

仏教学のためのデジタル学術編集システムの構築に向けたモデルの提案と実装

永崎 研宣^{1,a)} 大向 一輝² 下田 正弘²

受付日 2021年5月18日, 採録日 2021年11月2日

概要: 仏教学のための知識基盤の構築にあたり, テキストの共有は重要な課題である. これを効率的・効果的に実現することを目指すデジタル学術編集システムにおいて, 近年普及しつつある国際的な画像共有の枠組みである IIIF (International Image Interoperability Framework) は実装と運用の側面を改善する役割を果たしうる. 筆者らは, 仏教学のためのデジタル学術編集システムのモデルを設計し, そのモデルにおいて実装・運用面を改善する手段として IIIF を位置づけ, 仏典全文テキストデータベース SAT2018 においてそれに基づく実装を行った. これを通じて, デジタル学術編集システムの運用におけるスケーラビリティと持続可能性に関する課題を明らかにした.

キーワード: IIIF, 大蔵経, TEI, 学術編集版, SAT DB

A Model and an Implementation of Digital Scholarly Editing System for Buddhist Studies

KIYONORI NAGASAKI^{1,a)} IKKI OHMUKAI² MASAHIRO SHIMODA²

Received: May 18, 2021, Accepted: November 2, 2021

Abstract: Text sharing is an important issue in the construction of a knowledge base for Buddhist studies. The International Image Interoperability Framework (IIIF), which is an international framework for image sharing, can play an eminent role in improving the implementation and operational aspects of digital scholarly editing systems that aim to achieve this efficiently and effectively. The authors designed a model of a digital scholarly editing system for Buddhist studies, positioned IIIF as a means to improve the implementation and operational aspects of the model, and implemented it based on the model in the full-text Buddhist text database SAT2018. Through this study, we clarified the issues of scalability and sustainability in the operation of digital scholarly editing systems.

Keywords: IIIF, Daizokyo, TEI, scholarly edition, SAT DB

1. はじめに

筆者らは, デジタル時代の到来を受けた仏教学のための新たな知識基盤のモデル構築に向けた研究および実践を行っている. その営みにおいて中核的な役割を担うのがテキストを適切に共有するためのシステムであり, それ

を学術的に信頼できかつ利便性が高いものにしていくこと, そしてそれを可能な限り低コストで運用・改善していくことは喫緊の課題である. 本稿では, このテキスト共有システムを, 仏教学のためのデジタル学術編集システム (Digital Scholarly Editing System for Buddhist Studies 以下, DSES-BS. なお, デジタル学術編集システム一般を指す場合には DSES と表記する) と位置づけ, そのモデルを提案するとともに, 実装にあたっての現段階での成果と課

¹ 一般財団法人人文情報学研究所
International Institute for Digital Humanities, Bunkyo,
Tokyo 113-0033, Japan

² 東京大学大学院人文社会系研究科
The University of Tokyo, Bunkyo, Tokyo 113-0033, Japan

a) nagasaki@dhii.jp

題について、筆者らの実践をふまえて論ずる*1。

2. デジタル学術編集システムについて

DSES は、人文学の基盤を構成する要素としてこれまで世界中で多くの取り組みがなされてきた。ここでは、Peter Schillingsburg の議論 [1] をふまえつつ、その現況について概観してみよう。

とりわけテキスト研究においては、誰もが共通して依拠しうる資料を構築し共有することは研究の場を形成するうえで欠かせないものである。1つの作品があったとき、近代であれば、草稿があり、編集が加わり、校正が行われ、最終的に出版された後にもさらに修正が加わることがある。あるいは、中世以前の資料であれば、作者自身の手で書かれた資料は残っていないことが多く、各地で書写され残された写本や木版・活版の印刷本、石刻資料等に基づいてテキスト資料を再構成することになる。聖書や仏典、源氏物語等、このような資料は枚挙に暇がない。いずれにおいても、テキスト資料が1つの資料に依拠するだけで成立するわけではなく、多様な資料を収集し、対比するなかで、テキスト資料が構築されていくことになる。紙媒体の時代であれば、主に物理的な制約により、何らかのルールに基づいて1つの「本文」を決定し、それに枝をつけるような形で他の資料における異同を注記していくのが一般的であった。Netsle-Arland 版新約聖書 [2]、大正新脩大藏經（以下、大正蔵）[3]、源氏物語大成 [4] をはじめ、この種の多くのテキストはその方法を踏襲し、研究者はその種の資料を通じて研究対象としての作品を扱うという構図になっていた。なかには、本文を再構築するためのルールを分野として共有しようとする場合もあり、仏教学の一部分野においてもその手法が採られていた [5]。そして、現代語に翻訳される場合には、そのようにして構築された「本文」が翻訳されるのが通例であった。校訂テキスト・校本等、いくつかの呼称があるが、1つの作品は、そのようにして広く深く研究されつつ、その成果の精髓としてのテキストが人口に膾炙してきたとみることができる。

近年、デジタル媒体がコモディティ化する過程で、この種のテキストをよりニュートラルに扱おうとする動きが出てくる。これまでは紙資料において「本文」と「脚注における異同」として構成されレイアウトされていたものを、分量やレイアウト方法に制約が少ないデジタル媒体上では、対照資料それぞれを等価な形で扱えるようにしたうえで、それぞれの異同を動的な仕方で参照できるようにしようというのである。この動きは、「正しい本文」を追求することでテキストにおける作成プロセスや時間/地域の

