UNEP

CBD

CONVENTION ON BIOLOGICAL DIVERSITY

Distr.
GENERAL

UNEP/CBD/BS/TE-BCH/1/3
4 August 2000

ORIGINAL: ENGLISH

MEETING OF TECHNICAL EXPERTS ON
THE BIOSAFETY CLEARING-HOUSE
Montreal, 11-13 September 2000
Item 3.2 of the provisional agenda*

OPERATION OF THE BIOSAFETY CLEARING-HOUSE

Note by the Executive Secretary

Contents

I. INTRODUCTION

1. The purpose of the present note is to assist the Meeting of Technical Experts in its consideration of the items in the work plan of the Intergovernmental Committee for the Cartagena Protocol (ICCP) required to make the Biosafety Clearing-House operational, with a focus on design features of an electronic processing unit at the core of the system. A short analysis is provided of information management issues (including common formats, data-input systems, and quality-assurance procedures), considerations in designing the system architecture, means to protect confidential data, and security issues.

* UNEP/CBD/BS/TE-BCH/1/1.

/…

II. INFORMATION MANAGEMENT ISSUES

A. *Information requirements*

2. The information-exchange system of the Biosafety Clearing-House will need to meet the information needs of a wide and diverse audience, including competent national authorities and national focal points of Parties and other Governments, international governmental organizations, national regulatory agencies, industry developers, non-governmental organizations, members of the public, and the Convention Secretariat.

3. The note by the Executive Secretary on the establishment of the Biosafety Clearing-House (UNEP/CBD/BS/TE-BCH/1/2) describes the types of information that will be processed by the Biosafety Clearing-House. The communication of information required to operate the advance informed agreement procedure and procedures for processing living modified organisms for direct use for food, feed or processing (LMO-FFPs) through the Biosafety Clearing-House will be its most important function in the early stages. Thus, the main characteristics of the Biosafety Clearing-House information-exchange system will be:

(a) To allow validated data to be submitted to the system;

(b) To store and/or provide access to this data;

(c) To present the data so that it can be easily found;

(d) To control access to confidential data in the system; and

(e) To protect the data submitted to the system.

B. *Common formats*

4. In order to allow for proper data browsing and querying of the database, it is essential that all reports submitted should share a common format—either a single common format or common open formats where the tools to read the documentation are freely available over the Internet. A lack of common standards will stunt the growth of an effective information-exchange mechanism, and the ability of the Biosafety Clearing-House to further develop and exploit information-exchange opportunities.

5. Therefore, agreement must be reached on common formats for sharing and exchanging information and data for inclusion in the Biosafety Clearing-House. Samples of common formats for data exchange, based on the requirements of the Protocol, will be presented to the Meeting of Technical Experts for further consideration.

C. *Data input*

6. Metadata (i.e., information about the data, such as the owner and content) will be required as part of the Biosafety Clearing-House to inventory what information is available in the system and where it is located. The content provider would supply the initial metadata that describes the information. It would then be valuable if automatic analysis of the information submitted through the database could allow some metadata to be created automatically (for example: dates of submission; keyword indexing). However, since the machine interpretation of data cannot as yet reliably answer common questions that users are likely to ask, the validation process will probably also need to allow a human operator to enter additional information.

7. In order to facilitate entry into a database, documents should be submitted in electronic format as far as practicable. The geographic extent of the interested area, the enormous differences in computer technology and the different organizations and regulatory frameworks that exist among countries suggest the implementation of a flexible system accessible to all users.

8. A few standard, widely used file formats, are discussed below. At the simplest level of data-exchange, ASCII (American Standard Code for Information Interchange) is the most common format for encoding text files in computers and on the Internet, although standard ASCII does not allow for complex formatting or use of diacritics. RTF (Rich Text Format) is a file format that enables the exchange of text files between different word processors, while Unicode is a relatively new system for interchanging written text in 24 supported language scripts.

