が大きくなります。
- **フォーマットの不統一：**電話番号や住所、日付などの表記ルールが統一されているかを調べます（例：「2024/04/01」「2024-04-01」「240401」の混在）。

2. 内容分析 (Content Analysis)

- **値の頻度分析：**特定の項目（例：取引先分類、支払い条件）について、値の種類（ディステINCTな値）とそれぞれの出現頻度を調べます。これにより、使用されていない古いコード値や、移行後の SAP コード体系へのマッピングが困難な独自のコードを特定します。
- **データの分布と範囲：**数値データ（例：在庫数量、単価）について、最大値、最小値、平均値、標準偏差などを確認し、異常値（**Outlier**）を検出します。異常値は、データ入力ミスや、過去の特別な取引を示すことが多く、クレンジングの対象となります。
- **ドメイン値の逸脱：**項目が取り得るべき値の範囲（ドメイン）から外れた値（例：性別の項目に「X」が入っている）を特定します。

3. 関連性分析 (Relationship Analysis)

- **参照整合性のチェック：**親データ（例：顧客マスタ）が存在しないのに、子データ（例：受注伝票）が存在している、というデータの不整合をチェックします。SAP は厳格な参照整合性を要求するため、この問題が最も致命的なエラーの原因となります。
 - **事例：**存在しない仕入先コードを参照している購買発注データは、SAP へロードできません。
- **機能的依存性のチェック：**ある項目の値が、別の項目の値を一意に決定しているか（例：品目コードが決まると、必ず単位が決まる）を確認します。この依存関係が崩れている場合、データ設計そのものに問題があることを示します。
- **冗長性の検出 (重複分析)：**同一のエンティティ（例：顧客、仕入先、品目）が、異なる ID で複数登録されているケース（**重複データ**）を検出します。この作業には、**ファジーマッチング**（少しの表記揺れがあつて

も同一と見なす技術）が用いられます。

4.1.3 プロファイリング結果の活用

プロファイリングの結果は、以下の重要な成果物に反映されます。

- **データクレンジング計画**：どのデータ項目を、どの程度修正する必要があるか、修正に必要な工数とツールを決定します。
- **マッピング定義の洗練**：旧データの複雑な実態に合わせて、変換ロジックをより詳細かつ正確に定義します。
- **データオーナーシップの再定義**：品質問題が顕著なデータのオーナー（業務部門の責任者）を明確にし、クレンジング作業への関与を促します。

4.2 データクレンジングと標準化の手法

データプロファイリングによって品質問題が特定されたら、次に必要なのが、それらの問題を修正し、データを新しい SAP 環境の要件に合わせて整えるデータクレンジング（Data Cleansing）とデータ標準化（Data Standardization）です。この作業こそが、SAP 移行の品質を物理的に担保する核となります。

4.2.1 データクレンジングの原則とプロセス

データクレンジングは、単にデータを修正するだけでなく、**修正の履歴と根拠**を保持することが重要です。

クレンジングの原則：

- **データソースの変更は避ける**：原則として、旧システム内のオリジナルデータは変更せず、抽出したデータを一時作業領域（Staging Area）でコピーし、そこで修正作業を行います。これにより、万が一修正ミスがあってもオリジナルデータが保護されます。
- **ルールベースの自動化を優先**：手作業による修正は、工数がかかり、ヒューマンエラーのリスクも高いため、可能な限り**クレンジングルール**を定義し、ツールによる自動処理を優先します。
- **業務部門による最終承認**：データの内容に関わる重要な修正（例：重複データの統合、主要コード値の変更）は、必ず**データオーナー**である業務部門の承認を得る必要があります。

クレンジングの主要プロセス：

1. **重複排除（Deduplication）**：ファジーマッチングなどの手法を用いて重複データを特定し、ゴールデンレコード（Golden Record）を一つ選定して統合します。統合の際、重複レコードに含まれていた重要な情報は、ゴールデンレコードに

集約されます。

2. **欠損データの補完 (Imputation)**：必須項目が欠損しているデータについて、業務ルールに基づき、適切なデフォルト値の適用、関連データからの推定、または業務担当者による手作業での入力を実施します。
3. **データ変換と標準化**：
 - **コード変換**：旧システム固有のコード体系（例：支払条件コード「001」）を、新しい SAP のコード体系（例：「NET30」）にマッピングして変換します。
 - **フォーマット統一**：表記揺れ（例：住所の「株」と「株式会社」）を統一し、SAP のデータ型要件（例：日付形式）に合わせます。
 - **データエンリッチメント**：外部データ（例：最新の郵便番号データ、企業情報データベース）を参照して、旧データに不足していた情報を補完します。

4.2.2 クレンジングの具体的な手法とツール

1. ルールエンジンと自動処理

- **正規化ルール**：住所表記の全角・半角変換、不要な記号の削除、カタカナ表記の統一など、一律に適用できるルールを設定し、大量のデータを自動処理します。
- **辞書マッピング**：旧システムの曖昧な値（例：担当者名、短い製品説明など）を、新しい SAP の標準マスター値に変換するための**辞書**（マッピングテーブル）を作成し、ツールに読み込ませて自動適用します。

