2 ENHSIN in the context of the evolving global biological collections information system

Walter G. Berendsohn

This article provides a review of international networking efforts and the role of the European Natural History Specimen Information Network in that context. *ENHSIN* is one of several international initiatives that have emerged over the past decade and that are concerned with electronic access to the information held in biological collections. Taken together, these initiatives prepare the ground for a global collection information system providing data access and interoperability. To provide an effective service, a consensus about technical protocols and semantic data definitions must be reached. *ENHSIN* has contributed to this aim by prototyping data access to distributed non-homogeneous databases via XML and HTTP and by introducing the principle of variable atomisation into the ongoing efforts for a collection data standard.

Introduction

Biological collections and their role

Biological collections in a wider sense encompass living collections such as botanical or zoological gardens and culture collections, as well as the preserved specimen collections in museums, universities, and ecological research facilities (see Berendsohn, 2000, for a justification of this wide definition of biological collections).

Biological collections provide the basis for research in biological systematics, but their relevance is not restricted to systematics. For many sub-disciplines of biology the only way to ensure the possibility for scientific scrutiny of research results is by means of deposited conserved specimens (see Huber, 1998, for a summary of the importance of voucher specimens). In addition, living collections in particular represent a considerable reservoir of genetic resources and provide raw material for applied genetic and other biotechnological research. Collections have an even broader role within the emerging global system of biodiversity information. The specimens themselves as well as associated data (labels, field ledgers, publications) carry a significant amount of primary information on the structure, composition, geographical distribution, and lifestyle of individual organisms as well as of the ecosystems to which they belonged. Of particular importance is the temporal dimension: by combining data from hundreds of millions of specimens accumulated over the past centuries we can achieve a solid, vouchered (and thus falsifiable) picture of changes in ecosystem composition for many geographical regions. Collections thus have the potential to make a significant contribution to the solution of problems resulting from global change, in particular towards the conservation and sustained use of natural biodiversity within a global framework (Berendsohn, Häuser, & Lampe, 2000).

Recognising this role of their holdings and the need to collaborate in order to realize the full potential of collections, the members of the Consortium of European Taxonomic Facilities initiated the *ENHSIN* project. Providing access to a broad audience via the Internet is a key component of their plan. As Maurer, Firestone & Scriver (2000) put it: 'Society cannot get full value for its investment in science unless anyone desiring existing data actually gets them'.

The emerging collection information system

Governments and funding agencies are also increasingly taking notice of the general importance of collections as a primary source of biodiversity information. The most significant event to-date in that respect was the decision taken by OECD science ministers (OECD, 1999) to give the go-ahead for the Global Biodiversity Information Facility (*GBIF*), based on the proposal elaborated by the Working Group on Biodiversity Informatics of the OECD Megascience Forum (Anon, 1999). *GBIF* recognises natural history collections information as one of its four central themes, the other three being a catalogue of names of organisms, outreach and capacity building, and interoperability and networking. For the first time in history, *GBIF* will provide a government-supported framework for biodiversity information on a global scale.

Early international efforts to network collection information are clearly concentrated in the area of living collections and *in situ* recording of occurrence records. A few examples:

- The International Species Information System (ISIS) for **zoological gardens** and comparable institutions is a truly global network and comprises a computer-based information system for wild animal species held in captivity. Since its foundation in 1973, ISIS has grown to a membership of 576 institutions in 54 countries. The ISIS central database currently holds information on 1.44 million individual animals representing nearly 8,000 species. Information is collected almost exclusively by using a common software, the ISIS Animal Records Keeping System (arks) for institutional animal records of isis members (ISIS, 2002).

