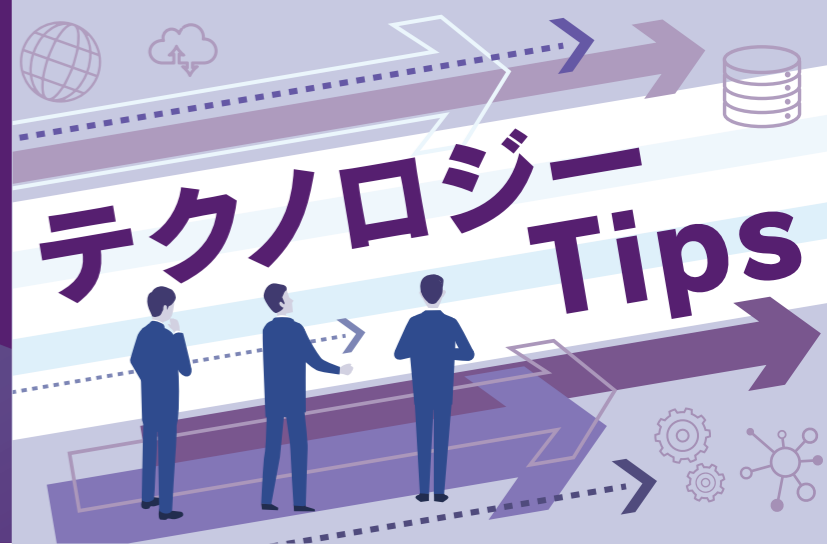


当たり前ですが、DXの基盤はツールやソフトウェアではありません。「データ」ですよ!

# わかっているようでわかっていないかも? 「データの本質」「課題」を整理します!

マスターデータとトランザクションデータの理解がデータ活用には重要です。  
基本から整理して、お客様のDX基盤構築、データドリブンにお役に立てれば嬉しいです!



今こそ知りたい  
Tech Tips!  
vol.17  
(Ver2.0)

## データの定義と種類

いまさら…かもですが、  
データの定義を明確にします。

### データ (data) とは?

物事の推論の基礎となる事実。また、参考となる資料・情報。事実や観測結果を数値や文字、記号、音声、画像などで表現。  
**一言で! 「判断材料になるモノ」**



## ●データの種類はどんなものがある? 代表的なカテゴリは?

<b>数値データ</b> 売上高、在庫数、顧客数などの数値 	<b>テキストデータ</b> 議事録、日報、メール、メモなどの文章 	<b>画像データ</b> 製品の写真、スキャンした書類などの画像、PDFなど 	<b>音声データ</b> 録音された会話、音楽などの音声 
--------------------------------------	--	---	-------------------------------------

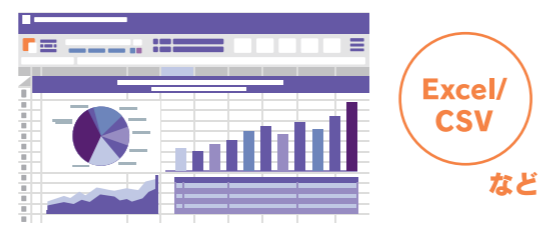
## 最近のデータ事情：非構造化データへの注目と評価

データには、構造化データと非構造化データがあります!

### 構造化データとは?

まずは一言で言いますね!**「元々データの構造が決まっている、行と列のある表形式で管理するのが簡単そうなデータ」**です! 決まった形式や構造に従って整理されます。例えば、Excelの表やデータベースのテーブルのように、行と列で整理されています。各列には特定の種類の情報(例えば、名前、年齢、住所など)が含まれ、各行は一つのレコード(例えば、一人の人の情報)を表します。

#### ●構造化データのイメージ



対して、非構造化データって?

### 非構造化データとは?

まずは一言で言いますね!**「表形式で管理してよ!」**と言われても**「えっ、どうやって?」**となるようなデータのことです。つまり、決まった形式や構造がないデータです。例えば、テキストファイル、画像、ビデオ、音声ファイル、ソーシャルメディアの投稿などがこれに当たります。これらのデータは、特定の形式に従って整理されていないため、分析や検索が難しいことがあります。

#### ●非構造化データのイメージ



**例** メール本文、写真やビデオ、PDFデータ、ソーシャルメディアの投稿、営業日報の報告文書、お客様からの要望や音声データ

<b>違いを整理しておきますね。</b>	<b>形式</b> 構造化データは決まった形式に従って整理されているのに対し、非構造化データは形式が決まっています。	<b>検索と分析</b> 構造化データは簡単に検索や分析ができるのに対し、非構造化データはそのままでは難しいことが多いです。	<b>保存方法</b> 構造化データはデータベースやスプレッドシートに保存されることが多いですが、非構造化データはファイル形式のままサーバーやNAS、パソコンの記憶デバイス、オンラインストレージに保存されることが多いです。
----------------------	---	---	--

最近では、SNSや動画の利用が拡大し、ビジネスにおいてもウェブマーケティングや顧客体験の向上に「非構造化データ」が活用されています。これにより、非構造化データの重要性がますます高まっています。しかし、構造化データと比べると、非構造化データはファイル形式が多様であり、管理、保存、検索、分析、再利用、信頼性や正確性の確保が難しいのが一般的です。非構造化データを「どのように保存し、管理し、扱うか」がデータドリブンの重要なポイントとなっています。

