



## **Networking Activity D**

**Task 2.3 Biography: Create a data standard for holding biographical data about workers in the biodiversity / conservation subject area, leading to an authoritative Union List of Naturalists' Names, directory of expertise or possible third-party name service**

Deliverable D 2.3.2

Report on the development of BiogML: an XML Schema for biographical information as a Synthesys NA-D activity

**Neil Thomson**

The Natural History Museum, London

June 2005

## Contents

1. Acknowledgements
2. Background and requirement
3. Development process
4. Data elements and element groups
5. Selected biographical material about naturalists
6. Appendix 1: XSD file availability
7. Appendix 2: Text version of the Schema

## 1. Acknowledgments

I wish to acknowledge the help and guidance provided by the following people in the preparation of this schema:

Walter Berendsohn, Rachel Perkins, Diane Tough, Laszlò Peregovits, Barbara Mathé, Günter Waibel.

## 2. Background and requirement

This project arose out of the European Union Framework V BioCASE (Biodiversity Access Service for Europe) project, which ran from November 2001 until early 2005.

The creation of collection description records for that project showed how important biographical details are to the history of collections, yet there appeared to be no XML Schema available for gathering together these details in a standard format that would facilitate their re-use in a variety of contexts.

This schema is intended primarily for holding biographical information about naturalists, but may also be applicable in other subject areas. It links with the Natural Collections Descriptions (NCD) schema, which is the second schema developed under the auspices of the Sythesys NA-D programme. NCD extends the collection description schema from BioCASE from being concerned only with specimen collections, to being able to deal with any type of collection of natural material. It is especially of use with mixed collections, such as those that were formed during the great voyages of exploration and discovery.

It should thus be possible for these two standards to interoperate, providing a comprehensive means of recording information about collections of natural material and the people associated with them.

Later stages in the project will deal with the population of a database of the biographies of naturalists from both published and unpublished sources, but this report is concerned just with the development of the data standard itself.

## 3. Development process

The first step in the development process was to determine whether any schemas for biography already existed that were available in the public domain or under a Science Commons style of licence.

Only one schema was found, called Bio. This schema was published as version 0.1 early in 2004 and may be found at: <http://purl.org/vocab/bio/>

Bio appears to be intended for use with Friend-of-a-friend (FoaF) and RSS feeds, with some usage in genealogy. It did not appear to be sufficiently comprehensive for the purposes envisaged for the Synthsys project and so the decision was taken to create a separate standard called BiogML.

The basic structure for BiogML (see Figure 1) was determined by examining the categories of information recorded in a wide variety of biographical dictionaries and other publications.

Additional elements catered for interoperability and record metadata. Links to biographical information that may exist elsewhere can be made through a repeatable element that records other identifiers for the person, along with the source of the ID as an attribute. Similarly, if a family member has their own biography, a link may be recorded using the same ID and source pairing.

The link through to collections with which the subject is associated also uses the same mechanism. In this case, the ID could be that of a collection description record held in the NCD standard, but it could equally be simply a collection identifier in a particular organisation.

