Received: 18 March 2022 | Revised: 4 December 2022 | Accepted: 7 March 2023

DOI: 10.1002/gdj3.210

DATA SERVICES ARTICLE

Geoscience Data Journal RMetS WILEY

Online data service for geologic formations (Lexicons) of China, India, Vietnam and Thailand with one-click visualizations onto East Asia plate reconstructions

Wen Du[1,2] | **Suyash Mishra**[3] | **James G. Ogg**[1,2,3,4] | **Yuzheng Qian**[3,5] | **Sabrina Chang**[3] | **Karan Oberoi**[3] | **Aaron Ault**[3] | **Sabin Zahirovic**[6] | **Hongfei Hou**[7] | **D. S. N. Raju**[8,†] | **O'Neil Mamallapalli**[8] | **Gabriele M. Ogg**[4] | **Haipeng Li**[2] | **Christopher R. Scotese**[9] | **Bui Dong**[10]

[1]Institute of Sedimentary Geology, Chengdu University of Technology, Chengdu, Sichuan, China

[2]Deep-time Digital Earth (DDE) Research Center of Excellence (Suzhou), International Union of Geological Sciences (IUGS), Kunshan, Jiangsu, China

[3]Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA

[4]Geologic TimeScale Foundation, West Lafayette, Indiana, USA

[5]Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

[6]EarthByte Group, School of Geosciences, The University of Sydney, Sydney, New South Wales, Australia

[7]Institute of Geology, Chinese Academy of Geological Sciences, Beijing, China

[8]Oil and Natural Gas Corporation (ONGC), Dehradun, Uttarakand, India

[9]Paleomap Project, Department of Earth and Planetary Sciences, Northwestern University, Evanston, Illinois, USA

[10]Faculty of Geology, Vietnam National University of Science, Hanoi, Vietnam

Correspondence
Wen Du and James G. Ogg, Institute of Sedimentary Geology, Chengdu University of Technology, Chengdu, Sichuan 610059, China.
Emails: wendu_0911@icloud.com and jogg@purdue.edu

Funding information
Geologic TimeScale Foundation; Grant (Jiangsu, China), Grant/Award Number: BZ2022057 and KC2218;

Abstract

Paleogeography is the merger of sediment and volcanic facies, depositional settings, tectonic plate movements, topography, climate patterns and ecosystems through time. The construction of paleogeographic maps on tectonic plate reconstruction models requires a team effort to compile databases, data sharing standards and map projection methods. Two goals of the Paleogeography Working Group of the International Union of Geological Sciences (IUGS) program for Deep-Time Digital Earth (DDE) are: (1) to interlink online national lexicons for

Data set Included within each website URL.

Identifier URL: http://chinalex.geolex.org/; http://indplex.geolex.org/; http://vietlex.geolex.org/; http://thailex.geolex.org/

Creator: Earth History Visualization team of Purdue University & The Paleogeography Working Group of the IUGS Deep-time Digital Earth program.

Title: Online data service for geologic formations of China, India, Vietnam and Thailand.

Publisher: Geoscience Data Journal.

Publication year: 2023.

[†]Deceased.

Chengdu University of Technology (Sichuan, China); Purdue University (Indiana, USA); IUGS Deep-time Digital Earth (DDE)

all sedimentary and volcanic formations, and develop online ones for nations that currently lack these; (2) to target specific regions and intervals for testing/showcasing paleogeography output from the merger of these databases. Following those goals, we developed and applied new cloud-based lexicon data services and interactive visualization techniques to regions in East Asia. This has been a successful collaboration among computer engineers and plate modellers and has involved stratigraphers in India (ONGC), China (Chengdu Univ. Tech., and Chinese Acad. Geol. Sci.), United States (Paleomap Project, and Purdue Univ.), Australia (GPlates visualization team) and Vietnam (Vietnam Nat. Univ.). Independent online lexicons with map-based and stratigraphy-based user interfaces have been developed (as of the date of this submission in March 2022) for all Proterozoic to Quaternary formations on the Indian Plate (over 800) and Vietnam (over 200), the majority of the Devonian through Neogene of China (ca. 2000) and partially for Thailand. A multi-database search system returns all geologic formations of a desired geologic time from these four independent databases. With one click, users can plot the regional extent of one or of all of those regional formations on different plate reconstruction models of that desired age, and these polygons are filled with the lithologic facies pattern.

KEYWORDS

Earth history, Gondwana, GPlates, Paleogeography, Stratigraphy, TimeScale Creator

1 | INTRODUCTION AND OBJECTIVES

The initial challenge to the preparation of a regional or global paleogeographic map for a specific geologic age is having a database of the rock types (carbonates, siliciclastics, volcanics, etc.) with their geographic location/extent and interpreted depositional setting for that desired geologic age. The next challenge is to plot the available data onto different plate reconstruction models for that specific geologic age, so that one can make interpretations about the patterns and trends of depositional environments, paleotopography, basin history, paleoclimate and paleoceanography proxies, etc.

Many nations have online compilations of the nomenclature for their geologic formations with generalized descriptions of lithology, nature of upper/lower contacts, regional extent and other details. The user interfaces to those geological lexicons are quite varied and are usually oriented towards delivering extensive details about a specific formation but without overviews of its stratigraphic or regional context. Therefore, the geological surveys or stratigraphic commissions in some countries (e.g. Australia, Belgium, Canada, New Zealand, United Kingdom) have developed external interfaces using generalized stratigraphic graphics and map views in conjunction with the 'Earth History Visualization' team at Purdue University and the Geologic TimeScale Foundation. Each formation entry in the '*TimeScale Creator*' map-stratigraphic interface has URL links directly into its national lexicon page (Figure 1). The age span of each geologic formation has been provided and updated by those geological surveys and is converted to numerical dates using the age model of *Geologic Time Scale 2020* (Gradstein et al., 2020) and plotted using standardized lithology patterns. An advantage of using the free *TimeScale Creator* visualization system (https://timescalecreator.org/) is that its internal database also includes all the main biostratigraphic zonations, regional chronostratigraphic scales, interpreted eustatic sea-level oscillations, stable-isotope curves, global climatic trends, major large igneous provinces and other graphics of the Earth's history to enable a user to place the regional stratigraphic units into a global context (e.g. Zehady et al., 2020). These collaborative *TimeScale Creator* regional mappacks are publicly available under 'datapacks' at the timescalecreator.org website. Similar *TSCreator* mappacks for geologic formations throughout South America, Africa and Middle East do not currently have URL links into formal national lexicons, because most of the nations in these regions did not yet have online Lexicons.

During its initial workshop in 2019, the Paleogeography Working Group of the International Union of Geological Sciences (IUGS) program for

