Geochemistry Geophysics Geosystems G3

Article
Volume 12, Number 6
24 June 2011
Q0AA19, doi:10.1029/2010GC003479
ISSN: 1525-2027

Published by AGU and the Geochemical Society

Engineering cyber infrastructure for U-Pb geochronology: Tripoli and U-Pb_Redux

J. F. Bowring
Department of Computer Science, College of Charleston, 66 George Street, Charleston, South Carolina 29424, USA (bowringj@cofc.edu)

N. M. McLean and S. A. Bowring
Department of Earth, Atmospheric, and Planetary Sciences, MIT, Building 54-1126, Cambridge, Massachusetts 02139, USA (nmclean@mit.edu; sbowring@mit.edu)

[1] In the past decade, major advancements in precision and accuracy of U-Pb geochronology, which stem from improved sample pretreatment and refined measurement techniques, have revealed previously unresolvable discrepancies among analyses from different laboratories. One solution to evaluating and resolving many of these discrepancies is the adoption of a common software platform that standardizes data-processing protocols, enabling robust interlaboratory comparisons. We present the results of a collaboration to develop cyber infrastructure for high-precision U-Pb geochronology based on analyzing accessory minerals by isotope dilution-thermal ionization mass spectrometry. This cyber infrastructure implements an architecture specifying the workflows of data acquisition, statistical filtering, analysis and interpretation, publication, community-based archiving, and the compilation and comparison of data from different laboratories. The backbone of the cyber infrastructure consists of two open-source software programs: Tripoli and U-Pb_Redux. Tripoli interfaces with commercially available mass spectrometers using standardized protocols, statistical filtering, and interactive visualizations to aid the analyst in preparing raw data for analysis in U-Pb_Redux. U-Pb_Redux implements the architecture by orchestrating the analyst's workflow with interactive visualizations and provides data reduction and uncertainty propagation that support data interpretations. Finally, U-Pb_Redux enables production of publication-ready graphics and data tables, the archiving of results, and the comparative compilation of archived results to support cooperative science.

Components: 10,600 words, 10 figures.

Keywords: cyber infrastructure; geochronology; software; uncertainty propagation; uranium lead.

Index Terms: 1115 Geochronology: Radioisotope geochronology; 1908 Informatics: Cyberinfrastructure; 1920 Informatics: Emerging informatics technologies.

Received 21 December 2010; **Revised** 23 March 2011; **Accepted** 5 April 2011; **Published** 24 June 2011.

Bowring, J. F., N. M. McLean, and S. A. Bowring (2011), Engineering cyber infrastructure for U-Pb geochronology: Tripoli and U-Pb_Redux, *Geochem. Geophys. Geosyst.*, *12*, Q0AA19, doi:10.1029/2010GC003479.

Theme: EarthTime: Advances in Geochronological Technique

1. Introduction

[2] The EARTHTIME (http://www.earth-time.org) initiative began in 2004 as a way facilitate cooperation among geochronologists with the goal of sharing ideas and approaches for eliminating or minimizing interlaboratory and intertechnique biases. A major component of this effort has been the manufacture, calibration, and distribution of a mixed ^{233}U-^{235}U-^{205}Pb tracer, effectively removing tracer calibration as a cause of interlaboratory bias. As a result, scrutiny of precise data sets has revealed that treatment of random and systematic uncertainties is not uniform and that a transparent approach to methods of data acquisition, processing, reduction, and reporting would benefit our science as a whole.

[3] The pioneering paper by *Ludwig* [1980] and the associated software packages such as 'PbDat' [*Ludwig*, 1985] and 'Isoplot' [*Ludwig*, 1991] made available along with it, described methods for the propagation of uncertainties and for plotting data associated with U-Pb ID-TIMS geochronology. Ludwig's treatment has been adopted by most U-Pb geochronologists either by use of 'PbDat' or derivative programs that use the same algorithms. These programs evolved from code written for HP computers in the 1980s to a QUICKBASIC® program to a MSDOS® version. Alternative treatments [e.g., *Davis*, 1982; *Roddick*, 1987] are also used at present, but these are not licensed or open source. Since the publication of these landmark papers, advances in U-Pb geochronology, including sample pretreatment, data acquisition strategies, and use of high-purity tracers that often contain two enriched isotopes (^{202}Pb and ^{205}Pb; ^{233}U and ^{235}U or ^{236}U) necessitate a new approach to data reduction. *Schmitz and Schoene* [2007] presented a modification and expansion of Ludwig's algorithms for error propagation in an Microsoft Excel®–based approach in support of the EARTHTIME initiative.

[4] One outgrowth of the EARTHTIME initiative is to propose the use of a common, open-source software platform or cyber infrastructure for the Earth sciences. An ideal cyber infrastructure would provide standardized data handling and data reduction protocols facilitating interlaboratory comparisons. Furthermore, it is essential that the software and its development processes be completely transparent, able to respond to community input and to serve as a teaching tool. In our experience, many practitioners (especially students), desire a more detailed understanding of all aspects of error propagation. Therefore this software should also function as an interactive tool with all of the relevant equations easily available. Our novel approach to creating this cyber infrastructure is an ongoing and close collaboration between software engineers (CIRDLES, http://cirdles.org) and geochemists (EARTHTIME, http://earth-time.org) that provides continuous, iterative development based on user feedback.

[5] We argue that this proposed cyber infrastructure solution requires several key properties: (1) a cohesive architecture that supports collaborative science by integrating and automating the end-to-end analysis workflow from measurement to archived results, (2) a set of well-defined interfaces supporting interoperability among various mass spectrometers and software packages (3) an architecture that can easily be modified to accommodate new insights and approaches to error propagation, output, and archiving, and (4) a set of software components that replace as much as possible the data reduction calculations done by proprietary mass spectrometer software, provide automated visual assistance for filtering and preparing raw data, use standardized algorithms to perform robust data reduction, provide interactive visualizations to facilitate sensitivity analysis and to aid the interpretation of dates and other results, aid in automating the production of publication-ready graphics and data tables, and the archiving of results, and support the retrieval, compilation, and comparison of analytical results from different sources.

[6] In this paper, we present the results of our ongoing collaboration to develop an exemplar instance of this cyber infrastructure. Our initial system, fulfilling these criteria, is designed expressly for the community of laboratories participating in high-precision U-Pb geochronology of accessory minerals known as isotope-dilution thermal ionization mass spectrometry (ID-TIMS).

[7] The backbone of this cyber infrastructure consists of two open-source software programs: Tripoli and U-Pb_Redux. These programs integrate the requisite scientific processes with dynamic interfaces to data reduction protocols, algorithms, and archived results.

[8] The platform-independent software U-Pb_Redux is central to the new architecture by providing seamless interoperability throughout the data acquisition, interpretation, publication, and archiving processes. U-Pb_Redux orchestrates this analytical workflow with the support of interactive visualiza-

tions. Robust interpretation and intercomparison of these data are facilitated by standardized and open techniques for data reduction and uncertainty propagation [*McLean et al.*, 2011]. For example, the software includes interactive graphical sensitivity testing during data acquisition. Finally, U-Pb_Redux supports the automated production of customizable, publication-ready data tables, concordia plots, and weighted mean plots. It supports community-based archiving including the comparative compilation of archived results.

[9] Tripoli is a Windows®–based program that imports mass spectrometer data files and provides interactive visualizations for data review. The user can reject data based on visual inspection, or by using a statistical outlier detection algorithm. Tripoli also implements a variety of point-by-point corrections for time-dependent mass fractionation for U and Pb using double spikes and isobaric interferences including uranium oxides, thallium, and barium phosphate. The filtered data can be exported to U-Pb_Redux or other analytical programs.

2. Motivation

[10] Geochemists rely on a variety of data reduction software, ranging from proprietary products included with mass spectrometers to specialized Microsoft Excel® spreadsheets shared by the community. In general, individual geochemists develop and maintain these specialized applications to support their own idiosyncratic workflow, and their community often appreciates and adopts these tools. The adopting scientists depend on the developer scientists for the quality of these tools based solely on this ad hoc community consensus.

[11] Our vision is to initiate and develop collaboration among computer scientists and geochemists in an interdisciplinary effort to create data reduction software that is open source. This collaborative effort was in part motivated by the geochemical community's apparent lack of a recognized role for software engineers and for formal software development processes. Software engineering processes and associated techniques can generate robust and tested software systems that support managed change and maintenance activities.

3. Architecture for Cyber Infrastructure

[12] The development of end-to-end data processing systems (raw data to archived analysis results) is an emerging activity in many scientific disciplines [e.g., *National Science Foundation*, 2007]. In this section, we detail our design of an architecture for the ID-TIMS community that can serve as a template for other related communities.

[13] Our architecture encodes the analyst's workflow as stages in data transformation. Each transformation incorporates additional data such as tracer compositions and values for physical constants and introduces new relationships among the data. Automated data handling at each stage requires a logical data structure that supports transmission, manipulation and storage. The design and implementation of this logical structure is commonly known as data modeling [e.g., *Simsion*, 2007].

[14] In sections 3.1–3.3 we first describe the dynamic data flows in our architecture, then we describe data modeling and our specific data models that together represent a static view of our architecture.

3.1. Data Flows

[15] Figure 1 illustrates our architecture with the data flows along directed arcs between the key stages, summarized below.

3.1.1. Sample

[16] The architecture centers on rock samples and derivative aliquots that may range from a whole rock powder to a zircon separate. There are two data flow arrows emanating from the sample stage, one to the sample registry stage, illustrated by System for Earth Sample Registration (SESAR), and the other to the mass spectrometer stage.

3.1.2. Sample Registry

[17] SESAR (http://www.geosamples.org) is an example of a centralized registry that provides unique identifiers for samples: the International Geo Sample Number or "IGSN". SESAR or any other sample registry is independent of the work described herein. Use of a unique sample identifier helps to ensure that all metadata for a sample is globally available.

