

Reproducible Pharmacometrics

Justin J. Wilkins
SGS Exprimo NV, Mechelen, Belgium
justin.wilkins@exprimo.com

E. Niclas Jonsson
Pharmetheus AB, Uppsala, Sweden
niclas.jonsson@pharmetheus.com

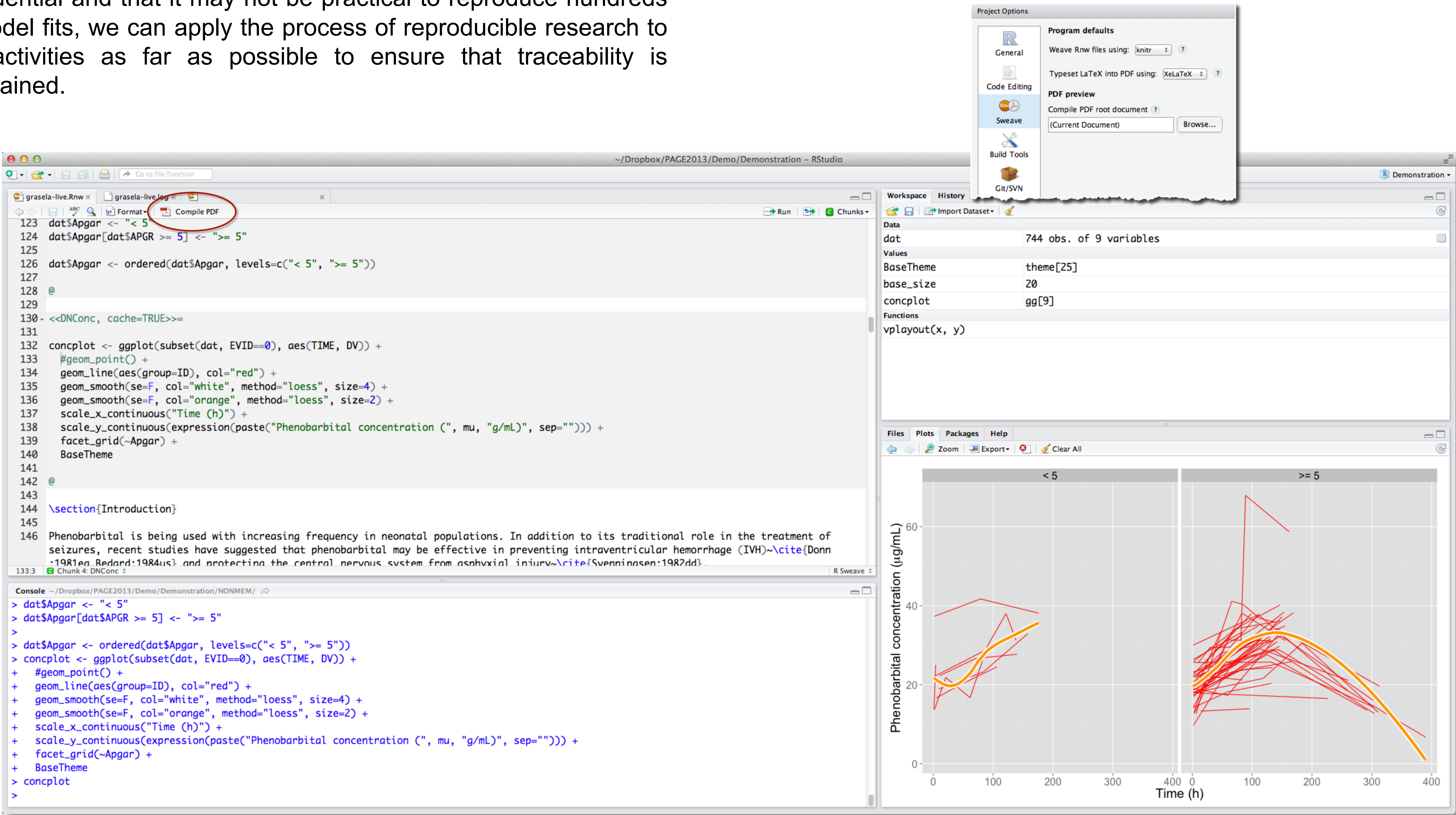
Reproducibility is the cornerstone of scientific research, but is nonetheless a challenging area in pharmacometric data analysis. The large number of intermediate steps required, often involving multiple versions of datasets, combined with a mixture of software tools and the substantial quantity of results that must be tracked and summarized renders traceability an onerous and time-consuming business.

The concept of “reproducible research” is that the final product of scientific research is not just the text of a report or research article, but should also include the full computational environment used to produce the results, including all the associated code and data – and that this bundle of data and scripts should be shared with others who wish to reproduce these results. Although this is not often possible in pharmacometrics, given that data are usually confidential and that it may not be practical to reproduce hundreds of model fits, we can apply the process of reproducible research to our activities as far as possible to ensure that traceability is maintained.

