The Tombstone Protocol: An Undertaking for Unfortunate Events

Alexander Ball

University of Bath

It takes a range of stakeholders and a great deal of effort and coordination to create, develop, and secure the long-term usability of digital content. This fact is rightly used as an argument for the value of digital collections and the credit due to all those involved. But it also follows that there are many points of failure along the way, and things will not always go to plan.

The University of Bath operates a Research Data Archive through which it publishes mainly long-tail data underpinning published research. These datasets are assigned DOIs and we do our utmost to ensure that researchers quote those DOIs in their data access statements and reference lists. In the vast majority of cases, this works exactly as intended.

Sadly, things do not always work out like that. In the early days of the Archive, in order to drive adoption, we would reserve DOIs for researchers as soon as they had provided a basic record. They could quote that DOI in their journal submission and worry about completing the dataset later. A sizeable minority of depositors did not worry about it at all, and so when their papers were published quoting the dataset DOI, there was no dataset for us to put there. In most cases the researchers responded to our requests to complete the deposit process, but by no means all.

It was because of cases like this that we realised we needed a comprehensive and considered set of procedures for dealing with the various ways the deposit process might go wrong. It also seemed like a good opportunity to prepare for the natural end-of-life for our datasets. Thus the Tombstone Protocol was born.

The protocol handles three classes of scenario:

1. The dataset record was started but never deposited, and the associated paper was published without a data access statement.

2. The published paper quotes the DOI, but the dataset is not ready to be published; maybe no data files were uploaded, or data files were uploaded but the record was never deposited, or the dataset is in review and the depositor has not responded to our feedback.

Draft from 10th October 2019

3. The DOI has been minted, but the data files cannot be made available. This might be because the data files are not there at all (e.g. the wrong dataset was published), or because the dataset needs to be withdrawn early, or because it has exceeded its retention period.

The Tombstone Protocol sets out each scenario in turn, and provides both a high-level summary and detailed steps of how to handle it. It includes timescales for attempting to get the deposit back on track, escalation procedures and, if all else fails, the wording for non-availability notices and the visibility settings to use in downstream discovery services.

While the protocol is not invoked often, it was an invaluable tool for clearing our systems of problematic cases, and helps us provide a consistent service in even the most unfortunate of circumstances.