

メタデータに関する 概念整理

プレスト資料

2025/11/27
MCM-JP CG

データの種別

1. コンテンツ (Content)

- **体験の主な対象**となるデータ
 - 放送: 映像フレーム、音声トラック
 - 出版: 本文テキスト、イラスト
 - 音楽: 音声波形、MIDI

区分は、データの中身によって決まるのではなく、利用目的で決まる。
(例: 字幕テキスト、書籍の目次は色んな役割)

2. メタデータ (Metadata)

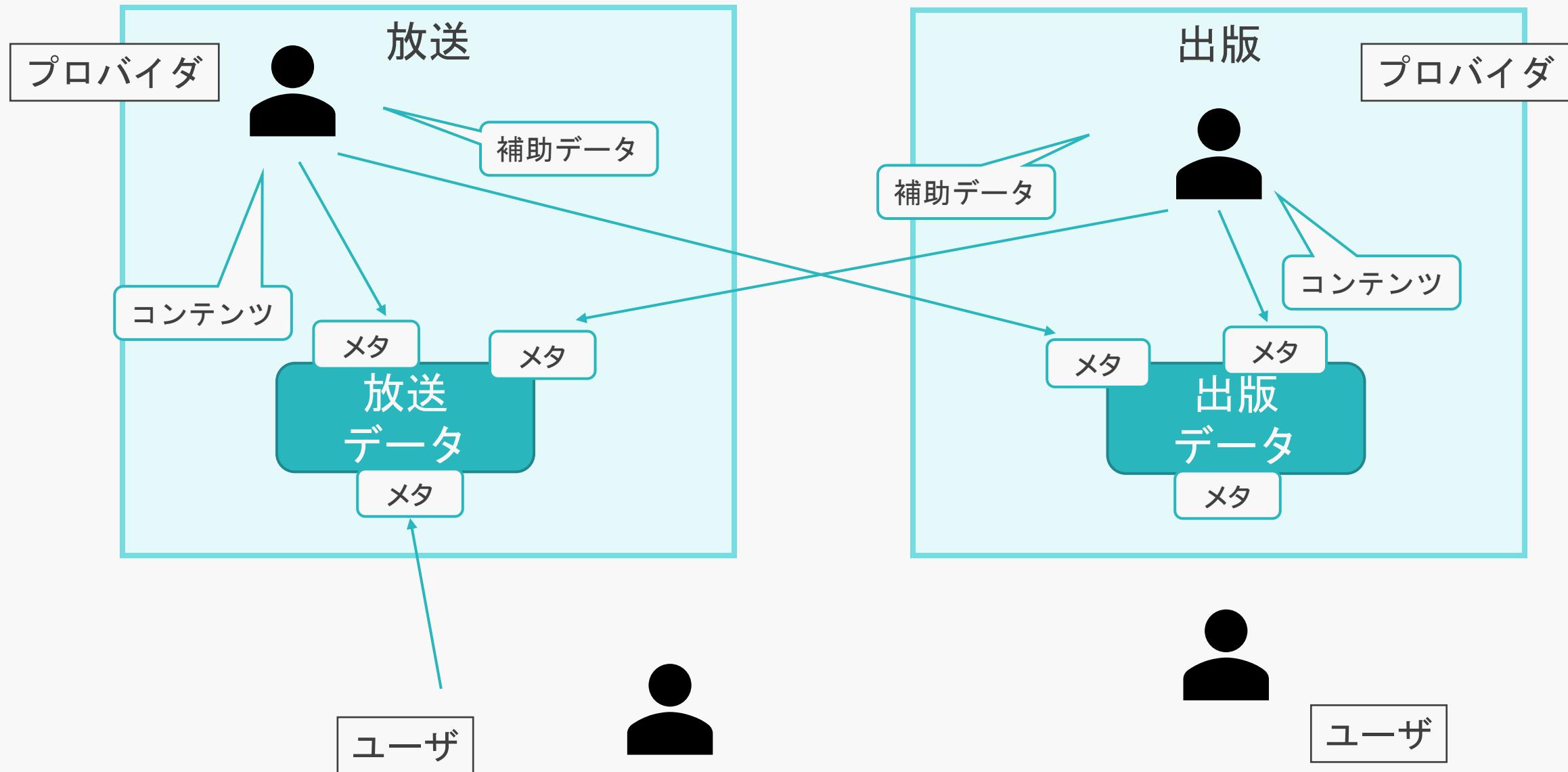
- コンテンツを **識別・記述・管理**する情報
 - タイトル、著者、ジャンル、権利情報、技術仕様、ISBN、DOI

