

仏教系デジタルアーカイブ の事例報告

SAT大蔵経テキストデータベース研究会の取組み

一般財団法人人文情報学研究所

永崎研宣

SAT大蔵経テキストデータベース 研究会の沿革

- 代表：下田正弘東京大学教授
- 1994年、東京大学文学部 故江島恵教教授により開始。
- 1999年、故江島教授の逝去に伴い下田正弘 同助教授（当時）が代表を継承
- 2007年、テキストデータ完成
- 2008年、Webサービスの公開
- 2012年、Webサービスの大幅改訂
 - 頁画像提示、英語対訳コーパスの提示機能等
- 2015年、Webサービスの大幅改訂、萬暦版大蔵経画像DB公開
 - 日本撰述部の大規模校正、他の画像公開サイトとのディープリンク等
- 2016年、SAT大正蔵図像DB公開、萬暦版大蔵経POD販売開始

主なプロダクト

(ここの文脈でデジタルアーカイブに含まれるもの)

- SAT大蔵経テキストデータベース (SAT DB)
 - 大正末～昭和初期に刊行された「大正新脩大蔵経」85巻分をデジタル化。
 - 1億字超の中国語・日本語等のテキストデータベースと全文検索・絞込検索機能
 - 仏典研究に有益な様々な機能を提供
 - 他のWebコンテンツとの連携を重視
- ITLR (Indo-Tibetan Lexical Resources)
 - ハンブルク大学との共同プロジェクト
 - 世界各地の研究者のコラボレーションによるレクシコンの作成
 - 様々なデータをつなぐ基盤となるデータ
- 萬暦版大蔵経画像データベース
 - 東京大学総合図書館に所蔵されている、明代末に刊行された萬暦版大蔵経を高精細画像として撮影しDeep Zoom形式でCC BYライセンスにより公開。頁画像ごとのパーマリンクを付与。紙版をPODで販売開始。
- SAT大正蔵図像DB
 - 大正新脩大蔵経図像編12巻分をデジタル化・タグ付け。
 - IIIFを採用したCC BY-SAの画像データベース

SAT大蔵経テキストデータベース

<http://21dzk.l.u-tokyo.ac.jp/SAT/satdb2015.php>

[illegible]

ITLR

<http://www.itlr.net/test.php?md=view>



← → ⓘ www.itlr.net/testb.php?md=view

Indo-Tibetan Lexical Resource (Beta version)

Home History Aims People Institutions **ITLR Search** History: 30 [ākāśa\(1\)](#) Terms of Use

Input search word: [Diacritical]
Search option [Open]

3 hits

1. [ākāśa\(1\)](#)
2. [cetanā\(1\)](#)
3. [nivaraga\(1\)](#)

Word/Term/Phrase:
ākāśa (1) [References](#) [URL of this headword](#)

Linked by: Key Information: [vyoman](#)

- **Grammatical Info.**
 1. n.,
[History](#)
- **Taxonomical Labels**
 1. Abhidharma; cosmology; dharma theory; entity;
[History](#)
- **Non-Sanskrit Indic Language**
 1. ākāśa (m. n.)
(Pāli) {[Cone 2001](#): s.v. ākāśa}
[History](#)
- **Tibetan Rendering/s of the Skt Headword**
 - In Tibetan Translations with Available Skt
 - 1. nam mkha' {[Prasannapadā](#): 24.8; [Prasannapadā-Tib](#): P 9a2}
Remarks: The passage is citing MMK 5.1 [History](#)
- **Tibetan Rendering/s of the Skt Headword**
 - In Sources with Bi/Multilingual Attestations
 - 1. nam mkha' {[Mvy.1](#): 7399; [Nobel 1950](#)}
[History](#)
 - 2. go 'byed {[Mvy.1](#): 7399}
[History](#)
- **Modern Rendering/s**
 1. space, ether
(Translation into English) {[Schmithausen 2007](#): p. 95}
[History](#)
 2. ether, air, sky, space, open space
(Translation into English) {[Cone 2001](#): s.v. ākāśa}
[History](#)
 3. Raum
(Translation into German) {[Schmithausen 2002](#): 65}
[History](#)
 4. Luftraum
(Translation into German) {[Nobel 1950](#): s.v. nam mkha'}
[History](#)
- **Key Information**
 1. In the system of the Sarvāstivādin, ākāśa is regarded as an unconditioned
[History](#)
See also: [asamskṛta\(1\)](#)
- **Occurrences in Primary Sources: Indic**
 1. Pāli
 - puna c'aparam, bhikkhave, bhikkhu sabbaso rūpasahñānam samatikkamā paṭighasahñānam atthaṅgamā nānattasahñānam amanasikārā 'ananto

Relationship of entries via bibliographical sources (Experimental!!)

萬曆版大藏經画像データベース

http://dzkings.l.u-tokyo.ac.jp/utlib_kakouzou.php

萬曆版大藏經

万曆版大藏經(嘉興藏)画像データベース(試験公開版)

[東京大学総合図書館](#)所蔵
大藏經研究推進会議・SAT大藏經テキストデータベース研究会
検索は[SAT大藏經DB](#)からお願ひいたします。

本データベースは、東京大学総合図書館所蔵の万曆版大藏經(嘉興藏)を公開することにより、これに関連する多様な学術研究を推進することを目的として、大藏經研究推進会議の支援の下、SAT大藏經テキストデータベース研究会によって作成されています。なお、本データベースでは、オープンサイエンス・オープンデータの国際的な動向に鑑み、大藏經のみならずデジタル文化資料そのものの広汎な活用を促進することを視野に入れ、搭載されているデジタル画像をCC BYライセンスにて公開いたしました。今後は、順次、PDF版等も提供していく予定です。