- In the realm of **botanical gardens** (about 1,600 institutions worldwide), Botanic Gardens Conservation International (*BGCI*, 2002) has developed a computer database on the rare plants in over 300 institutions to bring worldwide co-ordination to the individual efforts of each garden. This effort started in the mid-1980s and has, for example, spawned the first community-recognised standard, the International Transfer Format for Botanical Garden Records or *ITF* (*BGCS*, 1987). However, international associations of botanical gardens have until recently not seen information exchange as a priority and have lacked the coherence displayed by their zoological counterparts. As a result, where they exist, information networks for botanical garden accessions are still mostly national in scope, and that in spite of the early existence of a suitable software (BGBase, Walter, 1989; Walter & O'Neil, 2000).
- Data collections of floristic and faunistic mapping records and other **organism observation records** (bird ringing, migratory animals, etc.), although restricted to national and regional scope, represent huge data collections, often based on several decades of recording. For example, the German Database of Vascular Plants (BfN, 2001) contains more than 13 million records of high-quality grid-based observations (R. May, pers. comm.).
- The *CABRI* (Common Access to Biological Resources and Information) system allows access to collection catalogues of quality-controlled **biotechnological resources** covering animal and human cell lines, bacteria and archaea, fungi and yeasts, plasmids, phages, DNA probes, plant cells and viruses. The eight European centres currently participating encompass 26 collections holding nearly 140,000 deposits (90,000 of which are currently included in the catalogues), which represent approximately half of the deposits available in Europe. The system was set up as a eu-funded project and is now maintained by the participants (*CABRI*, 2002).
- In the field of **natural history collections**, international networking was comparatively weak until the end of the 1990s, in spite of a high level of knowledge about data structures and several existing standard formats for data exchange. A number of reasons account for this situation: (i) access to natural history collection information traditionally relies on physical examination of sorted physical objects, so that computerisation of specimens was not a priority; (ii) the sheer size of the task of digitisation – billions of specimens – impeded its execution; (iii) the global scope: lack of easily accessible authority files, especially for taxa and for geographic areas adds considerably to the

cost of data capture; (iv) data capture by means of text input often results in datasets of low quality, especially if done by non-scientific personnel; and (v) lack of suitable software, i.e. programs that are able to cope with the complexity of specimen information (cf. Berendsohn & Nimis, 2000) and at the same time offer an easy-to-use interface for data capture.

However, in some countries funding agencies, local or state governments, or individual institutions recognised the importance of the information contained in natural history collections and proceeded to collect or network data on specimens pertaining to their area of interest. Many such efforts exist, and it is obvious that persistent efforts are now starting to pay off. Examples of such unit-level networks are noted below by region:

- South Africa. The precis database was started in the early 1970s and is now used by the majority of herbaria in the region (57%) as well as being endorsed by the regional Botanical Diversity Network (*SABONET*, 2002).

- Mesoamerica. The Mexican Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (*CONABIO*) uses its BIOTICA software and now collaborates with the Costa Rican Instituto Nacional de Biodiversidad (*INBIO*) to form a specimen access network (*CONABIO*, 2002a).

- Australia. The Australian Biological Resources Study network (*ABRS*, 2002) presents an impressive range of activities, based on earlier networking efforts started in the 1980s (Environmental Resources Information Network, *ERIN*). Funding of specimen-capture software forms an integral part of the effort. At least for botanical collections, a decision has been taken to establish a complete inventory of specimens held in the nation's herbaria, the Australian Virtual Herbarium (*ABRS*, 2001).

- Usa Collection digitisation has been promoted through grant programmes of the National Science Foundation, either directly for specimen data capture (see NSF, 1998), or indirectly by supporting software development such as Biota (Colwell, 2002), SPECIFY (Anon., 2002), and various institutional databases as well as networking efforts, e.g. Species Analyst (Anon, 2002a).

