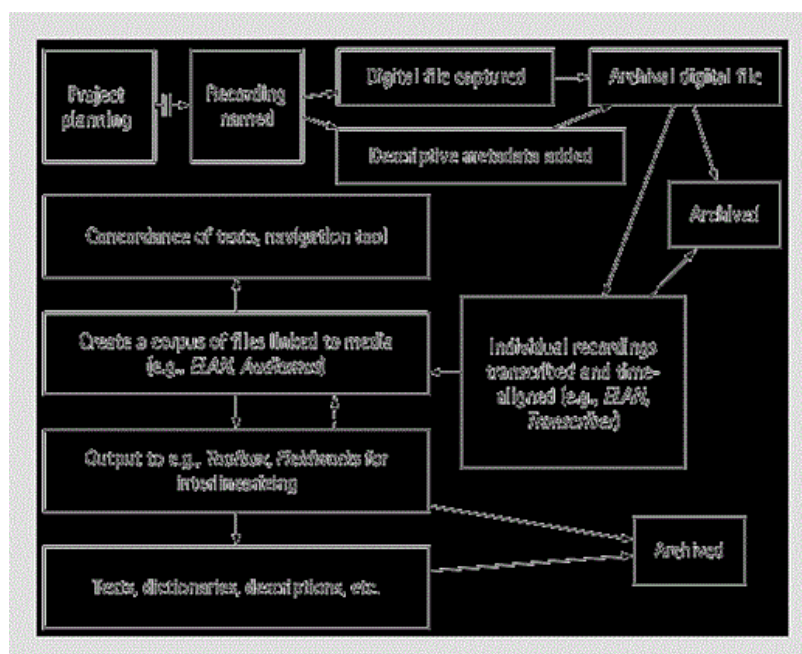


p. 98 become less of a problem over time). A useful way of converting text into new formats is to use regular expression functions as discussed in §4.3.4.

Figure 4.3.



Typical workflow resulting in well-formed linguistic data (adapted from Thieberger 2004).

A final potential problem with workflows is that you need to keep track of which parts of your corpus have undergone which steps, with the risk that you may end up with different versions of the same file on your hard disk. If this happens, you may accidentally edit two different versions of the same file and subsequently have to use a ‘compare documents’ function to keep all of the edits you have made. For this reason it is also helpful to keep *administrative* metadata (which describes your processing of the data) to track items through the steps in your workflow, as well as *descriptive* metadata (descriptions of the data itself) to describe the items.

Care on your part and an awareness of what is lost and what is gained by the use of a particular tool can mitigate all these dangers. Understanding what the various tasks along the workflow involve can make your life easier, and will also determine what tools are applicable at each stage of the workflow.

## 4.2 Advance Planning for Data Management Before You Leave for the Field

While you are planning the logistical details of your upcoming fieldwork (and see Bower 2008: 222 or Fox<sup>10</sup> for lists of things to take and do before departing), one component of your preparation should be planning for the management of the data you will be collecting in the field. Some advance planning before you leave will save time and effort later and will ensure you have a good strategy in place when you arrive.