APPENDIX

Tracking Yu: Developing a Data System for Yellow River History

BY RUTH MOSTERN AND RYAN M. HORNE

When Ruth Mostern began conducting research into Yellow River history, she discovered that most publications about the Yellow River from China after the 1930s were claiming that, counting from the time of the legendary Yu the Great, the Yellow River had experienced 1,500 floods and twenty-six course changes. Those assertions were ubiquitous and consistent but seldom footnoted. When they were referenced, they routinely pointed toward a 1935 book called *The Yellow River Annals* (*Huanghe nianbiao*).[1] *The Yellow River Annals* is a chronological list of three thousand years of documented floodplain history. Each entry in the annals consists of a date, an event type, and a quotation from one or more primary sources. It seems that this work is the corpus upon which most assertions about the Yellow River have rested for the past eighty years. That makes it very influential, given the extent to which assumptions about the unstable and perilous character of the Yellow River have shaped warfare, large-scale engineering works, and even national identity during that time.

The Yellow River Annals focuses on the history of the alluvial plain, but it does not document erosion at its point of origin. Moreover, it has rarely been used as the basis for a database management system that takes advantage of an era of contemporary computing, the places named in it have never been georeferenced, and it has never been integrated with other annalistic lists of historical events or with other spatial and environmental information.[2] Mostern and Horne have therefore created their own chronicle of Yellow River history, with the conviction that the pursuit would inspire people interested in various fields, including the history of China, the world environment (because Yellow River records are the world's longest continuous collection of documentary observations about entanglements of people, water, and sediment), and the digital humanities. The remainder of this appendix describes the creation of the information system that forms the foundation of Mostern's research: the Tracks of Yu Digital Atlas (TYDA).

STEP 1: MAKING SPREADSHEETS AND SHAPEFILES

The TYDA aspires to be a comprehensive and interactive resource for identifying disruptive events, slow processes of change, and places of human activity throughout the entire Yellow River watershed and floodplain during all of recorded history. Its objective is to offer volume at the scale of the watershed and the Holocene rather than detail at the scale of the historical case study. For that reason, it is based on published sources with vast coverage rather than new primary source research. Inevitably, the TYDA inherits mistakes and biases from those sources. The project mitigates that problem by relying on sources that clearly reveal their own connection to the documentary and scientific record, and by cross-referencing multiple sources against one another. Moreover, we aim to make the data publicly available, thus allowing other people to improve its accuracy in the future.

THE MIDDLE COURSE

Four students at the University of California, Merced, and the University of Pittsburgh worked to create, merge, and clean information from print publications about settlements on the Loess Plateau, which is the erosion zone in the middle reaches of the river. Acknowledging them by name and in the order in which they worked on the project, they are Edward Lanfranco, Rocco Bowman, Zhifeng Shen, and Shaobai Xiong. Erin Mutch, the manager of the Spatial

Analysis and Research Center at the University of California, Merced, developed a work plan and helped to supervise Lanfranco and Bowman. Ryan Horne, digital history postdoctoral fellow at the University of Pittsburgh's World History Center, worked with Shen and Xiong.

The erosion zone component of TYDA is a database that is optimized for tracking changes to the distribution of settlements on the Loess Plateau, located primarily in Shaanxi province in northwestern China, which is the source of the sediment that periodically inundates the floodplain. Developing this content was time consuming but relatively straightforward. Mostern and her team georectified all of the maps in each volume of the *Historical Atlas of China* (*Zhongguo lishi ditu ji*) that overlapped the Loess Plateau.[3] Next they georeferenced all the units below county rank and created attribute tables that associated each unit with its parent prefecture and that also included the historical date referenced on the map in question. They supplemented the data from the *Historical Atlas* with information from a table of garrisons established during the Northern Song, because Mostern had identified that period as the most important turning point in the history of erosion on the Loess Plateau.[4] Next, they merged the attribute tables associated with each separate map to create a time series of establishment and disestablishment that spanned the whole imperial era. They identified continuous existence for any place that had the same name as a place that appeared on a previous map and that was within 5 kilometers of the identically named place.[5] Finally, they clipped the data to a shapefile that identified the extent of the Loess Plateau.[6]

THE FLOODPLAIN

The floodplain component of the TYDA database tracks waterworks constructions and repairs (e.g., levees, canals) and river disasters (e.g., floods, breaches). It comprises information from sixty-two lists and tables spread across ten Chinese publications about Yellow River history, agrarian history, and the history of natural disasters. *The Yellow River Annals* is one of those ten sources.[7] Every attested event in those lists and tables is referenced to one or more historical documents, and each one is also attributed to a single year and, where possible, at least one named location: 81 percent of entities in the database include an attested location.