3.1.3. U-Pb_Redux Workflow Management

[18] Before making and editing isotopic measurements using Tripoli, the analyst uses U-Pb_Redux to organize and name the aliquots and fractions, and to specify various parameters for data reduction such as the tracer and tracer mass. The analyst also specifies the network paths that support real-time

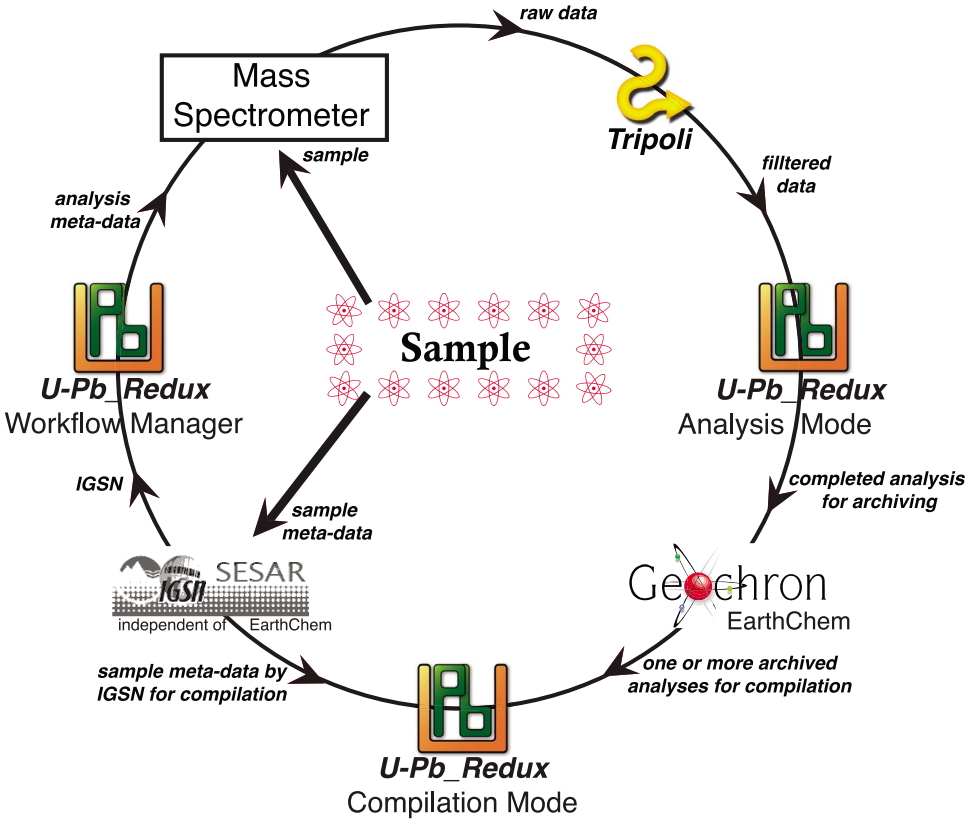

Figure 1. Data flow architecture for U-Pb cyber infrastructure. First, the sample is submitted to a sample registry (e.g., SESAR), which provides a unique sample identifier to the U-Pb_Redux workflow manager. Using the workflow manager, the analyst sets up the sample, aliquots, fractions, and parameters that will be part of the life cycle of this analysis (see section 4.1). Mass spectrometer data files are input to Tripoli for visual inspection, and the filtered data are exported to U-Pb_Redux for analysis (see section 4.2), for production of publication-ready artifacts, and for electronic archiving, e.g., to Geochron (see section 3.1.7). In the compilation stage archived analyses can be combined for analysis (see section 4.3).

interaction between the two programs. The workflow for a sample never terminates, as after the analysis results are archived, they remain available for review, additions and compilation indefinitely. We describe the workflow management mode of U-Pb_Redux in detail in section 4.1.

3.1.4. Mass Spectrometer

[19] The architecture provides an interface to any mass spectrometer capable of producing the requisite raw data, enabling interoperability among different laboratories.

3.1.5. Tripoli

[20] Tripoli imports and processes mass spectrometer raw data files and exports the processed isotopic data to U-Pb_Redux for analysis, as described in detail in section 5.

3.1.6. U-Pb_Redux Analysis Mode

[21] In analysis mode, U-Pb_Redux receives Tripoli-exported data for analysis, and provides full uranium-lead data reduction and uncertainty propagation for any U-bearing phase. U-Pb_Redux outputs aliquot analysis results according to the aliquot schema presented in section 3.3 for export to EarthChem's Geochron database, for example. We describe the analysis mode of U-Pb_Redux in detail in section 4.2.

3.1.7. Geochron

[22] Geochron (http://geochronportal.org) is a community-based database and World Wide Web portal for data access that serves U-Pb geochronology. It is a joint project of EARTHTIME and EarthChem.

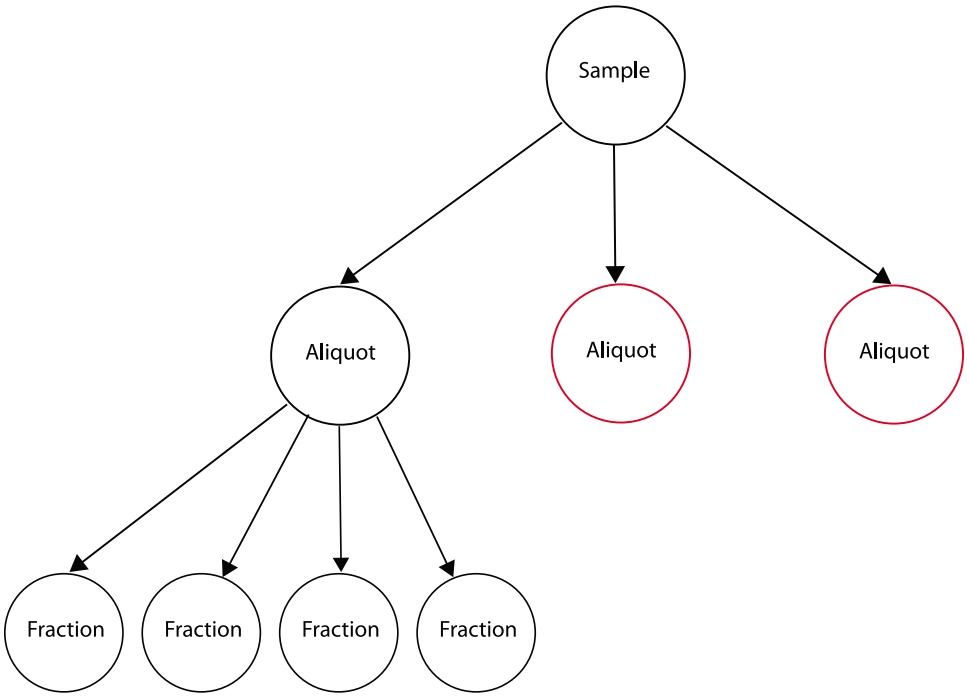

Figure 2. Example of rooted directed tree structure. The root here is a sample with three children, each an aliquot of the sample. The left-hand piece has four children, each a fraction. The red-colored nodes are leaves, which have no children in this illustration. However, each leaf can be replaced with a subtree.

[23] Geochron's database archives completed analyses from a variety of sources as either public or private records. Users can upload or search for and retrieve archived data directly through a Web browser or by using other software that accesses the data with Web services. For example, U-Pb_Redux produces and retrieves archival files for Geochron using its Web services as detailed in section 4. Alternatively, U-Pb_Redux can produce and retrieve archival files from any database designed for this purpose.

3.1.8. U-Pb_Redux Compilation Mode

[24] In compilation mode, U-Pb_Redux can import and compile archived results, allowing the analyst to compare results within and among laboratories. We describe this compilation mode in detail in section 4.3.

3.2. Data Modeling

[25] A data model represents an abstraction of a complex system that, in this case, contains geochemical data and associated analysis workflows.

[26] Data modeling involves specifying a hierarchy of data models, starting with a conceptual data model that defines the mathematical structures organizing the data. The data is specified by a data dictionary that provides a unique identifier and definition for each data element to be encoded in the data models. These elements may include inputs, named intermediate calculated results, outputs, and metadata about data elements and their relationships. Next in the hierarchy is the set of logical data models (LDMs) that implement the conceptual data model by specifying the relationships among the data elements in the data dictionary. At the bottom of the hierarchy, we populate LDMs with actual data to produce defined data models.

[27] Geochronological data have an inherent hierarchical nature. In the case of ID-TIMS, the hierarchy is sample-aliquot-fraction, as explained in section 3.3.

[28] The conceptual data model we create encodes measured isotope ratios and data introduced in our workflow such as physical constants, analytical decisions, and interpretations. The data encoded in any specific analysis workflow will not require every data element specified in the data dictionary. Furthermore, additional data elements may be added to the data dictionary to accommodate new discoveries.

[29] This potential sparseness of encoded data and the possibility of additional defined data elements

fit the definition of semistructured data, which are often best modeled as a rooted directed tree. Figure 2 illustrates a rooted directed tree based on the sample-aliquot-fraction example. From the root node labeled "Sample" emanate a number of directed graph edges, each leading to a child node that represents an aliquot of the sample. Each aliquot child node is either a leaf node (depicted as a red circle), denoting that no fractions were identified for the aliquot, or it is the root of another tree whose child nodes each represent the identified fractions of that aliquot. This tree structure repeats as needed.

[30] One important fact about a directed tree is that there is a maximum of one path between any two nodes. This restriction mirrors the physical world, where each fraction is part of only one aliquot, which is part of one sample. The elegance of this conceptual model allows us to encapsulate data at any level of granularity in a tree structure and to compose more complex trees from other trees. For example, in section 3.3.1 we describe the LDM of a Pb blank. This model of a Pb blank can be inserted into the overall LDM of an aliquot as one child of a node that is itself the root of a subtree containing the various corrections applied during the analysis.

[31] Using this conceptual data model and the data dictionary, we implement a set of LDMs that together include the data elements from the data dictionary and define their interrelationships. Various modeling languages exist for implementing LDMs. We chose the Extensible Markup Language (XML) schema language. XML is an emerging standard for modeling data because it provides for precise declarations of data that are independent of any specific presentation of data, any specific software, or any specific hardware. The World Wide Web Consortium (W3C) defines XML at http://www.w3.org/XML.

