Annotation as Process: A Vital Information Seeking Activity in Historical Geographic Research

Mary B. Ruvane

School of Information and Library Science, University of North Carolina at Chapel Hill, CB# 3360, 100 Manning Hall, Chapel Hill, NC 27599-3360.

ruvane@email.unc.edu

This paper describes the role annotations played in evolving and growing the value of a research database in the framework of an historical geographer's information seeking process. The database was designed as a "note-taking" tool for gathering historical evidence from primary source documents. The individual facts collected at first provided little utility to the researcher, indicative of Buckland's (1991) information-as-thing. As the process of collecting data began to amass a large body of material the geographer's information needs grew as new connections were made between the accumulating annotations.

Introduction

Annotations played a key role in the information-seeking process of an historical geographer, as observed during a nearly two-year collaboration. From the standpoint of an historical geographer's information requirements few studies appear to address this topic specifically -- how do they become informed? Several resources touch upon issues related to historians' information behavior; describing the characteristics of source material (Brundage, 2002), research motives and methods (Case, 1991), historians' use of computers (Lewis & Lloyd-Jones, 1996), and the relationship between geography and history (Baker, 2003; Ogborn, 1999).

For a glimpse into the specific interests of historical geographers Sauer (1940) aptly

summarized the nature of this discipline in his address to the Association of American Geographers. Baker (2003) offers alternative understandings depending on the perspective taken and circumstance. Other papers impart the challenges faced when researching topics in historical geography, specifically when the findings being collected are intended for presentation in a digital format within a geographic information system (GIS) (Boonstra, Breure, & Doorn, 2004; Gregory, Kemp, & Mostern, 2003), as is the case in this study.

There have been numerous books and articles written describing the characteristics of information, information behavior, information seeking, and information-use (Buckland, 1991; Case, 2002, Jarvelin & Wilson, 2003; Marchionini, 1995; Wilson, 1999). One aspect of this research deals with the issue of information seeking in context (Dervin, 1997; Solomon, 2002) defined as the observation of an actual information-need and the processes undertaken by the 'seeker' to satisfy their need. In the context of this study, annotations were a crucial aid in the process of information discovery for the historical geographer.

In this paper I first relate how the data in an historical database, used for gathering historical evidence, corresponds to the range of dimensions of annotation outlined by Marshall (1998). These annotations played a key role in the process of becoming informed, which corresponds nicely with Buckland's (1991) analysis of information-as-process (or information-as-evidence) and MacMullen's (2005) paper correlating annotation-as-thing. I follow this with a synopsis of the ongoing collaboration between Rebecca Dobbs and me, an historical geographer and information scientist respectively, as we continued to incorporate changing information needs into a database to maintain her research objectives.

Dimensions of Annotation

Marshall (1998) contends "...annotation is a key way in which hypertexts grow and increase in value." It also appears annotations play an essential role in improving and adding value to an evolving research database used for collecting historical evidence. Annotations can take on many forms and serve various purposes. The following

provides a summary of Marshall's dimensions accompanied by a description of analogous characteristics found within an historical database.

Dimension 1: Formal vs. informal annotations

Metadata is given as an example at the extreme end of formal annotation, a structured method that employs descriptors adhering to strict standards in terminology for describing a document or data. The objective of formal (or structured) notation is to insure interoperability and optimize query performance. On the opposite end would be informal annotations, such as notes made to oneself in the margins of a text or document during the course of reading.

In an historical database, fields with structured controls are representative of Marshall's formal dimension. For example fields programmatically restricted to a particular format of data, such as a survey's north-south/east-west directional coordinates and related numerical boundary lengths, or limited to a list of predefined choices, the equivalent of keywords one might underline in a text. On the other end of this spectrum would be informal notes found in a comment or note field, designed for capturing unstructured free text to satisfy the readers need or desire to add personal or explanatory comments.

Dimension 2: Explicit vs. tacit annotations

On one end of this dimension annotations may be easy-to-read, intended for others to understand and interpret. On the opposite end they may solely be for personal use, such as notes written in a cryptic style decipherable only by the author. The former represents an explicit annotation, structured for sharing; the latter would be tacit, useful only to the note taker.

In an historical database, explicit fields would be those designated for capturing categorized content such as a type of date, a person's name, a particular event, or a name and class type of a feature. These fields typically share qualities with dimension

one, on the metadata end of the scale. Tacit notes are more often found in comment or note fields, frequently used as the catchall for the reader's reaction to a document's content and intended for personal clarification.

Dimension 3: Annotation as reading vs. annotation as writing

Annotations may simply be a fleeting act during the process of reading, such as unconsciously highlighting or doodling as one progresses through the text, to be forgotten and lost once a passage has been examined. Or they could be construed as a form of writing that purposely adds value to an existing text or document.

In an historical database a variety of field formats may take on the dimension of annotation as writing. For instance, unstructured note and memo fields may contain translations of archaic prose, point to related sources of interest, or include author observations. Controlled fields can be manipulated (say by concatenation) to organize and present previously disjointed concepts more clearly.

The opposite dimension, annotation as reading, appears less conducive in the context of reacting to primary source documents when employing a database for note taking - spontaneous interaction is unwieldy, requiring the reader to disengage from the act of reading if the urge to doodle strikes them. Although a few examples might be interpreted as electronic doodling, particularly one employing an emoticon, in most cases the author confirmed these were conscious acts of reading, intended to remind her of documents deemed visually interesting or unique.

Dimension 4: Hyperextensive vs. extensive vs. intensive annotation

Marshall (1998) makes use of Levy's (1997) distinctions between the terms hyperextensive, extensive, and intensive. Hyperextensive annotations could be likened to the links in a web page leading you to successively more fragments in the chain, either related or tangential. Extensive annotations are associations recorded for

analytical use in the comparison of many documents. Intensive annotation would be a meticulous response to a single text.

Extensive annotations appear to be the predominant dimension found in an historical geographer's database because the objective is to correlate evidence found across numerous documents. For example, identifying features common to numerous parcels. Yet concurrently individual records in the database could be construed as intensive, since each focuses on details contained in a particular document. The numerous references to related material, both within and external to the Dobbs' database, may be representative of the hyperextensive dimension. For example, there are frequent pointers to an external historical gazetteer, an often referred to geographical dictionary of place names used for translating past vernacular terms into present day equivalents.

Dimension 5: Transient vs. Permanent annotations

Annotations might simply be an ephemeral interaction with a text or document during the act of reading, only useful at the time of engagement and of little value on a return visit. Yet in some instances what remains may unintentionally be informative to subsequent readers, including the original annotator.

In an historical database note and comment fields often provide a vehicle for capturing a reader's peculiar observations, initially transient in nature. What at first may appear to be a singular worthless remark, over time could prove valuable both for the annotator and later readers. As notes continue to accumulate in the database, and external resources are consulted, seemingly transient annotations may eventually shift into the realm of the permanent dimension.

Dimension 6: Published vs. private annotations

One's personal reading material may be heavily annotated with thoughts motivated during the course of reading. These notes are typically not intended for sharing,

although once a colleague is given access to this copy the annotations shift from the private to published dimension. There are also annotations intentionally written for publication, such as scholarly works aimed at deciphering the meaning of previously published material.

In an historical database private annotations are often incorporated within the comment or note fields, yet they unintentionally move to the 'published' dimension once the database is shared with a collaborator or other interested party. In Dobbs' database there are many fields intended from the onset for publication, such as those purposely formatted using strict controls for use in later resource discovery (e.g., metadata), comparative analysis, and the full text transcriptions of each document examined.

Dimension 7: Global vs. institutional vs. workgroup vs. personal annotations

The range of circumstances in which annotations are employed can be various. From the global application of hypertext links leading to related web pages, to those found in communications between shared communities (such as on an institutional or small workgroup level), or simply created for personal use. Each situation may incorporate one or a mixture of the annotation dimensions described.

In this paper the particular historical database being studied was foremost a tool for the researcher's personal annotations. After the data is analyzed a selection of annotations will be made available on a global scale in a variety of formats; a relational database, as metadata linked to GIS compatible features, and in a publishable format (both online and in print). The annotations generated from this research have the potential of being used in a range of situations representative of this dimension.

Annotation = Information (as-thing, as- process, as-knowledge)

Many have grappled with the definition of information offering numerous characterizations. Webster's dictionary (Friend & Guralnik, 1956) suggests that

"...information applies to facts that are gathered in any way, as by reading, observation, hearsay, etc. and does not necessarily connote validity (inaccurate information);...". It further defines knowledge as "...[that which] applies to any body of facts gathered by study, observation, etc. and to the ideas inferred from these facts, and connotes an understanding of what is known (man's knowledge of the universe);..." Both of these fail to impart the subtle transition, or dimensions, that exists between the terms fact (or data) and knowledge.

Here Buckland's (1991) definition of information helps to distinguish this imperceptible scale (e.g., "information-as-thing", "information-as-process," and "information-as-knowledge"), which seems to share facets of Marshall's seven dimensions of annotation. Buckland makes a distinction between three levels of information: the embryonic stage, or basic building block, of "information-as-thing" represented in its simplest form as datum of interest, the middle stage of "information-as-process" where the act of collecting evidence aids in the process of becoming informed, and the final stage of "information-as-knowledge" -- perhaps the culmination of the information seeking process where the 'seeker's' need has been satisfied and understanding achieved (see: Figure 1).