Mostern and her graduate student Kaiqi Hua surveyed thirty-three published articles, books, and book series with structured information about Yellow River history. Hua made a data inventory spreadsheet based on these thirty-three publications. He assigned source codes, analyzed the content of each source, and classified each table in each source as either interesting, optional, important, or critical. In total, the sources comprised 1,017 pages and 8,711 entries. Mostern and Hua ultimately decided to include the important and critical tables: 3,407 entries. Hua digitized the contents primarily via hand typing after determining that optical character recognition for Chinese, including literary Chinese, formatted in ruled tables, was too prone to errors. His initial data entry parsed each of the sources into separate Microsoft Word documents with columns and rows that precisely reproduced the original form of each published table so as to maintain fidelity to the source material. He then converted each Word document into an Excel spreadsheet and assigned unique IDs for each published source table and each entity in each table. Next, maintaining those IDs, he reformatted each spreadsheet into a standardized column structure and then merged the spreadsheets into a first draft master list.

The next step, cleaning and parsing data, required significant historical expertise and semantic judgment. Under Mostern's direction, Hua removed records that were unrelated to the Yellow River or any other floodplain water system, excised events that occurred after the end of the imperial era in 1911, identified and merged records that were redundant across more than one of the ten core sources (taking care to maintain unique information and original ID numbers as he did so), and as necessary, parsed information across multiple spreadsheet columns.

Subsequently, he merged and split data to ensure that there was a single unique record for each event. The goal was a master table in which each event was associated with a specific date (at least a single year), at least one location, and a description in at least one primary source. This work relied primarily on Hua's exceptional skills of data selection and historical judgment. In some cases, historical sources conflated a series of events that transpired in the same year (e.g., a breach followed by flood followed by a repair), and we subdivided them accordingly. At the end of this task, we identified more duplicates and removed or reconciled them.

Next, we standardized event types. We began with the forty-nine event types in *The Yellow River Annals*. Some of these referred to dozens of events, and others to only a handful; some were distinctive, and others were nearly synonymous. With Mostern's input, Hua split, merged, and selected among them to create a vocabulary of twenty event types: ten items that refer to types of fluvial disasters and ten items that refer to river management activities. This set of categories allows for meaningful distinctions while still also permitting meaningful query and visualization. Where necessary, we have assigned multiple event types to individual events. The explanations of all the event types appear in table A.1, along with information about which synonymous terms from primary source and data tables are subsumed under each one. This table in itself is an accomplishment for river history research.

The final data development task for the floodplain tables in the TYDA was to identify all places named in the historical documents and event summaries. This was another task that required significant specialist research and aptitude in literary Chinese. First, Hua made a master list of the places named in river events. He followed the imperial spatial hierarchy. Sub-administrative places such as villages, levees, and buildings are all situated in counties, which nest into prefectures, then provinces. He assigned an ID to each unique place-name, and then linked event IDs and place IDs in many-to-many relationships (a single event may involve multiple places, and a single place may be the site of many events). He determined the latitude and longitude of each place with as much precision as possible, and then identified modern place-name matches for all historical place-names in the table. There are 1,642 unique places in the downstream place table in TYDA. These are categorized into four levels. When the sources name sub-administrative places, like towns or dams, these are Level 1 places, and there are 750 of these in the TYDA. The remainder of the places are counties or prefectures. A few sources name only provinces. In total, the TYDA floodplain data consist of 1,642 places and 3,754 events.

