



StraboSpot data system for structural geology

J. Douglas Walker¹, Basil Tikoff², Julie Newman³, Ryan Clark⁴, Jason Ash¹, Jessica Good¹, Emily G. Bunse¹, Andreas Möller¹, Maureen Kahn², Randolph T. Williams², Zachary Michels^{2,*}, Joseph E. Andrew¹, and Carson Ruffledt¹

¹Department of Geology, University of Kansas, Lawrence, Kansas 66045, USA

²Department of Geoscience, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA

³Department of Geology and Geophysics, Texas A&M University, College Station, Texas 77843, USA

⁴MapBox, Washington, D.C., USA

ABSTRACT

StraboSpot is a geologic data system that allows researchers to digitally collect, store, and share both field and laboratory data. StraboSpot is based on how geologists actually work to collect field data; although initially developed for the structural geology research community, the approach is easily extensible to other disciplines. The data system uses two main concepts to organize data: *spots* and *tags*. A spot is any observation that characterizes a specific area, a concept applicable at any spatial scale from regional to microscopic. Spots are related in a purely spatial manner, and consequently, one spot can enclose multiple other spots that themselves contain other spots. In contrast, tags provide conceptual grouping of spots, allowing linkages between spots that are independent of their spatial position.

The StraboSpot data system uses a graph database, rather than a relational database approach, to increase flexibility and to track geologically complex relationships. StraboSpot operates on two different platform types: (1) a field-based application that runs on iOS and Android mobile devices, which can function in either Internet-connected or disconnected environments; and (2) a web application that runs only in Internet-connected settings. We are presently engaged in incorporating microstructural data into StraboSpot, as well as expanding to include additional field-based (sedimentology, petrology) and lab-based (experimental rock deformation) data. The StraboSpot database will be linked to other existing and future databases in order to provide integration with other digital efforts in the geological sciences and allow researchers to do types of science that were not possible without easy access to digital data.

INTRODUCTION

Structural geology stands at a crossroads. For more than a century, practitioners in the field have collected data with pencil, paper, and analog tools. The discovery of original data was almost impossible, and without firsthand knowledge of the geologist who collected the data, it was difficult to divine the intent and competence of that person from published work. This approach

will not work in the future. Structural geology data must be collected in or converted to a digital format to become widely available and profitably used in the future. One approach is to simply render digitally our field notebooks and streamline our data collection, meeting data archiving requirements solely by posting spreadsheets to servers of uncertain lifetime. Instead, we have opted to use the critical analog-to-digital transition as an opportunity to reimagine how data collection and archiving could work with modern computational tools that have become available in the last few decades. We present here a new paradigm—StraboSpot—for field data collection that is designed for structural geologists but is easily extensible to other disciplines.

StraboSpot is an attempt to reconceptualize field data collection, allowing the structural geology community to digitally collect, store, and share both field and laboratory data (<https://strabospot.org>). The current work was motivated by the recognition that field scientists had not yet joined the EarthCube¹ (<https://earthcube.org>) effort to transform science through the development of infrastructure enabling sharing of data, because the field sciences lack community databases and have minimal reporting standards. This situation was confirmed by the U.S. structural geology and tectonics community (<http://earthcube.org/document/2012/structural-geology-tectonics-end-user-workshop-report>). The primary reason is the inherent nature of field data; they are heterogeneous, sparse, and—importantly—not instrumentally collected, making them notoriously difficult to digitize (e.g., Laxton and Becken, 1996; Walker et al., 1996).

The StraboSpot digital data system is an attempt to build a *geologic* data system, not a *geographic* information system (GIS), to address the difficulties of digitizing field-based data. This paradigm of a geologic data system is based on how geologists actually work, rather than trying to shoehorn their workflows into poorly fitting computational templates. As such, it requires the introduction of a few key concepts. The *spot* concept is foundational to the StraboSpot data system, as it captures the scale-dependent and hierarchical data collected by geologists. A spot is an observation with a location and area of significance. Spots are inherently spatial, so we group them into *nests* that accommodate the hierarchical nature of geologic observations while giving them real-world coordinates. Conceptually related spots may be linked through *tags*, a flexible and powerful way to apply geologic attributes to any observation (spot),

*Now at Department of Earth Sciences, University of Minnesota–Twin Cities, Minneapolis, Minnesota 55455, USA

¹EarthCube is a National Science Foundation program to engage computer and Earth scientists in building a cyberinfrastructure for the science.

consistent with how structural geologists group and organize data. While spots are inherently spatially referenced, tags allow conceptual labeling of data. *Relationships* between spots, tags, and/or measurements establish aspects of space-for-time substitution, such as cross-cutting relationships and superposition of fabrics. Finally, the *purpose* for collecting data must be specified to provide the context for observations and measurements.

In this contribution, we fully describe the StraboSpot data system. First, we begin with a discussion of two critical components of field data collection, workflow and scale, and describe their roles in the organization of the StraboSpot data system. The emphasis is field-based data collection, because this aspect of StraboSpot is best developed and builds on the decades of work for digital field data collection for mapping (e.g., Walker and Black, 2000; Pavlis et al., 2010; https://serc.carleton.edu/research_education/geopad). Second, we explain why we chose a graph database to capture geologic field data. A graph approach arranges data into what is termed a tree structure. This is much more flexible than a relational approach in that it easily accommodates modifications to terminology and has no fixed data structure other than that data are connected either hierarchically or logically. Graphs are described as NoSQL databases (Not Only SQL), meaning they employ other methods of querying than structured query language (SQL). The complex nature of geologic data limits its representation and linking using standard data methods (e.g., relational databases or simple flat files). A graph database approach provides the flexibility to accommodate new and complex relationships, such as space-for-time substitutions. We describe the overall framework of the data system, which incorporates both a front-end user interface to collect data digitally as well as a back-end database to facilitate sharing of data within the geologic community. Finally, we describe our future plans for the StraboSpot data system.

■ WORKFLOWS AND SCALE IN GEOLOGIC FIELD WORK

Geologists collect data in the field using a variety of different protocols or workflows (e.g., Shipley and Tikoff, 2018). Workflows during geologic field studies fall into four major categories or conceptual modes (below). These workflows are similar to the scopes or types of geologic maps as reviewed in Barnes and Lisle (2004, p. 21–22), although we add some texture to their description. We have designed StraboSpot so that users can transition seamlessly between workflows or modes of data collection.

Field Study Workflows

Reconnaissance Mode

This is the most generalized form of field study, and may not even require leaving the office. Reconnaissance is done to get a very general overview of a field area and is commonly done using remote-sensing imagery (satellite or

photos). Only the most general information is recorded. The purpose of this work can be poorly defined (are there any likely areas of outcrop?) to fairly specific (are there potential locations of fault scarps to examine in the field?). Reconnaissance mode is used also for basic logistics, such as accessibility of an area. Reconnaissance studies typically apply to a single physical scale of observation.

Mapping Mode

This is the workflow we consider typical for geological mapping and teaching students field geology, where geologists walk and collect data on rocks they encounter in outcrops. The features normally present on a geologic map, such as contacts and attitudes, are recorded by direct observation or inference. Location is important, but most observations are applicable at similar spatial scales of significance, that is, they represent the local attitude of rocks or representative rock types and textures. The data are typically represented on a map of some sort and commonly presented at scales of 1:1000–1:100,000. Sample collection and outcrop photos are commonly incorporated into this mode. Lastly, mapping is commonly done at scales where individual observations are locatable by GPS measurements, although with the advent of drone imagery and software such as structure from motion (e.g., Westoby et al., 2012), some mapping is done at scales much finer than even 1:100.

Multi-Scale Mode

This is perhaps the most common and complex workflow we encounter today in research-oriented studies for structural geology and field geology. This workflow spans many orders of magnitude in scale (thin section to mountainside) as well as potentially large gaps in periods of time. In this mode, a field geologist will use an existing map or reconnaissance and identify locales for more detailed study. The detailed study is commonly done in the field at spatial scales below the resolution of most mapping or GPS methods. However, the retention of the spatial relations of observations is critical to correctly represent, record, and interpret all measurements. Samples collected in the field must also retain these spatial relations, which must be carried through to further subdivisions of the sample (e.g., thin sections) for additional analyses. Samples collected therefore inherit the field context (location, orientation) and then acquire more detailed information upon further analyses that may span scales from the whole sample to the atomic level. At every scale, each location and relative position is important.

Multi-Measurement Mode

In this mode, the field scientist typically studies a specific feature or set of features where the actual positions of observations are not important, but

the (conceptual) relations between observations must be preserved. For example, when collecting orientation data around a fold, single measurements are critical information, while their specific locations within the outcrop may not be. In this example, multiple orientation measurements are attributable to a single outcrop, although the spatial relation of these orientation measurements to each other is not significant. In this mode, a rich data set consisting of numerous observations is collected that corresponds to a single, somewhat generalized location.

