

Dataset-JSON

1.0 ⇒ 1.1 への主要な変更点

イーピーエス株式会社
森岡裕

1.0でFDAのパイロットスタディを経て，そのフィードバックを受けて

1.1となり，これをStable Versionと定めているので

基本的にはこれからDataset-JSONをつくるなら1.1に対応させた方がよい

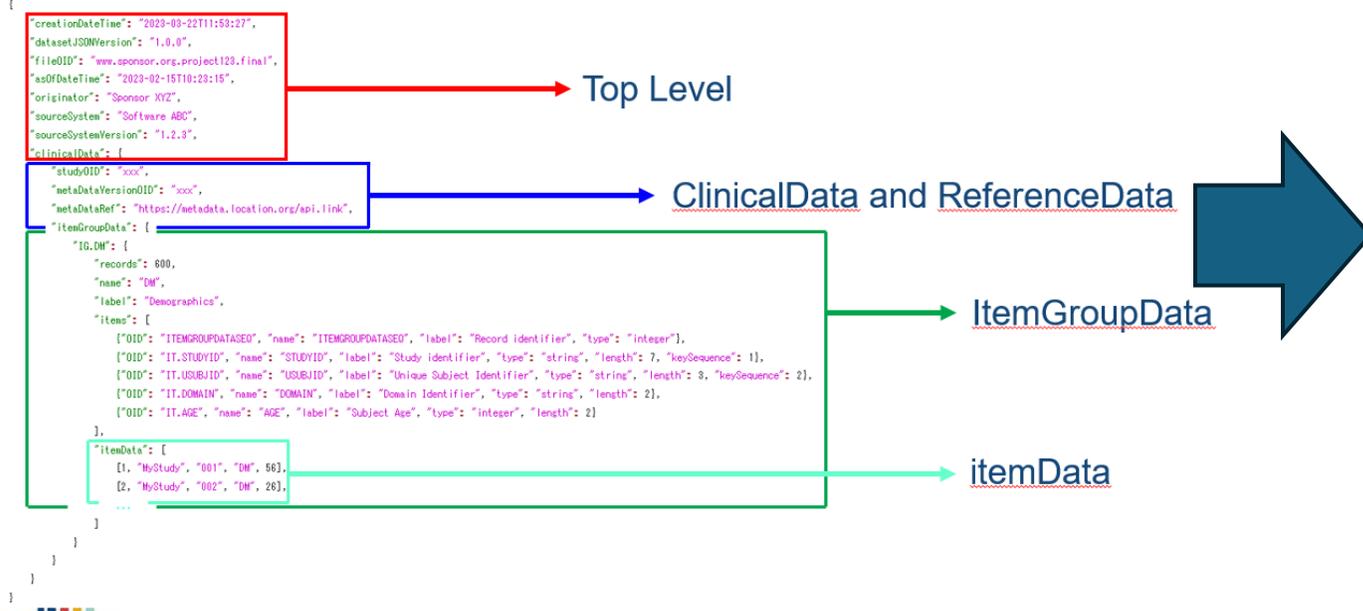
Dataset-JSON1.1 仕様

<https://cdisc-org.github.io/DataExchange-DatasetJson/doc/dataset-json1-1.html>

JSON構造のフラット化

Dataset-JSON全体の構造が平坦化され、処理しやすい形に変更されました。
たとえば、カラム（変数）メタデータを表すオブジェクト名が従来のitemsからシンプルにcolumnsに変更され、データレコード部分もitemDataからrowsに改名されています。つまり、変数定義は"columns"配下のリストで表現され、各行データは"rows"配列内のリストとして表現されます。これらの変更により、Dataset-JSON v1.1の構造はv1.0よりもフラットで理解しやすく、処理も簡易になっています

