Python Program to Input, Sort, View, and Store Detrital Zircon Geochronological Data in a SQL Database

BURGES, Jarrod[1,2*], METCALF, Kathryn[1], GOFMAN, Mikhail[2]
Department of Geological Sciences, California State University, Fullerton[1], Department of Computer Science, California State University, Fullerton[2]

Email: jburges@csu.fullerton.edu* | Github: jarrod-burges*

GEOLOGICAL SCIENCES CSUF

CALIFORNIA STATE UNIVERSITY FULLERTON COLLEGE OF ENGINEERING AND COMPUTER SCIENCE

The Problem

Many geologists struggle with the task of manually organizing the exponentially growing detrital zircon datasets used to constrain the age and provenance of rocks

Detrital Zircon

- Index mineral to determine age signatures in a rock sample
- Zircons extracted from sample by using standard methods of mineral isolation
  - Crushing, density, and magnetic separation
- Mass spectrometry analysis produces large datasets of U-Pb geochronologic data, such as isotopic ratios and apparent ages (Ma)
- Analytical methods becoming cheaper to produce, leading to more data being created
  - Approx. $5/analysis for LA-ICP-MS

Issues With Detrital Zircon Formatting

- There is a lack of standardized formatting for lab-produced detrital zircon datasets
- Figs. 1-5 are examples of detrital zircon datasets of Cretacous sandstones from the India-Asia suture zone in southern Tibet, China
- If a standardized structured format to display data would arise already published data would be incompatible
- Unlikely an approved format would be adopted by all labs worldwide
  - Seen in Fig. 1 & 5, datasets produced at the same university differ

Analysis	CORRECTED RATIOS									CORRECTED AGES (Ma)								Disc.	Best	1δ
NO.	$^{207}Pb/^{206}Pb$	1δ	$^{207}Pb/^{235}U$	1δ	$^{206}Pb/^{238}U$	1δ	$^{208}Pb/^{232}Th$	1δ		$^{207}Pb/^{206}Pb$	1δ	$^{207}Pb/^{235}U$	1δ	$^{206}Pb/^{238}U$	1δ	$^{208}Pb/^{232}Th$	1δ	(%)	ages	
	Turbidite in Section A 08LGC01																			
08LGC01-47	0.05056	0.0023	0.08742	0.00387	0.01254	0.00021	0.00273	0.00017		221	71	85	4	80	1	55	3	6.25	80	1
08LGC01-49	0.04523	0.00177	0.07843	0.00302	0.01258	0.00019	0.00283	0.00012		-8	51	77	3	81	1	57	2	-4.94	81	1

Figure 1 LA-ICP-MS formed detrital zircon U-Pb data produced at State Key Laboratory of Mineral Deposits Research, Nanjing University, China with isotropic ratios, sigma values, apparent ages, discordance, and best ages from An et al. (2018)

06T01 lithic arkose

Isotope ratios								Age estimates (ma)								
Analysis_#	$^{207}Pb/^{206}Pb$	1σ	$^{207}Pb/^{235}U$	1σ	$^{206}Pb/^{238}U$	1σ	rho	$^{207}Pb/^{206}Pb$	1σ	$^{207}Pb/^{235}U$	1σ	$^{206}Pb/^{238}U$	1σ	age	1σ	% disc
06T01-01	0.04715	0.00201	0.0567	0.00263	0.00872	0.00027	0.00236	56.3	99.08	56	2.52	56	1.74	56	1.74	0
06T01-02	0.04831	0.0014	0.09704	0.00339	0.01457	0.00044	0.00295	114.5	67.13	94	3.13	93.2	2.82	93.2	2.82	1
06T01-03	0.05399	0.00124	0.11505	0.00348	0.01546	0.00047	0.00301	370.6	50.9	110.6	3.17	98.9	2.97	98.9	2.97	11
06T01-04	0.0488	0.00161	0.10121	0.00385	0.01504	0.00046	0.00339	138.2	75.66	97.9	3.55	96.3	2.94	96.3	2.94	2
~~06T01-05~~	~~0.0767~~	~~0.00216~~	~~0.08891~~	~~0.00302~~	~~0.00841~~	~~0.00026~~	~~0.00276~~	~~1113.5~~	~~55.27~~	~~86.5~~	~~2.82~~	~~54~~	~~1.65~~	~~54~~	~~1.65~~	~~38~~