多様性をともすれば見逃してしまいがちであったことへの反省と軌を一にしている面があるように思われ、校訂テキスト (Critical Text, Critical Edition) といった呼称よりも学術編集版 (Scholarly Edition) というテキストのバリエーションに対してニュートラルな言葉が使われるようになってきているのはその点を反映しているとみてよいだろう。デジタル媒体で学術編集版を実現するにあたっては、国内外を問わずすでに多くの試みがなされてきているが [6]、その中でも再利用性が高く国際的に普及している記述手法として Text Encoding Initiative ガイドライン (以下、TEI ガイドライン) [7] における対照資料間の異同を記述する手法は第 12 章 Critical Apparatus [8] に提示されている。ここでは、3種類の手法があげられており、そのうちの2つ、parallel segmentation method と double-end-point-attached method では、異同情報とテキスト中の位置情報を用いて各対照資料のテキスト全文を自動的に再構成することが可能である。これは、比較的ニュートラルに各資料を扱おうとする傾向に対応するものであり、実際に、Versioning Machine [9] や Critical Apparatus Toolbox [10] では、それが実現されており、より重量級のツールとしては Edition Visualization Technology [11] も西洋中世写本研究を中心に用いられている。筆者らは、double-end-point-attached method に沿ったマークアップを実現すべくシステム開発を行ってきた [12], [13]。

3. デジタル学術編集システムにおける画像

一方、この校訂テキストから学術編集版に至る流れにおいては、活字、あるいは符号化された文字を用いる（あるいは、置き換える）ことで利用者の便を図ってきたものの、一方で「資料では実際にはどう書かれているか」という確認がこれまでは困難な場合が少なくなかった。高額な影印本（資料を撮影したものをそのまま印刷した本）が刊行されていれば、あとは費用と置き場所・閲覧しやすさを工夫することでなんとか対応できるものの、そうでない資料は現地で閲覧しなければならず、閲覧許可を得たり交通費を確保したりすることをはじめとして膨大な手間を前提とせざるをえない面があった。あるいは、デジタル画像が公開されている場合でも、それぞれの Web サイトにアクセスして資料を探して当該箇所を閲覧するという手間をかけなければならなかった。筆者らは、2007年頃にこれを解決するためのシステムを構築したものの [14]、各 Web サイトでの画像公開手法の違いをはじめとする多くの課題を前に、このシステムを本格的に基盤として稼働させることはできなかった。そのような状況下、2015年頃から世界的に急速に広まりつつある国際的な Web 画像共有の枠組みである IIIF (International Image Interoperability Framework) [15] は、この課題を実装レベルで解決するための有力な手段となりつつある。

*1 本稿の一部は、以下の研究報告を発展させたものである。
永崎研宣、下田正弘：国際的な画像共有に基づくデジタル学術編集版の構築 SAT 大藏經テキストデータベースの事例を通じて、情報処理学会研究報告人文学とコンピュータ (CH), 2019-CH-120(4), pp.1-6 (2019)。

解釈 (interpretation) が施された注釈書 (Commentary) が作成され、それぞれが翻訳されることもあり、やがて木版で刊行されることもある。それらを批判的に編集 (critically editing) したものとして現代の学術編集版が作成されているのである。

DSES-BS ではこのような関係を研究者が効率的に扱えるようにすることが目標となる。

4.2 関連する先行モデル

DSES-BS が対象とする仏典テキストのこのような様態をふまえると、そのモデルもまた様々な観点から構築することができる。紙媒体を前提としてサンスクリットテキストの校訂において重要な役割を担う Steinkellner [5] が提示したモデルはデジタルにおいても有効であり、DSES-BS においても重要な一部をなす。また、Wittern が提示した 8 次元のテキスト・モデル [22] は仏典だけでなく東洋学全般を対象とし時空間までも含む包括的かつ概念的なものであり、実装面において異なる点はあるものの、モデルとしては DSES はほぼ互換性を持ち、それを現在の技術的進歩をふまえてより詳細化したものという位置づけになる。

また、これらの資料やそこに含まれる実体等の関係を記述するための語彙体系、Buddhist Digital Ontology vocabulary [23] を BDRC (Buddhist Digital Resource Center) が開発・公開しており、これはチベット大蔵経を対象として開発されたものの、あらゆる仏典に適用可能とすることを目指しており、この語彙体系との接続は今後の重要な課題の 1 つである。

4.3 DSES-BS のためのデータに関する近年の動向

DSES-BS におけるテキストデータについては、筆者らが取り組む SAT 大蔵経テキストデータベース研究会 (以下、SAT 研究会) により大正蔵の約 1 億字のテキストデータを構築済みである。文字に関しては、UCS (Universal Coded Character Set) 未符号化文字の扱いが課題となっている。これにはいくつかの解決策があり、TEI ガイドラインにおける外字モジュールの活用と、UCS 符号化の提案、という 2 つの選択肢が有力である。筆者らは、後者を行うことによってこれを解決することを目指し、すでに 3,000 字種以上の UCS 符号化を行ってきた [24] が、その結果、潜在していた課題に直面することになり、現在は前者を組み合わせ解決方法を検討している。これについては別稿を期したい。

より高度な DSES のためには TEI ガイドラインに準拠することが有用である。TEI ガイドラインの具体的な適用手法を仏教学資料に特化して定める必要はあるものの、それについては検討を進めているところである [25]。

仏典画像のデータについては、SAT 研究会では 30 万点以上の IIF 対応仏典デジタル画像を作成・公開しており、

さらに、IIF の普及とともに世界各地の研究図書館等から Web 公開される仏典画像が急速に増加している。国立国会図書館、国文学研究資料館はもとより、東京大学や京都大学、ハーバード大学の図書館、そして、米国議会図書館、フランス国立図書館、バイエルン州立図書館等、各地の機関が Web 公開するデジタルコレクションに仏典画像が少なからず含まれるようになってきている。しかも、その数は着々と増えてきており、すでに公開された画像に対応するだけでなく、今後増えていくものに対しても対応できるような仕組みが必要となっている。

5. DSES-BS のモデルの提案

ここまでみてきた状況に対応可能なシステムとして DSES-BS を構築するにあたり、筆者らはそのモデルとして「仏教学典籍モデル」を提案する。仏教学典籍モデルは、典籍単位のモデルである「単典籍モデル」と、単典籍モデルどうしが相互に関連するモデルである「典籍関係モデル」で構成される。それぞれについて以下に説明する。