⇒ 嘉興藏について
⇒ データベースについて

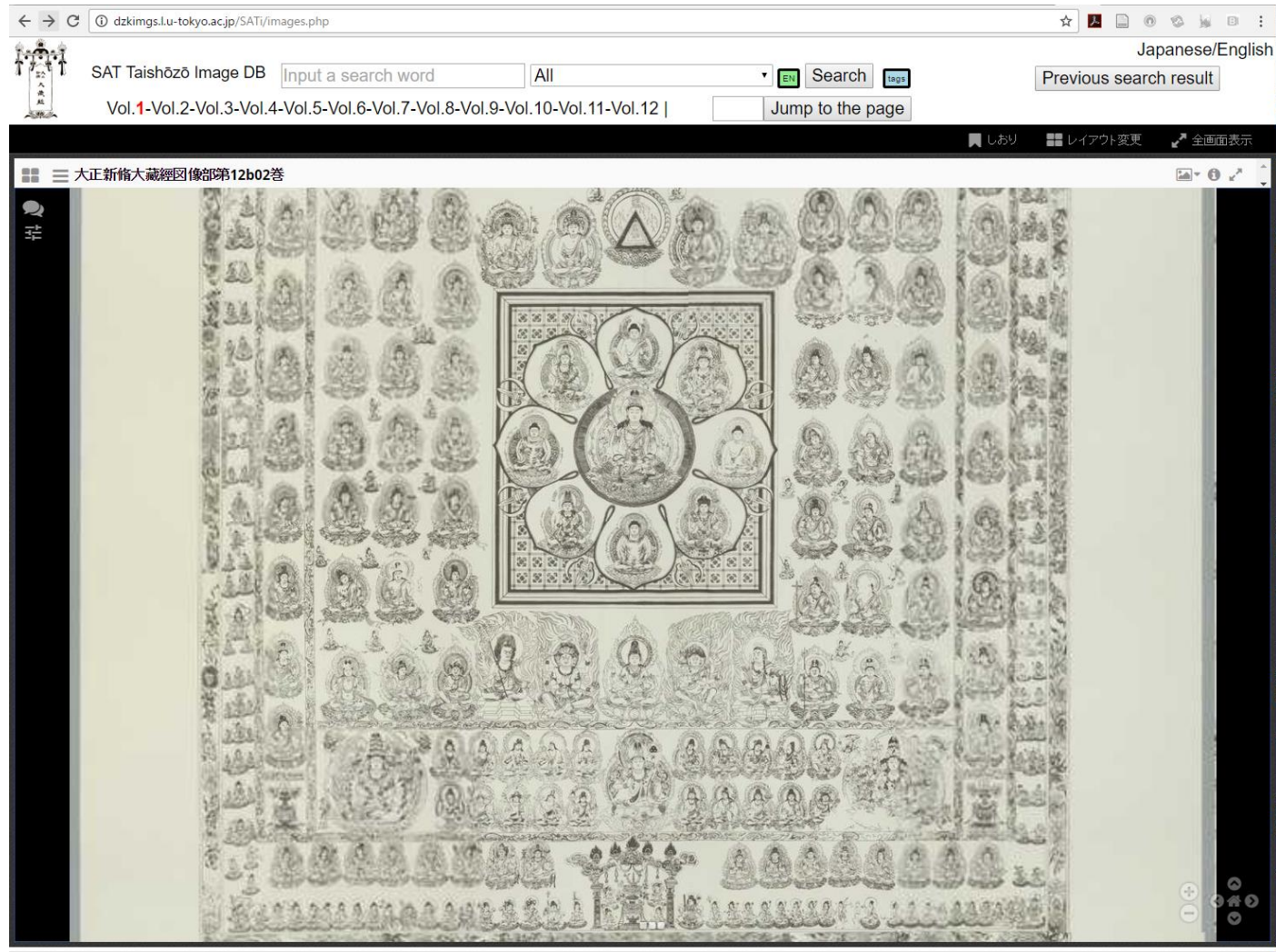
なお、この嘉興藏の特徴や来歴等についての詳細は、横手裕、末木文美土、渡辺麻里子、菊地大樹監修『[東京大学総合図書館所蔵嘉興大藏經：目録と研究](#)』(2010年9月発行)をご参照ください。

嘉興藏

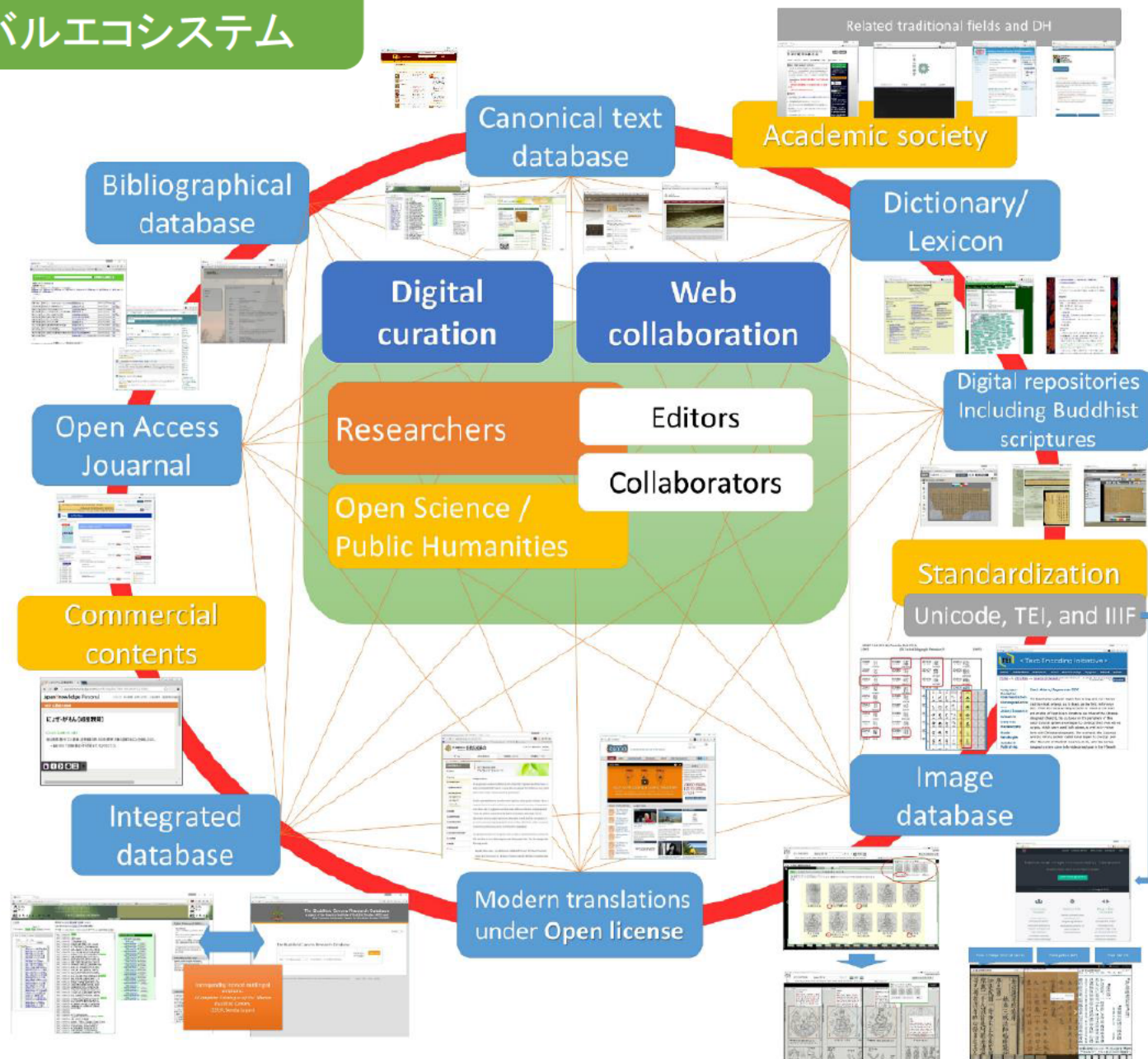
第1巻	第16巻	第31巻	第46巻	第61巻	第76巻	第91巻	第106巻	第121巻	第135巻	第173巻	続110巻
第2巻	第17巻	第32巻	第47巻	第62巻	第77巻	第92巻	第107巻	第122巻	第136巻	第187巻	
第3巻	第18巻	第33巻	第48巻	第63巻	第78巻	第93巻	第108巻	第123巻	第137巻	第190巻	
第4巻	第19巻	第34巻	第49巻	第64巻	第79巻	第94巻	第109巻	第124巻	第138巻	第191巻	
第5巻	第20巻	第35巻	第50巻	第65巻	第80巻	第95巻	第110巻	第125A巻	第139巻	第201巻	
第6巻	第21巻	第36巻	第51巻	第66巻	第81巻	第96巻	第111巻	第125巻	第140巻	第206巻	
第7巻	第22巻	第37巻	第52巻	第67巻	第82巻	第97巻	第112巻	第126巻	第141巻	第209巻	
第8巻	第23巻	第38巻	第53巻	第68巻	第83巻	第98巻	第113巻	第127巻	第142巻	第210巻	
第9巻	第24巻	第39巻	第54巻	第69巻	第84巻	第99巻	第114巻	第128巻	第155巻	第212巻	