[32] XML schema documents are also rooted directed trees and hence a good match for our requirements. In practical terms, extensible means that we can identify our data elements with names from the data dictionary or any other source and then specify the data element relationships using predefined XML constructs such as basic data types.

[33] In the next section, we describe our conceptual data model and data dictionary for all phases of ID-TIMS U-Pb geochronology and show how we implement specific LDMs based on this conceptual data model. Specific examples of defined data models based on the logical models are also presented.

3.3. Data Models for ID-TIMS Geochronology

[34] The first step in designing our data models was to achieve some level of consensus within the ID-TIMS community on a minimal listing of the data elements required for a complete analysis of a sample and a meaningful archiving of that analysis.

[35] We adopted naming conventions to guarantee that each name would be legal in any computer programming language as well as human readable. For example, the name of radiogenic $^{206}Pb/^{204}Pb$ is listed in the data dictionary as "r206_204r", where the leading "r" denotes ratio and the trailing "r" denotes radiogenic. Our current data dictionary is summarized in Appendix A.

[36] An analysis of the data dictionary provided the basic elements of the conceptual data model. We modeled three abstractions: sample, aliquot and fraction.

[37] 1. A sample is a single collection of a geologic material (usually a rock or mineral) from one location. A sample has a lab-specific name and a unique identifier such as the IGSN provided by SESAR (see section 3.1.2).

[38] 2. An aliquot is some derivative of the sample that is analyzed, such as a mineral separate. If a piece of sample is held in reserve for future analysis, it is a unique aliquot. An aliquot is the basic unit of aggregate analysis for a sample that our architecture archives. Aliquots and their associated data analyses and interpretations form the basis for the archival records in our architecture. Aliquot records can be compiled together as described in section 4.3. U-Pb_Redux supports the analysis of multiple aliquots in the same session.

[39] 3. A fraction encapsulates the isotopic data for an analyzed volume of a mineral without being lab specific. For U-Pb, the paired isotopic data are used to calculate a date and uncertainty.

[40] We implemented the LDMs starting with the aliquot because the LDM of the sample is implemented in the sample registry (see section 3.1.2). The LDM of an aliquot is defined as an XML schema at http://earth-time.org/projects/upb/public_data/XSD/AliquotXMLSchema.xsd. Figure 3 displays part of the rooted directed tree structure of the aliquot schema. The red leaf nodes represent actual data dictionary elements. The black nodes represent LDMs for complex entities, such as tracers and Pb blanks.

[41] There are four categories of LDMs currently associated with the LDM of an aliquot: (1) parameter

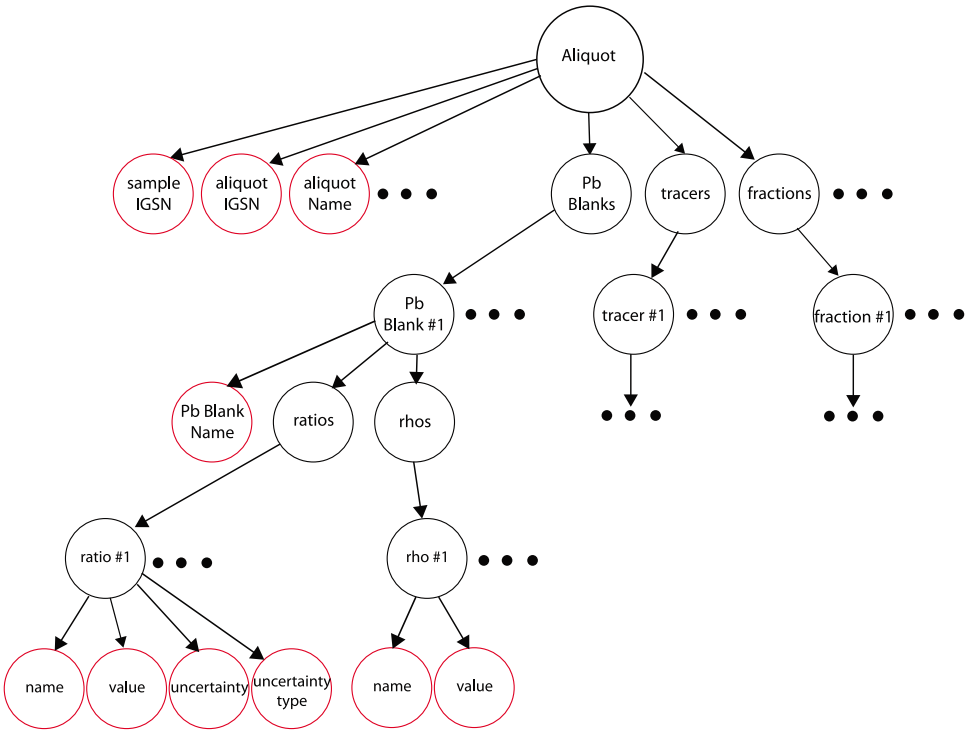

Figure 3. Partial view of an aliquot XML schema shown as a rooted directed tree. The tree is rooted at the aliquot, and the first set of children include both simple data elements (shown as red-colored leaves without children) and complex elements representing both sets of parameter models and the aliquot's fractions. The child node labeled "Pb Blanks" has several children itself, each a named defined data model. The child labeled "Pb Blank #1" follows the design description in section 3.3.1 and has children for its name and for the set of ratios and correlation coefficients (rhos). The ellipsis after a node represents elided nodes.

LDMs, (2) fraction LDMs, 3) mineral standard LDMs, and 4) date interpretation LDMs. Each LDM is defined with an XML schema at http://www.earth-time.org/projects/upb/public_data/XSD.

3.3.1. Parameter LDMs

[42] Currently, five types of parameter LDMs for ID-TIMS data are included in an aliquot LDM, but our architecture allows for additional LDMs as needed; each is summarized as follows: (1) Pb blank pertains to isotopic ratios with uncertainties and uncertainty correlations of the laboratory Pb blank, (2) tracer pertains to isotopic ratios with uncertainties, uncertainty correlations, and concentrations with uncertainties of Pb and U, (3) alpha Pb pertains to magnitude and uncertainty of the Pb fractionation for a monoisotopic Pb tracer, (4) alpha U pertains to the magnitude and uncertainty for U fractionation correction for a monoisotopic U tracer, and (5) physical constants pertains to decay constants with uncertainties and atomic molar masses.

[43] As an example, the LDM for a Pb blank has an XML schema at http://earth-time.org/projects/upb/public_data/XSD/PbBlankXMLSchema.xsd, shown in Figure 4. The actual data for a specific Pb blank is a defined data model encoded as an XML document arranged according to this schema. Note that Figure 3 includes the rooted directed tree of the Pb blank schema whose data elements are a set of ratios and a set of uncertainty correlations.

[44] As another example, EARTHTIME provides and certifies several defined data models for tracers at http://earth-time.org/projects/upb/public_data/EARTHTIME_tracers/XML. Each of these defined data models has a unique name and version number combination for identification, thus supporting later recalibration of the tracer. Sets of parameters are included at the aliquot level and each included fraction refers to the appropriate parameter model by name and version.

3.3.2. Fraction LDMs

[45] A fraction LDM encapsulates the data from the measured ratios and the selected parameter models. The complex data elements in a fraction LDM fall into six categories named to represent their data

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
        targetNamespace="http://www.earth-time.org"
        xmlns="http://www.earth-time.org"
        elementFormDefault="qualified">