Substituting Space for Time

In addition to tracking observations across scales, geologists use spatial patterns and relations between features to make inferences about time and event order using cross-cutting or other relationships. This inferred sequential development of primary and deformation features is referred to as a space-for-time substitution or record (e.g., Jenny, 1941). Examples include documenting a dike that intrudes across deformation fabrics or the superposition of fabrics (S_1 , S_2 , etc.): there is a hierarchical and complex arrangement of fabric and geometric elements from the outcrop to the grain scale. Because of these relationships, individual observations collected in the modes described above may represent a complex set of relations and hierarchies.

Incorporating the Field Geologist's Workflow into Digital Data Collection

As a result of the inherent complexity and heterogeneity of data collection, the vast ranges in scales of observation, and the interpretation of relations observed in the field and in the laboratory, the field sciences have been slow to adopt digital databases. There have been attempts to describe this span of data in a single schema (the data backend; e.g., GeoSciML: <http://www.geosci.ml.org> and Sen and Duffy, 2005; and GeMS: USGS NCGMP, 2018) and in a GIS database (Tomlinson, 1974; Walker et al., 1996, 2002; ESRI Geology Data Model: ESRI, 2018), but a usable interface or front end for the field and laboratory geologist has proven elusive and a problem for geologists in moving forward in the digital era. The StraboSpot system is intended to work across the range of workflows described above using a single front end.

A GIS can have a very rich level of content and can enforce a host of spatial and logical connections between data (e.g., topology and spatial joins), and most systems can perform a variety of data analyses (area calculations) and have presentation and production capabilities. Many structural geologists have resisted the use of GIS platforms for a variety of reasons: their interfaces and tools are very extensive but mostly not suited to geologic applications, they are expensive to acquire, they are not easy to use when first starting (i.e., have a very steep learning curve), and most run only on Windows platforms. In particular, ArcGIS by ESRI (<https://www.esri.com/en-us/arcgis/about-arcgis/overview>)

is not palatable to many geoscientists because the software is proprietary and for profit. Open-source software such as QGIS (<https://qgis.org>) is available but has a much smaller group of structural geologists using it. Again, both ArcGIS and QGIS are complicated to learn and use. There are some available mobile applications, such as ArcPad (<https://www.esri.com/en-us/arcgis/products/arcpad/overview>), Collector for ArcGIS (<https://www.esri.com/en-us/arcgis/products/collector-for-arcgis/overview>), Mappt (<http://www.mappt.com.au>), and Fieldmove Clino (<https://www.mve.com/digital-mapping>), but these packages are very limited in capabilities and, in part, output data in proprietary formats unreadable in other packages. Finally, none of these software packages archives data into a shared database available to the geologic community.

A GIS and relational database function very well for documenting field data on maps in reconnaissance and mapping modes of work, but are less than ideal for detailed studies in the multi-scale and multi-measurement modes. Although a GIS by definition does not need a specified scale, the fact that it relies on a relational approach (e.g., rows and tables) makes radical scale shifts difficult. That is, data almost always have an intended scale of use in a GIS system. For a single base map, e.g., a topographic map or satellite image, the GIS approach is very efficient and fast. However, when the user wants to switch between base maps and scales, the tracking of data becomes cumbersome and inefficient. For example, zoning observed in a garnet may be as important as field relations on a 10 km² map, but the two have completely different data types, organizations, and base images. The relational approach demands that we define and declare the structure at the start. Geologic interpretations, which use observations taken at a variety of spatial scales, inherently rely on smooth access to data from a map to a thin section.

The multi-measurement mode also presents problems to a typical GIS approach. Each data entry in a GIS is essentially a line in a table with a unique identity. When taking numerous observations of the same type at the same location, the user of a GIS must continually add more features to a single location. The points are redundant in location as observed on a map, but the GIS does not capture the logic of multiple observations at a single point taken for a single reason. Again, we designed StraboSpot and its underlying data structure to accommodate multi-measurement work.

■ STRABOSPOT DIGITAL DATA SYSTEM

We have developed a data system that accommodates the workflows described above and facilitates tracking hierarchical and spatial relations between structures at all scales as well as documenting cross-cutting relations. The three principal challenges are to: (1) capture and track the hierarchical and spatial relations between structures from the scale of a mountainside to the scale of a thin section; (2) preserve the relationships between observations such as fabric superposition or grouping of bedding measurements to define a fold axis; and (3) allow description and documentation of the structures present so that other workers can understand and reuse data collected (e.g., strike and

dip of foliation with an associated trend and plunge of a lineation). None of these challenges is particularly well met by a GIS-based approach; they need an approach that is based on principles that explicitly preserve geological relationships and temporal information. As a result, we reconceptualized the collection of field data that can address the above challenges, starting with the foundational spot concept. This approach provides a simple and powerful way to link map-scale, mesoscale, and microscale observations (reconnaissance to multi-scale) that can also provide for documentation of observations and space-time substitutions.

Spot Concept: Collecting and Coordinating Data at Different Spatial Scales

StraboSpot is based on the concept of a *spot*. A spot is an area of significance within which a set of observations is relevant. A spot can contain a single measurement or an aggregation of individual measurements to characterize a geologic feature or interpret a geologic concept. A single spot is associated with a user-defined area over which a measurement or quantity is most applicable. For example, a strike and dip measurement may be applied to a meter area or tens-of-meters-wide area of tilted beds; at the laboratory scale, a laser ablation spot might reflect the date of material sampled on one portion of a zircon grain. A spot can be a *point* with a radius of significance, a *line* with a buffer region around it, or a *polygon*. Below, we present a field example that demonstrates how the spot concept organizes data, both spatially and conceptually. The main motivation here is to document observations and the relationships between them across many scales of observation.

Spot Concept: Example from the Twin Sisters Ultramafic Complex, Washington State

Tikoff et al. (2010) reported on the deformation and rheological behavior in the Twin Sisters ultramafic complex, Washington State, USA (Fig. 1). These peridotites are characterized by alternating, subparallel bands of dunite and harzburgite, which host orthopyroxenite bands and dikes that are generally folded or boudinaged. Tikoff et al. (2010) mapped a 100 × 140 m area in detail, measured fabrics (lineations, foliations), mapped normal faults, and documented the shortening or elongation of the orthopyroxenite dikes (Fig. 1). In the lab, they conducted wavelength and thickness analyses of the orthopyroxenite dikes as well as microstructural analyses of the orthopyroxenites and host peridotites.

The data were used for multiple purposes: to determine finite strain and to estimate the relative rheologies of the different rock types. In the Tikoff et al. (2010) study, finite strain came from aggregation of data from all of the folded dikes in the study area and applies to the entire 100 × 140 m field area (Fig. 1, Spot I, largest circle). This is the appropriate spot size for this analysis (~50 m radius circle) and the area over which the analysis is representative. The

overall finite strain is based on individual measurements from folded dikes (Fig. 1, Spots 1, 2, 3, etc.). Each folded orthopyroxenite dike is a spot. The spot size for each fold varies from decimeters to meters, depending on the size of each folded dike. Going down in scale, each orthopyroxenite dike contains several folds (e.g., Fig. 2A). Plunge and azimuth were measured at every fold hinge, and a fold axial plane was estimated where possible from limb orientations for each fold in each dike. These individual measurements define new spots within the spot for each dike (Fig. 2A, Spots A, B, C, etc.). All spots are recorded in a hierarchical relationship that spans from the smaller-scale structures (i.e., fold hinges) to the larger-scale 50-m-radius circle that defines the area for which these measurements are relevant (Fig. 1).

To accomplish the second purpose of this study, which is to estimate relative rheologies, Tikoff et al. (2010) collected samples at the outcrop (e.g., Fig. 2A) and studied them at the thin-section scale (Fig. 2B). Individual grains and microstructural features are smaller-scale observations (Fig. 2B, Spots i, ii, iii, etc.) that are linked to all of the other relevant structures from the thin-section scale (Fig. 2B, Spot A) through the outcrop scale (Fig. 2A, Spot 1) through the map scale (Fig. 1, Spot I). Simply put, the aggregate of all these spots is relevant to understanding the finite strain and relative rheologies within the area represented by the larger, 50-m-radius Spots I and II in Figure 1. The spatially cascading relationships between the various levels of spots are shown in Figure 3A. All of the spots thus distinguish the areas over which the measurements are representative, and track the spatial and hierarchical relationships between them. One of the side benefits of this approach is that it accommodates and integrates areas of heterogeneous and homogeneous deformations, and users can more easily understand how areas with variable deformation can be viewed in terms of an overall amount of finite strain.

Polygons and Lines Are Spots

In the above example, spots are represented as circles with a given radius centered on a point. Spots can have any size, shape, or topology. Conceptually, a polygon is not significantly different from a circle: it applies an arbitrary shape to the area of significance, and can surround other, smaller-scale spots. A very common example is the extent of a formation or rock unit on a map. The unit, or spot, has an area of significance that is defined by its contacts with other units. Lines or traces with a buffer region form a spot that may appear at first to be different from points or polygons, but is in fact conceptually similar. Consider the example of a fault mapped in the field that contains locations along its trace where kinematic indicators are present (Fig. 3B). The fault trace can be a spot, and we may choose to capture the kinematic indicators' locations as additional separate spots, associated with the fault trace spot. The indicators might be on the fault surface itself (slickenlines) or in the volume adjacent to the fault (Riedel shears in rocks next to the fault). In other words, the fault spot may encompass the trace alone, or the scientist may choose for the fault spot to encompass both the trace and other spots associated with it.