Next, Mostern and Horne imported all the locations in the place table into GIS along with China River Basins shapefiles.[8] The contemporary Yellow River floodplain is channelized between levees. To encompass the historical extent of the floodplain as it existed in all incarnations of the river course, we created one study area shapefile that includes the Hai River and Huai River basins as well as the modern Yellow River basin. That allowed us to identify places in the TYDA events database that lay outside of the study region and events that lacked a coordinate. Working with Mostern and Horne, undergraduate student Shaobai Xiong determined locations for all unreferenced places. In addition, for events located at places that fell outside the designated floodplain region, he determined whether the initially assigned place was mistaken and needed to be corrected, whether the original location was correct and should be maintained (we decided to retain a number of places at the edge of the floodplain that were technically outside the study area), or whether the event in question was outside the scope of this research and should be removed from the database. This activity—another task that required expert judgment and exceptional literary Chinese—also served to validate Hua's work.

CONTEXTUAL DATA

The previous section of this appendix describes two new databases created under Mostern and Horne's supervision. In addition, the full TYDA system also includes several previously published data sources, which support contextual queries and maps linking Yellow River history with enduring topics concerning Chinese historical geography and environmental history. Some of these resources are spatial data native to GIS systems, and some are data tables. Table A.2 is a table listing all these sources.

STEP 2: DESIGNING AND VALIDATING A SYSTEM

To finalize the TYDA and prepare it for eventual web publication, and to facilitate queries, visualizations, and other explorations of the data set, Horne integrated the various data sources into a PostGIS relational database. Several design considerations animated this stage of the project. One objective was to permit eventual interoperation with Linked Open Data projects like the World Historical Gazetteer, a digital spatial ecosystem for integrating specialist information about places from multiple contributors. Another was to create a union gazetteer, a merged list of all places relevant to the study, along with a list of all individual events and their types. A third was to be able to capture changes to the nomenclature, status, administration, and even location of individual places.

TABLE A.1
Contextual Data in the Tracks of Yu Digital Atlas

Category	Source
China Historical GIS	Historical Counties and Prefectures[a]
Lakes and Open Water	Harvard ChinaMap
Rivers	China Rivers Sorted by Basin[b]
Historic Courses of the Grand Canal	Harvard ChinaMap
Basemap	Ancient World Mapping Center, Natural Earth, NASA SRTM[c]
Vegetation	Harvard ChinaMap (2000)
Historical Courses of the Yellow River	Chen Yunzhen[d]
Dynasties and Reigns	Wikipedia
Mountain Peaks and Passes	Harvard ChinaMap
Ecosystems	World Wildlife Fund
Moisture Index	Monsoon Asia Drought Atlas[e]
Sedimentation Rate	Xu Jiongxin
Historical Census Data	Harvard ChinaMap

[a]Fairbank Center et al., "China Historical GIS."
[b]Berman, "China River Basins."
[c]Horne, "Map Tiles"; NASA, SRTM.
[d]Chen, "Course Change Shapefiles." Thank you to Chen Yunzhen for sharing this data with me and allowing me to use it.
[e]Cook et al., "Asian Monsoon Failure and Megadrought." Thank you to Amy Hessl for preparing this data and discussing it with me.

The initial design of the database remained largely unchanged throughout this process. The database was designed to have a Places table that would serve as a digital gazetteer for the project, which would include a single identifier, representative location, and source, and a Chinese-language title that originated from either the source material or from a modern place-name. Places associated with specialized or more verbose data would be linked to other tables that held that data, as the Places table was designed to focus on the smallest number of common elements between the various data sources.

The Events table was another key component of the database design. Just like entries in the Places table, each individual event needed to have its own unique identifier, a date, and a link to any number of types describing the event. Because multiple types could be assigned to any individual event, the types themselves need to reside in another table that expresses the relationship between an event and any number of event types. These types were themselves described in another table, as were the historical sources that describe each event. These are also in a many-to-many relationship, as any event may be described in multiple sources, and many events are listed in each source.

When Horne began his work, most of the data sets for this project were in the form of individual spreadsheets with their own internal schemas for generating IDs. This made a simple compilation of data impossible. Furthermore, the collected data, fields, and formats differed substantially between the upriver data (a collection of named places and dates at which they were attested) and the downriver data (a collection of historical events that transpired at particular places). Therefore, Horne created a series of staging tables to import data and align new standardized IDs with original ones while preserving the relationships specified in the tables themselves. He imported each data table into the system, disassembled it into its various relationships and components, and then entered it into the new database schema. This required significant trial and error, as some of the spreadsheets had conflicting formats within the same table, multiple character encodings, and potential data errors.