9. HTML (Hypertext Markup Language) is a "presentation", or "formatting" language—a set of "markup" symbols or codes inserted in a file intended for display on a World Wide Web browser—and is a standard recommended by the World Wide Web Consortium (W3C). The current version of HTML is HTML 4, but HTML 3.2 is the most widely supported version. However, the major Web browsers (Microsoft's Internet Explorer and Netscape's Navigator) implement some features differently and provide non-standard extensions for the language.

10. XML (Extensible Markup Language) is a "data description" language and is currently a formal recommendation from W3C. XML is similar to the language of today's Web pages, HTML. Both XML and HTML contain mark-up symbols to describe the contents of a page or file. HTML describes the content of a Web page (mainly text and graphic images) only in terms of how it is to be displayed and interacted with. In contrast, an XML file can be processed purely as data by a program or it can be stored with similar data on another computer or, like an HTML file, can be displayed and therefore be used to exchange information with a database.

11. Proprietary file formats are also often used for documents such as those produced using popular word processors like Microsoft Word and WordPerfect. Also produced by a specific commercial application, but widely used on Internet, are PDF (Portable Document Format) files, produced by Adobe through Acrobat applications; its main commercial advantage consists in the free distribution via Internet of the "Reader" application.

D. *Common language*

12. An important issue in accepting submissions to the Biosafety Clearing-House will be the language of submission. The simplest solution for exchanging information through the Biosafety Clearing-House system would be the adoption of a compulsory unique language to be used in all information submitted to the Biosafety Clearing-House. (This may have resource implications for many regions, particularly with regard to the time-limits required under the Protocol for submission of certain types of information to the Biosafety Clearing-House.)

13. A functional alternative, but one that would be more limited for search and retrieval functions, would be to allow the submission of exhaustive documents in a common language, synthesizing and referring to original documents provided as annexes. More limited again would be to provide only abstracts and metadata in a common language.

14. Further consideration could be given to the use of a controlled vocabulary for keyword indexing of information to be processed by the Biosafety Clearing-House in a number of different languages. In the absence of a standard vocabulary it would be very difficult to compile meaningful datasets and information products, let alone exchange them in an efficient and harmonized manner. However, use of publishing reference tools may facilitate activities in this regard, for example, by using the UNEP multilingual thesaurus of environmental terms, EnVoc (Environmental Vocabulary). EnVoc is published in all six official United Nations languages (Arabic, Chinese, English, French, Russian and Spanish), and a number of other Governments have undertaken the translation of the thesaurus into their national languages.

E. *Content validation and quality assurance*

15. To be effective in an increasingly large-scale information-exchange service, content validation must proceed with minimal manual intervention. It is unlikely to be economically feasible for every piece of content submitted to the Biosafety Clearing-House to be checked manually. Therefore, metadata information will be needed to describe both the content and the validation criteria.

16. There are likely to be a number of parts to a validation service, for example:

(a) Syntactic validation checks the technical correctness of the content, for example that all links in a Web page are valid;

(b) Semantic validation determines if the content matter is correct in its current use context, for example, that this is the correct risk assessment for the LMO concerned. The answers to these questions are supplied by the content metadata, either generated automatically or else supplied by a human user. (It is to be hoped that intelligent media processing can replace the human user.)

(c) Additional metadata can be added to the content to record the results of the syntactic and semantic validation. (This may be particularly important if a human user has contributed to validation, since the same questions can then be answered automatically in future.)

(d) Finally, the content can be secured so that if it is transferred to another organization or process, the metadata can be trusted and potentially the content used without any further need for validation.

17. Nevertheless, each Party should be considered totally responsible for its own submissions.

F. *Data-reporting*

18. The reporting role of the Biosafety Clearing-House will be to:

(a) Make its information accessible to all users,

(b) Facilitate the process of both integrating and summarizing the information to the extent desired by decision makers and the public;

(c) Sift through this information to find that information specifically requested by decision-makers and facilitate getting the information to them; and

(d) Ensure presentation in a format that is clear and understandable to decision makers.