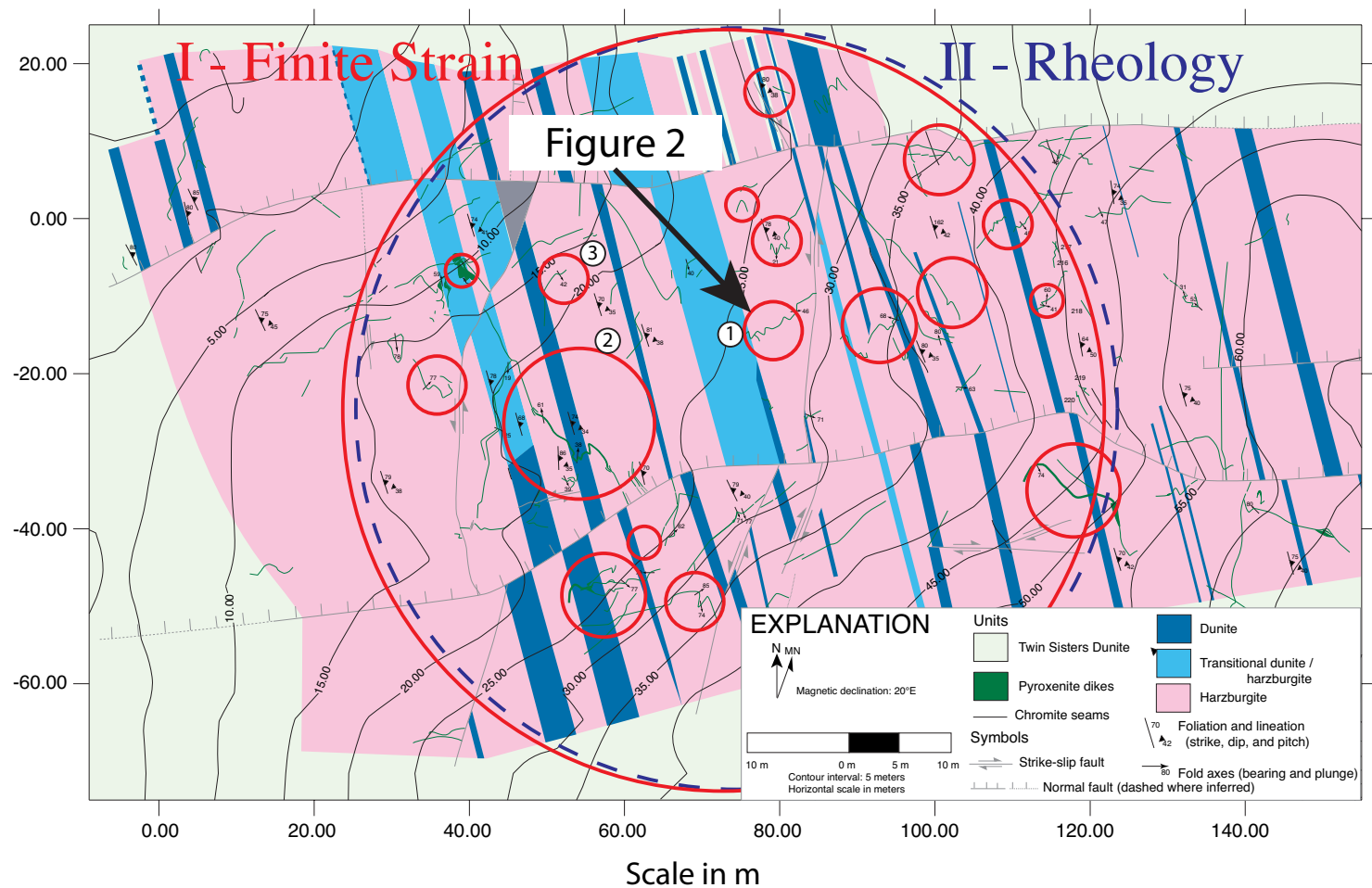


Figure 1. Geological map of a portion of the Twin Sisters ultramafic body (Washington State, USA), modified from Tikoff et al. (2010). This figure, combined with Figure 2, illustrates the concept of spots as a spatial method for grouping data. A *spot* is an area that is characterized by specific observations. Spots can be points, lines, or polygons; here they are shown as circles with differing radii. Each spot can enclose multiple other spots. As a result, spots are an efficient way of recording observations—and grouping them—over a wide variety of spatial scales. Also shown in Roman numerals are two purposes of collecting the field data: I—Finite Strain, and II—Rheology. Purpose is an attribute tied to a project. In this example, the same data are used for two different purposes. Map grid is arbitrary and is in meters (see Tikoff et al, 2010).

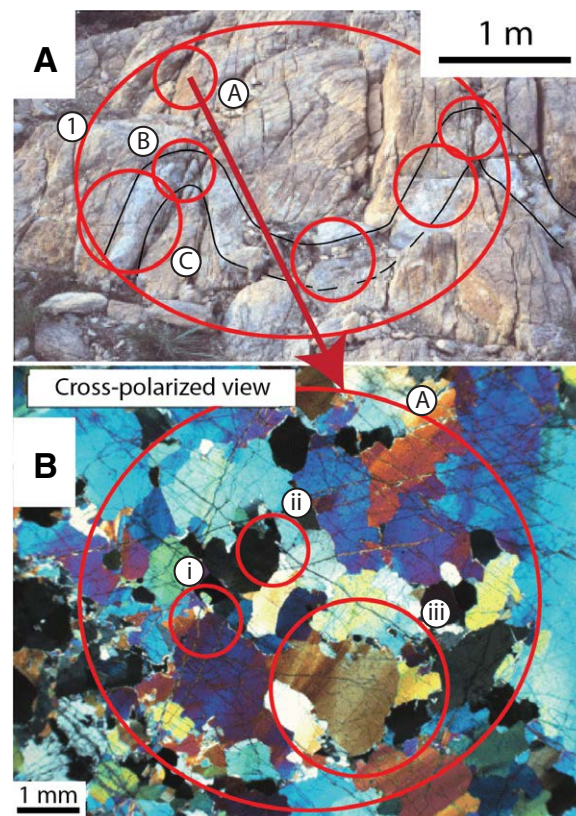


Figure 2. Images from the field area shown in Figure 1 (Spot 1). (A) Outcrop photograph. (B) Thin-section image (cross-polarized light). This diagram illustrates the utility of the spot concept below the resolution of hand-held GPS. Both A and B are image basemaps, meaning that they are capable of containing spots. The spots on the thin section are image basemaps, which are used when higher spatial resolution is required. Each spot is referenced upward (in a spatial scale), so that a spatial hierarchy is recorded; for example, the thin section spot (shown in B) is part of a specific outcrop (shown in A), which is part of the project delineated by Figure 1.

Multiple Measurements within a Spot

There are occasions in which a structural geologist will want to make a series of measurements within a given area, but does not wish to keep track of the location of the measurements with respect to each other: the multi-measurement mode. For example, within the fault described above (Fig. 3B), a worker may choose an area of study (a spot) and make numerous measurements of the fault surface and Riedel shears within that spot, but not record

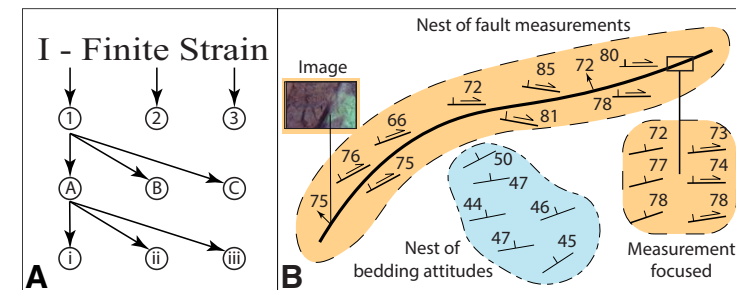


Figure 3. Illustration of the connection of spots with nests, and an example of measurement-focused work. (A) Nesting of spots shown in Figures 1 and 2. Note that the StraboSpot system is built to automatically track the topology of the spatial hierarchy. In this way, we follow spots within spots across all scales. (B) Examples of nests and measurement-focused spots. The fault is a line, but has an area of importance as shown by adjacent shears in the orange field. Hence the fault is a spot containing many other spots. At the lower right are sets of measurements taken at a general point on the fault in measurement-focused work. The blue nest shows the attitude of layers near the fault.

the location of each measurement individually on a map. In this case, each of the individual measurements is not a spot, because it does not have a unique location. Rather, these data will be associated with the larger spot, with both a location and areal extent, that contains them (Fig. 3B). This approach is similar to taking multiple measurements of the orientation of layering or axes of minor folds associated with a larger-scale structure. This method of data collection is perfectly permissible within the StraboSpot data system, although the spatial relations between the different measurements are lost in this tradeoff for efficiency. We call this approach “measurement focused,” to distinguish it from a spot-based approach.