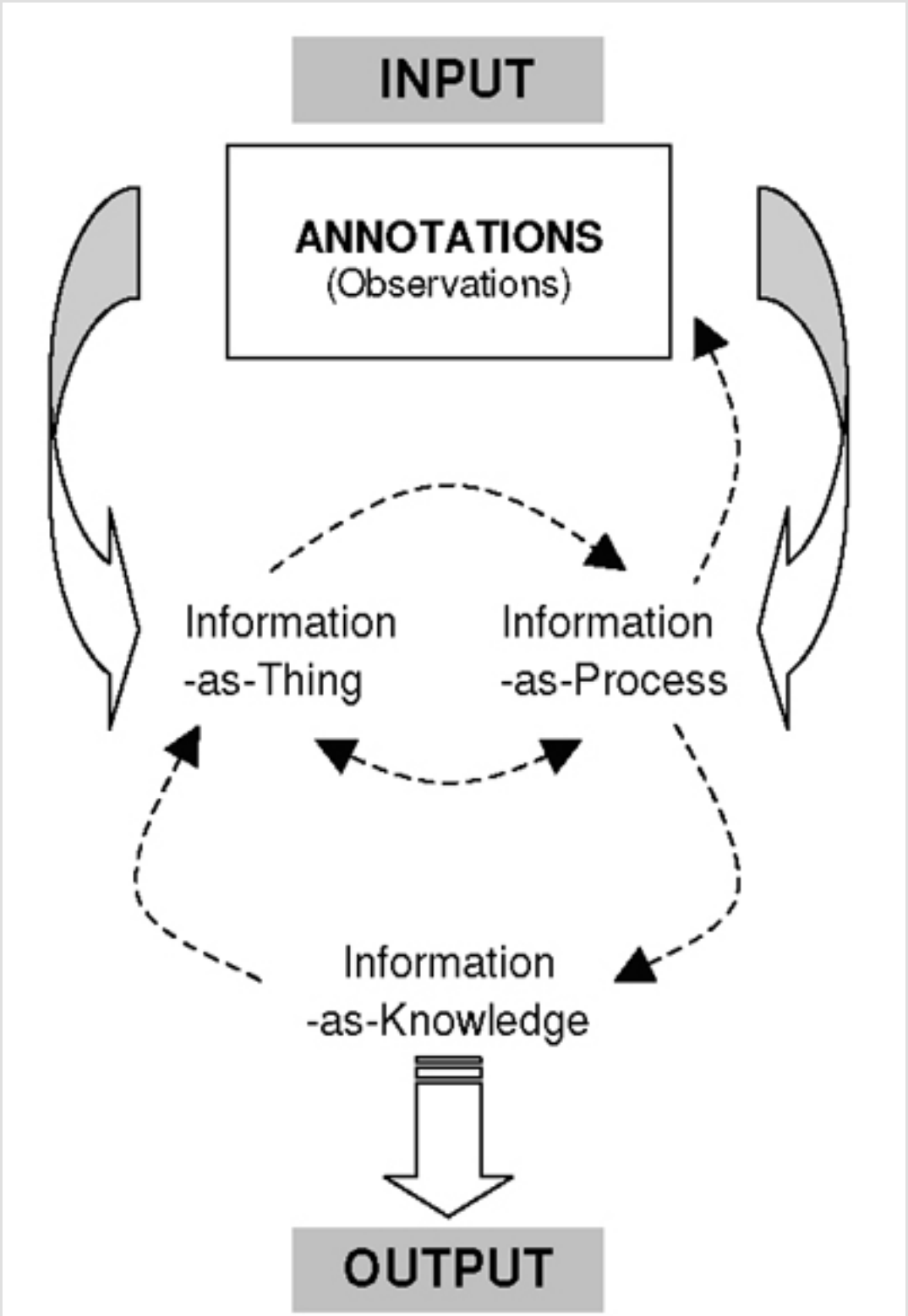


Figure 1. Annotation = Information

As one begins to read a new document or text the range of information initially

collected could represent any one, or a combination, of the seven dimensions of annotation depending on the reader's level of preexisting knowledge and intended objective. Annotations may start out as a vague response to the process of reading or observing, falling under the umbrella of "information-as-thing"; the stage lacking an informative aspect where it may remain or in time shift into the next level of "information-as-process."

As reading continues and annotations (or facts) accumulate the information seeker becomes more informed, transitioning into the next degree of "information-as-process." At this stage the types of data or annotations being collected (both past and future) may need to be adjusted, expanded, or added to in order to incorporate additional dimensions more appropriate for analyzing the evolving inferences.

Historical Geography: Information Seeking in Context

In this section I provide an overview of Dobbs' information-seeking process and share some observations regarding our nearing two-year collaboration (Ruvane & Dobbs, 2005). My role was to develop a flexible database for her use as a "note-taking" tool to collect and organize historical evidence in support of her dissertation research.

In this role I was not only an observer of the information-seeking process but an active participant in the direction and evolution of the added value she sought.

Research Setting: Context of Information-Need

The objective of Dobbs' research was to demonstrate the influence a prominent transportation route, the Indian Trading Path, had on settlement patterns in the mid 18th century and the consequent emergence of today's urban centers. Part of this research would entail identifying the land occupied in the central piedmont region of North Carolina between 1748 and 1763, followed by creating a digital multi-media map to illustrate the findings.

Land in this region, during this time period, was under the control of two separate administrations; Lord Granville held the rights to the northern half of the state, the Colony of North Carolina the southern portion. The materials she deemed most useful to her study were original land grant documents. These hand written papers recorded the four stages (e.g., entry, warrant, survey, deed) of a process that culminated in the issuance of a deed, a grant authorizing a settler the right to occupy a tract of land in return for an annual quit-rent.

Land grant documents contain various levels of detail (or lack thereof) ranging from broad general descriptions of a parcel, such as the estimated acreage and county it resides in, to very detailed descriptions including surveyed plats accompanied by narrative describing the land's relation to geographic features; such as rivers, transportation routes, cultural features and bounding neighbors. Despite the abundance of evidence in these historic records the handwriting styles, use of vernacular names, and cryptic short hand leaves a great deal of uncertainty when deciphering.

Annotations: Use & Motivation

Dobbs wanted to compile notes that would serve as an aid for building a spatially accurate map of the land parcels occupied. To illustrate her findings she would be using a Geographic Information System (GIS). The evidence essential to her research would be survey measurements, key for recreating the shape of each parcel. Also important would be any clues that could help in deducing the physical relationship between individual properties (e.g., the position of adjoining tracts).

Historical Geography: Database Annotation Uitlity

Conceptualizing & Implementing

From the onset Dobbs determined a database would be ideal for recording and organizing her research notes, since writing in the margins of primary source material was not an option and making copies of each document to do so would be prohibitive.

After reading through a sample of survey documents (e.g., recordings of the third stage in the land grant process) she identified several categories of information to collect in the database. These included the parcel survey measurements, descriptive narratives, people names, geographic features (e.g., rivers, roads, cultural), administrative characteristics (e.g., county, grantor), and related dates.

Using a relational database application (e.g., MS Access) Dobbs created a simple table (e.g., flat file) and data entry form for recording her observations. In its original format this "annotation tool" proved extremely efficient for taking notes while she read through the survey documents.

Method & Use

Dobbs recorded her research notes in the database concurrently as she viewed microfilm representations of the land grant documents or related material. The primary reason initially for using a database was to capture the survey measurements. These would be exported to another application for building GIS compatible "shape files" for illustrating the size and outline of each parcel (NWF/DEM, 2001). Once the shapes were generated Dobbs' original intent was to print 3 x 5 cards containing the annotative content recorded describing each parcel. This would allow her to sort the cards manually into groups, based on location clues, to assist with positioning each parcel in real time and space on her digital map. Groupings could be sorted into parcels sharing common neighbors, or tracts linked by connecting geographic features, such as rivers or paths, or into clusters of properties within proximity of a familiar cultural amenity or other type of bounding evidence.

What had not been anticipated was how the growing body of data being collected would reshape Dobbs' information-needs. While it was apparent the evidence she was collecting was relevant to her needs, the original database model and field formats were hindering her ability to easily restructure her observations to improve query analysis. Many of her fields had been informally structured for collecting a mixture of personal commentary, reflecting a range of Marshall's dimensions. As the process

evolved patterns began to emerge indicating that critical keywords had been interspersed within unstructured annotation fields. These newly discovered categories of evidence needed to be shifted into a more structured dimension to be of value.

The technical complexity of her dilemma was beyond what Dobbs was willing to take on. It was at this stage I offered to design a new database (e.g., relational) capable of capturing, comparing, and linking the mounting number of unforeseen and overlapping clues buried within her unstructured annotation fields.

Collaboration: Informative value of annotations

Dobbs' original database consisted of 43 fields. Over half, 24, were formal annotations (e.g., metadata) representing parcel survey measurements. In reality only three fields were needed for capturing this information (e.g., sequence, measurement angle, and length), the rest were redundant duplicate columns containing the same category of information; an inefficient method inherently found in a flat file. This type of redundancy, using multiple fields for one entity type, or commingling unique data within one field, was evident across several fields employed for recording evidence in the original database.

The non-survey measurement information Dobbs was collecting primarily represented informal annotations that lacked structure and embodied a variety of Marshall's dimensions. For example, people names were entered into memo fields intermixed with related commentary or clarifications. The same situation existed for geographic feature names , which were recorded along with directional clues such as '...on the north east side of the Yadkin River, above John Smith's property....'. In some cases annotations were cryptic while others were clear.

By sharing her database with me I was able to develop a better understanding of Dobbs' shifting information needs. The informal annotations she had been collecting provided the most insight, illustrating the kinds of observations she had made and how they had changed over time. After I read through a sample of these we met several times to discuss the value and meaning of various notes and what had prompted their

evolution. Our discussions helped to shape the blueprint I used for building a new and more flexible database.

To improve Dobbs' annotations, from an interoperability standpoint, I first focused on analyzing the contents to identify recurring themes appropriate for transitioning into a more formal dimension (e.g., suitable for controlled fields, categorization, new tables, etc.). For example people names had not originally been perceived as a critical piece of evidence for analyzing, so they had been recorded informally and commingled with related commentary as a future memory aid. As Dobbs continued to collect data she began to notice that the roles people played in the transaction of a particular tract of land were an invaluable clue for locating parcels in relation to each other (e.g., adjacent neighbors, near neighbors, likely neighbors, etc.). It was now important to repurpose these annotations to facilitate relationship queries.

In the new database people names were moved to more formal fields and linked to an expandable "type" category to identify the role each played in a land transaction (e.g., grantee, assignee, neighbor, chain carrier, etc.). A separate memo field was provided for Dobbs to add informal clarification notes to any given name. The same approach was taken with feature names, dates, and parcel descriptions to improve interoperability, provide flexibility for adding new "types", and offer space for assorted author observations. The first rendition of the new database consisted of over 25 formal annotation fields, of which 3 were for survey measurements, and approximately six informal fields.

