David Sworn
Sturgeon Falls Mill Closing Database
Supervisor: Steven High
Thus far, the InterClipper database development has been a process of trial and error and decisions about the direction of the project have often been made out of necessity rather than sound calculation; at a number of points, project benchmarks should have been modified and the way that the overall project was being tackled should have been rethought. There are a number of reasons for these missteps: inexperience on my part, time-consuming technical difficulties, and the fact that this has been a largely one-person project. Nevertheless, the project has been successful in a number of areas and, in any event, this has been a learning process and there are a number of ways in which key stages may be tackled more effectively in future.

This report lays out the methods by which the database, in its current incarnation, has been developed and it outlines the logic behind these. Equal attention, however, is devoted to how these methods may be improved in future – my suggestions for “best practices,” to employ an unfortunate but highly appropriate management buzzword. To put this in concrete terms, the development of the database can be divided into two broad stages: the clipping stage, in which the interviews were “divided” into short segments, and the indexing stage, in which these were thematically categorized. The clipping techniques have been systematized and, with some qualification, could be called efficient and effective, while the indexing methods have largely been haphazard and were frequently ill-conceived.

The Blank Database

The project involves a dozen full interviews, in all over 15 hours of video. Each interview is divided into separate video files which are between eight and twenty minutes long. In the database itself, these separate video files are titled “Childhood,” “Life at Mill,” “Management,”
and so on, but these titles often misleadingly suggest thematic distinctions between the segments; it may be more appropriate to title them “Part 1,” “Part 2,” et cetera.¹

The interviews are segmented in this way so that they are easier to work with in the timeline. It isn’t clear, however, that this segmentation is justified: precise work in the timeline is normally only necessary at the clipping stage and, in any case, InterClipper is equipped with several useful magnifying features which allow the user to zoom in to a tiny segment of the timeline with a click of the mouse. The fact that interviews are always segmented when they are streamed into the database suggests the extent to which the researchers in Buffalo have been developing InterClipper with individual (exported) clips in mind. Because we have been concerned with maintaining the integrity of complete interviews, it would be useful for us to experiment with full, un-segmented interviews.

Starting Out

The sheer size of a database like this can be quite intimidating and at the early stages it should be approached systematically on an interview-by-interview basis. Working alone on the project, I tended to think of it too holistically, particularly at the beginning; I would watch and clip an interview for the first time and then move on to the next one, thinking about how near or far I was from completing the clipping for all the interviews. The problems with this approach are manifold. Watching a single 90-minute interview with 15 hours of interview in mind makes the task seem more daunting and exhausting than it actually is. Although the interviews remained distinct in my mind even at this early stage, the tasks of note-taking and clipping tended to obscure, rather than elucidate, my overall impressions of each interview and it was often weeks before I watched lengthy sections of each interview for a second time. Also, I think it would

¹ This change could be made quite easily to the current database.
have been extremely useful – perhaps more useful than much of the material that is currently included in the notes section of the database – to have written a short reflection on each interview after watching it.

Thus, my suggestions for approaching a set of recorded interviews for indexing are as follows:

- *Researchers should meet and discuss each interview.* The interviews should be divided between two or more researchers who should meet after completing the clipping and note-taking for each interview. If possible, the project leader and/or the interviewer should also meet with the researchers.2

  These meetings need not be formal or lengthy – probably no more than half an hour – but they would give the researchers the opportunity to share their impressions of the interview and, importantly, to offer their colleagues an outline sketch of the interviewee and his or her narrative (this will make it easier to develop short biographies and interview outlines which, if possible, should be completed at this stage). Crucially, these meetings will allow the researchers to discuss any patterns or anomalies in the interviews and to begin compiling tentative methods of categorization and potential indexing tags and topics; likewise, any foreseen problems with indexing could be discussed and the project leader would be able to ensure that the direction of the project is in keeping with its research goals.

- *Researchers should write reflective summaries of each interview after at least two viewing sessions.* Researchers may wish to jot down a few reflections after the first viewing session and after completing the clipping and note-taking, but the short reflection on the first viewing is not as useful as it is after the final viewing.