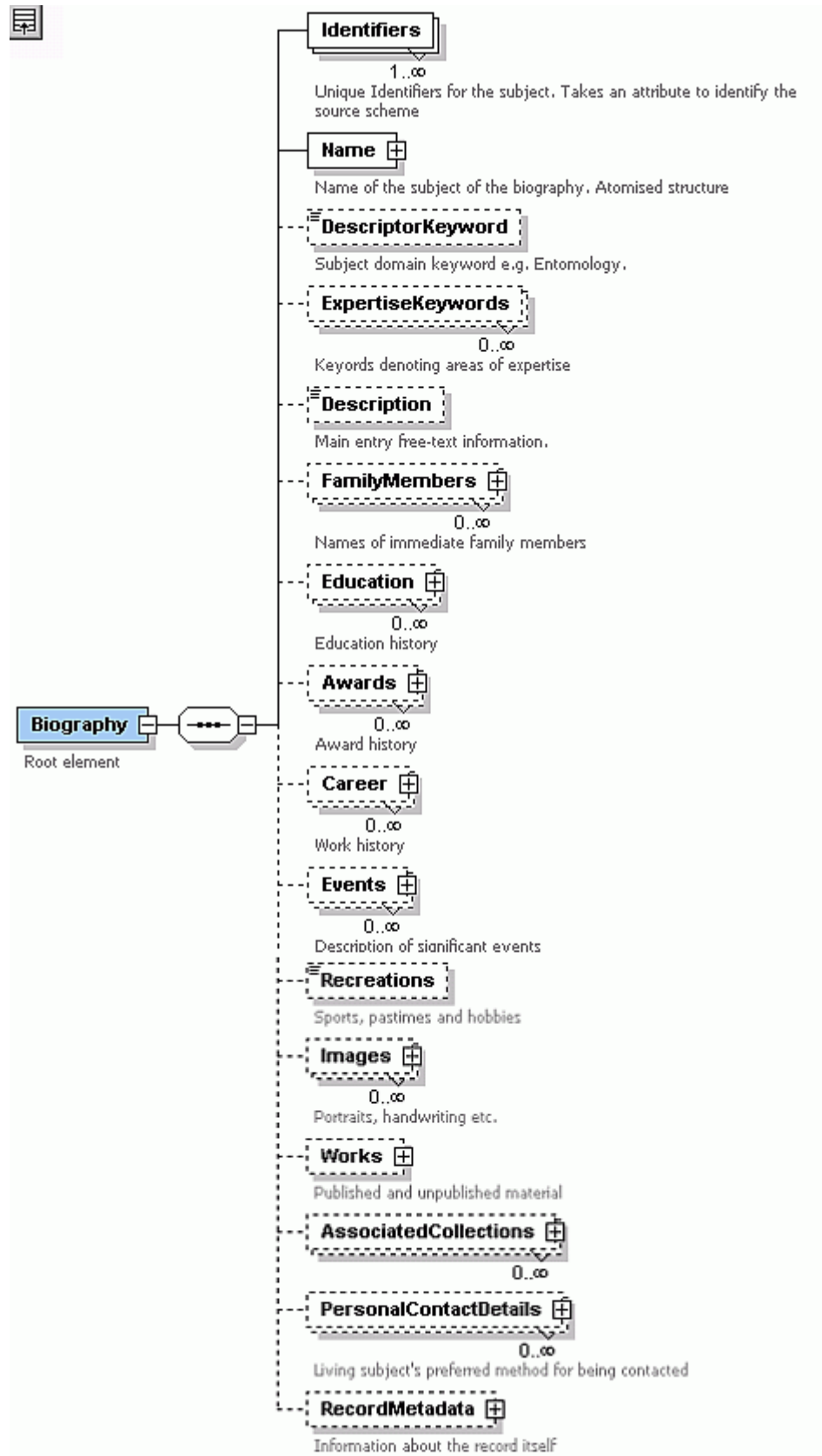
A group of elements were added to record the source of the biography, if the record is not an original creation. Also recorded is the creator and date of creation, which should never change, and an edit history where corrections or additions can be attributed correctly.

BiogML progressed from its earliest incarnation as version 0.1 through to version 0.5, which was sent for comment to several people that had expressed interest in helping with the development, noted in the acknowledgements section.

Now that version 1.0 has been reached, BiogML is ready for a wider airing and this is likely to be through the Taxonomic Databases Working Group (TDWG - about which more information may be found at <http://www.tdwg.org> ) along with the complementary NCD data standard.

TDWG is developing a suite of data standards on behalf of the Global Biodiversity Information Framework (GBIF – about which more information can be found at <http://www.gbif.org> ). GBIF promote the use of the TDWG standards throughout the world once they have reached an accepted level of stability.

## 4. Data elements and element groups



**Figure 1:** Top level structure. Note that an element in a dotted-line box is optional, elements with a double box are repeatable and a plus sign indicates an element group.

This section contains notes for the element groups, interspersed with the GIF images from XML Spy. Figure 1 (above) shows the top-level structure, which consists of 11 element groups, denoted by the plus sign in a small box at the right-hand end of the element, along with 5 simple elements.

### Simple elements

The <Description> element is the main element, being the one that will contain the free-text description of the subject in anything from a few lines to an entire essay.