Although much of the original annotations' content were shifted into categorized fields it did not eliminate the need to continue providing space for recording free text observations. Indeed, the number of informal comment, note, and semi-controlled fields employed grew with each successive modification of the database (e.g., currently v10.3). These were typically linked to a particular formal field, such as a person or feature name. From the beginning, and throughout our collaboration, the informal fields played a critical role in guiding the direction of each database revision. They continually pointed to new patterns worth formalizing into categories within the database, ever improving the value of the researcher's "note taking" tool, both for her and future

readers.

Conclusions

Marshall's introduction to her paper seems to mirror my observations of Dobbs' information seeking process. By taking a little license with Marshall's words I have repurposed them to explain the similar value of annotations to an historical geographer:

> **As observed, a database for collecting historical evidence will grow and change by way of addition - for example Dobbs responded to her initial database by adding commentary, making new connections (discovering unexpected relationships) and creating new pathways (fields, tables, and links) in the process of gathering and interpreting the material she was reading. Her activity encouraged the expansion of both the database's structure and content. In so doing, she added crucial value to an existing body of interrelated historical material.**

The value of incorporating various dimensions of annotation into a database (or similar electronic tool) designed for collecting historical evidence seems crucial, especially placeholders conducive to recording informal commentary. The historical researcher's process of becoming informed is nonlinear; uncertainty persists throughout the information-seeking activity, especially when one is examining imprecise primary source documents that are inconsistent in both quality and clarity.

For any "note taking" tool to be of value to an historical geographer it should provide built in flexibility and ample space for informal observations. Too much structure, without a method to impart outstanding uncertainty, could render the collective findings unfit for use in future research. Alternatively, as informal annotations grow in volume they may need to be divided into more formal dimensions (e.g., metadata) as new patterns emerge suggesting improvements to interoperability.

Future Research

This study forms the basis for future research into the types of database models best suited for recording imprecise and time dependent historical evidence; both from an individual researcher's needs and those required in a distributed community environment. There are a growing number of collaborative initiatives looking at methods and standards for incorporating the "accurate" display of geographic content, as described in historical resources, with the use of digital mapping tools (e.g., GIS).

Acknowledgements

This work was partially funded by an unrestricted research gift from Microsoft Research to the Annotation of Structured Data research team in the School of Information and Library Science at the University of North Carolina at Chapel Hill, whose members contributed to this work: Gary Marchionini, Paul Solomon, and Catherine Blake, co-PIs; with team members Tom Ciszek, Xin Fu, Lili Luo, W. John MacMullen, Cathy Marshall, David West, and Megan Winget. The project's website is available at: http://ils.unc.edu/annotation

References

Boonstra, O., Breure, L., & Doorn, P. (2004). *Past, Present and Future of Historical Information Science* (Report/Study). Amsterdam, Netherlands: Netherlands Institute for Scientific Information.

Baker, A.R.H. (2003). Geography and History: Bridging the Divide. *Cambridge studies in historical geography, 36*. Cambridge, UK: Cambridge University Press.

Brundage, A. (2002). Chapter Two: The Nature and Variety of Historical Sources. In *Going to the Sources : A Guide to Historical Research and Writing* (3rd ed.). Wheeling, Ill.: Harlan Davidison.

Buckland, M. K. (1991). Information as Thing. *Journal of the American Society for Information Science, 42*(5), 351-360.

Case, D. O. (1991). The Collection and Uses of Information by Some American Historians: A Study of Motives and Methods. *Library Quarterly, 61*(1), 61-82.

Case, D. O. (2002). *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. New York: Academic Press.

Dervin, B. (1997). *Given a Context by any Other Name: Methodological Tools for Taming the Unruly Beast*. Paper presented at the Information Seeking in Context: proceedings of an International Conference on Research in Information Needs, Seeking and Use in Different Contexts, 14-16 August, 1996, Tampere, Finland.

Friend, J. H., & Guralnik, D. B. (Eds.). (1956). *Webster's New World Dictionary* (College ed.). New York, NY: The World Publishing Company.

Gregory I.N., Kemp, K. and Mostern R. (2003). Geographical Information and Historical Research: Current Progress and Future Directions. *History and Computing, 13*, pp. 7-22.

Järvelin, K., & Wilson, T. D. (2003). On Conceptual Models for Information Seeking and Retrieval Research. *Information Research, 9*(1) paper 163.

Levy, D. M. (1997). I Read the News Today Oh Boy: Reading and Attention in the Digital Library. In *Proceedings of the second ACM international conference on Digital libraries*, pp. 202-211.

Lewis, M. J., & Lloyd-Jones, R. (1996). *Using Computers in History: A Practical Guide*. London; New York: Routledge.

MacMullen, W. J. (2005). *Annotation as Process, Thing, and Knowledge: Multi-domain studies of structured data annotation*. SILS Technical Report TR-2005-02. Chapel Hill: University of North Carolina, School of Information and Library Science, Technical Report Series.

Marchionini, G. (1995). Information-Seeking Perspective and Framework. In *Information-Seeking in Electronic Environments* (Chapter 3, pp. 27-60). New York, NY: Cambridge University Press.

Marshall, C. C. (1998). *Toward an Ecology of Hypertext Annotation*. Paper presented

at the HyperText 98 Conference, Pittsburgh, PA.

NWF/DEM Data Editing Extension for ArcView. (Sept. 2001). Developed through a cooperative project between the Norcross Wildlife Foundation and Massachusetts Department of Environmental Management, v1.2. Available at: http://www.mass.gov/dem/programs/gis/de/help/help_frame.htm

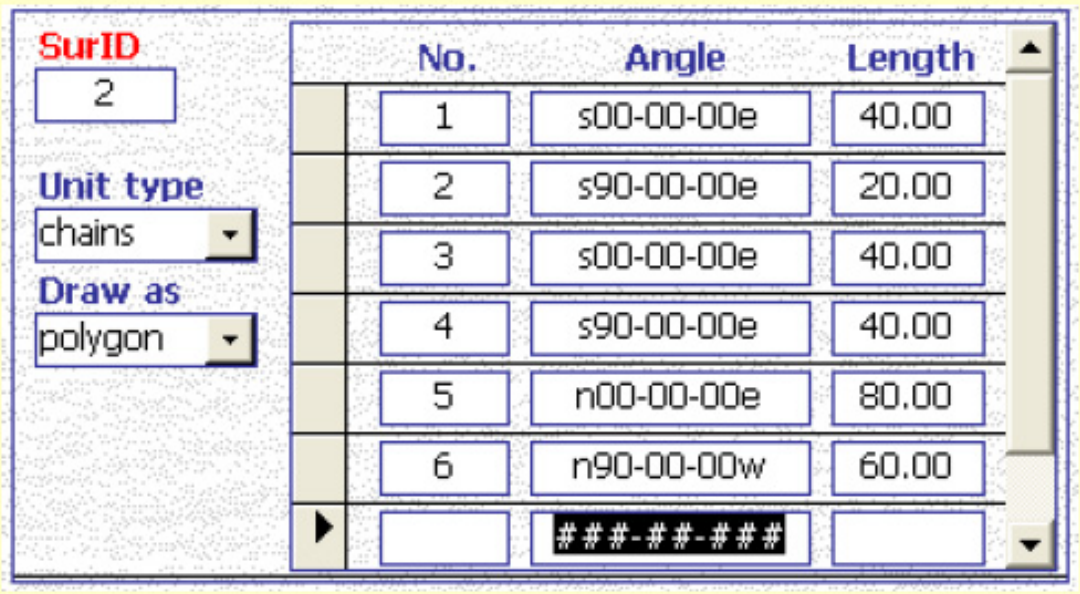
Ogborn, M. (1999). The Relations Between Geography and History: Work in Historical Geography in 1997. *Progress in Human Geography, 23*(1).

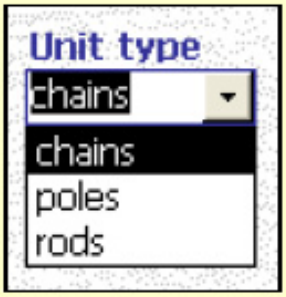
Ruvane, M. B. & Dobbs, G. R. (2005). *Interdisciplinary Collaboration and Database Modeling for Historical GIS: Structured Annotation for Land Grant Research*. Slide show presented at the Association of American Geographers, 101st Annual Meeting, Denver, CO (April 5-9, 2005).

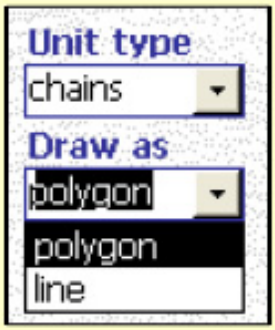
Sauer, C. O. (1940, Dec.). *Foreword to Historical Geography*. Paper presented at the Association of American Geographers, Baton Rouge, LA. [Online text converted copy].

Solomon, P. (2002). Discovering Information in Context. In B. Cronin (Ed.), *Annual Review of Information Science and Technology* (Vol. 36, pp. 229-264).

Wilson, T. D. (1999). Models in Information Behaviour Research. *Journal of Documentation, 55*(3), 249-2.