SAT大正蔵図像DB

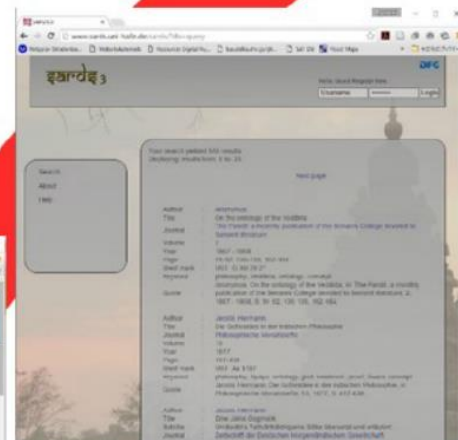
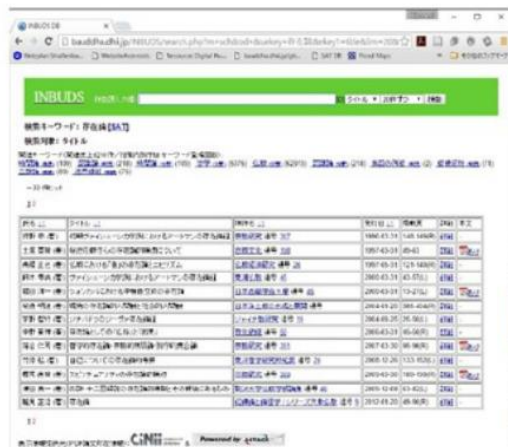
<http://dzkimgs.l.u-tokyo.ac.jp/SATi/images.php>



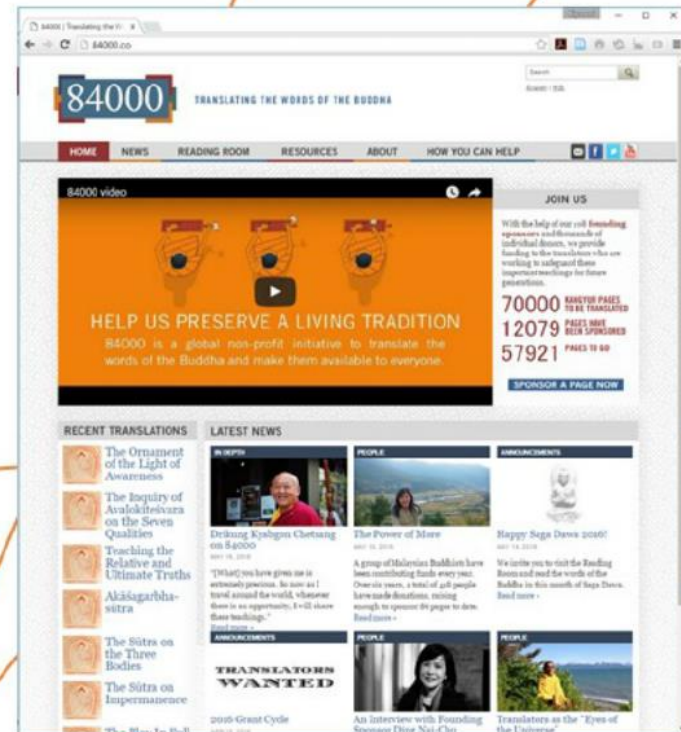
仏教学デジタルアーカイブの グローバルエコシステム



Bibliographical database



[illegible]



Modern translations
under Open license

20047 D-38

20988 EIC



Ontologies
Scholarly
Publishing

syllabic writing system called kana began to develop soon after the turn of the first millennium CE, and the Korean *hangeul* system came into widespread use in the fifteenth

Image database



From U-Tokyo General Library

From gallica (BnF)

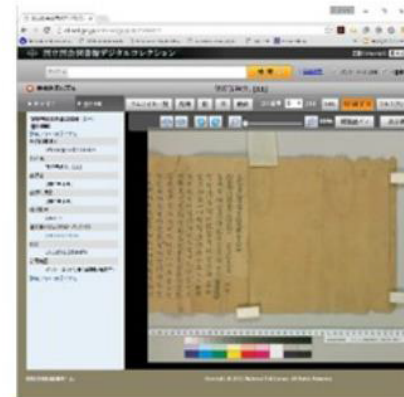
From SAT DB



Commercial contents



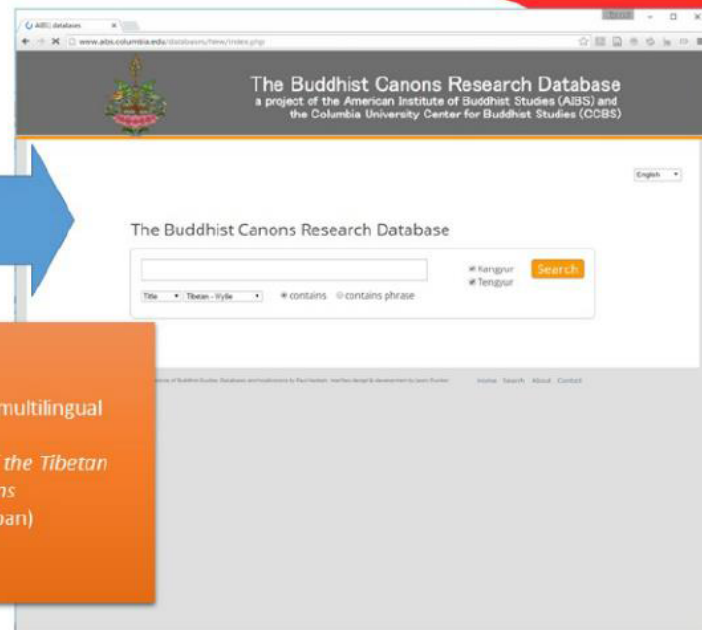
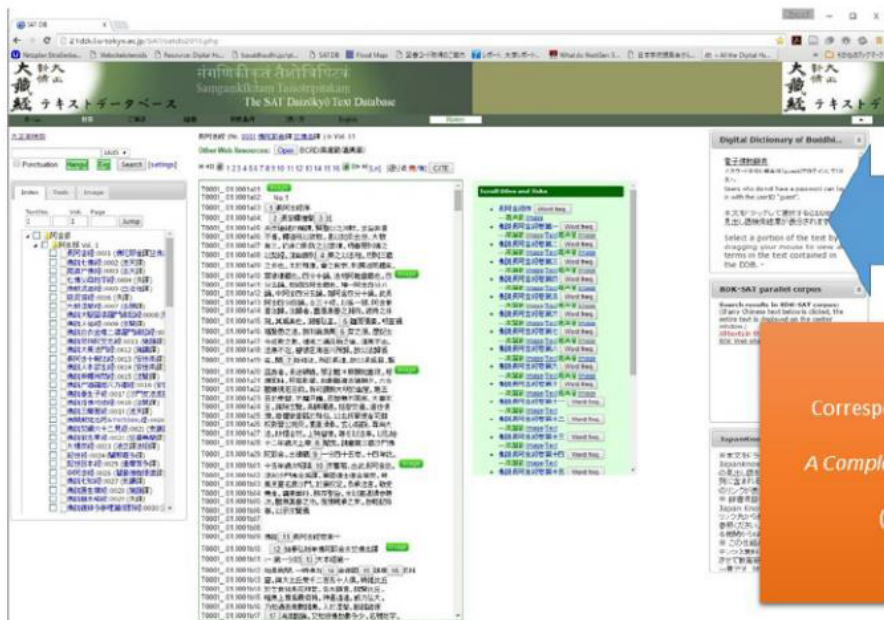
Digital repositories Including Buddhist scriptures