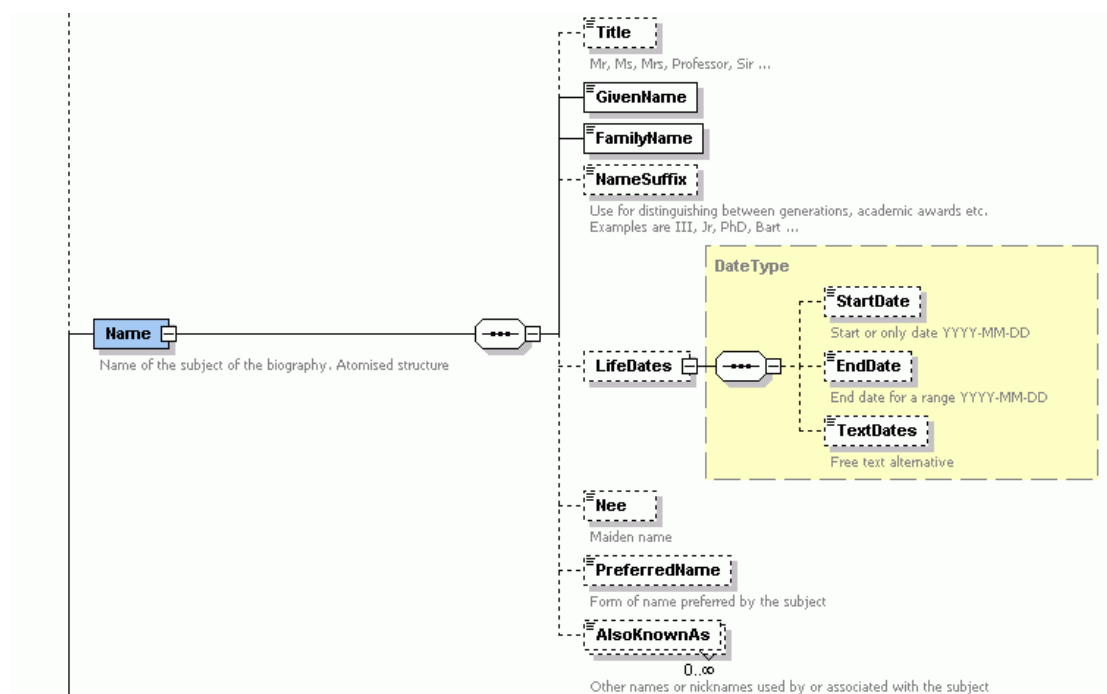
Only an <Identifier> and certain elements of the subject's name are mandatory, to give as much flexibility as possible for incorporating data from a wide variety of sources. A record can thus start out as a stem, with the details filled in at a later time.

The <Identifier> element, along with a few others, takes an attribute that holds the source of the identifier.

The <DescriptorKeyword> and <ExpertiseKeywords> elements give respectively a course and fine indication of the subject domain and specific areas of expertise of the subject. These keywords may be drawn from controlled terminology sources, in which case the term identifier and source can be recorded as attributes.

The final simple element is <Recreations>.

### Name Element Group



**Figure 2: Name**

The name of the subject is atomised into its individual components, so that they may be re-assembled in whatever order is required.

The <LifeDates> complex type is used to record date of birth and, if applicable, death. If these are unknown or inexact, a text alternative is provided for recording phrases such as “lived during the early 20<sup>th</sup> century”. The <Nee> element records the maiden name of a married woman.

There is a <PreferredName> element that can be used to record the subject’s own preferred form of name. This may be a nickname or indicates the preferred use of a “middle” name over the main given name. Finally, a series of other names by which the subject is known can be recorded, which may be useful for those that write under a “nom de plume”.

### Family Members Element Group

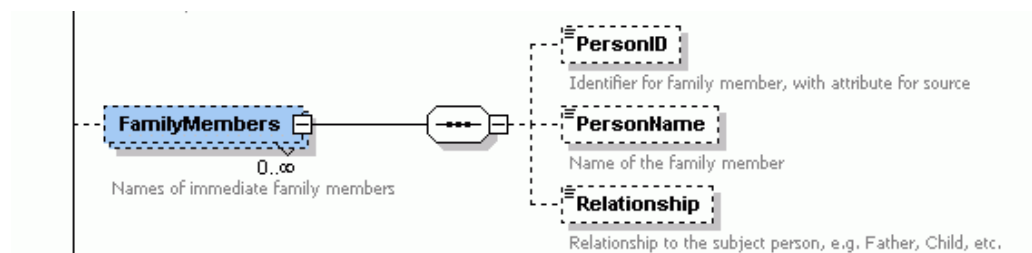


Figure 3: Family Members

This repeating group can hold the basic details of the family members of a subject. Just the person’s name and their relationship to the subject is held, but if there is a separate biographical record somewhere, it can be linked through the <PersonID> element, using the attribute to record the source of the ID.