 Geoscience Data Journal RMetS

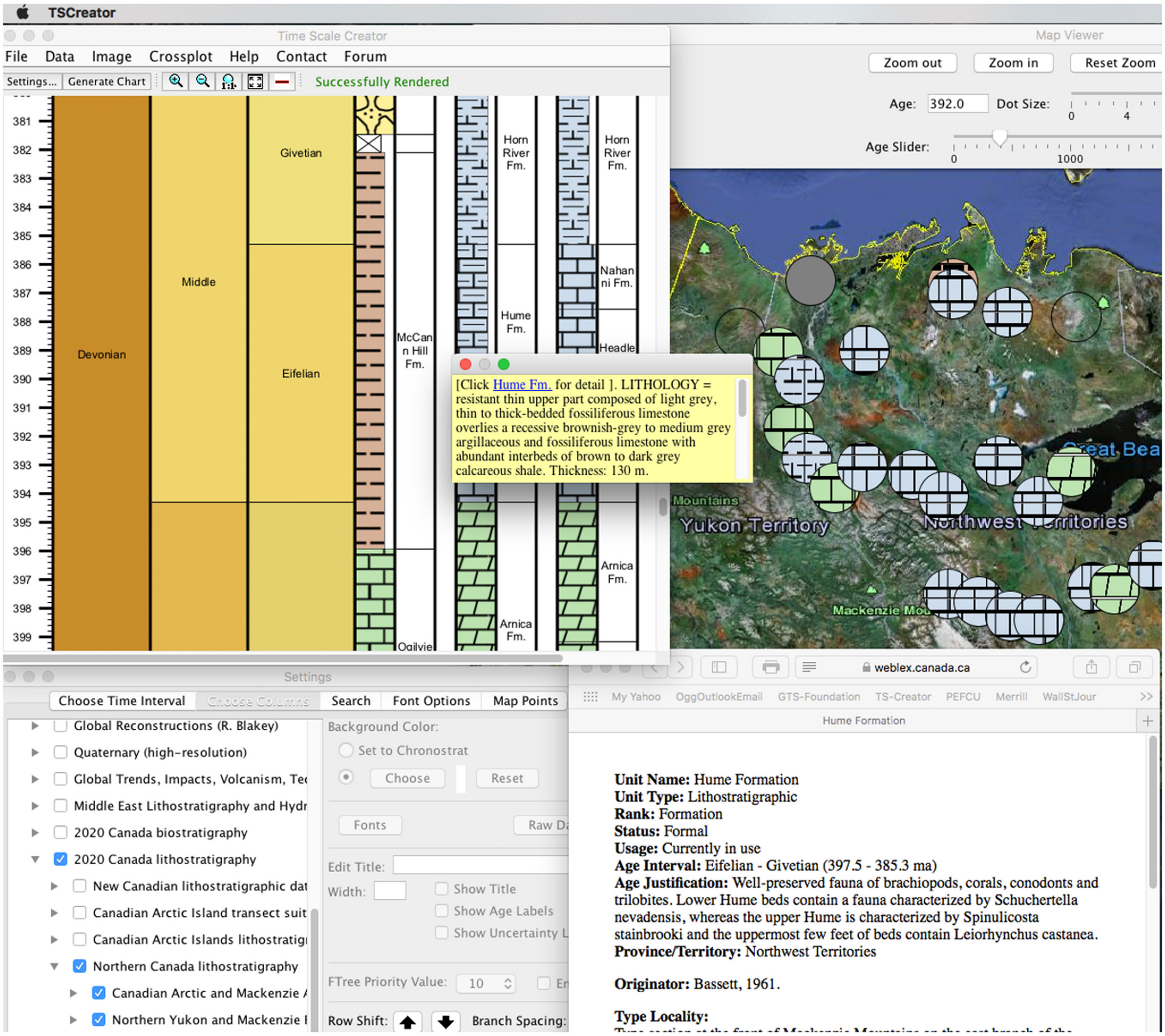


FIGURE 1 Excerpts from *TimeScale Creator* mappack of the Geological Survey of Canada ('2020 Canada Datapack' at https://timescalecreator.org/datapack/datapack.php) showing three mid-Devonian stratigraphic sections from the Northwest Territories (upper left; with TSCreator menu at lower left). A mouse-click on Hume Formation (blue-colored clayey limestone in middle stratigraphic section) accesses the information popup (yellow-colored central image) that has a URL link (blue 'Hume Fm' in the top line) to the GSC Lexicon webpage (https://weblex.canada.ca/html/006000/GSCC00053006812.html) for more details (lower right). Upper right image is the *TSCreator* schematic display of the lithology of the regional stratigraphic columns during the early Middle Devonian (392 Ma).

Deep-Time Digital Earth (DDE) recognized the critical need for easy user access to the detailed and updated geologic and depositional history of each region of our planet without the current restriction of having to visit each nation's lexicon system (if online), pay for access to specific basin history databases compiled by commercial consultants (e.g. IHS, Neftex-Halliburton) and search the voluminous recent geologic literature. Therefore, two goals of the Paleogeography Working Group are: (1) to interlink the current suite of online national lexicons for all sedimentary and volcanic formations, and develop new online ones for nations that currently lack these; and (2) to target specific regions and intervals for testing/showcasing paleogeography output from the merger of these databases. The initial focus would be on East Asia, which has an intriguing and debated plate-tectonic history, a wealth of recently updated dating and interpretation of local geologic formations, but a lack of relevant online databases.

These goals coincided with the initiation of independent programmes by the National Commission on Stratigraphy of China (hosted by the Chinese Academy of Geological Sciences) and by Dr. Raju's team at India's Oil and Natural Gas Corporation (ONGC) to compile and publish updated English-language lexicons for the geologic formations of all terranes that comprise modern China and for the entire India Plate respectively. The stratigraphy group at the Vietnam National University for Science was also eager to update and convert their non-digital book-form geologic lexicons (Thanh & Khuc, 2011; Van Tri & Khuc, 2011) to a public online system.

Therefore, our international collaboration was to aid in compiling detailed regional lexicons for East Asia geologic history and to design a data service to host and display this extensive information with user-friendly map-or-time visualization interfaces. This system was enhanced to enable the display of the location of formations onto appropriate plate-tectonic reconstructions with a single 'click'. This ambitious project has achieved most of its technical goals, and the new online lexicon systems for China, Indian Plate, Vietnam and Thailand are being progressively populated. At the time of this writing (mid-March 2022), there were about 2,500 individual formations that have been standardized and recalibrated in these four public online databases, and about 1,000 of these formation entries include areal-extent GeoJSON codes for one-click display onto any of three plate-tectonic models.

2 | MATERIALS AND METHODS

2.1 | Data collection and upload system

The most critical part of an authoritative database is the compilation and revision of the source information. For these lexicons, the details for each individual geologic formation were either being actively compiled for a planned set of publications for the Indian Plate (D.S.N. Raju et al., *South Asia and Myanmar Stratigraphic Lexicon [SAMSLEX]*, in prep) and for China (Hongfei Hou et al., *Stratigraphic Lexicon of China, in* prep.), or had been previously detailed in books on the stratigraphy of that region (Ridd et al., 2011; Thanh & Khuc, 2011; Van Tri & Khuc, 2011).

Hongfei Hou coordinated the China Lexicon compilation by expert teams with coordinators for each geologic period: Precambrian (Linzhi Gao), Cambrian (Shanchi Peng), Ordovician–Silurian (Xiaofeng Wang, Chuanshang Wang, Jianpo Wang), Devonian (Hongfei Hou), Carboniferous (Xiangdong Wang), Permian (Yue Wang), Triassic (Jinnan Tong), Jurassic (Jingeng Sha), Cretaceous (Xiaoqiao Wan), Cenozoic (Tao Deng [in chief]), Palaeogene (Qian Li), Neogene (Boyang Sun), Quaternary (Yong Wang).

D.S.N. Raju supervised the Indian Plate Lexicon compilation by expert teams with coordinators for each major basin region: Pakistan (Nusrat K. Siddiqui), Himalaya (O.N. Bhargava, Birendra Singh), Karakoram (K.P. Juyal), Kashmir (Rajeev Patnaik), Gondwana basins (G.V.R. Prasad and Varun Parmar), NE India (Kapesa Lokho), Cauvery (A.N. Reddy, R. Nagendra), with D.S.N. Raju's team covering most of the other basins from the ONGC studies or from external syntheses (e.g. Barber, 2017; Dhital, 2015; Garzanti, 1999; Gradstein et al., 1991; Long et al., 2011; Martin, 2017; Valdiya, 2016). The compilations were also partly based on his previous group compilation for ONGC (Raju & Misra, 2009). Upon Dr. Raju's unfortunate death in the summer of 2021, the completion of the Lexicon was coordinated by O'Neil Mamallapalli and A. Nailapa Reddy.

The initial Vietnam database based on Tong-Dzuy Thanh and Vu Khuc (2011) was extensively revised during a 1-week focused workshop at Vietnam National University (Hanoi) coordinated by Nguyen Thuy Duong and Bui Dong that involved compilations for Neoproterozoic–Mesozoic (Tran Van Tri, Dang Tran Huyen, Nguyen Van Vuong, Tong Day Thanh, Ta Hoa Phuong and Nguyen Nuu Hung) and Cenozoic (Tran Nghi) with additional guidance from the Vietnam Stratigraphic Commission (Tong-Dzuy Thanh's team). The current Thailand compilation was compiled by Wen Du, based largely on Ridd et al. (2011) and other publications.