3. 補助データ (Ancillary / Sidecar / Timed Metadata)

- コンテンツの**理解・利用・検索・同期などを補佐**する付随情報
 - 字幕、チャプター、タイムコード、広告挿入点、サムネイル、特徴量

構造化データの役割

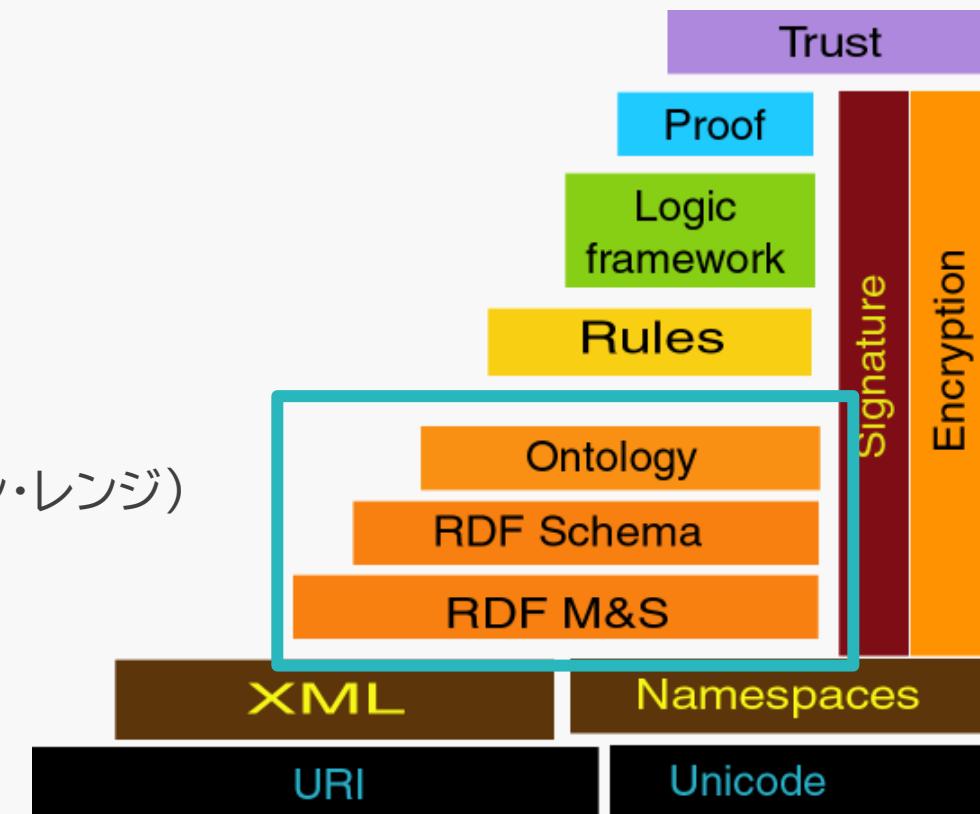
人・サービス・システムを相互につなぐ情報（ルール）



データ構造化とセマンティックレイヤー

- Ontology層（オントロジー層）
 - 役割：語彙間の関係性や制約を形式的に定義
 - schema:TVEpisode owl:equivalentClass ebu:Programme
 - 「ニュース番組は必ずジャンル=ニュースを持つ」
- Vocabulary層（語彙層）
 - 役割：メタデータ記述に使う用語集合
 - Dublin Core: dc:title, dc:creator
 - Schema.org: schema:name, schema:datePublished
- Schema層（スキーマ層）
 - 役割：語彙の構造や型を定義(クラス、プロパティ、ドメイン・レンジ)
 - schema:TVEpisode rdfs:subClassOf schema:CreativeWork

