How Students Find, Evaluate and Utilize Peer-Collected Annotated Multimedia Data in Science Inquiry with Zydeco

Alex Kuhn¹, Brenna McNally², Shannon Schmoll², Clara Cahill², Wan-Tzu Lo², Chris Quintana², Ibrahim Delen²

¹Department of EECS
University of Michigan
2260 Hayward St.
Ann Arbor, MI 48109
kuhnalex@umich.edu

²School of Education
University of Michigan
610 East University
Ann Arbor, MI 48109
brenna.mcnelly@gmail.com, {schmoll, claracah, loclair, quintana, idel}@umich.edu

ABSTRACT
Scientific inquiry can be more authentic and meaningful to students when using personal and peer-collected data. The challenges of organizing and evaluating a potentially large amount of data can be overcome through the use of annotations (title, tags, and audio notes). We created Zydeco, a multi-component system that students use to collect annotated multimedia data from a museum (using a smartphone app), and then create a scientific explanation with their personal and peers’ data (using a tablet app). We ran a classroom study with 54 students (ages 11-13) investigating how students searched for, evaluated, and used annotated data to construct a scientific explanation. We found that tags supported data interpretation, while title searching and panning through the unfiltered data set supported finding and using data.

Author Keywords
Learner-centered design; tagging; mobile computing; children and technology; science inquiry

ACM Classification Keywords
K.3.1 [Computers and Education]: Computer Uses in Education – Computer-assisted instruction
H.5.2 [Information Interfaces and Presentation]: User Interfaces - User-centered design

General Terms
Design

INTRODUCTION
Science inquiry, the act of posing questions and gathering and evaluating empirical evidence, is a critical component of science education in K-12 settings [1]. One important aspect of science inquiry involves collecting, managing, and analyzing different types of data to construct a scientific explanation that addresses the question that has been posed.

Traditionally, many K-12 science inquiry activities have been focused within the classroom. Explanation construction tools for science inquiry, such as SenseMaker and Belvedere, have been designed to support students in utilizing curated or web data in this context [18]. Field trips or similar experiences that take learners into informal environments are infrequent, and their complexity can make it difficult to engage in substantive scientific data collection. The new generation of mobile technologies, coupled with contemporary networking and cloud technologies, promise to provide new and enhanced opportunities for data collection by students and teachers in informal environments, such as museums, nature parks, etc. For example, students can collect and annotate multimedia data in the field and then return to the classroom to access their data as well as data collected by their peers, bringing the richness of these out-of-class experiences.

Using personal and peer-collected data can make inquiry tasks more authentic, motivating [15] and can enable students to view data they otherwise wouldn't have encountered in the field. The Participate project involved data collected by and shared amongst students (ages 13-15); using Google Earth and graphing software on a laptop in conjunction with data collection equipment, a GPS device, and a mobile phone [20], students were able to explore visualizations showing large amounts of numerical data. They found this student generated data was very engaging back in the classroom and provoked a lot of discussion. Fraser’s SENSE project, which had students (ages 10-14) collect environmental data from the field and share this information back in the classroom with other students, found the process of viewing another classes’ data helped students reflect and understand new perspectives on the data [9].

While using peer-collected data during science inquiry can be beneficial to students, and it may be advantageous to expand the ways that students can collect and review scientific data, using self-collected and peer-collected data...
can also pose distinct challenges to students who are not experienced with the science inquiry process. In addition to the constant challenge of reflecting upon and interpreting data for their explanations—where students must also deal with management issues as they organize and search through the large amounts of data—the entire process is made more difficult since the data is messier and potentially inaccurate due to being collected by peers, as opposed to utilizing a curated data source [3].

Mediating the search process with curated data sources or Internet data searches for young learners has been widely researched [8, 12]. Studies looking at how children search for information on the Internet have found them to have different strategies than adults, often having chaotic search habits that employ a variety of strategies, though they typically prefer and are more successful at browsing for data instead of using keyword searches [11]. Bilal ran three different trials with children (aged 12-13) searching on Yahoooligans (a site designed for 7-12 year olds), looking at them doing a fact-based search, research-based search, and defining their own search task [4]. She found that students fared better with browsing strategies (misspelling and syntax issues hindered keyword searches), and also had a greater interest and task completion success when they defined their own search task versus being assigned a task. However, while students are best at browsing by categories, students would need to accurately assign large data sets into categories or otherwise organize the data during collection for this to be useful.