Figure 2 Standard techniques of seperation created by Strong and Discoll (2016) formed detrital zircon U-Pb data produced at Langfang Laboratory of Beijing University with isotropic ratios, sigma values, apparent ages, sigma values, discordance, and best ages with strike ommited data from Atichison et al. (2011)

Analysis	U	206Pb	U/Th	206Pb*	±	207Pb*	±	206Pb*	±	error	206Pb*	±	207Pb*	±	206Pb*	±	Best age	±	Conc
	(ppm)	204Pb		207Pb*	(%)	235U*	(%)	238U	(%)	corr.	238U*	(Ma)	235U	(Ma)	207Pb*	(Ma)	(Ma)	(Ma)	(%)
6.7.12.1KM-58 <>	161.833951	14503.4111	0.86340944	20.9356298	8.55500765	0.22046587	8.6468657	0.03347543	1.25703246	0.14537435	212.263161	2.62476656	202.297451	15.8613326	87.6916388	203.13789	212.263161	2.62476656	NA
6.7.12.1KM-64 <>	133.75583	9228.95052	1.04384768	19.6711539	14.6133905	0.23494203	14.7637803	0.03351886	2.10191032	0.14236938	212.534095	4.39443784	214.270223	28.526985	233.416522	338.861713	212.534095	4.39443784	NA

Figure 3 LA-ICP-MS formed detrital zircon U-Pb data produced at Arizona LaserChron Cente with element ppm, isotropic ratios, sigma values as a percent, error correction, apparent ages, discordance, and best ages with NA data from Metcalf and Kapp (2019)

Sample	Latitude	Longitude	Elevation (m)	Unit	Type
6.7.12.1KM	29.14381	87.61359	4058	Tethyan strata	bed
6.7.12.3KM	29.17129	87.60725	4633	Tethyan strata	bed
6.7.12.4KM	29.18911	87.60182	4923	Tethyan strata	bed

Figure 4 Sample location data in same workbook as Figure 3 but in different sheet Metcalf and Kapp (2019)

Spot No.	Th	U	Th/U	Isotopic ratios						Age (Ma)						Disc (%)	Accepted age (Ma)	
	ppm	ppm		$^{207}Pb/^{206}Pb$	1σ	$^{207}Pb/^{235}U$	1σ	$^{206}Pb/^{238}U$	1σ	$^{207}Pb/^{206}Pb$	1σ	$^{207}Pb/^{235}U$	1σ	$^{206}Pb/^{238}U$	1σ		Age	1σ
TL18-12ZK Quartz schist. (GPS: N29° 0'38.00", E93° 0'2.00")																		
TL18-1262	704	257	2.74	0.10201	0.00156	3.89966	0.06748	0.27723	0.00395	1661	29	1614	14	1577	20	5%	1661	29
TL18-1263	39.1	19.8	1.52	0.5326	0.27143	1.33483	0.07836	0.01815	0.00485	4336	1597	861	252	116	31	642%		
TL24-01ZK Quartz schist. (GPS: N29°59'4.00", E93° 9'22.00")																		
2401ZK01	48.8	79.5	1.63	0.04925	0.00664	0.10885	0.01446	0.01603	0.00047	160	275	105	13	103	3	2%	103	3
2401ZK2	159	295	1.86	0.04991	0.00287	0.11305	0.00641	0.01643	0.00033	191	132	109	6	105	2	4%	105	2

Figure 5 LA-ICP-MS formed detrital zircon U-Pb data produced at State Key Laboratory of Mineral Deposits Research, Nanjing University element ppm, isotropic ratios, sigma values, apparent ages, discordance, and accepted or best ages with multiple samples in same file separated by row headers. Sample headers also contain identification, rock classification, and GPS location. from Wei et al. (2020)

Formatting cont.

