

Using Real-Time 'Freeform' Annotations as Qualitative e-Research Metadata

Mike Fraser, Muneeb Shaukat, Serguei Timakov,
Phillip Smith, Jon Hindmarsh, Dylan Tutt, Christian
Heath, Anne Manuel, Marie Gibbs and Sally Barnes

MiMeG - Key areas of research

- Building on VidGrid pilot project reported in last year's conference
- Identifying needs of qualitative social science user communities
- Understanding social scientific practice
- Support for mixed media qualitative analysis
- Capacity building across the social sciences, software roll-out programme

Annotations in MiMeG - Transcripts

- Commonly used form of annotation in many types of qualitative analysis to illustrate talk
- Layouts, format and schemas vary methodologically and individually
- Transcription of talk may be conceived of as an analytic step in itself, i.e. transcribing the talk gives the researcher insight into the data.

Annotations in MiMeG - Transcripts

- Provide an important collaborative resource
 - encourage others to notice action that occurs around the words or utterances
 - can be crucial in reaching a shared perspective
- Our software supports timestamped text events
 - import existing text files as transcripts to be shown alongside the video
 - shared between distributed researchers by normal electronic means (e.g. email, web), or through the application via a database

Annotations in MiMeG - Scribbles

- Scribbles allow distributed researchers to communicate by freeform data annotation
- Scribbles produced directly over a video window by any users, who can all publish and subscribe to annotation events, and are assigned their own colour

Annotations in MiMeG - Scribbles

- Participants may highlight a phenomenon of interest for one another.
 - e.g. indicate or circle the area of the data under inspection, or mimic a gesture or movement that features in the video data with the scribble.
- Mimicking gestures are not providing ‘exact copies’ of conduct, but render relevant action visible as well as the analytic point that is being made about that action, by exaggerating or transforming the on-screen conduct

Realtime Research-driven Annotation

- Two forms of data mark-up for video data
 - researcher-produced insights or information for the purposes of sharing perspectives with colleagues
- We have not invented the need which these kinds of annotation are addressing, we have derived them from studying our particular domain of interest

Realtime Research-driven Annotation

- Are there other research domains which use transformed representations of primary data during analysis and require at least some communication around data with research colleagues or others?
- deriving our annotation mechanisms from domain study rather than from technology, has produced mark-up tools unlike metadata systems as traditionally conceived within e-Science

E-Science metadata storage and automation

- Automatic metadata
 - Metadata production is laborious/difficult
 - Errors or omissions are problematic
 - => Automate where possible, take the hit where necessary

OR

- Treat the creation of metadata as an important feature of the work of research

E-Science metadata – reuse across boundaries

- 1. *provide and share data across geographical and institutional* boundaries by engineering datasets which can be used by multiple stakeholders.
- metadata must be of reproducible and generaliseable form too
- e.g. semantic ontologies of the domain applied to each new dataset
- significant investment in development work (of the ontology as well as the software)

E-Science metadata – reuse across boundaries

- precludes research that does not fit the predetermined semantics of the domain without adapting the ontology independently of doing that research
- new research findings will reorganise domain semantics.
- Arbitrary distribution of datasets is inappropriate for many qualitative social science studies
- Data is rarely and carefully distributed in raw primary forms due to ethical, legal and consent constraints.

E-Science metadata - Reorganising workflow

- 2. *allow* independent modules of data to be processed together in any organization of workflow, allowing the application of old data to new problems
- To allow such module reorganisation, one must provide predictable inputs and outputs to the module, which in turn requires uniformity in that module's dataset. Again, conforming to a domain ontology guarantees this data integrity.

