

# 5

## TOWARDS A MANAGEMENT AND BUSINESS PLAN

WOUTER LOS AND HERMAN DE JONG

76

### INTRODUCTION

Sharing digitised specimen information requires co-operation between those institutes acting as custodians of such information. Such co-operation needs to be supported by agreed protocols on financial, legal and management issues. Within *ENHSIN*, the University of Amsterdam (Zoological Museum) was responsible for designing a Management Model for a Memorandum of Understanding, and for drafting a strategy document on resources. The Royal Botanic Gardens, Kew, was responsible for examining the legal aspects of the Network (see Owens, this volume).

To inform this study, an inventory was made of existing institutional policy documents. This exercise facilitated analysis of various management issues, a review of alternative models, and enabled key elements for the Memorandum of Understanding to be listed. A number of printed institutional policy documents and mission statements were obtained by special request from libraries, while additional information was gleaned from the Internet. Explicitly, formulated and specific guidelines for sharing and transferring electronic collection data relevant to *ENHSIN* were usually lacking, although in many cases policy codes for electronic data exchange were said to be in preparation. Nevertheless, some relevant documentation could be gathered to identify the main themes to be addressed in the *ENHSIN* Management Model. Topics covered included the management structure (organisation, legal position, stratification of decision powers), management of authorisation of data access by partners and users, data ownership and copyrights, data quality and validation, data maintenance and updating, and future support.

The inventory of potentially complementary initiatives revealed that the realm of biodiversity informatics is in a state of rapid flux. Amid all the biodiversity informatics infrastructures, *ENHSIN* occupies a unique position by taking into consideration all scientific, technical, organisational, management and legal issues. The conclusions of the project with respect to the outline of a management and business plan are described below. However, although this effort was directed primarily towards *ENHSIN*, the results are also being used to explore management complexities more comprehensively in other initiatives, notably *BioCASE* and *ENBI*, which have many more partners affiliated across Europe.

## CHARACTERISTICS OF ENHSIN

The objective of *ENHSIN* is to study the structure of an information service by sharing specimen-related data from distributed sources (databases). It is not intended to be a system operated from a single server or institution, but rather a network structure that can be visited from any site by means of a common access system. Database custodians (museums, herbariums, etc.) typically have in the past developed their own systems and online services. But with modern information technology, seamless integration of these different data formats can be enabled across the Internet. The *ENHSIN* demonstration model shows that it is possible to provide a distributed service, regardless of the way in which each organisation manages and displays its data. The advantage of such a system is that each participant keeps control of its own data, in terms of specimens selected, data fields for each specimen, and the format in which the data are stored locally.

77

## IMPLICATIONS FOR THE MANAGEMENT OF A SHARED SERVICE

Although the possibility of sharing specimen data presents many opportunities, it raises a number of questions for those organisations wishing to maintain such a common virtual infrastructure: who will manage the service?; to what extent is there a joint responsibility for the service (for technical mechanism and for the data)?; what are the financial implications? Comparable services and initiatives in biology have been studied (e.g. stocks, genomic data, brain data), but do not appear to be applicable to natural history specimen data.

It is clear, nevertheless, that almost all biological information services with digital data face comparable problems. The solutions to these problems vary between the objectives and function of the database services. In the case of digitised information about living objects (for example genetically modified organisms, or strains), the services act as a shop-window to order specimens. Plainly, such data providers have no difficulty in maintaining and paying for the showcase, as it helps their income generation.

The situation is different for database services where the data refer to objects that are not, largely, commercially exploitable objects. The data themselves have limited commercial value or are considered as belonging to the public domain. The latter situation is exemplified by the service of the European Molecular Biology Network (*EMBNet*), which deals with molecular sequences of proteins and DNA. It is considered that such sequence data are the results of scientific work that is publicly funded and managed and should, therefore, be placed in the public domain. Other (mostly private) organisations consider their digital sequence data as property, and only available publicly after patenting. It is not the purpose of the authors of this chapter to make a judgment on these positions, but it should be noted that the positions taken by natural history institutions also vary. However, when the digital data held by these institutions are made accessible to other users through a common, shared service, it is essential that a common approach exists. Such issues will also have to be addressed by the Global Biodiversity Information Facility (*GBIF*). Since these matters are not part of the current workplan of *GBIF*, it is hoped that results from the *ENHSIN* study will be found to be of value to the global initiative represented by *GBIF*.