The resulting detailed formation-by-formation descriptions for China and for the Indian Plate were submitted by the different contributors as Microsoft Word documents using a standardized format for publication. Therefore, we designed a parser for directly uploading a slightly modified form of those same Word documents into the online database. The parsing code of the online system recognizes the header for each subsection of the description if that header consists of certain words and was separated from other text by a paragraph mark (Figure 2). A separator line of a dozen asterisks is the flag to the parser that the next item will be the name of a new formation or group (if the name on the next line ends in 'Fm' for Formation, or 'Gr' for Group) followed by its set of headers. An occurrence of 'Fm' or 'Gr' for a unit name within the bodies of the description texts are flags to the database display system to automatically insert appropriate URL links during the generation of the on-the-fly webpage if those formation names are to other database entries.

The data fields are Name (the entry after asterisks that ended in Fm or Gr); **Period**; Age Interval (Map column); **Province**; Type Locality and Naming; Lithology and Thickness; *Lithology-pattern*; Relationships and Distribution: Lower contact, Upper contact, Regional extent; *GeoJSON*; Fossils; Age (discussion); *Age Span: Beginning stage, Fraction up in beginning stage, Beginning date (Ma); Ending stage, Fraction up in ending stage, Ending date (Ma)*; Depositional Setting; *Depositional-pattern*; Additional Information; and Compiler. The data fields marked by **Bold** in the previous sentence are automatically inserted into the two pull-down menus for searching each database, if these were not previously entered. The items in italics are for graphic output (e.g. filling the GeoJSON polygon of the areal extent with an appropriate lithology pattern or depositional pattern) or for computing the formation's age span based on the reference table for the ages of geologic stages. The date in Ma is computed upon uploading according to the entered percent-up in the geologic stage (e.g. begins at 25% up in Oxfordian); and all dates in the database are recomputed if a new reference table (Excel format) is uploaded.

2.2 | Database, data display, user interfaces and administrative functions

The online lexicons utilize the MySQL relational database management system to store data from uploaded Microsoft Word documents and to retrieve data for 'on the fly' webpages. The primary key is the Formation-name (or Group-name), rather than an obscure entry number system. A user has many ways to locate formations within the database (Figure 3). The standard methods are (1) by name, (2) by part of a name, (3) by period with pull-down menu, (4) by province with pull-down menu, (5) by beginning or ending date (in Ma) and (6) by a combination of filters.

When a search request is made, a GET request extracts the formation names that satisfy those criteria, and hotlinks to individual formations are generated (through PHP). The returned information from a filter lists the

****

Chaiburi Fm

Period: Triassic

Age Interval (Map column): Lower to Upper Triassic

Province: Sibumasu - Inthanon Zone: Lower Peninsula

Type Locality and Naming: Type area: Named after the largest carbonate mountain in Phatthalung province.

Synonym: หมวดหินชัยบุรี

[Figure 1 : Generalized distribution of different Triassic sedimentary environments and some of the more important lithostratigraphic units, and their relations to the terranes. (The geology of Thailand, 2011) page 146]

Lithology and Thickness: Thin- to thick-bedded and massive dolomite, laminated limestone, and bioclastic limestone. Consists of 3 members (Subdivision: Up to bottom: Phukhaothong Dolomite, Chiak Limestone, and Phanomwang Limestone Members.
Thickness: 400-500 m

Lithology-pattern: Dolomite

FIGURE 2 A subset of a Word document with data on a formation in Thailand for upload into the database. The top row of asterisks is the flag to the parser that the next line is a formation/group name for insertion as a row into the database. Headers followed by colons (period, age interval [map column], province, etc.) are flags to the parser to insert the subsections under those fields (columns) in the database table. The synonym in this example is in the Thai language.

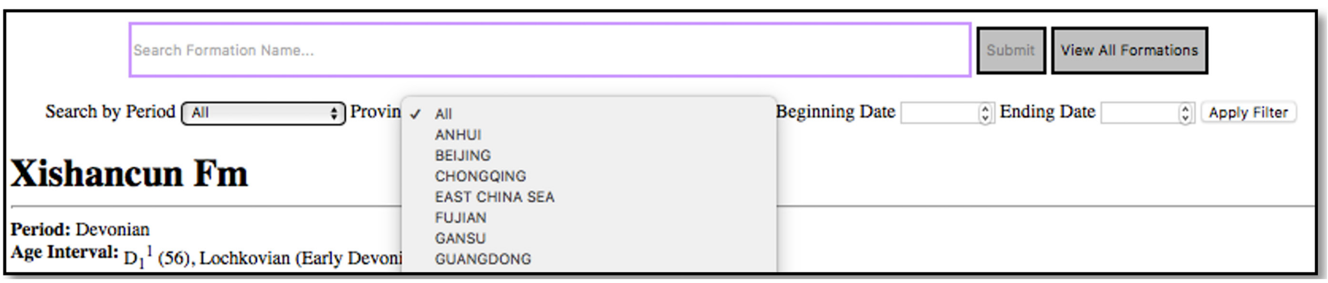


FIGURE 3 A portion of the standard search options for the ChinaLex database.

corresponding formations—either in stratigraphic order with the background of each URL link corresponding to the standard international colour of its beginning geologic stage (Figure 4), or in alphabetical order.

The hotlinks, which clicked, generate a URL of the form: http://chinalex.geolex.org/displayInfo.php?formation=Xishancun%20Fm. This enables an easy insertion (or deletion) of a formation name in the database, but does require that each formation have a unique name. Fortunately, that uniqueness is also required by the national lexicon standards.

The generated hotlink, through PHP programming, creates an 'on demand' webpage (Figure 5).

However, it is often more useful if users are able to make selections based on the geographic and stratigraphic context of a suite of formations. For example, if one was interested in the Devonian of Tibet, then it is often more user-friendly to be able to select formations based on the following type of inter-linked map and lithology interfaces (Figure 6).

After preparing graphic artwork for the location maps and array of stratigraphic sections in Adobe Illustrator, the webpage interfaces were added by using Adobe Dreamweaver. This required overlying and linking a graphic rectangle on each formation stratigraphic pattern to the corresponding appropriate database URL.

Of course, any database also requires an administrative interface. After an authorized regional expert logs into the lexicon admin webpage, the interface enables them to (1) upload additional or revised Word documents with formation information, (2) insert images under headers within a formation entry, (3) directly edit the text of a formation entry, (4) delete a formation and (5) change the default age look-up table for calculating the beginning/ending dates of all formations.

2.3 | Projecting the location of geologic formation onto a plate reconstruction

One of the powerful toolsets within the GPlates platform (https://www.gplates.org) is the ability to rotate the outline of a region on the present Earth's surface, such as tectonic plate or outline of a continental basin, to its ancient position using different models of the plate-tectonic motions through the past billion years. The Python version of GPlates (pyGPlates) was used to rotate simplified outlines (in GeoJSON format) of the estimated extent of geologic