The spot concept facilitates data collection, regardless of scale or workflow. The examples above show how it is possible to move through the spatial hierarchy. At each scale, the spot can be sized and shaped accordingly to reflect the area of significance of measurements and observations. Spots then fit easily into the reconnaissance, mapping, multi-scale, and multi-measurement modes of data collection.

Relationships between Spots

Nesting and Tag Concepts for Grouping Data

One of the powerful aspects of the StraboSpot data system is the ability to group or establish explicit relationships between spots at any spatial or conceptual level. Spots are inherently spatial in nature, and the system intrinsically tracks the spatial relationships between spots. Purposeful spatial grouping of spots is done via what we term *nesting*. Conceptual relationships between spots, such as measurements aggregated to understand a single structure

or assigned to a single rock unit or geologic concept, are tracked using tags. A *tag* is a label applied to spots.

Nesting, the grouping of points to create larger spots, refers strictly to spatial groupings of spots. A new spot, with a larger spatial extent, can be made if multiple existing spots are grouped (e.g., Figs. 1–3). Spots can also be added at smaller scales as more data are collected within an existing spot. Figure 3A shows the nesting for Figures 1 and 2. Spot I, whose purpose is to determine finite strain, is the largest spot, containing observations at smaller spatial scales. In this way, thin-section data (Spots i, ii, and iii in Fig. 2B) are nested within a sample (sample taken at Spot A in Fig. 2A), which is nested within a part of the outcrop (whole spot in Fig. 2A), which is nested within the finite strain Spot I (Fig. 1). In a second example (Fig. 3B), we show that measurements explicitly made for a fault (in this case, fault plane measurements and adjacent Riedel shears or minor faults) can be nested within a single spot. This spatial grouping is also applicable to adjacent bedding measurements. Observations from a limited area on the fault are also put into the larger nest of the fault.

Tags are used in cases where grouping is determined based on a logical or conceptual framework. An example of a conceptual group in structural geology is the designation of a generation of fabrics—foliation or lineation (e.g., S_1 , L_1). In these cases, there may be no meaningful *spatial* correspondence between the spots (other than that they are in the study area), but they have a *conceptual* association based on the geologist's workflow of mapping different generations of fabrics and geological structures. We use a tag as the tool to conceptually link data. Perhaps the simplest example of a tag applies to spots within a single geologic unit or formation. Any field geologist would tag or label these observations with the same unit name.

One or more tags can be assigned to any spot. Consider the following example: A fold has defined layering, measurable fold hinge lines, and modest axial planar fractures (Figs. 4A, 4B). Attributes of layering, foliation, and hinge lines defining the fold are assigned to spots (Fig. 4C). These spots can be nested together to form a larger spot that defines the area of the fold itself, as in Figure 4B. We could add tags to document that the spots shown in Figure 4B are all part of the same unique fold, as in Figure 4D, or a particular fold, fold 1, as in Figure 4B. In this way, the particular structure resides in the overall area of Figure 4A. Individual folded layers can be tagged to show measurements taken on the same layer to distinguish geometry or competency (Figs. 4E, 4F). The axial planar fracture and fold axes define different data from layering, but can also be tagged fold 1. Using this approach, we can specify all appropriate measurements as belonging to a single structure. Hence the attitudes around the fold will be both grouped conceptually (tagged as same fold or same layer) and spatially (nested into a location). In the StraboSpot data system, the information is categorized at both the tag and location level.

Tags can be assigned throughout a field study. For example, if we establish the sequence of folding, then structures and fabrics across an area that reflect a folding generation (e.g., F_2) could be tagged as such. Alternatively, a fold could be tagged as associated with a specific orogenic event (e.g., Laramide) or a particular type of axial planar foliation (e.g., centimeter-scale banded axial planar foliation).

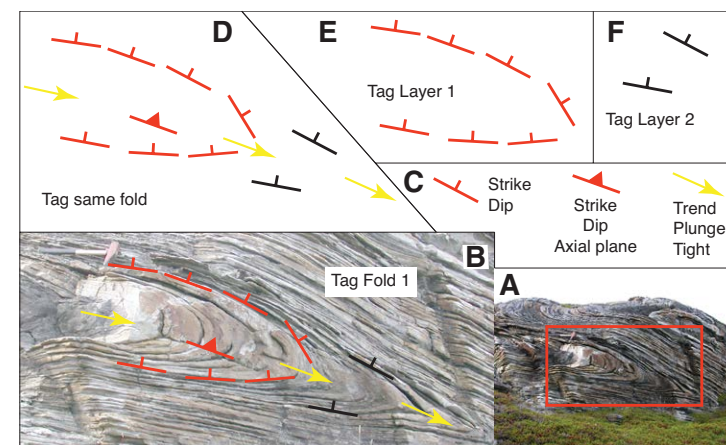


Figure 4. Example of using tags for a fold. (A) Outcrop photo of folded rocks showing setting and scale of subsequent diagrams. Note the rock hammer for scale. (B) A closer view with layering attitudes (red and black), axial plane attitude (red), and minor fold hinge lines (yellow) ornamented with typical geological symbology. (C) Attributes associated with the measurements of the layers (strike and dip layers), foliation (strike and dip and the fact that the foliation was taken along the axial plane or surface), and fold axes (trend and plunge as well as the classification of each fold as tight). (D) Diagram showing that all spots can be tagged as the same fold even though they contain different attributes. In addition, we could add, for example, an axial trace line feature and still add the same tag to the features. This would be cumbersome if using standard GIS techniques. (E and F) Insets showing that different spots can be tagged as different layers, as well as being tagged as the same fold or structure. Images courtesy of Stephen Daly, University College Dublin.

The advantage of tags is their flexibility, in that they can be completely defined by individual scientists. Critically, they are independent of the spatial scale of the observation. Tags also can be used for more complex and complete descriptions in the way lookup tables are used with GIS and databases. Examples include the naming of orogenic events, e.g., Taconic, Acadian, and Alleghanian, or usage such as enveloping surfaces for *en echelon* geological structures. Lastly, tags can be changed or modified as needed as additional information is collected. For example, consider the situation where an igneous unit receives a U-Pb zircon age. This new geochronology information can be assigned to the tag for that unit. If the sampling locality is the critical (e.g., “Rosetta”) outcrop that shows that the previously assigned S_1 fabric is actually S_2 , the user need only update this single tag rather than editing many individual spots.

Relationships

The ability to document space-for-time substitutions is critical in the analysis of geologic structures and tectonics, and is, in fact, inherent in most fields of the geosciences (e.g., principle of vertical succession in stratigraphy). This

ability is built into StraboSpot using what we call *relationships*. Relationships are established explicitly between spots, features, or tags, and consist of such concepts as “cross-cuts,” “includes,” “is included within,” and so on (Fig. 5). Relationships are established between two or more spots, spot features (orientations, samples, etc.), or tags. For example, in Figure 5, the felsic unit contains blocks of mafic rocks and is cross-cut by a later dike. This relation implies that the mafic dike is explicitly younger than the felsic unit, and by direct inference, also younger than the mafic blocks. These types of inferences can be discovered or inferred from information and relationships in StraboSpot, and are listed as implicit relationships in Figure 5. Users can also define new relationships as needed to document the observed geology.

Summary of Approach

The approach outlined above fairly and sufficiently describes data from structural geology (and perhaps most field sciences), and fits or can fit most workflows used by field geologists. A spot is any observation or group of observations with an areal extent and attributes, such as single or multiple strike and dip measurements. Nests organize spots by location, and naturally accommodate changes in scale from the mountainside to the thin section. Tags act to conceptually group data. Any spot can have as many tags as needed to describe its attributes and associations with other spots, both spatial and conceptual. Relationships describe space-for-time substitution used in interpreting geological history.

Using StraboSpot

StraboSpot is a digital system available as both mobile and online applications. These applications fully implement the organization of the spots, as described above (nesting, tags, and relationships). Spots are organized into *data sets*, a collection of spots, and into *projects*, a collection of data sets. Projects are the container for all information. In this way, a project in StraboSpot is similar to a geodatabase in ArcGIS in that the latter holds master information about the schema and logic of the contained features. The main metadata about the spots that form the data is documented at the project level, such as dates of the study, who collected the data, and the *purpose* of the study (i.e., why the data were collected). Data sets are similar to feature data sets in ArcGIS. We use data sets and projects so that users can organize, aggregate, and manage information easily rather than having to navigate or copy individual or larger-scale spots.

Defining the Purpose

Users are asked to define the purpose of each project. Most structural geologists collect data for a specific purpose, which influences the choice of

data collected as well as the context within which the data are understood. For example, Tikoff et al. (2010) attempted to estimate the rheology of the folded orthopyroxene dikes relative to the dunite host rocks. Therefore, they collected data on fold axes and fold limb orientations, but ignored joints. The choice of data collected was determined by the purpose of this project. Thus, the purpose is stated at the project level (in the detailed project descriptions), not at the spot level. Scientists typically do not list the sorts of features ignored, and it is up to the user to infer the relevant data collected from the information given.