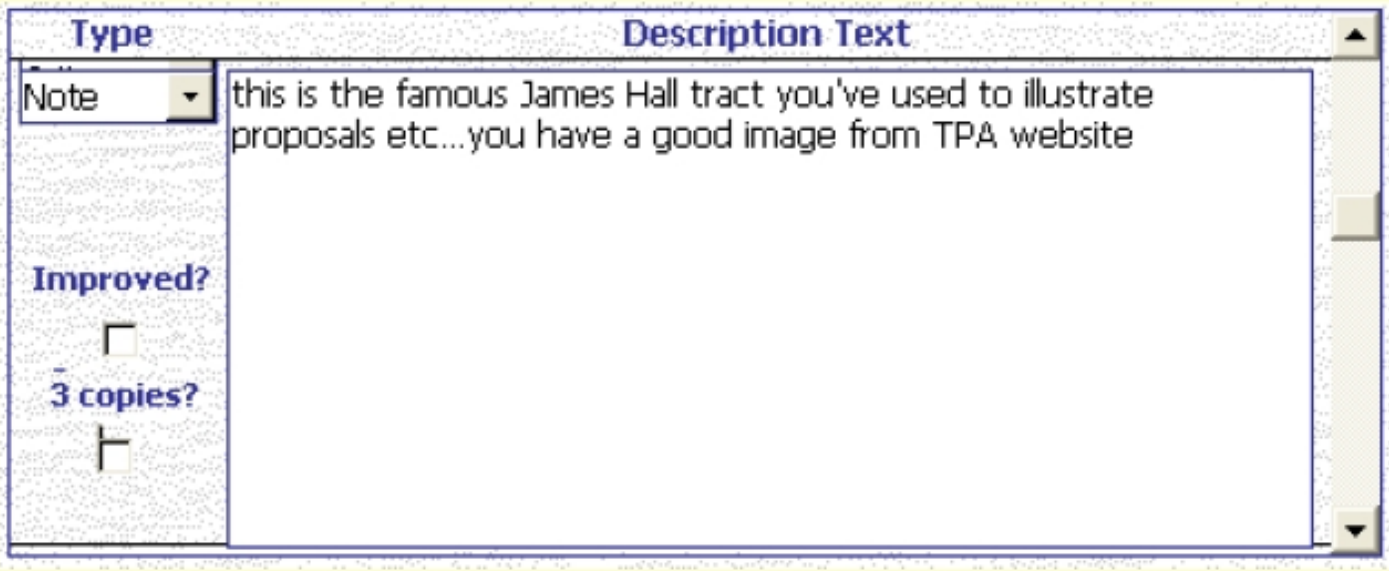




a. Survey dimensions restricted by : auto numbers for sequentially pairing angles and lengths, controlled data formats for north-south/east-west directional coordinates, and numerical lengths.

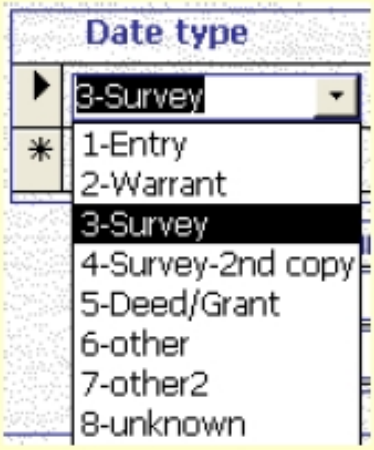
b. Restricted by a predefined selection; unit type, draw as

FIG. 1. Formal annotations – restricted to a particular format of data

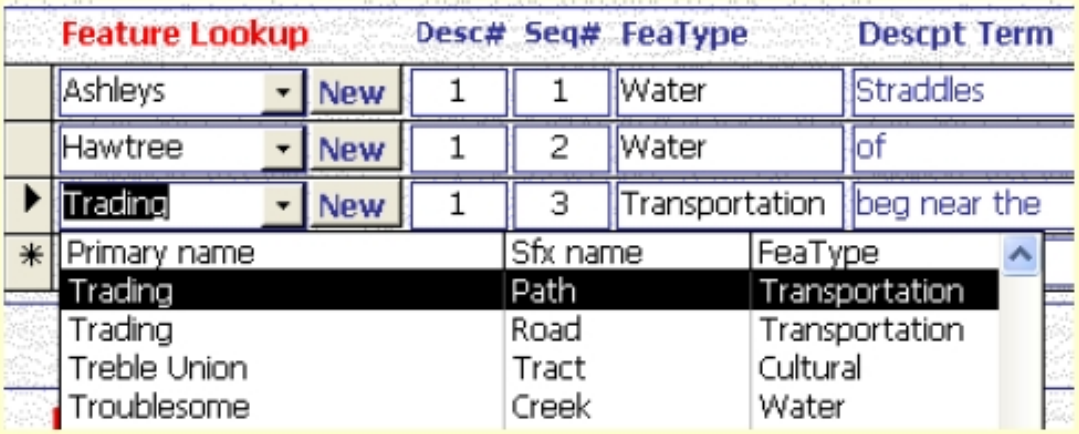


Informal annotations may be recorded in a comment field. In this example "Type: Note" was designed for the author's reflections on a particular parcel. Here unstructured free text can be recorded to satisfy the readers need or desire to add personal or explanatory comments.

FIG. 2. Informal annotations – unstructured free text

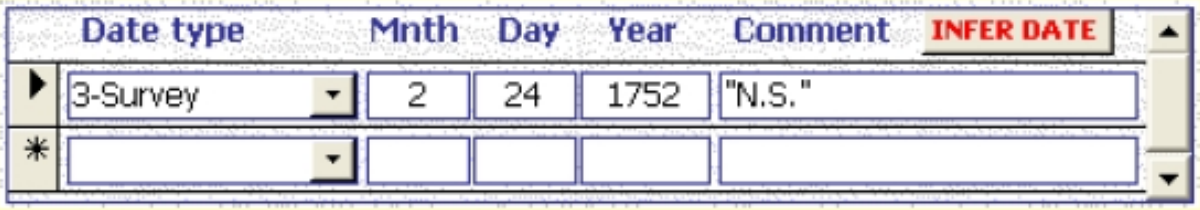


a. Type of document linked to filing date



b. Selectable names and class types for recording features mentioned in each land grant document

FIG. 3. Explicit annotations – intended for others to read



The date of this parcel's physical survey is clear, but only the author knows what "N.S." means.

FIG. 4. Tacit annotations – useful only to the note taker

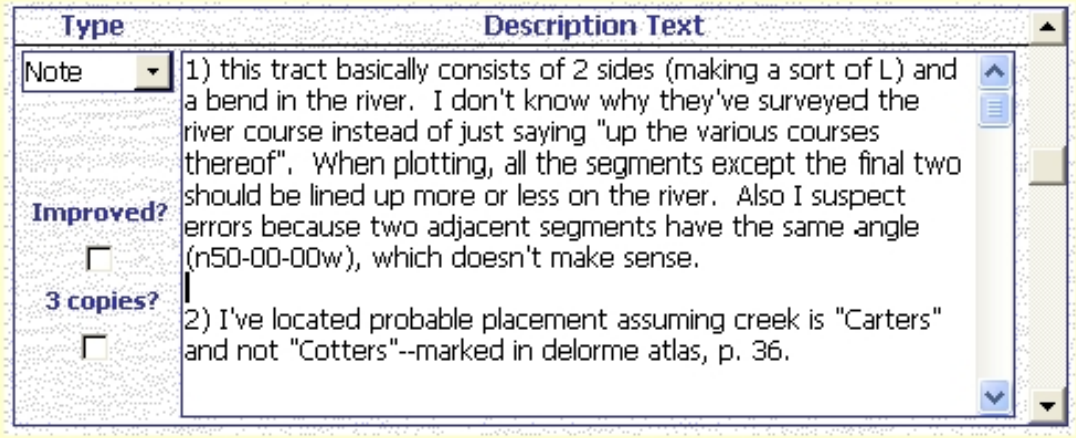


FIG. 5. Annotation as writing –
purposely adds value to an existing document

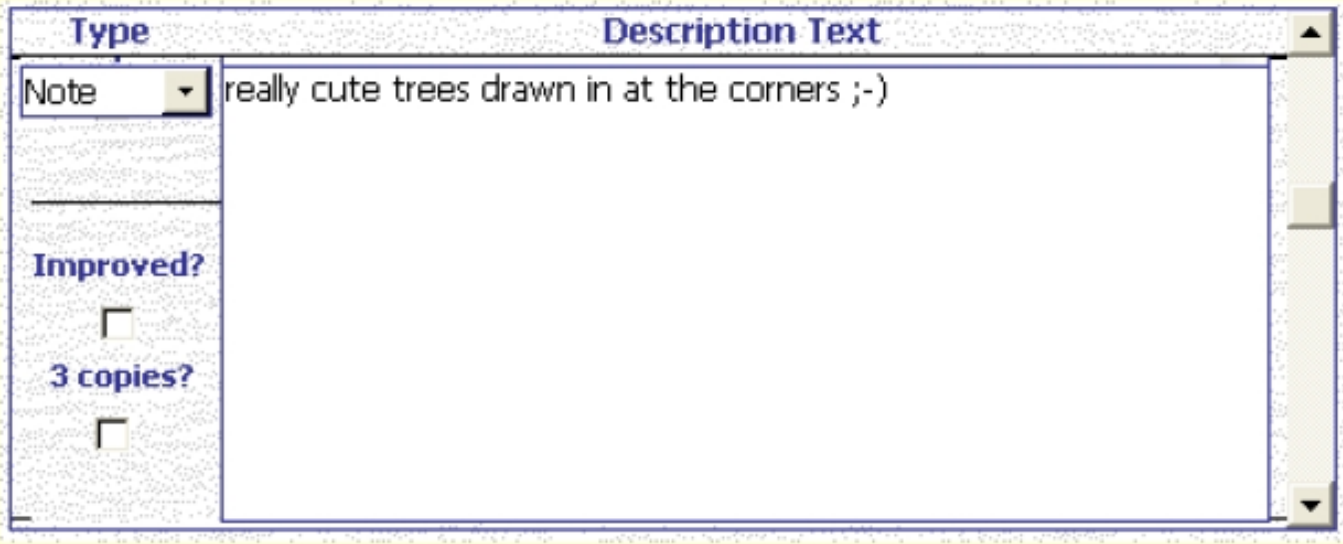


FIG. 6. Annotation as reading –
unconscious highlighting or doodling during act of reading