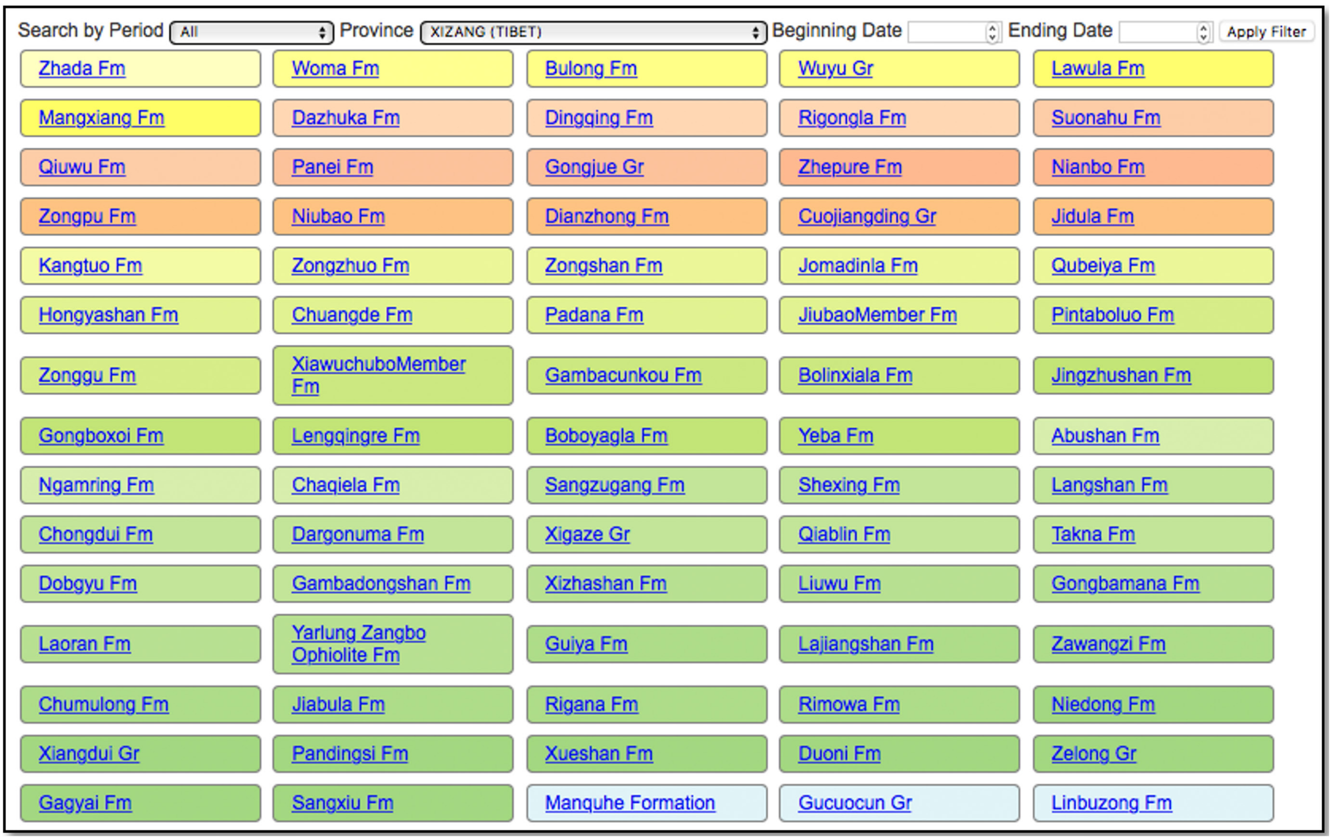
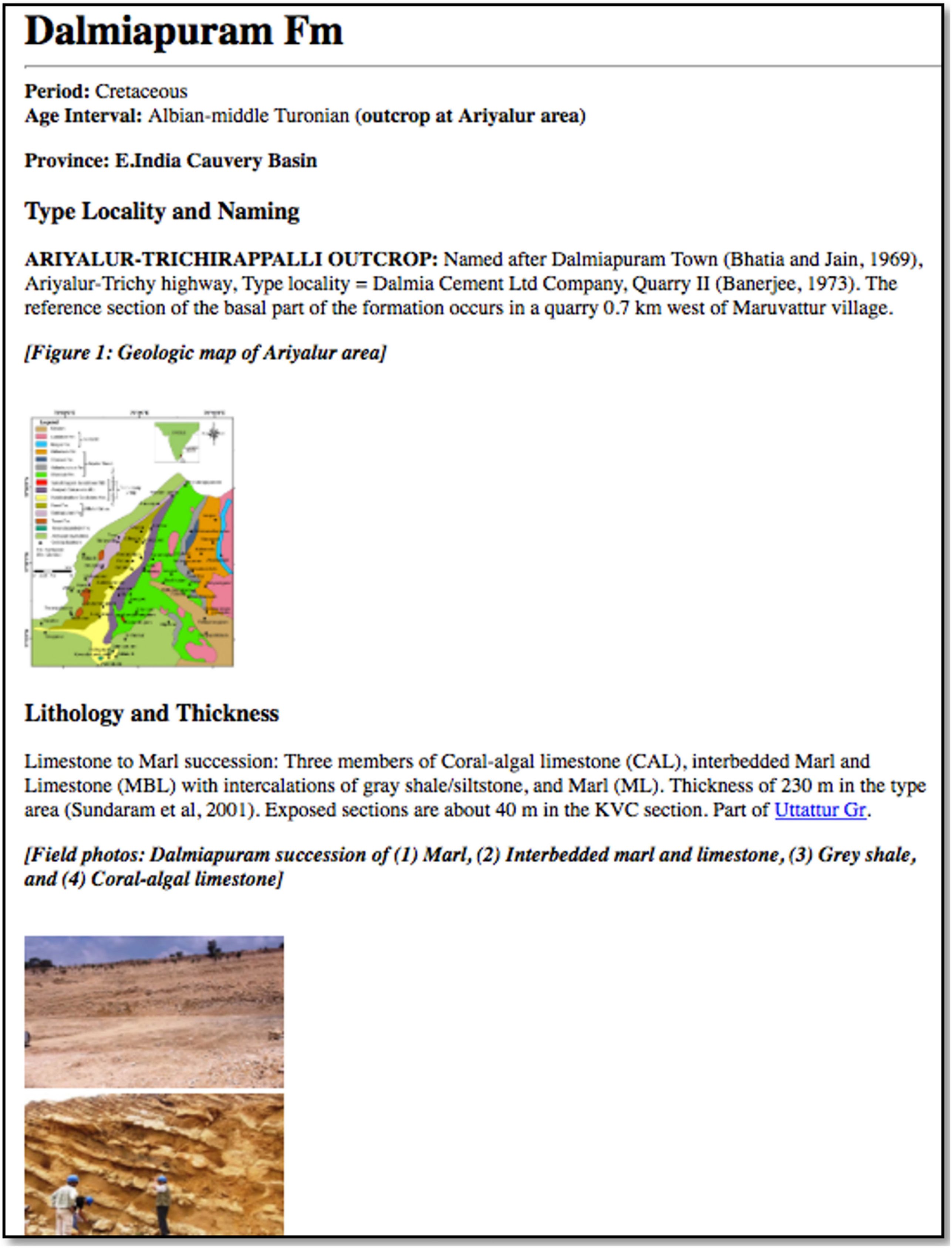


FIGURE 4 A portion of the formations of Xizang (Tibet) in the ChinaLex showing Neogene (yellow), Paleogene (orange), Cretaceous (green) and some Jurassic (blue) links. When one of these buttons are clicked, an on-demand webpage from that Lexicon will open (Figure 5).



Dalmiapuram Fm

Period: Cretaceous
Age Interval: Albian-middle Turonian (outcrop at Ariyalur area)

Province: E.India Cauvery Basin

Type Locality and Naming

ARIYALUR-TRICHIRAPPALLI OUTCROP: Named after Dalmiapuram Town (Bhatia and Jain, 1969), Ariyalur-Trichy highway, Type locality = Dalmia Cement Ltd Company, Quarry II (Banerjee, 1973). The reference section of the basal part of the formation occurs in a quarry 0.7 km west of Maruvattur village.

[Figure 1: Geologic map of Ariyalur area]

Lithology and Thickness

Limestone to Marl succession: Three members of Coral-algal limestone (CAL), interbedded Marl and Limestone (MBL) with intercalations of gray shale/siltstone, and Marl (ML). Thickness of 230 m in the type area (Sundaram et al, 2001). Exposed sections are about 40 m in the KVC section. Part of Uttattur Gr.

[Field photos: Dalmiapuram succession of (1) Marl, (2) Interbedded marl and limestone, (3) Grey shale, and (4) Coral-algal limestone]

FIGURE 5 Portion of an on-demand webpage with data extracted from the IndpLex database. Note that last sentence in 'Lithology and Thickness' has a URL link to the component Group entry; and similar URL links are to overlying/underlying formations in the next section (not shown). Images were uploaded under the appropriate headers through the Admin interface, and clicking on an image thumbnail opens the full size diagram as another window.