19. The reporting system should be characterized by transparency, accessibility, objectivity, reliability, high quality and rapid reporting of results.

III. SYSTEM ARCHITECTURE

20. A first important issue to consider in designing the system is related to whether information is maintained in a decentralized network or a central database.

A. *Decentralized network*

21. One option for the operation of the Biosafety Clearing-House would be to implement a decentralized information-exchange system, based on new and existing autonomous systems for storing and distributing data. If such an approach were to be adopted, the most important consideration would be technical inter-operability between the systems. It is likely that data would need to be both exchanged between different information systems, and also shared or “pooled” at a central location in order to achieve synergy and added value.

22. Ensuring technical inter-operability places detailed demands at multiple levels, which range from physical interconnection to correct interpretation by applications of data that is provided by other applications. For two information systems to inter-operate effectively, they must be able not only to exchange relevant information but also to interpret the information they exchange according to consistent definitions—merely providing information in digital form does not necessarily mean that it can be readily shared between systems. Inter-operability would also require that systems be inter-operable at the data level—that the format and semantics of the data are also coordinated so as to permit interoperation.

23. If desired, there are a number of approaches by which current autonomous systems, not designed up-front for inter-operability, could be made to inter-operate to exchange information:

(a) *The data "bus" approach.* Each system uses its own data definitions internally. However, exchanges of data with other systems are conducted through a "bus", that is, a common data standard into which data must be translated before being transmitted to another system. Any system wishing to use this data then downloads it from the "bus" and retranslates the data into locally meaningful terms before that data is used;

(b) *The data-dictionary approach.* Each system has a published data dictionary and a simple query-response mechanism to access the data with published message formats. Given a later need to inter-operate, another supplier could build to that embedded base interface and access the system's data. A system with this capability may cost more than a closed system, and additional security issues may need to be addressed with this approach;

(c) *The data-translator approach.* Two systems that need to inter-operate have a translator that converts one set of data definitions into the other. This approach preserves the internal integrity of the data, but the translators may be slow and, more importantly, may not preserve the original semantics of the underlying data;

(d) *The data-server approach.* Data and processing are separated. When a system requires data, it connects to a data server that provides the data. Enforcement of definitions can thus be limited to just a few servers rather than a myriad of applications. By moving the data into a system separate from the individual applications, this approach facilitates reuse of data in new, unanticipated ways.

24. Benefits of the decentralized model would include more timely data-sharing, as the original data providers would not have to go through this extra step of circulating data to a central repository. The result would be a need for fewer resources.

25. However, the establishment of such a system would involve major trade-offs between inter-operability and security. Inter-operability can promote an attacker's access to diverse systems, thus facilitating the rapid spread of attacks. In addition, ad hoc work-arounds to overcome a lack of inherent inter-operability could introduce many hard-to-manage security problems. Another trade-off is the potential for inter-operability problems posed by the introduction of new security features into part of a larger system of systems.

26. Apart from technical challenges involved in ensuring inter-operability, other disadvantages with establishing a decentralized system may be exclusive availability to Internet users, limits in access to strategic data where parts of the network have less reliable telecommunications infrastructure, and a significant increase in the time to carry out data queries. A decentralized system may also suffer from a lack of coordination in data reporting, quality assurance, and database management, making it difficult to combine data across systems and make regional information available quickly.

B. *Centralized database*

27. An alternative strategy to overcome limitations such as lack of coordination in data-reporting, and to increase data-quality assurance, could be to create a centralized database that contains all the core data submitted under the Protocol.

28. In addition to storing essential and official information in a central repository, it would perhaps be desirable to include, in the searchable data, all the relevant references (and links) to other optional information available on other systems. The synergy of the core database with the other information systems, maintained by the Parties or other international stakeholders, would contribute to develop a neutral, transparent, cost–effective, efficient, accessible and decentralized system, in harmony with the design of the clearing-house mechanism of the Convention (as discussed in the note by the Executive Secretary on the establishment of the Biosafety Clearing-House (UNEP/CBD/BS/TE-BCH/1/2)).