One method used to address the challenge of organizing and searching data has been the use of titles and tags as metadata, which is commonly used on social media websites such as Flickr and Del.icio.us [2]. There is a growing amount of research looking at the use of tags, the motivation behind why people use tags, and the impact of tags [2, 5, 10]. For example, Ames found that tags people applied to photos with Flickr and Zonetags helped people organize and describe their images, and tags help communicate the purpose of the image, both for personal recall later or to others viewing the image [2]. Without these identifiers, access to data can become tedious. OOKL (previously called Myartspace) is a handheld program that enabled users to record photos, audio or textual notes in a museum and access them later, though they were unable to link the text, audio, or photos to each other and had no metadata beyond the title [19]. While students enjoyed the activity, the lack of identifying metadata caused them to take a great deal of time reviewing data—more than was available for the activity. MobiTOP explored a very different approach: using hierarchical tagging to enable users to comment on photos by having a tag within a tag [17]. The tagging system received a mixed response from the college students who used it to collect data at geological sites while peers were in a lab suggesting tests. The complications of the tagging process and poor image quality made it difficult for students not involved with the collection process to understand the data.

But while research has focused on designing tagging systems and on the organizational and management aspects of tags, there has been little work on how learners search for, evaluate, and utilize peer-collected annotated data in an educational context. Tags have the potential to not only support data organization and retrieval, but also to support the evaluation and analytical tasks that students struggle with during the inquiry process. In this way, they could become an invaluable tool in connecting formal and informal learning contexts during inquiry, while supporting students in utilizing their peers’ data.

THE ZYDECO PROJECT

We are exploring these possibilities within a project called Zydeco, which is examining the use of mobile computers (e.g., iOS devices) and cloud technologies to support students in engaging in science inquiry activities across formal classroom and informal field contexts, such as science museums, nature parks, etc. [6, 13, 14]. Zydeco consists of three components: (1) a website where students set up and access their investigation questions, hypotheses, and data, (2) a data collection/annotation component implemented as an iPhone/iPod app, and (3) an explanation construction component implemented as an iPad app where students view and use the data they and their peers have collected to construct a scientific explanation addressing the questions they are exploring.

We have implemented a version of the data collection component where students take photos or videos of interesting field artifacts and phenomena as appropriate, and then annotate those artifacts with (1) textual tags (previously created in the classroom with teachers or devised in the field), (2) textual titles (choosing from a pre-defined list or creating their own), or (3) audio notes. In previous work, we found that students were largely able to reflect upon the data they need to collect and tag accurate and potentially useful data with the Zydeco data collection app [13], and that recording audio notes encouraged some students to be more mindful while annotating their data [7]. We also found that the term “tags” was confusing to students and the term “labels” was more understandable.

We are now doing more work to explore the impact of these annotations (titles, tags, and audio notes), not only in terms of data management and organization, but also in terms of the reflective and analytical aspects of inquiry, where students need to assess their data in order to see whether it can be used for their scientific explanation, and how it might connect to the explanation. In this paper, we report on a study where we explored the following questions:

- What is the impact of annotations on how students search through and evaluate data?
- How do students find and choose the evidence they use in their explanation?
How do Students Find and Choose the Evidence They Use in Their Explanation?

We ran a study in three public-school classrooms with middle school students who used Zydeco to collect data from a local natural history museum. Students spent 4 days (1 hour a day) using the explanation construction component to evaluate their data and use that data to create their explanations. We found that the annotations supported inquiry activities: tags supported data interpretation, titles supported data identification and searching. However, while audio notes showed the potential to convey additional meaning we found they were often not used appropriately.

USE SCENARIO

In the following scenario, we illustrate how a pair of students generated their explanation using the explanation construction component of Zydeco. While the focus is on how students manipulate and interpret data, we provide perspective for the activity by summarizing the preparation and inquiry leading up to explanation construction, as well as the activities after developing explanation, which includes peer critique and explanation presentation.