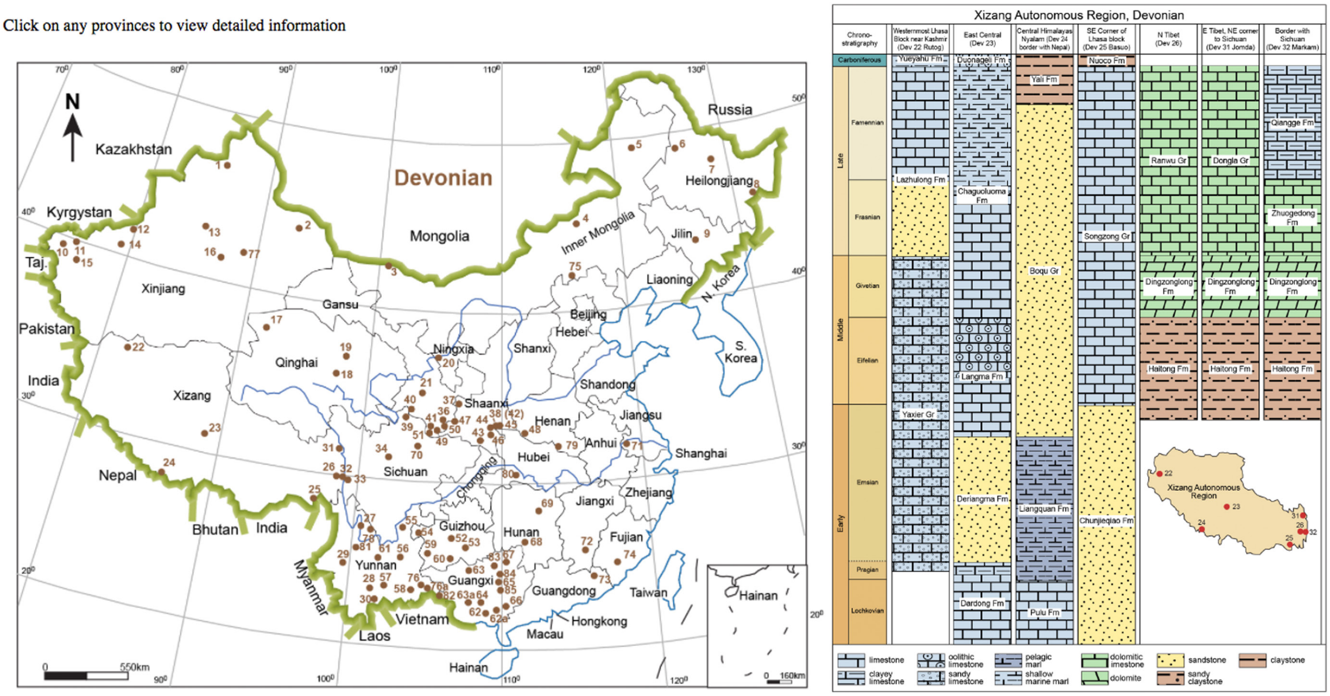


FIGURE 6 Devonian interface for ChinaLex. Clicking on Xizang (Tibet) province on the China map (left image) opens the generalized stratigraphic columns with their locations within Xizang (right image). Clicking on any formation in that diagram will open its webpage with details (similar to what is shown in Figure 5). [The lithology columns were initially produced using the associated *TimeScale Creator* mappack for China, then compacted.]

formations to the ancient coordinates according to its assigned geologic age, and then superimposed upon the host tectonic plate outlines or other GPlate libraries of graphics. The rotated tectonic plate and formation outlines are plotted using Python's version of the Generic Mapping Tool package (pyGMT; https://www.generic-mapping-tools.org) to display on a choice of map projections (e.g. Mollweide, orthographic etc.).

The GeoJSON digital coding of the simplified areal extent of individual formations utilized either the QGIS software or the geojson.io website tool. Basin map images from the literature were georeferenced in QGIS, then the component strata were digitized into polygons, and the two-decimal output in GeoJSON format was inserted into the formation's Word file for upload to the database (Figure 7). The coastline and plate boundary files used for georeferencing are from Matthews et al. (2016) and a QGIS plugin 'QuickMapServices'. During the upload process, the parsing procedure inserts 'FROMAGE' and 'TOAGE' time span items into the GeoJSON code according to the formation's age span that was calculated using percent-up in the geologic stage. That GeoJSON code contains all the necessary information for its pyGPlate rotation process and pyGMT plotting.

The integrated Lexicon–pyGPlates–pyGMT processes of rotation and plotting of geologic formations onto ancient plate reconstructions happen when the user selects which plate-tectonic model to apply, and clicks the orange button beside it (upper left image in Figure 8). The corresponding paleogeographic image is returned with the overlay of the formation's rotated GeoJSON polygon (examples in Figure 8).

The current (March 2022) Lexicon–pyGPlates interface has three current plate-tectonic and visual display options: (1) the GPlates default (Merdith et al., 2021) of only the full tectonic plate outlines surrounded by blue-coloured oceanic crust, (2) an interpreted continental-flooding history of each plate (Marcilly et al., 2021) using plate-tectonic rotations of Torsvik and Cocks (2017) and (3) Chris Scotese's suite of paleotopography graphics superimposed on his plate-rotation model (Scotese, 2021). These models span the Phanerozoic (past ca. 600 Myr), and the Merdith et al.'s (2021) model extends back to 1 billion years.

2.4 | Accessing multiple Lexicon sites to plot all formations onto plate reconstructions

Each Lexicon website includes a 'Multi-Country Search' that accesses the databases of multiple independent lexicon websites (currently China—http://chinalex.geolex.org/, Indian Plate—http://indplex.geolex.org/,

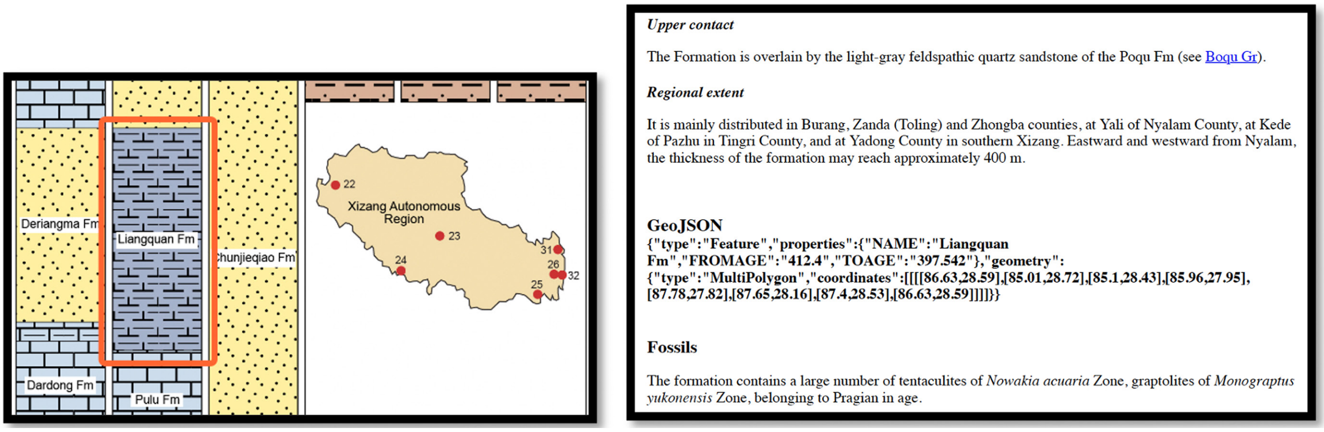


FIGURE 7 Example of estimated areal extent expressed as GeoJSON polygon for the Liangquan Formation of southern Tibet.