2. 業務部門との協働による修正（ワークフロー）

自動化で対応できない複雑な問題、特に**重複データの統合**や**論理的不整合**の修正は、業務知識を持つ担当者による判断が必要です。

- **例外レポートの出力**：移行ツールや DQ ツールは、定義されたルールで修正できないデータや、修正に業務判断が必要なデータ（**例外データ**）をレポートとして出力します。
- **ワークフローによる修正依頼**：この例外レポートに基づき、データオーナーである業務担当者に修正依頼を出し、修正後の値を移行チームにフィードバックする**修正ワークフロー**を確立します。このワークフローは、プロジェクトの進捗を左右するため、迅速かつ正確に行われる体制が必須です。

事例：総合商社 M 社のデータクレンジング

M 社は、SAP S/4HANA への移行において、仕入先マスタの重複が深刻な問題でした。プロファイリングの結果、約 20,000 件の仕入先マスタのうち、約 15%が重複して

いることが判明しました。M社は、サードパーティのDQツールを導入し、仕入先名、住所、電話番号のファジーマッチングを行い、重複候補を自動抽出しました。その後、購買部門の担当者を巻き込み、この候補リストを基に、手作業で統合すべきレコードとそうでないレコードを判断させました。この協働体制により、わずか2ヶ月で仕入先マスタの品質を劇的に向上させ、移行後の購買業務の効率化を実現しました。

4.3 マスターデータ管理 (MDM) の重要性

データ移行は一時的なイベントですが、データ品質の維持は永続的な課題です。移行後にデータ品質が再び低下するのを防ぐために不可欠なのが、マスターデータ管理 (MDM: Master Data Management) です。MDMは、SAP移行で得られた「クリーンなデータ」を維持・発展させるための企業のガバナンス体制そのものです。

4.3.1 MDM とは何か？なぜ SAP 移行で不可欠か？

MDMとは、企業全体で共有される最も重要なデータ (マスターデータ: 顧客、品目、仕入先、勘定科目など) を、**単一の、正確で、一貫性のある情報源**として管理するための**プロセス、ガバナンス、およびテクノロジー**を指します。

SAP 移行における MDM の必要性:

1. **品質の永続化:** クレンジングによって一時的にきれいになったデータも、運用フェーズでルールやプロセスがなければ、必ず時間とともに品質が低下します (データの「エントロピーの法則」)。MDMは、この品質低下を防ぐための防波堤となります。
2. **統合化の担保:** SAPは統合システムであるため、品目マスタの変更が販売、購買、生産、会計のすべてに影響します。MDMは、この**部門横断的なデータの整合性**を担保するための中心的な役割を果たします。
3. **S/4HANA への対応:** SAP S/4HANAでは、従来の取引先マスタ (顧客/仕入先) がビジネスパートナー (BP) という単一のエンティティに統合されています。このBPモデルは、以前よりも厳格なデータガバナンスと、MDMプロセスを要求します。

4.3.2 MDM を実現するための3つの要素

MDMを成功させるには、「人 (ガバナンス)」「プロセス」「テクノロジー」の三位一体の取り組みが必要です。

1. データガバナンス (人): 組織と役割の定義

- **データオーナー (Data Owner):** マスターデータの定義、品質基準、利用ルールに関する最終責任を負う業務部門の役職者。
- **データスチュワード (Data Steward):** 日々のデータ登録・更新・承認プロセス

を実行・監視し、データ品質を維持する実務担当者。部門横断的な連携を担います。

- **データガバナンス評議会**：部門間のデータ定義やルールのコンフリクトを解決し、全社的な MDM 戦略を決定する、経営層を含む組織体。

2. データプロセス：ライフサイクル管理の確立

- **データ作成・変更プロセス**：新しいマスターデータを登録・変更する際の、申請、承認、検証、システム登録という一連のワークフローを厳格に定義します。このプロセスにおいて、データスチュワードが必ず検証ステップを挟むようにします。
- **データ品質モニタリング**：稼働後も、定期的にデータ品質の KPI（例：欠損率、重複率）を測定し、問題が発見された場合には自動的にデータオーナーに通知・是正を促す仕組みを構築します。

3. データテクノロジー：ツールの活用

- **MDM ハブ**：マスターデータを一元管理し、SAP や周辺システムに配信するための中心的なシステム（MDM Hub）を導入します（例：SAP MDG - Master Data Governance）。
- **データ統合ツール**：MDM Hub と周辺システム間のデータ連携を自動化し、データの変更がリアルタイムで各システムに反映される仕組みを構築します。