- Europe. Several examples exist on the institutional or national level, but in spite of the immense global significance of European natural history collections, research funding agencies have been reluctant to support the digitisation of specimens up to the present day. The EU has started backing collection networking as part of its support for the construction of a European research infrastructure. Underlining the need for improving communication between European collections, including issues of digitisation, has been the foundation of the

Consortium of European Taxonomic Facilities (CETAF), a body that has served to anchor the *ENHSIN* initiative. *ENHSIN* partners restricted the project's scope to the level of specimen (unit) information, and to natural history collections. These restrictions were agreed in view of the existing eu-funded *BioCISE* project, which focused on metadata about collections in Europe (*BioCISE*, 1999). *ENHSIN* and *BioCISE* partners collaborated to establish a Biological Collections Access System for Europe (*BioCASE*), a project that commenced in January 2002. *BioCASE* is attempting to unite collection-level and unit-level (i.e., specimen-based or observation-based) information in a common access system (Berendsohn Häuser & Lampe, 2000).

Standardising collection access

Once the decision to share data is taken, the fundamental prerequisites for interoperability are compatibility of data definitions and a common access protocol.

Standardised data access protocol

Protocol development. With regard to communication and access protocols, the Internet and World Wide Web provide a solid base. The present consensus is that any system to be created in the biological domain should be based on established Internet and W3-standards, particularly hypertext transfer protocol (HTTP) and extensible markup language (XML). The decision to develop the *ENHSIN* prototype (Güntsch, this volume) exactly along these lines was taken after evaluating the Species Analyst prototype (Vieglais, 1999), which at that time used the Z39.50 protocol.

The core purpose of *ENHSIN* is to construct a demonstrator access system to natural history collections throughout Europe, which implies the eventual incorporation of at least several hundred collection databases into the network. This is a considerable technological challenge because the existing prototypes cannot handle a network of this size. Moreover, to truly realise the envisioned role of biological collections it is necessary to significantly widen the scope of data points that can be queried, or at least the number of defined data items the network can return to the user. XML schemas are used to standardise the returned data, and schemas developed by other communities will be incorporated. The data access protocol should be independent of the schema underlying the returned data. A standard query format, preferably also in XML, has to be developed. Considerable standardisation and development efforts are necessary to reach these goals.

Steps towards developing a data access protocol based on HTTP and XML are taken by the *CODATA/TDWG* (Committee On Data for Science and Technology/Taxonomic Database Working Group) working group on access to biological collection data (*ABCD*, 2002), and particularly the distributed generic information retrieval project (*DIGIR*, 2002). In the European context, both the *BioCASE* project and *ENBI* (European Network for Biodiversity Information, which at the time of writing is in contract negotiation) are pursuing similar aims. It is also expected that the Global Biodiversity Information Facility (*GBIF*) will play a substantial role in fomenting and accelerating the protocol standardisation process.

Standardisation process. Apart from providing the technological base for networking and interoperability, the World Wide Web also provides a model for a process of supported, but in essence non-governmental, self-organisation and standardisation. It is expected that the Global Biodiversity Information Facility, *GBIF*, will join forces with community initiatives such as tdwg and ongoing networking projects to facilitate this process. The development of a protocol for data access to heterogeneous distributed information resources is a field of wide potential application, so that it makes sense to collaborate with, for example, the library community. The responsibility for the development of standardised data definitions, however, clearly rests with the biological scientific community.

Data definition

Information models. A prerequisite for standardising data definitions is a thorough knowledge of information structures in the subject area. During the 1990s, much effort was dedicated to the task of information modelling. Many models of implemented systems were published (see *TDWG* reference list, Berendsohn, 1999), but rather few research models exist. Research or 'reference models' look at the structure of the information itself, without being bound by specific implementation restrictions and deadlines. The us Association of Systematics Collections held a workshop in 1992, which produced a core model (ASC, 1993). The ASC Model has been adopted at least in part by many database systems in natural history collections. The transaction model of the Smithsonian Institution's National Museum of Natural History (CRIS, Anon, 1994) was based on an extensive review of legal context and best practices in museum collection management and has been used in several implementations and information models. In Europe, two EU-financed Concerted Action projects helped to produce and publish a general reference model for biological collections (Berendsohn *et al*., 1999). For observation data, the British Recorder Model (Copp, 2000) is essential reading. For botanical taxa the *IOPI* Model was developed (Berendsohn, 1997). It has been further developed in the course of the Euro+Med project (Güntsch, Li & Berendsohn, 2002) and could easily be adapted to cover zoological taxa.