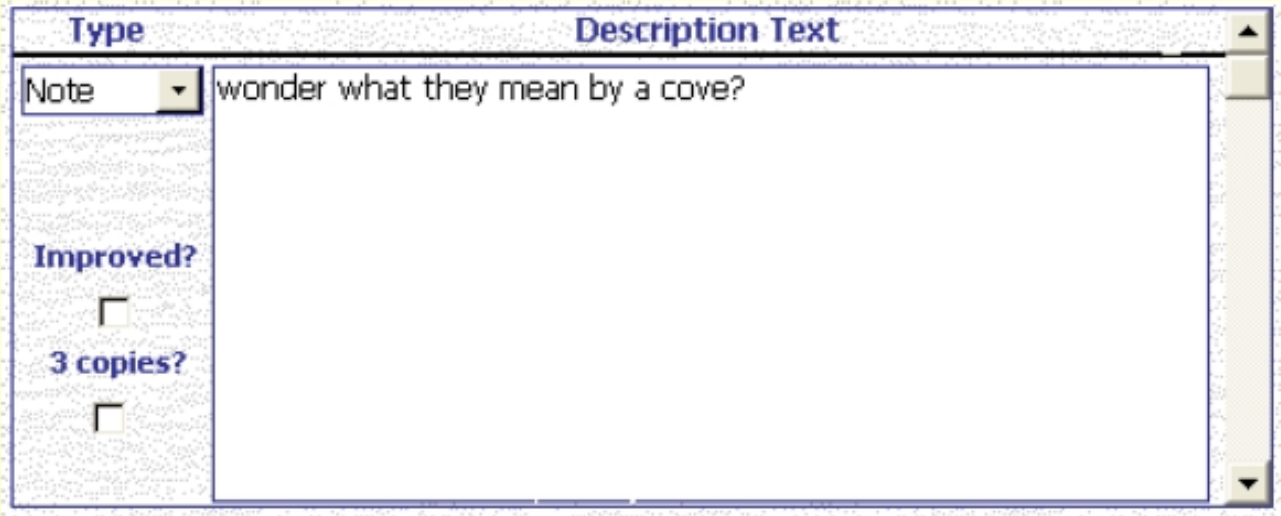
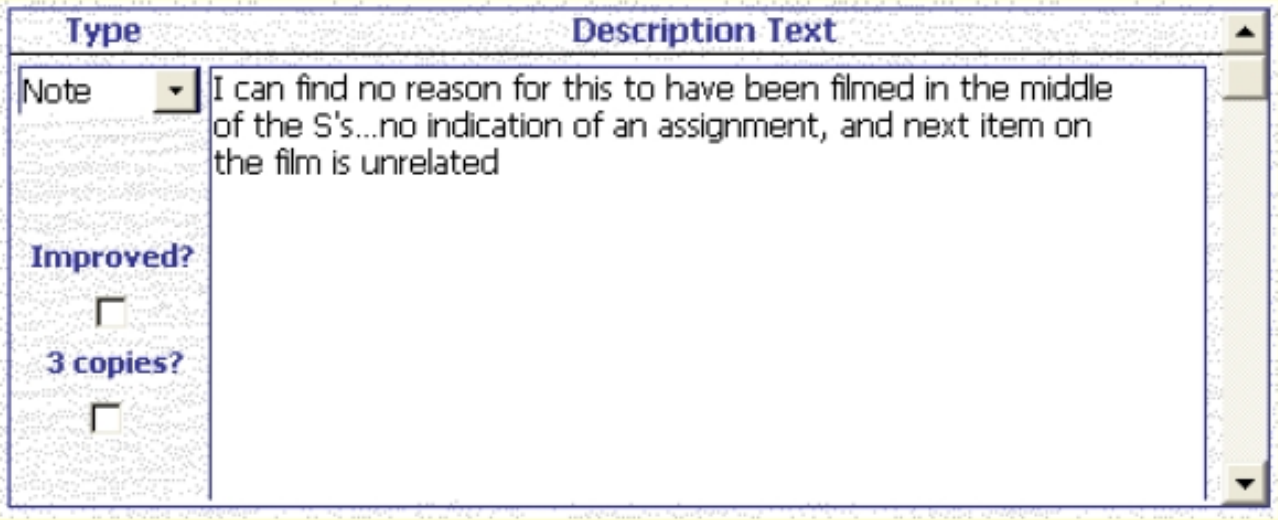


FIG. 7. Transient annotations –
ephemeral interaction with a text; of little value on a return visit



This comment adds little value to Dobbs' research objective. But for subsequent researchers being aware that documents may have been misfiled by the original colonial administration, or perhaps the archival institution itself, can be very informative. (Doc ID: 3523)

FIG. 8. Permanent annotations –
transient notes that unintentionally become informative

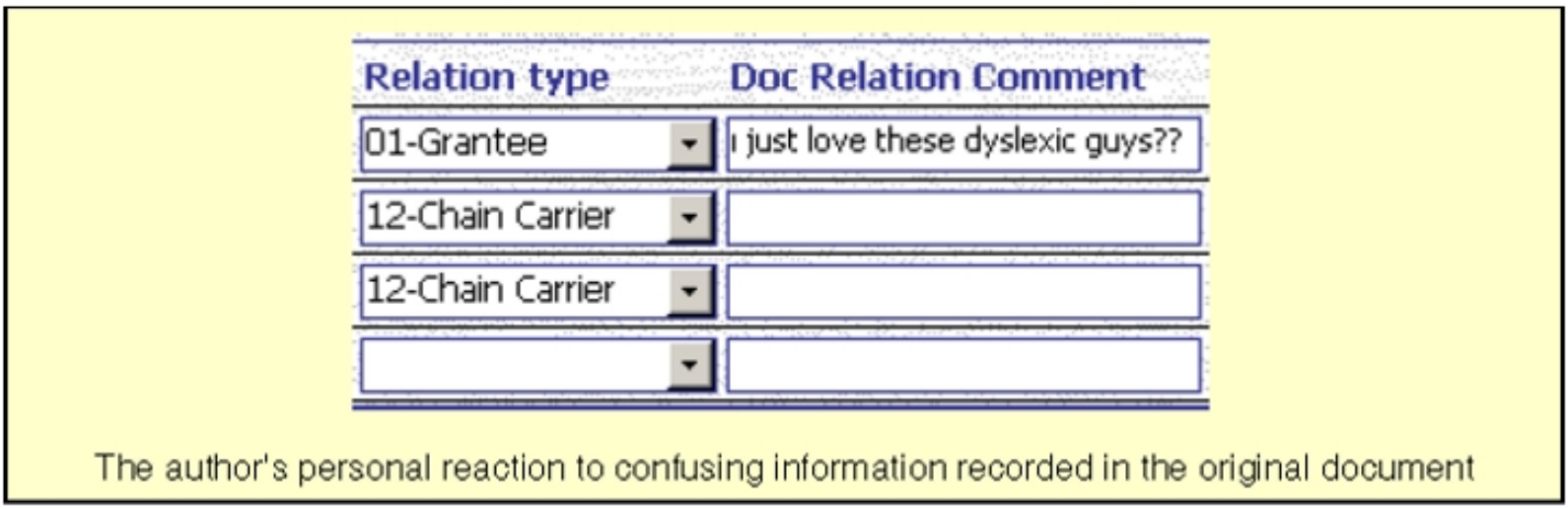
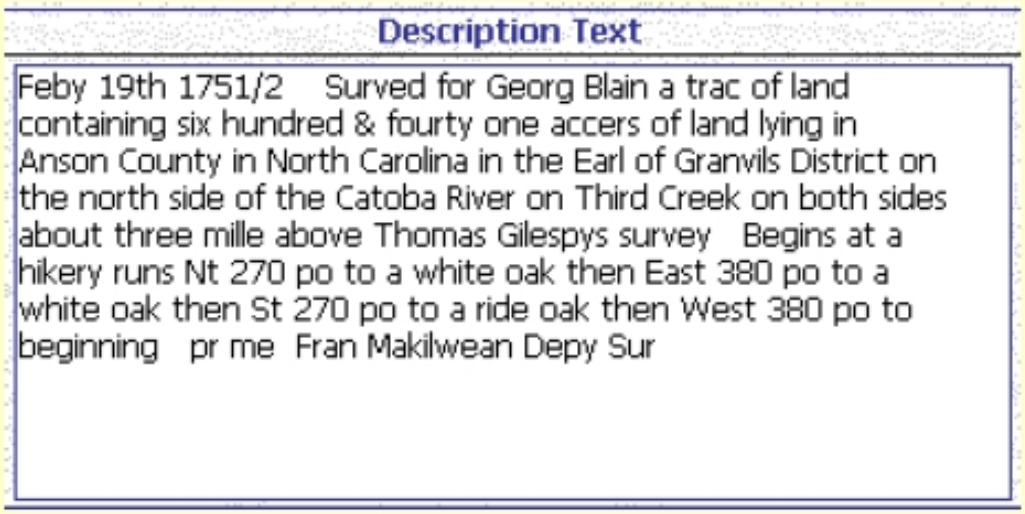