### Education, Awards, Career and Events Element Groups

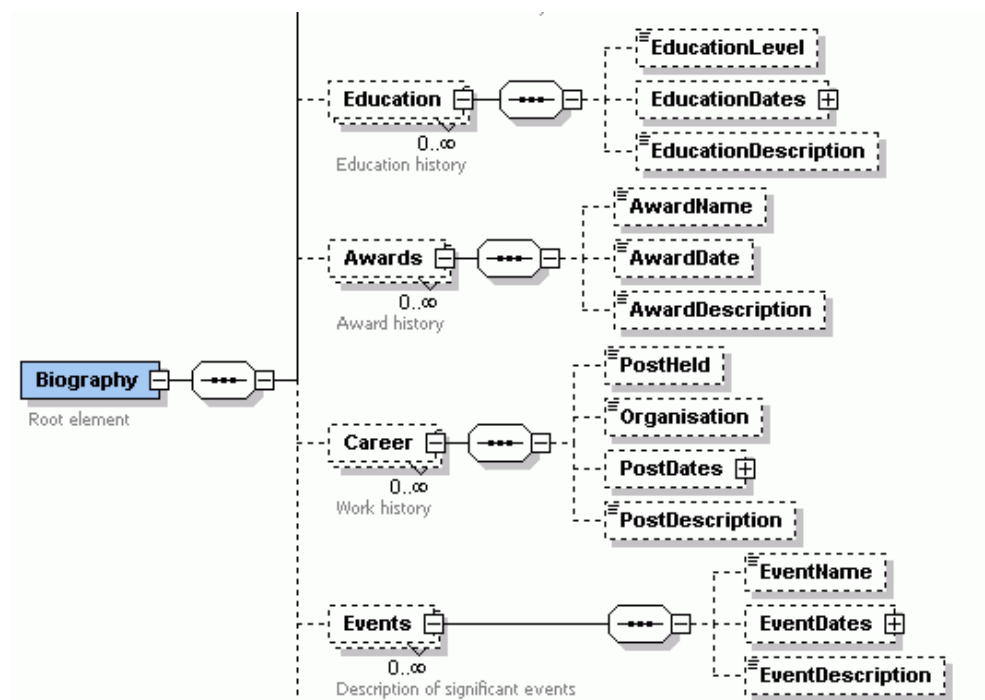
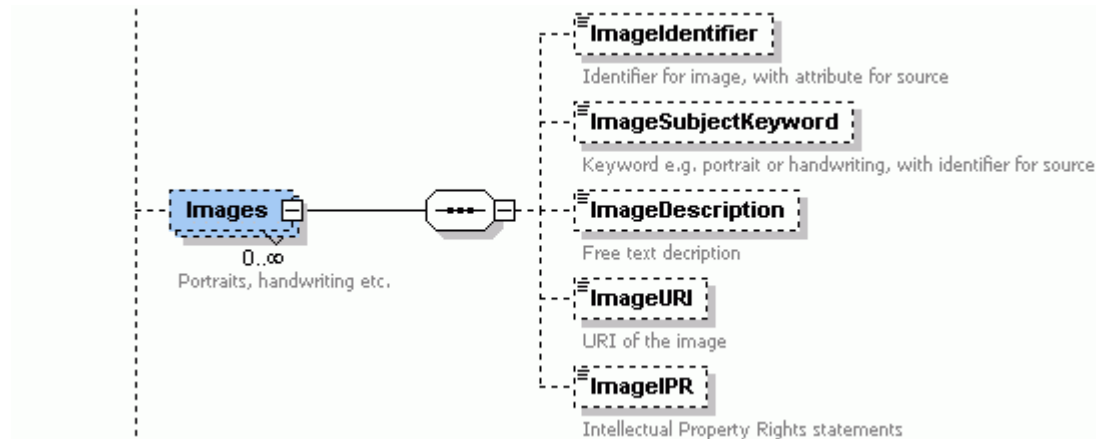


Figure 4: Education, Awards, Career and Events

Figure 4 shows the structure of the four element groups that deal with the subject's education, awards, career and other significant events. Each is a repeating group and, with the exception of <Awards>, each uses the <DateType> structure, as shown in Figure 2.