29. Problems could be encountered with such a system because of the need for data providers to turn over their data to a centralized database, a process that can be time-consuming. The process of making corrections to the centralized database is likely to be slower and may result in multiple versions of the same data set—one set on the data provider's computer system and a second in the centralized database.

30. These problems may be overcome by the use of a database-management system that would allow individual data collectors and data providers to manage their own data locally, while providing a centralized means of uploading the data into a larger database. These data could be fully protected by the data-management structure, and only the data provider would be permitted to make changes. Data in the centralized database would then be available for comprehensive analysis and reporting.

C. *Combination model*

31. A combination of systems may also be considered. Depending on its sophistication and design, a combined model may offer the Biosafety Clearing-House the necessary flexibility for better coordination of the submission of data while ensuring timeliness and links to complementary information distributed. Once the issues of security and validation of information are resolved, the system could be designed to deal with different types of data having different levels of confidentiality and validation needs. In this manner, through a combined model, it may be easier to target, administer and make available the data that is required for the development of the Biosafety Clearing-House.

IV. CONFIDENTIALITY CONSIDERATIONS

32. Article 21 of the Protocol provides for the notifier to identify information submitted under the procedures of the Protocol that is to be treated as confidential, and Parties are required to ensure that they have procedures to protect such information.

33. Article 21, paragraph 6, of the Protocol clearly defines the type of information that shall not be considered confidential:

(a) The name and address of the notifier;

(b) A general description of the living modified organism or organisms;

(c) A summary of the risk assessment of the effects on the conservation and sustainable use of biological diversity, taking also into account risks to human health; and

(d) Any methods and plans for emergency response.

34. Any other information therefore, could potentially be classified as confidential (with appropriate justification) and would need to be sufficiently protected when being circulated through the Biosafety Clearing-House.

35. In order to allow for maximum transparency of the Biosafety Clearing-House, and to ensure that only a minimal amount of information is classified as confidential, such information should not be mixed with non-confidential information and could be contained exclusively in separate files annexed to the main document.

A. *A sample approach for dealing with confidential information*

36. One example of confidential data being circulated through an information clearing-house is that relating to experimental field trials of genetically modified organisms conducted in the European Union under directive 90/220/EEC. A demonstration of the system will be given at the Meeting of Technical Experts.

37. The main operational procedures of the system are as follows:

(a) The competent authorities submit a notification to the Commission by registered mail. A fax or an electronic mail is addressed to the Commission from the competent authority as evidence that the notification has been sent;

(b) On receiving the notification, the Commission records the date of receipt on the document and also numbers the pages received. The competent authority submitting the document is informed by the Commission (by fax) of the date of receipt and the date of circulation to other competent authorities;

(c) The circulation will, as a rule, be carried out once a week.

38. The system allows for greater security if confidential business information is being distributed to the member States. Firstly, member States must comply with the general requirements for receiving confidential business information, namely, that all individuals (within the member States and within the European Commission) handling such information must have received clearance to do so and that dossiers must be stored in secure places and distributed by diplomatic bag rather than by post or courier. Finally, mail must be registered at all stages of circulation to allow the security officer of the European Commission to be informed of the whereabouts of the confidential information at all times.

39. An overview of the content of the database can be found at http://food.jrc.it/gmo/. When the project was established in 1991, it consisted solely of a system for the exchange of printed dossiers. In 1996, an electronic system was developed and implemented in all member States of the European Union. However, information was exchanged exclusively through the distribution of diskettes and not through electronic mail, while confidential business information continued to be distributed on paper in line with the original distribution mechanism.

40. As with Article 21 of the Protocol, article 19 of directive 90/220/EEC clearly defines the type of information in notifications that may or may not be kept confidential. At present, only a limited amount of information in the database (144 out of 1569 notifications (9.2 per cent)) is labelled as confidential business information. In most cases, the confidential business information is restricted to the molecular characterization of the insert. The fact that so little information is confidential does not necessarily mean, however, that all of the non-confidential information is readily accessible. The availability of this information is dependent upon national decision-making processes.

