

音楽と画像コンテンツへのメタデータ技術の応用 Metadata Application Technologies for Music and Video Contents

新 麗

Ray ATARASHI

株式会社 IJ イノベーションインスティテュート

IJ Innovation Institute Inc.

E-mail: ray@ijlab.net

1. はじめに

スマートフォンのようなデバイスが普及して、誰もがビデオカメラを持ち歩くような状況になり、音声や画像コンテンツが簡易に作成できるようになった。また、パノラマ撮影ができるカメラや 360 度カメラなど新しい記録デバイスも登場している。音については立体音響のような技術が普及し、臨場感のある音の再現が可能となってきた。

さらにネットワークの普及によりコンテンツを共有することが容易になった。これらは視聴するだけでなく、編集機能の簡易化・高機能化により、共有されたコンテンツを元に新しいコンテンツを作成することも多く行われている。現在のところは 2 次元のコンテンツが主流であるが、いずれ 3 次元のコンテンツの流通が増えることが予想される。

このような画像と音声のそれぞれで 3 次元情報化技術の進歩を受けて、さらにそれをバーチャル空間上で組み合わせて再現する試みが、SDM (Software Defined Media) コンソーシアムで行われている [1][2][3]。実際のライブや音楽コンサートのデータを記録して、画像と音楽とを融合させたコンテンツを作成し、一部のデータは公開している。これまでの数回の実験を通じて、コンテンツの効率的な作成にはデータおよびメタデータの管理が重要であることがわかってきた。記録したデータを組み合わせることでコンテンツ作成は可能であるが、再利用性や拡張性に欠け、そこから新しいコンテンツを作成することは困難である。音声・画像だけでなく加工や公開まで行うには、各種標準や流通方式までを意識したデータ設計が必要となる。本稿では、記録作成した音楽コンサートのデータを対象とし、既存のメタデータ技術の調査を行い、新しいコンテンツ作成とその流通のためのメタデータとその応用について考察する。

本稿では、まず第 2 章で SDM の概要とデータセットについて紹介し、第 3 章でデータモデルとその関連技術、第 4 章でメタデータの標準を調査しつつデータ

セットへの適用を検討する。第 5 章で三次元情報のデータ管理について考察する。

2. Software Defined Media

本節では、SDM コンソーシアムと本稿で対象とする音楽コンサートのデータについて説明する。

2.1 Software Define Media の概要

現在の映像システムや音響システムは、コンサート会場やスタジオなど特定の空間で設計・設置されている。これらのシステムは、各種の録画・録音システム、編集システムなどのサブシステムで構成されており、サブシステム間はアナログ線で接続されている。しかし IP ネットワークの広帯域化により、IP による接続が可能となってきつつある。

Software Define Media コンソーシアムは、映像・音響システムの IP ネットワーク化を受けて、これらのサブシステムを IP 上の機器として一つのシステムとして扱い、ソフトウェアにより制御するための試みを行っている。機器の機能を抽象化、仮想化し、サービスとしての映像・音響を提供するというアプローチをとり、新しいデジタルメディアの創造をめざすのが目的である。スタジオ内だけでなく、実際のコンサートホールやライブ会場での録画・録音も行い、コンテンツの制作実験を行なっている。

2.2 コンサートのデータセット

本稿で扱うデータは、2016 年 1 月 10 日に収録されたものである。会場は、慶應義塾大学日吉キャンパス内に建ち 509 席を有する藤原記念ホールで、演奏は慶應義塾大学コレギウムムジクム古楽アカデミーである。収録対象は、17 世紀のドイツの宮廷音楽であるヨハン・フリードリッヒ・ファッシュの「管弦楽組曲ト長調 FaWV K:G2」などヨーロッパの古楽の分野に属する曲で、最大 24 名の演奏家によるアコースティックの演奏である。楽器は、現代のオーケストラではあまり使用されない、バロック・ヴァイオリン、バロック・オ

ーボエ、ヴィオラ・ダ・ガンバ、チェンバロ、テオルボなどで、17世紀当時の奏法にならって演奏された。

録音は、舞台上、客席に各所に配置された合計 82 チャンネルのさまざまな方式のマイク、360度カメラ、4K カメラなどを用いて行われた[4]。この中には業務用、実験用の機材からスマートフォンまでが含まれる。SDM ではこれらのデータを統合し、位置情報に基づいて音源と映像をリンクさせ、インタラクティブに自由な場所で音を聞くことができるアプリケーション、“360Square”を開発した。観客席だけでなく、普通は聞くことのできない舞台上での音を体験でき、また聞きたい場所だけを選択したり複数地点の音だけを同時に聞いたりすることも可能である。