A particularly complex operation was sorting and normalizing the upstream data, as many places were present on more than one of the maps and other sources that constituted the data set. Shaobai Xiong disambiguated this part of the data, and Horne then placed it into a staging table where each place entry was assigned a unique ID. At that point he checked each place to see if it was present in multiple sources at different times. If so, the first chronological entry was treated as the "primary" place, with the information from other sources placed in another table and linked accordingly. The final list of upstream places served as the core table of places, to which downstream places were added later.

In addition to the Places table, Horne constructed a new version of the Events table and established a many-to-many relationship with the Places table, as any event can have an impact on multiple places, and any place can be the subject of multiple events. Once again, he used a staging table to ensure that each events entry was comparable with the new database schema, and that the old linkages from the individual spreadsheets were preserved with the appropriate new IDs. He also linked each event to the table housing the different primary sources describing it and linked each upstream place to a table containing all its historical information. All the linkages are treated as foreign keys in the database, which ensures that each linked value actually exists in its respective table.

The result is a stable system that allows for visualization via simple PostGIS query and in which all events are associated with geolocated places. All the data from the various spreadsheet inputs has been normalized with sanitized and stable IDs and can be linked to external Linked Open Data resources.

WHAT'S IN HERE?

By the numbers, the TYDA database includes 3,657 unique places, of which 1,642 are downstream and 2,015 are upstream. There are 2,533 total upstream attestations in the data, but many places occur more than once in the data set. There are 3,754 unique events. Two of these lack source connections, so there are actually 3,752 unique event IDs in the source connections. Mostern began this project in part to test and validate *The Yellow River Annals*, which seems to have remained unanalyzed since its publication in the 1930s despite its importance to the study of Yellow River history and even to public policy and cultural imagination about the river. We can now report that the *Annals* appears 2,139 times in the TYDA, and it serves as a source for 2,095 unique events of the 3,754 events recorded in the TYDA. (The difference between those numbers represents forty-four cases in which the TYDA merges multiple events from the *Annals* into a single event.) Among these, there are 482 instances in the TYDA in which the *Annals* and other sources describe the same event, among which are 453 distinct events. The *Annals* is therefore the sole source for 1,642 events in the TYDA.

There are also 2,068 instances in the TYDA for which the *Annals* is not a source, representing 1,657 unique events that exist in the TYDA but not in the *Annals*. This means that more than 44 percent of the events in the TYDA are not visible in the *Annals*. In short, although the *Annals* remains an important source for understanding flooding and flood control on the imperial Yellow River floodplain, it should no longer be considered comprehensive.

Specifically looking at data about floods and course changes, 667 instances in the TYDA appear in the *Annals,* representing 628 unique events. Of these, 419 instances (403 events) appear in no other sources and 248 instances (225 events) appear in at least one other source. Among the 1,645 (1,032 events) instances of floods and course changes that appear in other sources, 1,069 (807 events) do not appear in the *Annals*, and 576 (225 events, the same set as above) include the *Annals* among their sources. Here, too, the TYDA is a more comprehensive source than the *Annals.* It should also be clear from this discussion that it is always possible to query the TYDA for results from any particular historical compilation: for purposes of comparison like this one, or because a particular TYDA user may find a particular part of its source base particularly interesting or reliable. The TYDA retains all the information needed to make such selections possible.

CREATING A DIGITAL SYSTEM

Although the TYDA database is a comprehensive and powerful resource, there was still a need to visualize the data, to create printable maps and charts for this publication, and to enable nontechnical specialists to query and interact with the data. In the summer of 2019, Horne and Mostern decided that using the Python programming language through Jupyter Notebooks (jupyter.org) provided the best mix of power, flexibility, and usability for the TYDA project.

A Jupyter Notebook is an open-source application that allows a user to run and modify code, embed narrative text, and present data visualizations. By using software libraries like Matplotlib and Pandas/GeoPandas, Horne created new methodologies for generating consistently styled and visually appealing charts and maps. The basic functionality of the Jupyter Notebook for TYDA is provided by almost one thousand lines of code, which represent various function calls and database queries. These are designed to make extremely complex operations simple for any end user. For instance, Horne wrote a query to combine events, places, places to events, event types, floodplain geography, and an arbitrary point and buffer. This makes it possible to find all information about all places impacted by an event of a certain type, between an arbitrary number of years, and within a set distance from a point. A user simply needs to input particular years, distances, and point values into a function call. The final notebook contains more than five hundred such queries and calls. Each individual query can be styled and adjusted as needed. Box A.1 depicts an example of the query logic in Jupyter, and Figure A.1 depicts a query call.