2 “Researcher,” here and below, refers to the individuals working on the database and “project leader” refers to the individual supervising the project. In the future and particularly for those using Zotero, there may be only one person (probably a student) indexing interviews for their own project, in which case there will be no need to make sure that others are on the same page. Nevertheless, it will be helpful to talk about the individual interviews with colleagues that are familiar with the software and/or the research project.
viewing and the initial clipping is complete, but generally full reflections should be completed after a second viewing session, preferably on another day. This is because the clipping stage is quite intensive and there is a tendency to fall into a kind of trance, focusing on making precise clips and accurate notes and paying less attention to the nuances of the interview. It is important to recognize that because the notes are taken very quickly and, as much as possible, without stopping the video, the process is not conducive to the kind of deep listening that transcription often involves (see the section on clipping and note taking). Therefore, the reflections should be completed after a second viewing.

This does not mean, however, that each interview needs to be re-watched in its entirety: the researcher should take advantage of the clips and notes that he or she has made during the first session to determine which sections demand close attention. In many cases, the earlier parts of an interview will require less re-watching than later sections, because at the beginning the interviewee is getting comfortable and finding his or her voice.³

At the re-viewing stage researchers may consider the effectiveness of the clips and how they might be improved (with embeds, for example, or by dividing), but they should really be concentrating on the interview itself: they should be looking for narrative patterns, bodily gestures and facial expressions. Having concentrated on the timeline and the notes section in the first viewing, the researcher can now pay attention to the video

---

³ In the Sturgeon Falls database, there are long sections (occasionally over fifteen minutes) at the start of each recording that include an explanation of the project, the consent form and so on, these of course do not need to be watched more than once. Likewise, responses to early biographical questions tend to be short and concise; it takes a few questions before the interviewee begins to provide long and complex answers. This is not always the case, however, but the interview should still be fresh in the researcher’s mind and it will be easy to remember which sections need closer attention and which do not.
itself. Researchers should take notes on their meta-linguistic observations in a few words at the end of the notes section.

- **Compiling the reflective outlines and adding them to the database.** The reflective outlines should act as introductory sketches for each interview and should be no more than a few short paragraphs long. The first paragraph might give the relevant biographical details of the interviewee and how his or her narrative relates to others in the database. Subsequent paragraphs should enumerate the major themes of the interview and outline the interviewee’s views on key issues as well as the places, events and subjects that are of particular concern to him or her. Information about the interviewee’s emotions, language patterns and bodily expressions should also be included here.

Researchers should make use of clip ID numbers to refer to specific clips; for example, if an interviewee exhibits an interesting pattern of laughter, it would be appropriate to note the ID numbers of some or all of the clips in which it appears. In this regard, it will be helpful to export the notes from InterClipper to a Word document, compile the reflective summaries in Word and then cut and paste them into the database. A suitable field would be the notes section of the Session Info box (View >> Session Info), which allows some formatting (paragraph breaks) and can be modified at a later stage (currently this field contains technical information from the video streaming process which is no longer relevant to the project).

These reports would be an extremely useful addition to the database but, more importantly, this process would help maintain the independence of each interview. Because the InterClipper database includes an unchangeable number of interviews, it is easy to view them largely in relation to one another and to over-emphasize cross-referentiality, this becomes a more obvious

---

4 This feature exports the clip ID numbers, the clip labels, and either the notes or the transcript fields.
problem at the indexing stage, but it begins at this early stage when the database is approached as a whole, rather than on an interview by interview basis.

This may be less of a problem in the future, however. Ideally, interviews would be clipped very soon after being conducted and would exist independently of one another, to be included in any number of user-defined databases (this would be the case with Zotero, for example). In essence, this brings to mind a rather obvious point, but one which I’m afraid I haven’t always borne in mind and one which the InterClipper database tends to occlude: it is quite likely that a user will want to concentrate on one interview alone rather than the full database and the way that the database is organized shouldn’t impede this kind of approach.

Clipping

The previous section dealt with the clipping stage in broad terms, this section focuses on clipping in greater detail and outlines the question-answer approach used for the Sturgeon Falls database.