Dictionary/ Lexicon



Integrated database



Corresponding index of multilingual scriptures:
A Complete Catalogue of the Tibetan Buddhist Canons
(1934, Sendai Japan)

Digital curation

Web collaboration

Researchers

Editors

Open Science /
Public Humanities

Collaborators

Standardization

Unicode, TEI, and IJF

ISO/IEC 10646:2016 (E) Committee Draft (CD.2)

2D99E

CJK Unified Ideographs Extension F

2D9ED

2D99E 𠂔 G2-491108	2D9B2 𠂔 USA1-00153	2D9C5 𠂔 USA1-00811	2D9DA 𠂔 KC-06187
2D99F 𠂔 G2-491109	2D9B3 𠂔 USA1-00154	2D9C7 𠂔 USA1-00812	2D9DB 𠂔 G2-241932
2D9A0 𠂔 USA1-00546	2D9B4 𠂔 USA1-00155	2D9C8 𠂔 USA1-00813	2D9DC 𠂔 G2-241933
2D9A1 𠂔 JAU-050609	2D9B5 𠂔 USA1-00156	2D9C9 𠂔 USA1-00814	2D9DD 𠂔 G2-241934
2D9A2 𠂔 KC-06188	2D9B6 𠂔 KC-06189	2D9CA 𠂔 G2-241935	2D9DE 𠂔 G2-241935
2D9A3 𠂔 USA1-00157	2D9B7 𠂔 KC-06190	2D9CB 𠂔 G2-241936	2D9DF 𠂔 G2-241936
2D9A4 𠂔 KC-06191	2D9B8 𠂔 G2-241937	2D9CC 𠂔 G2-241937	2D9E0 𠂔 G2-241938
2D9A5 𠂔 KC-06192	2D9B9 𠂔 G2-241938	2D9CD 𠂔 G2-241939	2D9E1 𠂔 G2-241939
2D9A6 𠂔 USA1-00158	2D9BA 𠂔 USA1-00159	2D9CE 𠂔 G2-241940	2D9E2 𠂔 G2-241940
2D9A7 𠂔 G2-241941	2D9BB 𠂔 G2-241941	2D9CF 𠂔 G2-241942	2D9E3 𠂔 G2-241942

TEI < Text Encoding Initiative >

Home Guidelines Activities Tools Membership Support About News

Home -> Activities -> Special Interest Groups (SIGs) -> East Asian/Japanese

Computer-Mediated Communication
Correspondence
East Asian/Japanese
Education
Libraries
Manuscripts
Music
Ontologies
Scholarly Publishing

East Asian/Japanese SIG

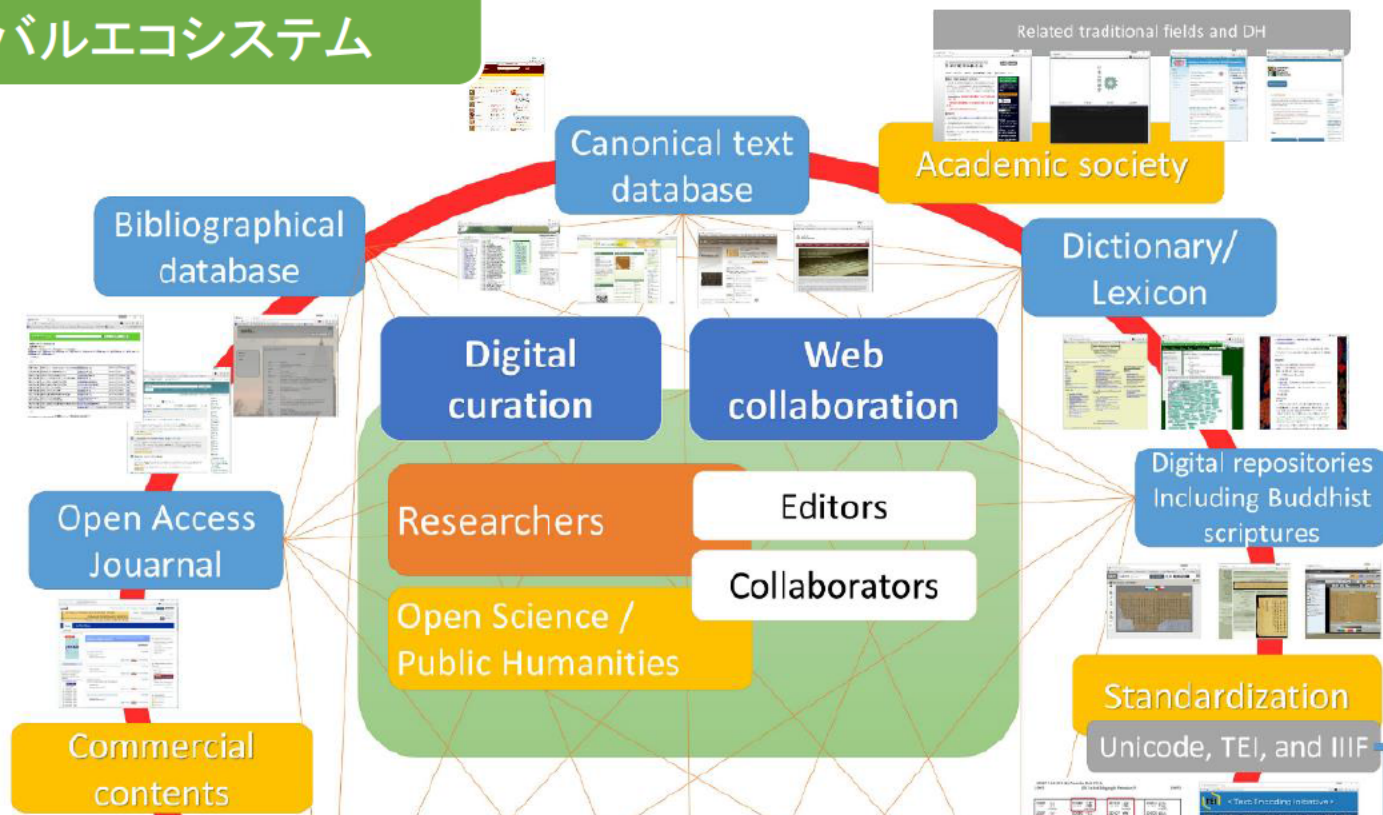
The East Asian cultural region has a long and rich literary tradition that extends as far back as the first millennium BCE. While the basic writing system for most of the early and middle of East Asian literature was that of the Chinese ideograph (*hanzi*), the cultures on the periphery of the *hanzi* cultural sphere also began to develop their own native scripts, which were used both alone, as well as in mixed form with Chinese ideographs. For example, the Japanese syllabic writing system called *kana* began to develop soon after the turn of the first millennium CE, and the Korean *hangeul* system came into widespread use in the fifteenth