- Seen in Figs. 1-5 Detrital Zircon datasets are dynamic in their data reporting
- Data may not be in the same unit format:
  - Figure 5 as element ratios, Th/U and Figure 3 as U/Th
  - Figure 5 as location data, DMS and Figure 4 as DD
  - Have to convert units to a single type
- Data may also be located in the paper resulting in further disconnect of location of data

Effects on Geochronologists

- Increased time spent organizing detrital zircon datasets
- Lack of a(n):
  - Centralized organized data in a single database
  - Tagging data with cross-referenceable tags
  - Searching based on any variable across all files
  - Multi-variable filtering using logical operators (e.g. AND, OR, or NOT)
  - Quick-view plot from data selected
  - Easy sharing or exporting of data
  - Source tracker

Our Solution

Python application to assist geologists in importing detrital zircon datasets into a localized personal database that can then be manipulated, sorted, filtered, and exported with built-in tools

Importing

- User can import any Excel file (.xlsx, .xls) file into the import wizard
- Import wizard is a step-by-step prompt that assists the user in linking where data is located within the file
  - Such as Best Age, Isotopic Ratios, or Element Concentration
- Sample names must be unqiue within the database
  - Conflicting entries will be prompted with automatic renaming, custom renaming, or skipping that entry
- All linked data is stored into the database
- Imported samples have option to be stored with a linked copy of the source Excel file
- After import, linked settings are stored for future use
- If on next import the Excel file matches a previously imported format, automatically applies linked values to data columns and values
- Prompts user to double check assumed linked values

Sorting

- Tables separate important data
  - Sources, Samples, U-Pb data, Units, Regions, etc.
- In each table, column based ascending/descending
- Default sorting method for all tables can be set or remembered from previous setting
- Single cell editing done from main window UI

Searching

- Search bar at top of table
- Regular expression (regex) matching syntax available to be applied to all data within the table
- Data shown as filtered selection of rows
- Normal editorial data options are available
  - Delete, edit by column, add data to all

Filtering

- Searching is for quick finds of data, not designed for complex searches
- Filtering handles complex logical-operator-based variable searching
  - Multi-variable capable
  - AND, OR, NOT
  - Numerical
    - Equals, Greater than, or Less than
  - Categorical
    - Filter by groups
  - Word/String-based filtering
    - Contains, Match, Starts, or Ends
- Filter results appear in new tab
- Save filters for later viewing

Custom Tagging

- Context Tags used to classify important data
- Customized user specified contexts can be created for Samples, Aliquots, and Spots
- Can be any text length that a researcher wants to tag with
- Tags must be unique
  - Non-case-sensitive, 'Tibet' & 'tibet' not allowed
  - 'Tibet 23' & 'Tibet 22', allowed
- Multiple tags can be applied to any single context
- Any tag can be tied to multiple contexts
  - E.g. 1 Sample - 5 Tags
  - E.g. 5 Samples - 1 Tag
- Works with searching functions
- Tags can be viewed and edited in separate tables

New Labs, Sources, Tags

- New Labs, Sources, and Tags are added in specific user-interfaces with automatic searching of already added info; However, must be unique or will conflict
- See Figs. 6-9

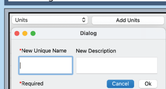

Figure 6 Add new Unit dialog, showing parent unit, unit name, and unit description

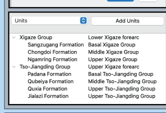

Figure 7 Expandable unit tree-based organization with add unit button

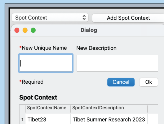

Figure 8 Add new Spot Context dialog, showing custom context tag "Tibet23" and description

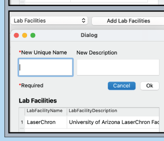

Figure 9 Add new Lab Facility dialog. UPb Data contains what lab produced the data.