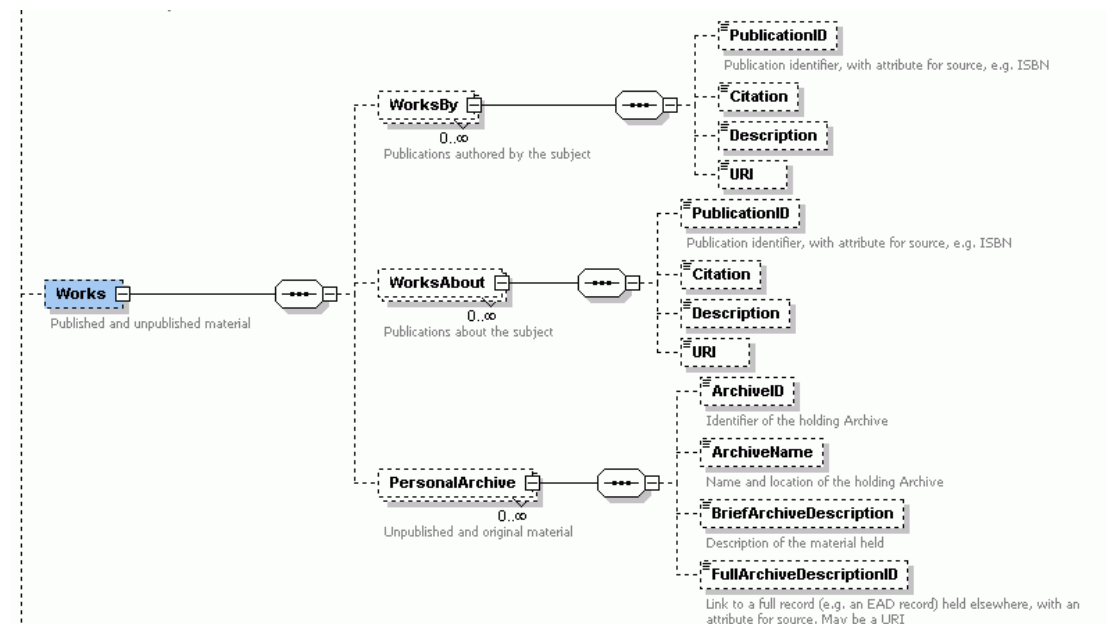
### Images Element Group



**Figure 5: Images**

Images are an important part of biography, especially portraits and handwriting samples. The schema does not accommodate binary files directly, only via a URI.

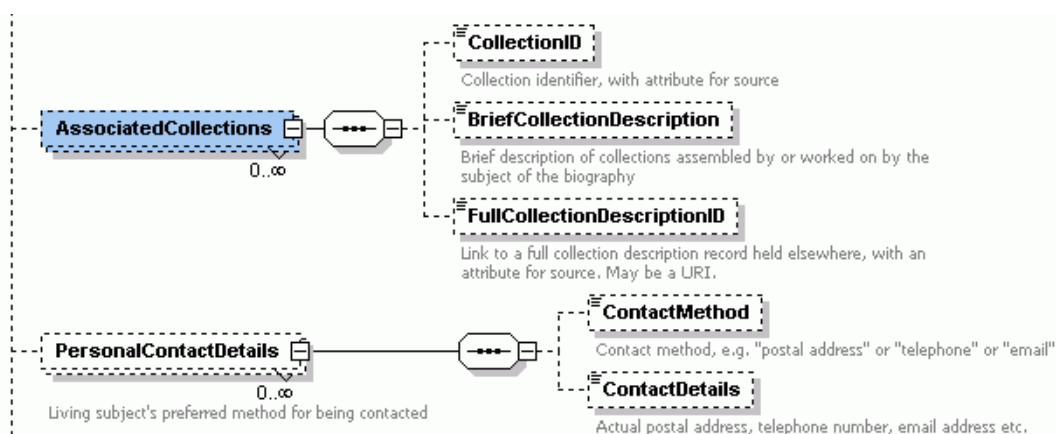
### Works Element Group



**Figure 6: Works**

This section of BiogML deals with works authored by the subject or authored by others about the subject. Where an official archive exists, the location of this may be recorded along with a brief description and a link to a fuller descriptive record, should one exist elsewhere.

## Associated Collections and Contact Details Element Groups



**Figure 7: Collections and Contact Details**

The <AssociatedCollections> group allows the interlinking of collection description records, if they exist, such as those recorded in the Natural Collections Descriptions (NCD) format, which is also being developed under the auspices of Synthesys NA-D. Optionally, a brief description may be given here, which may be particularly useful if no external records are available.

### 5. Selected biographical material about naturalists

To give an indication of the wealth of biographical material about naturalists that is available, the following sections provide selected references to published and unpublished material revealed through use of the library catalogue for the Natural History Museum, London. This catalogue is available online and may be searched from <http://www.nhm.ac.uk/library/index.html>

#### Part 1: Sources to published material

##### Articles

There are approximately 2,500 entries on the Natural History Museum Library catalogue pointing the user to biographical information on naturalists in journal articles. All are searchable under “biography” in the library catalogue.

##### Biographies – General

*Australian Dictionary of Biography*. 14 volumes, General editor: John Ritchie. Carlton, Victoria: Melbourne University Press, 1966-1996.

*Dictionary of American Biography*. 10 volumes, edited by A. Johnson; [vol. 2-4 ed. by A. Johnson & D. Malone; vol. 5-10 ed. by D. Malone.] New York, Charles Scribner’s Sons, 1927-1930.

*Dictionary of New Zealand Biography*. 5 Volumes, edited by W.H. Oliver and Claudia Orange. Auckland : Auckland University Press/Department of Internal Affairs, 2000.