2D99E 𠂔 GZ-4191108	2D9B2 𠂔 USAT-00150	2D9C6 𠂔 USAT-00851	2D9DA 𠂔 KC-05187
2D99F 𠂔 GZ-4191109	2D9B3 𠂔 USAT-00151	2D9C7 𠂔 USAT-02937	2D9DB 𠂔 GZ-2481502
2D9A0 𠂔 USAT-05546	2D9B4 𠂔 USAT-05177	2D9C8 𠂔 USAT-03680	2D9DC 𠂔 JMJ-059957
2D9A1 𠂔 JMJ-059959	2D9B5 𠂔 USAT-05206	2D9C9 𠂔	2D9DD 𠂔
2D9A2 𠂔 KC-05126	2D9B6 𠂔 KC-05182		
2D9A3 𠂔 USAT-02100	2D9B7 𠂔 KC-05181		
2D9A4 𠂔 KC-05156	2D9B8 𠂔 GZ-3561200		
2D9A5 𠂔 KC-05155	2D9B9 𠂔 GZ-3561200		
2D9A6 𠂔 USAT-02208	2D9BA 𠂔 USAT-01705		
2D9A7 𠂔	2D9BB 𠂔		

	11087	1108F	110A7	110B7	110C7
8	𠂔	𠂔	𠂔	𠂔	𠂔
9	𠂔	𠂔	𠂔	𠂔	𠂔
A	𠂔	𠂔	𠂔	𠂔	𠂔
B	𠂔	𠂔	𠂔	𠂔	𠂔
C	𠂔	𠂔	𠂔	𠂔	𠂔
D	𠂔	𠂔	𠂔	𠂔	𠂔

Academic society

仏教学デジタルアーカイブの グローバルエコシステム



SAT大蔵経テキストデータベース研究会は
このエコシステムを支える主要プロジェクト
の一つとして国際的な役割を担っている

SAT DBの主な相互連携

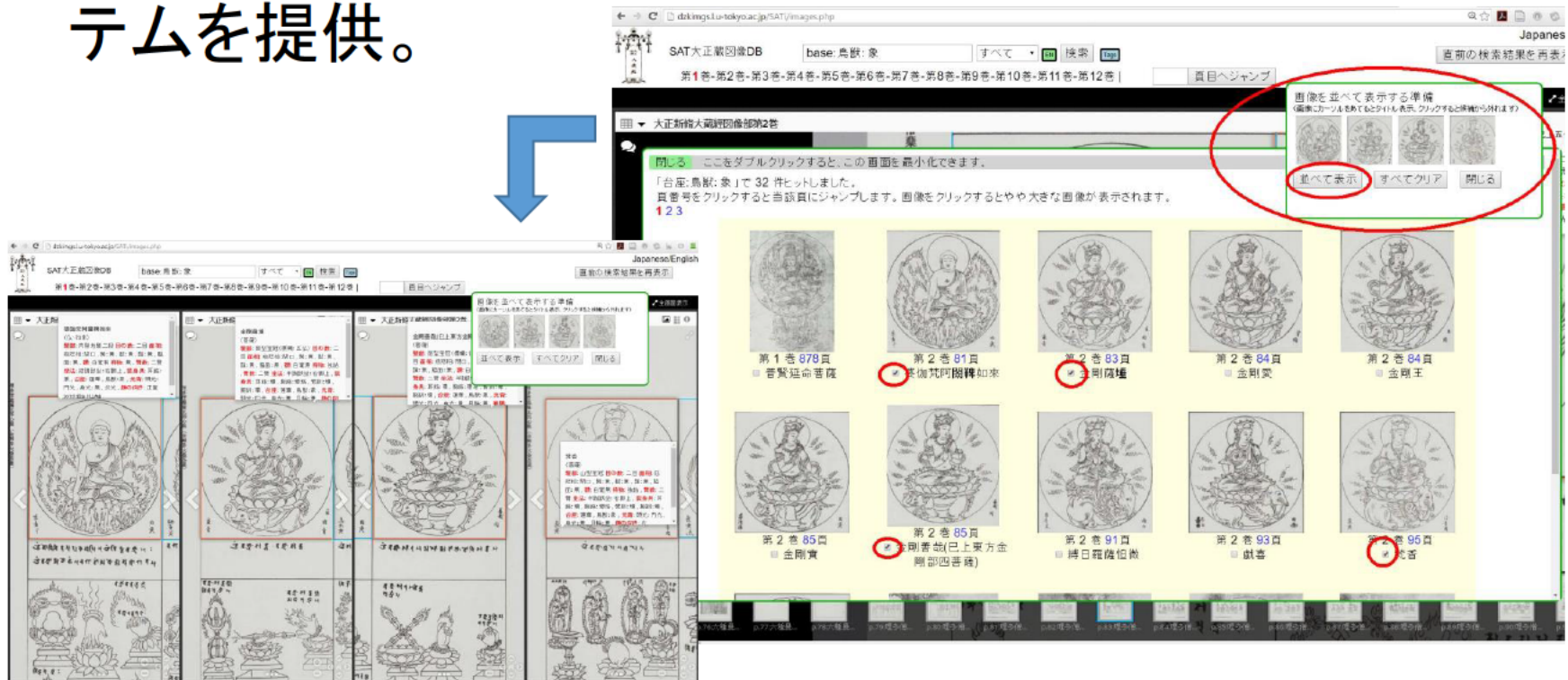
- 相互連携(先方からも同様のリンク等あり)
 - コロンビア大学 Buddhist Canons Research Database
 - 経典単位で漢訳中国大蔵経とチベット大蔵経をリンク
 - 高麗大蔵経研究所(ソウル)
 - 高麗大蔵経・大正新脩大蔵経を経典巻単位でリンク
 - 東京大学総合図書館
 - 萬暦版大蔵経画像DBと経典巻単位でリンク
 - 電子仏教辞典
 - テキスト本文をドラッグすると最長一致で英語版電子仏教辞典の見出し語と意味等を表示。
 - ジャパンナレッジ
 - テキスト本文をドラッグすると最長一致で仏教語大辞典の見出し語と読み、ジャパンナレッジの当該単語頁へのリンクを表示。
 - 公益財団法人仏教伝道協会
 - 漢訳大蔵経と英訳大蔵経の対訳コーパスに基づいて文章単位での相互リンクや英訳用例の検索機能を提供
 - 日本印度学仏教学会
 - 1987年から継続して作成されている専門分野論文書誌データベース + CiNii/J-Stage論文所在情報サイトを画面上で検索

他のコンテンツへの一方向リンク

- 世界中のデジタル画像公開サイトにおける仏典画像へ経典単位でリンクを表示(一部は経典巻単位)
 - リンク先:
 - 国立国会図書館、フランス国立図書館、英国図書館、国文学研究資料館、早稲田大学、HathiTrust、e国宝、龍谷大学、立命館大学、CADAL、東京大学、京都大学、ケンブリッジ大学等
 - このリンクがなければ、すべてのサイトを各ユーザが見て回らなければならない。
 - ⇒一部の機関がIIIF対応を始めているので、こちらの表示システムもIIIFへ移行中。
- 専門分野論文書誌データベース(ハレ大学)
- CiNii(論文書誌情報検索)、CHISE(漢字情報検索)等

IIIFの採用

- 画像共有の国際規格として普及しつつあるIIIFをSAT大正蔵画像DBの公開時点で採用した。
- IIIFのアノテーション機能を活用した検索閲覧システムを提供。



POD版の刊行

- 萬曆版大藏經に関して、プロジェクト内で版下を作成し、Amazon、Honto等でプリントオンデマンドの販売を開始。
- 紙媒体での利用を希望する人向け。
 - ページ番号で参照したがる人もいる。
- プロジェクトとしては利益を出さないようにして、実費のみで提供。
 - 通常の影印版に比較してかなり安価に提供。
 - 紙質、印刷の質はそれほど高度ではないがある程度の実用には耐え得る。
- 徐々に売れている模様。