このアプリケーションは、音楽コンサートだけでなくスポーツや会議、パーティーなどさまざまなユースケースへの適用が可能である。観客席からは見られない映像や音をさまざまな角度から得ることができ、体験の幅が広がることが期待される。現在、さまざまな三次元情報の記録とコンテンツ作成を試行している。

3. 三次元情報の管理の問題点

三次元情報の記録には、これまで特別の録音・録画機材とその調整のためのさまざまな機器、位置を調整するためのスタンド等、さらにそれらをつなぐアナログ線などの大規模の設備が必要であった。しかし前述のようにスマートフォンや新しいデバイスの普及により、安価に多チャンネルの情報が記録可能となり、さらにネットワークで収集して再構築することが可能となっている。さらに API 機能の搭載などで遠隔操作が可能な機器もあり、操作性が向上している。

しかし、録音・録画機材のデータの記録方式は統一されているとは言えず、そもそも解像度などの品質も異なる。また、記録に関するメタデータは、一部自動的に付与されるものもあるが、その後の編集には不十分である。スマートフォンの位置情報などは GPS の精度では不足であるし、室内での記録には利用できない。また、演奏された曲目や会場の情報などは機材とは別に付与する必要がある。

記録した情報から一つのコンテンツを制作する場合には、記録の状況を熟知した人間がメタデータを補完し編集することが可能であるが、インターネット等で共有されて再利用を促進する場合にはメタデータ情報を完備することが重要である。また、人手でのコンテンツ制作も短時間かつ限られた範囲であれば可能であるが、機材がさらに増えて複雑になっていった場合には、効率の良い検索や関連情報の提示などが欠かせなくなる。

以上の状況に対応するためには、関連しあう情報を

効率よく管理できるデータモデルとメタデータ設計が必要となる。メタデータ研究は歴史が長く、すでにさまざまな場面で利用されており、既存の標準を利用しなければ検索性が落ちるため、新規設計は慎重にすべきである。そこで本稿では、データモデルとメタデータ技術の最新動向を調査し、三次元情報への応用を検討する。

3. データモデル技術

データモデル技術は各分野において行われているが、本稿で求めるのはデータベース技術ではなく関連するデータを接続し共有していく技術である。これについては、Web の技術とともに発展してきた Linked Data の概念が最適であると思われる。本節では、Linked Data とそれを支える関連技術について調査する。

3.1 Linked Data

Linked Data とは、World Wide Web を考案した Tim Berners-Lee が提唱した、Semantic Web の仕組みを使ってデータをつなげる思想である。具体的には Web 上にあるリソースの記述方式である枠組みである RDF (Resource Description Framework) を利用して、データのネットワークを作っていく。仕組みとして持つのはデータを結合する技術だが、オープンなデータを増やしてそれをつないで初めて価値が生まれるという思想のもと、Linked Open Data としての活動が行われている。

Linked Data の基本となる RDF にはさまざまな記述方式があり、目的やツールのサポートに応じて利用されてきた。これまでに RDF/XML, RDF/Turtle, RDFa などが提案され、利用されてきたが、近年の Web と REST 技術の普及に合わせ、JSON-LD[6]という形式が登場し標準化された。JSON をベースにして拡張した形式であり、Web プログラムと相性が良い。

SDM で記録したデータは実験データであり、コンソール内での利用だけでなく公開して再利用を促進することも考えている。Linked Open Data の思想に基づき、RDF で記述しデータを結合できるようにする必要がある。また、拡張性を考え、JSON-LD の採用を検討している。

3.2 グラフデータベース

Web で公開するリソースは、データベースに格納されているものも多い。RDF はグラフ化されたデータモデルであり、直感的にリソースを格納するにはグラフ型のデータベースが適するが、グラフデータベースは格納方式が複雑になり、検索が難しいため、大容量のデータを扱える性能が出なかった。しかし近年、コン

コンピュータの高速化とメモリが大容量化したことから、実用に耐えるグラフデータベースが登場してきている。

GraphDB[5]は、Ontotext 社が開発したグラフデータベースである。Open Linked Data への対応を掲げており、RDF やオントロジ記述言語である OWL に対応しているほか、RDF のクエリ言語である SPARQL にも対応している。サポートしている RDF の記述形式は、RDF/Turtle, N-Triple などであり、JSON-LD はないが、変換することで利用することは可能である。

