

データの利活用推進に向けた技術ガイドライン付録 データ品質の評価項目

ガイドライン本文の「2.2 データの品質」では、デジタル庁の「データ品質管理ガイドブック」で示されているデータ品質の 15 項目のうち重要 5 項目を取り上げた。ここでは、「データ品質管理ガイドブック」からの引用を基に全 15 項目を概説する。

① 正確性（重点項目）

データの基本であり、データの正しさを表すもの。データが正確でなければ、判断に誤りが生じる恐れがある。

- 書式が正しいか
- 誤字脱字などはないか
- 意味的な誤りがないか
- データに誤りはないか

② 完全性（重点項目）

データが完全であること。データは目的に応じて抜け漏れなくあることで、詳細な分析をすることが可能となる。

- 用途に応じて必要な項目が網羅されているか
- 必須項目に空欄が含まれていないか

③ 一貫性（重点項目）

データに矛盾がないこと。データには整合性や一貫性が必要で、データ内の項目や値に矛盾があると自部署内において、連続性のある利活用ができない(エラー処理が必要となる)。

- データセット¹内でデータに矛盾はないか
- データセット間でデータに矛盾はないか

④ 信憑性

データは意思決定に使われることも多く、信頼できるデータである必要がある。

- データの出所が明示されているか
- データの更新日が明示されているか
- 改ざん防止策が施されてあるか

¹ データセットとは、何かしらの目的の下で収集したデータを 1 つの表やファイル等にまとめた、データの集合を意味する。

⑤ 最新性（重点項目）

データが十分に新しいものに維持されていることデータが最新版に更新されていることで、誤処理や再処理が不要となる。

- 公開データの更新サイクルは元データの更新サイクルに対して適切か
- データは収集時から十分に短い期間で公開されているか
- 最終更新日時及び最新版の所在等、更新版の有無が確認可能か

⑥ アクセシビリティ

データを受け取った人がそのデータを活用できるようにする必要がある。データが誰でも使用できるものになっているか。

- ファイルで提供している場合、データの使用权を持つすべての人が利用できるようになっているか
- ソフトウェアを通して提供している場合、そのソフトウェアはISO/IEC40500に準拠しているか
- 使用している文字セット（常用漢字など）は正しいか

⑦ 標準適合性

データは入力ルール等の一定のルールにより管理されており、そのルールを守ることで円滑に処理することができる。

- データの書式は標準に準拠しているか
- 使用している文字セットは正しいか
- 選択項目に、指定された選択肢以外のデータが入っていないか

⑧ 機密性

データによっては機密情報を含むものもあり、目的に応じた機密性が確保される必要がある。

- データにアクセスできるのは、アクセスを許可された者に限定されているか
- 利用者を制限する場合、暗号化やハッキング対策等が行われているか

⑨ 効率性

データは効率的に処理される必要があり、そのためにコードを割り当てる等の対応を行う。

- データの内容に重複などがいないか
- データは効率的に処理できるようになっているか

- コードを効果的に使用しているか
- データに一貫性はあるか

⑩ 精度

データには使用目的に応じて必要な精度がある。また、精度の異なるデータを一体として扱う時に精度の調整が必要になる。

- データの精度は適正に設定されているか
- データの精度が揃っているか
- データの精度が示されているか

⑪ 追跡可能性

データに疑義が生じた時に、データの原典などを参照する場合がある。

- 外部データが明確になっているか
- データの変更の際に、変更者、変更日などを記録しているか

⑫ 理解性

データを活用するときには、データの項目を正しく理解して活用する必要がある。

- データ全体及びその各項目が意味するものを利用者が理解できるようになっているか
- データ全体や必要に応じてその各項目にメタデータ²が提供されているか
- 共通語彙基盤³のような意味を定めたものに関連付けがされているか

⑬ 可用性（重点項目）

データが利用可能な状態にあること。データは必要な時に使えるようになっていることが重要である。

- 必要な時にいつでもデータにアクセスできるようになっているか
- データを公開するシステムは常時稼働しているか

⑭ 移植性

² データを説明するデータ。例えば、写真データの場合、撮影日や写真につけた名前、撮影者等のデータがメタデータとなる

³ 情報処理推進機構が提供する共通的なデータ項目を定義するための基盤

システムの入れ替えやシステム間の連携を行う際には、データを簡易に移行できる必要がある。

- 標準的なフォーマットで出力できないソフトウェアに依存していないか
- データを管理するシステムから標準的な形式によりデータを出力することができるか

⑮ 回復性

データセンターなどで事故が起こった時に、そのデータが早急に復元される必要がある。

- データのバックアップが保存されているか
- システム障害が発生した場合であっても継続してデータを提供するバックアップシステムが存在するか