These modelling efforts have been very extensive and thorough, and, as a result, the information domain of collections and taxa should be considered as well understood.

Data standards and standard data, past efforts. The Taxonomic Databases Working Group (*TDWG*), a voluntary community organisation established as a commission under the International Union of Biological Sciences, was created in 1985 to foment standardisation efforts in (original botanical) databases. Annual meetings drove the process, and working groups were formed to tackle different problem domains. Several data-exchange standards were developed or adopted, among them that for botanical gardens (*ITF*, Wyse Jackson, 1997), for botanical names (Bisby, 1995), and for herbarium collections (*HISPID*, Conn, 1996). Some of the 'standards' adopted by *TDWG* are in reality standard data catalogues (authority files), which had been developed as community efforts or within separate projects. This includes a biogeographical scheme (Hollis & Brummitt, 1992) and standard abbreviations for taxon authors (Brummitt & Powell, 1992) and for bibliographic title citations (e.g. Bridson & Smith, 1991).

The process of standardisation was hampered by the very nature of tdwg as a community organisation. With its changing membership it essentially relied on voluntary efforts and the willingness of its members to attribute their efforts to the group. However, that the task of defining a data dictionary for biological collections is indeed a complicated one, is proven by the drawn out discussions that took place in existing networks and within virtually all institutions developing their collection information system. With the possibility of large-scale networking, this discussion has achieved a new urgency, so that finally some resources are made available to support standardisation efforts to achieve compatibility between different data sources. *ENHSIN* has been one of the projects contributing to this aim. *TDWG* is currently taking steps to revise its procedures and adapt to its newly won recognition by international organisations as an appropriate body for the standardisation process.

COMPATIBILITY OF DATA DEFINITIONS. This can be separated into structural compatibility and semantic compatibility. Traditionally, semantics and structure have been combined, sometimes with elements of protocol (e.g. *HISPID*). Exchange standards were used at first primarily to extract data from existing datasets or to specifically prepare data for a central system (e.g. *ITF*).

To achieve *semantic compatibility*, i.e. the adherence to a definition of the contents of a data element, we need to describe and define data elements for the entire domain. Fortunately, much of this work has already been done so that the remaining work consists of collecting the existing definitions and feeding them into a community-wide discussion process.

Over the past two decades it became clearly apparent that full *structural compatibility* (i.e. use of the same fields and relations, the same database schema) is impossible to achieve in all but a few exceptions. Too many systems are already in existence and databases may vary widely in the coverage of the domain and thus actually need very different structures. An exception to this general position exists for many Mexican collections, which, thanks to an early start, a co-ordinated effort, and sufficient and constant support, are using the Biotica database, devised and developed by *CONABIO* (2002b). The specify database system (Anon., 2002) developed at the University of Kansas and used in over 50 collections should also be mentioned in this context.

Most *exchange standards* or access 'profiles' have circumvented the problem of relational structures by de-normalising the data to form a single flat-file (or two files, with the second one covering the metadata describing the dataset). The fields in the file describing the unit (specimen or recorded species) are uniquely defined, i.e. a specific datum is normally only found in one place. Two strategies have been developed to overcome the problem of different cover for data items: either the common profiles or exchange standards were focused on a few data items considered most important and common (minimal common denominator approach, e.g. 'Darwin Core' or '*ENHSIN* element set'), or they have been devised in a drawn out laborious process adapted to – and involving changes in – the participating databases ('adaptive collaboration', e.g. among Australian herbaria in the process of hispid development).