Classroom Preparation and Inquiry

Alice and her partner Bob are assigned an animal, the Sandhill Crane, and two investigations in their 6th grade science class: Determine how three other animals are related to your animal, and Describe how these animals could survive in Michigan. After spending several days doing pre-activities to learn about animal traits with their class, Alice and Bob research the internal and external traits of the Sandhill Crane. They create a list of these traits, which is combined with the lists from all of the other groups that were assigned the Sandhill Crane. This master list of traits is input into the Zydeco system as potential tags for students to use in the museum. After collecting a piece of data, they are provided the list of potential tags and can tap on a tag to apply it, or create their own tag. When adding a data title, the students can type in a title or select from a filterable list of the 165 exhibits at the museum that the students may visit, which provide potential title choices.

At the museum, Alice and Bob are each given an iPod Touch to use during for data collection at the museum and they each collect 8 pieces of data. They tag each piece of data with 2-3 tags, and they each have two pieces of data that are additionally annotated with an audio note.

Explanation Synthesis

The teacher introduces the students to the claim—evidence—reasoning framework, which was developed to assist students in constructing explanations by breaking the task into manageable parts: making a claim, providing evidence to support the claim, and presenting reasoning that links the evidence to the claim. Then each group is given an iPad loaded with all their peers’ data in the Zydeco explanation constructor system. After being introduced to the system and completing a short tutorial, Alice and Bob begin working on the first investigation, Determine how three other animals are related to your animal, which is always visible at the top of the screen. They are taken to the investigation guide, where they can see the steps they need to do to complete their investigation. Alice hits “Go Here” next to the prompt that tells them to review the data that was collected, and is taken to the Evidence page (Figure 1).

As Alice is scrolling down through thumbnails of the data collected by the class (Figure 1), Bob suggests that they look at the data they collected. By tapping “group” at the top of the screen, Alice and Bob are able to look at the 16 pieces of data that they cumulatively collected on their field trip. Just like in the class data, they are able to view the titles and tags they applied to each piece of data below each image thumbnail.
Bob wants to take a better look at the data he collected on a Moa. He taps on the thumbnail to enlarge the image and can now listen to the audio note he took in the museum (Figure 2). Alice really likes the tag "small brain" that Bob had put on the Moa, and tells him that they should search for that tag in the class data. Scrolling through the tag cloud on the left of the screen (Figure 1), Alice finds and applies the tag “Small brain” as a search term. All the data annotated with that tag appear, including an image titled Deinonychus, which also has tags “backwards knees” and “feathers”—which matches the Sandhill Crane! Bob gets excited and does a title search for more Deinonychus data. There are two data with really good pictures of the animal, but one of them has a tag “fluffy” that Alice thinks is silly, so Bob selects the other picture. The tags on this picture are “backwards knees”, “hollow bones”, and “feathers”; Alice and Bob agree that this is a good piece of data because the tags match both the image and their animal and so they select “use”. They create a group called “Deinonychus” and are taken back to their explanation page.

Alice and Bob continue searching for new evidence until they have three groups: Deinonychus, Moa and Pigeon. Alice taps on the Claim tab and, following the prompt, writes that, “The mostly closely related animal to the Sandhill Crane that we have is the Deinonychus, followed by the Moa and then the Pigeon”. Bob agrees with her, and in the Reasoning text area writes that, “Because animals that have the most traits in common are more closely related, and the Deinonychus has six traits in common with the Sandhill Crane—such as backwards knees, eyes on side and feathers—that it is more closely related than the Moa or Pigeon because they have fewer in common. The Moa is next closest because it has five traits in common and the pigeon is last because it only has two.” They now have finished their first draft of their explanation.

The evidence groups and reasoning from an explanation completed by a group in our trial can be seen in Figure 3.

**Feedback and Presentations**

Alice and Bob do peer critique with another group, using the system to provide feedback on specific components of the explanation their peers created. Once they received feedback on their explanation and made revisions for the relatedness investigation, they went on to the second investigation, which looked at how the animals they chose in the relatedness investigation would survive in Michigan. At the end of the week they presented both of their explanations to their peers.