41. The member States of the European Union have expressed an interest in having not only electronic access to the information relating to small-scale field trials, but also dossiers submitted for authorization of commercial releases under part C of the directive. These dossiers very often contain

confidential business information and, therefore, a special system had to be set-up to allow for secure on-line access.

42. The two biggest security threats to this system are possibly unauthorized access to corporate assets (both from outside the network and from within), and the threat of damage and loss through viral infection. The following methods as a means to secure the system are currently being considered.

B. *Extranets*

43. One option for facilitating secure transfer of confidential information is to limit its exposure through use of an extranet*. An extranet is a private network that uses Internet protocols and the public telecommunication system to securely share part of an organization's information or operations with key stakeholders. An extranet can be viewed as part of an organization's intranet that is extended to specified users outside the organization.

44. An extranet requires security and privacy. These require firewall-server management, the issuance and use of digital certificates or similar means of user authentication, encryption of messages, and the use of virtual private networks (VPNs) that tunnel through the public network.

45. When considering secure data transfer via the network, it is important to understand that there are a number of security issues and risks associated with extranets. The keyword for all extranet applications is "sharing": sharing databases, sharing information, sharing documents, etc. It is also a tool for efficient collaboration, since extranet users can actively participate in the process of sharing information.

C. *Firewalls and proxy servers*

46. The most common method of securing an extranet system is through use of "firewalls". Firewalls are hardware/software combinations configured to control the information that can flow in and out of the extranet. All data passing in and out of the Internet is transmitted through "routers" and these play a major role in firewalls. Routers act as packet filters (i.e. filtering units of data) and, based on a set of rules established by the system administrator, the router will allow certain packets in but will reject the input of others.

47. Proxy servers are another important tool for the maintenance of extranet security. The proxy server acts as an intermediary between the extranet and the Internet. It evaluates all requests for information from an authorization database and, if the request is acceptable, the proxy contacts the Internet. The returning page also passes through the proxy server from the Internet. In this way, the proxy server can keep a record of all transactions, and provides a trail to track any kind of attack. The proxy server also shields the extranet from the Internet given that the only Internet Protocol (IP) address transmitted to the Internet is that of the proxy server. Against this background, individuals pretending to be legitimate clients and trying to capture IP addresses for a "spoofing" attack (pretending to be a legitimate client) are not able to "see" the originating IP addresses, which are hidden inside the network.

48. Firewalls and proxy servers are an effective "barrier" method of controlling the passage of information in and out of an extranet, but do not address the issue of maintaining data integrity before or

* **Intranets** are secured areas that utilize Internet and WWW standards and technologies to conduct internal communication and collaboration activities. Adopted by companies at a phenomenal rate, intranets have produced efficiencies for businesses that allow users to manage their organizations more efficiently and effectively "behind the firewall". **Extranets** represent the bridge between the public Internet and the private corporate intranet. Extranets connect *multiple* and diverse organizations on-line behind virtual firewalls, where those who share in trusted circles can network in order to achieve commerce-oriented objectives.

after transmission. They also cannot address the integrity of the individuals sending or receiving information although encryption and authentication systems are available for this purpose.

D. *Encryption*

49. Encryption is a sophisticated method of encoding or "scrambling" data so that the data can only be decoded or unscrambled by the party for whom the message is intended. While encryption is a very powerful method for securing data, it neither offers positive proof of the identity of the sender nor verifies whether or not information has been tampered with or somehow altered in transmission.

E. *Authentication*

50. Authentication adds another layer of security to the system by providing positive identification of the sender of the information. Traditional authentication systems include the widely used password authorization methods. However, in today's robust computing environment, more sophisticated methods of authentication are necessary to ensure the integrity of data and to eliminate or reduce the possibility of fraud.