5.1 単典籍モデル

まず、単典籍モデル (図 2) については、IFLA Library Reference Model (LRM) [26] を参照しつつ、以下のように位置づける。なお、ここでの LRM との対応は、本モデルを説明するための解釈の一例であり、別の対応の仕方もありうることに留意されたい。

各典籍 (work) は、サンスクリット語をはじめとする各種インド語だけでなく漢文やチベット語、西夏語、モンゴル語等、様々な言語に翻訳され、あるいはそういった言語で著され、さらに現代語に翻訳される (expression)。それは書写の対象となり、あるいは木版として開版されるなどして (manifestation)、さらにそれが様々な個別の資料 (item) として我々の目にするところとなる。1 つの典籍に関する item は多くの場合複数存在することになる。DSES-BS においては、この item それぞれの画像データが存在することが望ましく、それを証拠資料としてテキストデータが作成される。同じ言語で書かれた item 群、すなわち、expression を同じくする item 群については、item

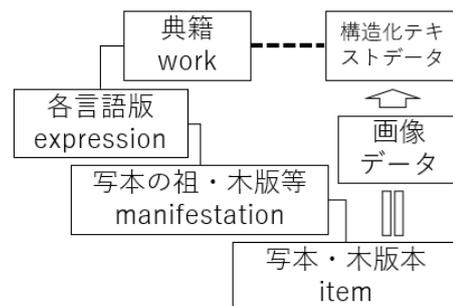


図 2 単典籍モデル (典籍単位でのモデル)
Fig. 2 A model of a work.

の数だけ、それに応じたテキストどうしの異同の記述が行われることになる。言語が異なる版については、各言語版どうしの対応関係が記述される。この、itemの複数性をCritical apparatus等の記述方法で可能な限り如実に反映したテキストとしての構造化テキストデータが1つの典籍に対応し、画像データも含めて1つの典籍を表すモデルということになる。

5.2 典籍間の関係のモデル

次に、単典籍モデルで表現される典籍どうしの関係についてのモデル(典籍関係モデル, 図3)を検討する。図1では典籍どうしの関係としては注釈の例のみをあげているが、それ以外にも、引用や参照といった関係がある。これらは、単に論を補強するために行われることもあるが、必ずしも明示的ではないにせよ、先行する典籍の影響を受けた痕跡として観察できるものもある。一方、先行する論を批判することも特に仏教論書においては広く行われており、これも関係として記述できるだろう。

これらの関係は、典籍単位で行われることは少なく、むしろ、1つの典籍において様々な典籍が批判・注釈・引用・参照される事例がよく観察される。したがって、このモデルにおける関係は、複層的なものとなる。

こういった関係を持つ個々の典籍が、どのような時空間に関連付けられるか、ということも重要な課題である。中国・チベットの典籍は比較的時空間を明確にしやすい傾向にあるが、インドの典籍はこれが明確でないことが多く、むしろ典籍間の関係が相対年代を決定することになる場合もある。

典籍間の関係は、関係を見出した研究者の解釈に依拠することも少なくない。したがって、この関係は、DSES-BSにおいては、一意に決定すべきものではなく、むしろ、任意の解釈であることがつねに明示されている必要がある。時間軸についても、上述のように、必ずしも明確なものではなく、解釈に依存する場合があるため、これもDSES-BSでは複数解釈を共存させられる仕組みが提供される必要がある。

典籍関係モデルがこのようにして典籍を連結していくことによって、DSES-BSにおいて典籍の全体像が形成され

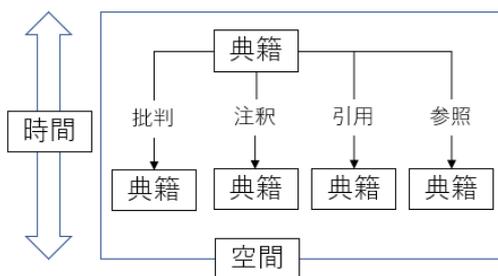


図3 典籍関係モデル(典籍間の関係のモデル)

Fig. 3 A model of relationship of works.

ることになる。

6. 仏教典籍モデルの実装上の課題

6.1 初期段階の実装と課題

このような検討をふまえ、筆者らは、SAT研究会が開発・公開する仏典の全文テキストデータベースであるSAT大蔵経テキストデータベース(以下、SAT DB)[27]に仏教典籍モデルを実装することを目指した。

2015年に公開した、itemに対応する世界各地の仏典デジタル画像をテキストデータベースと仏典単位でリンクする機能[28]は単典籍モデルの部分的な実装であるといえる。そして、2018年には増加し続ける各地のIIF対応仏典画像に対して仏典の巻単位でのリンクを継続的かつ効率的に実現するシステムIIF-BS(IIF Manifests for Buddhist Studies)を開発・公開するとともに、SAT DB2018の初期版において、IIF-BSのAPIを活用して典籍画像群を典籍単位でリンクするサービスを提供した[29]。なお、IIF-BSにおいて集約された仏典画像は2021年5月16日時点で8,147件であり、その多くがSAT DBから1クリックで巻単位で閲覧できることとなった。

しかしながら、IIF-BSまではいずれも長ければ500行を超える「巻」の単位でのリンクが基本であったため、閲覧者は画像との対応を確認する際にその数百行にわたる仏典画像の中から自分が確認したい箇所を探さねばならず、精通している典籍以外については該当箇所を探すのに少々時間がかかってしまうことが研究者による利用に際しての大きな課題となっていた。