According to Doug Lambert, the method of clipping used in Buffalo involves watching interviews and making a clip when the interviewee says something particularly significant. The clip is then adjusted to its proper length and given an appropriate title (or “clip label”) at a second stage. This might be called a “selective” approach to clipping because large portions of each interview are not clipped; the interviewer’s questions, for example, are not normally included in the clips. For the Sturgeon Falls project, however, we decided that the indexes should be as comprehensive as possible and that all sections of the interview should be clipped. To do

---

5 Initially I clipped all sections of the video including interruptions, pre-interview small talk and so on. Eventually I removed these extraneous clips, but pertinent information from the pre-interview could still be clipped. In many cases, the final clip in the interview ends just before the camera was turned off – most post-interview conversation should be clipped.
this, I have started clips at the beginning of substantial (non-follow-up) questions, and ended
them just before the interviewer asks a new question. I like this method because I think it creates
clip patterns that provide an interesting “map” of each interview. I also feel that, generally
speaking, dividing the interviewee’s narrative into sections that are smaller than a single
response risks fragmenting and de-contextualizing it. That having been said, responses are
frequently long and often cover a number of topics, in these cases embedded clips have been
used so that the user can watch sections of a response and then refer to the full response to see
the full context.

Indexed clips (long clips that contain multiple clips are not indexed) range from less than
30 seconds to one clip which is over seven minutes; the majority of the clips are more than a
minute and less than five. Generally speaking, an average 90 minute interview would have
between 30 and 40 clips but the deviation from this average is quite large due to the ways that
interviewees choose to respond to questions and to the way that questions are phrased.
Responses tend to be short at the beginning of interviews and longer towards the end, thus there
tend to be a disproportionate number of clips at the beginning of the interviews.

Clipping is a three-part process, it involves making and adjusting the length of the clip,
taking notes on the clip, and giving it a title.

- **Working in the timeline to make and adjust clips.** This is probably the technique with
which I am happiest; it is easy and systematic and ensures that all parts of the interview
are clipped. Because one clip ends where the next begins, it’s easy to incorporate this
process into the first viewing of an interview and, at the end, it provides a grid of clips
that are ready for indexing. If a user would like a “sound bite” or short quotation clip for
use in a presentation or website, it is a very easy process to make a clip, adjust its length
and export it. In my view, the database should help users locate and re-locate such quotes where necessary, it should not be a compilation of sound bites.

This process could be improved, however, by making more embedded clips. Embeds can function as very effective clips in their own right, but they can also help to overcome the limit of one index tag per column, per clip. Short embeds could be a way of point to a specific topic within a clip, as long as it was clear to users that they were being used for this purpose and that they should really refer to the longer clip.\(^6\) In the later stages of developing the database, I found myself adding quite a few new embeds (and removing some, too). Most embeds should be made at the first and second viewing, but some may need to be added at the indexing stage. The clip references in the reflective outlines may have to be updated when new embeds are added.

- **Taking Notes.** Having set the starting point of a clip, the researcher takes notes on what is being said in the recording. This process allows some variability between researchers and depends, in part, on their typing skills. But I think note taking needs to be made more systematic and more thought needs to be given to the purpose that notes are going to serve in the finished database. I mostly took notes thinking that they would be helpful at the indexing stage – and they were – but notes need to be taken on meta-linguistic information as well and this should be done during the second viewing.

  Notes should include names, dates, important events, opinions and so on, but these should be described in a few words, rather than a few sentences. Most of my notes are too long (although some of my early notes are too short): point form should be used as much as possible and it may be preferable to avoid paraphrasing what the interviewee says (which is what I’ve done). A key could be developed by researchers; for example, an

---

\(^6\) This method seems preferable to simply duplicating clips so that they can be given more index tags.
asterisk could be used to mark important or interesting points (although such markers should be used sparingly) or “H” could mark hesitation – in my own note taking I’ve frequently used “R” to show when logical reasoning dominates the narrative.

Unfortunately there is no formatting in the notes field so it is impossible to make paragraph breaks, this hasn’t been a problem for me, but researchers could used two forward slashes to separate content notes from meta-linguistic notes where necessary. Square brackets should also be used for researcher or user comments.

Currently the transcript field has not been used. If researchers choose to work in this field in the future, they should bear in mind that only the notes or the transcripts can be exported to a single text document.

- Creating clip titles, or “clip labels.” When the interviewee has finished responding to a question, the researcher stretches the clip to the appropriate length and decides whether embeds are necessary, before moving on they should come up with a title. This is a process that also needs systematization.