現在グラフデータベースとして評価が高いものには、Neo4j[7]がある。大規模なグラフ解析が可能であり、Social Network の解析などに実績があり、REST 対応など Web 技術との親和性は高い。しかし RDF 形式には対応しておらず、クエリ言語は cypher と呼ばれる独自のものである。

3.3 SPARQL endpoint

データを RDF で表記して公開するにあたっては、外部から検索するための入り口となる SPARQL endpoint が必要となる。基本的には RDF を格納し、SPARQL のクエリを HTTP で受け付ける Web サーバである。過去から開発が行われてきたが、Linked Open Data の普及とともに性能が向上してきている。Web 対応と拡張可能性への要求が高く、JSON-LD への対応も進んでいる。

主な SPARQL endpoint には、StarDog[8]、Apache のモジュールとして提供されている Apache Jena Fuseki [9]、AllegroGraph [10] などがある。今回は、サポートする RDF 記述形式の数、オープンソースである点、サポートするプラットフォームなどを比較検討した結果、StarDog を利用する予定である。

4. メタデータ

メタデータとはデータについてのデータ、つまり、データに付随して持つその付加的なデータを指す。Web の登場でデータの相互流通が促進され、メタデータおよびその語彙を標準化・公開することの必要性が重視されるようになった。しかし、メタデータはデータの管理団体の管理方針により定められていることが多く、通信プロトコルのような世界標準を定めるのは困難である。

そのなかで、共通の認識となる語彙となるよう設計された基本語彙セットとそれらをサポートするメタデータ語彙を定義しているのが Dublin Core[11]である。他のメタデータ語彙を定義する場合に、共通化できるものがあれば Dublin Core を利用することが推奨されている。ただし Dublin Core は基本語彙のみのため、特定の分野のデータ管理に使用するには語彙が不足することが多い。そこで、それぞれの団体や組織でメタ

データやその概念表現であるオントロジが定義され公開されている。

本稿では音楽コンサートのメタデータ語彙の設計を目標としているため、音楽情報を記述する Music Ontology、および近年の Web データに利用される語彙である schema.org を調査する。

4.1 Music Ontology

Music Ontology[12]は、Web 上の音楽に関するデータを公開し結合していくための語彙の定義である。RDF をベースに定義されており、RDFa, RDF/Turtle, JSON-LD で記述することを意識している。FOAF (Friend of a Friend)[13], The Event Ontology [14], The Timeline Ontology [15], FRBR ontology [16]の4つのオントロジを利用し、その上で音楽に必要な語彙を規定している。定義されているのは、MusicArtist や MusicalWork など 53 のクラスと activity_start/end や composed_in など 153 のプロパティである。

MusicArtist などは汎用性が高く、共通化しておくことで Web における検索性が向上しデータの結合を推進しやすいと考え、採用を検討中である。

4.2 Schema.org

メタデータ、オントロジは、各分野、各組織で目的に合わせて定義されており、公開されているものだけでもかなりの数にのぼる。図書館や博物館のように大規模な所蔵物があり、その中での検索が行われている場合には、横断検索を実現しようとしたとしても、対象となる組織数がある程度わかっているため、対処の方法がある。しかし Web 上に公開されているデータの検索サービスのようにならぬ多数の組織から情報を集める場合には、検索システム側で語彙変換や統一を行うのはほぼ不可能である。

2011 年頃から、Google, Microsoft, Yahoo, Yandex が共同で語彙統一の作業をしているのが schema.org[17]である。2015 年からは W3C の schema.org Community Group となり、作業を続けている。Schema.org は、microdata, RDFa, JSON-LD での記述が可能である。

Schema.org は、リソース作成側が Web の検索エンジンに対してメタデータを付与する仕組みであり、作成側が対応することで検索のヒット率が上がる可能性がある。検索エンジン最適化(SEO)技術として注目されているが、Linked Open Data の目的でも利用できると考えられる。録音・録画データにおいても、例えば曲目などの Web での検索性を要求されるメタデータは、Schema.org を採用する可能性を検討している。

5. 今後の展望

本稿では、新しい録音・録画デバイスの登場を受けて、画像と音声を融合させた新しいコンテンツを作成する試みについて紹介した。その際、メタデータ情報が重要性を増していることを述べ、既存の技術の適用可能性について調査した結果、基本技術や標準、ツールが進展し、実用化に近づいているという結論を得た。