51. Digital signatures or "digital IDs" have brought this level of sophistication to the computing arena. Digital IDs incorporate a public/private key pair that is generated and bound to a user's name (and other identifying information) by a trusted third party certification authority, which issues the digital ID to the user. This ID can be attached to an encrypted message to assure the recipient of the correct identity of the sender. It can also be installed in a Web browser to be used in place of a password dialogue for information and services that require membership or restrict access to particular users. Since the slightest change in a "digitally signed" document will cause the digital signal verification process to fail, this method of authentication also allows people to check the integrity of signed documents.

E. *Viruses*

52. Viruses are a major concern for the integrity of an extranet. An appropriate way to deal with this problem is to run virus-checking software specifically designed for extranets. This software operates on a server, and checks files for viruses as they are sent to the extranet. Files are only accepted if they are virus-free and are blocked if they appear to be contaminated.

53. There are at present a number of available hardware/software packages that provide extremely high levels of security. The ultimate choice depends upon a number of factors including the type of operating system, the cost, the number of users, the speed of access required, etc.

F. *Implications for the Biosafety Clearing-House*

54. It is technically possible to bring together a large number of users under a secure extranet, and this option could be considered for the exchange of confidential information through the Biosafety Clearing-House.

55. The cost for the design, development and maintenance of a system for exchange of information is very high, particularly in terms of human resources, and it is therefore imperative to design a system that is both functional and sufficiently flexible.

56. It is also essential for any information-exchange system to have a clear structure defined at the outset. This should include precise knowledge of the fields of information needed and the details of the information required. It is also important to design a streamlined system for the import of data and to train essential that all users in its use. The system must also be functional and user-friendly and should clearly meet the expectations of the user.

57. The amount of confidential information should be limited to the minimum in line with existing legal requirements. Eventually, a layered system could be developed that limits the number of authorized people who have access to all layers. The second layer would contain no confidential information and would be intended for personnel who need to work with the system without necessarily requiring knowledge of all the data.

V. ESTABLISHMENT OF A PILOT PHASE OF THE BIOSAFETY CLEARING-HOUSE

58. In its decision V/1, adopting the work plan of the Intergovernmental Committee for the Cartagena Protocol, the Conference of the Parties emphasized that it was a matter of priority to launch the Biosafety Clearing-House no later than the entry into force of the Protocol. However, the complexity associated with designing an information-exchange system to achieve a grand universal solution must not be underestimated.

59. Given the need for the Biosafety Clearing-House to be operational as soon as possible, the scope of the initial establishment phase will need to be smaller in scale and less complex, and should concentrate on implementing core activities of the Protocol. Consideration could be given to developing a pilot phase of the Biosafety Clearing-House, similar to that developed by the clearing-house mechanism of the Convention. The clearing-house mechanism's long-term programme of work and strategic plan (adopted by the fifth meeting of the Conference of the Parties, in May 2000) were based on the results of the independent review of the pilot phase.

VI. POSSIBLE ISSUES FOR FURTHER DISCUSSION BY THE MEETING OF TECHNICAL EXPERTS

60. The Meeting of Technical Experts may wish to further discuss the following issues under this item:

(a) System architecture design issues: for example, centralized versus decentralized systems, including the possible establishment of a central database for the Biosafety Clearing-House, receiving official submissions by Parties and other stakeholders and allowing references (and links) to other external, accessible information exchange mechanisms;

(b) Mechanisms for data input and validation, including control and development of metadata; election of standard electronic formats for documents submission compatible with the platform of choice; election of authorized methods of submissions; and definition of validating methods according to defined security standards;

(c) Authentication of contributions: such as definition of a list of acknowledged contributors, and the inclusion of a clearly labelled, public board for unacknowledged, relevant contributions;

(d) Issues associated with data-browsing and querying, such as selection of common language for submissions (and definition of the core information that must be provided in this language) and a classification system and document layout for standard submissions;

(e) Handling of confidential data: definition of security standards to protect the integrity of the system and definition of procedures to prevent unauthorized access to classified data.