 <xs:include schemaLocation="http://earth-time.org/projects/upb/public_data/XSD/ValueModelXMLSchema.xsd"/>
 <xs:element name="PbBlank">
  <xs:complexType>
   <xs:sequence>
    <xs:element name="name" type="xs:string"/>
    <xs:element name="ratios">
     <xs:complexType>
      <xs:sequence minOccurs="3" maxOccurs="4">
       <xs:element name="ValueModel" type="PbBlankRatio">
       </xs:element>
      </xs:sequence>
     </xs:complexType>
    </xs:element>
    <xs:element name="rhoCorrelations">
     <xs:complexType>
      <xs:sequence minOccurs="3" maxOccurs="3">
       <xs:element name="ValueModel" type="PbBlankRhoCorrelation">
       </xs:element>
      </xs:sequence>
     </xs:complexType>
    </xs:element>
   </xs:sequence>
  </xs:complexType>
 </xs:element>
</xs:schema>
```

Figure 4. Pb blank schema shown in XML schema language. The formatting of this schema is purely for readability and has no impact on its meaning. The schema is written in XML and consists of a series of nested and balanced named tags such as the tag "element" near the top written: ⟨*xs:element name* = "*PbBlank*"⟩. This tag is closed near the end of the schema with ⟨*/xs:element*⟩. The "element" tag is defined in the specifications for XML schema at http://www.w3.org/XML/Schema.

reduction role and summarized using examples of relevant data dictionary elements: (1) analysis measures (tracer mass, fraction mass, U blank mass, Pb blank mass, etc.), (2) measured ratios (r207_205m, r208_205m, r202_205m, r238_235m, r233_235m, etc.), (3) radiogenic isotopes ratios (r206_204r, r208_206r, r206_238r, r207_235r, etc.), (4) radiogenic isotope dates (date206_238r, date207_235r, date207_206r, etc.), (5) composition measures (U, Pb, Th concentrations; radiogenic and common Pb masses), and (6) sample isochron ratios (r204_206s, r238_204s, r235_204s, r204_207s, etc., where the suffix "s" refers to the sample corrected for tracer and fractionation only).

[46] Each data dictionary element is implemented with our lowest-level LDM: a ValueModel, containing four data values: a data element name, a value of specified or arbitrary precision, an uncertainty, and uncertainty type, e.g., absolute or percent.

3.3.3. Mineral Standard LDMs

[47] Repeated analysis of a mineral standard of known age is useful for determining the overall performance of a U-Pb ID-TIMS laboratory. The accuracy of an average measured age can be used to determine the quality of unknown analyses and the internal reproducibility of the standard gauges the laboratory's long-term performance.

[48] We implemented a LDM for mineral standards that provides for a customized model name and a set of data elements. These elements include the mineral standard, such as "Temora", a standard mineral name such as "zircon" or "apatite", a true age, and a measured age. Each age is modeled as a ValueModel plus a literature reference. The analyst assigns one or more of these mineral standards to each aliquot analysis as supporting documentation for their interpretations.

3.3.4. Date Interpretation LDMs

[49] Finally, we implemented a LDM to represent interpreted dates. These are either single grain dates or weighted mean dates calculated from some subset of the fractions in the aliquot. The analyst can include any number of date interpretations with the aliquot data but must select a "preferred" date interpretation. Our architecture currently supports models for the following dates: $^{206}Pb/^{238}U$, $^{207}Pb/^{235}U$, $^{207}Pb/^{206}Pb$,

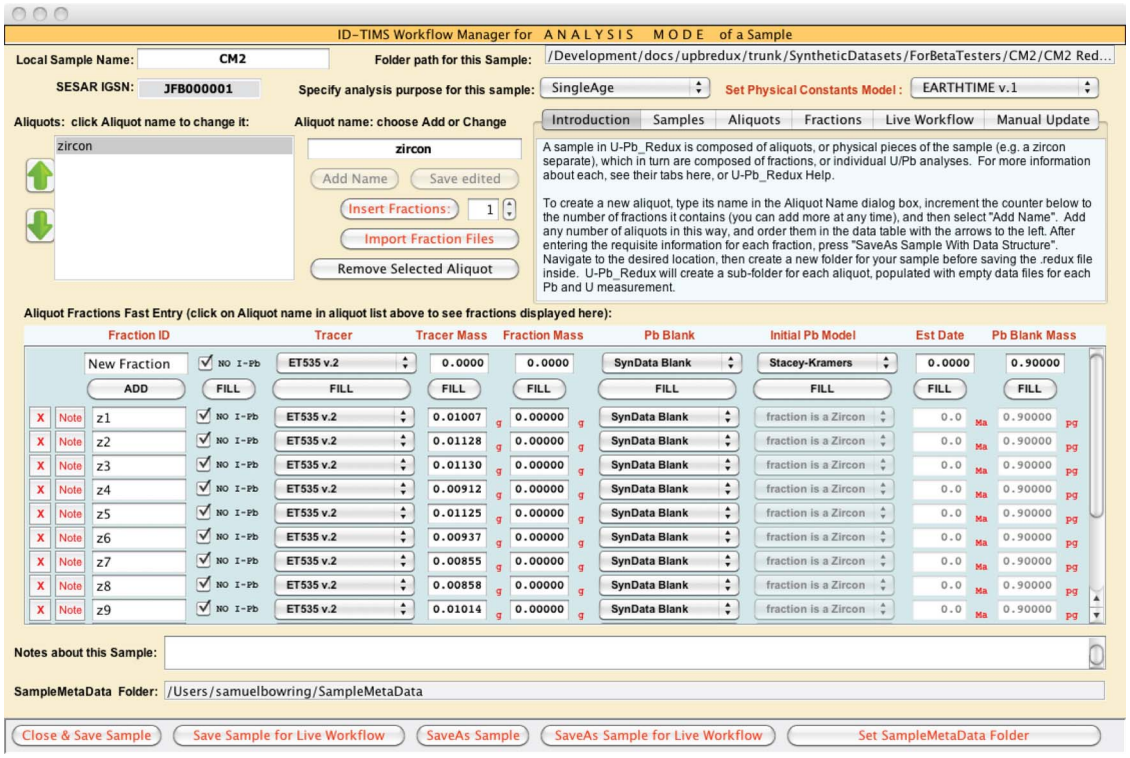

Figure 5. The U-Pb_Redux ID-TIMS sample manager provides the user with an interactive form in which the sample, aliquots, and fractions can be named and the reduction parameters specified. At the top left are text boxes for the lab's local sample name, the IGSN from SESAR (see section 3.1.2), and the local path for the sample's U-Pb_Redux file. The manager provides for adding and naming aliquots and for populating aliquots with placeholder fractions or with Tripoli-exported fraction files. The fractions panel across the bottom of the manager is an abbreviated instance of the aliquot manager's Fraction Fast Details tab (see section 4.2.2).

$^{206}Pb/^{238}U$ (Th corrected), $^{207}Pb/^{235}U$ (Pa corrected), $^{207}Pb/^{206}Pb$ (Th corrected), $^{207}Pb/^{206}Pb$ (Pa corrected), and $^{207}Pb/^{206}Pb$ (Th corrected and Pa corrected).

4. U-Pb_Redux

[50] U-Pb_Redux provides workflow management, data analysis, and analysis archiving and compilation.

4.1. Workflow Management

[51] High-precision geochronology depends on a well-defined workflow that facilitates rigorous and reproducible data handling and reduction by organizing the processing of each input data element and its transformation into the output geologic interpretations.

[52] Initially, the workflow manager in U-Pb_Redux supports the organization of input parameters before measurement. The analyst opens the sample manager (Figure 5) and specifies the sample name and IGSN (see section 3.1.2) plus the name of each aliquot. Before populating each aliquot with fractions, the analyst decides whether to use Live Workflow (section 4.1.1) or the default Manual Workflow (section 4.1.2). Then the analyst can set fraction parameters (see section 3.3.1) such as the tracer's defined data model, tracer mass, and fraction mass.

4.1.1. Live Workflow Fractions

[53] Live Workflow mode in U-Pb_Redux creates a shared workspace as a file structure on disk that models the hierarchy of physical sample, aliquots and fractions. U-Pb_Redux shares this workspace with Tripoli so that the two programs can interact in real time. When Tripoli is also in Live Workflow mode, it updates files in the shared workspace as data is collected from the mass spectrometer and processed by the analyst.

4.1.2. Manual Workflow Fractions

[54] In Manual Workflow mode the analyst imports measured data from fraction files (saved as XML documents), such as those produced by Tripoli

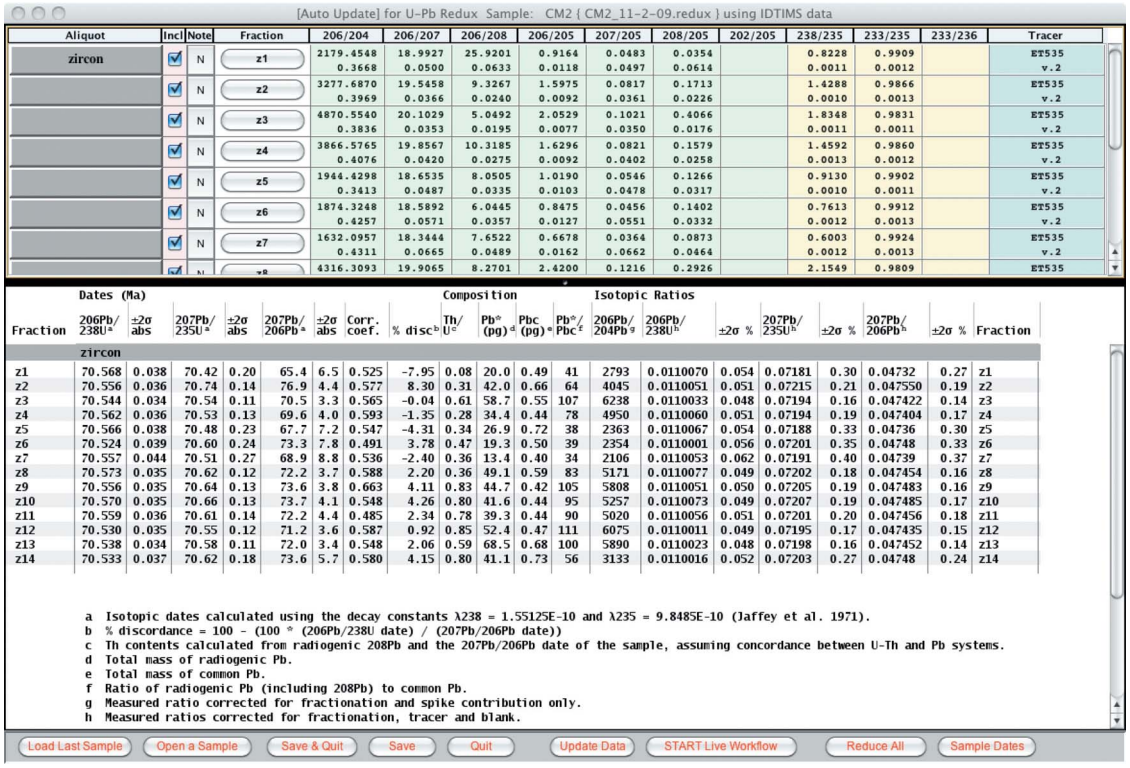

Figure 6. The main U-Pb_Redux window features two horizontal panels. The top panel displays aliquots, fractions, and their measured isotope ratios. The aliquot and fraction buttons open their respective managers. The bottom panel displays the data table for the reduced data (see section 4.4).

(see section 5). U-Pb_Redux also automatically retrieves metadata such as the sample name from these files.

4.2. Analysis Mode

[55] In analysis mode, U-Pb_Redux supports data reduction and interrogation of new U-Pb analyses and provides tools for the interpretation of geochronological data. After the analyst has loaded data for a sample, the main U-Pb_Redux window shows the measured data and a data table, as illustrated in Figure 6. The analyst organizes the data reduction using specialized managers for lab data, aliquots, fractions and sample dates.

4.2.1. Lab Data Manager

[56] Using the Lab Data Manager, the analyst can import, create, and manage defined data models of data reduction parameters (see section 3.3.1), such as for tracers, Pb and U fractionation, laboratory Pb blank and initial common Pb isotopic compositions, physical constants, and mineral standards. Because U-Pb_Redux saves the selected models with each sample file, they are available to anyone accessing the file with U-Pb_Redux.

4.2.2. Aliquot Manager

[57] The Aliquot Manager organizes the input parameters for all fractions in an aliquot, collates notes and metadata about the aliquot, and provides an interface to archive reduced data to a database. The Fraction Fast Details tab in the Aliquot Manager contains a row for each fraction and a column for each possible input reduction parameter, organized into four groups: Laboratory Data, Common Pb, Uranium, and Initial Isotopic Disequilibrium, as detailed below.

[58] 1. The Laboratory Data group contains the following: (1) *Fraction Name* that is unique within the sample, (2) *No Initial Pb* that is false when common Pb is apportioned between laboratory blank and initial Pbc and is true when common Pb is assumed to be laboratory blank only, (3) *Tracer model* by name, (4) *Tracer mass, uncertainty* values, (5) *Fraction Mass* used to calculate the concentration of Pb, U, or Th or set to zero to suppress this calculation, (6) *Pb Fractionation model* used only

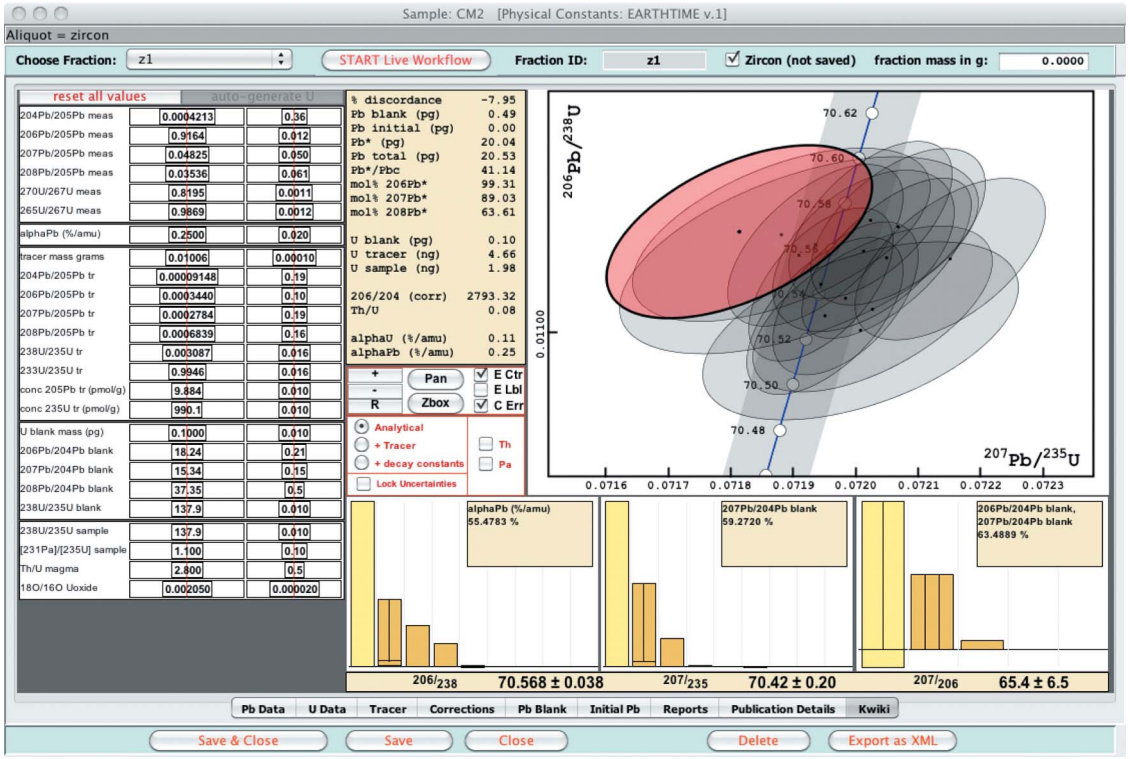

Figure 7. The Kwiki tab provides a number of interactive visualizations, including an interface to explore the sensitivity of each date and its uncertainty to each of its input parameters and their uncertainties. Section 4.2.3 explains in detail each of the five main panels.