The purpose can also be a critical attribute when searching for data in the shared database, helping to refine the data types searched. Future studies may incorporate the data collected with new studies with similar or different purposes. However, the initial purpose will have influenced the types of data available in the original data set. Reuse of data is a critical reason for establishing appropriate metadata and descriptions into the digital data system.

Incorporating Images and Image Base Maps

The incorporation of photographs and other images into data collection for structural geology and the field sciences is becoming more common since the advent of digital cameras and cameras incorporated into mobile devices. The StraboSpot system gives users the ability to add any number of images to a spot (Figs. 2, 3, and 6) and to document the orientation and scale of the image. Any type of image may be used, including a photograph or sketch. At this time, the location and orientation information is entered by the user, but will be eventually captured into StraboSpot using the Exchange Image File Format (EXIF) data from the mobile device. In this way, the outcrop or field relation can be recorded and documented at any location. In addition, users can annotate or perform analysis of any image using applications outside of StraboSpot and still incorporate them into the attributes of a spot (Fig. 6). The system is designed to incorporate images at all scales, from mountainside to thin section, so that the scientist can share observations at any scale, from the field to the laboratory.

It is often critical to preserve spatial relations at a scale below that of a handheld GPS device. For this purpose, we introduce the concept of an *image basemap*. Besides providing rich documentation of field relations, images can be used at any scale for mapping, using an image as a base map. It is common for structural geologists to use handheld GPS to determine location for field data collection. The StraboSpot workflow accommodates a direct link to geolocate spots using the internal GPS of mobile devices, but this approach only works well at a scale above the uncertainty associated with the measurement (typically a few meters). Below the resolution of GPS, a spot or spots can be located directly on an image basemap (Fig. 6). Any type of image (e.g., photograph or sketch) can be used as an image basemap. Mapping on the image basemap is done using spots in the same manner and with the same set of available attributes that one would use with a topographic map.

An image basemap may also be used to locate other images or pictures that in turn can be used as base maps to collect new data as spots. Thus, images

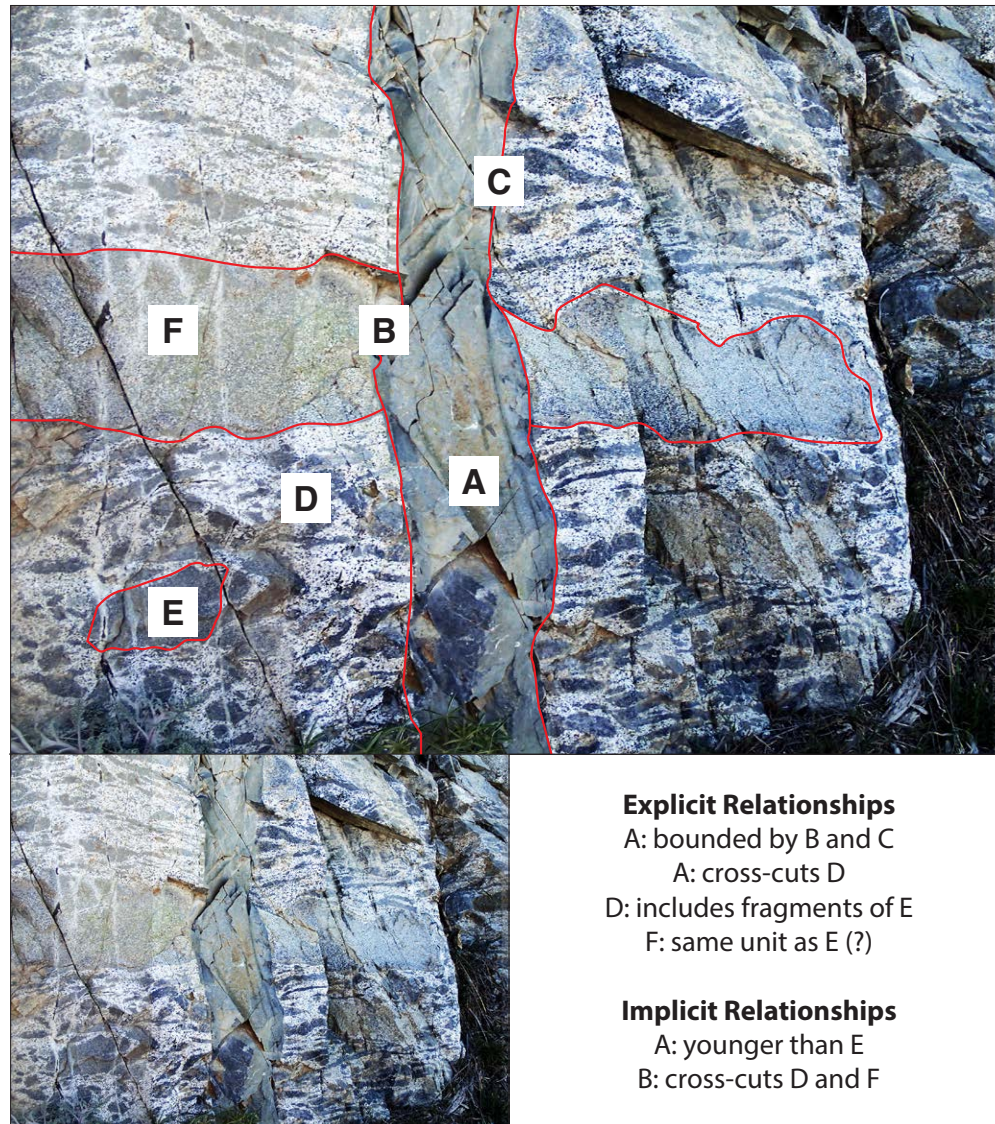


Figure 5. Examples of relationships. Lower left is an outcrop photo. Outcrop is 2 m high. Upper part is the photo with spots added for contacts. Lower right are the relationships documented between spots. Note that we make some relations explicit, that is, they are set up and defined in the StraboSpot application. Others can be figured out from the explicitly stated ones, and we call these implicit or implied relations. For example, in this figure, A cross-cuts D, and D includes fragments of E. Therefore A is younger than E.

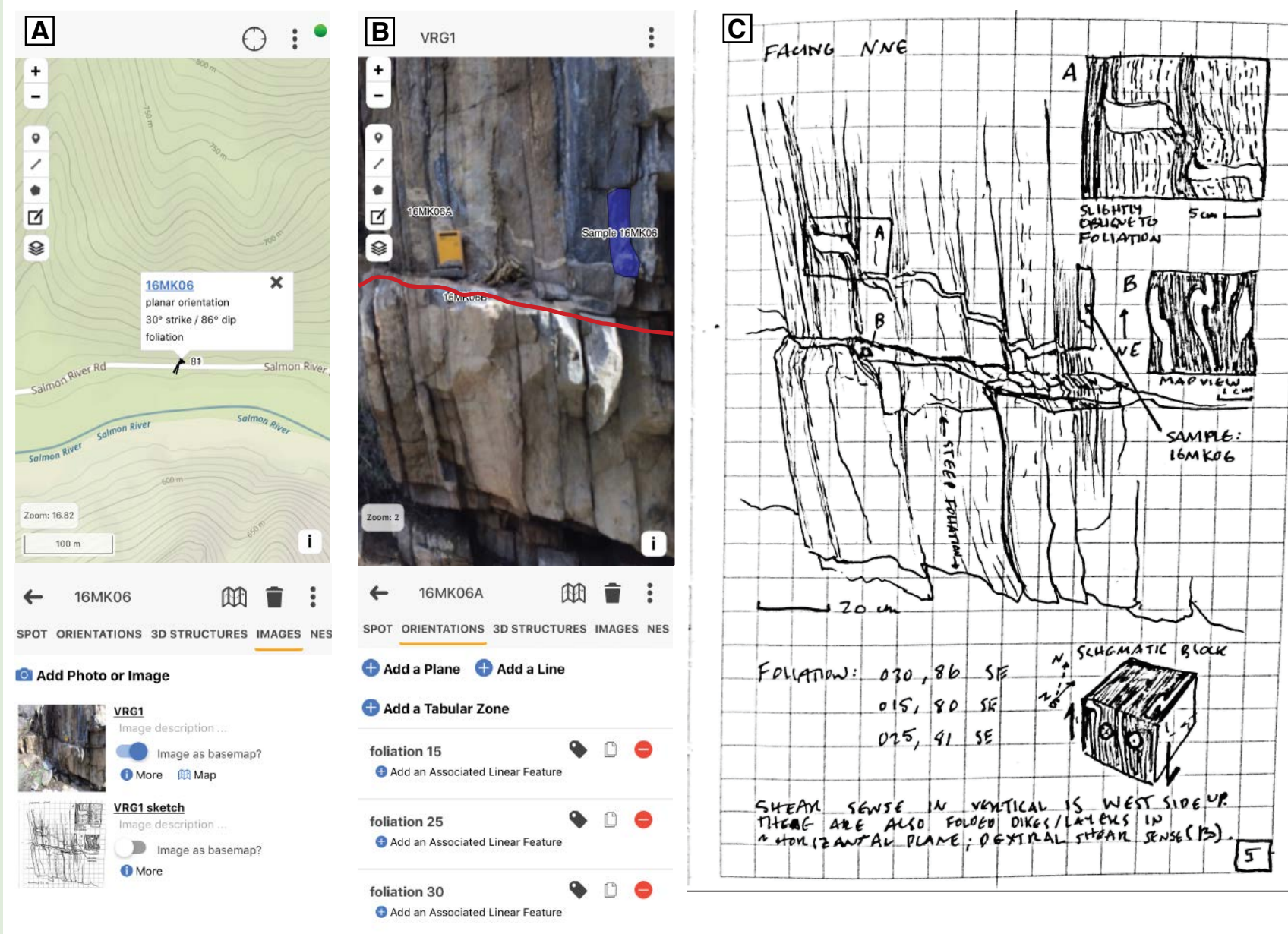


Figure 6. Photograph and field sketch illustrating spatial relationships between images and other field data that can be captured by using image basemaps in StraboSpot. (A) Spot location on a map in the StraboSpot mobile application containing multiple images. (B) Photograph used as a base map for additional spots and associated data. Note that the red line is also a spot on this image. (C) Field sketch with associated data written in a field book. This image is also part of the information for the spot defined in A.

