

画像データについて

1 JPEG2000の概要と国立公文書館の選択

JPEG2000は、JPEGを発展させた仕様で画像符号化に関する国際標準化機構（ISO）と国際電気通信連合・電気通信標準化セクタ（ITU-TS）の検討によって、2001年1月に規格化された¹。従来のJPEGよりも高圧縮かつ高品質なデータ処理が可能なウェーブレット変換で圧縮する方式を採用したことがその特徴の一つである。こうした圧縮方式の採用もあって、従来と比較して圧縮効率が大きく向上すると共に、それまで高圧縮時に問題となっていたノイズの発生が無くなり、さらに、全くデータ損失がない可逆圧縮も可能となった。また、近年、重要視されているセキュリティやコンテンツ保護、メタ情報の格納などの観点から、「電子透かし」やXML等への対応も標準化されるなど、新たな試みを取り入れられている。さらに、JPEG2000ではデータが圧縮された状態でさらに内部的に圧縮されているデータから必要な部分のみを抽出することが可能なため、一つの画像データから、高画質な画像、通常画質の画像、大きなサイズの画像、携帯電話向けの小さな画像など、様々な解像度、サイズの画像データを動的に取り出すことが出来る。

こうした画像データ自身に多様な特徴を持つJPEG2000であるが、さらなる特徴と

¹ JPEG2000は、ISO/IEC 15444シリーズとして標準化されている。その概要について次の文献をあげておく。David S. Taubman and Michael W. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*, in *Kluwer International Series in Engineering and Computer Science*, Secs 642 (Boston: Kluwer Academic, 2002), see also, Tinku Acharya and Ping-Sing Tsai, *JPEG2000 Standard for Image Compression: Concepts, Algorithms and VLSI Architectures* (Hoboken: John Wiley & Sons, 2005). 小野定康、鈴木純司『わかりやすいJPEG2000の技術』オーム社、2003年。福原隆浩、板倉英三郎『JPEG2000詳細解説——JPEG2000/Motion-JPEG2000/ストリーミング』CQ出版社、2004年。標準化の動向については、JPEG-JBIGの公式ウェブサイトを参照せよ。Official site of the Joint Photographic Experts Group, JPEG and Joint Bi-level Image Experts Group, JBIG. <<http://www.jpeg.org/>> (August 31, 2005). その他、略号は以下のとおり。ISO: International Standard Organization, IEC: International Electrotechnical Commission, ITU-TS: International Telecommunication Union Telecommunication Standardization sector.

² ISO/IEC FDIS 15444-9, Information Technology – JPEG 2000 Image Coding System – Part 9: Interactivity Tools, APIs and Protocols. For example see, David S. Taubman, Robert Prandolini, "Architecture, Philosophy, and Performance of JPIP: Internet Protocol Standard for JPEG 2000," *Proceedings- SPIE The International Society for Optical Engineering*, Vol. 5150 (Jun 2003), 791-805.

して、インターネット配信に関する新たなプロトコル（通信規約）であるJPEG2000 Internet Protocol (JPIP) が標準規格の中に含まれ、画像データの提供側と受け取り側との間における双方向通信が規定された²。このプロトコルを通じて、適宜、クライアント端末が求めるデータのみをサーバーから端末へ送信し、またユーザーが閲覧するデータをそれに先だってキャッシュする仕組みを取り入れたことで、大きなデータであってもスムーズに配信し閲覧できる条件が整ったのである。以上のように、JPEG2000は、従来の様々なデジタル画像データ形式、配信・閲覧システムがそれぞれ独自に持っていた利点を総合化しながら、従来の画像配信・閲覧システムが持つ様々な弱点を補正したものであると言える。そして、現在、静止画像、動画 (Motion-JPEG2000)、データ配信プロトコルなど、全体のプロセスを包括的に定義し国際標準化が行われた唯一の規格であり、「ユビキタス社会」を迎える21世紀の幕開けに相応しい次世代規格であると言えるだろう³。