when external fractionation correction is absent; and (7) *U Fractionation model* used only when external fractionation correction is absent.

[59] 2. The Common Pb group contains the following: (1) *Pb Blank IC model* that contains the laboratory Pb blank isotopic composition subtracted from all analyses, (2) *Pb Blank mass, uncertainty* subtracted from each analysis if *No Initial Pb* is true (the rest of the common Pb is assumed to have the initial Pbc isotopic composition), (3) *Initial Pbc IC model*, which refers to either a published model [e.g., *Stacey and Kramers*, 1975] or a custom model determined from a low mu phase, (4) *Estimated date* used as input to terrestrial Pb ore models that calculate age-dependent compositions, and (5) *Pbc model uncertainty*, which is the uncertainty in the terrestrial Pb ore model Pb IC.

[60] 3. The Uranium group contains the following: (1) Sample $^{238}U/^{235}U$, uncertainty, (2) *Blank* $^{238}U/^{235}U$, uncertainty, (3) *U Blank* mass, uncertainty, and (4) *Uoxide* $^{18}O/^{16}O$, uncertainty. If the U measurement is made on the UO_2^+ species with a mixed ^{233}U-^{235}U or ^{233}U-^{236}U tracer, an oxide correction must be made for an isobaric interference. The magnitude of this correction depends on the (often depleted) $^{18}O/^{16}O$ ratio in the UO_2^+ ion.

[61] 4. The Initial Isotopic Disequilibrium group contains the following: (1) *Th/U of magma*, uncertainty, required to correct for initial ^{230}Th disequilibrium in the ^{238}U decay scheme and (2) $^{231}Pa/^{235}U$ *activity ratio*, uncertainty, required to correct for initial ^{231}Pa disequilibrium.

4.2.3. Fraction Manager's Kwiki Sensitivity Testing

[62] The Fraction Manager provides detailed information about each paired U-Pb analysis organized into tabs such as the measured Pb data tab and the corrections tab (see section 4.2.5). The most informative tab, called Kwiki, provides several visualizations, described below and illustrated in Figure 7.

[63] Panel 1, the left-hand panel of the Kwiki tab, is a table with a row for each input parameter. The columns display the name of the input reduction parameter, its value and its uncertainty. Both the values and their uncertainties are displayed using miniature sliders. As each value and/or uncertainty slider is adjusted by mouse gesture, the data is

rereduced on-the-fly and the resulting changes appear in the three other data panels. Any number of sliders can be moved, but their positions are temporary, intended for hypothesis testing and data exploration only: their values are reset either by the user or upon exiting the Kwiki tab, with no way to effect a permanent change in the data.

[64] Panel 2, the top center panel, presents a synopsis of important data reduction parameters, including, for example, how the analysis is apportioned between laboratory blank, initial common Pb, and parent or daughter Pb or U. These values change in response to the value sliders.

[65] Panel 3, the top right panel, is a U-Pb concordia plot that displays uncertainty ellipses for all fractions in the aliquot, with the current fraction's ellipse highlighted. This ellipse responds in real time to movement of both value and uncertainty sliders. The concordia plot is navigable by panning and zooming with the mouse or using conventional navigation tools in a control panel (#5) to the left.

[66] Panel 4, the bottom panel shows the three calculated dates ($^{206}Pb/^{238}U$, $^{207}Pb/^{235}U$, and $^{207}Pb/^{206}Pb$) for the fraction, their uncertainties, and a graphical breakdown of each date's variance by the contribution from each input parameter. These graphs respond to both value sliders and uncertainty sliders.

[67] This novel uncertainty breakdown visualization, called sails and centerboard, leverages the fact that variance terms (1σ uncertainties, squared) are additive. The variance of an output parameter is the sum of the variances of its inputs, weighted by the squared derivatives of the outputs with respect to the inputs [e.g., *McLean et al.*, 2011]. Within each of the three graphs, vertical bars (sails) show the relative variance contribution of the input parameters as a vertical height above a horizontal centerline. Parameters with correlated uncertainties, such as the measured isotope ratios, each make a positive contribution to the variance, and their sum is displayed as a single segment divided into two or more evenly spaced zones. The variance contribution from the correlation between these variables can be positive or negative, and is displayed as a single segment (centerboard) above the horizontal line for a positive contribution and below for a negative contribution, illustrated in Figure 7.

[68] The total contribution for a set of correlated parameters is mathematically required to be greater than zero, so a negative contribution from the correlation is always smaller than the positive contribution from the variances of the individual parameters. Selecting a contribution bar highlights it in yellow and displays the corresponding variable name, or names in the case of covariance, and the percent contribution to the total variance. For example, in Figure 7, alphaPb (the fractionation correction coefficient for Pb), contributes 55% of the variance in the $^{206}Pb/^{238}U$ date. In Live Workflow, this tool is useful in determining whether to stop data acquisition.

[69] In panel 5, centered in the tab are two toolboxes: one for manipulating the concordia plot and the other for manipulating the uncertainty contributions and viewed corrections.

[70] The Kwiki tab also provides a live view of the data reduction progress in Live Workflow mode while Tripoli is updating data for U-Pb_Redux in real time, as explained in section 4.1. In particular, when only the Pb portion of a fraction is present, the user can use independent constraints on the $^{206}Pb/^{238}U$ date to assess discordance and uncertainty budgets with the "autogenerate uranium" button at the top left of the tab. The uranium ratios and uncertainties are populated from estimates of the date and alpha U entered by the analyst on the "U Data" tab of the fraction manager, as described in section 4.2.4.

4.2.4. Fraction Manager's Other Tabs

[71] In addition to the Kwiki tab, the fraction manager has tabs for "Pb Data," "U Data," "Tracer," "Corrections," "Pb Blank," "Initial Pb," "Reports" and "Archiving Details." The "Pb Data" and "U Data" tabs provide a view of the input ratios and their uncertainties. The user can compose fraction-specific notes and see the file path of the imported data. On the "U data" tab, the user can also enter an estimated date, uncertainty, and U fractionation magnitude to estimate U isotope ratios for visualization in the Kwiki tab before the U measurement is made.

[72] The "Tracer" tab displays the isotopic composition and U/Pb ratio of the selected tracer defined data model, as well as the mass of tracer added to the selected fraction and its uncertainty. This tracer may be imported from Tripoli (see section 5), or chosen from a list at the top of the tab.

[73] The "Corrections" tab displays information about Pb and U fractionation corrections, oxide correction, and U sample components. The magnitude of the fractionation correction applied to the Pb and U measurements are shown, and one of four

possible techniques used to determine them is highlighted with a red box: (1) Tripoli performed fractionation correction point by point using the measured ratios from a double spike, (2) an average fractionation value was calculated from the mean double-spike ratio, (3) for a monoisotopic tracer, a fractionation model was applied, or (4) the fraction is not corrected for fractionation.

[74] Oxide correction parameters for uranium oxide analyses are displayed in the middle of the Corrections tab. Tripoli performs oxide corrections automatically, but U-Pb_Redux allows the user to recorrect the data by supplying a different $^{18}O/^{16}O$ from the value used by Tripoli and to specify the uncertainty.

[75] The "Pb Blank" tab shows the details of the selected laboratory Pb blank defined data model.

[76] When the fraction is identified as having no initial Pb, i.e., the common Pb is considered to be a mixture of laboratory blank and initial common Pb, the "Initial Pb" tab displays details of the defined data model chosen for the initial common Pb IC. There are two ways to specify the initial common Pb IC: (1) use a common Pb ore evolution model, or (2) select a custom defined data model. For the first option, a work area appears in which the user can explore the effect of the estimated date on the calculated Pb IC.

