Taxonomy in Europe in the 21st century

Report to the Board of Directors
European Distributed Institute of Taxonomy

Chair: Richard Lane, the Natural History Museum, London

Table of contents

Executive Summary

This report was commissioned by the Board of Directors of EDIT (the European Distributed Institute of Taxonomy) in February 2007 as part of their process to develop a mission for EDIT. Richard Lane of the Natural History Museum, London chaired the report.

To facilitate development of the report, a two-day meeting of invited participants to scope the possible environments in which taxonomy could be operating in 10 or more years time was held in Oxford, UK in December 2007 (see appendix). Participants were familiar with different areas of biodiversity research, developing technologies and applications or social-political environments. Those not able to attend were interviewed beforehand and their views incorporated into discussions.

The overall conclusions of the report are:

- That taxonomy faces exciting challenges and opportunities in the future to meet the demand for an ever more profound understanding of the diversity of life on this planet, how it developed and the impact of increasingly destructive human activity including climate change, factors that are predicted to have an enormous negative influence on the diversity and distribution of biodiversity (the biodiversity crisis)
- Pivotal to the development of taxonomy are the rapidly expanding fields of high throughput DNA sequencing, automated digital data-gathering and biodiversity informatics. Incorporating these technologies will be critical to the science of taxonomy.
- Scientific collaborators and users of taxonomy will require new ways of working and interacting with taxonomists. It is essential that taxonomists and their users respond to this need. Taxonomists integrated into interdisciplinary teams will be an essential way of working.
- Although an ever expanding repertoire of theoretical and practical tools is available to taxonomists, unheralded in the history of the subject, there will have to be substantial, even radical, changes in how taxonomy is done and its supporting infrastructure operated, to exploit these opportunities to the full. “Business as usual”, even if scaled up, is simply not an option.

Report

Three main areas are considered here, the science of taxonomy, technological developments and the socio-political environment in the medium term (i.e. 10+ years) which are addressed as a series of questions.

What will taxonomy look like as a subject?

1. Taxonomy as a discipline will underpin efforts to produce a credible tree of life. In this, taxonomic expertise will be essential in defining and parameterising the mode and tempo of evolutionary change in organisms and the processes of speciation. These strengths will remain at the core of the subject, enhanced rather than replaced by new developments.
2. Much previously unknown biodiversity will be found in the micro-world of invisible and barely-visible organisms. Emerging molecular tools will make discoveries possible at a remarkable scale and speed.
3. Incorporating taxonomic outputs into forecasting and modelling the impact of environmental change on biodiversity will be a growth area. It will require integrating knowledge of natural ecosystems with human systems and social impacts.
4. The role of taxonomy as an information science will increase greatly, most likely as a primarily web-based science. Tools for the management of information will be central to taxonomic work. This is likely to take two forms:
   - The web will serve as a clearing-house in which vast quantities of information, currently fragmented in innumerable and badly-mapped locations, can be marshalled.
   - Web-based tools will enable a transformation in the scale of data analysis so that huge data sets can be analysed for patterns. Well verified taxonomic data, professionally audited, will be fundamental to this macroscope approach.
5. Specimens, or vouchers, will remain a critical part of taxonomic science as hypotheses are based on them, but our concept of 'specimens' will include surrogates such as digital images, bioacoustic data and molecular sequences.
6. Current approaches to taxon description will need to be radically reviewed. The current approach is inadequate to meet needs so simply ramping up productivity using existing nomenclatural and publication tools will not suffice. Formal description might only be used in taxa or instances where a formal name is essential. Emerging biodiversity informatics techniques can associate different kinds of information with unique identifiers that do not require a formal name. These changes need to be led from within taxonomy.
7. An ecosystem approach based on metagenomics will need to be incorporated into the portfolio of approaches to describing life on earth, particularly for the micro world.
8. Organisationally, parts of taxonomy will move from an 'artisanal' to an 'industrial' scale: massive sequencing and other data capture, massive output and analysis and high throughput identifications. This will increasingly require working in integrated teams rather than individually. Concomitantly, a change in the systems for the assignment and evaluation of scientific products is required.
9. Sociologically, the taxonomic community is fragmented and has not built a sense of cohesion and common purpose that exploit the excitement, vitality and dynamism of the subject. The community needs to focus on achievements and articulate clearly defined "big science" projects which could be achieved with specified new resources. The challenges here include prioritisation and building consortia embracing taxonomists and other biologists.