*A Dictionary of Twentieth-Century World Biograph*. edited by Asa Briggs. Oxford: Oxford University Press, 1993.

*The International Who's Who, 1998-99.* 62<sup>nd</sup> ed. London: Europa Publications, 1998.

*Neue Deutsche Biographie.* Berlin: Duncker & Humblot, 1994.

*Oxford Dictionary of National Biography.* Oxford: Oxford University Press, 2004.

*Royal Commonwealth Society – Biography Catalogue.* London: Royal Commonwealth Society, 1961,

*Who's who in Scandinavia.* edited by Karl Strute and Theodor Doelken. München, Who's Who 1981.

*Who's Who.* London: A & C Black, 2004.

*Who Was Who 1991-1995 : A companion to Who's Who containing the biographies of those who died during the period 1991-1995.* New York: St. Martin's Press, 1996.

### **Biographies - Science**

African Academy of Sciences, *Profiles of African Scientists.* Nairobi: African Academy of Sciences, 1990.

*American Men and Women of Science.* 17<sup>th</sup> ed., 8 vols. New York: R. R. Bowker, 1989-1990.

Creese, Mary R. S., *Ladies in the Laboratory? : American and British in science, 1800-1900 : a survey of their contributions to research.* Lanham, Maryland: Scarecrow Press, 1998.

Daintith, John et al... , ed., *Biographical Encyclopaedia of Scientist.* 2<sup>nd</sup>. ed. Bristol: Institute of Physics, 1994.

Haines, Catherine M. C. with Stevens, Helen M., *International Women in Science.* Santa Barbara, California: ABC-CLIO, c2001.

Hellman, Hal, *Great Feuds in Science : ten of the liveliest disputes ever.* Chichester: Wiley, 1998.

Holmes, Frederic L., ed., *Dictionary of Scientific Biography.* New York: Charles Scribner's Sons, 1991.

Howsam, Leslie, *Scientists since 1660 : a bibliography of biographies.* Aldershot, Hampshire: Ashgate, 1997.

Lightman, Bernard, ed., *Dictionary of Nineteenth –Century British Scientists.* Bristol: Thoemmes Continuum, 2004.

Martin, Paula, *Lives with Science : profiles of senior New Zealand women in science.* Wellington: Museum of New Zealand Te Papa Tongarewa, 1993.

Millar, David, ed., *The Cambridge Dictionary of Scientists*. 2<sup>nd</sup> ed. Cambridge: Cambridge University Press, 2002.

Newbold, Heather, ed., *Life Stories : world-renowned scientists reflect on their lives and on the future of life on earth*. Berkeley: University of California Press, 2000.

Turkevich, John, *Soviet men of Science : academicians & corresponding members of the Academy of Sciences of the USSR*. Princeton, New Jersey: D. Van Nostrand, 1963.

*Who's who in world Agriculture : a biographical guide in the agricultural and veterinary sciences*. 2<sup>nd</sup> ed. Harlow, Essex : Longman, 1985.

*Who's who in Science in Europe : a biographical guide in science, technology, agriculture and medicine*. 5<sup>th</sup> ed., Harlow : Longman, 1987.

### **Biographies - Natural History**

Arnett, Ross H. and Mary E., *The Naturalists' directory and Almanac (International): an index to contemporary naturalists of the world and their special interests*. 44<sup>th</sup> ed., Gainesville, Florida: Flora & Fauna Publications, 1985.

Axelrod, Alan and Charles Phillips, *The Environmentalists : a biographical dictionary from the 17<sup>th</sup> century to the present*. New York: Facts on File, 1993.

Curmie, Lynn M., *World Who is Who and does What in Environment and Conservation*. London: Earthscan for the Foundation for Environmental Conservation, 1997.

Davies, Matthew et al, *Who's Who in the Environment : Wales = Pwy yw pwy yn yr amgylchedd : Cymru*. London: Environmental Council, 1995.

Desmond, Ray, *Great Natural History Books and their Creators*. London: British Library, 2003.

Dunn Ewan, Joseph and Nesta, *Biographical Dictionary of Rocky Mountain Naturalists : a guide to the writings and collections of botanists, zoologists, geologists, artists and photographs, 1682-1932*. Utrecht : Bohn, Scheltema & Holkema ; The Hague : Junk, 1981.

Gates, Barbara T., *Kindred nature : Victorian and Edwardian women Embrace the Living World*. Chicago: University of Chicago Press, 1998.

Grinstein, Louise S., Biermann, Carol A. and Ross, Rose K., eds., *Women in the Biological Sciences : a bibliographic sourcebook*. Westport, Connecticut: Greenwood Press, 1997.