1.0



1.1

```
{
  "datasetJSONCreationDateTime": "2023-03-22T11:53:27",
  "datasetJSONVersion": "1.1.0",
  "fileOID": "www.sponsor.xyz.org.project123.final",
  "dbLastModifiedDateTime": "2023-02-15T10:23:15",
  "originator": "Sponsor XYZ",
  "sourceSystem": {
    "name": "Software ABC",
    "version": "1.0.0"
  },
  "studyOID": "xxx",
  "metaDataVersionOID": "xxx.y",
  "metaDataRef": "https://metadata.location.org/api.link",
  "itemGroupOID": "IG.DM",
  "records": 100,
  "name": "DM",
  "label": "Demographics",
  "columns": [ ... ],
  "rows": [ ... ]
}
```

OIDの必須要件緩和（転送にdefine.xmlを必要としない）

v1.1では、多くのOID（識別子）がオプション扱いになりました（v1.0では必須項目でした）。例えば、各データセットや変数に対応するOID（例: studyOID, itemOID等）は、Define-XMLとの連携に必要な場合のみ使用し、そうでなければ省略可能です。実運用上、Define-XML（データ定義XML）が存在する場合はDataset-JSON内のOID値をDefine-XMLと一致させる必要がありますが、Define-XML自体はDataset-JSONの必須要件ではありません。この変更により、Define-XMLを参照しないシンプルなデータ交換シナリオでは、OIDを省略してより簡潔なJSONを書けるようになりました。

データ型の拡充

扱えるデータ型が増えました。特に、Decimal（高精度10進数）型が新たにサポートされ、Boolean（論理値）型やDateTime（日時）型も明確に定義されました。これについては後述しますが、v1.1ではODM（Operational Data Model）のデータ型を取り込みつつ、新しいデータ型を導入することで、XPT（SAS v5 XPORTフォーマット）では表現できなかったデータも正確に扱えるようになっていきます。

追加メタデータ属性の導入

TargetDataType属性が新設されました。この属性は、JSON上の値をどのデータ型に変換すべきかを示すものです。例えば「数値として格納されている日付」をやり取りする場合、JSONではISO8601形式の文字列に変換して保持し、targetDataTypeで「本来の型（Integerなど）」を指定する、といった使い方をします（詳細は後述）。

日付の表現方法の変更

日付/日時データはISO 8601形式の文字列で表現する方針が取られました。v1.0では日付を数値（日数や秒数）で保持するケースがありましたが、相互運用時に異なる起点（例えばSASは1960年基準、Rは1970年基準）によるズレが生じる問題が指摘されました。v1.1では日時データは統一的にISO標準の日付文字列で格納し、必要に応じてtargetDataTypeで元の数値型への変換情報を付与することで、この問題に対処しています

Decimal（高精度10進数）データ型の追加

Decimal型は、浮動小数点数では表現が難しい高精度な数値を扱うために新設されたデータ型です。従来、XPT形式では数値は全て倍精度浮動小数点（約15桁精度）に丸められており、例えば有効数字が多い測定値や通貨計算などでわずかな丸め誤差が生じうる問題がありました。Dataset-JSON v1.1ではODM由来のDecimal型を導入することで、非循環小数（finite decimal）を丸めず正確に表現できます

例えば、30.8983333232059など15桁以上の数値を正しく転送したい場合などは以下のように定義します

```
{
  "itemOID": "IT.ADSL.BMIBL",
  "name": "BMIBL",
  "label": "Baseline BMI (kg/m^2)",
  "dataType": "decimal",
  "length": 16,
  "targetDataType": "decimal"
}
```

Date/Time（日時）データ型の扱い強化

バージョン1.1では日付・時刻系データの取り扱いが大きく改善されています。具体的には、新しいデータ型として**date**, time, **datetime**が（明示的に）サポートされ、併せてこれらをISO 8601形式で表現するルールが導入されました

