

IMAGES, COPYRIGHT AND BIOLOGICAL DATABASES.

Jeff Christiansen¹, Mags McGeever², Lorna Richardson¹, Shanmugasundaram Venkataraman¹, Richard Baldock¹, Duncan Davidson¹. 1 - EMAGE, MRC Human Genetics Unit, Western General Hospital, Edinburgh, UK, EH4 2XU. 2 - Digital Curation Centre and SCRIPT Centre, University of Edinburgh, Edinburgh, UK, EH8 9LE.

An important source of data for many biological databases is the scientific literature. Most commonly, the data types extracted and standardized are alpha-numerical in format (e.g. nucleotide sequences or text-based descriptions). However, a significant amount of biological data is also image based and at EMAGE, images are central to our curation/annotation process. We collect images from the literature and use these as a source of raw data for a spatial annotation process. Upon deposition of the data in EMAGE, reproduction of the original image in the database entry is desirable as it allows our users to view the original data, even when they do not have access to the source journal.

When images are sourced from the literature, these are subject to copyright. We routinely assess data images from over 150 journals for entry into EMAGE. Of these, <5% publish using a Creative Commons Attribution License, which allows the free use of data in any way as long as the source is properly attributed. For the remaining 95% of journals, we have explored several avenues to allow image reproduction in EMAGE, including use under fair-use / fair-dealing, which has a significant amount of risk involved, and setting up of individual agreements with each journal, which is time consuming for our curation staff.

We will discuss these experiences, and what we have learned about the various meanings of “open access”. We feel it is necessary to raise this issue within the biocuration, publishing and biology communities, so that both curators and our end-users may access the data they require with the minimum of unnecessary effort.