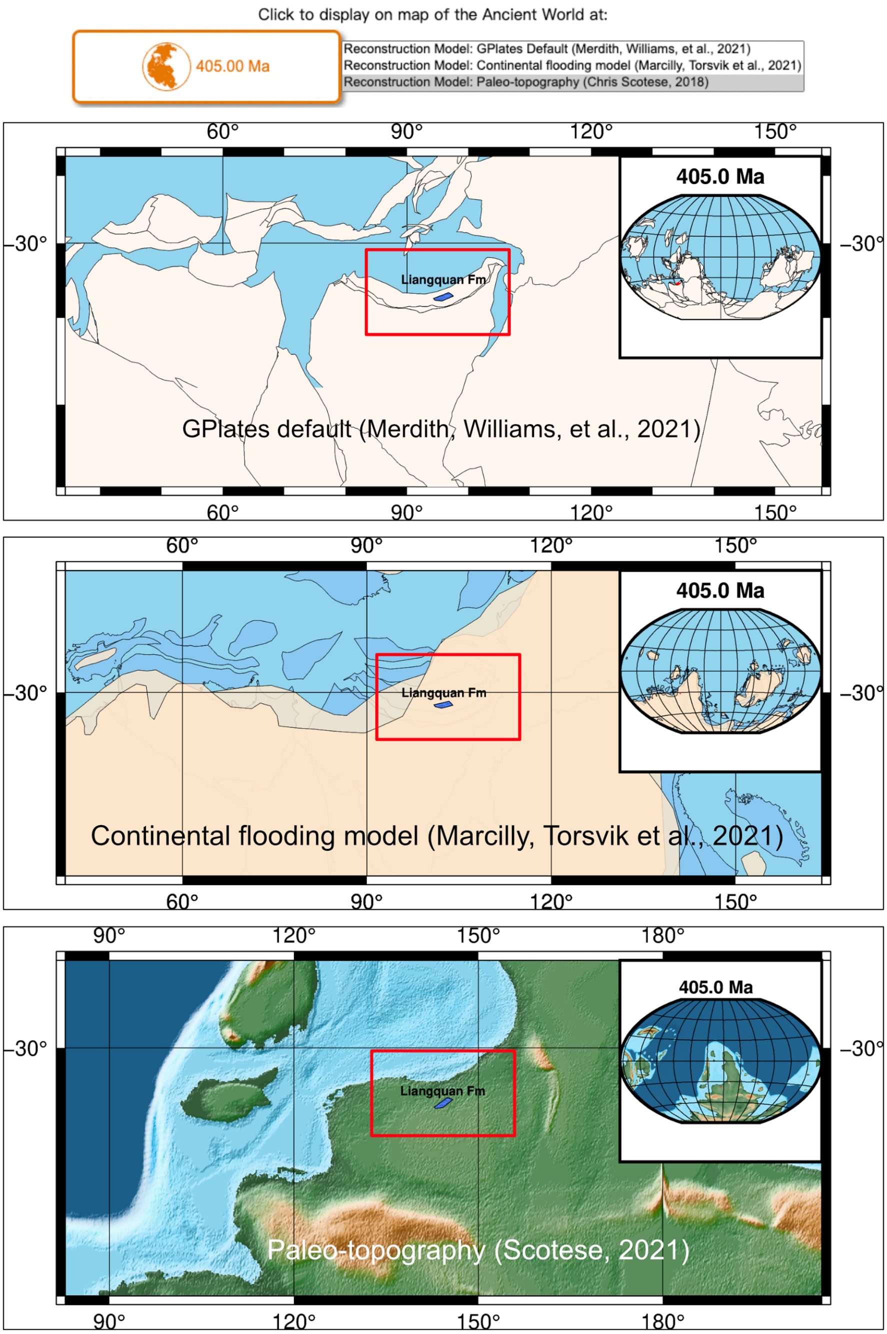


FIGURE 8 One-click display of the Liangquan Fm (south Tibet) onto three different plate reconstruction models for the Pragian Stage of Early Devonian. The Lexicon user selects among the options for plate reconstruction models (upper right image), and the areal extend of the Liangquan Formation in southern Tibet is rotated to fit the plate rotation model for its age. The output is shown for the GPlates default model (Merdith et al., 2021), continental flooding model (Marcilly et al., 2021) and schematic paleotopography (Scotese, 2021).

Thailand—http://thailex.geolex.org/, and Vietnam—http://vietlex.geolex.org/). The search includes options to return all formations whose span includes a specified geologic stage (e.g. Moscovian) or a specific date span (e.g. between 170 and 150 Ma; Figure 9); and then, the user can click on the individual formation names to go to the appropriate website page.

The returned information also includes the set of GeoJSON information. Therefore, the user can send the entire array of GeoJSONs into the linked pyGPlate–pyGMT system to plot the positions of all these formations on any of the three plate-tectonic models (similar to Figure 8) for the beginning, middle or ending date of the specified age range. One can also restrict the merged data to a single region (e.g. plot only the earliest Devonian of China's blocks onto a continental flooding model).

2.5 | Portability of these Lexicons to other systems

The current set of four individual Lexicons are hosted at an informal main cloud website (geolex.org). It is intended that each Lexicon database, user interfaces and linking to pyGPlates and pyGMT will be transferred to the appropriate country for its continued enhancement and updating (e.g. ChinaLEX to the National Commission on Stratigraphy of China). For this purpose, each Lexicon system is self-contained in a Docker module that can be installed on other servers. For the multi-country search function, it is important that external access be maintained and the web address changes be updated in the other national modules.

One goal of the IUGS's Deep-time Digital Earth (DDE) paleogeography program is to enable digital access to information on all formations and associated interpretations of sediment deposition and igneous eruptive environments through the Earth's history. Therefore, to maintain convenient global access, and to enable fluid merger of database information for insertion into community tools and visualization, it is anticipated that these Lexicons will always be mirrored within the IUGS-DDE cloud system.

3 | LIMITATIONS AND FUTURE DIRECTIONS

3.1 | Limitations and caveats

Even through the current Lexicon system (March 2022) contains over 2,500 formations, these suites for each region did not attempt to include every local-named geologic formation, nor to include many of the poorly dated Precambrian strata or any post-deposition intrusions. As of the time of this submission (March 2022), the National Commission on Stratigrahy of China has not yet completed their careful updating of early Paleozoic (Cambrian, Ordovician, Silurian) or early Mesozoic (Triassic, Jurassic) geologic formations. [However, after this submission was accepted, all of those periods, were added to the ChinaLEX during the remainder of 2022 and the beginning 2023, in addition to the insertion of the majority of Precambrian ones; for a total addition of another 1,000 formations.] The Indian Plate lexicon (IndpLEX) only contains about half of the named formations of Nepal. The Vietnam lexicon (VietLEX) contains all of the Phanerozoic formations for that region, including offshore. As of March 2022, the ThaiLEX set currently only contains Mesozoic formations. The preparation and insertion of the areal-extent GeoJSON for each of the eventual ca. 7,000 formations will take significantly longer, because this digital areal-extent information was not part of the documentation provided by the different stratigraphic commissions.

There is a tendency for novice users to assume that 'national' lexicons of geologic formations that are accompanied by colourful graphical user interfaces are authoritative 'final answer' representations of that region's geologic history. This is always a challenge in science—how to vividly present our current 'knowledge' for public use, but also to convey that it is a simplified capture of an evolving understanding of the true nature of our world. This is especially true for the estimated age spans of the formations in these lexicons. For example, the Jaishidanda Formation of the Lesser Himalaya in Bhutan had been considered as possibly Cryogenian in age (ca. 680 Ma) from its glacial facies,

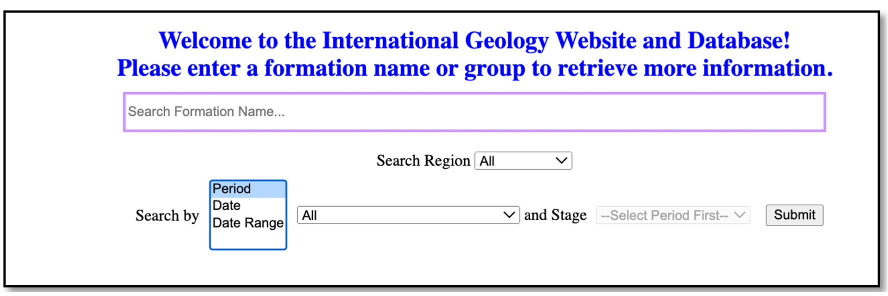


FIGURE 9 The search options for multiregion (currently the four independent lexicon websites of Indian Plate, Thailand, China and Vietnam) include listing all formations that fall within a desired geologic stage (or period) or within a desired date range.

but is now (Martin, 2017) assigned to the widespread late Carboniferous–early Permian glaciation of Gondwana (ca. 280 Ma); and that ca. 400-Myr shift in its age required reinterpretation of its adjacent strata. Many stratigraphic columns fail to indicate potential major hiatuses between strata, but draw the successions as continuous deposition with the bases of formations coinciding with the beginnings of geologic stages. Interpretation of depositional environments is fraught with unstated uncertainties. Former areal extent of the facies of a formation, which is crucial for making paleogeographic maps, is rarely studied. Field geologists and stratigraphers know about these caveats and other limitations, but the Lexicon entries provided by the regional stratigraphers generally do not include such uncertainties.

