



Project no. 018340

**Project acronym: EDIT**

**Project title: Toward the European Distributed Institute of Taxonomy**

Instrument: Network of Excellence

Thematic Priority: Sub-Priority 1.1.6.3: "Global Change and Ecosystems"

## **C6.09 Report resulting from a workshop on web publication issues on IPR**

Due date of component: Month 26

Actual submission date: Month 26

Start date of project: 01/03/2006

Duration: 5 years

Organisation name of lead contractor for this component: 10 NHML

Revision: final

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
<b>PU</b>	Public	<b>X</b>
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission Services)	
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	

## C6.09 Report resulting from a workshop on web publication issues on IPR

**WP6 Workshop: 20 February 2008, Royal Botanical Gardens, Kew, UK**  
**IPR and the web: challenges for taxonomy**

<http://www.editwebrevisions.info/content/meeting-report>

Workshop Organizers: Dr David Roberts (Natural History Museum), Dr Soraya Villalba (Royal Botanic Gardens, Kew), Dr William Baker (Royal Botanic Gardens, Kew), Dr Jane Smith (Natural History Museum), Melinda Trudgen (Royal Botanic Gardens, Kew), Dr Simon Mayo (Royal Botanic Gardens, Kew).

List of speakers, rapporteurs and participants please see webpage (link above)

### Introduction

Under the paradigm of journal publication, author's intellectual property (IP) has traditionally been respected by citation and copyright has been handled by publishers. Few authors, however, have understood these procedures in detail. Moving taxonomy onto [Web 2.0](#) demands that we clarify these relationships and presents entirely new problems in managing authors' rights. Following the EDIT Forward Scoping Group Report (Taxonomy in Europe in the 21st Century) we expect that taxonomy will move from an "artisanal" to an "industrial" scale of production, which means that many individuals will deliver components to be assembled into a finished treatment.

This context lends urgency to finding a solution for these problems and the workshop aims to identify those issues that may inhibit the progress of using web tools to facilitate collaborative working, i.e. the potential Intellectual Property (IP) obstacles to achieving EDIT's goals.

### Using existing data

Traditional taxonomic procedures have produced more than 90 million printed pages of information covering descriptions and the relationships between organisms. The rate of production has followed a power law, so by far the greater part of this is covered by copyright and only a tiny fraction is electronically accessible. There is no central index to this information and it is dispersed into libraries across the world, leading to a generally very poor level of accessibility and contributing directly to the 'taxonomic impediment'.

These data, descriptions and statements of relationship, are factual and are not, therefore, 'original work' in the sense of copyright so there is no legal barrier to the work being abstracted for re-presentation. Copyright in this case only prohibits the reproduction of the work 'as it is'. It is an almost inviolate convention that the original author be credited with the description and relationship statements or parts thereof.

Opinion differs on the extent to which copyright is a major barrier, according to national culture and legal systems. In general some level of risk assessment is necessary and a sensible way forward for EDIT is to formulate guidelines that taxonomists and their institutions can use to implement standard procedures.

Printed publications are structured to be read by humans not machines, so technical barriers exist to solving the access problem.

[XML schemas](#), such as [TaxonX](#) and [TaXMLit](#), can be applied to human-readable text to indicate to computers which data element is which, for instance what is a taxon name, what is a

taxonomic authority and what is a description. Such mark-up makes the text available to computer searching and retrieval. Tools to add these marks to the text are being made available, such as the [GoldenGATE project](#).

There is no reasonable doubt that large homogeneous data sets are of greater value in modern science than small heterogeneous sets and these tools present the potential to create a very large, searchable information source directly from the existing literature. Agosti pointed out that an NLM/taxonx module is in production, and PLOS-One, BMC and Zootaxa are working on implementing taxonomic specific mark-up into their production XML. Technical constraints mean that taxon-xml might be added as additional material and not integral to the publication.

Technical barriers are the least likely to be problematic in the medium and long term since there is so much innovation under way in the new collaborative Web 2.0 environment. Taxonomic institutions rather than individual taxonomists will have to bear the burden of adopting and maintaining technical infrastructures to make a global "Taxonomic Grid" feasible. This is a political and financial problem which a consortium like EDIT could help to resolve through concerted action, including major grant proposals.

### **"If you can't cite it, it's not science"**

Credit is used here to mean acknowledging the work someone has put into a resource, specifically here a piece of taxonomic information. The conventional way to give credit is by citation of the original work in subsequent derivative works. Traditionally both the original work and the derivative work would be published on paper and the citation index has been adopted as a convenient metric of credit, both for the author and their employer.

An immediate consequence of citing a web page is that the cited version should not change. In a dynamic web site designed to reflect current opinions, as envisaged in the EDIT project, this means that pages need to be archived so that following the citation leads the reader back to the same page that the person quoting the citation actually saw.

The clear identification of data contributors is a statement of responsibility and lends authority to a data set or taxonomic statement and it generates kudos for the authors. We wish to measure kudos because it provides evidence for status and can be used for career advancement.

The measurement of kudos is related to the level of use to which a datum is put. Use can be measured in the number of times a page is read, the number of times it is downloaded, the number of times it is cited, (both in traditional publishing and in other web sites) and the number of times it is used in other derivative works. All of these measures are technically straightforward but are contingent on common standards for making the attribution information accessible to subsequent users and the general adoption of systems to allow "deep citation", as described by Chavan, which would attach [metadata](#) to a datum accessible no matter how many times it has been copied and re-used. The image format [Exif](#) is an example that can be used as a means of storing machine-readable copyright information and permissions to images mounted on the web. An obvious candidate to hold and resolve such metadata would be [LSIDs](#).

