It is perhaps an intuitive thought that the more heavily technologised a research situation is the more of a team effort it requires, whereas projects dealing with simpler technologies (e.g. email) are more easily within the capacity of a well-organised individual. To an extent this is true, but you can still carry out small-scale research in the more multimodal media if your aims are well defined (see the introduction to chapter 15) and your expectations are realistic. Expectations can only be kept at a realistic level if you are aware of the demands of the overall environment, even if you do not have to work with all aspects of it. This is why in this chapter we review data-related issues that might affect teams or individuals.

16.1 What counts as data?

Researchers operating in non-electronic environments have to devote much time to producing videos and transcripts. Automatic recording, tracing and archiving make this aspect of CMCL researchers’ work much easier than their non-networked colleagues’, particularly those working with text-based data. Yet the technology cannot currently automatise all the processes. Furthermore, it brings with it additional processual issues. As an example of the first point, consider how audio and video data, once digitally recorded, have to be transcribed, to clear and exacting standards, as do recordings from face-to-face events. The second point relates to the multimodal nature of CMCL data. To capture the communicative actions of participants in modes other than the linguistic, researchers must be able to trace and synchronise new types of data. As McCambridge remarks in her study of language learning by the deaf in multimodal environments, in such projects a conversation may ‘consist
of a picture or webcam picture, text, smileys, animations, sent files, and links' (2006: np).

For example, researchers may want to keep track of which hot buttons and icons users clicked and in which order, or how users built up a shared text or drawing. In a virtual world using avatars, the researchers may want precise logs of how the users move their avatars around the space, and what elements of mimicry (nodding, smiling, blinking, etc.) they made the avatar perform. This not only necessitates the use of screen capture software, but also requires that the screen of each user be captured, as not all users see exactly the same sequence of actions on their screen. The users' posture and body language in front of the computer may also be of interest (see Garcia and Jacobs, 1999, for asynchronous CMC). In this case, again, video is required. If you are planning to use this medium, we recommend Mondada’s (2006) study, which offers much to those filming classroom interactions as well as to the CMCL community.

As alluded to in the previous paragraph, research teams have different data processing (and collecting) priorities depending on their research orientation. This specialisation allows for the otherwise daunting prospects of processing multimodal events to be considerably mitigated, as long as sufficient time has been devoted to agreeing exactly what each researcher wants to focus on and to deciding precisely which traces, interactions and data should be collected (both during and after the project) for each team. Digital tools may allow for easy saving of online events in their raw form, but it is unrealistic and unnecessary to expect to process all the material.

To give an idea of the chain of data treatment that we anticipate a CMCL project would need to include, below is a list from Mondada (2005). She distinguishes between primary and secondary data. Primary data include recordings and documents ancillary to recordings; secondary data comprise everything to do with transcripts and meta-data. The crucial concern is to ensure that synchronisation between primary and secondary information is never lost, no matter how many transformations the data undergo during the research process.

To this we would expect to add a set of documents commenting on, and structuring access to, all of the files above, organising them into a coherent and legible whole which should be able to reconstitute for researchers, in as many ways as desired, information about the original experience.