and associated spots are nested in a documented hierarchy. The highest-level image basemap is located in a spot that has GPS or other real-world coordinates, and all smaller-scale spots will be tied to this location through nested images. Images and spots on the smaller-scale images have cascading pixel coordinates related back to the primary location. However, any image can be given refined coordinates, and the user can specify the orientation and scale of any image.

■ UNDERLYING DATABASE STRUCTURE FOR STRABOSPOT

Most attempts at developing geological databases have relied on fitting the field and map information into a relational database structure. This structure is based largely on a controlled vocabulary and individual tables that efficiently store single and unique values in a database. This efficient storage is referred to as normalization. This approach is a very powerful for many types of data, but can be somewhat limiting for information that contains critical spatial relations and hierarchy, and cannot easily accommodate new relationships between data. For that reason, we explored and implemented a different system based on a graph database rather than a relational database.

Graph Database

A graph database is built using the concepts of nodes and edges (Robinson et al., 2015; Fig. 7A). A node in StraboSpot is a spot representing data that have an area over which the data are valid. An edge is a relationship or connection between spots (nodes). Edges can also have descriptions or attributes associated with them. In a graph database, the search or linking between different nodes is done by what is called a transversal of the database. Such transversals are optimized for speed and performance and can easily go across many levels of relationships quite quickly. Transversing from spot to spot is extremely fast. The similar operation in a relational database is called a join. Joins across several levels of information or tables can become quite slow and/or impossible to complete in a reasonable time (Vicknair et al., 2010). For that reason, relational databases are commonly completely denormalized into one or two tables that can be indexed and quickly searched.

The node and edge topology is applied in Figure 7B to the geologic structure shown in Figure 4. It is clear that many nodes have multiple relationships (edges) with other nodes (e.g., same fold, same layer, etc.). The nodes also form an enveloping spot, indicating that they are related to the same structure, similar to the data described along the fault in Figure 3. Using a graph approach allows as many relations as needed to be defined easily and while working (“on the fly”). In contrast, recording the data and their relations using a strictly relational approach is difficult, as mixing data geometries (e.g., fold axial trace lines with points and polygon areas) is not possible. In addition, StraboSpot allows the user to work across a range of scales, going both up and down in scale as necessitated by observations. A relational approach would require the

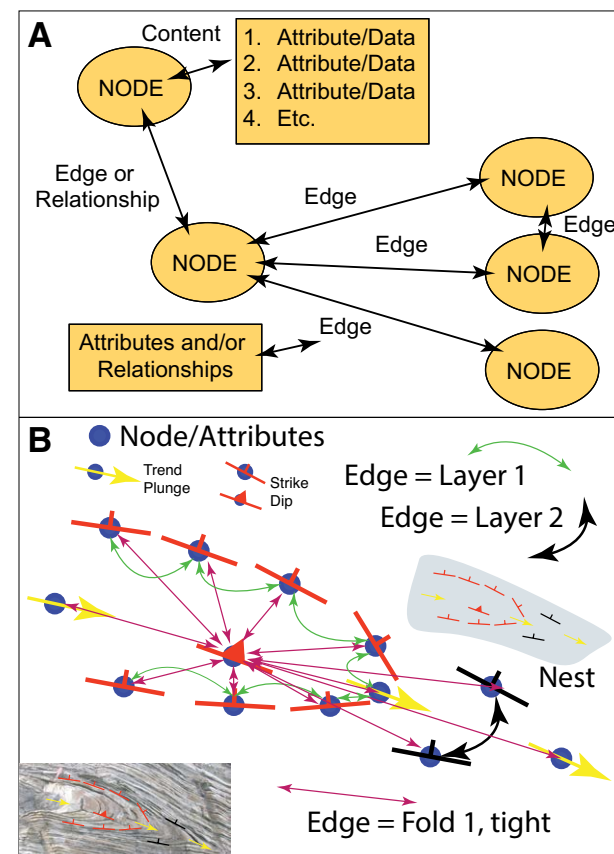


Figure 7. Graph database approach. (A) Topology of nodes and edges, and information associated with both. Nodes and edges can both contain a series of attributes, but edges can also express relationships between nodes. (B) Example from Figure 4 showing how nodes and edges can be used. These are similar to the groupings shown in Figure 4, but show more explicitly the topology of edges between nodes. Image courtesy of Stephen Daly, University College Dublin.

user to choose a starting point, not always possible if we do not know at first where we are in the hierarchy of structures. The graph approach of StraboSpot allows seamless data collection, even when data span a range of spatial scales.

Framework of the StraboSpot Data System

Spurred on by the EarthCube report, the three senior authors organized multiple workshops for the structural geology and tectonics communities.

These workshops focused on how geologists collect field information on (1) shear zones, (2) ductile and three-dimensional structures, (3) faults, and (4) pluton fabrics and migmatites to identify and synthesize the commonalities of scientific data workflows and vocabulary. Each workshop involved geologists from different academic backgrounds to partly alleviate the issue that there are almost as many ways to collect field information as there are practitioners. Recognizing the inherent flexibility needed for collecting data, we organized data into the graph database described above, using Neo4j (<https://neo4j.com>) for the graph database persistence layer. We further recognized the importance of the system for data input (or front end) for users to interact with the database. Such a system needs to function on mobile and desktop devices, be platform independent, and be as interoperable as possible with other systems (e.g., ArcGIS, QGIS, and applications such as Stereonet Mobile [<http://www.geo.cornell.edu/geology/faculty/RWA/programs/stereonet-mobile.html>]). The system must function in both online and offline settings while giving users as much power and flexibility as possible. It must also be open source. Below we describe the technology behind the system as well as capabilities and uses.

Field Application—StraboSpot on Mobile Devices

To accommodate all mobile users, the StraboSpot application runs on both iOS and Android devices. We leverage modern web technologies and develop the code using AngularJS (<https://angularjs.org>) as the structural framework and the Ionic framework (<https://ionicframework.com>) for the user interface so the application operates the same on both platforms. The Apache Cordova framework (<https://cordova.apache.org>) provides the appropriate wrappers for iOS and Android. For this reason, StraboSpot is considered a hybrid application. (All code for StraboSpot is open source, and can be accessed at <https://github.com/StraboSpot>.) The field application stores data locally in an SQLite database. Information can be saved locally to the device in GeoJSON format (<http://geojson.org>), with links to associated images. Data transfer from device to the backend database is made via the REST protocol using the GeoJSON file format.

An animation is included with this paper that shows the use of the mobile application and gives a bit more background on StraboSpot (Supplemental Video²).

The application runs in both online and offline settings. While online, the user can access base images and maps from a variety of sources and can connect directly to the StraboSpot server via an application program interface (API). Offline usage is more complicated in that maps and images must be downloaded to the device while connected to a network. The preloading of georeferenced maps and images is functionality built into the application. Note, however, that the application can still access the mobile device's GPS for real-time locations while offline. Because StraboSpot is a mobile application, offline imagery is downloaded as image tiles from a tile service (Sample and loup, 2010). Users have the option of downloading at many scales (meters to kilometers). The map images are saved locally and are loaded while the device is in offline mode. Image resolution is limited to the scale that the user

downloads. Tile services at MapBox (<https://www.mapbox.com>), MapWarper (<https://mapwarper.net>), and StraboSpot (<https://www.strabospot.org>) can incorporate user-prepared images (e.g., local geological maps). All images and the StraboSpot GPS are referenced to the WGS84 datum.

Web Application

The StraboSpot online version is built using the same codebase as the mobile application. For this reason it has all the data entry and editing functionality offered in the field while giving direct access to the database. It also leverages online connectivity and speed to perform a host of search, download, and export tasks not practical on a mobile device. The online application also allows users to manage and edit stored data, as well as giving them the ability to share their projects with the wider community. Stored information is automatically versioned upon editing, and users can recover any older version as needed.