What facilities will we use?

1. Cheap and fast, high-throughput DNA sequencing will lead to the industrial scale production of data, needing analysis with a scaled-up biodiversity informatics capability. Costs are falling as capacity is increasing.
2. Biodiversity informatics capability will need to be strengthened considerably to handle different types of data. Existing developments in computer storage and computation capacity can cope with the expected industrial scale acquisition of taxonomic data but facilities will need to be expanded. Open access is a critical principle.
    a. Taxonomic information must be in the public domain and machine-readable to remain productive and relevant. The considerable volume of heritage data must be transformed into digital form using emerging technologies.
    b. Standardised ontologies will be necessary for linking information. This is a necessity not only for the management of vast amounts of data, but also for the meaningful and objective treatment of scientific judgments.
3. A unique identifier system will underpin different types of data; different naming systems will coexist.
4. Moving taxonomies to the web will be critical (distinct from just digitising data), but will require real investment in taxonomists.
5. Collecting will remain an essential component of taxonomy, though emerging technologies (e.g. digital imaging, bioacoustics) might automate collection of 'specimens' and make field trips more productive. There is a pressing need for 'pocket identifiers' (digital, molecular, bioacoustic) for use in the field.
6. New storage facilities that relate more to tissues and microbial diversity than whole organisms (likely to be frozen but new technologies, such as polymers, may obviate this)
7. Georeferencing specimens will become central to their use, to enable data input to distribution mapping and subsequent modelling. Open access sources of geographical data will need to be developed with and agreed structure so that collection data from diverse sources can be viewed as a whole.
8. Most large publications will be interactive on the web, and in real time.

What expertise will we need?

1. With many new practical and theoretical tools becoming available, taxonomists will increasingly need to be broadly based and to keep abreast of technological developments. The ability to integrate different types of data will be critical as will flexibility, such as transferring skills from one taxon to another. Taxonomists will increase their capacity to interact both within and outside taxonomy, and ready to be hired in cognate fields (ecology, conservation, etc.).
2. As is the case for many areas of science, the future availability of expertise will be shaped by available funding and an ability to contribute to the major scientific issues of the day. Currently the number of full-time taxonomists is declining, as universities and other organisations continue to reduce positions.
3. There will be a continuing need to train students in taxonomy, focusing on exploring and understanding biodiversity. The training of taxonomists is likely to be undertaken by specialised institutions, but with this comes the risk of a narrow perspective. It will be essential to engage the emerging cadres of students produced by bioinformatics, computer science, and ecological training schemes with the major challenges and opportunities of taxonomy.

4. With the development of new cross-cutting technologies is important that taxon-based expertise should not be lost. The taxonomic contribution to research may increasingly focus on natural history, that is, taxon-based integrative knowledge. While internet based tools will correlate and organise information, human editorial control will be necessary and this is likely to be a key role for tomorrow's taxonomists.
5. The relative importance of amateurs or citizen scientists will increase and the challenge for the professional community is to provide opportunities for their effective contribution.
6. Technology problems can be solved (relatively) easily; sociological problems are more difficult and should receive serious attention.

Who will be using taxonomy/systematics and how?

1. The community of users and collaborators will broaden provided taxonomic information and new tools are made more readily available Users will increasingly include developing countries. Customisation of taxonomic information will be essential. However, the way in which information is provided may well change, using a just-in-time approach, on demand from users, rather than the current just-in-case approach, determined by suppliers. The maintenance of up-to-date information, and therefore rapid response to new information and mediated taxonomic judgment, will be vital
2. The needs of users, especially field-based rapid biodiversity assessments, will be a major driving force in the development of the subject
3. Compliance with an increasingly complex framework of international laws and regulations will be essential to ensure continued access to biodiversity samples as well as to collaborators. Information Technology will track source and use of samples.
4. Virtual research environments and collections will increasingly enable the public use of taxonomy as human populations become more urbanised and may be their primary contact with Nature.
5. The user base will access taxonomic information through the internet, requiring online diagnostic services based on morphological data (automated analyses of digital images), sequence data (automated assignment of specimens to taxa via sequence) and other modes (e.g. chemistry, bioacoustics).
6. More political, ie user driven, selection of research priorities will become increasingly necessary with the consequence that taxonomists will be integrated further into different disciplines, genomics, conservation, ecology, etc., in order to realise their potential. Inventorying will require teamwork between institutions.