ISO 8601形式への統一: v1.1では日付や日時を表す値は**"YYYY-MM-DD"や"YYYY-MM-DDThh:mm:ss"といったISO標準フォーマットの文字列としてJSONに格納します。例えばSDTMの開始日AESTDTCや終了日AEENDTCであれば、"2013-05-20"のような文字列になります。このように人間が読める標準表記**に統一することで、異なるシステム間で日付の解釈を誤るリスクを下げます

targetDataTypeによる元の型情報保持: 特にADaMなどで日付を数値（例えば治験開始日からの経過日数や秒数）で管理している場合、単に文字列化するだけでは元の値への復元ができません。そこでv1.1では、targetDataType属性を使って「この日付は元々数値型だった」ことを明示できます。targetDataTypeは必要な場合にのみ指定し、JSON上の形式（文字列）と論理型が異なる場合に使います

```
{
  "name": "TRTSDT",
  "label": "Date of First Exposure to Treatment",
  "dataType": "date",
  "targetDataType": "integer",
  "displayFormat": "E8601DA."
}
```

例: ADaMの治療開始日（TRTSDT）がSAS日付（1960年1月1日基点の日数）で保持されているとします。この変数はDataset-JSONでは例えば以下のように定義されます

targetDataType

v1.1では**targetDataTypeという新しいメタデータ属性が導入されました。この属性は「変数の値を最終的に変換すべきデータ型」**を示します。通常、dataType属性に論理型（そのデータが本来意味する型）を指定しますが、実際のJSON上の型との違いがある場合にtargetDataTypeが活躍します

FDAパイロットテストでは、変数のフォーマット情報（displayFormat）だけでは「この数値は日付として解釈すべき」などの指示が不十分であることが指摘されました。

targetDataTypeはこのギャップを埋めるために導入されたもので、データ交換後の再構築を正確に行うための**変更指示書**として機能します

NDJSON (改行区切りJSON) 形式のサポート

NDJSON (Newline Delimited JSON) は、巨大なJSONデータを効率よく扱うためのフォーマットです。Dataset-JSON v1.1ではこのNDJSONでデータセットを表現できるオプションが追加されました

形式の概要: NDJSONでは、各レコード（または各JSONオブジェクト）を1行ずつに記述し、改行で区切ります。一般的なJSONファイルは単一の巨大な配列やオブジェクトとして全データを含みますが、NDJSONは行単位で完結したJSON構造を持つため、ストリーミング処理が可能です