## ASSUMPTIONS FOR A MANAGEMENT MODEL

It should be appreciated that in *ENHSIN*, efforts were directed at designing the infrastructure, not at prescribing the kind of formats of the data to be shared. As long as there is basic similarity of structure within the networked databases, network functionality is not compromised (see Güntsch, this volume). Thus neither the infrastructure nor its network of participants shares either the specimens or the ownership of specimen data. The shared infrastructure should be perceived only as the mechanism permitting access to the integrated information. The extent of the detail made available to the network is determined by the policy of each member institute of the infrastructure. In general, these data are already available via the institute's web pages. From this observation, it may be implied that an information infrastructure acts only as the messenger, the tool by means of which the data may be shared: it does not determine the data contents.

In designing the management of such an infrastructure, the legal entity (or entities) needed to maintain it should be specified. Should that entity be a *single* institute or company charged with maintaining and exploiting the infrastructure, or should it be regarded as a *joint* service? The first alternative is analogous with the position held by a publisher. Given the nature of the scientific data and the public funding that typically enables its compilation, it is likely that a partnership will prefer, and indeed be under pressure, to keep control over its information service.

Based on the considerations above, the main issues on which partners in a common information infrastructure should agree can be identified as follows:

- ♦ Criteria or procedures required of an institution to participate in the infrastructure network.
- ♦ Ownership of the data and of the infrastructure.
- ♦ Definition of the data type and standards to be accepted in the infrastructure.
- ♦ Infrastructure standards and protocols.
- ♦ Policy for the provision of data access.
- ♦ Governance: maintenance, management and finance.
- ♦ Procedures allowing changes in the agreement.

Before examining the implications of these issues on management and business plans, we consider in more detail the extent to which there is shared responsibility for data contents in the distributed infrastructure.

## WHEN IS THERE A SHARED RESPONSIBILITY FOR DATA CONTENTS?

It has been argued above that an infrastructure providing access to shared data, chosen and offered to the network by independent data providers, has no responsibility for these data. Nevertheless, there are circumstances when such an infrastructure may need to consider accepting partial responsibility.

79

## DATA QUALITY

The significance of a joint information service depends strongly on the size of the data set and its quality. Data quality can be valued in different ways. For some applications it is essential to be able to rely on data of a high level of accuracy. In other applications, it is more important to know where particular specimens are housed, even if the level of accuracy of their identification is not absolute. It is assumed that each institute already participating in the joint information service is capable of making its own judgments about the validity and quality of the shared data, but which will be based on common curatorial standards. If a third party asks to share its data, the partnership will have to decide whether this party offers data appropriate to the infrastructure. For example, it is unlikely for example that a custodian of a database of postcards with flower images will be accepted. The decision on how to react to such a request implies a degree of decision making by the infrastructure over data content, but this responsibility is, in fact, focused rather on the custodians that curate the data. The infrastructure needs to be aware of this issue when drawing up policy.

## LIABILITY IN REGARD TO OWNERSHIP AND IN REGARD TO THIRD PARTY RIGHTS

A legal entity may argue that data from third-party material should not be shared in a common information service. The Convention on Biological Diversity could be interpreted as saying that permission is required if data about specimens from a specific country or origin are to be made accessible through the Internet. The question then arises as to whether this issue is one falling upon the institute that houses the specimen and associated digitised information, or whether it is the responsibility of the messenger – namely the shared service.

A further issue is that third party rights may be infringed by the information in the shared information service. A well-known example concerns the citation of localities of endangered species on the Internet. Who is responsible if the joint service transmits this information, but the information originates from an independent partner?

The growing liability of Internet providers on the information they transmit suggests that it would be wise for a common policy to be agreed for the transmission of data about natural history specimens across the Internet. The minimal requirement would be the provision of a disclaimer with respect to the data of each separate partner in the network. But the partnership may also benefit from a code of conduct. Models for such a code are available through the International Council of Museums (*ICOM*).

80

## SPECIAL SERVICES

An effective information service will help satisfy requests from special users. These users may wish to have access to a specific set of data, or combination of data, that are not directly available from the information service. In such a case the service could act as an intermediate to prepare a contract between the user and the delivering institutes.

## MEMORANDUM OF UNDERSTANDING

A number of issues raised in the previous paragraphs suggest the need for a special kind of agreement between the partners of a network. As the information service will act as a joint infrastructure, allowing new partners to join easily, a Memorandum of Understanding provides a suitable facilitating mechanism. The different items of such a Memorandum of Understanding are considered in this section.

## THE PARTNERSHIP

Since the sole purpose of the joint infrastructure is to provide access to shared information about natural history specimens, it may be perceived as an open network. Each member institute, or other body, with online and available digital specimen data shares its information in the infrastructure. The members

may be large natural history institutions, smaller museums or laboratories with far fewer data, agricultural and other specialised institutes, or companies.