FIG. 9. Private annotations



Description Text

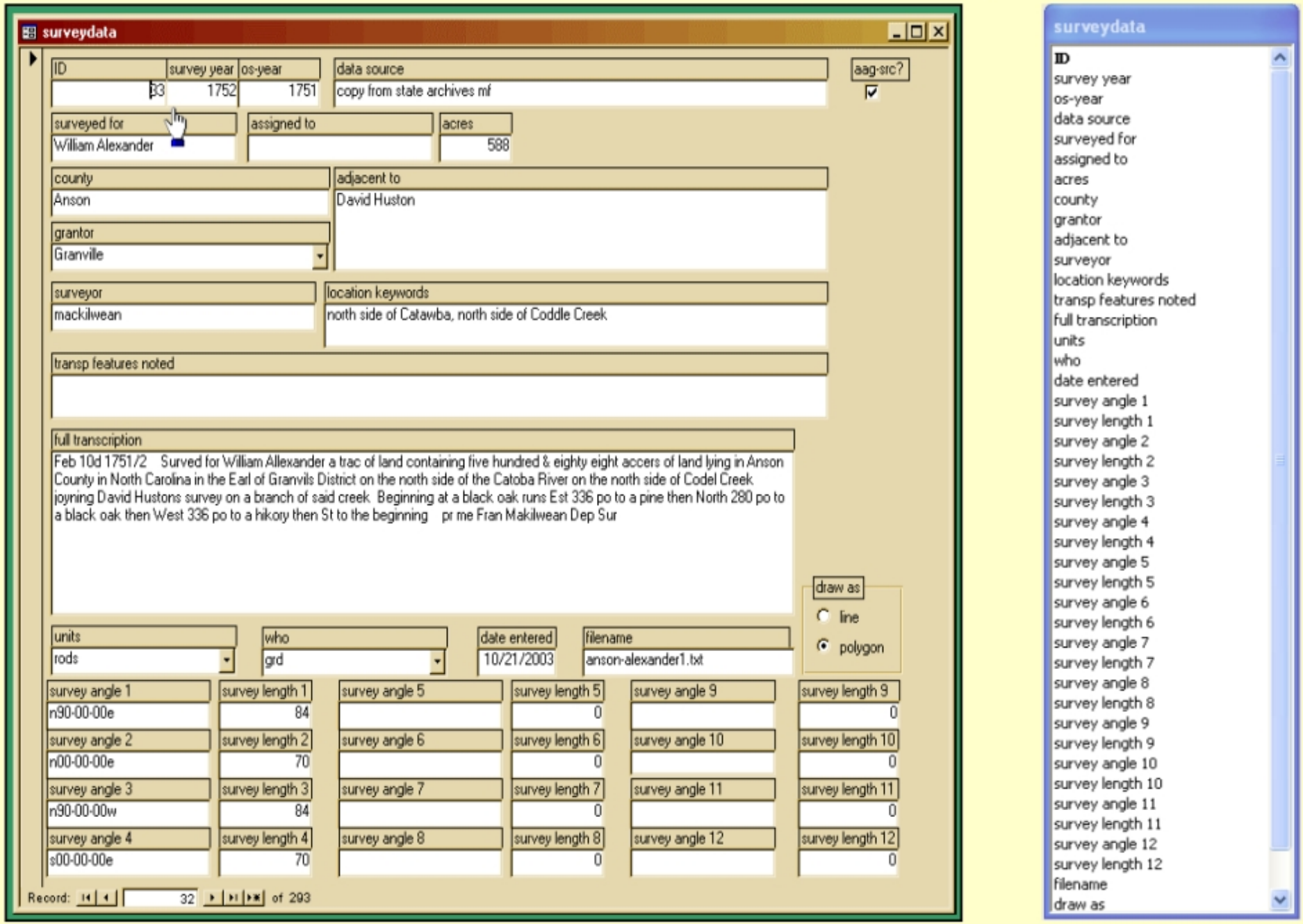
Feby 19th 1751/2 Surved for Georg Blain a trac of land containing six hundred & fourty one accers of land lying in Anson County in North Carolina in the Earl of Granvils District on the north side of the Catoba River on Third Creek on both sides about three mille above Thomas Gilespys survey Begins at a hikery runs Nt 270 po to a white oak then East 380 po to a white oak then St 270 po to a ride oak then West 380 po to beginning pr me Fran Makilwean Depy Sur

a. Full transcription of a survey document, as recorded in the main data entry form: Doc ID 32.

Desc#	Seq#	FeaType	Descpt Term	Primary name	Sfx name
1	1	Water	On N side of	Catawba	River
1	2	Water	Straddles	Third	Creek
1	3	Parcel	about 3 miles above	Gilespy's, Thomas	land

b. A "Feature Lookup" sub-form, part of the main data entry screen, allows Dobbs' to parse keywords of interest from the full transcription. Sequential numbering allows for concatenation into full sentences (e.g. George Blain's parcel is 'On N side of Catawba River, straddles Third Creek, about 3 miles above Thomas Gilespy's land.')

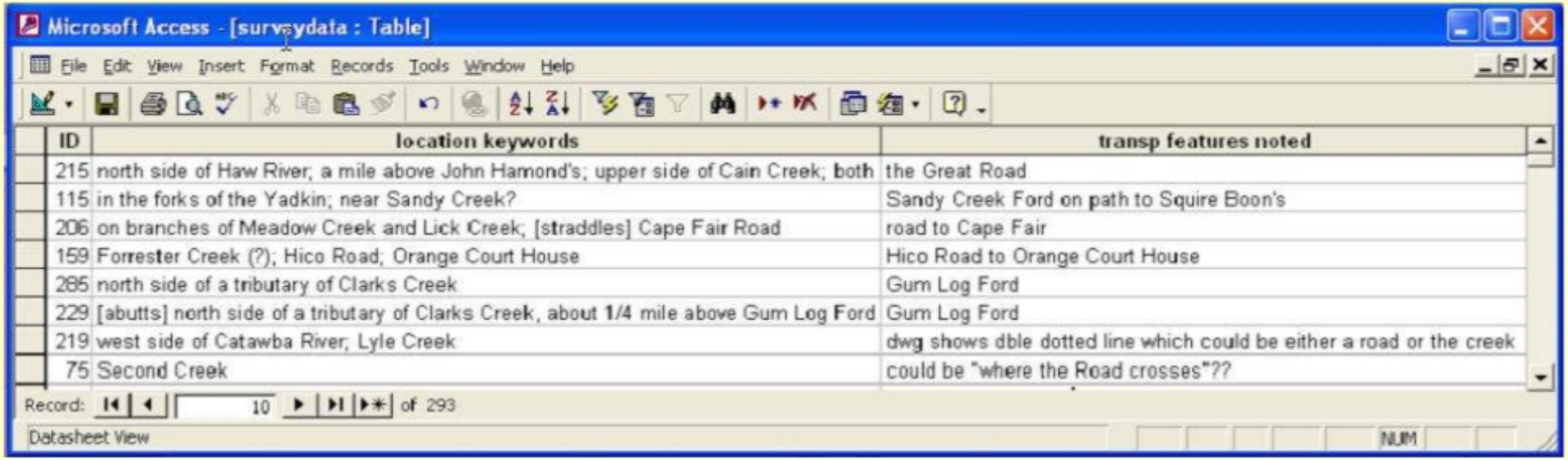
FIG. 10. Published annotations – intended for publication