目的: 大規模データセットの処理性能向上

v1.1仕様では通常のJSONとNDJSONのどちらを使ってもよいので、目的に応じて使い分けることができます

<https://cdisc-org.github.io/DataExchange-DatasetJson/doc/dataset-json-ndjson1-1.html>

```
1 datasetJSONCreationDateTime: 2024-11-11T15:09:14, datasetJSONVersion: 1.1.0, fileOID: www.cdisc.org/study/M8GZ2/Define-XML/2.1.0/2024-11-11/
2 ["CDISCPIL001", "AE", "CDIS0001", 1, "1", "INJECTION SITE REACTION", "MODERATE", "N", "DOSE NOT CHANGED", "UNLIKELY RELA
3 ["CDISCPIL001", "AE", "CDIS0001", 2, "2", "FATIGUE", "MODERATE", "N", "DOSE NOT CHANGED", "UNLIKELY RELA
4 ["CDISCPIL001", "AE", "CDIS0002", 1, "1", "INJECTION SITE REACTION", "MILD", "N", "DOSE NOT CHANGED", "F
5 ["CDISCPIL001", "AE", "CDIS0002", 2, "2", "SHOULDER PAIN", "MILD", "N", "DOSE NOT CHANGED", "NOT RELATEC
6 ["CDISCPIL001", "AE", "CDIS0002", 3, "3", "JOINT DISLOCATION", "MODERATE", "N", "DOSE NOT CHANGED", "NOT
7 ["CDISCPIL001", "AE", "CDIS0002", 4, "4", "INCONTINENCE", "MILD", "N", "DOSE NOT CHANGED", "NOT RELATEC
8 ["CDISCPIL001", "AE", "CDIS0002", 5, "5", "INJECTION SITE REACTION", "MILD", "N", "DOSE NOT CHANGED", "F
9 ["CDISCPIL001", "AE", "CDIS0002", 6, "6", "SKIN LACERATION", "MILD", "N", "DOSE NOT CHANGED", "NOT RELAT
10 ["CDISCPIL001", "AE", "CDIS0002", 7, "7", "CONFUSIONAL STATE", "MODERATE", "N", "DOSE NOT CHANGED", "NOT
11 ["CDISCPIL001", "AE", "CDIS0002", 8, "8", "DYSPNICIA", "MODERATE", "N", "DOSE NOT CHANGED", "NOT RELATEC
12 ["CDISCPIL001", "AE", "CDIS0002", 9, "9", "SUDDEN DEATH", "SEVERE", "N", "DRUG WITHDRAWN", "NOT RELATEC
13 ["CDISCPIL001", "AE", "CDIS0003", 1, "1", "HYPERHIDROSIS", "MILD", "N", "DOSE NOT CHANGED", "POSSIBLY RE
14 ["CDISCPIL001", "AE", "CDIS0003", 2, "2", "STOMACH DISCOMFORT", "MILD", "N", "DOSE NOT CHANGED", "POSSIE
15 ["CDISCPIL001", "AE", "CDIS0003", 3, "3", "PAIN", "MILD", "N", "DOSE NOT CHANGED", "POSSIBLY RELATED", "
16 ["CDISCPIL001", "AE", "CDIS0003", 4, "4", "NASAL CONGESTION", "MILD", "N", "DOSE NOT CHANGED", "NOT RELA
17 ["CDISCPIL001", "AE", "CDIS0003", 5, "5", "MALAISE", "MODERATE", "N", "DOSE NOT CHANGED", "UNLIKELY RELA
18 ["CDISCPIL001", "AE", "CDIS0003", 6, "6", "MYALGIA", "MILD", "N", "DOSE NOT CHANGED", "POSSIBLY RELATED",
19 ["CDISCPIL001", "AE", "CDIS0003", 7, "7", "PHARYNGOLARYNGEAL PAIN", "MILD", "N", "DOSE NOT CHANGED", "NE
20 ["CDISCPIL001", "AE", "CDIS0003", 8, "8", "PRODUCTIVE COUGH", "MILD", "N", "DOSE NOT CHANGED", "NOT RELA
21 ["CDISCPIL001", "AE", "CDIS0003", 9, "9", "DIZZINESS", "MODERATE", "N", "DOSE NOT CHANGED", "UNLIKELY RELA
22 ["CDISCPIL001", "AE", "CDIS0003", 10, "10", "DIZZINESS", "MILD", "N", "DOSE NOT CHANGED", "POSSIBLY RELA
23 ["CDISCPIL001", "AE", "CDIS0003", 11, "11", "CONFUSIONAL STATE", "MILD", "N", "DOSE NOT CHANGED", "UNLIKELY RELA
```

```
{
  "datasetJSONCreationDateTime": "2023-03-22T11:53:27",
  "datasetJSONVersion": "1.1.0",
  "fileOID": "www.sponsor.xyz.org.project123.final",
  "dbLastModifiedDateTime": "2023-02-15T10:23:15",
  "originator": "Sponsor XYZ",
  "sourceSystem": {
    "name": "Software ABC",
    "version": "1.0.0"
  },
  "studyOID": "xxx",
  "metaDataVersionOID": "xyz",
  "metaDataRef": "https://metadata.location.org/api.link",
  "itemGroupOID": "IG.DM",
  "records": 100,
  "name": "DM",
  "label": "Demographics",
  "columns": [ ... ],
  "rows": [ ... ]
}
```