3.2 | Future directions

There are many future enhancements, user search functions and visualization items that we wish to add to the growing Lexicon database system.

1. Synonyms: The nomenclature and English-language spelling of Asian formations are not standardized in many cases. This is especially true for those of Xinjiang (Tibet) and Inner Mongolia, where the original local formation name has been converted into Chinese characters, and then into Pinyin English spelling. Similar situations occur for conversion of names from Hindi, Thai, etc. For the Himalaya regions, traditional descriptive formation names in the literature of the past century, such as the mid-Jurassic widespread 'Ferruginous Oolite Formation', are now being slowly renamed using a type locality and omission of the lithologic label as required by the international stratigraphic code. These present a challenge in enabling locating information on formation names from older literature, and for national users to find the formation name in Chinese, Vietnam or Hindi character spelling. Therefore, this synonym-search capability will be added during the summer of 2022. The current lexicon includes a 'synonym' header, but there has not yet been an effort to populate these with the proper Chinese, Hindi or other spellings. Similarly, there has not yet been an effort to translate all the entries into the local languages (e.g. a Hindi version of their Lexicon).
2. Keyword search: As of March 2022, the search functions only access formation names (or partial name), ages and provinces. It would be useful to enable searches of the text descriptions (e.g. 'coal' or 'reef'), fossil lists and other details to filter regional and inter-regional lexicons. Therefore, this keyword-search capability will be added to all websites during the summer of 2022.
3. Display of lithology-patterns into the GeoJSON polygons for both ancient and modern positions: The current pyGPlate–pyGMT display system only supports filling a set of polygons with one colour. Working with the GPlates and GMT software teams, we are enabling the display of the polygons with fillings by the appropriate lithology pattern (or depositional pattern). This will produce time slice displays of geologic facies, similar to this version for Vietnam during the Middle Permian (Figure 10).
4. Minor enhancements include adding URL links to the associated TSCreator mappacks when the individual Lexicon databases have a permanent home, adding additional recent plate-tectonic models for the pyGPlate options, enabling zoom-in on the returned plate-tectonic displays of formation locations and installing capability to export the entire database system in other formats for use in future applications (e.g. the FAIR goal of findable, accessible, interoperable and reusable).
5. International Lexicon digitization. During approximately 1956–1963, a major *Lexique Stratigraphique International* was compiled by a Subcommission for Stratigraphic Lexicon of the IUGS. Each continent-region 'volume' of this project encompasses several parts, and the entire suite spans nearly 3 m of library shelf space. Unfortunately, this significant achievement, which had attempted to summarize every named geologic formation of every nation, is nearly impossible to find at university libraries and has never been digitized. Even though some of the age spans, portions of nomenclature and some interpretations of depositional environments are no longer valid, a reformatting and upload of this invaluable source would provide a framework for many nations to rapidly create their own public lexicon websites with our Lexicon code for parsing into a standardized database, followed by preparation of graphic interfaces.

Most of these 'future directions' will be accomplished by eager and innovative computer science students during the remainder of 2022 and early 2023; whereas the enhancements of the Lexicon databases and the preparation of new ones with graphic interfaces require extensive months of careful work by geoscientists.

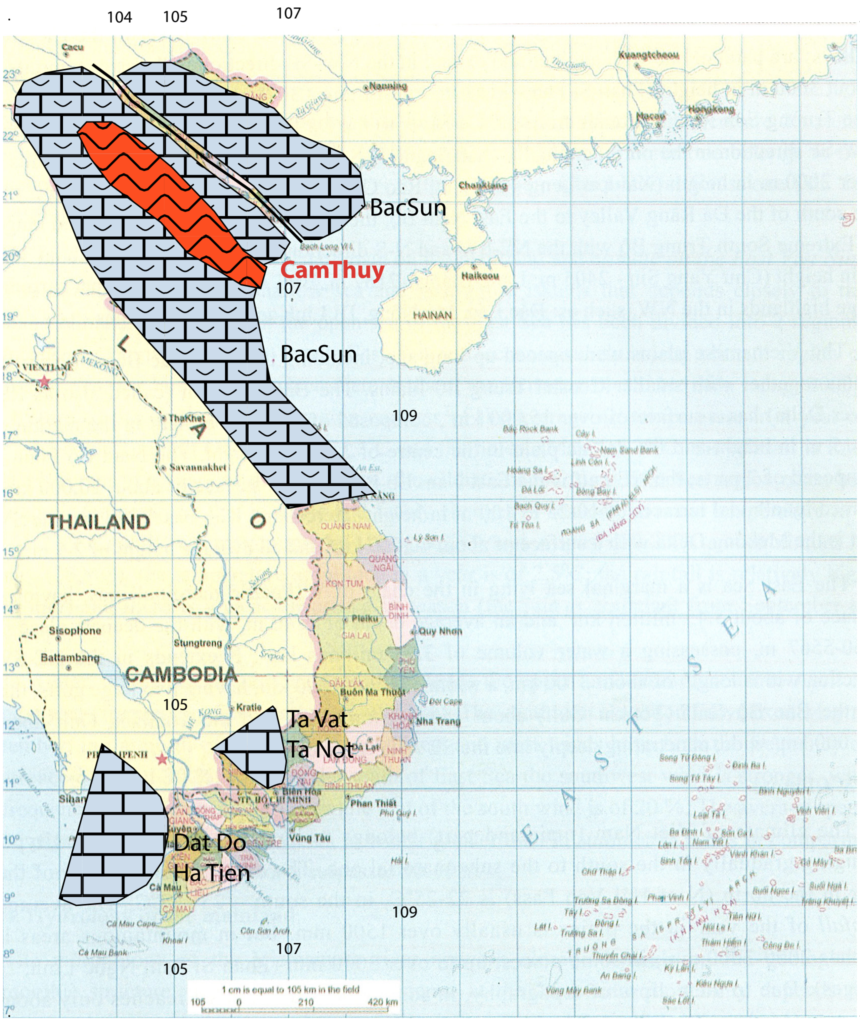


FIGURE 10 Generalized location and facies of Middle Permian formations in Vietnam. A similar database-generated time slice display directly from information in the regional Lexicons is one of our next goals.

4 | CONCLUDING REMARKS

The cloud-based data service websites for the Lexicons of China, Indian Plate, Vietnam and Thailand, which are temporally hosted at chinalex.geolex.org, indplex.geolex.org, vietlex.geolex.org and thailex.geolex.org, contain a total of over 3,000 geologic formations (status of March 2022). These databases include location, age, lithology, contacts, fossils and other standard information for each formation. In addition to basic search functions of name, partial name, province and geologic age, there are graphic interfaces of regional maps linked to composite stratigraphic columns. The areal extent of a formation can be projected as a polygon on any of three plate-tectonic reconstruction models for its geologic age. Multiregional searches of all four lexicon sites can merge all formations of a specific age and plot the entire array of polygons on these plate reconstruction models.

The current project was part of a programme to improve the paleogeographic maps of East Asia. However, the Lexicon website code and its linking to pyGPlate and pyGMT plate reconstruction and display platforms are designed to as a data service be installed on any server to enable other national geologic surveys or regional stratigraphic commissions to create similar cloud-based public websites. In addition to allowing user-friendly graphic-based access to the rock record of the region's geologic history, the standardized multi-website data merger and GeoJSON format will assist in compiling multiregional paleogeographic maps. IUGS's Deep-time Digital Earth project teams are eager to assist other national stratigraphic groups to achieve this goal of making our planet's fascinating story-in-the-rocks accessible for both public enlightenment and research discovery use.