研究活動としての展開

- 個別発表

- デジタル系・デジタル学際系学会

- 情報処理学会(人文科学とコンピュータ研究会・じんもんこんシンポジウム)

- Alliance of Digital Humanities Organizations (ADHO)

- 専門分野(仏教)系学会

- 日本印度学仏教学会、日本宗教学会、International Association for Buddhist Studies

- パネル発表

- 日本印度学仏教学会、東方学会、ADHO

- その他、各地のシンポジウムで招待講演

近年の科研費取得状況

- 「デジタル化」のプロジェクトではなく「デジタル化の研究」のプロジェクトとして展開しているもの
 - 2015-2019 科研費基盤研究(S)
 - 「仏教学新知識基盤の構築—次世代人文学の先進的モデルの提示」
 - 2010-2014 科研費基盤研究(A)
 - 「国際連携による仏教学術知識基盤の形成 — 次世代人文学のモデル構築」
 - 2008-2009 科研費萌芽研究
 - 「次世代新大蔵経編纂スキームの構築」
- データベース科研
 - 2015「SAT大蔵経テキストデータベース圖像編」
- ※他のプロジェクトとの連携によってコンテンツの厚みを増しているため、実際にWebサービス全体にかかっている費用はこの限りではない。

デジタルアーカイブ関連の国際標準への成果の反映

- Unicode (ISO/IEC 10646)への反映
 - 悉曇文字の異体字の符号化(登録) Unicode8.0
 - 仏典外字漢字2800字超の符号化 Unicode10.0(予定)
- Text Encoding Initiative (TEI)への対応
 - TEI Consortium におけるEast Asian/ Japanese SIGの設置
 - 日本語・漢字文化圏のテキストをより適切に扱えるようにする
- IIF Consortiumへの加盟
 - 加盟したのは東京大学大学院人文社会系研究科次世代人文学開発センター人文情報学拠点だが、加盟の基礎になっている活動はSATプロジェクトによるもの。

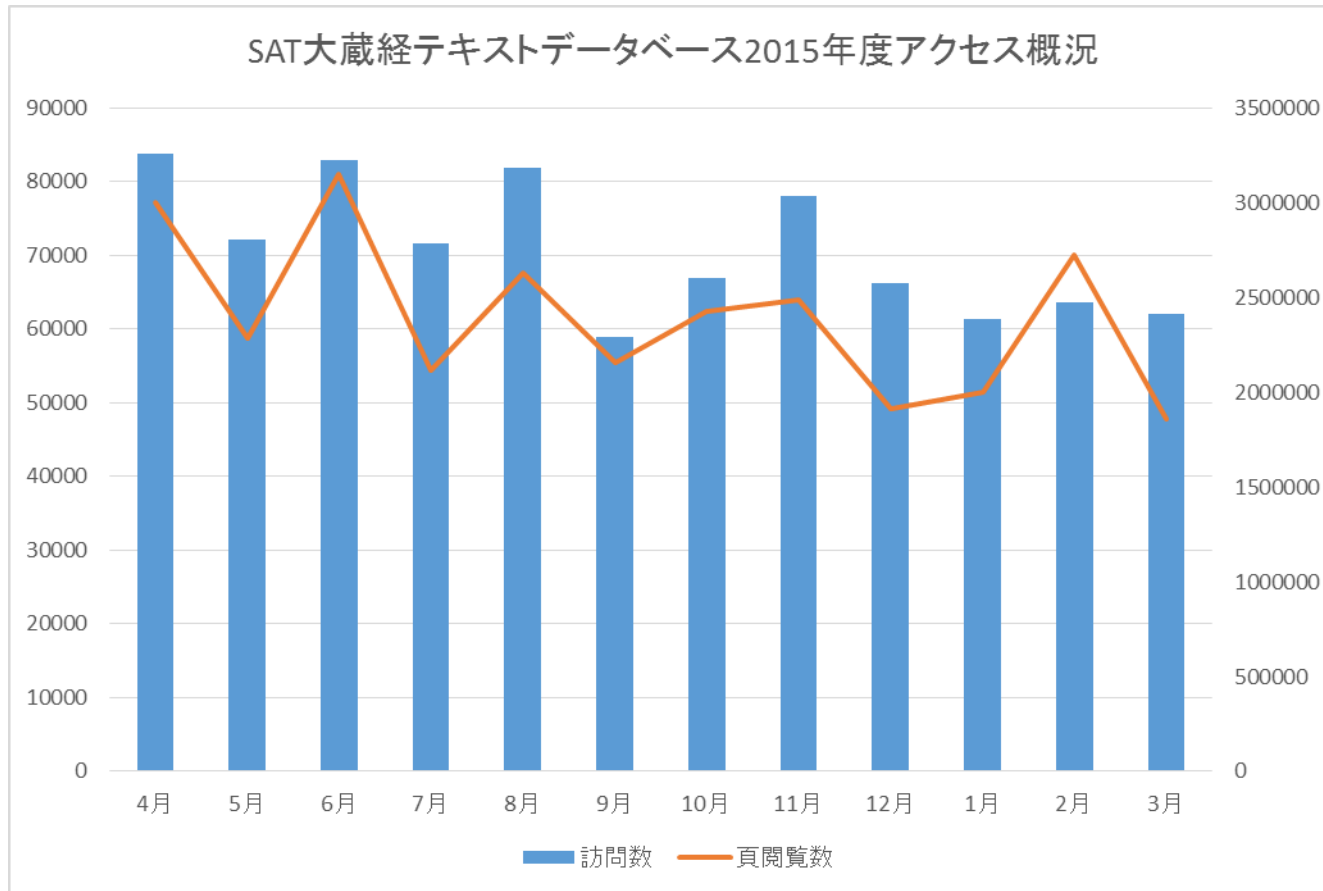
利活用促進の例：利用者講習会

- 利用方法の詳細についてハンズオンで解説。
 - 集まった聴衆に応じた活用例を提示。
- アンケート用紙に記述してもらう形で利用者の声・要望も同時に収集。
- 各地で開催してきている。
 - 国内
 - 北海道大学/東京大学/京都大学/駒沢大学/大正大学/大谷大学/国際仏教学大学院大学/浄土宗総合研究所/浄土真宗本願寺派研究所/曹洞宗総合研究センター/真宗大谷派しんらん交流館/全日本仏教会事務局
 - 海外
 - UCB/UCLA/ミシガン大学/ブリティッシュコロンビア大学/ライデン大学/ラトガーズ大学/浙江大学