I had imagined that users would look at the index tags and then refer to the clip label before watching a clip, although I’m not sure that this is how they have been used in practice. In any case, the clip labels are too haphazard at the moment – some are good but quite a few seem like they might be useless. Presumably clip labels should be more specific than index tags but more general than the notes. More experimentation might be helpful in developing a good system for clip labels, but I think it would make sense to use the “clip label” field for keywords arranged in descending order of relevance. This would probably be more useful than “titles” which risk being overly vague.
The clip labels tend to be quite long which can be a problem because it means that they take up a lot of room on the grid (even if the Quick Sort\textsuperscript{7} feature is being used). A neat solution for this would be to copy the clip labels into the beginning of the transcript field (especially if it isn’t being used for anything else); this would make it easy for users to refer to the full clip labels without using precious space on the grid.\textsuperscript{8}

\textit{Indexing}

This has probably been the least successful part of developing the Sturgeon Falls database; it has certainly been the most difficult. Because I feel less confident about this area, my recommendations are more general than those of the previous sections, although I’m sure that certain practices will lead to far more effective indexing in the future.

My process for indexing Topic A, a field which literally indexes the topic of each clip, was to move from notes to clip label to index tag. As result, my first “draft” of Topic A was like a generalized list of clip labels; there were nearly 150 in all.\textsuperscript{9} These were then aggregated into fewer than twenty index tags. This process probably sacrificed too much specificity in favor of generalized tags, but nevertheless I feel that Topic A is the most effective of the indexical categories.

Topic B ultimately followed a similar pattern. I tried to have this category address broader historical themes, but the result may have been even greater over-generalization. It would have been more effective for Topic B to have dealt more clearly with research question associated with the book. Indeed the database could only be interesting if Topic B involved a

\textsuperscript{7} Quick Sort allows the group the clips by one or more column headers.
\textsuperscript{8} If the clip label was copied into the transcript field, users could also take advantage of the Preview feature: right-clicking on any clip on the grid and selecting Preview >> Transcript opens two lines of blue text from the transcript field directly on the grid, this would be perfect for clip labels.
\textsuperscript{9} This draft is included on the accompanying DVD.
more radically interpretative approach, as long as it was appropriately framed (possibly using the reflective passages).

Topic C has, of course, been the least successful category. In retrospect I should have used the method outlined above to develop a second tier of notes on meta-linguistic data and then used these to produce tags. Early in the clipping stage I became concerned about the challenges associated with indexing body language and speech patterns and, unfortunately, I let my trepidation get the better of me: I came up with preconceived tags that were far too general and added nothing to the database.

The indexing process would have been more productive had the following techniques been effectively employed:

- Establish clear aims regarding what the database should include and who will be using it. It may seem that proper goals should be established earlier in the project – they should – but I feel that the clipping processes are well-suited for various goals and, in fact, that a properly clipped, annotated and labeled database would be extremely useful even if it wasn’t indexed. Although clearly defined research questions would help researchers write effective reflective outlines, it’s at the indexing stage that goals become crucial because decisions about categorization frequently determine the direction of the database in concrete ways.

Equally important is the question of who will be using the database and how they will be using it. At one extreme there is the hypothetical lone researcher who develops a database as a methodological tool and has very specific research questions in mind, in

---

10 It would be interesting to compare student responses to an “un-indexed” database with responses to the “complete” database. For the former, users could look at the clip labels of each interview and view clips based on these and the notes – with an average of less than 40 clips per interview, this would hardly be overwhelming. Clearly such a database would include no cross-referentiality between interviews, but the user could look for patterns themselves and use the tags to mark related clips in different interviews.
this case there may be relatively little indexing necessary. At the other extreme is the database for public or near-public use, this would include extensive indexing and, presumably, would have to lend itself to a great variety of research questions. Somewhere between these two extremes is the database developed for archiving. My feeling is that a database developed by the “lone researcher” would be very useful for archival purposes, if not as useful as one that is fully indexed. In any event, it is very difficult to anticipate how a database will be used, but it seems clear that certain methods of categorization facilitate some ways of approaching interviews while inhibiting others; the details of how the construction of databases really determines the way they are used remains to be seen, however.