How will we manage taxonomy in Europe?

1. The key is continued integration to provide critical mass and delivery of ambitious projects. It will be critical for the large museums and herbaria, the major long-term players, to clearly articulate how they will integrate collections, biodiversity informatics provision and research capability.
2. Internationally, taxonomic efforts will be increasingly organised through globally relevant bodies such as the Global Biodiversity Information Facility, the Consortium for the Barcoding of Life, the Tree of Life and the Encyclopaedia of Life, and similar large infrastructures.
3. Within Europe, a broadly based and forward looking coordinating mechanism, will be important, perhaps centred around a revised CETAF (Consortium for Taxonomic Facilities) incorporating national taxonomical societies and associated users.

4. If taxonomic facilities follow the pattern of major DNA sequencing centres, resources for data extraction, storage and analysis will be concentrated in a few regional centres. Public support of these centres will be necessary. Similarly, rationalisation of collections to provide critical mass for addressing large-scale programmes will need to be considered. The criteria for continued collection development and management should be based on gap analysis, risk assessment and future scoping of usage.
5. Information management systems will be critical in driving the management and evaluation of taxonomy within Europe. Disparate efforts for digital taxonomic information infrastructure will need to be harmonised: the priority will turn to standardisation over development. An open-source approach will enhance this development.
6. To measure the impact of taxonomy, new evaluation metrics will be developed that more effectively recognise taxonomic contributions to knowledge and world science. It will be essential that the intellectual content of synthetic taxonomic publications, especially floras, faunas and revisions, is recognised and hence are more valued.
7. Europe has a rich amateur sector that needs encouragement and further integration to exisiting and developing infrastructures. This may be a major challenge.
8. The training of new taxonomists within Europe could benefit from a distributed model but again will have to be integrated. The role of Universities in both teaching and research cannot be underestimated for introducing new blood as well as integrating taxonomy into other subjects. These roles cannot be left to specialised institutions alone.

Appendix: Participants

Michael Ashburner, University of Cambridge, UK
Jesse H. Ausubel, Program for the Human Environment, Rockefeller University, NY, USA
Mark Blaxter, University of Edinburgh, UK
Mark Costello, Leigh Marine Laboratory, University of Auckland, New Zealand
Benoît Dayrat, University of California, Merced, California, USA
Michael Donoghue, Peabody Museum of Natural History, Yale University, USA
James Edwards, Encyclopaedia of Life, c/o National Museum of Natural History, Washington, USA
Leonard Hirsch, Smithsonian Institution, USA
Line Le Gall, Muséum National d'Histoire Naturelle, Paris, France
Charles Godfray, Department of Zoology, University of Oxford, UK
Ken Johnson, Dept of Palaeontology, Natural History Museum, London, UK
Sandy Knapp, Dept of Botany, Natural History Museum, London, UK
Len Krishtalka, Biodiversity Institute, Dept of Ecology and Evolutionary Biology, University of Kansas, USA.
Matjaz Kuntner, Jovan Hadži Institute of Biology, Slovenia
Robert May, Department of Zoology, University of Oxford, UK
Jeffrey McNeely, IUCN, Gland, Switzerland
David Remsen, Global Biodiversity Information Facility, Copenhagen, Denmark
Richard Smith, BIONET International, c/o CABI, Wallingford, Oxfs., UK
Simon Tillier, Muséum National d'Histoire Naturelle, Paris, France
Wolfgang Wägele, Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany

Rapporteur: Gaël Lancelot, EDIT Communication Officer, Muséum National d'Histoire Naturelle, Paris, France