Hunt Institute for Botanical Documentation, *Catalogue of Portraits of Naturalists, mostly Botanists, in the collections of the Hunt Institute, the Linnean Society of London and the Conservatoire et Jardin Botaniques de la Ville de Geneve*. Pittsburgh: Hunt Institute for Botanical Documentation, Carnegie-Mellon University, 1988.

Howgego, Raymond John, *Encyclopedia of Exploration to 1800 : a reference guide to the history and literature of exploration, travel and colonization from the earliest times to the year 1800*. Sydney: Hodern House, 2003.

Lucas, Eileen, *Naturalists, Conservationists and Environmentalists*. New York: Facts on File, 1994.

Moring, John, *Early American Naturalists : exploring the American west. 1804-1900*. New York: Cooper Square Press, 2002.

Pavillion, Jean-François, *Annuaire des Chercheurs Francophone en Sciences de l'eau = Directory of Fresh-water and Marine French speaking Scientists*. 3<sup>rd</sup> ed. Paris: Union des Oceanographes de France, 2000.

Poggendorff, J. C. *Biographisch-literarisches Handwörterbuch*. Leipzig: Johann Ambrosius, 1863 – 1904.

Shearer, Benjamin F. and Shearer, Barbara S., eds., *Notable women in the Life Sciences : a biographical dictionary*. Westport, Connecticut: Greenwood Press, 1996.

Siegel, Patricia Joan and Finley, Kay Thomas, *Women in the Scientific Search : an American bio-bibliography : 1724-1975*. London: Scarecrow Press, 1985.

Sterling, Keir B., ed. et al, *Biographical dictionary of American and Canadian Naturalists and Environmentalists*. Westport, Connecticut: Greenwood Press, 1997.

Varley, Allen, ed., *Who's Who in Ocean and Freshwater Science*. Harlow: Hodgson, 1978.

### **Biographies – Botany**

Desmond, Ray, *Dictionary of British and Irish Botanists and Horticulturalists : including plant collectors, flower painters and garden designers*. London: Taylor & Francis, 1994.

Funk, V. A. and Scott, A. Mori, *A Bibliography of Plant Collectors in Bolivia*. Washington D. C.: Smithsonian Institution Press, 1989.

Kelly, Howard Atwood, *Some American Medical Botanists Commemorated in our Botanical Nomenclature*. New York ; London: D. Appleton & Co., 1929.

Stafleu, Frans A. and Richard S. Cowan, eds., *Taxonomic Literature: a selective guide to botanical publications and collections with dates, commentaries and types*. 7 volumes. Utrecht : Bohn, Scheltema & Holkema, 1976-88. PLUS 6 supplements 1992-2000.

Vernon, R., ed., *Directory of Research Workers in Agriculture and Allied Sciences*, Wallingford, Oxfordshire: CAB International, 1989.

## **Biographies - Zoology**

Blackwelder, Richard E. & Ruth M., *Directory of Zoological Taxonomists of the World*. Carbondale, Illinois: Southern Illinois University Press for The Society of Systematic Zoology, 1961.

Friedländer, R. & Sohn eds., *Zoologische Adressbuch: international Zoologists's Directory*. Berlin: R. Friedländer & Sohn, 1875.

Mearns, Barbara & Richard, *Biographies for Birdwatchers: lives of those commemorated in Western Palearctic bird names*. London: Academic Press, 1988.

Pemberton, John E., ed., *Who's Who in Ornithology*. Buckingham: Buckingham Press, 1997.

Science Museum Library, *A Biobibliography of Geologists who died between 1850 and 1900*. London: Science Museum, 1972.

Zoologist Society of India, *Who's who of Zoologists of India*. Bhubanewar, India: B. K. Behura on behalf of the Zoologist Society of India, 1980.

## **Biographies - Entomology**

Carlucci, Lisa and Sagle, Leslie E., *Record Unit 7323 : Systematic Entomology Laboratory, USDA : photographs of and biographical information on entomologists, 1797-1988 and undated*. Washington, D.C.: Smithsonian Institution, [1989?].

Gilbert, Pamela, *A Compendium of the Biographical Literature on Deceased Entomologists*, London: Natural History Museum, 1977.

Kulshrestha, S. K., ed., *Directory of Indian Entomologists*. Dehra Dun, India: Surya International, 1997.

Silbermann, Gustave, *Enumeration des Entomologistes Vivans*. Paris: Chez Roret, 1835.

## **Biographies - Earth Sciences**

Cleevely, R. J., *World Palaeontological Collections*. London : BMNH; Mansell, 1983.

Gates, Alexander E., *A - Z of Earth Scientists*. New York: Facts on File, 2003.