a. Original data entry from for recording information from survey documents

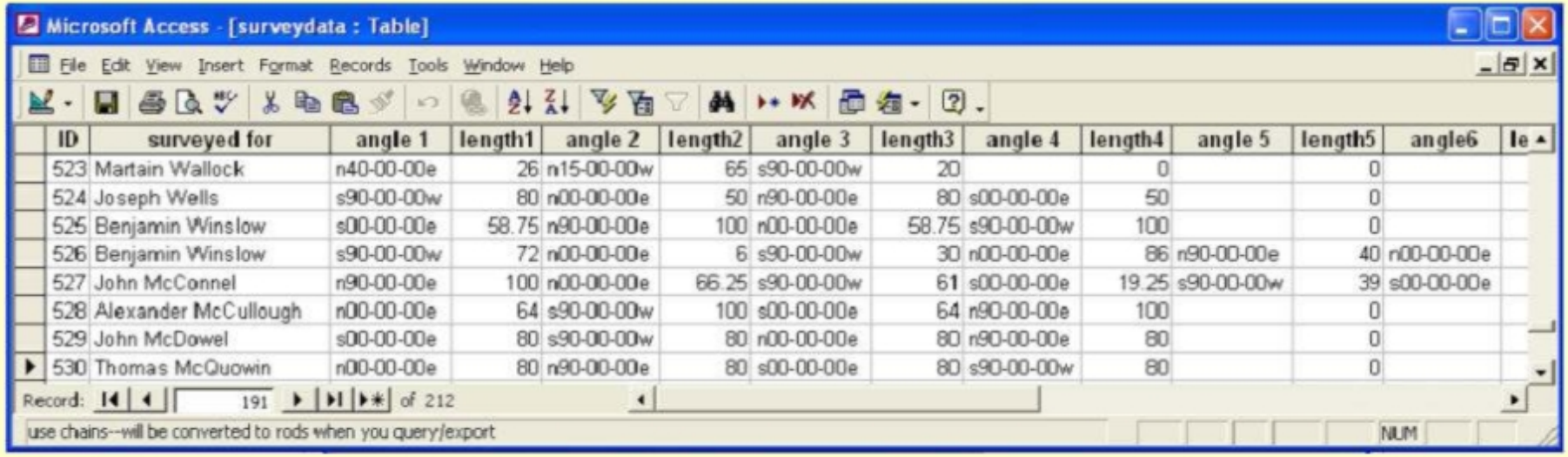
b. Underlying single table

FIG. 11. Original data entry form and single database table

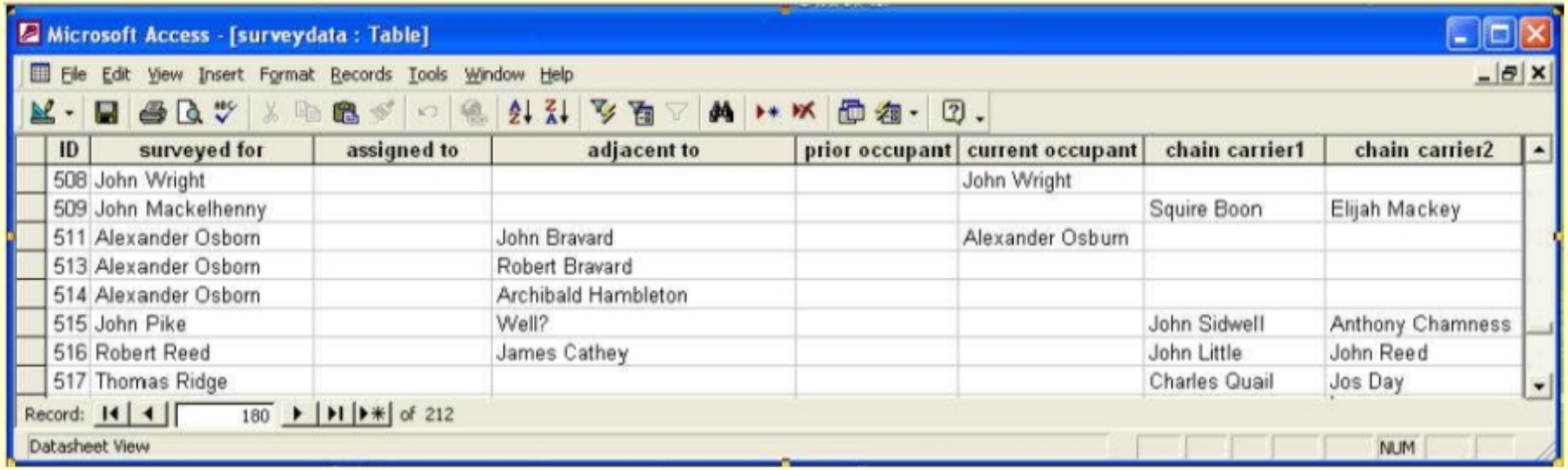


a. Searchable keywords commingled with unstructured annotation

FIG. 12. Original database retrieval issues - annotation mixed with search terms



a. Twenty-four redundant fields used for two entity types: angles and lengths



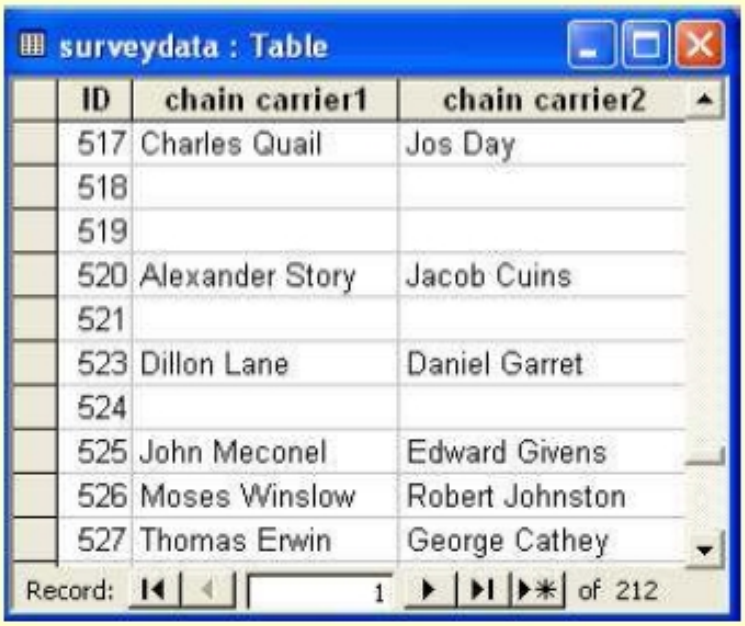
b. Seven redundant fields used for one entity type: people names

One cause of retrieval problems in a flat file are the "data redundancies", inherent in a single table database design. Shown here are two views showing 19 out of the 43 fields in Rebecca's original database. The top illustration shows the use of multiple columns for collecting sequentially paired survey measurements: angle and distance (the sequence identified by: 1, 2, 3, etc.). The lower image shows multiple columns being used for collecting different types of people names: such as the person the land was surveyed for, assigned to, adjacent to, who the prior occupant was, or current occupant, and who the chain carriers were, etc..

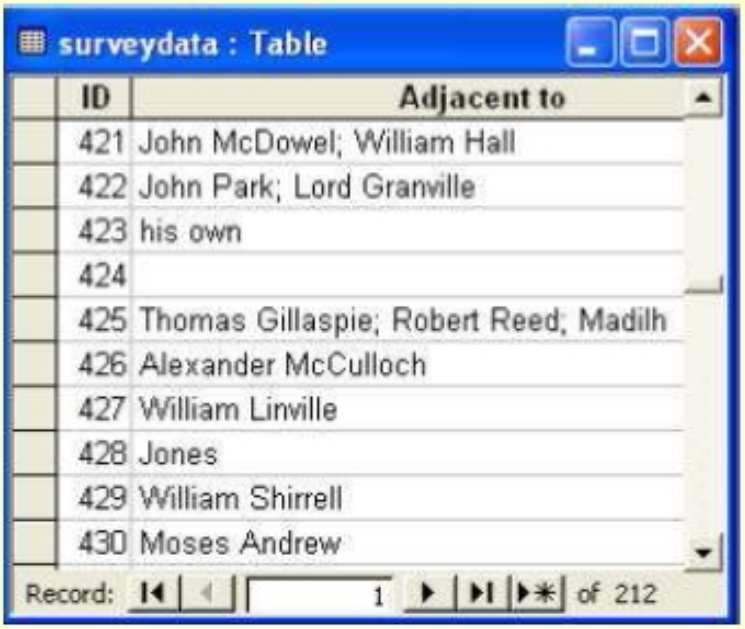
Creating separate fields (or columns) for entities that share common attributes makes it very difficult to retrieve and compare information. For instance, when looking for people with the same name an advanced union query would be necessary to join the seven individual 'person name' columns into one list for comparison. To make matters more complicated, the content entered into these fields was inconsistently formatted.

FIG. 13 Original database retrieval issues - multiple columns

FIG. 13. Original database retrieval issues - multiple columns

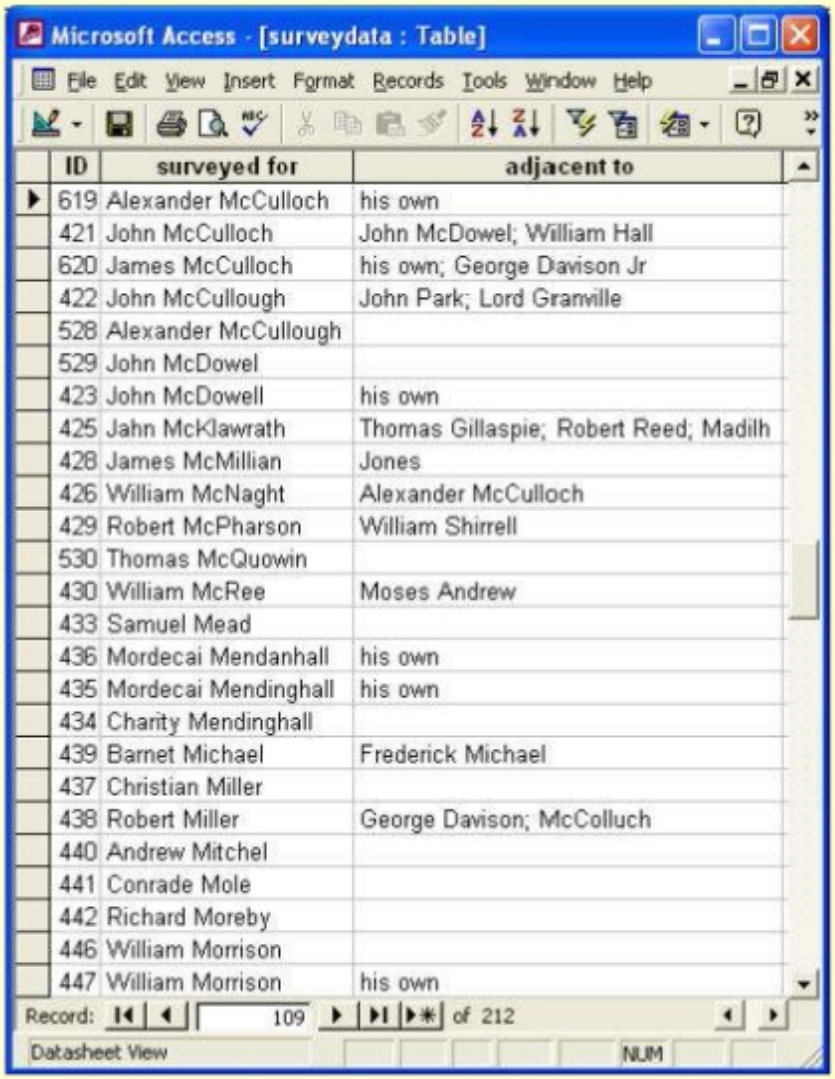


a. Multiple columns, across row

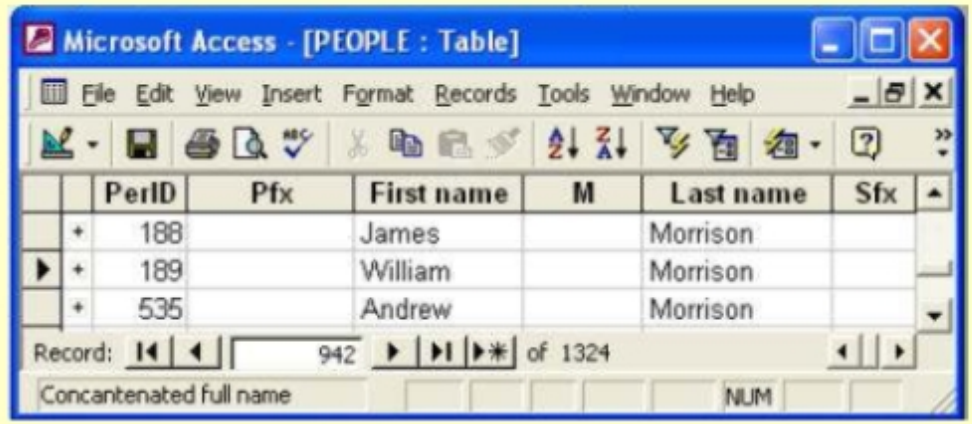


b. One column, names commingled

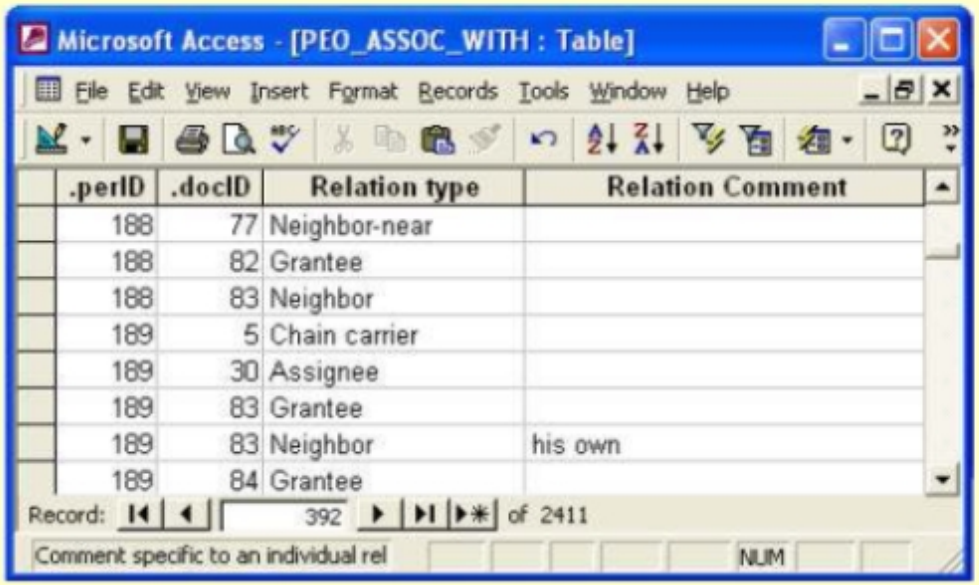
FIG. 14. Original database retrieval issues – multiple values & complex attributes



a. BEFORE: Portion of original single table database



b. AFTER: new PEOPLE table



c. AFTER: new PEO_ASSOC_WITH table

On the left (a.) is a small portion of Dobbs' original single table database. These columns represent just two, out of the seven, she had been using to record different categories of people; who a property had been "surveyed for" and who it was "adjacent to". This format made it impossible to sort by last name, and was difficult to query for a particular person that could exist across multiple columns.