Box A.1

A Query to the Tracks of Yu Digital Atlas

This generic query combines the events, places, places to events, event types, floodplain geography, and an arbitrary point and buffer. It generates all information about places that were affected by an event of a certain type, between an arbitrary number of years, and within a set distance from a point. Users need simply input the years, distance, and point values of interest into a function call. The final notebook contains more than five hundred such queries and calls, with each individual query capable of being styled and adjusted as needed.

```
def makeManInfoCount(startdate, enddate, category, area):
    sql ="""
SELECT
  c.events_id,
  c.w_date,
  c.events_type,
  d.type,
  c.events_old_name_c,
  c.events_old_name_p,
  c.places_c_title,
  c.places_en_tite,
  c.places_current_loc,
  c.places_hz_title,
  c.places_py_title,
  c.places_id,
  c.geom
FROM
  (
    SELECT
      a.events_id AS events_id,
      a.w_date,
      a.type_c AS events_type,
      a.old_name_c AS events_old_name_c,
      a.old_name_p AS events_old_name_p,
      b.c_title AS places_c_title,
      b.en_tite AS places_en_tite,
```

Box A.1 (continued)

```
      b.current_loc AS places_current_loc,
      b.hz_title AS places_hz_title,
      b.py_title AS places_py_title,
      b.geom AS geom,
      a.place_id AS places_id
     FROM
      (
       SELECT
        events.old_name_c,
        events.old_name_p,
        places_to_events.place_id,
        events.type_c,
        events.w_date,
        events.id AS events_id
       FROM
        events
        LEFT JOIN places_to_events ON events.id = places_
     to_events.event_id
       WHERE
        events.id IN (
         SELECT
          events_to_type.id
         FROM
          events_to_type
          JOIN event_types ON event_types.ch_title =
     events_to_type.type_c
         WHERE
          event_types.en_cat = {category}
         GROUP BY
          (events_to_type.id)
        )
      ) a
      LEFT JOIN (
       SELECT
        places_to_events.place_id,
        places.c_title,
```

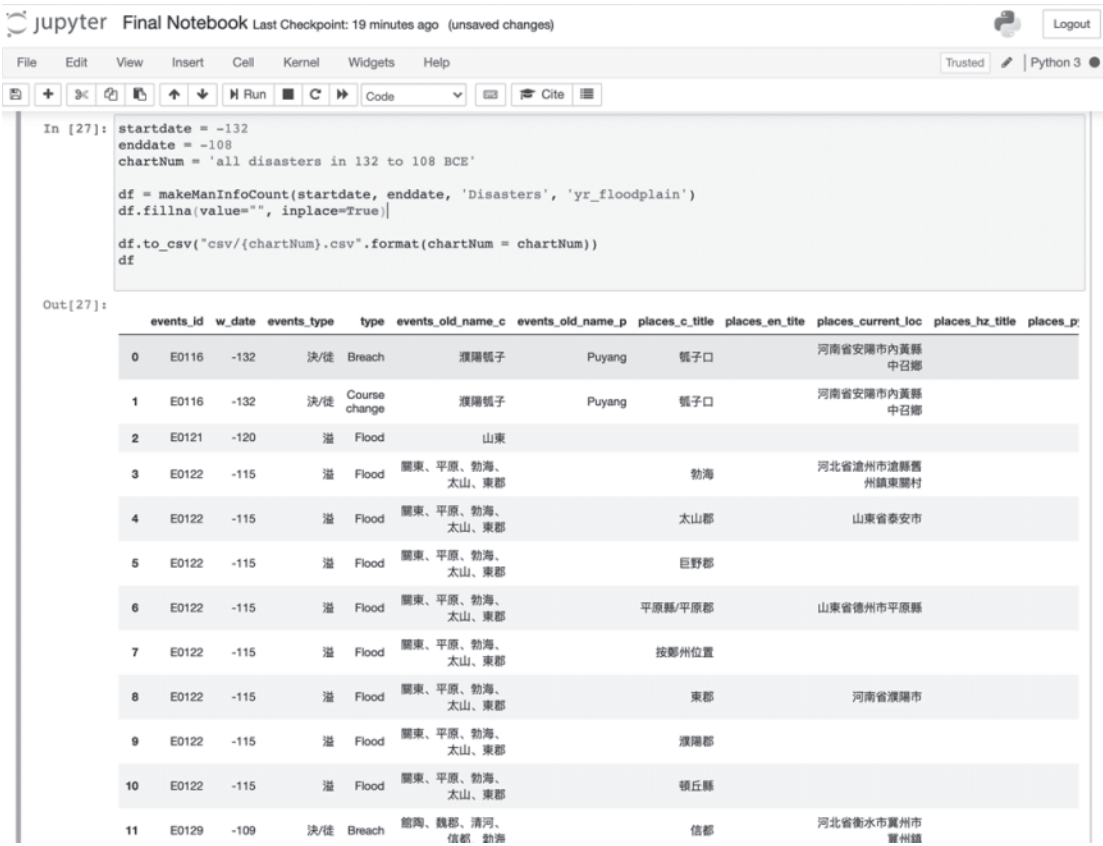
```
          places.en_tite,
          places.current_loc,
          places.hz_title,
          places.py_title,
          places.geom
        FROM
          places
          LEFT JOIN places_to_events ON places.id = places_
    to_events.place_id
        WHERE
          st_intersects(
            places.geom,
            (
              SELECT
                ST_Union(
                  ST_Buffer({area}.geom :: geography, 180000)::
    geometry
                )
              from
                {area}
            )
          ) = true
      ) b ON a.place_id = b.place_id
    WHERE
      a.w_date BETWEEN {startdate}
      AND {enddate}
  ) c
  JOIN (
    SELECT
      event_types.type,
      events_to_type.id AS event_id
    FROM
      events_to_type
      JOIN event_types ON event_types.ch_title = events_to_
    type.type_c
    WHERE
      event_types.en_cat = {category}
```