[77] The "Reports" tab currently provides for writing two files: (1) a file containing all the inputs, intermediate variables, outputs, and their uncertainties per the data dictionary in Appendix A and (2) a file containing all of the covariance and Jacobian matrices used in the uncertainty propagation algorithms (see section 4.2.5). This tab will eventually host interactive views of the data reduction and uncertainty propagation equations to aid the analyst and student alike.

[78] Finally, the "Archiving Details" tab serves as the preparation portal for each fraction to be archived as part of the aliquot, described in section 4.3. Annotations such as the standard mineral, the setting type, whether it was physically abraded or chemically purified, for example, are made here. In addition, an image of the fraction can be chosen for upload and the analyst's comments recorded.

4.2.5. Data Reduction and Uncertainty Propagation

[79] Data reduction and uncertainty propagation in U-Pb_Redux are performed with the algorithms of *McLean et al.* [2011]. Building on the contributions of *Ludwig* [1980], *Roddick* [1987], and *Schmitz and Schoene* [2007], McLean et al. provide data reduction equations for several commonly used mixed U-Pb tracers.

[80] In U-Pb_Redux, the subset of data reduction equations from *McLean et al.* [2011] that is applied to each fraction is determined by the user's choice of tracer and initial common Pb case. Because Tripoli exports the average magnitude of each correction it has applied, such as mass fractionation or isobaric interferences, U-Pb_Redux can safeguard against erroneously making the correction again.

[81] The uncertainty propagation algorithm developed by *McLean et al.* [2011] linearizes the functions of the output variables in the vicinity of the measured input parameters. The output variables are expressed as a series of intermediate calculations whose partial derivatives are hard coded into U-Pb_Redux. This linear algebraic formulation offers four important advantages over previously published approaches. First, the algorithms efficiently calculate many outputs simultaneously, such as U/Pb and Pb/Pb ratios and dates and their uncertainties along with correlations used for plotting. Second, the derivatives can be systematically organized into matrices by U-Pb_Redux, eliminating the requirement that long uncertainty propagation equations be programmed manually and thus removing the consequent possibility of typographic errors. The product of these matrices is the total derivative required for linear uncertainty propagation, and is efficiently calculated using the JAMA matrix math package from http://math.nist.gov/javanumerics/jama/. Third, the calculation results, which are organized matrices of uncertainties and derivatives, are in a convenient format for potential use in plotting and in weighted mean calculations. Finally, the efficiency of this approach provides for the rapid calculations required to drive the interactive visualizations provided on the "Kwiki" tab and the Date Interpretations manager (see section 4.2.6).

4.2.6. Date Interpretations Window

[82] The Date Interpretations Manager in U-Pb_Redux, depicted in Figure 8, is streamlined and interactive. The left-hand panel initially presents a tree-based hierarchy of the sample and its aliquots. The right-hand panel has a number of vectorized, full-color visualizations, including concordia and weighted mean plots. The additional tabs shown are in development and will feature other interactive visualizations. Upon selection of an aliquot in the left

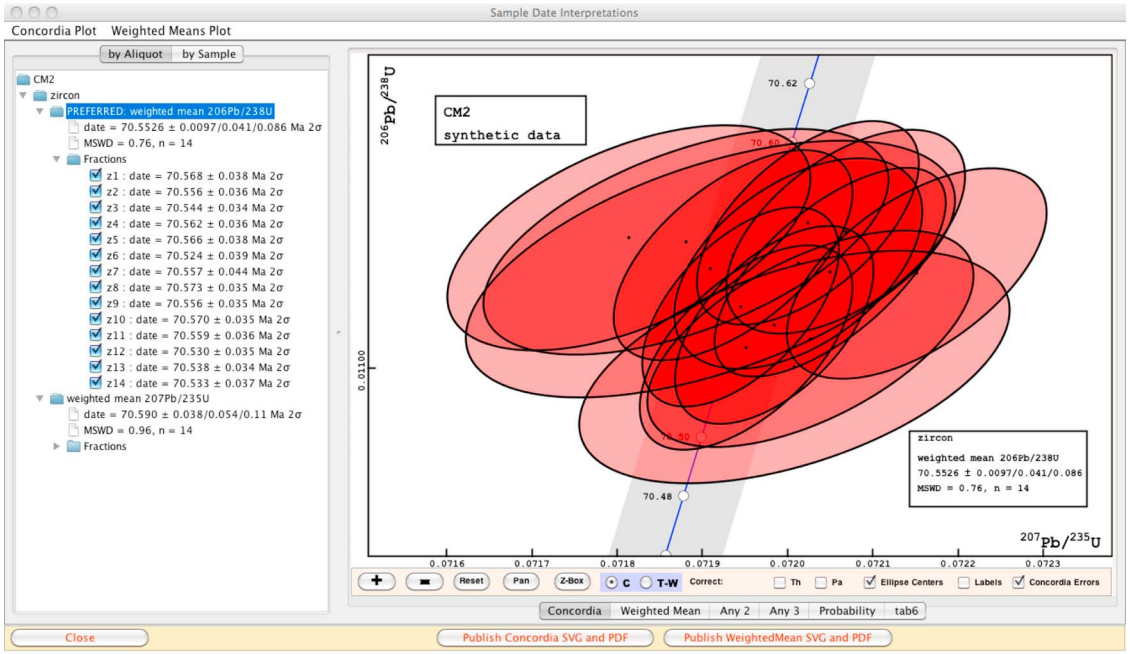

Figure 8. Sample date manager showing the concordia panel. The left-hand panel presents a tree-based hierarchy of the sample and its aliquots. The right-hand panel has a number of tabbed full-color views, including concordia and weighted mean plots. Upon selection of an aliquot in the left panel, the analyst can choose one or more defined date interpretations, including weighted mean and upper/lower intercept dates. For each selected date interpretation, the analyst can choose the subset of fractions for inclusion in the calculation, which is performed and displayed on-the-fly in both panels (section 4.2.6).

panel with a right-click of the mouse, the analyst can choose one or more defined date interpretations. These are enumerated in section 3.3.4 and include single grain, weighted mean, and upper/lower intercept dates. The user selects one of the interpretations as "preferred" and all chosen interpretations become part of the aliquot analysis. For each selected date interpretation, as the analyst chooses the subset of fractions to include, the requisite calculations are performed and the results displayed on-the-fly in both panels. The results include the calculated date with the uncertainty shown in the form ±X/Y/Z [e.g., *Davydov et al.*, 2010]. Here, X refers to the analytical uncertainty alone, Y to the analytical and tracer uncertainty combined, and Z to the analytical, tracer, and decay constant uncertainties. Details of how these are calculated appear in the study by *McLean et al.* [2011]. Also shown are the MSWD and the count of fractions used to calculate it.

4.2.6.1. Concordia Tab

[83] The concordia tab shows an interactive concordia as either a conventional or Terra-Wasserberg plot. The appearance and scaling of the ellipses and axes can be fully customized. The user can elect to display ellipse centers, labels, or the concordia error envelope, and can plot the result of thorium and protactinium corrections.

4.2.6.2. Weighted Means Tab

[84] The user can select which weighted means to display side-by-side as, for example, the two weighted means shown in Figure 9. Each vertical rectangle at the top represents a fraction, centered at the mean date of the fraction and extending 2σ above and below the mean. The vertical scale on the left is in millions of years. The horizontal line through the rectangles is the weighted mean, and is shaded at both 1σ and 2σ intervals for reference.

[85] Below the plotted fractions is a box showing the statistics for the weighted mean with the uncertainty expressed as the triple ±X/Y/Z described above. Below this is a graphical depiction of the weight assigned to each fraction using a filled square with side length proportional to $1/\sigma$. At the bottom of the window is a plot of the probability distribution function of the MSWD, which depends on the number of fractions included. The mean of the curve is always 1, and the peak width decreases as n increases; the vertical black line represents the

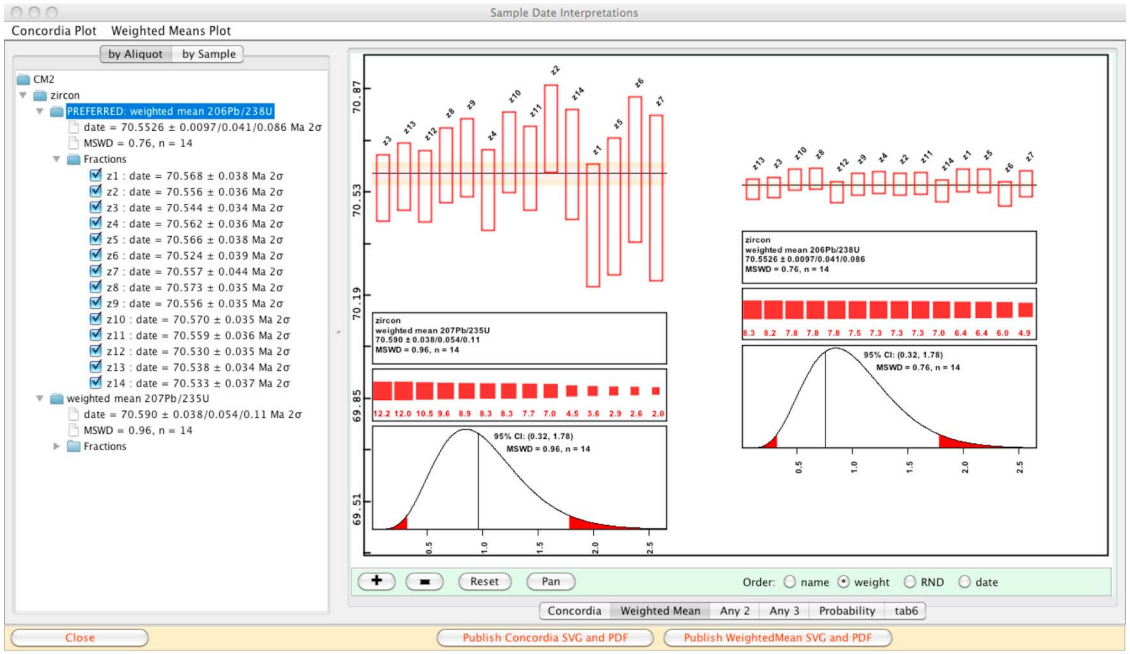

Figure 9. Sample date manager, showing the weighted means panel. Each weighted mean is represented by four vertically stacked panels. From top to bottom they are the fractions, plotted as $\pm 2\sigma$ rectangles with increasing date on the vertical axis, the weighted mean and its three levels of uncertainty, a visualization of the weight assigned to each fraction, and the probability distribution function of the MSWD statistic. These are explained in detail in section 4.2.6.2.

current MSWD. The user can elect to order the weighted means by name, weight, date or randomly.

4.3. Archiving and Compilation Mode

[86] U-Pb_Redux supports the complete archiving and retrieval, including compilation, of aliquot analysis details and visualizations. The archived data are stored in an XML document (explained in section 3.3) that can be saved anywhere, including the Geochron database at http://geochronportal.org. The compilation of several archived aliquots from different labs into a "super sample" empowers the analyst to interpret large data sets by viewing multiple samples side-by-side.

[87] The archiving functionality is accessed through the tabs of the aliquot manager described in section 4.2.2. The "Measured Ratio Check and Metadata" tab checks the consistency of the measured data by testing whether the correlation coefficients among inputs are within [−1.0,..,1.0]. If not, U-Pb_Redux details the violations and recommends against archiving. In addition, the analyst can enter references and comments as well as choose one or more Mineral Standards models (see section 3.3.3). The analyst uses the "Archive" tab to write an XML document of the current analysis to a disk drive or to Geochron for later access and possible compilation with other archived analyses. U-Pb_Redux also supports the archiving and compilation of legacy data.