**FINDINGS**

Our results come from a study with 54 students (ages 11-13) from three different classes at a diverse suburban school where 62% of the students get free or reduced lunch. Students had 6 hours of pre-activities to learn about animal traits (1 hour a day), 1 hour of data collection in a museum, and 5 hours (1 hour a day) of post-activities to create an explanation, four of which were spent using the tablet.

Of the 54 students, six pairs in each class (18 total pairs out of 27 across the three classes) were given approximately five-minute daily in-process interviews using a semi-structured interview protocol while they were using the tablet. This interview included daily questions and one-time questions asking about various aspects of the system. Fifteen of the pairs were randomly selected, while one pair from each class was chosen by the teacher for their ability to provide good feedback. Due to time constraints, only 18 pairs were selected for interviews and we were unable to get responses to every question from each pair.

Usage logs of the system, generated by all 27 pairs, were analyzed along with the field notes from four researchers who were present in the classroom. The data the students collected in the museum was analyzed and the students’ explanations were graded following a rubric; the explanation grades are presented in this paper to provide context for the rest of the study.

**Characteristics of the Data**

In order to provide context for how the students were interpreting and using data, we first examined the quality and characteristics of the data set to assess the difficulties students would encounter finding and analyzing the data.

The three classes collected 676 pieces of data in total. The data included work from students in the class who were not part of this research study, which was excluded from further analysis. This left 474 pieces of data, 40 of which were audio notes that we requested of the students at the end of the field trip to tell us what they learned. Those 40 pieces of data were also excluded from analysis. Of the remaining 434 pieces of data, 327 were photos (75%), 9 were audio notes (2%), and 98 (23%) were a photo that had an audio note appended to it. There were a total of 1134 tags applied
to the data (average of 2.6 tags per object) and out of these tags, 945 were tags generated during pre-activities (83%).

Tag Characteristics
Each of the tags was analyzed by a member of the research team and coded as to whether it was inaccurate or nonsensical in its representation of the data, accurate to the data, or if there was uncertainty regarding whether it was accurate or not. The breakdown of the tags was 1016 (90.1%) accurate, 84 (7.4%) inaccurate or nonsensical, and 27 (2.4%) uncertain (Cohen's Kappa = 0.82).

The accurate tags were then analyzed as to whether they could potentially be useful to the investigation or not. For instance the tag “New to me”, while accurate from the student’s perspective, is not helpful in ranking animal relatedness or determining survivability. In fact, we found that all of the accurate tags were potentially useful to the investigation except for the one instance of “New to me” (100% IRR agreement). Overall the data was tagged well, the vast majority (90%) of the tags were both accurate and potentially useful to the investigation; though that left a portion of data (~10%) that students would be searching through that was incorrect or nonsensical.

Title Characteristics
Unlike tagging, students preferred to create their own titles, with only 152 (35%) data items being titled with the predefined exhibit names. Since some of the methods of searching through the data relied on accurate and correctly spelled titles, the titles were coded as to whether the title accurately described the animal and whether it was spelled correctly. All of the titles were first coded if it was giving the name of the animal (duck, t-rex, dino mastodon), a general family name for the animal (bird, dinosaur, dino), a general descriptor (predator, large), or gibberish (amm, ch) and had a Cohen’s Kappa = 0.93. The vast majority (80%) of the data was titled with the name of the animal, which we expected and encouraged given the activity and since it increased the ease of searching for the data. Of the remaining data, 5.3% were titled with a more generic classification of the animal (fish, canine), 9% were either named after a trait or short description of the animal, and 5.7% of the titles were gibberish (amm, ch).

Next all the titles (except those that were gibberish) were analyzed as to whether they were accurate (either naming the correct animal, family, or the description was accurate), not accurate, or unclear. Of these, 367 (89.7%) of the titles were accurate, 22 (5.4%) were unclear, and 20 (4.9%) of the titles were not accurate (Cohen’s Kappa = 0.86). Often the animal was incorrectly titled (for example, a possum was titled “owl”). In this case, extra effort and knowledge is needed on the students’ part to identify and discard the data in these cases to avoid using incorrect evidence.