事例：グローバル企業 N 社の MDM 戦略

N 社は、SAP グローバル統合に際し、各国でバラバラだった品目マスタの統合が課題でした。彼らは、単なる移行作業で終わらせず、SAP MDG を導入し、「マスターデータは MDG でしか作成・変更できない」というルールを徹底しました。これにより、各国拠点からの品目登録申請はすべて MDG に集約され、グローバルデータスチュワードチームが定義された標準ルールに基づいて検証・承認することで、移行後のデータ品質の低下を完全に防ぐことに成功しました。

4.4 不要データの削減とアーカイブ戦略

データ移行プロジェクトは、「どのデータをきれいにするか」だけでなく、「**どのデータを移行しないか**」を決定するプロセスでもあります。不要なデータを SAP に移行しないことは、プロジェクトの工数削減、システムパフォーマンスの維持、そして将来的な運用コストの低減に直結します。

4.4.1 不要データの定義と削減のメリット

不要データとは、以下のいずれかに該当するデータです。

1. **利用頻度の低いデータ**：過去 5 年以上利用されていない、参照頻度が極めて低

いトランザクションデータ。

2. **法的・会計的義務が終了したデータ**：法令上の保持義務期間を過ぎたデータ。
3. **技術的に不要なデータ**：旧システムのアドオン機能にのみ紐づいていたデータ項目や、新しい SAP の業務プロセスでは管理対象外となるデータ。

不要データ削減のメリット：

- **移行コストの削減**：データ抽出、変換、ロード、検証の工数を大幅に削減できます。
- **システムパフォーマンスの向上**：SAP S/4HANA 環境におけるデータベースサイズを小さく保ち、日常業務（トランザクション処理、レポート作成）の応答速度を向上させます。
- **運用コストの低減**：データベースのバックアップ時間、システムパッチ適用時間、インメモリデータベース（HANA）のメモリ使用量を削減し、全体的な TCO（総保有コスト）を低減します。

4.4.2 データアーカイブの戦略的アプローチ

不要なデータを単に「捨てる」のではなく、将来的な参照の可能性や法的義務を考慮し、**戦略的にアーカイブ**することが重要です。

1. 保持ポリシーの定義

- **法的要件の確認**：国や地域ごとの会計法、税法、個人情報保護法などにに基づき、各データカテゴリ（例：会計伝票、購買発注、顧客データ）の**最低保持期間**を明確に定義します。
- **業務上の必要性の確認**：データが過去何年分まで業務上（例：製品の保証期間、カスタマーサポート）必要とされるかを業務部門と協議し、保持期間を決定します。

2. アーカイブ対象の選定とタイミング

- 保持ポリシーに基づき、移行時に「**SAP へ移行するデータ**」「**アーカイブシステムへ格納するデータ**」「**完全に破棄するデータ**」の 3 つに分類します。
- アーカイブのタイミングは、**移行前**（旧システムから直接アーカイブ）と**移行後**（SAP 稼働後に SAP 標準機能でアーカイブ）の 2 つがあります。
 - **移行前アーカイブ**：移行工数を最大化したい場合に有効。旧システム内で不要データを抽出し、アーカイブシステムへ格納します。
 - **移行後アーカイブ**：データの完全性を担保したい場合に有効。すべてのデータを SAP へ移行した後、SAP 標準のデータアーカ

イブツール（ADK など）を用いてアーカイブ処理を行います。

3. アーカイブシステムの選定

- アーカイブされたデータは、参照の必要が生じた際に、迅速かつ安全に参照できる必要があります。
- **参照機能の確保**：アーカイブシステムは、SAP の画面から直接、過去のアーカイブデータを参照できる機能（例：SAP Information Lifecycle Management - ILM）を提供していることが望ましいです。
- **長期保管とコスト効率**：アーカイブデータは頻繁には利用されないため、低コストで大容量のデータを安全に保管できるストレージ技術を選定します（例：クラウドストレージ、専用のアーカイブデータベース）。

事例：製薬会社 P 社のアーカイブ戦略

製薬会社 P 社は、厳格な規制要件（GxP など）に基づき、製品の製造記録や品質管理データを数十年にわたり保持する必要がありました。SAP S/4HANA への移行にあたり、P 社は過去 10 年分のトランザクションデータのうち、規制遵守に必要なデータを特定し、**SAP ILM**（Information Lifecycle Management）を導入しました。移行対象を直近 3 年分のデータに絞り、残りの 7 年分と古い過去データは、移行前に専用のアーカイブシステムに格納しました。この戦略により、移行工数を約 40%削減できただけでなく、S/4HANA の本番環境のデータベースサイズを最適化し、高いパフォーマンスを維持することに成功しました。

データ品質管理のこれらのステップは、SAP 移行プロジェクトを単なるシステム更新ではなく、**企業価値を向上させるデータ変革プロジェクト**へと昇華させます。データプロファイリングで真実を知り、クレンジングでデータを磨き、MDM でその輝きを維持し、アーカイブで不要な負荷を取り除く。この一連のアプローチこそが、移行成功の鍵となります。