Conclusions. Both approaches are impractical for our purpose. The minimal common denominator approach, although technically enticing and leading to quick initial successes, is overly restricting the content that may be provided for direct analysis. The prediction is that it will soon fail to fulfil the expectations of research users of the network. The second approach – adaptive collaboration – is not possible given the number and variation among data providers participating in a global collection information system covering all disciplines. What is needed is a flexible approach, allowing for efficient integration of the data resources produced by the community. A wide range of data items needs to be defined semantically to ensure their compatibility when received by client systems. Of course, some mandatory elements and data access points must be defined, and this is where the common denominator approach will prevail.

BioCISE surveyed and informally compared existing database systems, which led to *ENHSIN's* recognition of the need for **variable atomisation**, i.e. the possibility of representing data both in structured and less structured form. For example, many data files or even databases created by taxonomic experts in their course of work may be excluded from public access simply because they do not adhere to a certain level of atomisation (and the experts are not willing to invest the time to re-structure their data). On the other hand, to be able to deduct content relations from structure, certain rules must apply to the relationships among such data items. A compromise allowing for variable atomisation but providing a structure open to machine analysis is the hierarchical approach. Using hierarchical decomposition as the method for analysis (cf. Berendsohn, 1997), we achieve a tree of data items where the nodes represent differing atomisation levels of the data. It is read from top to bottom as a series of 'consists of' relationships. Such a structure can be expressed as a tree in an XML document, with an alternative free text element provided for the summary of the contents of data elements of a certain level.

Of course, the relationship in content can only be ensured by means of semantic definition, integrity rules can only apply to highly atomised data items. However, the *ENHSIN* prototype has shown how a query client can use alternative text structures for controlled 'fuzzy' queries (see Güntsch, this volume).

Further needs

Present efforts are focused on data access and retrieval. For the data providers, the reverse route, standardised feedback, will be increasingly important. For example, a suitable remote annotation scheme for natural history specimens and specimen information is needed. Where specimens are routinely duplicated and distributed to several collections (e.g. botanical specimens), a 'harvesting' scheme to take advantage of data capture elsewhere can lead to considerable synergies (J. Croft, pers. comm.). Standardised feedback will also contribute to overcome one of the most serious impediments for effective use of collection information resources: the issue of data quality in collection databases.

ENHSIN's contribution to the evolving collections information system

The ABCD Working Group is in the process of developing a protocol specification as well as a comprehensive XML Schema including data element definitions, which are to be proposed to *TDWG* and *GBIF* as a standard. *ENHSIN* has actively supported this process by producing a first draft Document Type Definition (DTD) in 2001. This has been adopted by the *BioCASE* project (the proposal of which was a joint initiative of *BioCISE* and *ENHSIN*). *ENHSIN* has also been instrumental in proposing an implementation within the European Network for Biodiversity Information (*ENBI*) using the emerging standard. An *ENBI* workpackage is now dedicated to collections and within that, the *ENHSIN* prototype is to be extended both in the number of databases accessed and in the scope of data items provided. Work on the *ABCD* standard currently proceeds in a workpackage within *BioCASE* (Thomson, 2002). Within the subgroup tackling the protocol development, the Berlin team of *ENHSIN* contributed the experience gained while implementing the *ENHSIN* prototype. All this clearly shows the important role EU supported projects play within this global drive towards interoperability of collection data resources.

So, apart from aforementioned European collaborative aspects, the contributions in the fields of intellectual property rights (Owens, this volume), user interests (Calbuig *et al*., this volume), and business models (Los & de Jong, this volume), the *ENHSIN* project and the prototype network have demonstrated the viability of a network using a hierarchical structure of data items with variable atomisation in the access system. This approach will considerably widen the scope of potential data sources that can be integrated into the network.

References

ABCD, 2002. Working Group on Access to Biological Collection Data. www.bgbm.org/TDWG/CODATA/

ABRS, 2001. Australia's Virtual Herbarium. *Biologue* **24**: 9.