Lastly, we looked at the spelling of the titles. Most of the titles were correctly spelled, with 371 (90.7%) of the titles having the correct spelling, though the remaining objects were incorrectly spelled which would lower the chance that they would show up in the search results if looking for a specific animal (100% agreement between reviewers). The high percentage of accurate, correctly spelled titles that are the name of the animal will make the task of searching by title easier on the students as they don’t have to try and guess the title of the data they desire, increasing the utility of searching on titles.

Audio Note Characteristics
The pieces of data that had an audio note were analyzed for accuracy and potential usefulness of the data. Each audio note was labeled as to whether what was said was accurate (no statements made that were inaccurate), partially accurate, or not accurate. Additionally, these notes were analyzed as to whether they were potentially helpful to the investigations or not helpful at all. For the audio notes, 81 (76.4%) were accurate, with 14 (13.2%) partially accurate and 11 (10.3%) inaccurate (Cohen’s Kappa = 0.79). From these notes, 69 (67%) were found to be potentially helpful in answering the investigation (Cohen’s Kappa = 0.89). Because the audio notes had lower accuracy and usefulness compared to titles and tags, it may be less helpful for students to review and listen to audio when evaluating data.

In summary, the data was overall quite accurate and in over 80% of the cases the data was annotated in the manner that the system was structured around supporting (accurate tags, titles with the name of the animal that are accurate and correctly spelled). However, in each of these categories there was still a fair amount of data that was not accurate or more difficult to search and find due to the nonstandard labeling or incorrect spelling. The audio notes were also less useful for interpretation of the data as they had the lowest accuracy (76.4%) and only 67% of the audio notes were potentially helpful in the investigation.

What Is the Impact of Annotations on How Students Search Through and Evaluate Data?
Students could use the data annotations as a mechanism to organize and filter the data, as well as analyze the data by looking at or listening to the annotations. Below we present the strategies students used to search and evaluate the data.

Annotations Impacting How Students Search
The different search strategies that were available to the students in this trial included: searching by the title of the data, searching by tags present in the data, looking at their own data, or looking at the entire data set. The groups employed a mix of strategies and spent radically different amounts of time viewing data (ranging from 16 to 113 minutes, median of 48 minutes). We analyzed the usage logs to determine the amount of time each group spent searching via each annotation type as well as without any filtering to see the effect annotations had on searching.

Using tags was the least common method of searching. While all but 2 groups tried tag searching, only 12 groups spent at least 5 minutes using this method. Only two groups
spent over 25% of their time looking at data by tags (the heaviest user spending 38% of their time looking at data, a little over 22 minutes).

The next method that students used was browsing through their own data. Because of the small amount (average 16 pieces) of data collected by each group, all of the data could be viewed onscreen quickly and so annotations did not play a large role in search habits. Two groups spent the majority of their time browsing their own data, with a total of 14 groups spending 5 or more minutes browsing through their own data. All but one group spent time reviewing their own data, with 20 of the groups spending at least two minutes reviewing their data. The class average for time spent looking at their own data was 17% of their total time looking at data.

The next most common strategy that students used was title searching, with 17 of the groups spending 5 or more minutes searching by the title and a class average of 22% of their time looking at data being employed this way. Seven groups used title searching as their primary means of locating data (spending up to 42 minutes looking at data in this fashion). However, not all groups took this approach, one never using this feature and another five groups spent less than two minutes searching in this fashion.

With an average of 54% of their data browsing time being spent in this way, panning through tiled images of the entire data set without any filters applied was the most common method of reviewing data. Seventeen groups used this as their primary means of perusing data, and all groups spent at least 5 minutes using this tactic. Nineteen groups spent 20 or more minutes reviewing data via this method (two groups spending over an hour). This method did not use annotations for search purposes. Some students browsed through the entire data set to see what their peers’ collected and found it enjoyable to see unrelated data, which could account for the high amount of time spent here.