Box A.1 (continued)

```
    ) d ON c.events_id = d.event_id
   GROUP BY
    c.events_id,
    c.places_c_title,
    c.places_en_tite,
    c.places_current_loc,
    c.places_hz_title,
    c.places_py_title,
    c.places_id,
    c.w_date,
    c.events_type,
    c.events_old_name_c,
    c.events_old_name_p,
    d.type,
    c.geom
   ORDER BY
    c.w_date ASC=""".format(startdate = startdate, enddate =
     enddate, category=category, area=area)
     dfManInfo = pd.read_sql_query(sql, cnx)
     return dfManInfo
```

The TYDA Jupyter Notebook was a shared resource between Mostern and Horne. This allowed us to jointly track data requests and tweaks to the different visualizations and to make code modifications that could be run and analyzed in real time. Using this system, Mostern was able to change queries, request different permutations of visualizations, and ask new questions of the data, to which Horne could respond quickly by simply modifying the inputs of different Jupyter functions. Using the GeoPandas library, these queries could be exported into formats that were readable by the QGIS geographic information system application, where Horne could apply a "house style" and prepare maps for publication.

Although popular in the field of data science, the use of Jupyter Notebooks in humanities research and computing is relatively new, although the Programming Historian website added a lesson about



jupyter Final Notebook Last Checkpoint: 19 minutes ago (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run | Code | Cite

In [27]:

```
startdate = -132
enddate = -108
chartNum = 'all disasters in 132 to 108 BCE'

df = makeManInfoCount(startdate, enddate, 'Disasters', 'yr_floodplain')
df.fillna(value="", inplace=True)

df.to_csv("csv/{chartNum}.csv".format(chartNum = chartNum))
df
```