ABRS, 2002 [9 Jul]. About the Australian Biological Resources Study. www.ea.gov.au/biodiversity/abrs/

Anonymous, 1994 [5 May]. Logical data model for museum collections transaction management, Version 1.0. nmnhwww.si.edu/gopher-menus/LogicalDataModelfor MuseumCollections TransactionManagement.html

Anonymous, 1999 [Jan]. Final report of the OECD Megascience Forum working group on biological Informatics. www.bgbm.org/biodivinf/docs/OECDMSWGBI.pdf

Anonymous, 2002: I specify. usobi.org/specify/

Anonymous, 2002a. The Species Analyst. tsadev.speciesanalyst.net/

ASC, 1993. An information model for Biological Collections (Draft). *Report of the Biological Collections Data Standards Workshop*, August 18–24, 1992. Association of Systematic Collections, Committee on Computerisation and Networking. gopher://kaw.keil.ukans.edu:70/11/standards/asc

Berendsohn, W.G. 1997. A taxonomic information model for botanical databases: The IOPI Model. *Taxon* **46**: 283–309.

Berendsohn, W.G., A. Anagnostopoulos, G. Hagedorn, J. Jakupovic, P.L. Nimis, B. Valdés, A. Güntsch, R.J. Pankhurst & R.J. White. 1999. A comprehensive reference model for biological collections and surveys. *Taxon* **48**: 511–562. www.bgbm.org/biodivinf/docs/CollectionModel/

Berendsohn, W.G. (ed.) 1999. Standards, Information Models, and Data Dictionaries for Biological Collections. TDWG Subgroup on Accession Data, Berlin. www.bgbm.org/TDWG/acc/Referenc.htm

Berendsohn, W.G. 2000. The *BioCISE* project. Pp. 1–4 in: Berendsohn, W.G. (ed.), *Resource Identification for a Biological Collection Information Service in Europe (BioCISE)*. Berlin.

Berendsohn, W.G. & Nimis, P. L. 2000. The complexity of collection information. Pp. 13–18 in: Berendsohn, W.G. (ed.), *Resource Identification for a Biological Collection Information Service in Europe (BioCISE)*. Berlin.

Berendsohn, W.G., Häuser, C.L. & Lampe, K.-H. 2000 ['1999']. Biodiversitätsinformatik in Deutschland: Bestandsaufnahme und Perspektiven. *Bonner Zool. Monogr*. 45: 1–62. www.bgbm.org/BioDivInf/Docs/BinD/

Berendsohn, W.G., Costello, M.J., Emblow, C., Güntsch, A., Hahn, A., Koenemann, J., Thomas, C., Thomson, N. & White, R. 2000. Concepts for a European Portal to Biological Collections. Pp. 59–70 in: Berendsohn, W. G. (ed.), *Resource Identification for a Biological Collection Information Service in Europe (BioCISE)*. Berlin.

BfN, 2001. Floraweb. Daten und Informationen zu Wildpflanzen und zur Vegetation Deutschlands. *Bundesamt für Naturschutz*. www.floraweb.de

BGCI, 2002 [Mar 30]. Botanic Gardens Conservation International; An International Clearing House Mechanism for Botanic Gardens. www.bgci.org.uk/

BGCS, 1987. Botanic Gardens Conservation Secretariat, International transfer format for botanic garden plant records. TDWG Plant Taxonomic Database Standards No. 1, Hunt Institute for Botanical Documentation, Pittsburgh.

Bisby, F. 1995. Plant names in botanical databases. Plant Taxonomic Database Standards No. 3, International Working Group on Taxonomic Databases for Plant Sciences (TDWG). Pittsburgh.

BioCISE, 1999. *BioCISE*, *Resource Identification for a Collection Information Service in Europe*. Botanischer Garten und Botanisches Museum Berlin-Dahlem. www.bgbm.fu-berlin.de/biocise/

Bridson, G.D.R. & Smith, E.R. 1991. Botanico-Periodicum-Huntianum/supplementum. Hunt Institute for Botanical Documentation, Pittsburgh.