While more students browsed through the entire data set without applying any annotations as search filters, a substantial portion used title searching to examine interesting data. Searching by tags was only popular with a small set of students, though the distribution of search strategies could be the result of the activity structure of finding similar animals: tag searches would be most useful to discover an animal related to their favorite animal, while title searches would locate an animal in mind.

Annotations Impacting How Students Evaluate Data
In order to see if any style of annotations helped students analyze data, we asked them, “How do you determine if the information you and your classmates collected is good?” Here, the tag annotations had a great effect. The majority, nine of the twelve groups responding to this question, indicated that they compared the tags to the associated images in order to evaluate their peers’ data.

   Student 1: It has good tags.
   Student 2: It has good tags and we kinda like look at the picture and see if it’s like if its right.
   Student 1: If it goes with the animal.
   Student 2: Yep.

After this, discussion between partners (4 groups) and using tags to find traits that were similar to their own animal (3 groups) were the most common approaches toward determining how good their classmates’ data was.

   Student 1: You should check the features of the animal the same traits that one animal has over the other...
   Student 2: You ah use the picture you click that you’re gonna use it ah look for like the traits that the picture has ah if it has the same trait as your animal then that those are the right traits.

To further look at the utility of tags (called labels to the students), and any effect they had on analyzing the data, we asked the students, “Did you find the labels describing individual animals helpful? Why?” All but one of the groups responding to this question reported the tags were helpful, primarily to see similarities in their animal (6 groups) or get more information than would be present in the image alone (5 groups). The group who didn’t find the labels helpful felt this way, “because, most of the stuff I would see by yourself so you don’t really need labels.”

   Student: If there were no labels I would think what’s the picture like what’s the like what’s all the things about it cuz I don’t know anything about if there’s no labels. Like turn that picture to the animal. Ok, then I didn’t know what it was, there’s no label on it so I don’t know what it is still, and I don’t know any facts about it or nothing.

This implies that the tagged data greatly assisted students in evaluating their evidence, with some students believing they could not evaluate the data without tags being present.

However, audio notes were not mentioned in how students determined good data and we wanted to see if the students used this annotation. In looking at the usage logs, we discovered that while all but one group listened to audio notes, there was a high variance in audio playback. During the trial we had noted many students seemed to enjoy listening to audio notes other students collected, often wanting to identify who made the note. Additionally, there were two pieces of off topic audio data that several students found entertaining (one being a photo of a duck titled “Aflack” with an audio note that said Aflack, mimicking a TV commercial). Several students repeatedly played these audio notes; one group listened to the Aflack audio 85 times. This had caused a disruption to students nearby at
one point and the group was instructed to stop playing the audio by the teacher.

The breakdown of how many times each pair listened to audio notes can be viewed in Figure 4, which also separates out the two off-topic audio notes.

When students were asked, “Did you find audio playback on this page useful? Why?,” several students found it useful, one student mentioned “because it explained details” and another group liked their own audio notes “because it kinda refreshed me why I kind of took it, the picture”. However, one group mentioned not finding them useful because the audio notes say the same thing as the tags, showing a mixed impression of the utility of this annotation. Future work needs to be done to look at why particular groups of students avoided the audio notes and further exploring the effect that listening to audio notes has on evaluating and utilizing data.

**How do Students Find and Choose the Evidence They Use in Their Explanation?**

To understand how students found the data they applied to their explanation and how they ultimately chose this particular evidence, we examined how they first discovered each unique piece of evidence. In order to understand what they found important about the pieces of evidence included in their explanation, we asked them (1) how they might choose data they would want to use and, later, (2) how they chose the data that they did use in their investigation.

**How Students Discovered Data**

When we examined all the unique pieces of data students chose to use in their explanation, we categorized it by which search or browsing method they used to find the data (breakdown of the methods by group in Figure 5).

Only a few groups had spent any substantial amount of time looking at and evaluating their own data, and their usage also shows a lack of interest towards their own data, with only five groups using more than 50% of their own data while nine groups never used any of their personally collected data. This implies that most groups preferred evaluating and utilizing their peers’ data, several mentioning that it was fun to look at what their peers collected. However, a few groups had heavy utilization of their personally collected data, showing it is valuable to support the filtering between personal and peers’ data.
Finding data by searching titles was the most common means of adding data to the explanations, which groups did for a total of 88 times. Though students spent more time browsing the unfiltered full data set, the groups only utilized 57 pieces of data from this search method, showing the reduced effectiveness of trying to browse the unfiltered data set. As with time spent searching, tags were not used to discover much data they used, with only 12 pieces selected in this fashion.