E-Science metadata - Reorganising workflow

- more typical in qualitative research, e.g. digging out old data which in retrospect may have relevance to a new interest
- Do we have to introduce of domain constraints or additional effort in identifying and reorganizing categories of metadata?
- Rely on individuals who own or who have encountered data being the primary users of retrieval mechanisms at potentially significantly later dates (which is inevitable where ethical constraints on data distribution apply)
- Rely on shared/stored memory of producing real-time annotations – as part of communication or transcription in our case – to recall the related data e.g. Pattern matching algorithms for that medium narrowed by broad categories of browsing

Shared annotations in MiMeG

The screenshot displays the VidGrid v1.0 software interface. The main window shows a video of two men in an interview setting. Red annotations (circles and arrows) are overlaid on the video. A central log window displays the following text:

```
00:00:00.000000000 Start of interview
00:00:01.057000000 Interviewer points left
00:00:05.853000000 Interviewer points right
00:00:19.542999872 See where interviewee is looking, and placement of the hands
```

Other visible windows include:

- File selection:** A table with columns 'File', 'Mime type', and 'Frozen'. It lists 'interview1.mpg' (video/mpeg, No) and 'interview1_log' (text/plain).
- Annotation settings:** Sliders for 'Persistence in seconds' and 'Brush width', and buttons for 'Clear', 'Clear all', 'Save', and 'Load'. The 'User Key' is set to 'mke2'.
- Play controls:** A timeline with buttons for 'Restart', 'Back 2s', '-0.1s', 'Stop', '+0.1s', 'Play', and 'Last play'. The current time is 00:00:19.542999872.
- Volume control:** A 'Mute' button and a volume slider.
- Offline users:** A list showing 'mke' as an offline user.
- Select Annotations:** A tree view showing 'Interview1.mpg' with sub-items for annotations by Mike Fraser at different times and locations.

The MiMeG logo is visible in the center-right of the interface.

Shared annotations in MiMeG

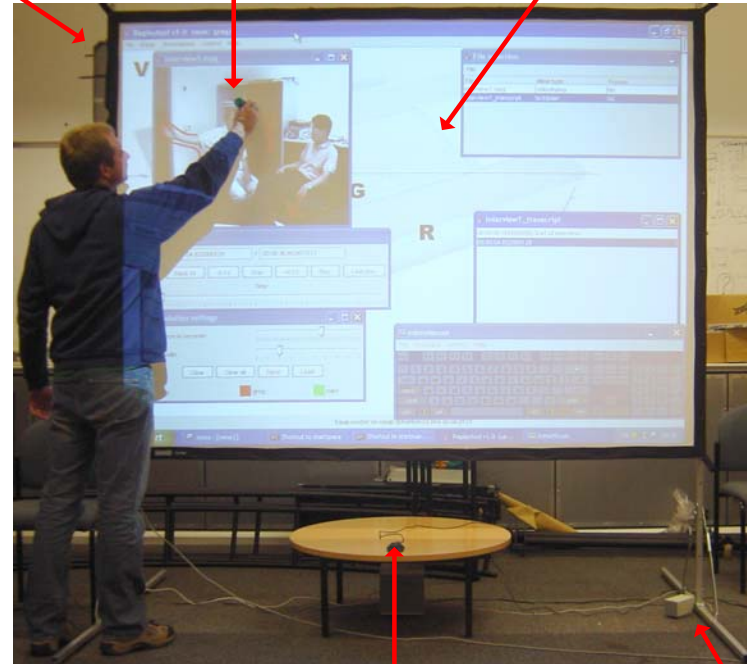
- stores scribbles according to broad categories (user log-in, self-declared physical user location, project, data file annotated and time/date)
- browse for known previous annotations, for example by retrieving the annotations of the most recent previous data session working on this data as a means to remind the researchers of their interim conclusions
- narrow search possibilities by linking searches of annotations, e.g. allow freeform annotations to be produced over transcripts as well as video and then pattern matching on those