## REQUIREMENTS FOR PARTNERS

A fundamental expectation of each partner is that it will be prepared to share its data or parts of its data, and that there are no special restrictions placed upon it to prevent these data from being placed in the public domain. Each partner is also expected to accept its responsibility for the information contents (as agreed in the disclaimer for the whole infrastructure). This expectation implies that each legal partner will need to agree a 'code of conduct' with respect to (a) level of accuracy of specimen data, (b) ownership or legal custodianship of the data, (c) ethical and other restrictions to putting specific data in the public domain (see Owens, this volume). Finally, the maintenance of a 24/24 hardware and software environment has to be promoted by each partner to provide continuous information access.

## LEGAL POSITION AND OWNERSHIP

Since the tool to share and disseminate the specimen information was developed by the partnership, it is expected that the contractors of the project would collectively own the results of the *ENHSIN* architecture, the software and some hardware. Further implementation of the information service will require a joint commitment to keeping the infrastructure alive and to developing it. Nevertheless, the physical infrastructure (architecture, software, hardware) has to be placed somewhere. It could be a separate legal entity, but the provision of a hosting legal body (one of the partners) seems an adequate solution for the time being. As such, the partnership should agree that this hosting legal body should maintain the physical infrastructure, within the conditions of the management of this infrastructure (see below).

81

## MANAGEMENT

The arguments given already imply that each organisation that (a) can be regarded as a reliable provider of natural history specimen data, and (b) is willing to agree with the Memorandum of Understanding, belongs to the partnership. Implicitly, the partnership agrees on maintaining the information infrastructure, although the physical infrastructure will be hosted by one of the partners. The commitment of the partnership will be reflected in the management and the finance of the infrastructure. A practical approach to achieving this end is to keep the day-to-day management (personnel, hardware, software) with the hosting partner. A small elected/appointed executive committee on behalf of the partnership may oversee the day-to-day management, but – more importantly – would be responsible for defining the policies for the development, scope, and finance of the infrastructure.

## INFRASTRUCTURE STANDARDS AND PROTOCOLS

Such issues are those of common concern to the partnership. These have been developed in a number of international networks and in eu projects. The *ENHSIN* project builds on these standards and protocols, which include:

- ♦ Information exchange and access.
- ♦ Modes of data presentation.
- ♦ Interoperability with other information systems.

The partnership needs to agree the general principles developed, while accepting that the executive committee is mandated to decide on further developments. The *BioCASE* project is already being run along these lines, which shows that the large European partnership in this project agrees with common standards and protocols.

## POLICY TO PROVIDE ACCESS TO THE INFORMATION SERVICE

A fundamental requirement of the information service is that it will put the shared information in the public domain. However, the Memorandum of Understanding should contain a provision for the partnership to agree a policy for restricting access rights (for example preventing certain databases from being down-loadable). This will satisfy any need to protect non-authorized data or information restricted by commercial agreement. It will also be necessary to ensure that a disclaimer is included in the Memorandum of Understanding to prevent collective liability of the common service. This disclaimer would state that each partner is to decide individually on those data that, for any reason, should not reside in the public domain (at least not until permission from an entitled third party is obtained). Provision should be made, however, for such data to be accessible to the partnership alone.

82

## MAINTENANCE AND FINANCE

Maintenance of the infrastructure implies the need for costs for hardware upgrading, software development and supporting personnel. The partnership may also wish to invest in further developments and innovation (such as interoperability with ecological databases), and in helpdesk support or training for individual partners. Arrangements for cost recovery will be discussed on p.83. It should be noted that the principle of providing shared data in the public domain also implies that partners will not charge for their investments in digitisation. Copyright for data provided by each partners will remain with that partner. The issue of separate or common arrangements for copyrights is discussed by Owens (this volume).

## PROCEDURE TO ALLOW FOR CHANGES IN THE MOU

If the executive committee has reasons to propose changes in the text of the Memorandum of Understanding, this should be possible by a substantial majority vote in the partnership (ballot via email). Each partner will have to accept this procedure when agreeing the Memorandum of Understanding.

## OPTIONAL ITEMS

In addition to the previous items for a Memorandum of Understanding, there should be provision for an option on extra activities of the shared infrastructure. These optional extra activities are summarised below.

## RESEARCH AND FURTHER DEVELOPMENTS

Given the rapid rate of innovation in information technology, the partnership should be ready to identify opportunities so as to improve the performance of the infrastructure. Users of the information service are likely to generate requests for new applications. The infrastructure should have a research and development plan to meet these requests, provided funding is available to implement it.