When we looked at the characteristics of how students searched for data when they then chose to use a piece of data, several trends emerged. Whenever students found a piece of data to use with a tag applied, the tag was always a trait their favorite animal had, which is an ideal search behavior for discovering related animals.

For title searching, every search that resulted in a student adding data to their explanation began by looking up the animal name except for five cases. Three separate groups made queries on a more generic animal type (animal, fish, birds) and two different groups only typed in part of the animal name to retrieve the result (Q to find Quetzalcoatlus and gir for Giraffe). The system evaluated the search term looking to see if it matched the prefix of any word in the title, which in this case helped the student avoid spelling out Quetzalcoatlus (and avoid misspelling it in the process).

**Student Perceptions on Choosing Data**

While comparing pictures to tags was the main mode of judgment for students in determining what “good” data was, the criterion for, “How do you pick out which data you want to use?” was different. Though comparing pictures to their tags was still a notable strategy, the more prevalent strategy was comparing the tags to the traits of their chosen animal. Seven pairs explicitly named this as their strategy.

We expected that if students were being reflective about their work, then these methods of comparing the labels to the image, or to their chosen animal, would be their main strategies. Note that while image quality and asking an expert (such as a teacher or researcher) received mention as ways to determine good data and pick out what data to use, reviewing audio was never mentioned as a potential tool for evaluating the data. This may be due to the lower quality of the audio notes or the students’ interpretation of the question, but it could indicate that audio notes are not seen as a useful means to interpret the data in this context.

An interesting phenomenon occurred when the students were asked, “How did you choose the specific pieces of data that you used in your explanation?” Several groups reported that picture quality became an important factor, even potentially surpassing the quality of the data.

**Student: Well, its really one thing like this salamander picture it’s actually a good picture… you can actually get better like you can get a bit more detail. Some other animals may be more closely related but still this gets back to the quality of the pictures**

This suggests that it would be helpful to have a method to swap the image that is associated with the annotations if a goal is to encourage using data with the best annotations.

In summary, students ended up finding the pieces of data they used for their explanations primarily from searching by titles, which was a more efficient use of time than browsing the unfiltered data set. The tags were primarily used to interpret the quality of the data, however a trend was noted that image quality was seemingly more important than quality of annotations for the data several groups included in their explanation.

**Overview of Student Performance in Constructing Scientific Explanations**

To provide context for the results presented in this paper about students’ ability to find and use data, we evaluated their final explanations. Overall, students were able to find, evaluate, and utilize peer-collected, annotated, multimedia data to construct coherent scientific explanations for the two investigations they were assigned during this study.

**Expectations**

Previous research has shown students typically can articulate a claim and use some supporting evidence, but greatly struggle with providing reasoning that connects evidence to their claim [3]. We were working with students who have never engaged in such science inquiry investigations in school, only had a day introduction to the claims—evidence—reasoning framework, and were working in a difficult context with large amounts of peer collected data. Nonetheless, we wanted performance with the Zydeco system to be at least comparable to what is found in previous literature, particularly for the first investigation on animal relatedness due to it being an easier task for students to make a claim.

A rubric was created for each investigation modeled off McNeill’s generalized claim—evidence—reasoning grading format [16]. We were defining success in this context as the majority of students completing the activity above a specific performance threshold. In investigation one the students were expected to create a complete claim, a piece of evidence for each ranked animal, and a reasoning that at least links the claim and evidence by repeating the evidence (Table 1). In the second investigation students were expected to have a claim answering the question, have at least one piece of appropriate evidence, and reasoning that at least links the claim and evidence by repeating the evidence. For each investigation, two reviewers independently graded 20% of the explanations and checked inter-rater reliability by percent agreements. Our inter-rater agreement was 100% for claim, above 94% for evidence, and above 83% for reasoning on each investigation.