さて、こうした画像フォーマットを選択した「国立公文書館デジタルアーカイブ」であるが、他の画像データ配信・閲覧システムに比較した技術的な先進性、有意性のみを理由としたわけではない。当館の「利用と保存」という基本的な目的がその鍵となる。JPEG2000では、情報ロスを生じない可逆圧縮が可能であり、非可逆圧縮データを再変換する際にもデコードせずに行われるため、圧縮時の誤差が累積しない。そして、そうしたデコードまたはエンコードが、単一のコーデックで行われ、さらに、様々な色表現に対応している。つまり、デジタル画像データとして媒体変換が行われた後、その原本となるデータが持つ本来の情報を維持し、その情報を失わずに利用可能なことが重視されたのである。当館では、マイクロフィルムやカラーポジフィルムからデジタル化を行い、その一次デジタル化データを可逆圧縮のデータとして保管している。またJPEG2000の仕様は、国際標準規格として公開され、またリファレンス・ライブラリも提供されている。従って、特定の組織や技術、製品の存続に依存せず、長期にわたって、デジタルデータの利活用性を維持出来ることになる。こうして、国立公文書館に求められている過去から未来、紙から電子媒体に至る多様な資料の長期にわたる利用と保存を実現していくための基盤技術の一つとして、JPEG2000に注目したのである。

2 画像データの作成と配信機能の概要

それでは、実際に「国立公文書館デジタルアーカイブ」で画像データの作成、提供について紹介する。まず、公文書などの資料のデジタル画像データの作成であるが、

³ ISO/IEC 15444-3:2002, Information Technology – JPEG 2000 Image Coding System – Part 3: Motion JPEG 2000.

当館では、所蔵資料のマイクロフィルムへの媒体変換を行っており、それを元にデジタル化を行っている。公文書などの16mmマイクロフィルムから、8ビット、グレースケールのA3再現サイズで可逆圧縮のJPEG2000 (part 1) を作成し、このデジタル画像データを保管用データとして、さらにデータのトランスコード或いはエンコードの際に元データとして利用している⁴。実際に画像配信システムに搭載する画像データは、この元データから作成した非可逆のJPEG2000 (part 1) のデータである。また、同じ元データから、地図や写真と文字が混在する文書などの圧縮に適した拡張規格であるJPEG2000 (part 6) を用いてPDF形式へ変換したデータも合わせて提供している⁵。前者では、マイクロフィルムの1コマが1ファイルとして作成されており、1ページ毎のデータ或いはクライアント端末から要求があった表示部分のみのデータが配信される。これに対し後者では、簿冊中の一件毎に1ファイルのデータとしてダウンロードされ、クライアント端末に保存、表示される仕組みとなっている。こうした特徴が示唆するように、JPEG2000はブロードバンド環境を前提とした仕組みであることから、当然、そうした条件での閲覧に向いており、PDF形式のデータはダウンロードして保管する場合や印刷などローカル環境でのデータ利用に向いていると言えるだろう。また後者では、画像データ内部にレイヤーを持つことが可能であり、現在、試験的に、背景画像データ層と資料画像部分のデータ層でレイヤーを構成しており、それぞれグレイとモノクロで選択的に表現できる。

次に、貴重な歴史資料のデジタル化について紹介する。当館では、こうした資料を4×5インチのカラーポジフィルムに撮影し保管している。マイクロフィルムの場合と異なり、国絵図のような大判の資料など、その一部は分割して撮影しているので、デジタル化を行う際には接合する必要がある。資料によっては、20カットを超える場合もあり、合成された画像データは、最大で約9GB、画素数も30億画素を超える場合がある。こうした画像データは、圧縮した上でシステムに搭載するが、それでも1ファイルが数百MBというサイズになる。従来の画像データ配信・閲覧の仕組みでは、こうしたファイルを予め細分化し、ビューアーで表示するファイルのみを配信するなどしていた。JPEG2000では、ビューアー側の要求に従って、表示に必要なデータを画像データ内部の論理的に分割されたタイル単位で配信している。つまり、閲覧者が画像表示の移動や拡大・縮小を行うと、そのために必要なデータのみをサーバーから

⁴ ISO/IEC 15444-1: 2004, Information Technology - JPEG 2000 Image Coding System - Part 1: Core Coding System.