参照するための
識別情報、名前空間



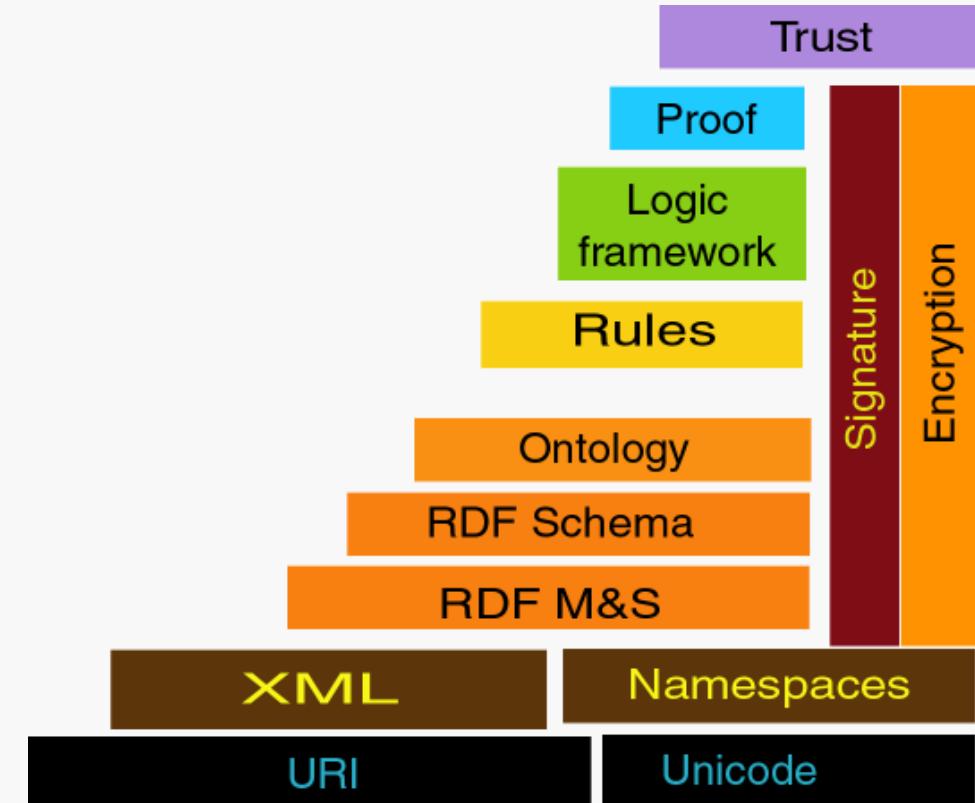
The Semantic Web "Layer Cake", Tim Berners-Lee
<https://www.w3.org/2004/Talks/0412-RDF-functions/slide4-0.html>

構造化データとはとAI時代での役割

- 語彙: title, creator
- スキーマ: TVEpisode
- オントロジー:
- 確率的推定には、必須ではない
- しかし、**あると精度が向上する**
 - 研究の時代
 - スケールの時代
 - 研究の時代

今後は効率化の時代。
構造化データ需要はさらに高まる

今後は効率化の時代。需要はさらに高まる
Ilya Sutskever



The Semantic Web "Layer Cake", Tim Berners-Lee

<https://www.w3.org/2004/Talks/0412-RDF-functions/slide4-0.html>

AIの2分類（ざっくり）

- 知識ベース型AI (Symbolic AI)
 - Deterministic Reasoning(決定的推論)
 - 論理ルール・オントロジーに基づく決定的な推論
- 機械学習型AI
 - Probabilistic Reasoning(確率的推論)
 - 統計モデルやニューラルネットによる推定