**Performance**

In the first investigation, “Determine how three other animals are related to your animal,” the results exceeded our expectations with 18 (67%) of the pairs being
through the use of systems built around variety of create search through and evaluate their peers’ multimedia data to collected image data, along with an effective interface how students, particularly those new to science inquiry, educational activities. It presents one of the first and utilizing personal and peer.

CONCLUSION
This work contributes to research on how to support finding and utilizing personal and peer-collected data for educational activities. It presents one of the first studies on how students, particularly those new to science inquiry, think about and utilize large quantities of annotated peer-collected image data, along with an effective interface design that enabled students to explore the data and create an explanation on a tablet.

Students were able to use different aspects of Zydeco to search through and evaluate their peers’ multimedia data to create scientific explanations. This shows the potential for a variety of new rich activities for students, as they are able to engage in more authentic and personal inquiry activities through the use of systems built around their personal and peers’ annotated photographic data. While they had the option to use their own data, almost every group primarily searched their peers’ data set and used that data to make an explanation. The students’ performance as they managed the over 650 pieces of data produced during this study implies that inquiry and collection activities can run concurrently across multiple groups (as was the case in this trial), or have activities that run over a longer duration and generate similarly large quantities of data.

Some lessons learned from this study were:
- Tags are a valuable tool in aiding students to evaluate peer-collected data
- Title searching and panning through the unfiltered data set are the preferred methods for finding data
- Evidence used in explanations are rarely the students own but primarily from their peers
- While audio notes were generally on topic, students saw mixed results in their utility for interpreting data

Previous research found that students were varied in their search strategies on the web [4] and they were better at browsing than keyword searching [4, 11]. This study suggests that while browsing is still the preferred method of exploring data, students were more efficient at finding data to use when they searched via the title. However, students had varying usage and preferences towards the different annotations for different tasks; for developers of data-rich systems, this can indicate there seems to be no all-encompassing annotation or search type and such systems would benefit from a range of annotations to assist students in finding, evaluating, and using data. Some students chose data based on the image quality instead of the annotation quality. For designers seeking to encourage appropriate annotation use, it would be useful to provide a mechanism to swap the image associated with annotations, to enable students to get the high quality image they desire matched with the appropriate annotations.

This work has some limitations in that it was only investigating one particular activity context (collecting data from a history museum investigating animal relatedness and survival), and future work will need to investigate how the

<table>
<thead>
<tr>
<th>Claim</th>
<th>Evidence</th>
<th>Reasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: Nothing/ not a claim</td>
<td>0: No evidence/ evidence not related to claim</td>
<td>0: Does not provide reasoning/ reasoning does not connect the evidence to the claim/ does not make sense</td>
</tr>
<tr>
<td>1: Partial claim/ ranks fewer than three animals</td>
<td>1: Some ranked animals without evidence</td>
<td>1: Provides reasoning that links the claim and the evidence by repeating the evidence</td>
</tr>
<tr>
<td>2: Complete claim- includes three ranked animals</td>
<td>2: One evidence for each ranked animal</td>
<td>2: Provides reasoning that explains why the evidence (traits) supports the claim</td>
</tr>
<tr>
<td>3: Multiple evidence for some ranked animals</td>
<td>3: Provides reasoning that includes why the animals share traits through a scientific principle (i.e. shared ancestors)</td>
<td></td>
</tr>
<tr>
<td>4: Multiple evidence for all animals</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Grading rubric for the first investigation. Grey cells denote the minimum performance threshold for success.
use of annotations changes in varying contexts. Additionally, we will be adding in data probes to the system so students can gather numerical data and exploring how to support explanations using a combination of this data along with multimedia data. Future work will also include adding methods to incorporate information and articles from the Internet to further supplement the field data.

ACKNOWLEDGMENTS
We would like to thank Kira Berman and Amy Harris at the Exhibit Museum of Natural History for their support during our trials, the aid of team members on the Zydeco project, and the teachers and students who gave their time and advice to the project. This material is based on work supported by the National Science Foundation under Grant No. DRL-1020027. Any opinions and findings expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation.

REFERENCES