Users can also upload other data into StraboSpot from the web application. For example, maps or data stored in shapefile format (which is very common in GIS and other applications) can be imported to a project. In this case, the user is led through the steps of aligning the file attributes and vocabulary with that of StraboSpot.

Unique Identifiers

Every spot, tag, nest, or relationship is assigned a unique identification number (ID). The ID is based on time in Unix milliseconds with added random digits, and should be unique. In addition, every Data Set and project are given a unique ID, assuring that the combination of IDs, in practical terms, is never repeated. The ID can be used to fully document the provenance of observations, as it is associated with the person collecting the data.

Extracting Data from StraboSpot

Most of our efforts on the StraboSpot data system have been on developing domain vocabulary, the interface for the mobile device, and the details of managing a graph data system. However, extraction of data from StraboSpot is a critical component for users. Downloading one's own data is relatively straightforward, with an option for downloading a digital chronological log (in PDF file format) of one's data. We call this approach "field book format," because it mimics a traditional field book. However, the power of the database lies in downloading other data and conducting simple searches. Users can currently use the StraboSpot website for searches based on the presence or absence of data types, such as spots with images or orientations. Results can be downloaded in several formats: shapefile, KMZ, Microsoft Excel spreadsheets, and text files formatted for Richard Allmendinger's Stereonet program

Example of StraboSpot Application

Made on an iPad Mini4 using StraboSpot v1.4.1,

This show demonstrates briefly the basic functionality of StraboSpot and some of the operations for adding and relating data. It demonstrates the basic concepts of Spot, Tag, Purpose, and Relationship. The show also gives insight into project and dataset management.

The application runs on either iOS or Android devices. It can be downloaded from either the Apple App Store or Google Play. More information is available at <https://strabospot.org>. An online version is also available at this site.

The application is free. It was developed using funding from the National Science Foundation.

²Supplemental Video. Video shows basic use of the StraboSpot mobile application. Please visit <https://doi.org/10.1130/GES02039.S1> or access the full-text article on www.gsapubs.org to view the Supplemental Video.

(<http://www.geo.cornell.edu/geology/faculty/RWA/programs/stereonet.html>). In all of these formats, as well as the PDF file output mentioned above, all spot data are included with thumbnails of images.

FUTURE PLANS

Working with sedimentary geologists and petrologists, we are starting to add functionality to the StraboSpot system for these field-based geologic disciplines. For petrology, vocabulary and interfaces to collect mineral information will be added to the current system. For sedimentary geology, in addition to new vocabulary, StraboSpot requires the functionality to collect data in stratigraphic columns in addition to on maps. In this case, the column is the equivalent of a map, as described above for field mapping using StraboSpot. To facilitate stratigraphic columns, we are creating a column-based profile in the system to reflect the workflow of the scientist. For samples, we plan on implementing seamless interaction between StraboSpot and the System for Earth Sample Registration (<http://www.geosamples.org>), a registry for the International Geo Sample Number (IGSN) (<http://www.igns.org>).

Microstructures

We are expanding StraboSpot to a desktop environment optimized for the incorporation of microstructural images and data, deformed in nature and in rock-deformation experiments. Naturally deformed rocks will be linked to their field data so that deformation microstructures and interpretations are recorded across scales. Experimentally deformed samples will be linked to experimental metadata and mechanical results. By developing a single digital data system for rock microstructures deformed in experiment and in nature, we can enable the critical interaction between practitioners of experimental deformation and those studying natural deformation.

The advantage of using image basemaps is obvious when moving to the subsample and micrograph (thin-section) scale. Data (e.g., electron backscatter diffraction, cathodoluminescence imaging, geochemical) can be tracked at the micrograph scale and linked to its field and/or experimental context. A difficulty with this approach is developing community standards for how to orient thin sections in space, a topic addressed by Tikoff et al. (2019).

Sustainability

A major impediment for any digital data system is to ensure its long-term stability. We note that most community-driven database efforts are ultimately successful in finding a means for long-term storage. Early successful databases include MagIC (paleomagnetic database; <https://earthref.org/MagIC>), North American pollen database (now incorporated into the Neotoma database;

<https://www.neotomadb.org>), and EarthChem (now part of the Interdisciplinary Earth Data Alliance; <https://www.earthchem.org>). Because of the larger file requirements associated with images and microstructural data, it is likely that StraboSpot will have to look for a different model for long-term storage. While we have not yet determined how to best ensure the long-term stability of the StraboSpot digital data system, we are exploring possibilities, including partnering with a member-supported organization (e.g., Geological Society of America) or joining with existing organizations funded by the U.S. National Science Foundation (NSF) that already store data (e.g., Incorporated Research Institutions for Seismology [IRIS]). An alternative approach is a distributed system that uses, for example, state or university servers. A commercial approach is also possible as long as an open-source and an academic version can be maintained.

Quality Control

A common concern of users of publicly available data is what sorts of quality control and quality assurance measures were applied during data entry and submission. This problem is difficult to address because checks of incoming data are time consuming and almost always beyond the scope of the effort (true here for StraboSpot). Approaches that are scalable to large data sets usually involve guidelines to submitters, outlier checks of information uploaded to the data system, and warnings to users to use such factors as the amount and specificity of metadata to evaluate quality (see discussions by DataONE [2012] and IRIS [2018]). Some materials submitted to StraboSpot will be associated with a peer-reviewed publication. In this case, the publication should be acknowledged and referenced in the contributed data. There may also be cases in which maps are digitized and submitted by someone other than the original mapper or author. For these sources, the original map should be acknowledged as the originator and the role (e.g., digitizer, compiler, etc.) and methods of the contributor clearly specified.

Using Artificial Intelligence and Machine Learning in StraboSpot

Moving forward, the StraboSpot system will take advantage of aspects of artificial intelligence (AI) and machine learning. AI can be utilized to understand better the lexicon of geoscientists and perhaps interpret and render drawn or dictated material into appropriate data entries for the system. Matty Mookerjee and Gurman Gill (Sonoma State University, Rohnert Park, California, USA), in collaboration with StraboSpot, have already started working on machine learning focused on recognition and classification of microstructures (M. Mookerjee, 2018, personal commun.). In the future, data and images in StraboSpot that are sufficiently well described can be the basis for machine learning on geological structures. In some sense, StraboSpot allows the structural geology (and, more broadly, the field geology) community to participate in the “big data” approach that characterizes efforts like the NSF EarthCube initiative.

Student Learning

The approach and application of the StraboSpot data system will also facilitate student learning. First, the integration of multiple geological fields into a single data system provides a single, familiar platform to collect data. In this way, student learning can be focused on the science and less on the application. Second, the dialog boxes and prompts in StraboSpot can be utilized as a means for students to both organize their observations and view field information as data. Third, the fixed vocabulary will act as a guide to help students determine what is useful information to collect in the field (e.g., interlimb angle of a fold). Fourth, although not currently developed, the application will ultimately allow group projects to be completed, in which multiple students can work together on a project. At present, we cannot predict the ways in which StraboSpot will be used for pedagogy. Our focus has primarily been on disciplinary research, but we anticipate that it can be used as an innovative tool for geological education.

■ WHY THE NAME STRABOSPOT?

We adopted the name StraboSpot³ for this effort. The choice of spot comes from our use of the spot concept to organize data. Strabo was a Greek geographer who lived from ca. 63 B.C. to A.D. 24. Strabo's *Geographica* was a descriptive and encyclopedic attempt to characterize the geography of Europe and the near Middle East (e.g., the known world for Strabo), among other knowledge, and the StraboSpot database system is arguably an updated version of this approach. Further, a case could be made that Strabo was the first structural geologist and tectonicist. His writings include the following:

Some, however, may be disinclined to admit this explanation, and would rather have proof from things more manifest to the senses, and which seem to meet us at every turn. Now deluges, earthquakes, eruptions of wind, and risings in the bed of the sea, these things cause the rising of the ocean, as sinking of the bottom causes it to become lower. It is not the case that small volcanic or other islands can be raised up from the sea, and not large ones, nor that all islands can, but not continents, since extensive sinkings of the land no less than small ones have been known; witness the yawning of those chasms which have engulfed whole districts no less than their cities, as is said to have happened to Bura, Bizone, and many other towns at the time of earthquakes: and there is no more reason why one should rather think Sicily to have been disjoined from the main-land of Italy than cast up from the bottom of the sea by the fires of Ætna, as the Lipari and Pithecussan Isles have been.

—Strabo's *Geographica*, 1.3.10 (in Hamilton, 1892, p. 84)

³We originally intended to use simply Strabo as the name of the application, but the name Strabo.org was taken. Moreover, it would be unusual to name a data system after a real person, so we added the "Spot" to the end. The added "Spot" does emphasize the fundamental aspect of the spot concept to the data system.