La Rocque, A. ed., *Biographies of Geologists Materials for the Study of the History of Geology, prepared in Geology 851: a seminar in the history of geology 1953 to 1958*. Columbus, Ohio: Ohio State University of Geology, 1961.

Sargeant, William A. S., *Geologists and the History of Geology: international bibliography from the origins to 1978*. 5 vols. London: Macmillan Press, 1980.

PLUS 2 supplements 1979-1984 (published in 1987) and 2 supplements in 3 volumes 1985-1995 (published in 1996). Supplements published Malabar, Florida: Krieger Publishing Company.

### **Biographies – Natural history Illustration**

Hardouin-Fugier, Elisabeth and Grafe, Etienne, *French Flower Painters of the Nineteenth Century : a dictionary*. London :Philip Wilson Publishers, 1989.

Horne, Alan, *Dictionary of 20<sup>th</sup> century British Book Illustrators*. Woodbridge, Suffolk: Antique Collectors' Club, 1994.

McKay, Judith, *Brilliant Careers : women collectors and illustrators in Queensland*. South Brisbane: Queensland Museum, 1997.

#### **Part 2: Unpublished material**

Most of the libraries hold biographical indexes on cards or paper. Some of the cards are handwritten, some typed and some are printed information cut out of publications and pasted onto cards. All give birth and death dates if known. Some cards will have more than one reference on a card.

- Earth Sciences Library – 2 card catalogue drawers of Mineralogy index (approx. 2,450 cards) & 7 card catalogue drawers of Palaeontology index (approx. 7,000 cards)
- Entomology Library – 12 drawers of catalogue cards compiled by Pamela Gilbert. The information contains references to books and journal articles and portraits. Pam is using this information to produce her book on Biographical information on Entomologists.
- General Library – 12 drawers of catalogue cards. Includes references to mainly journal articles, but some books.
- Tring Library – 10 drawers of catalogue cards
- Zoology Library – 40 drawers of catalogue cards

Botany library does not hold indexes to the library's biographical material but holds 35 red boxes of biographical material arranged by name plus 5 boxes of Professor W. Stearn's reprints. It also holds 10 boxes of information on British botanists compiled for Desmond's book.

#### **Portraits**

- Earth Sciences – 1 card catalogue drawer of Portrait indexes for Mineralogy (approx. 1,050) & 2 card catalogue drawers of Portrait indexes for Palaeontology (approx. 1,450)
- General – 22 catalogue drawers of Portrait indexes (approx. 21,000)

## **6. Appendix 1: XSD file**

An XML Schema file (.xsd) suitable for use with XML Spy or similar programs is available from the Synthesys Website.

## 7. Appendix 2: Plain version

### BiogML V1.0 Schema May 2005 for Synthesys NA-D

Identifiers (With Source) Repeatable Element  
Descriptor Keyword (With Source and ID in Source)  
Expertise Keywords Repeatable Element  
Description

#### Name Group

Title  
Given Name  
Family Name  
Name Suffix  
Nee  
Preferred Name  
Also Known As Repeatable Element  
**Life Dates Group**  
Start Date  
End Date  
Text Dates

#### Family Members Group

Person ID (With Source) Repeatable Group  
Person Name  
Relationship

#### Education Group

Education Level Repeatable Group  
Education Description  
**Education Dates Group**  
Start Date  
End Date  
Text Dates

#### Awards Group

Award Name Repeatable Group  
Award Date  
Award Description

#### Career Group

Post Held Repeatable Group  
Organisation  
Post Description  
**Post Dates Group**  
Start Date  
End Date  
Text Dates

#### Events Group

Event Name Repeatable Group  
Event Description  
**Event Dates Group**  
Start Date  
End Date  
Text Dates

## Recreations

### Images Group

Repeatable Group

Image Identifier (With Source)  
Image SubjectKeyword (With Source)  
Image Description  
Image URI  
Image IPR

### Works Group

#### Works By Group

Repeatable Group

Publication ID (With Source)  
Citation  
Description  
URI

#### Works About Group

Repeatable Group

Publication ID (With Source)  
Citation  
Description  
URI

#### Personal Archive Group

Repeatable Group

Archive ID (With Source)  
Archive Name  
Brief Archive Description  
Full Archive Description ID (With Source)

### Associated Collections Group

Repeatable Group

Collection ID (With Source)  
Brief Collection Description  
Full Collection Description ID (With Source)

### Personal Contact Details Group

Repeatable Group

Contact Method  
Contact Details

### Record Metadata Group

Record Source  
Record Creator  
Record IPR  
Record Created Date

#### RecordEdit History Group

Repeatable Group

Record Editor  
Record Edit Date  
Record Edit Comment