In the new database, two tables (b. and c.) were used to record information about people. The PEOPLE table (b.) stores unique names in a simple attribute format; the prefix, first name, middle name, last name, and suffix of each individual. This eliminated duplicate names found across columns in the old database and facilitated sorting.

The second table (c.), PEO_ASSOC_WITH, was linked to the PEOPLE table using a one to many relationship for identifying the "role" each unique name played in an individual land grant transaction; roles such as grantee, assignee, neighbor, chain carrier. As new roles were identified Dobbs could easily add a new category, as she did for the terms "prior claimant", "Attester" and "debtor" that appeared in later documents. In the old database she would have needed to add new columns. A place for recording unstructured annotative comments was also included (e.g., Relation comment).

was also included (e.g., Relation comment).

FIG. 15. Before and after methods for recording people names

surveydata : Table

ID	location keywords	transp features noted
157	McCulloh's line, Great Lick Creek, waters of Neuse	
158	Double Creek	
159	Forrester Creek (?); Hico Road; Orange Court House	Hico Road to Orange Court House
160	waters of Little River	
161	Lick Creek or Fork (waters of Cape Fear)	
162	[straddles] Shadocks [Shaddox?--see gazetteer] Creek, waters of Haw River	

Record: 162 of 293

a. BEFORE: Portion of original database

FEATURE : Table

FeaID	FeaType	Primary name	Sfx name
144	Cultural	McCulloh's	Line
145	Water	McDowells	Creek
146	Water	Meadow	Creek
147	Water	Mendinghales	Creek
148	Water	Middle	Creek
149	Water	Mill	Fork
150	Water	Mill	Pond
151	Cultural	Moravian	land
152	Water	Mud	Branch
153	Water	Mudlick	Creek
154	Water	Neuse River	Basin

Record: 154 of 411

b. AFTER: new FEATURE table

FEA_LOCATORS : Table

.feaID	.parID	Descpt#	Seq#	Descpt Term	Comment
144	157	3	1	abutts	
144	179	2	2	abutts	
144	187	3	1	Tract includes	
144	309	2	1	Tract includes	
144	1105	3	1	unknown relationship	prob joins, but unclear
145	284	1	1	On S fork of	
146	206	1	1	On branches of	
147	171	2	1	Straddles	below lower falls of
148	100	2	1	On W side of	
148	1057	1	1	beg in fork of	
148	1094	1	1	On N side of	

Record: 511 of 1089

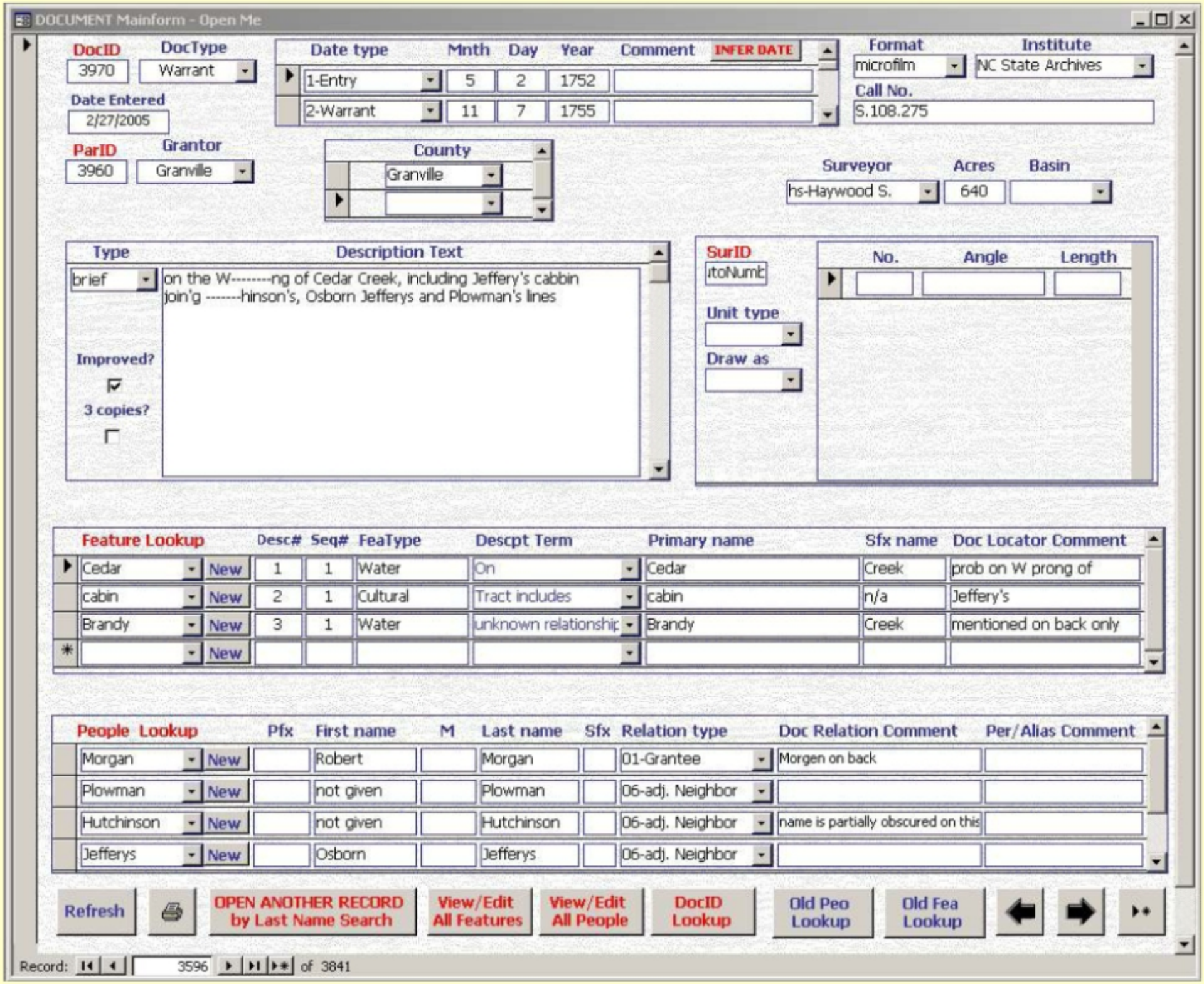
c. AFTER: new FEATURE LOCATORS table

Features were being recorded in Dobbs' original table in two separate columns (or fields); one designed for geographic evidence (e.g., location keywords) the other for storing transportation features (e.g., transp features noted). This format made it impossible to sort by feature name, and difficult to form queries for comparing between the two as well to other columns containing related content, such as people names.

In the new database, two tables (b. and c.) were used to consolidate both types of features in order to collect the data more efficiently. The new FEATURE table (b.) contains fields for the type of feature (e.g., water, cultural, transportation, etc), the feature's primary name and related "suffix" (e.g., creek, fork, pond, basin). The new FEATURE LOCATORS table (c.) is linked to the FEATURES table and stores directional terms or commentary that had previously been commingled with the feature's name. Removing the unstructured "annotation" from a feature's primary name greatly improved the database's search capabilities. As an added bonus Dobbs' could now search on a feature's primary name and a person's last name to find matching records, which is often helpful. For example "Cathey" is the last name of a grantee and also the name of a creek (e.g., Cathey's Creek).

a grantee and also the name of a creek (e.g., Cathey's Creek).

FIG. 16. Before and after methods for recording feature names


DOCUMENT Mainform - Open Me
DocID
3970
DocType
Warrant
Date Entered
2/27/2005
ParID
3960
Grantor
Granville
Date type
Mnth
Day
Year
Comment
INFER DATE
1-Entry
5
2
1752
2-Warrant
11
7
1755
Format
microfilm
Institute
NC State Archives
Call No.
S.108.275
County
Granville
Surveyor
hs-Haywood S.
Acres
640
Basin
Type
Description Text
brief
on the W--------ng of Cedar Creek, including Jeffery's cabbin
join'g -------hinson's, Osborn Jefferys and Plowman's lines
Improved?
3 copies?
SurID
itoNumb
Unit type
Draw as
No.
Angle
Length
Feature Lookup
Desc#
Seq#
FeaType
Descpt Term
Primary name
Sfx name
Doc Locator Comment
Cedar
New
1
1
Water
On
Cedar
Creek
prob on W prong of
cabin
New
2
1
Cultural
Tract includes
cabin
n/a
Jeffery's
Brandy
New
3
1
Water
unknown relationship
Brandy
Creek
mentioned on back only
New
People Lookup
Pfx
First name
M
Last name
Sfx
Relation type
Doc Relation Comment
Per/Alias Comment
Morgan
New
Robert
Morgan
01-Grantee
Morgen on back
Plowman
New
not given
Plowman
06-adj. Neighbor
Hutchinson
New
not given
Hutchinson
06-adj. Neighbor
name is partially obscured on this
Jefferys
New
Osborn
Jefferys
06-adj. Neighbor
Refresh
OPEN ANOTHER RECORD by Last Name Search
View/Edit All Features
View/Edit All People
DocID Lookup
Old Peo Lookup
Old Fea Lookup
Record:
3596
of 3841

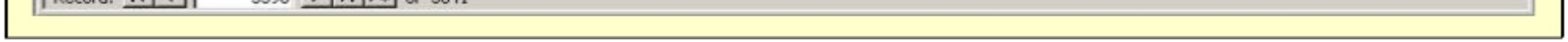


Fig 17. Current data entry form – populates multiple relational tables