The Art of Annotation: Digital Micropublications

Michael Cysouw
Ludwig Maximilians University Munich, Germany
cysouw@lmu.de

A major part of research in the humanities consists of the interpretation of extant resources. To facilitate this process, many initiatives are underway using digital technologies to make the resources themselves available to as wide an audience as possible. A central challenge of this process is to convince scholars and other rights holders to allow access to their resources under a permissive license. One central argument to be repeated over and over to scholars is that licensing is independent of scientific recognition: a permissive license does not imply that the resource can suddenly be used outside of the accepted paths of scholarly acknowledgement.

However great the challenge of making the resources available, this process is well-understood and ongoing. In contrast, the development of tools to systematize the *interpretation* of the resources is still in its infancy. Such tools would be especially valuable to fields of research where the resources are not (easily) replicable, either because of high costs associated with their collection or because the resources are historically bound (which is the typical scenario in the humanities, but which also is true for e.g. archaeological artifacts, fossils, or biological specimens). In such cases, the same resources will often be subject to different interpretations by different researchers, or even by the same researcher at different points of time, and consensus can only be achieved by discussing and comparing these interpretations, not by collecting more resources. Normally, the bulk of the content of such interpretations is inaccessible, since only summaries of them make their way into the published scientific literature. Furthermore, there is generally no way for an outsider to discover ways in which a given set of interpretations evolved over time. What is needed are tools that allow such interpretations of resources to be systematically documented and kept track of, so that all of their content, and not just small pieces, can be made available for inspection by other researchers.

The bare embodiment of such interpretations in the humanities are *annotations*, be they handwritten marginalia on a printed sheet, collections of filing cards, or notes kept in word-files or in more sophisticated note-keeping software or databases. Basically, such annotations consists of a selection of specific parts of a resource and an interpretation of this information in the light of any research question at hand. Annotations are mash-ups of text, quotations, figures and tables, and can even include image-snippets of the source itself. Typically, such annotations will find their way into publications in the form of examples, footnotes, or simply as a reference to the source without further explanation. However, such annotations are the real research data of scholarship in the humanities and should be considered on a par with empirical research data in the natural sciences.

In contrast to the common practice of publishing and discussing research results, most scholars (and natural scientists alike) currently do not disclose their underlying research data, i.e. their annotations. They do not make them available to a wider audience because of various reasons, such as:

- failure to see broader applicability of data (“Why would anybody be interested in this?”);
- insufficient quality (e.g. data collection is not finished or not properly cross-checked, or the data is incomplete);
- fear of plagiarism (others might not properly acknowledge the data);
- loss of control over interpretation (others might misunderstand the data, with undeserved blame being cast on the original creator of the data);
- loss of primacy of discovery (others might come up with important discoveries that the original creator of the data also observed, but did not have time to work out and publish);
- lack of suitable publications to publish the data (most publishers are not interested in publishing large tables of raw data);
- lack of technical knowledge about how to make data available;
- limited scientific recognition for making data available.

All these—completely legitimate, though slightly short-sighted—reasons lead to the current situation in which annotations are mostly unavailable for inspection and scientific scrutiny, unavailable for reanalysis, and unavailable for meta-analysis. This is unfortunate, and should be changed. In practice, annotation also remain invisible to other researchers because there is no space to publish them in the traditional publication infrastructure. There is a need for a sustainable (self-)publication infrastructure for such annotations that can subsequently be (re)used as underlying data in structured databases or as examples in traditional publications.

The solution proposed here is to consider each annotation to be a separate small publication, to be called *micropublications*. Such micropublications are texts too small and too numerous to be published in the traditional context of physically printed material, but too valuable to simply be discarded. Recent developments in computational infrastructure (“web 2.0”, “blogs”) have created new possibilities for the exchange of such micropublications. By providing researchers with a simple and lightweight online platform to create, store, publish, and archive annotations (including snippets of the original source) it will become possible to provide an explicit trail of interpretation to which links can be made in regular publications. It will even allow for researchers to refer to annotations made by others, thus reusing earlier compilations of information.

An important new possibility for scientific publishing offered by the (online) electronic format is that publication and quality control can be separated from one another. In a publication system in which each publication is costly, quality control has to precede the physical publication. In contrast, in electronic form the cost of each publication is small (the main costs relate to the maintenance of the overall system, not to the individual item published). This allows for a system in which publication itself (i.e “making available”) can happen independent of the assessment of the quality (“peer review”). Micropublications can thus be published before quality control, and only in case of successful submission to a review process they obtain a seal of approval (or even multiple seals from different review processes).