6.2 課題解決のためのディープリンクの実装

上記の課題について、IIF対応画像に関しては、1つの資料におけるCanvasの任意の箇所に対するAnnotationという形で、より細かい粒度でのリンクを設定することがAPIの仕様上は可能である。そこで、テキストデータの任意の箇所とWeb公開画像の任意の領域を任意の単位でリンクする仕組み(これを本稿では「ディープリンク」と呼ぶ)を開発することで、単典籍モデルの実効性を高めることでDSES-BSとしての利便性をより向上させることを目指した。そして、このディープリンク機能をSAT DBの2018年版(SAT2018)に対して2018年12月に追加実装した。以下、これについて説明する。

6.3 テキストと画像のリンクとその要件

単典籍モデルにおける画像データとテキストデータの間を、ディープリンクをふまえて適切に実装するために必要となる、両者の関連付けに関わる要件(ここでは「リンク要件」と呼ぶ)については、現時点では以下の項目が考えられる。

要件1. 全文テキストデータと対応する対照資料画像との

間の容易な相互参照

要件 2. 対照資料どうしの異同箇所等についての注記と対応する資料画像の簡単な一覧

要件 3. データ作成者によるテキスト・画像間の容易な関連付け注記

要件 4. 関連付け注記の適切な分類

要件 5. 閲覧者によるデータの取捨選択

要件 1. に関する有用性については上で述べたとおりである。参照された資料や関連する資料においてテキストがどのように書かれているかを確認するには多大な手間と時間を必要としてきたものが、1画面上でクリックするだけでかなりの程度できることになり、対象となる典籍に関心を持つすべての研究者にとってこの機能は有益である。

要件 2. の有用性は、資料どうしの記述や字形が異なっているかどうかの判断を他者に委ねる形になるため、文献研究におけるステークホルダの在り方 [13] として論じたように、対象典籍そのものを研究対象とする研究者には必ずしも有用ではない場合があるが、一方、対象典籍を直接には専門としないが対象典籍を子細に検討する必要がある場合があるタイプの研究者にとっては、元資料の参照すべき箇所を容易に参照し検証できるようになるため、この機能は研究を効率化するものとなる。

要件 3. については、前 2 者が利用者から見た DSES-BS の要件であるのに対して、この要件は、作成者、あるいは提供者にとっての学術編集版を実現するためのものである。多くの研究者にとって、典籍への関わり方には、個別の写本・版本等の証拠資料まで参照する程度に深く研究しているものから、当該典籍の専門家の責任において提示されたテキストをただ参照するといったグラデーションがあり、前者については作成者・提供者であると同時に後者については利用者になるという互酬性が成り立ちうるものである。

要件 4. については、データと画像の関連付けは、そのリンクの性質に応じて分類しておく必要がある。まず、同じ典籍として伝承される複数資料の間での異同である。たとえば、4 点の写本・木版本等の資料があったとして、テキスト中のある箇所について 2 点は「涅槃」、他の 2 点は「泥洹」と記述されていたとする。両者は意味としては同じだが、記述としては異なっている。そして、たとえばこの箇所について、SAT DB のテキストデータにおいて「涅槃」となっている場合には、あくまでも便宜的にだが、「涅槃」と書かれている 2 点の資料は「同一」であり「泥洹」と書かれている 2 点の資料は「異読 (variant reading, 異文と呼ぶこともある)」と分類されることになる。あるいは、別の資料で、この箇所が欠けている場合は「不足」「削除」といった記述をすることになるだろう。そして、逆に、「涅槃」にさらに何か文字が加えられている場合には、「追加」「増広」ということになるだろう。ただしこれはあくま

でも大正蔵から見た場合の話であり、そのようにリンクした後、表示に際しては、「同一」と「異読」等の関係を逆に表示したり、単に並列に表示したりすることも技術的には問題なく可能である。

要件 5. については、研究者によって異同や関連付けについての判断が異なる場合がある。たとえばテキスト中のある箇所の各対照資料間の異同について、文字の判定基準をどうするかによって異同の判定が変わってくる。各関連付けデータに個々の作成者名が付与されるようにしておくことで、どの作成者の判断に依拠するかを利用者が取捨選択できる機能を付与することが可能である。システム提供者側による公式な関連付けデータ、利用者兼作成者の個人の判断によるデータ、さらには、作成者グループの判断によるデータ、という形での分類も用意しておくことで、プラットフォームとしてはシステム提供者側のデータに依拠しつつ、必要に応じて利用者兼作成者や作成者グループによるデータを取捨選択して利用することができるようになる。ただし、研究論文等から参照する場合には、「どのデータを利用したか」ということを容易に提示し共有できるような仕組みも必要になるだろう。

6.4 ディープリンクを構成するデータ

上記の要件を満たしたディープリンクを構成するために必要なデータ (以下、これをディープリンクデータと呼ぶ) について検討する。前述のように、SAT DB が依拠する大正蔵は、仏教学分野においては漢文仏典の標準テキストとされていることから、その巻・頁・行の番号 (以下、大正蔵行番号) を指定することでテキストを参照できることが国際的なデファクト標準になっている。SAT DB や CBETA [30] といった漢文仏典の電子テキストはいずれもこの番号に依拠することで国際的な仏教研究のデジタル基盤を形成してきている。SAT 研究会は大正蔵行番号を利用することで、これまで、この SAT DB における任意の典籍中の任意のテキスト部分 (文章や単語等) の位置情報を文字単位で取得し、そこに別の任意のテキストのデータの任意の箇所をリンクするという機能の実装を行ってきており [13]、それによって漢文・現代日本語訳、漢文・英語訳、漢文・チベット語訳等、複数言語間の文章単位でのテキストデータのリンクデータを構築し、対訳提示等の機能を提供している。今回のディープリンクにおけるテキストデータ側でのリンクの仕方については、それをそのまま継承している。そして、このリンクの片側を IIF 仏典画像にも対応させたのが今回の実装である。ここでの IIF 対応画像側のリンク情報として保持しているのは以下の情報である。なお、ここで、準拠しているのは、IIF Presentation API 2.1.1 [31] および IIF Image API 2.1.1 [32] である。