■ CONCLUSIONS

We present the organizational principles associated with the StraboSpot digital data system for structural geology data. The approach is built on the concept of a spot, which is a specific area that is characterized by one or more specific observations, including other spots. Any spot can have multiple measurements. Below GPS resolution, a spot can be tied to an image basemap (outcrop photo, sketch, etc.). The spot approach inherently allows spatial grouping (nesting) of data, and will be critical as we move to incorporation of microstructural data into the StraboSpot data system. Tags allow for conceptual grouping of data, defined by the user. Examples of tags are designating different generations of structures (e.g., F_2 folds), attributing orogenic timing to geological structures (e.g., Laramide orogeny), or designating shared attributes between spots (e.g., presence of alteration). Space-for-time substitutions are done by establishing relationships between spots. We utilized a graph database for the StraboSpot digital data, because of the increased efficiency of transversing across spots and the flexibility in specifications of the data.

The StraboSpot system is an attempt to move the structural geology and tectonics community into the era of big data. We have engaged the sedimentary geology and petrology communities to take this same approach to documenting data, and expansion of the capabilities and backend of StraboSpot are ongoing. Other disciplinary communities in the geological sciences (e.g., seismology, geodesy) have agreed to share data, with general consensus about data reporting and shared resources such as tools. While the original agreement within these fields was based on the necessity of sharing expensive equipment, the shared data have unambiguously improved the quality of science and expanded the possible types of science within these fields. It is our hope that the StraboSpot digital data system will have a similar positive effect on the field of structural geology and, more generally, on the larger field of geology.

ACKNOWLEDGMENTS

Discussions with many workers contributed to our development of StraboSpot. Rick Allmendinger provided inspiration on programming for mobile devices and an understanding of how scientists filter observations for a purpose. Danny Stockli helped with digital mapping at the University of Kansas for years and helped solidify the pedagogical aspects of our work. We acknowledge Jean Crespi, Chris Gerbi, and Emily Peterman for their contributions in coming up with the spot concept. We gratefully acknowledge all of the structural geologists who participated in the 2012 Structural Geology and Tectonics End User workshop and the small-group workshops that helped to define vocabulary. Katie Graham, Jeff Oalman, and Clay Campbell also provided valuable field testing. The paper is much improved owing to reviewer comments and suggestions from Terry Pavlis and Rick Allmendinger. We particularly thank Rick Allmendinger for the motivation to recast the introduction, and acknowledge his significant contribution to it. Support for this work was provided by U.S. National Science Foundation grants to Newman (EAR 1347323 and ICER 1639749), Tikoff (EAR 1347285, ICER 1639549, and ICER 1639748), and Walker (EAR 1252279, EAR 1347331, ICER 1639734, and ICER 1639738).

APPENDIX: GLOSSARY

GIS. Geographic information systems. A combination of hardware, software, and users that stores and exposes information about spatially referenced data. GIS typically store data in tables defined by the geometry of observations (point, line, or polygon).

Graph database. A data structure that consists of nodes and edges. Attributes at nodes are linked through edges, which express the relationship between the nodes. Edges can also have complex attributes. Graph databases are schemaless, meaning they accept any structure of data. A graph database works more like a social network than a GIS.

Nest. A spatial relation between spots, recording the hierarchy of observations. For example, all observations at an outcrop are nested within the area of the outcrop. Thin sections from a sample are nested directly with the sample and, at a higher level, within the outcrop from which the sample came. Nests are strictly spatial in organization.

Relational database. A database founded in relational algebra. Data are stored in tables that are related to each other by keys. For example, a person's address is an entry in one table with an entry number. The entry number serves as a key to another table that contains the person's name. The database is ideally constructed so that information only appears once in one table, a process called normalization.

Purpose. The reason a study is done; the purpose influences the choice of data acquired. For example, a geoscientist studying deformation structures may measure layering and fold axes, but ignore other features such as joints or trends in grain size.

Space-for-time substitution. Using physical relations or spatial patterns to infer time or sequence of events. In geology, this consists largely of cross-cutting relations, such as a dike cross-cutting other rocks, or overprinting of one deformation fabric on another.

Spot. A spot is an area of significance within which a set of observations is relevant. A spot can contain a single measurement or an aggregate of individual measurements to characterize a geologic feature or interpret a geologic concept. A single spot is associated with a user-defined area over which a measurement or quantity is most applicable.

Tag. A flexible way to assign conceptually related attributes shared by a number of spots. An example is a geological unit description, commonly shown as a unit label, that conveys information about rock type, formation, age, etc.

REFERENCES CITED

- Barnes, J.W., and Lisle, R.J., 2004, *Basic Geologic Mapping* (fourth edition): Chichester, UK, Wiley-Blackwell, 184 p.
- DataONE, 2012, Primer on data management: https://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf (accessed October 2018).
- ESRI, 2018, Geology data model: http://downloads2.esri.com/support/TechArticles/Geology_Data_Model.pdf (accessed October 2018).
- Hamilton, H.C., translator, 1892, *The Geography of Strabo*: London, George Bell and Sons, 519 p.
- IRIS (Incorporated Research Institutions for Seismology), 2018, Data quality assurance: <https://ds.iris.edu/ds/nodes/dmc/quality-assurance> (accessed October 2018).
- Jenny, H., 1941, *Factors of Soil Formation*: New York, McGraw-Hill, 281 p.

- Laxton, J.L., and Becken, K., 1996, The design and implementation of a spatial database for the production of geological maps: *Computers & Geosciences*, v. 22, p. 723–733, [https://doi.org/10.1016/0098-3004\(95\)00128-X](https://doi.org/10.1016/0098-3004(95)00128-X).
- Pavlis, T.L., Langford, R., Hurtado, J., and Serpa, L., 2010, Computer-based data acquisition and visualization systems in field geology: Results from 12 years of experimentation and future potential: *Geosphere*, v. 6, p. 275–294, <https://doi.org/10.1130/GES00503.1>.
- Robinson, I., Webber, J., and Eifrem, E., 2015, *Graph Databases: New Opportunities for Connected Data* (second edition): Sebastopol, California, O'Reilly Media, 238 p.
- Sample, J.T., and Ioup, E., 2010, *Tile-Based Geospatial Information Systems: Principles and Practices*: Boston, Springer Science+Business Media, 237 p., <https://doi.org/10.1007/978-1-4419-7631-4>.
- Sen, M., and Duffy, T., 2005, GeoSciML: Development of a generic GeoScience Markup Language: *Computers & Geosciences*, v. 31, p. 1095–1103, <https://doi.org/10.1016/j.cageo.2004.12.003>.
- Shipley, T.F., and Tikoff, B., 2018, Collaboration, cyberinfrastructure, and cognitive science: The role of databases and data guides in 21st century structural geology: *Journal of Structural Geology*, <https://doi.org/10.1016/j.jsg.2018.05.007>.
- Tikoff, B., Larson, C.E., Newman, J., and Little, T., 2010, Field-based constraints on rheology of the lithospheric mantle, Twin Sisters, Washington: *Lithosphere*, v. 2, p. 418–422, <https://doi.org/10.1130/L971>.
- Tikoff, B., Chatzaras, V., Newman, J., and Roberts, N.M., 2019, Big data in microstructure analysis: Building a universal orientation system for thin sections: *Journal of Structural Geology*, <https://doi.org/10.1016/j.jsg.2018.09.019> (in press).
- Tomlinson, R.F., 1974, The application of electronic computing methods and techniques to the storage, compilation and assessment of mapped data [Ph.D. thesis]: London, University College London (University of London), 444 p.
- Walker, J.D., and Black, R.A., 2000, Mapping the outcrop: *Geotimes*, v. 45, no. 11, p. 28–31.
- Walker, J.D., Black, R.A., Linn, J.K., Thomas, A.J., Wiseman, R., and D'Attilio, M.G., 1996, Development of geographic information systems-oriented databases for integrated geological and geophysical applications: *GSA Today*, v. 6, no. 3, p. 1–7.
- Walker, J.D., Black, R.A., Berry, A.K., Davis, P.J., Andrew, J.E., and Mitsdarfer, J.M., 2002, Geologic compilation map of the northern Mojave Desert and southwestern Basin and Range Province, California: Explanation of maps on CD-ROM, in Glazner, A.F., Walker, J.D., and Bartley, J.M., eds., *Geologic Evolution of the Mojave Desert and Southwestern Basin and Range*: Geological Society of America Memoir 195, p. 297–299, <https://doi.org/10.1130/0-8137-1195-9.297>.
- Westoby, M.J., Brasington, J., Glasser, N.F., Hambrey, M.J., and Reynolds, J.M., 2012, 'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications: *Geomorphology*, v. 179, p. 300–314, <https://doi.org/10.1016/j.geomorph.2012.08.021>.
- USGS NCGMP (U.S. Geological Survey National Cooperative Geologic Mapping Program), 2018, GeMS (Geologic Map Schema)—A standard format for digital publication of geologic maps: https://ngmdb.usgs.gov/Info/standards/GeMS/docs/GeMSv2_draft7g_ProvisionalRelease.pdf (accessed May 2018); see also <https://ngmdb.usgs.gov/Info/standards/GeMS/>.
- Vicknair, C., Macias, M., Zhao, Z., Nan, X., Chen, Y., and Wilkins, D., 2010, A comparison of a graph database and a relational database: A data provenance perspective: Article 42 in *Proceedings of the Association for Computing Machinery 48th Annual Southeast Regional Conference*, Oxford, Mississippi, 15–17 April.