83

## HELPDESK AND TRAINING

The system of a shared information service requires that each partner wraps its contributed databases with the wrappers written by the partnership. Provision of a central helpdesk to assist with problems or software updates is an option that might become relevant in future. If special applications are developed, consideration might be given to organising training courses.

## SPECIAL REQUESTS

The natural history specimen information service will promote users' requests for specific information leading to knowledge about, for example, the biodiversity of an area. Such users might contact the information service to request specific information. The executive committee should be allowed to prepare – in agreement with the partners contributing relevant data – a project and a contract should such a specialist user need data with added value.

## FINANCIAL RESOURCES

If it is accepted (see p.82) that certain basic costs are required to maintain collections infrastructures (notably for updating hardware and software and for employing supporting personnel), then a means for covering these costs needs to be found. One approach is that of 'the user pays', for example through fees or subscriptions. As the partners of the infrastructure are themselves users, a fixed contribution from each partner may be considered. Many

countries have signed the *GBIF* Memorandum of Understanding, and established funded national *GBIF*-nodes to act as the facility to share the national biodiversity information with *GBIF*. Since *ENHSIN*, and particularly *BioCASE*, its operational successor, will offer access to its information to *gbif*, *GBIF*-nodes might be prepared to pay for the continuation of the infrastructure.

There are also users beyond the partnership. For many of them, charging may, however, be prohibitively expensive. This is especially likely to be the case with developing countries. The preamble of the *GBIF*-Memorandum of Understanding states that all primary biodiversity data should be in the open domain. So another option is that the shared information service should be funded as a public service. While it may be expected that digitising information is a national responsibility, maintenance of the European publicly shared infrastructure might be regarded as a responsibility of the European Union.

A first step towards creating a financial policy would be for all partners and the *gbif* national nodes to agree about the maintenance of the common infrastructure. Before approaching the European Commission for support, it would be essential to clarify the extent to which the partnership or the national *GBIF*-nodes or both might contribute financially.

## VIEWS OF NATURAL HISTORY INSTITUTES

84

Views of the larger European natural history institutes, which are the custodians of large specimen databases, were discussed and collated at a meeting of CETAF – the Consortium of European Taxonomic Facilities. It is clear that at present there is no common view on the issue of cost recovery to maintain an infrastructure of specimen databases. Very few believed that a commercial system would be appropriate where any user would have to pay for the data contents and the mechanism for delivery of the data. Most institutes argue that these kinds of data generally are made available through public funds, and hence these data should be available to the public without extra charging. And since the common information service transmits these public data, it was felt that this service should also be maintained with public funding and that further charging would also be inappropriate.

Yet since the natural history institutes themselves are also important users of the common service, there is an argument for expecting the participating institutions at least partly to share the costs. Running costs of maintaining a minimal infrastructure will be approximately €100,000 per annum. If the larger institutes are expected to contribute and cover all costs, this would imply an average yearly institutional contribution of €3,000 to €4,000. A number of institutions viewed this figure as prohibitive, pointing out that they already frequently reduce their subscriptions to other information sources, such as expensive scientific journals. Other institutes, however, expressed willingness to pay such a fee for a suitable service. A further suggestion was to match fees to institutional size. Another was to create a system to host and maintain the service composed of a small number of institutes.

## RECOMMENDATIONS

1. A European natural history specimen information system should exist to organise and manage an infrastructure (hardware, software, support) to enable the online sharing of digital specimen data. It is not concerned with the data contents, which is the responsibility of the contributing partners.
2. While these specimen data exist in the public domain, partners should be aware of restrictions on access to those data where there are legal implications for their distribution (e.g. localities of endangered species).
3. The infrastructure, which enables online access to distributed specimen databases, should be perceived as a joint undertaking of those partners that participate in the data-sharing.
4. The financial resources needed to maintain and update the infrastructure may originate from the partnership (also users), from national gbif-nodes (facilities that share biodiversity data), and from the European Commission (maintenance of a European public service with data in the public domain). With the digital data in the public domain, it would be inappropriate for the partnership to charge users for access.
5. A Memorandum of Understanding is an appropriate mechanism to organise and manage the infrastructure. The Memorandum of Understanding should prescribe:
  - ♦ The basis on which organisations can join the partnership.
  - ♦ The partner that will host the physical infrastructure.
  - ♦ The way in which the infrastructure will be managed by the host, including the guidance to be given by a small, elected, executive committee drawn from the partnership.
  - ♦ Common policies with respect to best practice and data policies.
  - ♦ Financial commitment of the partners.
6. The *BioCASE* project will establish the Memorandum of Understanding, and develop the business plan for a pan-European and operational system of distributed specimen databases. This plan, together with experience gained of running the service, will provide a realistic picture of the level of services and of costs.