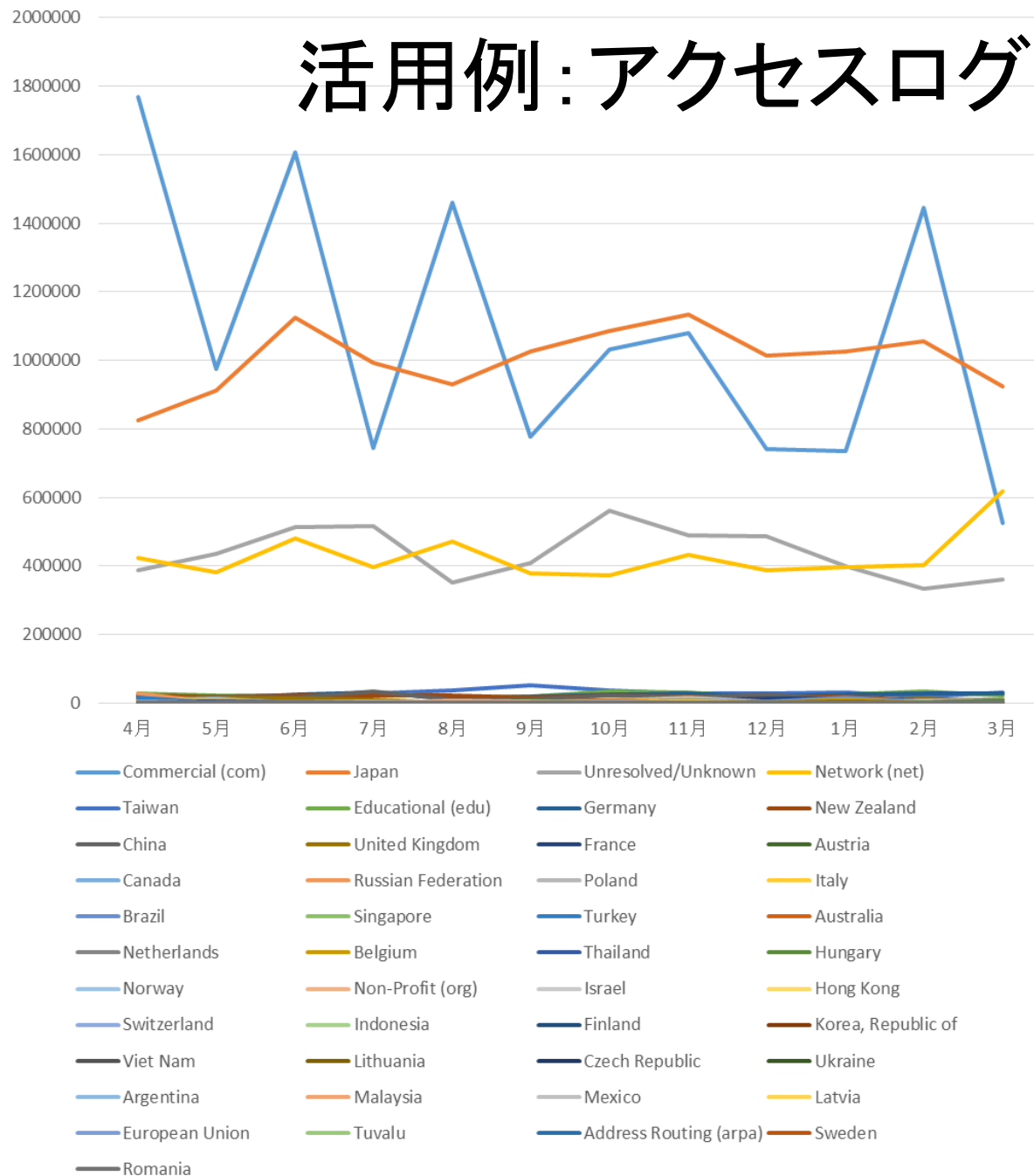
利活用促進の例：ファンディング

- http://butsugakushin.org/p4_josei_1.html
- 平成28年度一般財団法人仏教学術振興会研究助成募集要項より：
 - 2.研究分野
 - 以下のような広く新鮮な仏教研究を公募し、助成する。
 - a. SAT大蔵経データベースを直接の資料として用いた研究。
 - b.世界の思想・宗教状況を未来に向けて建設的に開いていく気鋭の論考。
- Cf. British Library Labs Awards
 - <http://labs.bl.uk/British+Library+Labs+Awards>

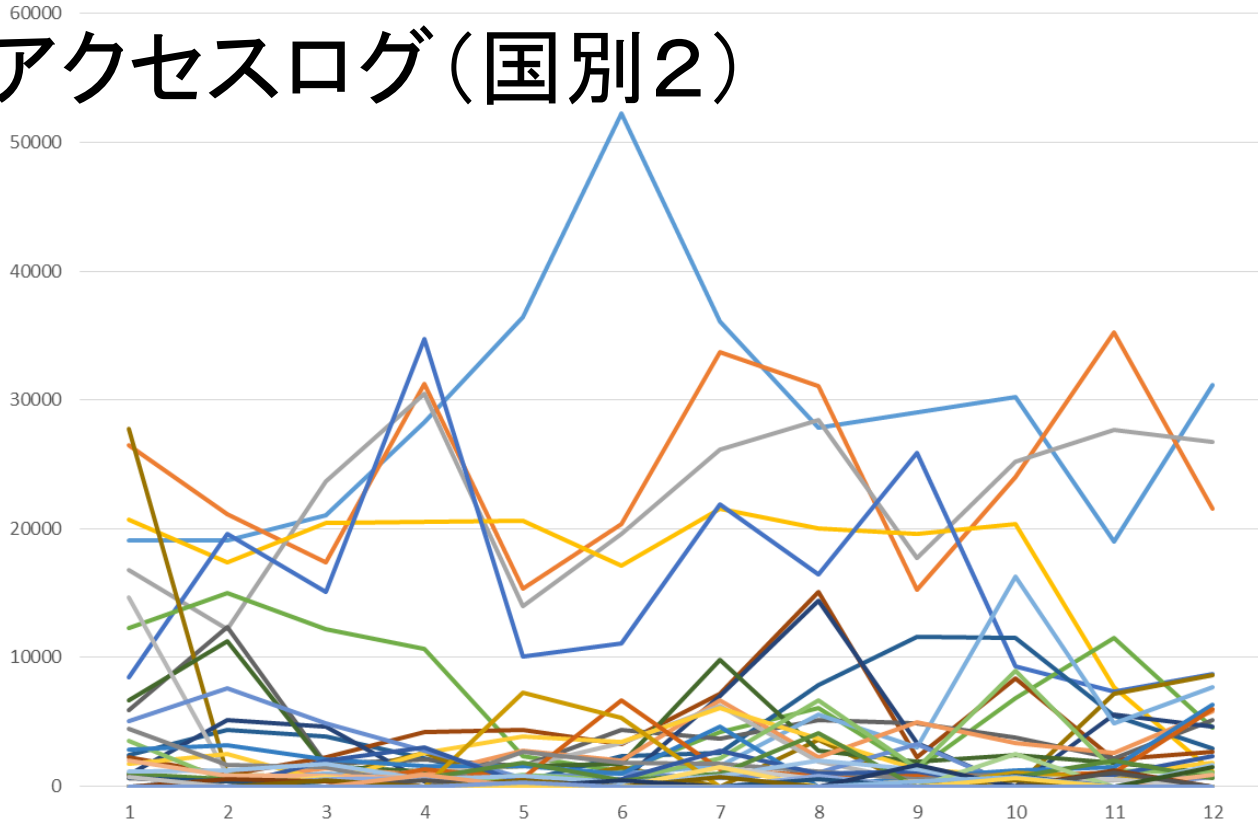
活用例：アクセスログ（総計）



活用例: アクセスログ(国別1)



活用例: アクセスログ(国別2)