4.4. Publication-Quality Documents

[88] U-Pb_Redux produces publication-quality documents including data tables, concordia and weighted mean plots. The user can fully customize each document on-the-fly including graphical elements and the display of numbers. Outputs are available as scalable vector graphics (SVG), portable document format (PDF), or text (TXT) files, or Microsoft Excel® spreadsheets (XLS), depending on context.

5. Tripoli

[89] As described in section 3, Tripoli is a program that imports mass spectrometer data files and supports interactive review and preparation of measured isotopic ratio data.

5.1. Importing and Reviewing Raw Data

[90] Tripoli currently recognizes and imports data files generated by Micromass, Fisons, VG/GV,

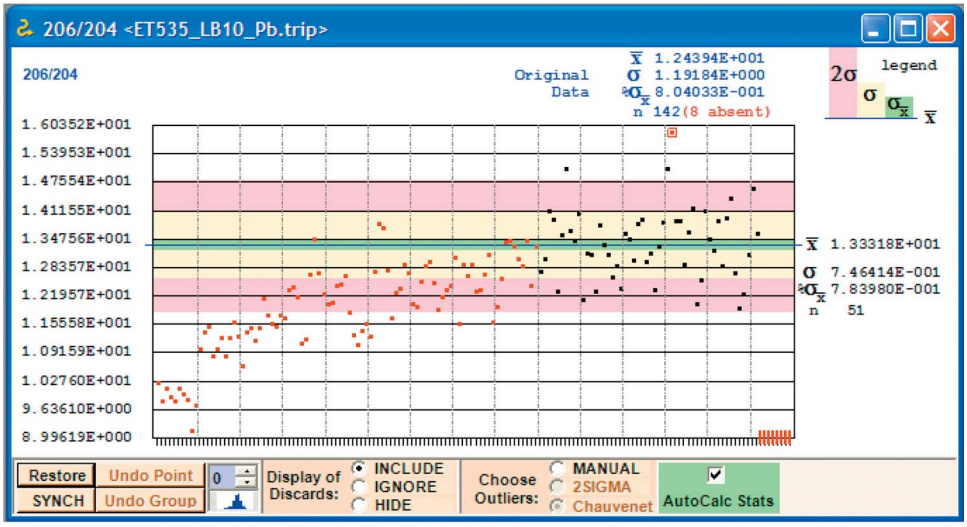

Figure 10. Tripoli time series graphs of measured ratio data are interactive, allowing the user to exclude data points and see the recalculated results on-the-fly. The red data points have been excluded from this analysis due to the presence of an isobaric interference, and the mean statistics of the black data points are presented on the right. The filtered data is then saved for export to U-Pb_Redux.

IsotopX, Finnigan and Thermo-Finnigan mass spectrometer software. Imported ratios for either Pb or U are listed in Tripoli's main window, with a button to the right that graphs each ratio measurement over time.

[91] Time series graphs for the measured ratios reveal temporal trends in the mean and variability of the data, as shown in Figure 10. Each data point and block of data can be manually toggled for inclusion, but Tripoli supports the more rigorous outlier filter, Chauvenets Criterion [*Chauvenet*, 1863], recommended when there is reason to believe the data are normally distributed. Parametric statistics are refreshed on-the-fly and shown on the right of each time series, with the original statistics at the top. The current mean, 1σ standard error, and 1σ and 2σ standard deviations about the mean are illustrated with a horizontal black line and green, yellow, and red bands, respectively. A histogram function overlays the graphed data with a user-selected bin count to provide a visual impression of the data distribution.

[92] Tripoli saves the state of all measured data and rejection choices as a work file with the file extension "[filename].trip". Thus, users can return to previous work or share raw data and interpretations with others. Before exporting to U-Pb_Redux, Tripoli can implement a number of point-by-point corrections to the measured data.

5.2. Corrections to Data

[93] Tripoli currently performs several ratio-by-ratio corrections for parameters that change during the course of analysis: (1) uranium oxide, (2) barium phosphate, (3) thallium isobaric interferences, and (4) isotopic fractionation (available for double-spiked U and Pb analyses).

5.2.1. Uranium Oxide Corrections

[94] An isobaric interference correction is required for the presence of $^{233}U^{18}O^{16}O$ under $^{235}U^{16}O^{16}O$, or $^{236}U^{18}O^{16}O$ under $^{238}U^{16}O^{16}O$ that utilize a mixed ^{233}U-^{235}U or ^{233}U-^{236}U tracer, respectively. Tripoli automatically performs this correction when it detects the presence of $^{265}(UO_2)/^{267}(UO_2)$ or $^{266}(UO_2)/^{268}(UO_2)$ ratios. In the "Corrections" menu, the user can specify the value of $^{18}O/^{16}O$ used, and this value may be later changed in U-Pb_Redux. Oxide-corrected uranium isotope ratios are flagged with a bluish tint on the front screen of Tripoli, and the $^{18}O/^{16}O$ used is displayed at the top of the window.

5.2.2. Fractionation Correction

[95] For Pb and U analyses performed with a double spike (e.g., a tracer enriched in ^{202}Pb and ^{205}Pb or ^{233}U and ^{235}U or ^{236}U), the magnitude of the measured isotopic fractionation can be deter-

mined as a function of the ratio of the two enriched isotopes. If this magnitude remains constant during the analysis, then the mean ratio of enriched isotopes may be used to calculate an average correction factor, α_U or α_{Pb}, applied to the measured isotope ratios.

[96] However, if this magnitude changes over time, it is preferable to correct each ratio, generating a time-resolved record of the magnitude of isotopic fractionation and a corrected data set with a meaningful mean and standard error. In Tripoli, ratio-by-ratio Pb fractionation requires specifying the tracer, and for U, the blank mass and U isotopic composition [see *McLean et al.* [2011]. Tripoli records the parameters used and the average magnitude of the correction, for export to U-Pb_Redux.

5.3. Data Export

[97] Tripoli exports data in either a Manual Workflow or a Live Workflow mode that interacts with U-Pb_Redux as detailed in section 4.1.1.

5.3.1. Manual Workflow

[98] From the Control Panel, the user can export to the system clipboard the means and standard errors of the ratios used by other data reduction spreadsheets as, such as "PbMacDat", or export the raw ratios and the current parametric statistics to a tab-delimited text file.

[99] The preferred mode of manual export is to write the data to an XML document for use by U-Pb_Redux. The XML schema is http://earth-time.org/projects/upb/public_data/XSD/UPb_ReduxInputXMLSchema.xsd. The exported files contain either the U data or the Pb data, the tracer's defined data model, if used, and data about any applied corrections.

5.3.2. Live Workflow

[100] We described Live Workflow from the perspective of U-Pb_Redux in section 4.1.1. In Tripoli the user manages Live Workflow through the Control Panel.

[101] To enable Live Workflow, the user specifies both the location of the "SampleMetaData" folder that was defined previously using U-Pb_Redux and the location of the folder used by the mass spectrometer for raw data files. Tripoli then automatically detects new or updated mass spectrometer data files, applies any active corrections and statistical filters to the new data, and refreshes its graphical displays. At any time during data acquisition and review, the analyst can transmit the current data to U-Pb_Redux, which rereduces the data and refreshes its displays in the Kwiki tab (section 4.2.3).

5.4. Tripoli History Files

[102] Tripoli can also archive and analyze the long-term reproducibility of isotopic standards and collector gains by reading folders of files and presenting them graphically, similar to other measured data. For collector gains, the folder must contain Excel files in a prescribed format. For standards, the folder must contain files produced using Tripoli that have the extension ".trip".

6. Software Quality and Evolution

[103] The quality of software is determined both objectively and subjectively. We strive for objective quality by adhering to the best practices of computer science and software engineering with regard to design, implementation, documentation, and testing. We provide U-Pb_Redux and Tripoli as open source software licensed under the Apache Software Foundation License, Version 2.0, found at http://apache.org/licenses/LICENSE-2.0.html. In this way, analysts and reviewers will have access to the full code base, its documentation, and testing mechanisms on request. Participants are free to alter and recompile the software for their own use. The authors control the official versions of the software and grant others permission to submit changes to us for review for possible inclusion.

[104] The subjective nature of software quality derives from how users view its reliability, robustness, security, and ability to accommodate change as the system evolves. In our case, there is an additional complication bearing on both development and the determination of quality attributes: the collaboration among scientists specializing in the two different knowledge domains of computer science and Earth science has exposed inherent difficulties in communicating between disciplines.

[105] The authors have found over the last 15 years of collaboration that as the software evolves to handle the current tasks of geochronology, new possibilities in refining our approach become evident. These possibilities are investigated and ultimately incorporated into the software to begin the cycle anew. The open source development model serves this collaborative evolution well because it provides transparency and, through con-

trolled repositories, preserves important aspects of project history.