- a. IIF Manifest URI
- b. IIF Canvas URI

- c. IIIF Image API
- d. Coordinate
- e. 作成者
- f. 作成日付
- g. 状態
- h. 関係
- i. 対象画像中の文字列の翻刻

IIIF API を介して画像上の任意の位置を指定するためには、上記の a.~d. の情報が必要になる。すなわち、資料の存在を示す IIIF Manifest URI, 頁等のまとまりを示し実際のコンテンツを載せる仮想的な共有キャンバスを示す Canvas URI, IIIF Image API に対応する画像の存在を示す URI, これに加えて、特定の画像上の任意の部分の座標情報である。これらの情報があれば、画像上の任意の領域へのリンク、すなわちディープリンクは成立する。ただし、共有キャンバスのサイズと載せる画像のサイズや座標開始位置が異なっている場合には座標情報を画像にあわせて取得するとビューワ上でズレが生じることがある。この場合には両者の値を取得して調整することになる。

以上の情報は、対象資料のデータの形式いかんにかかわらずリンク情報としてはある程度形式化できており大幅な変更は見込まれないことから、本実装ではデータの出入力に関して実績のあるリレーショナルデータベース管理システム PostgreSQL に格納している。

また、データベース格納の際に、d. の座標情報は PostgreSQL の Polygon 型で格納しており、座標情報としての距離や重複等の検索も効率的に可能となっている。この点は、今後の分析等で活かしていきたい。

h. の関係については、前章で述べた検討に基づき、位置合わせを確認するための「(テキストと画像の) 位置合わせ情報」、大正蔵の校訂本文との関係において意味内容が異なる「異読」、同一箇所について他の版本・写本が大正蔵校訂本文と同じものは「校訂本文と同一」、追記されているものは「追加」、不足しているものは「削除」、これに加えて「異体字」を用意している。ただし、異読か異体字かの区別は時折困難であるため、現在の方針としては、大正蔵の脚注であげられているものは異読、明らかに異読であるものも異読としたうえで、それ以外のもので、異体字として興味深いと思われたもののみを異体字として記載している。今後、この区別についてはより丁寧なルールを作っていく必要があるだろう。

i. に関しては、大正蔵校訂本文とは異なる文字・フレーズである場合に、単典籍モデルにおけるテキストデータと画像をリンクする際に必要となる。TEI ガイドラインに準拠したテキストデータの自動生成を行う際に、ここで翻刻テキストを入力しておくことは、異読テキスト、すなわち、TEI ガイドラインでいうところの <app> エレメントにおける <rdg> のテキストを提供できることになる。

7. 実装された仏教典籍モデル

7.1 データの操作とワークフロー

SAT DB2018 でのディープリンクの機能においては、編集用インタフェースを用いて IIIF 対応のデジタル仏典画像と大正蔵テキストデータベースの任意の箇所をリンクすることになる。IIIF 対応画像が SAT DB 外の Web サイトから提供されるものであっても SAT DB 内で提供するものであっても編集方法は同様である。ここでは、前者を例として、データの操作とワークフローについて、順を追って示す。なお、編集用インタフェースは SAT DB2018 に登録された編集者のみが利用可能である。

SAT DB 外で公開されている IIIF 対応画像を用いる場合、あらかじめ IIIF Manifest URI を IIIF-BS に登録する必要がある。この画像と任意のテキストデータをリンクするには、IIIF-BS の Web API を通じて巻単位で対応する IIIF Manifest URI が提供され、SAT DB 2018 の編集用インタフェースでは、この IIIF Manifest URI が指す IIIF Presentation API のデータを読み込み、Canvas URI 等に記載された画像を表示する。同じ画面には、全文検索ソフトウェア Apache Solr に格納された仏典のテキストデータが表示することができる。

そして、編集インタフェースを用いてリンク対象のテキスト・画像の対応箇所、両者の関係情報の選択操作をすることで、テキストデータ側の始点・終点の位置情報、ディープリンクされたテキスト自体、これに加えてディープリンクデータとして前節に示したデータが取得され、サーバ上の PostgreSQL のテーブルに 1 レコードとして格納される。なお、画像の対応箇所のテキストが大正蔵校訂本文と異なる場合には、そのテキストの翻刻を文字入力する必要がある。その場合も同様にサーバ上の PostgreSQL のテーブルに 1 レコードとして格納される。入力されたデータは管理者が確認し、必要に応じて削除・修正を行う。

閲覧時には、編集インタフェースを通じて蓄積されたデータが表示対象となる。SAT DB2018 において任意の仏典を開き、任意の巻の異文や異体字等の表示操作を行うと、その仏典に紐付けられたディープリンクデータが PostgreSQL の当該テーブルから取り出されて整形された後に Web クライアントに送出され、Web クライアント側ではそのデータを Web ブラウザ上に表示する。なお、一連の操作はサーバサイドでは PHP、Web クライアントでは主に JavaScript を用いて実現されている。

7.2 編集用インタフェース

インタフェースはテキストと画像の 2 画面を両側に並べる形で提供している。テキスト上での位置情報については従来どおりであり、画像側での任意の箇所の位置情報記述は、SAT 大正蔵画像 DB [33] におけるアノテーション機能



図 4 対応情報の選択画面
Fig. 4 A function of selecting a relation data.