- **Researchers should work in groups to establish categories and tags.** Categorization is a complex and difficult process and it requires not only trial and error, but a lot of brainstorming and experimentation. Developing and workshopping ideas with a group of people that are familiar with the project would be an extraordinarily helpful process. As much as possible, the project leader should be included in this group work. Categorizing and indexing have an obvious impact on the overall success of the project and it is important for researchers to be confident about the decisions they make and to feel that they have the support of colleagues and their supervisor.

- **Develop categories for whole databases but index on an interview-by-interview basis.** Over emphasizing common themes in different interviews can lead to unhelpful, overly-broad indexing tags, and can hide the crucial specificity of each narrative and each clip. Just as focusing on each interview, rather than the entire database, would be helpful at the early stages of development, so would considering the idiosyncrasies of each story better
serve the indexing process. Users should be able to study clip labels to get an impression of what an interviewee has to say, but they should be able to look over an interview’s tags to see what it has in common with the other interviews as well as how it differs from them. Thus, index tags should be developed for each interview and then minimal adjustments should be made to integrate them into the rest of the database. Likewise, users should be encouraged to start working with individual interviews and then to look at common themes in other recordings.

As we move away from InterClipper and towards Zotero, the databases we work on will become increasingly flexible and tagged interviews will frequently be integrated into new databases after they’ve been tagged. It will therefore become more important to index an interview as a discrete entity and not as part on an evermore arbitrary whole. Moreover, Zotero doesn’t use category fields the way that InterClipper does – a body language index tag would be interchangeable with a tag about a union busting – so indexing will become more important and the categorization of index tags will be obviated.

Clearly the databases we work on need to be more than the some of their parts, but sacrificing specificity for the sake of cross-referentiality can reduce the usefulness of the whole and its parts. Researchers should develop indexes for each interview not as part of a whole, but independently of the database; then they should workshop these index tags with their colleagues and input them into the database; finally they should tweak them so tags that are very similar to those of other interviews are made to refer to both recordings. This will, in the long run, make for more nuanced and interesting databases.
Thus there are two overriding recommendations here. Firstly, there should be as much collaboration as possible in database development – especially now when it is still new to most of us, and particularly at the indexing and categorization stages when researchers are not developing databases for themselves but are working for a supervisor on a project that needs to be accessible to others. Secondly, cross-referentiality should not be allowed to overshadow the differences between interviews; indexing should emerge as organically as possible from each interview and should then be carefully and partially integrated into the database. In retrospect it would have been useful for me to completely index a few interviews before I even watched the rest – to develop prototypes within the prototype, as it were – but my desire to refine the clipping techniques outweighed any instinct to move onto indexing at the early stages. Meanwhile, the question of how the databases we develop are actually used remains open. As we move from hypothetical, imagined users to actual students working with the Sturgeon Falls database, we need to think of new techniques for “training” these users. It seems to me that we should be consider a dual or even three part manual: one part would deal with the technical aspects of working with InterClipper (or Zotero), the second would give suggestions for how users can get the most out of the database, and the third would provide context for the interviews themselves.

It may seem that the approaches I’ve outlined above will make database development a long and drawn-out process. I am absolutely certain, however, that if these methods were systematically employed, database development would be much faster and easier than it has been in the past. I really feel that increased systematization is the way to go here and I think the recommendations I’ve made allow for more interpretation and critical engagement with interviews – and with the process of database development itself – on the part of researchers.
Coda: What now for the Sturgeon Falls database?

It may be that the Sturgeon Falls database needs a dramatic overhaul. If this is the case I hope that further develop brings if in line with the methods I’ve outlined above. I don’t think this will take very long, because the clipping will not have to be re-done, even if much of the indexing does.

Aside from this, I suggest that we move forward with the database we have. In outline, I think that we should really take advantage of InterClipper’s export features to set up a very simple web-based database using Dreamweaver. I imagine that this would look very much like the grid but the play buttons would link to video clips which would play in an external viewer. Different ways of categorizing the grid could be presented as different web pages. I think a prototype should be set up with one or two interviews. Also, we should import the clips from a few interviews into Zotero and start experimenting with its tag indexing and notes systems, after all these won’t change when the new video indexing features are added, and it would be useful to get familiar with them sooner rather than later, we may save ourselves some headaches in the long-run. These projects should progress in parallel with any re-working of the existing database.