メタデータはなるべくデータ作成時に付与するのが最も効率が良いが、それには作成者による記録が欠かせない。しかし、データ作成の現場は録画・録音に集中してしまい、メタデータ付与は不十分になることが多い。そこで、機器やソフトウェアでの対応が必要となる。デジタル写真については、exif[18]という標準があり、撮影日時やGPSによる撮影位置のほか、絞りなどカメラ側のパラメータも保存することで、編集や検索の効率を向上させている。しかし、動画や音声、さらに360度カメラのように新しいデバイスについては、標準規格のようなメタデータは付加されていない。

機器に付随するメタデータは、機器を操作できるAPIのパラメータと見ることもできる。現状は人手で調整している各種パラメータをAPIで操作しメタデータとして保存すると、次回は自動で設定することができるようになり、再利用性、再現性が格段に向上する。

また、デバイスが増え、公開によってアクセスできる機会が増えると、図1のように最終形は加工の加工という形で視聴者に届く。これが一度限りのデータでなく、さらに加工するために元データへ遡ったりするには、たどっていくことが可能な連結されたデータモデルが必要となる。

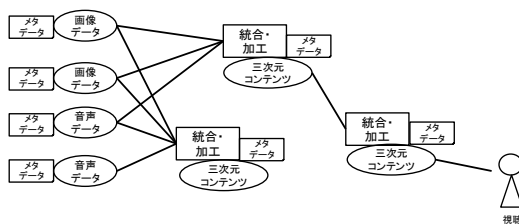


図 1. 三次元情報のデータ処理の概念図

メタデータやオントロジ技術は、図書館などのデータ管理や知識表現などを目的として発展してきた。ところが、デバイスのオープン化、APIによる機能公開の流れを受けて、ソフトウェアによるデバイスの制御

や管理にも利用できる可能性が出てきている。今後は、録音・録画データのメタデータだけでなく、このようなコンテンツをネットワークで配信する場合も視野に入れ、ネットワークのメタデータの検討および相互接続も検討していく予定である。

文 献

- [1] 塚田学, 小川景子, 池田雅弘, 曾根卓朗, 丹羽健太, 齊藤翔一郎, 粕谷貴司, 砂原秀樹, 江崎浩, “Software Defined Media: 視聴空間サービスのソフトウェア制御”, 日本ソフトウェア科学会学会誌「コンピュータソフトウェア」『ネットワーク技術』特集, 2017.
- [2] Manabu Tsukada, Keiko Ogawa, Masahiro Ikeda, Takuro Sone, Kenta Niwa, Shoichiro Saito, Takashi Kasuya, Hideki Sunahara, and Hiroshi Esaki “Software Defined Media: Virtualization of Audio-Visual Services”, IEEE International Conference on Communications (ICC2017), May 2017.
- [3] SDM (Software Defined Media) コンソーシアム, <http://sdm.wide.ad.jp/>.
- [4] Masahiro Ikeda, Takuro Sone, Kenta Niwa, Shoichiro Saito, Manabu Tsukada, and Hiroshi Esaki. New recording application for software defined media. In Audio Engineering Society Convention Paper, 141st AES Convention, Los Angeles, USA, September 2016.
- [5] Ontotext, <http://ontotext.com/>
- [6] JSON-LD 1.0, A JSON-based Serialization for Linked Data, W3C Recommendation 16 January 2014, <https://www.w3.org/TR/json-ld/>
- [7] Neo4j, <https://neo4j.com/>
- [8] StarDog, <http://www.stardog.com/>
- [9] Apache Jena Fuseki, https://jena.apache.org/documentation/serving_data/.
- [10] AllegroGraph, <https://franz.com/agraph/allegrograph/>
- [11] Dublin Core, <http://dublincore.org/>
- [12] Y. Raimond, S. Abdallah, M. Sandler, F. Giasson, “The music ontology”, In: Proceedings of the International Conference on Music Information Retrieval, pp. 417-422, September 2007
- [13] FOAF Vocabulary Specification 0.99, <http://xmlns.com/foaf/spec/>
- [14] The Event Ontology, <http://motools.sourceforge.net/event/event.html>
- [15] The Timeline Ontology, <http://motools.sourceforge.net/timeline/timeline.html>
- [16] The Functional Requirements for Bibliographic Records (FRBR) Ontology, <http://vocab.org/frbr/core.html>
- [17] <http://schema.org/>
- [18] デジタルスチルカメラ用画像ファイルフォーマット規格 Exif 2.21 統合版, JEITA CP-3451B