Out[27]:

	events_id	w_date	events_type	type	events_old_name_c	events_old_name_p	places_c_title	places_en_tite	places_current_loc	places_hz_title	places_p:
0	E0116	-132	決/徙	Breach	濮陽瓠子	Puyang	瓠子口		河南省安陽市內黃縣中召鄉		
1	E0116	-132	決/徙	Course change	濮陽瓠子	Puyang	瓠子口		河南省安陽市內黃縣中召鄉		
2	E0121	-120	溢	Flood	山東						
3	E0122	-115	溢	Flood	關東、平原、勃海、太山、東郡		勃海		河北省滄州市滄縣舊州鎮東關村		
4	E0122	-115	溢	Flood	關東、平原、勃海、太山、東郡		太山郡		山東省泰安市		
5	E0122	-115	溢	Flood	關東、平原、勃海、太山、東郡		巨野郡				
6	E0122	-115	溢	Flood	關東、平原、勃海、太山、東郡		平原縣/平原郡		山東省德州市平原縣		
7	E0122	-115	溢	Flood	關東、平原、勃海、太山、東郡		按鄭州位置				
8	E0122	-115	溢	Flood	關東、平原、勃海、太山、東郡		東郡		河南省濮陽市		
9	E0122	-115	溢	Flood	關東、平原、勃海、太山、東郡		濮陽郡				
10	E0122	-115	溢	Flood	關東、平原、勃海、太山、東郡		頓丘縣				
11	E0129	-109	決/徙	Breach	館陶、魏郡、清河、信都、勃海		信都		河北省衡水市冀州市冀州鎮		

FIGURE A.1
Query Call in Jupyter. The Jupyter collaboration environment allows for collaborative editing of code. It is possible to generate tables and maps live in the application. This image also depicts the complexity of the queries that lie behind the images and assertions in this book.

Jupyter Notebooks at the end of 2019.[9] The power and flexibility of this system enabled the efficient creation of compelling images and opened up new queries that combined spatial reasoning, data points, and historical analysis. Although software installation was not trivial, it would simply not have been possible to manage the sheer volume of tables, charts, maps, and queries for this book without the use of this system. Even visualizations that did not make it into this volume proved their worth as insights into the structure and contents of the database.

CONCLUSION

In the end, it required almost a decade of effort to complete the TYDA, with support from multiple funding bodies, a great deal of

TABLE A.2

Event Types in the TYDA Database and Their Most Common Synonyms in the Data Source

Chinese	Pinyin	Event type	Description
Disasters			
溢	*yi*	Flood	Floods may be signaled by the presence of terms such as *zhang* 漲 (raised water level) or *yan* 淹 (flood). *Daoguan* 倒灌 (water flowing against the current into a city or neighborhood) is another term that indicates the presence of a flood event.
决	*jue*	Breach	A synonym for *jue* is *beng* 崩 (a sudden opening). *Kui* 潰 is also a synonym for *jue* 决. A breach event (*jue* 决) occurs at a single location as a result of the rupture of a dike, whereas a flood event (*yi* 溢) involves a larger region.
旱	*han*	Drought	A drought may also be described by the synonyms *ganhe* 乾涸, *he* 涸, or *ku* 枯.
災	*zai*	Natural disaster	The natural disasters described as *zai* 災 in the sources are rainstorms, earthquakes, landslides, and locust swarms. If a rainstorm results in a flood, these are indicated in our database as two distinct events.
毀	*hui*	Anthropogenic disaster	This refers to an intentional dike breach, which may be performed by the government, usually during wartime, as a military tactic; by private individuals without approval for purposes of irrigation; or by mistake in a failed attempt to rectify an engineering error.
險	*xian*	Risky situation	The term *xian* 險 often co-occurs with the term *shixiu* 失修 (damage or disrepair), which is used to refer to poorly maintained waterworks. If a situation of disrepair becomes severe, it is referred to as a Risky situation.
兆	*zhao*	Omen	When the river is described as inexplicably running clear or, occasionally, dirty, this is referred to as an Omen.

Chinese	Pinyin	Event type	Description
徙	*xi*	Course change	This event type may also be signaled by the term *duo* 奪 (literally to seize), which indicates that one river has taken the course of another river or canal. Other specific types of course change are *fen* 分 (a river has divided in two) or *he* 合 (two rivers have come together as one). A Course change is the result of a Flood that occurs as a result of a Natural disaster or a disaster caused by humans.
絕	*jue*	River runs dry	This event type is also referred to by the synonym *duanliu* 斷流 (a break in the course). It is generally associated with Drought.
塞	*sai*	Blockage	The term *sai* 塞 often co-occurs with the word *yu* 淤 (silt), and *yusai* 淤塞 often appears as a compound term. *Yuzu* 淤阻 (silt obstruction) is another term for this event. *Se* 澀 is another synonym. In addition to silt, ice can also cause a blockage.