Viewing Detrital Zircon Data

- Filtered selection of data can be viewed in program without having to use external tools
- In-program-viewing not meant to replace popular data visualization tools
  - DZStats, IsoplotR, detritalPy, etc
- View by frequency histograms or probability density plots
  - Very basic with little visual customization
  - Not meant to be used as figures
- Save views for later using filters
- Allows quick-view graphs between filtered selections of data
- Future plans of adding georeferenced map to sample data
  - Draw transects on map and select all samples within approximate area
  - Lasso-tool-based data selection

Exporting

- Filtered selections of data can be exported as:
  - Excel (.xlsx, .xls)
  - detritalPy
  - DZStats, etc.
  - HafniumPlotter
- Options to include extra data for exporting as Excel:
  - Sample
  - UPb Data
  - Unit Descriptions
  - Context Tags
  - Sources
  - Settings
- Share and merge database
  - Database file format (.db)
  - Easy one-click importing
    - Conflict resolution prompt
  - Easy collaboration between researchers
- Split database
  - Remove selected data from current database; save into new database
  - Allows for easy management of database files

Current Workarounds

- Online databases such as GeoChron.org and United States Geological Survey (USGS)
- These databases are incomplete and USGS is US-based data only
- Cannot easily import new data without manual organizing and pending approval
- Difficult to incorporate other researchers' data
- Cannot customize the data with user inputted tags
- Online databases or a filtered subset is not easily shared and associated metadata may be lost
- Most researchers use localized Excel based methods of organization

Development

- Developed on Python (Van Rossum and Drake, 2009) and SQL (ISO) with library packages PyQt6 (UI development tool) and SQLite3
- Coded using Pycharm and DataGrip integrated development environment (IDE)
- Maintained on GitHub, popular online version management tool
  - Allows for simultaneous editing of code and tracker
  - Repository holds code; currently made private
  - Utilizes popular version control program, git
- Coding is done by Jarrod Burges and Kate Metcalf
  - Mikhail Gofman from CSUF Computer Science Department acting as advisor to Jarrod Burges

Target Release: Dec 2023

- Scan QR code below to be notified of public alpha release
- Suggestions or input are welcomed
- GitHub's issue tracker:
  - Track any issues/bugs with application, user submitted
  - Feature-enchancement tickets can be submitted
  - Pull-request system for coders to help contribute with development

Want to know when program is released? Or Feedback/Suggestions?

Microsoft Forms:
- Name
- Email
- University/Affiliation (optional)
- Feedback/Suggestions (optional)

No sign-in required
No extra data is collected

Artificial Intelligence (AI)

- Although a useful tool, not a solution for this problem since:
  - Lack of easily saving data
  - Criticalerrors occur
  - Incomplete responses
    - 31 analysis records were given to Chat GPT from Atichison et al. (2011) a summery of column names, total age range for 15 analysis were given as a response
  - Not able to add extra information/metadata
- Where AI is promising and to look forward to is:
  - Data visualization tools, seen in Excel's chart design
  - Automatic generation of long text descriptions or expansion of
  - Automatic data relevance grouping (E.g. filters)

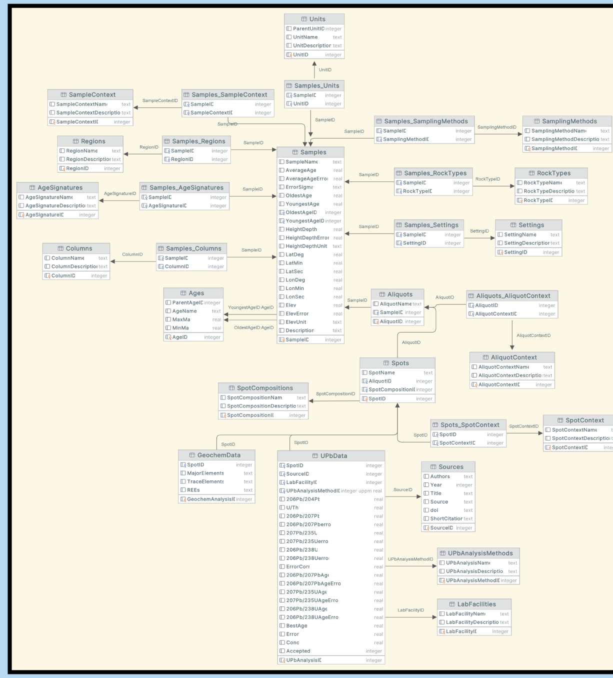

Figure 10 SQL Database Schema made using DataGrip (Jetbrains, 2023, DataGrip). Schema is the design of relationships between tables in the database. Comprised of 31 tables, 18 tables contain main user viewed data, 10 tables are many-to-many relational tables between main tables, 3 tables are Context Tags for Samples, Aliquots, and Spots. Arrows with labels display what variable relates two tables together.