A number of practical issues were raised that need resolution:

- how do we build a citation for a web page?
- how do we manage attribution on a dynamic webpage?
- how do we quantify and demonstrate credit? Page read? Page cited or linked? Citation downloaded?
- how do we apportion credit? Do all contributors deserve equal credit? Who decides whose name goes on the author list? Can this be done automatically by software? Should we credit pages or elements?
- Can we build it into a simple statistic?

### **Licensing and inhibitors to the release of data**

As a matter of principle, EDIT databases are managed repositories and do not own the rights to the data they contain. The ownership of the data remains with the data creator but such data should have a clear licence statement specifying the manner of use to which they may be put.

A clear data use policy is highly desirable to give confidence to data contributors. It is not yet clear how such policy apply to globally available information in the context of national legislation and therefore what level of protection licences really provide.

It is not at all clear how we should manage differently licensed materials within a single web page. If the web page is dynamic, then the licence should be accessible to the software building the page. For internal resources we need to determine whether there are any real conflicts and to determine the desired outcomes for both contributors and their employers. These issues are newly emerging and few people are in a position to engage in a policy discussion as yet. It will be of interest to find out whether a conflict exists between the desires of contributors and their employers. For external content, it is simply not feasible to expect the software engineer to have sought permission to use all resources that a dynamic page might recover.

There are different contractual issues relating to institutional contributions and potentially third parties, such as funding bodies. How is this represented downstream: ownership? control of use? attribution?

We look forward to receiving requests for commercial use of taxonomic data and having to deal with royalty payments! Current practice reserves all commercial rights for the data originator although we are not aware of a widespread mechanism to tag data elements with a rights statement.

One of the fundamental arguments about publishing on paper is long-term persistence. As we move to a web-based environment the question of long-term accessibility will need to be addressed. This may be a problem unique to taxonomy which has a need to find and use information from the time of Linneaus onwards.

### **How to use existing copyright material**

The principles and priorities of the scientific producers are that credit for authors and originators of data is much more important in science than copyright. Communication of scientific results should prevail over commercial considerations.

Practical advice for webmasters and contributors in building taxonomic sites:

- There is no infringement involved if authors copy and paste their own work and republish in a different form;
- Using descriptions extracted from published material is unlikely to provoke negative reactions, but there are copyright issues in doing this nevertheless. The best policy is to assess the risks and then negotiate with publishers;
- Webmasters need to react quickly to complaints about infringement of copyright, i.e. remove offending material from web pages if actual infringement has taken place.
- Different countries have different laws on rights. There is no single sure-fire solution. Best policy is to carry out risk management and implement procedures which diminish risk of infringement;
- Publication on the web could increase publisher's sales and provide incentive to grant permission.

### **Social changes**

There are a number of changes that speakers forecast would either take place or would increase in importance. It would be highly desirable for the [Codes](#) to allow e-publications. Social networks are becoming widely accepted and examples are beginning to appear in taxonomy, such as the

[Scratchpads](#). Most importantly, taxonomists need to be prepared to share their data in the way that happens in other areas, for instance physics and molecular biology.

The key social barrier is recognition for scientific work made available publicly. Technically, developing new crediting systems poses few problems, but the barrier is the current publication paradigm which is based on hypothesis-driven arguments and a particular, strictly defined citation system. Clear principles and methodologies regarding recognised publication of scientific data through the internet must be enunciated and accepted by the scientific community. Developing a workable peer review system for web-delivered data and publications is an indispensable element. EDIT will prepare prototype copyright policies for institutions which can be modified to suit local requirements.

Individually, taxonomists can:

- make sure that all you do is open access;
- understand copyright, do not be afraid of it and do not ignore it;
- self-archive ([the Green Road](#));
- don't sign any contracts giving away rights;
- urge your institution to adopt and build a repository for your research;
- talk to your scientific societies and museum to adopt a policy at least to allow self archiving;
- demonstrate the power of access through innovative research projects and data.

The practice of academic research will be the only motivation to change law and build up infrastructure.

## Conclusions

This workshop initiated a debate that could not be resolved within a single day. It is clear that the management of taxonomic data within current rights framework is complicated and uncertain. Much as taxonomists would like clear and simple guidance of how to stay within the rules of copyright, this is not possible largely because the copyright framework evolved before the digital revolution and is not keeping pace with current capability and practice.

The reality of the situation is that copyright management is a managerial judgement call. The majority of copyrighted taxonomic material is out of print and therefore not available for sale, so the amount of commercial damage that might be caused by extracting descriptive and nomenclatural information for re-structuring into a web page is negligible. There is, however, at least in the UK, a climate of fear that a breach of rights legislation may have dire consequences, possibly engendered by the strategy adopted by the music industry and the rules around photocopying enforced in libraries. It is not easy to see how this has come about because taxonomists are doing now what they have always done, but using tools that allow them to do it on a rather larger scale and publishing the outcome outside the commercial publishers.

The barriers to moving taxonomy to the web are now seen to be primarily sociological rather than technical. The crucial change that needs to be made is for employers of taxonomists to give career value to those contributing content to taxonomic web services. In this vein, grant-awarding bodies need to support data gathering with greater enthusiasm than they have already shown for infrastructure proposals.

To read Donat Agosti's blog on this meeting, [click here](#)