Management Events

Chinese	Pinyin	Event type	Description
議	*yi*	Proposal and discussion	This event type refers to several related activities: *yilun* 議論 (discussion), *jianyi* 建議 (a proposal), and *zouyi* 奏議 (a memorial that includes a proposal). Once a proposal is approved, it generally triggers a management event (*zhi* 治). Discussion and management are often co-occurring event types. The publication of a book or essay about the river is also described in the proposal and discussion event type.
遷	*qian*	Settlement relocation	A settlement relocation event occurs whenever people give up their residence. The term may refer to the relocation of a settlement or to the complete destruction or disappearance of a settlement. If a settlement ceases to exist altogether, it is generally referred to as *pi* 圮, or less commonly as *mo* 没. Settlement relocation is often indicated by references to the presence of refugees (*zaimin* 災民).

TABLE A.2 (*CONTINUED*)

Management Events			
放	*fang*	Dam or sluice opening	This event type is most commonly evoked in the Qing dynasty. The event is also referred to as *kaiba* 開壩 or *qiba* 啟壩.
疏	*shu*	Dredging	Dredging is the management event that rectifies the river event of blockage (*sai* 塞). It is synonymous with *jun* 浚 (dredging) or *fangyu* 放淤 (removal of silt). The term *shu* 疏 always refers to silt dredging. Other events associated with expanding water flow are covered by the dam or sluice opening event.
助	*zhu*	Using water for a benefit	This event type is the antonym of the disaster caused by humans in the river events. It refers to the use of river water for agriculture and transportation. In particular it encompasses sanctioned private actions that would generally be considered management (*zhi* 治) if they were government actions.
治	*zhi*	Management	Management is a catchall term that refers to events that cannot be specified more precisely from the source material or to complex acts of water conservancy with multiple components that cannot be separated into distinct events. The term generally refers to acts of policy implementation or policy enforcement. For instance, management of the river may mean that the emperor enforced a law, requisitioned a budget, appointed a supervisor, raised tax revenue, toured a site, or issued an edict (*zhao* 詔). Management may also refer to an engineering act intended to enhance the infrastructure system, such as opening a canal, building a dike, and repairing human damage and/or structures. However, it never refers to emergency repairs. Acts of management may be successful or unsuccessful. Management always refers to government activity, never to private activity.

Management Events			
修	*xiu*	Repair of structures	This refers to any kind of repair to the river infrastructure, such as repairs to dikes or embankments, and any efforts intended to return the river to an old course (*gudao* 故道 or *jiudao* 舊道). A synonym for *xiu* 修 is *fujian* 復建 (rebuilding).
救	*jiu*	Emergency repairs	An emergency repair is an attempt to respond quickly to a flood (*yi* 溢), a breach (*jue* 决), a settlement relocation (*qian* 遷), a blockage (*sai* 塞), a human disaster (*hui* 毀), a natural disaster (*zai* 災), or a risky situation (*xian* 險). The presence of terms such as *ji* 即 (promptly) and *xuan* 旋 (immediately) distinguish an emergency repair from a repair of structures. *Sai* 塞 as a verb rather than a noun refers to an intentional emergency blockage of a breach and is a kind of emergency repair.
建	*jian*	New construction	New construction can refer to construction of the components of the river management infrastructure: dikes (*di* 堤), dams (*ba* 霸), or canals. A synonym for *jian* is *zhu* 筑.
探	*tan*	Scientific survey	This event type, attested beginning in the nineteenth century, refers to surveys and field trips conducted in the context of modern science and engineering techniques.

expert assistance, and untold hours of student and professional labor. Like many data-intensive digital history projects that are based on content of diverse formats from heterogeneous sources and compiled by many hands, this project evolved slowly from a wide assortment of poorly cross-referenced spreadsheets and shapefiles. It would have been ideal to have designed a database structure early on, to have linked every location in it with a controlled vocabulary of counties and prefectures from CHGIS, and to have created a system for data entry that would have generated consistent ID formats. However, that would have required Mostern to make binding semantic decisions early in the project, long before she was prepared to do so. Like many such endeavors, this one proceeded stepwise, with data development and historical insights continuously informing one another. In the end, this completed system allows for a unique approach to historical research, one that merges close readings of individual historical texts and scrutiny of particular events with epochal scale conclusions about landscape and environmental change.