と同様に、IIIF Image API に対応し拡張性の高い Web 画像ビューワ OpenSeadragon [34] と、画像上の任意の矩形座標を取得するための Selection プラグイン [35] を用いている。そして、IIIF Manifest を読み込むと簡易なページめくりができるようになっており、作業内容の確認のために、対象となる IIIF Manifest URI, Canvas URI, そして Attribution の値、すなわち資料の所属を常時上部に表示するようにしている。さらに、本システム上で簡易に扱えるようにするために資料の略称を付記できるようにした。また、システム構築の簡素化と継承性の確保のためにコンテンツ管理システム Omeka とそのプラグインである Omeka IIIF Toolkit [36] の利用 [37] も検討したが、現時点ではシステム全体への組み込みがやや難しかったため、今回は見送り、独自のインタフェースを作成することとした。現在稼働中のものとしては、図 4 のように、右側のウィンドウでテキストを選択して文字位置を取得し、左側のウィンドウで対応する画像の座標情報を取得し、その関係について選択して保存ボタンをクリックすると、サーバ側にその情報が追記されるようになっている。

7.3 閲覧用インタフェース

閲覧にあたっては、現在のところ、「位置合わせ」「異読」「削除」「増補」「異体字」に対応しており、「削除」「増補」はラベルとしては異なるものの、異読と同じカテゴリで表示されるようになっている。

「位置合わせ」は図 5 のように、テキスト 10 行ごとに表示される IIIF アイコンに「+」のついた画像をクリックするとポップアップがひらき、そこに登録された資料の略称がリストされる。いずれかの略称をクリックすると、左側のウィンドウにその箇所が拡大される形で表示される。デフォルトでは、複数の IIIF 対応画像をそれぞれ拡大縮小可能な状態で並べたり、それぞれにアノテーションを表示したりする機能を持つ学術志向の IIIF 対応画像ビューワで



図 5 OpenSeadragon での位置合わせによる表示
Fig. 5 A representation of alignment on OpenSeadragon.



図 6 本文「而」に対する異読「而爲」の例
Fig. 6 An example of variant readings.

ある Mirador [38] で表示され、画像が順に横に並んでいくようになっているが、IIIF Link というタブを開いているときは、編集用の OpenSeadragon で 1 枚ずつ表示するようになっている。

「位置合わせ」以外に関しては、「variant」ボタンをクリックするとダイアログが表示され、そこにテキストと対応箇所の切り出し画像がリストされる。図 6 のように、テキストをクリックすると全文テキストデータベースの該当箇所が表示され、切り出し画像をクリックすると上述のルールで Mirador か OpenSeadragon 上にその箇所を中心とした画像が表示される。

いずれの場合も、編集者権限を持っている場合には、閲覧画面からそのまま編集作業に入ることが可能であり、知識を提供する側と享受する側がシームレスな双方向環境となっている。

表示のために必要なデータを Web インタフェース側で取得する際には、Web API として取得できる形にしており、一連のデータは大正蔵のテキスト番号や行番号、あるいは関係の種類等を指定すればそれに対応するものを JSON 形

式で取得できるようになっている。したがって、これを外部から利用して別のサービスを展開することも可能となっている。

8. 仏教典籍モデルの実装の評価

今回の実装では、DSES-BSの「単典籍モデル」における画像データとテキストデータの関係性を記述することができた。そして、IIIF-BSを介してリンクしたIIIF対応仏典画像を、さらに、IIIFがもたらすディープリンクの機能を利用することで、各機関のそれぞれの画像に対する統一的なインタフェースを提供し、それを通じて、これまで各対照資料の閲覧に要していた時間を大幅に短縮するとともに、各資料を容易に対比することを可能にした。このことは、文献研究における再検証の困難さを大きく改善しているという点で、IIIFを用いたDSES-BSの有用性を示すことになっている。そして、構築以降、1,280件の仏典画像が増加し、それに対応するリンクを継続的に実施できていることは、モデルの運用面も着実に機能しているといえる。また、IIIFを採用したことにより、ディープリンクによって指定した部分画像から頁全体・資料全体へとたどっていくことが容易であるため、部分的に切り出された断片を参照することにより視野狭窄的な判断に陥りやすくなることを防げるというのもメリットの1つだろう。なお、ここでは「リンク要件」の1~4までが公開されており、要件5については機能としては実装したものの未公開である。

9. おわりに

2章の最後に触れたように、DSESにおいては、何ができないか、ということをも可能な限り明示しておくことが、そのすべてを示すことは困難であるにせよ、できることを示すことと同様に重要である。そこで、DSES-BSにおける現時点での限界と課題について述べておきたい。

IIIFが持つ技術的に期待しうる可能性として、DSES-BSでは永続的なURIを前提としたディープリンクを構築する仕組みを運用している。しかしながら、実際のところ、IIIF-BSの公開以降、専門家コミュニティとして世界各地のIIIF仏典画像を扱ってきているなかで、すでにこのそれほど長くない期間に、主に国内の文化機関において、IIIF Manifest URIの変更やIIIF Manifestの内容変更、画像の差し替えといった事態に直面してきている。特に、IIIF Manifest URIの変更は可能な限り避けるべきだが、その内容、すなわち、Canvas ID等に関しても、外部からのディープリンクを維持して研究情報システムの持続性を担保しようとするなら、すでに公開された情報は可能な限り変更されるべきではない。画像の差し替えについても、たとえそれによって高精細画像になるとしても、座標情報が無意味になってしまいデータの有用性が著しく低下するため、機械的に修正可能でない限り、避けるべきである。その種

の変更に対しては、把握できたものについてはIIIF-BSやDSES-BS上での修正対応を行っているが、数が増えるとやがて困難になるだろう。また、画像の差し替えや消失等に対応するためには画像そのものをどこかに保存しておく必要もあるだろう。IIIFは画像公開を各組織がそれぞれ担う分散型の仕組みを前提としているため、それによるメリットは大きいものの、それゆえのデメリットもこのようにして現れる。技術的解決策としては、URLのホスト名や基礎部分に変更になったとしてもその箇所さえ修正すれば対応できるようにシステム内では相対リンクの利用を徹底し、さらには独自の名前解決手法を導入する方策や、あるいは、たとえばIPFS[39]をはじめ、URLに依存しないアドレッシングの手法を採用する選択肢もありうる。しかしながら、技術的可能性があったとしても、再配布可能なコンテンツでない場合には一次公開者による公開が終わってしまうとアクセスできなくなってしまう。また、上述のような画像の差し替えに対しては効力がない可能性が高い。IPFSは基盤レベルでは少しずつ広まりつつあるが、それがDSES-BSにおける実用レベルでの活用はまだ展開するには、たとえば、IIIFが安定した技術の組合せで構成されており技術的新規性はないものであったにもかかわらず、デジタル画像共有を主導する文化機関のコミュニティによって制定されたことで国際的な支持を集めデファクト標準としての有用性を獲得したように、新たな技術的枠組みが広まるにはコミュニティによる積極的な受容と普及のための体制の構築が必要だろう。こういった点については別稿を期したい。