AUTHOR CONTRIBUTIONS

Wen Du: Data curation (equal); methodology (equal); visualization (lead); writing – original draft (equal). **Suyash Mishra:** Data curation (equal); methodology (equal); software (equal); visualization (equal); writing – review and editing (equal). **James G. Ogg:** Data curation (equal); funding acquisition (equal); project administration (equal); resources (lead); supervision (equal); validation (lead); writing – original draft (lead); writing – review and editing (lead). **Yuzheng Qian:** Data curation (equal); methodology (equal); software (equal); visualization (supporting). **Sabrina Chang:** Software (lead); visualization (equal). **Karan Oberoi:** Data curation (equal); methodology (equal); software (equal). **Aaron Ault:** Conceptualization (equal); methodology (lead); project administration (equal); software (lead); supervision (lead). **Sabin Zahirovic:** Methodology (lead); software (lead); supervision (equal); visualization (equal). **Hongfei Hou:** Data curation (lead); investigation (lead); resources (lead). **D.S.N. Raju:** Data curation (lead); project administration (equal); resources (lead). **O'Neil Mamallapalli:** Data curation (lead); resources (equal). **Gabi Ogg:** Conceptualization (equal); software (equal); visualization (lead). **Haipeng Li:** Conceptualization (lead); supervision (equal); validation (equal). **Christopher R. Scotese:** Formal analysis (equal); resources (equal); supervision (equal); validation (equal); visualization (equal). **Bui Dong:** Investigation (equal); resources (equal).

ACKNOWLEDGEMENTS

We thank the College of Engineering (Purdue University) for supporting the Earth History Visualization teams under its Vertical Integrated Projects experience for undergraduates (https://engineering.purdue.edu/VIP). The prototype for the Lexicon database and interface was designed by Nikolas Ogg. Many of the compilations of the detailed information on the Asian formations for the Lexicons were contributed by numerous dedicated colleagues (especially retired professor volunteers) in India, China and elsewhere. Anqing Chen, Zheng'an Chen, Lei Zhang, Pengcheng Cai, Yingyi Ling, Yulu Shi and Sree Panuganti assisted on other aspects of the database compilations and translations.

FUNDING INFORMATION

The development of this service and preparation of this article was partially funded by Chengdu University of Technology (Sichuan, China), Purdue University (Indiana, USA), IUGS Deep-time Digital Earth (DDE) and the Geologic TimeScale Foundation (Indiana, USA), grants (BZ2022057, KC2218) (Jiangsu, China). This project is a part of the Deep-time Digital Earth (DDE) Big Science Program. Valuable reviews were received from Nicolas Flament, Shaunna Morrison and two anonymous reviewers.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

OPEN RESEARCH BADGES

This article has been awarded Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. Data is available at https://github.com/earthhistoryviz/geolex.

ORCID

Wen Du https://orcid.org/0000-0002-8087-1249
James G. Ogg https://orcid.org/0000-0001-5716-3511
Sabin Zahirovic https://orcid.org/0000-0002-6751-4976
Haipeng Li https://orcid.org/0000-0001-6443-6098
Christopher R. Scotese https://orcid.org/0000-0002-9742-3581

REFERENCES

Barber, A.J., Zaw, K. & Crow, M.J. (2017) *Myanmar: geology, resources and tectonics* Vol 48. London, UK: Geological Society, London, Memoirs, p. 712. Available from: https://doi.org/10.1144/M48

Dhital, M.R. (2015) *Geology of Nepal: regional perspective of the classic collided orogen, regional geology reviews*. Heidelberg, NY and Dordrecht, London: Springer Publ, p. 499. Available from: https://doi.org/10.1007/978-3-319-02496-7

Garzanti, E. (1999) Stratigraphy and sedimentary history of the Nepal Tethys Himalaya passive margin. *Journal of Asian Earth Sciences*, 17, 805–827. Available from: https://doi.org/10.1016/S1367-9120(99)00017-6

Gradstein, F.M., Gibling, M.R., Sarti, M., von Rad, U., Thurow, J.W., Ogg, J.G. et al. (1991) Mesozoic Tethyan strata of Thakkhola, Nepal: evidence for the drift and breakup of Gondwana. *Palaeogeography, Palaeoclimatology, Palaeoecolology*, 88, 193–218. Available from: https://doi.org/10.1016/0031-0182(91)90065-Y

Gradstein, F.M., Ogg, J.G., Schmitz, M.D. & Ogg, G.M. (coordinators). (2020) *A Geologic Time Scale 2020*, Vol. 2 volumes. Boston, Mo: Elsevier, p. 1357 https://www.sciencedirect.com/book/9780128243602/geologic-time-scale-2020

Long, S., McQuarrie, N., Tobgay, T., Grujic, D. & Hollister, L. (2011) Geologic map of Bhutan. *Journal of Maps*, 7, 184–192. Available from: https://doi.org/10.4113/jom.2011.1159

Marcilly, C.M., Torsvik, T.H., Domeier, M. & Royer, D.L. (2021) New paleogeographic and degassing parameters for long-term carbon cycle models. *Gondwana Research*, 97, 176–203. Available from: https://doi.org/10.1016/j.gr.2021.05.016 and Suppl. material.

Martin, A.J. (2017) A review of Himalayan stratigraphy, magmatism, and structure. *Gondwana Research*, 49, 42–80. Available from: https://doi.org/10.1016/j.gr.2017.04.031

Matthews, K.J., Maloney, K.T., Zahirovic, S., Williams, S.E., Seton, M. & Müller, R.D. (2016) Global plate boundary evolution and kinematics since the late Paleozoic. *Glob. Planet. Change*, 146, 226–250. Available from: https://doi.org/10.1016/j.gloplacha.2016.10.002

Merdith, A.S., Williams, S.E., Collins, A.S., Tetley, M.G., Mulder, J.A., Blades, M.L. et al. (2021) Extending full-plate tectonic models into deep time: linking the Neoproterozoic and the phanerozoic. *Earth-Science Reviews*, 214, 103477. Available from: https://doi.org/10.1016/j.earscirev.2020.103477

Raju, D.S.N. & Misra, R. (2009) Proterozoic and Phanerozoic integrated stratigraphy (south-East Asia): India, Pakistan, Bangladesh, Myanmar and Sri Lanka plus separate folio of fold-out charts. *ONGC Bulletin*, 44(2), 536.

Ridd, M.F., Barber, A.J. & Crow, M.J. (2011) *The geology of Thailand*. London: The Geological Society, p. 626.

Scotese, C.R. (2021) An atlas of Phanerozoic paleogeographic maps: the seas come in and the seas go out. *Annual Review of Earth and Planetary Sciences*, 49(1), 679–728.

Thanh, T.-D. & Khuc, V. (2011) *Stratigraphic units of Vietnam*, 2nd edition. Hanoi: Vietnam National University Publisher, p. 533.

Torsvik, T.H. & Cocks, L.R.M. (2017) *Earth history and Palaeogeography*. Cambridge, UK: Cambridge University Press, p. 317.

Valdiya, K.S. (2016) *The making of India: geodynamic evolution*, 2nd edition. Society of Earth Scientists Series. Heidelberg, NY and Dordrecht, London: Springer Publ, p. 945. Available from: https://doi.org/10.1007/978-3-319-25029-8

Van Tri, T. & Khuc, V. (2011) *Geology and earth resources of Viet Nam*. Hanoi: General Department of Geology and Minerals of Viet Nam, Publishing House for Science and Technology, p. 634.

Zehady, A.K., Ogg, J.G., Fordham, B.G., Palem, G., Bobick, J. & Ogg, G.M. (2020) Visualization of evolutionary relationships through geologic time in *Timescale creator*. *Applied Computing and Geosciences*, 8, 100037. Available from: https://doi.org/10.1016/j.acags.2020.100037

How to cite this article: Du, W., Mishra, S., Ogg, J.G., Qian, Y., Chang, S., Oberoi, K. et al. (2023) Online data service for geologic formations (Lexicons) of China, India, Vietnam and Thailand with one-click visualizations onto East Asia plate reconstructions. *Geoscience Data Journal*, 00, 1–14. Available from: https://doi.org/10.1002/gdj3.210