データ活用やデータの品質・正確性の維持/担保って難しいですね。  
 データクレンジングのポイント、そしてお客様が持つデータの「価値最大化」するためにはじめに注目すべきマスターデータについてもお話ししておきますね。

## データの整備・運用・活用の肝「マスターデータ」!

### マスターデータとは?トランザクションデータとは?

#### ●マスターデータ

マスターデータとは、ビジネスの基本的な情報を含むデータで、長期間にわたって安定しているものです。

例えば...	顧客情報	顧客の名前、住所、連絡先など
	製品情報	製品の名前、価格、仕様など
	従業員情報	従業員の名前、役職、給与など

マスターデータは、ビジネスの運営に必要な基盤となるデータであり、他のデータと結びつけて使用される。

#### ●マスターデータイメージ

社員マスタ

社員コード	氏名	部署
00001	AAAAA	XXXXXX
00002	BBBBB	YYYYY
00003	CCCCC	ZZZZZ

他のデータに属性を付与

勤怠データ

社員コード	出勤時間	退勤時間
00001	00:00	00:00
00002	00:00	00:00
00003	00:00	00:00

#### ●トランザクションデータ

一方で、トランザクションデータは、日々の業務活動や取引に関するデータです。

例えば...	販売データ	特定の顧客が特定の製品を購入した日時、数量、金額など
	営業日報データ	営業が訪問したお客様の状況、活動報告など
	支払いデータ	顧客からの支払い記録、支払い方法、支払い日時など

トランザクションデータは頻繁に更新され、マスターデータと結びついています。

#### ●トランザクションデータイメージ

社員マスタ

社員コード	氏名	部署
00001	AAAAA	XXXXXX
00002	BBBBB	YYYYY
00003	CCCCC	ZZZZZ

勤怠データ

社員コード	出勤時間	退勤時間
00001	00:00	00:00
00002	00:00	00:00
00003	00:00	00:00

日々追加更新

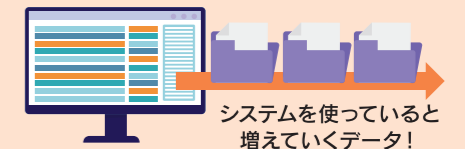
トランザクションデータ

#### 一言でまとめます!

##### マスターデータ



##### トランザクションデータ



#### ●勤怠管理の業務で考えてみましょう。

フロー1

タローさんが出勤してきて、社員番号「100」を入力すると、タローさんの名前と「出勤」「退勤」ボタンが表示されます。自分の名前間違いないことを確認したら「出勤」「退勤」ボタンを押します。これで、タローさんの出勤時刻が記録されます。

フロー2

このときのタローさんの出勤時刻を記録したデータがトランザクションデータです。他の社員さん全員も、「出勤」「退勤」ボタンをポチポチ押ししてシステムを使うことで蓄積されていくデータです。

フロー3

それに対して、社員番号と名前の対応データ「社員番号100の人の名前はタロー」という情報は、勤怠システムを動かす前の段階から用意しておく必要があります。用意しておかなかったら、最初の1人が社員番号を入力したときに、対応する名前を表示できませんからね。こちらはマスターデータです。

勤怠管理システムにおける出退勤情報をはじめ、売上管理システムにおける売上データや経理システムにおける仕分けデータのように、システムを使うことによって蓄積されていくデータはトランザクションデータです。一方で、勤怠管理システムにおける社員情報や売上管理システムにおける商品データ、経理システムにおける勘定科目のように、システムを使う最初の段階から無いと困るデータはマスターデータです。

Column 1

### データクレンジングって?

データクレンジングをわかりやすく説明すると、データの「お掃除」と考えると良いでしょう。データクレンジングは、データの中にある誤りや重複、不足している部分を修正して、データを正確で使いやすい状態にする作業です。例えば、名前のスペルミスや住所の入力ミス、同じデータが複数回登録されている場合などを修正します。

具体例

- 誤りの修正 ……例えば、「山田太郎」と「山田タロウ」が同じ人物である場合、表記を統一します。
- 重複の削除 ……同じ顧客が複数回登録されている場合、重複を削除して一つにまとめます。
- 不足部分の補完 ……住所の一部が欠けている場合、正しい情報を追加します。

Column 2

### ビッグデータとは?

日々生成される膨大な量の多種多様なデータのことを指します。例えば、SNSの投稿、GPSの位置情報、オンラインショッピングの履歴などなど…これらのデータは非常に大量で、多様性があり、リアルタイムで生成されることが特徴です。ビッグデータを活用することで、企業はお客様の行動パターンを分析したり、効率的なマーケティング戦略を立てたりすることができます。個人にとっても、健康管理や趣味の情報収集など、さまざまな場面で役立ちます。

#### ●ビッグデータイメージ



システムや業務の仕組み、ビジネス全体を加速させるためには、データが不可欠! 正確に進め、効果的に活用するためには、マスターデータがすべての基盤です。  
**まずは、マスターデータの「整備」と「クレンジング」を検討してみてもいかがでしょうか?**

当社がしっかりご提案、導入サポートいたします!

と当社にご相談ください!