Database Schema

- SQL based hierarchical relational model or commonly known as tree data structure
  - Samples Table is the root of the tree
    - Sample Name and other name/word-based variables must be unique to solve user-side confusion
      - E.g. Aliquot name per Sample, but Primary ID keys solve this, but leads to seemingly duplicate data being shown to user
- Utilizes one-to-many relationship (see Figure 10 UPbData/UPbAnalysisMethods)
  - Records of UPbData are unique, method of analysis is finite
- Utilizes many-to-many relationship (see Figure 10 Samples/SampleContext)
  - Allows for one Sample ID to be tied to multiple Context Tags or the inverse
  - For best database practice, split into two one-to-many relationships
    - Many-to-many relational table created to link tables, seen as Samples_Context
      - Bi-directional model
    - Samples table is one-to-many with Samples_Context
    - SampleContext table is one-to-many with Samples_Context

Future Work

- Database schema designed to allow for expansion of methods
- Geochemical Data
  - Major Elements
  - Trace Elements
  - Rare-Earth Elements
  - Quick view and export to external data visualization tools
- Future datasets:
  - Other mineral and dating methods
  - Petrography
- Groundwork done for stratigraphy
  - Columns
  - Column Depth
- Add files to data records
  - Images (maps, rocks, regions)
  - Pdfs, Documents, etc.
  - Suggestions?

References

- Aitchison, J.C., Xia, X., Baxter, A.T., and Ali, J.R., 2011, Detrital zircon U–Pb ages along the Yarlung-Tsangpo suture zone, Tibet: Implications for oblique convergence and collision between India and Asia: Gondwana Research, v. 20, p. 691–709, doi:10.1016/j.gr.2011.04.002.
- An, W., Hu, X., and Garzanti, E., 2018, Discovery of Upper Cretaceous Neo-Tethyan trench deposits in south Tibet (Luogangcuo Formation): Lithosphere, v. 10, p. 446–459, doi:10.1130/L690.1.
- ISO, ISO/IEC 9075–1:2016, Information technology - Database languages - SQL - Part 1: Framework (SQL/Framework)., DOI: https://doi.org/www.iso.org/standard/63555.html.
- JetBrains, 2023, DataGrip: JetBrains: https://www.jetbrains.com/pycharm (accessed 2022-2023).
- JetBrains, 2023, PyCharm: JetBrains: https://www.jetbrains.com/pycharm (accessed 2022-2023).
- Metcalf, K., and Kapp, P., 2017, The Yarlung suture mélange, Lopu Range, southern Tibet: Provenance of sandstone blocks and transition from oceanic subduction to continental collision: Gondwana Research, v. 48, p. 15–33, doi:10.1016/j.gr.2017.03.002.
- Strong, T.R., and Driscoll, R.L., 2016, A process for reducing rocks and concentrating heavy minerals: U.S. Geological Survey Open-File Report 2016–1022, 16 p., http://dx.doi.org/10.3133/ofr20161022.
- Van Rossum, G., and Drake, F.L., 2009, Python 3 Reference Manual: Scotts Valley, CA, CreateSpace, 242 p.
- Wei, Z., Li, X., Li, Y., Fan, X., Wang, J., Zhang, C., and Jiebu, G., 2020, Discovery of Vestige Sedimentary Archives of the India-Asia Collision in the Eastern Yarlung Zangbo Suture Zone: Journal of Geophysical Research: Solid Earth, v. 125, p. e2019JB018192, doi:10.1029/2019JB018192.