[106] Until now, the primary development decisions have occurred among the authors and their close colleagues directly. Our goal is to expand the software development to include more members of both the geochemistry and software engineering communities. To that end, we maintain a community Web site at http://cirdles.org, which hosts a number of projects including these two software products, community news, discussion groups, and bug reporting. We have also hosted a number of small hands-on workshops for interested users to contribute to the development process.

[107] The overarching consideration here is that both Tripoli and U-Pb_Redux are perpetual works-in-progress as we refine the algorithms and the feature set in response to progress in geochronology and as additional communities are included.

[108] The continuous evolution of software also requires continuous management of change. The best management techniques include continuous refactoring of the software, of the tests, and of the documentation. Refactoring is a technique for continuous software improvement formalized by *Fowler* [1999] and subsequently adopted and refined by the software engineering community.

[109] The problem of how to continuously evolve and maintain software in an academic setting and supported by external funding is ripe for solution. One of our long-term goals is to understand and communicate a set of best practices as a contribution to that solution.

7. Conclusions and Future Work

[110] Our cyber infrastructure architecture uses open-source software and software engineering processes to support the ID-TIMS analytical community. This includes standardized XML schema that undergird its interoperability. The software program U-Pb_Redux manages the workflows supporting our architecture as well as reducing data, propagating uncertainties, and visualizing the results. The software program Tripoli allows the user to interactively review measured isotopic data before export to U-Pb_Redux. This work will continue to evolve as software engineering and geochronology advance. In the near term, we plan a number improvements, including the adaptation of our architecture and software to related efforts.

[111] This work represents as far as we know a first-of-its-kind collaboration among computer scientists and Earth scientists. We have begun additional collaborative efforts to extend this cyber infrastructure to support other communities such as the LA-ICP-MS and U-series geochronology communities [*Horstwood et al.*, 2010]. We also plan to introduce several more interactive visualizations. One visualization will plot any two or three inputs or calculated parameters and their associated uncertainty ellipses or ellipsoids in two- or three-dimensional space. This visualization will aid in the evaluation of trends and in performing linear regression analysis. Another visualization will depict the underlying data reduction and uncertainty propagation algorithms as mathematical expressions with measured data from the sample, allowing the user to see how the inputs become geochronological data.

[112] Finally, we intend to further publicize our online community based at http://cirdles.org and to encourage users of the software to participate. As part of our development-as-evolution strategy, we will make continuous evaluations and improvements to the software. The most recent versions of the software are available at http://cirdles.org. The source code is available by request from the authors.

Appendix A: Summarized Data Dictionary

[113] This summary of the data dictionary lists the principal data elements by name only. As explained in section 3.3, the naming format is consistent with rules for naming variables in computer programming languages and is human readable. We use these names internally throughout the cyber infrastructure to provide consistency and clarity: (1) Measured Parameters (r206_204m, r207_204m, r208_204m, r207_206m, r206_208m, r204_205m, r206_205m, r207_205m, r208_205m, r202_205m, r238_235m, r233_235m, r238_233m, r233_236m), (2) Tracer Solution (r206_204t, r207_204t, r208_204t, r207_206t, r206_208t, r204_205t, r206_205t, r207_205t, r208_205t, r202_205t, r238_235t, r233_235t, r238_233t, r233_236t, r235_205t, concPb205t, concU235t), (3) Physical Constants (lambda230, lambda231, lambda232, lambda235, lambda238, gmol204, gmol205, gmol206, gmol207, gmol208, gmol235, gmol238), (4) Lab Parameters (r206_204b, r207_204b, r208_204b, rho206_204b_207_204b, rho206_204b_208_204b, rho207_204b_208_204b, r238_235b,

alphaPb, alphaU, labPbBlankMass, labUBlankMass), (5) Other Fraction-Specific Inputs (r238_235s, tracerMass, zirconCase, r206_204c, r207_204c, r208_204c, r206_207c, rTh_Umagma, commonPbAge, ar231_235sample, (6) Pb Analysis (tracer: alphaPb, molPb205t; Pb blank: r204_205fc, molPb204b, molPb206b, molPb207b, molPb208b, molPb204c, molPb206c, molPb207c, molPb208c, blankPbGramsMol, blankPbMass; radiogenic Pb and sample Pb: molPb206r, molPb207r, molPb208r, molPb206s, molPb207s, molPb208s), (7) U Analysis (tracer: molU235b, molU238b, molU235t, molU238t, molU233t, alphaU; sample U: molU235s, molU238s), (8) Dates (radiogenic isotope ratios: r206_238r, r207_235r, r207_206r; radiogenic isotope dates: date206_238r, date207_235r, date207_206r; Th and Pa correction: molsU, date206_238_Th, molTh232s, r207_206r_Th, date207_206_Th, date207_235_Pa, molPa231s, delta207_Pa, r207_206r_ThPa, date207_206_ThPa), (9) Isochron ratios (r207_206s, r204_206s, r204_207s, r238_206s, r238_207s, r235_207s, r238_204s, r235_204s), and (10) Outputs (blankPbMass, InitCommonPbMass, totCommonPbMass, radToCommonTotal, radToCommon206, radToCommon207, radToCommon208, percentDiscordance, concU, concPb, r208_206r_Th, rTh_Usample).

Acknowledgments

[114] Funding for this project was provided through a subaward from NSF award 0522222 (EarthChem) to Doug Walker at the University of Kansas and grant EAR-0930166 from the National Science Foundation. The authors thank Chris Starr and Jerry Boetje for helpful comments and Dan Condon and Don Davis for their constructive reviews. We thank Doug Walker for his steadfast and enthusiastic support and Drew Coleman, Matt Rioux, Robert Buchwaldt, Doug Walker, and Terry Blackburn for the many constructive suggestions made during their extensive use of the software. Undergraduate research assistants Brittany Johnson, Danielle Brandon, Hunter Hegler, and James Goodrich contributed to the software development.

References

Chauvenet, W. (1863), *A Manual of Spherical and Practical Astronomy*, vol. II, Lippincott, Philadelphia, Pa.

Davis, D. (1982), Optimum linear regression and error estimation applied to U-Pb data, *Can. J. Earth Sci.*, *19*(11), 2141–2149.

Davydov, V. I., J. L. Crowley, M. D. Schmitz, and V. I. Poletaev (2010), High-precision U-Pb zircon age calibration of the global Carboniferous time scale and Milankovitch band cyclicity in the Donets basin, eastern Ukraine, *Geochem. Geophys. Geosyst.*, *11*, Q0AA04, doi:10.1029/2009GC002736.

Fowler, M. (1999), *Refactoring: Improving the Design of Existing Code*, Addison-Wesley, Boston.

Horstwood, M., G. Gehrels, and J. Bowring (2010), Improving consistency in laser ablation geochronology, *Eos Trans. AGU*, *91*(28), 247.

Ludwig, K. (1980), Calculation of uncertainties of U-Pb isotope data, *Earth Planet. Sci. Lett.*, *46*(2), 212–220, doi:10.1016/0012-821X(80)90007-2.

Ludwig, K. (1985), PBDAT200, a computer program for processing raw Pb-U-Th isotope data, *U.S. Geol. Surv. Open File Rep., 85-547*, 90 pp.

Ludwig, K. (1991), Isoplot for MS-DOS, a plotting and regression program for radiogenic-isotope data, *U.S. Geol. Surv. Open File Rep., 88-557*, 39 pp.

McLean, N. M., J. F. Bowring, and S. A. Bowring (2011), An algorithm for U-Pb isotope dilution data reduction and uncertainty propagation, *Geochem. Geophys. Geosyst.*, doi:10.1029/2010GC003478, in press.

National Science Foundation (NSF) (2007), Cyberinfrastructure vision for 21st century discovery, *NSF 07-28*, Arlington, Va.

Roddick, J. C. (1987), Generalized numerical error analysis with applications to geochronology and thermodynamics, *Geochim. Cosmochim. Acta*, *51*(8), 2129–2135.

Schmitz, M. D., and B. Schoene (2007), Derivation of isotope ratios, errors, and error correlations for U-Pb geochronology using ^{205}Pb-^{235}U-(^{233}U)-spiked isotope dilution thermal ionization mass spectrometric data, *Geochem. Geophys. Geosyst.*, *8*, Q08006, doi:10.1029/2006GC001492.

Simsion, G. (2007), *Data Modeling Theory and Practice*, Technics, Bradley Beach, N. J.

Stacey, J., and J. Kramers (1975), Approximation of terrestrial lead isotope evolution by a two-stage model, *Earth Planet. Sci. Lett.*, *26*(2), 207–221, doi:10.1016/0012-821X(75)90088-6.