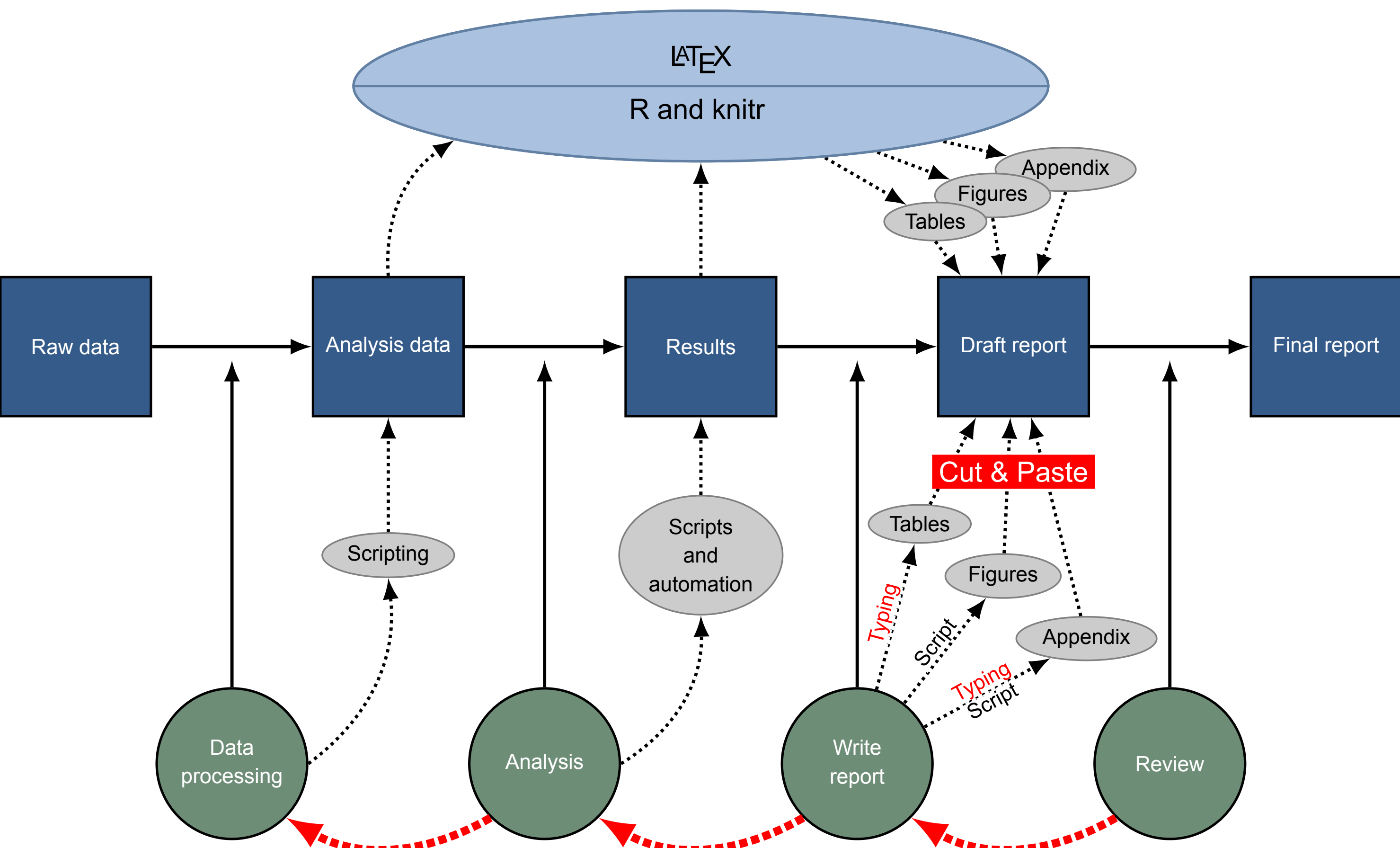
Although there are many approaches that may be taken to adopting this principle, we shall focus on the combination of **R**, **knitr** and **LaTeX**. These tools together enable the end-to-end scripting of data file creation, capture of results from external software tools and subsequent analyses, and can automate the creation of publication-quality reports, articles and slide decks.

We shall demonstrate that applying techniques such as these is not particularly difficult, especially now that they are coming into general use and support from software tools is maturing. We shall discuss the substantial benefits of doing so, which include increased accuracy, efficiency, reliability and credibility, elimination of transcription errors, built-in traceability, and the ability to reproduce an analysis, including article or report, in its entirety years later.

R · RStudio · knitr · LaTeX



Our example uses the open source tools R, RStudio, knitr, and LaTeX, but there are many other tools and approaches, both free and proprietary.



A typical analysis workflow, from raw data to final report. Although the process of dataset generation is often scripted and reproducible, report generation using desktop word processors still overwhelmingly depends on onerous copy-pasting of figures and tabulated results. When issues are discovered late (red dashed arrows), the burden of making changes is considerably greater than it is when literate programming techniques are used.

	L ^A T _E X	R	R	
	miktex.org			
	tug.org/mactex	www.r-project.org	rstudio.com	yihui.name/knitr
	tug.org/texlive			