⁵ ISO/IEC 15444-6: 2003, Information Technology - JPEG 2000 Image Coding System - Part 6: Compound Image File Format. なお、PDFは、Adobe社が提供している文書データの取扱いに適したフォーマットで、JPEG2000に対応している。

持って来るといふ仕組みである。以上からも明らかなように、JPEG2000の画像データ配信・閲覧の仕組みの基本的なコンセプトは、従来のそれに類似していると言えるが、そうした仕組みを1つのファイルという単位で実現し、さらに一つの国際標準として総合化されている点に従来との決定的な違いがある。

さて、JPEG2000の特徴の一つとして、メタデータなど画像以外のデータも合わせて格納できるという仕組みが標準化されていることにふれておく。こうしたメタデータを格納する機能により、デジタル画像自身のアイデンティフィケーション(Identification)が可能となると同時に、OCRや検索システムとの連携など、デジタル画像データ上のテキスト情報の利用という将来への可能性をもたらしめている。当館では、こうした特徴を活かした将来へ向けた試みとして、画像データ中に当該資料のメタデータの一部をXML形式で格納しているが、更なる利用方法については今後のテーマの一つとなっている。

以上、「国立公文書館デジタルアーカイブ」の画像データ、中でもJPEG2000を中心に紹介したが、当館デジタルアーカイブは、従来からの画像フォーマットであるJPEG形式でも提供している。従来、JPEGを用いた画像配信システムは、ファイルを予めサーバーに登載し、これをクライアント端末側にダウンロードする仕組みであった。しかし、当館のシステムでは、配信サーバー上で、JPEG2000の画像データ配信用ファイルからJPEG画像へ動的に変換し配信している。これにより、配信用のそれぞれのJPEGのファイルを予め作成するコストを全く必要ないものとし、かつデータ量が小さいためナローバンドの環境などでも効率的にデータを配信することが可能となった。この画期的なシステムにより、JPEG2000のデータを利用するには適さないケースにも対応が可能となり、将来的にはモバイル利用も視野に入ってきたと言えるだろう。

終わりに

以上のように「国立公文書館デジタルアーカイブ」は、あらゆる環境、あらゆる利用者、あらゆる利用方法を念頭においた「トータルサービス」である。そして、そうしたサービス基盤に載っている「コンテンツ」であるが、現在、提供しているデジタル画像データは、昨年度、当館で作成したJPEG2000データが約12万画像（コマ）とアジア歴史資料センター提供のDjVu形式の当館所蔵資料へのリンクを合わせて約182万画像である。当館では、歴史を跡付ける基本的な公文書や戦後の国家制度の変革に関係する資料などから順にデジタル化を開始した。その中には、憲法・法律・条約等の公布原本であり、天皇の御名・御璽がある「御署名原本」や、法律・勅令・政令・閣議決定等の閣議案件の原本を収録している「公文類聚」、歴代の官房総務課長等の手持ち参考資料として編綴された「内閣総理大臣官房総務課資料」などが含まれており、「日本国憲法」原本のデジタル画像はもとより、その制定に関係する資料や、いわゆる想定問答集など、多彩な資料がデジタル化されている。また、農地改革・財閥解体など、学校の教室で学んだ戦後改革の重要法令等の閣議書類が含まれている。また、カラーのデジタル画像で提供している画像データは、現在のところ、223画像（346カット）である。先述の現行憲法の他、「大日本帝国憲法」や「終戦の詔書」などの原本画像が提供されており、また徳川幕府が諸藩に命じ作成させた「国絵図」（国指定重要文化財）は、その天保年間作成の資料について、全国分を公開している。その他にも色彩豊かな歴史資料や物理的な制約から容易にご覧頂けない資料などを中心にデジタル化しており、さらに直感的に資料をご覧頂ける仕組みを用意し、印刷の際には資料の解説も添付されるなど、様々な工夫を施している。

当館では、今後も計画的に所蔵資料のデジタル化を進め、インターネットを通じてご覧頂ける資料画像の増加、充実に努める。資料をデジタル化するだけでも大変な労力と知識が求められるが、単にデジタル化するだけではなく、情報そのもののあり方、国が保存・提供するデジタル情報のあり方、インターネット・コンテンツの利用方法など、多様かつ深遠なテーマを念頭におきながら、将来を見据えた施策をとることも重視している。現に、電子文書の移管のように近々の課題も見えてきており、今後、多様なデジタルデータの保存と利用に関する対応が必要なことが明らかである。「国立公文書館デジタルアーカイブ」が、マルチメディア・デジタルアーカイブとして、サービスを開始することになる日もそう遠くはないかもしれない。当館では、更なるサービスの向上はもちろんのこと、デジタル化対応の基盤となる知識とノウハウを蓄積しつつ、将来を見据えて所蔵資料のデジタル画像データの提供を行っていきたい。