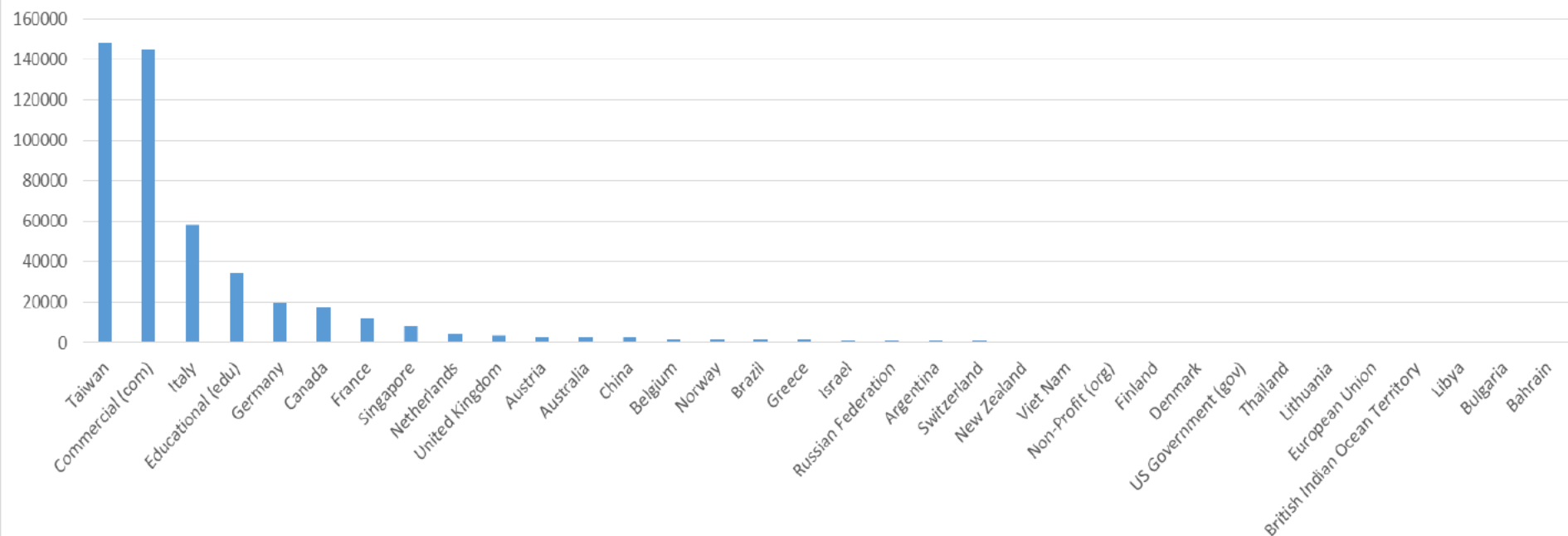


図像DB国別アクセス数上位3件は⇒

Japan	1620205
Network (net)	879988
Unresolved/Unknown	470815

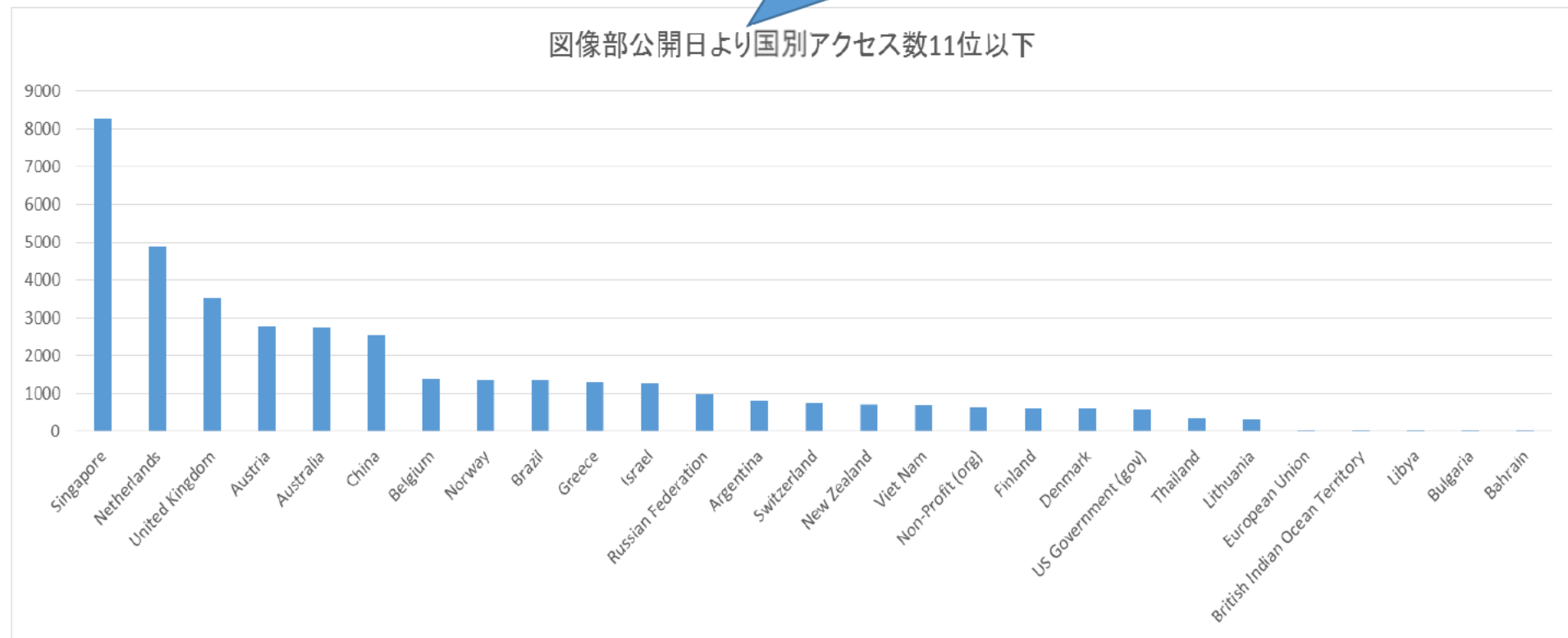
公開の翌月まで

図像DB公開日よりの国別アクセス数4
位以下



公開の翌月までの

図像部公開日より国別アクセス数11位以下



他分野からの利用例(1)

• 論文・エッセイ・発表等

- 池内 有為「オープンサイエンスの論点 Research Data Alliance第7回総会の議論と日本の趨勢」『情報管理』2016.7. vol. 59 no. 4, p. 241-249.
- 田中 政司「ジャパンナレッジの挑戦: 電子レファレンスツールの可能性」『情報管理』Vol. 59 (2016) No. 3 pp. 172-180.
- 生駒哲郎「大正新脩大蔵経テキストデータベース」『日本歴史』2010年1月号, 日本歴史学会編集, 吉川弘文館, pp. 122-123.
- 小口雅史「世界中に散らばった吐魯番出土文字資料断片を繋ぐ!」『科研費NEWS』2013 Vol.3, p. 4.
- Yokota-Carter, Keiko, 「Hathi Trust Research Center 101 : promoting Japanese studies in global digital environment」The 27th annual conference of the EAJRS, Leiden University, 2016/9/14-17.
- 楊曉捷 基調講演「デジタル時代と古典研究 ―画像資料のあり方を手がかりに―」大規模学術フロンティア促進事業「日本語の歴史的典籍の国際共同研究ネットワーク構築計画」第2回 日本語の歴史的典籍国際研究集会プログラム, 国文学研究資料館, 2016/7/29-30.
- 山本聡美「病草紙 ― 内なる異界としての身体 ―」 同上
- その他、確認中

他分野からの利用例(2)

IIIFの公式サイトにて、
IIIFの活用事例として
ショーケースに掲載