Brummit, R.K. & C.E. Powell, 1992. Authors of plant names. Royal Botanic Gardens, Kew. [TDWG Standard] Searchable database included in the IPNI database, see www.rbgkew.org.uk/web.dbs/authors.html

CABRI, 2002 [Mar 30]. Common Access to Biotechnological Resources and Information. The CABRI Consortium. www.cabri.org/

Colwell, R.K. 2002 [26 May]. Biota – The Biodiversity Database Manager.

CONABIO, 2002a [Apr 1]. REMIB: World Biodiversity Information Network. www.conabio.gob.mx/remib_ingles/doctos/remib_ing.html

CONABIO, 2002b [Apr 24]. El Sistema Nacional de Información sobre Biodiversidad de México. www.conabio.gob.mx/institucion/snib/doctos/acerca.html

Conn, B.J. (ed.), 1996. HISPID3. Herbarium Information Standards and Protocols for Interchange of Data. Version 3. Royal Botanic Gardens, Sydney. [TDWG standard] www.rbgsyd.gov.au/HISCOM/

Copp, C.J.T., 2000. The NBN data model and its implementation in Recorder 2000. *Environmental Information Management*, October 2000. www.bgbm.org/biodivinf/docs/archive/Copp_C_2000__NBN_Data_Model.pdf

DiGIR, 2002. Distributed generic information retrieval. digir.sourceforge.net/

Güntsch, A., Li, J. & Berendsohn, W.G. 2002. Euro+Med: Design of the Internet Taxonomic Sector Editor. www.bgbm.org/BioDivInf/Projects/Euro+Med/EditorDesign.pdf

Hollis, S. & Brummitt, R. 1992. World Geographical Scheme for Recording Plant Distributions. *Plant Taxonomic Database Standards No. 2, International Working Group on Taxonomic Databases for Plant Sciences (TDWG)*. Hunt Institute for Botanical Documentation, Pittsburgh.

Huber, J.T. 1998. The importance of voucher specimens, with practical guidelines for preserving specimens of the major invertebrate phyla for identification. *J. Nat. Hist*. **32**: 367–385.

ISIS, 2002 [Mar 30]. International Species Information System. www.isis.org/

Maurer, S.M., Firestone, R. B. & Scriver, C. R. 2000. Science's neglected legacy. *Nature* **405**: 117–120.

NSF, 1998. Biological Research Collections (BRC), Program Announcement (NSF 98–126). National Science Foundation, Directorate for Biological Sciences, Division of Biological Infrastructure.

OECD, 1999. Meeting of the OECD Committee for Scientific and Technological Policy at ministerial level, Paris, 22–23 June 1999. OECD, Paris. www1.oecd.org//subject/cstp/1999/body.htm

SABONET, 2002 [Apr 1]. Southern African Botanical Diversity Network – Herbarium computerisation. www.sabonet.org/countries/computerisation.html

Thomson, N. 2002 (ed.). The CODATA-TDWG-BioCASE schema. www.bgbm.fu-berlin.de/tdwg/codata/Schema/

Vieglas, D. 1999. Integrating disparate biodiversity resources using the information retrieval standard z39.50. TDWG 1999 Abstracts. Cambridge, USA, www.tdwg.org/rep1999.html#dave

Walter, K.S. 1989. Designing a computer-software application to meet the plant-record needs of the Arnold Arboretum. Arnoldia 49(1): 42–53.

Walter, K.S. & O'Neal, M.J., 2000. The BG-BASE user's manual. Kirtland, OH.

Wyse Jackson, D. (compiler) 1997. International Transfer Format for Botanic Garden Plant Records (version 2.00 draft 3.2.). Botanic Gardens Conservation International, Richmond. www.rbgkew.org.uk/BGCI/news.htm