Some further developments

Mimio receiver

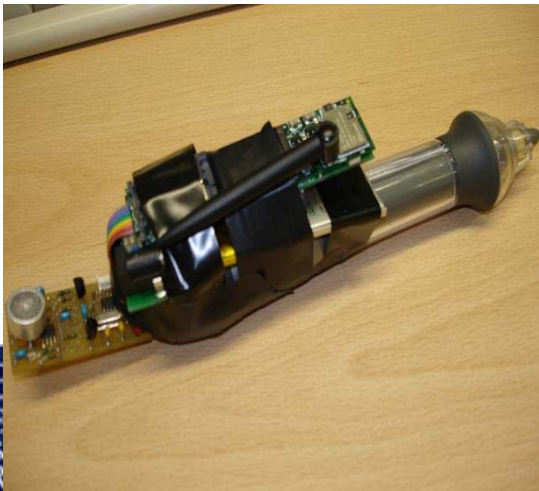
Mimio pen

Image projected to screen

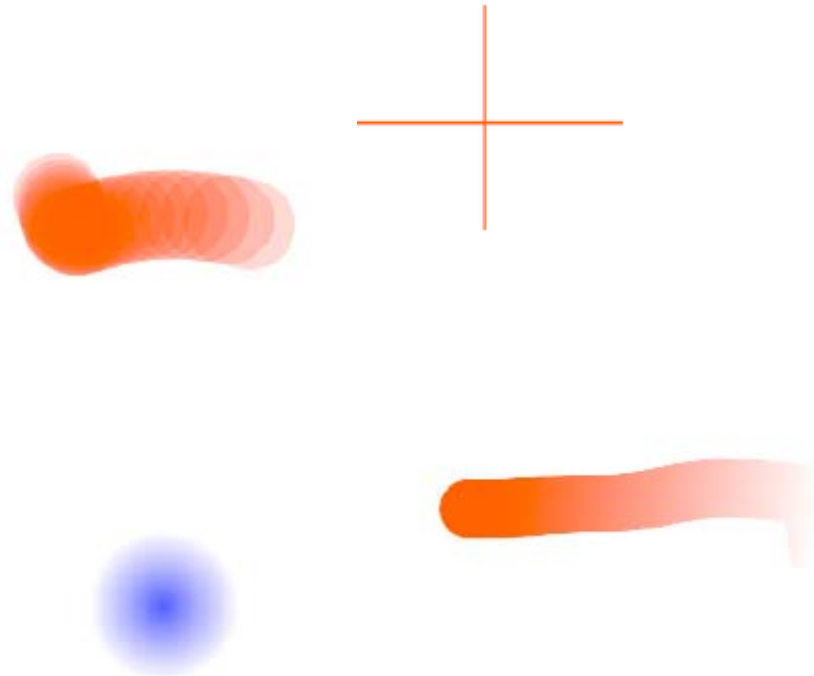
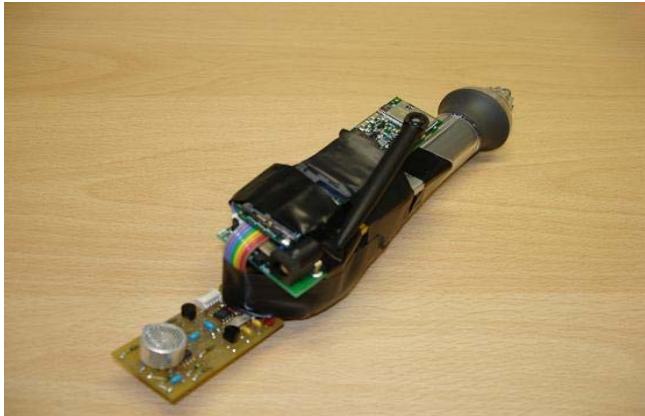


Boundary microphone

Speaker



Some further developments



Broader lessons

- Consider *intrinsic forms of metadata as well as extrinsic ones*
 - *Sharing policies should come from within a domain, not without*
- *One size does not fit all where research metadata context is concerned – some domains trade better on local reuse than global reuse*
 - *Identify where data might be repurposed locally as well as globally*

Next steps

- Observational studies of distributed analytic practices
 - Studies of MiMeG system in use in various domains
- Continued development of local metadata search techniques



Questions?