また、本稿では扱わなかったものの、TEIガイドラインに準拠した「典籍関係モデル」についての取組みも別途進めている[40]、そこでの各種の関係や時空間の記述は、「リンク要件」の5と共通の課題をはらむ。すなわち、研究者によって判定が異なる場合があるという問題である。それゆえ、DSES-BSでは、異なる判定によって作り出された複数の正しいデータを蓄積しておき、状況に応じてそれを選択的に提供できるようにする必要がある。このような状況は人文学資料においてはそれほど珍しいことではなく、記述手法としては個々の関係を判定した者を@resp属性で記述する方法がTEIガイドラインで提供されている。DSES-BSでは、これに準拠することで、利用者がデータを取捨選択する機能を実装しクローズドな運用を開始した。しかしながら、この機能は、広く公開した後に十分な有用性を発揮し続けられるかどうかという問題がある。判定した研究者を基準に選ぶのか、研究者グループ単位で選ぶのか、それとも個別に自身で判定を行うのかといった課題は、判定者(上記のリンク情報においてはe.作成者としている)や判定箇所がそれほど多くない間に対応可能だが、ここでもやはり、対象となる典籍についてどれくらいの専門性を有するかが関係してくることはあるにせよ、判定者

や判定箇所の増加等におけるスケーラビリティの問題は避けて通れない。一定程度の解決は可能であるにせよ、自らの判断に基づいた取舍選択を完全に行うことは不可能だろう。このことは、技術的な問題というよりも運用上の課題という側面が強い。利用傾向等に基づく判断の支援機能等を提供するという事も考えられるが、DSES-BS 提供者側だけでなく、むしろ利用者コミュニティとのコンセンサスを形成していくなかで判断されるべき課題だろう。

DSES-BS のモデルは仏教学における典籍の在り方とその研究の仕方を反映した基盤的なものであり、新たな技術や枠組みが登場することにより実装や運用が効率化される可能性はあるものの、目指すところが変化するわけではない。今後もより適切な実装と運用に向けて、実践を続けていきたい。この取り組みは、日本を含む東・南アジアの文献資料としての仏教典籍の特性をデジタル時代の文献研究に活かすための取り組みであり、同じ地域の様々な他分野の文献研究においても適用可能な側面が少なくない。仏教学だけでなく、他の関連する分野においても、筆者らの実践の成果が広く活用されることを期待したい。

謝辞 本研究は、SAT 研究会に関わる多くの方々の努力によって結実したものであり、深く感謝する。とりわけ、IIIF-BS のデータ構築にあたっては、村瀬友洋氏のご協力に多くを拠ったことを感謝とともに記しておく。本研究は、JSPS 科研費 JP18H03576, JP19H00516, JP19H00526, JP20H05830 の助成を受けたものである。

参考文献

- [1] Schillingsburg, P.: *From Gutenberg to Google: Electronic Representations of Literary Texts*, Cambridge University Press (2006). 明星聖子ほか (訳): ゲーテンベルクからグーグルへ, 慶應義塾大学出版 (2009).
- [2] Graece, N.T.: *Nestle Aland 28th Revised Ed. of the Greek New Testament, Standard Edition*, Amer Bible Society (2012).
- [3] 高楠順次郎 (編): 大正新脩大藏經, 大正一切経刊行会 (1924-1934).
- [4] 池田亀鑑 (編): 源氏物語大成, 中央公論社 (1953).
- [5] Steinkellner, E.: *Methodological Remarks on The Constitution of Sanskrit Texts from The Buddhist Pramāṇa-Tradition*, Wiener Zeitschrift für die Kunde Südasien, Band XXXII, pp.103-129 (1998).
- [6] 安永尚志: 日本古典文学本文データベース形成とデータ記述文法, 情報処理学会研究報告, 1990-CH-008, Vol.1991, No.20, pp.1-8 (1991).
- [7] Ide, N., Sperberg-McQueen, C.M. and Burnard, L.: TEI: それはどこからきたのか。そして、なぜ、今もなおここにあるのか?, デジタル・ヒューマニティーズ, Vol.1, pp.3-28 (2018).
- [8] Burnard, L. and Bauman, S.: *Critical Apparatus, P5: Guidelines for Electronic Text Encoding and Interchange* (2019), available from <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>.
- [9] Schreibman, S. et al.: *Versioning Machine 5.0* (2016), available from <http://v-machine.org/> (accessed 2021-05-16).
- [10] Burghart, M.: The TEI Critical Apparatus Toolbox: Empowering Textual Scholars through Display, Control, and Comparison Features, *Journal of the Text Encoding Initiative*, Issue 10 (online), DOI: 10.4000/jtei.1520 (2016).
- [11] Edition Visualization Technology, available from <http://evt.labcd.unipi.it/> (accessed 2021-05-16).
- [12] 永崎研宣: シラブルを最小単位とする仏教哲学文献データベースについて, 情報処理学会研究報告, 2006-CH-71, pp.33-40 (2006).
- [13] Nagasaki, K., Tomabechi, T. and Shimoda, M.: Towards a Digital Research Environment for Buddhist Studies, *Literary and Linguistic Computing*, Vol.28, No.2, pp.296-300, Oxford University Press (2013).
- [14] 永崎研宣: 人文科学のためのデジタル・アーカイブにおけるステイクホルダー—仏教文献デジタル・アーカイブを手掛かりとして, 人文科学とコンピュータシンポジウム論文集, pp.347-354 (2007).
- [15] International Image Interoperability Framework, available from <https://iiif.io/> (accessed 2021-05-16).
- [16] RDF Working Group: Resource Description Framework (RDF), available from <https://www.w3.org/RDF/>.
- [17] Sanderson, R., Ciccarese, P. and Young, B. (Eds.): *Web Annotation Data Model* (2017), available from <https://www.w3.org/TR/annotation-model/> (accessed 2021-08-22).
- [18] Sanderson, R. and Albritton, B. (Eds.): *Shared Canvas Data Model* (2013), available from <https://iiif.io/model/shared-canvas/1.0/> (accessed 2021-08-22).
- [19] 永崎研宣: IIIF の概要と主要 API バージョン 3.0 の公開, カレントアウェアネス, No.346, CA1989, pp.13-16, DOI: 10.11501/11596735 (2020).
- [20] Witt, J.C.: *Digital Scholarly Editions and API Consuming Applications: Digital Scholarly Editions as Interfaces*, Bleier, R. et al. (Eds.), Vol.12, pp.219-247, BoD (2018).
- [21] Monella, P. and Rosselli, R.D.T.: Extending the DSE: LOD Support and TEI-IIIF Integration in EVT, *Atti del IX Convegno Annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale*, pp.148-155 (2020).
- [22] ウィッテルン・クリスティアン, 中楯はまな: 新しいテキスト・モデルに基づいた東洋学文献研究の支援ツール, 京都大学人文科学研究所, 京都大学人文科学研究所, Vol.2007-CH-074, No.49 (May 2007).
- [23] *Buddhist Digital Ontology vocabulary*, available from <https://github.com/buda-base/owl-schema> (accessed 2021-05-16).
- [24] 下田正弘, 永崎研宣 (編): デジタル学術空間の作り方, 永崎研宣: デジタル学術空間の作り方, 文学通信, pp.88-93 (2019).
- [25] 渡邊要一郎, 永崎研宣, 朴賢珍, 王一凡, 村瀬友洋, 渡邊眞儀, 大向一輝, 下田正弘: 大正新脩大藏經の構造的記述に向けて, じんもんこん 2020 論文集, pp.61-66 (2020).
- [26] Riva, P., Le Boëuf, P. and Žumer, M.: *IFLA Library Reference Model, A Conceptual Model for Bibliographic Information*, IFLA (2017).
- [27] SAT 大藏經テキストデータベース, 入手先 <http://21dzk.l.u-tokyo.ac.jp/SAT/> (参照 2021-05-16).
- [28] 永崎研宣: 「デジタル・アーカイブ」の利活用可能性を高めるために—仏典画像統合検索 API の構築を通じて, 情報処理学会研究報告, Vol.2015-CH-107, No.3, pp.1-4 (2005).
- [29] 永崎研宣, 下田正弘, Muller A. Charles, 蓑輪顕量: 横

断型デジタル学術基盤を目指して—SAT2018 の構築を通じて、情報処理学会研究報告, 2018-CH-117(1), pp.1-7 (2018).

- [30] 中華電子仏典協会, 入手先 (<https://www.cbeta.org/>) (参照 2021-05-16).
- [31] Appleby, M. et al.: IIIF Presentation API 2.1.1 (2017), available from (<https://iiif.io/api/presentation/2.1/>) (accessed 2021-08-22).
- [32] Appleby, M. et al.: IIIF Image API 2.1.1 (2017), available from (<https://iiif.io/api/image/2.1/>) (accessed 2021-08-22).
- [33] 永崎研宣, 津田徹英, 下田正弘: SAT 大正蔵画像 DB をめぐるコラボレーションの可能性, 研究報告人文科学とコンピュータ (CH), Vol.2017-CH-113, No.8, pp.1-4 (2017).
- [34] OpenSeadragon, available from (<https://openseadragon.github.io/>) (accessed 2021-08-22).
- [35] OpenSeadragonSelection, available from (<https://github.com/picturae/openseadragonselection/>) (accessed 2021-08-22).
- [36] IIIF Toolkit, available from (<https://omeka.org/classic/plugins/IiifItems/>) (accessed 2021-05-16).
- [37] 中村 覚: IIIF とオープンデータを活用した『君拾帖』内容検索システムの開発, デジタルアーカイブ学会誌, pp.155-158 (2019).
- [38] Mirador, available from (<https://projectmirador.org/>) (accessed 2021-05-16).
- [39] IPFS Documentation, available from (<https://docs.ipfs.io/>) (accessed 2021-08-22).
- [40] SAT 大藏経テキストデータベース研究会: 聖徳太子御製『勝鬘經義疏』の TEI 版を公開しました (2020), 入手先 (https://21dzk.l.u-tokyo.ac.jp/SAT/sat_tei.html) (参照 2021-08-22).



下田 正弘 (正会員)

東京大学大学院人文社会系研究科教授。1989 年東京大学大学院人文科学研究所博士課程単位取得退学。東京大学博士(文学)。ロンドン大学(SOAS), ウィーン大学客員教授を務める。仏教学, 人文情報学の研究に従事。



永崎 研宣 (正会員)

一般財団法人人文情報学研究所主席研究員。2000 年筑波大学大学院博士課程哲学・思想研究科満期退学。博士(文化交渉学)。東京外国語大学アジア・アフリカ言語文化研究所, 山口県立大学を経て一般財団法人人文情報学研究所の設立に参画し, 現在に至る。人文情報学, 仏教学の研究に従事。



大向 一輝 (正会員)

東京大学大学院人文社会系研究科准教授。2005 年総合研究大学院大学複合科学研究科博士後期課程修了。博士(情報学)。同年国立情報学研究所助手, 2007 年同助教, 2009 年同准教授を経て, 2019 年より現職。人文情報学, ウェブ情報学, 学術コミュニケーションの研究に従事。