

ANNEX

REPORT OF THE SECOND INFORMAL WORKSHOP ON
INTELLECTUAL PROPERTY DIGITAL LIBRARIES (IPDL) STANDARDIZATION
(Geneva, February 25 to March 1, 2002)

BACKGROUND

1. In July 2000 the SCIT Plenary approved the establishment of a SCIT Task Force to support the IPDL Project (see document SCIT/5/10, paragraph 59). In order to enable the Task Force to begin its work, the Secretariat convened two informal standards workshops that included representatives of Member States together with industry experts.
2. The first workshop (Geneva, July 2001) conducted an analysis and recommended that upon concluding development of the WIPO IPDL system, the WIPO IPDL program should focus on the extension and coordination of standards, proof-of-concept validation projects, and related infrastructure to complement these standards and to support IPDLs of Member States.
3. The second workshop (Geneva, February 2002), the subject of this report, moves forward by recommending a specific set of simple preliminary steps, described below, that immediately raise the profile of the IPDL environment, stabilize its central mission of robust long-term access, and strategically position it to adapt to evolving technology and WIPO standards. In particular, these recommendations extend those of the first workshop by calling for establishment of a global service registry and development of simple, inexpensive prototype implementations to serve as proof-of-concept for the notions of persistent identifiers and resource location.

INTRODUCTION

4. The global IPDL envisions, as an important use of WIPO NET, a coherent set of IPDLs that globally distribute an appropriate balance of responsibility, authority, and independence among autonomous but interoperating national IPDLs. One application of the global IPDL is a user-friendly system that permits effective cross-searching of the different data collections provided in various IPDLs.
5. The use of agreed standards in the IPDL environment is critical as more IP offices wish to make their data available using alternative search systems or in a search environment more fully customized to suit local requirements and conditions. As foreseen by the SCIT, the burden of providing and maintaining timely, comprehensive access to centrally loaded data is too great to be borne by any one national or regional office or the International Bureau.
6. A fully standards-based approach is the most pragmatic way in which user-friendly interfaces to cross-collection IP search can be maintained in the long term. It permits data to reside as closely as possible to its most authoritative and up-to-date source. It also allows

multiple interfaces to be built and tailored to local or regional audiences, which makes it easier for sites to be independent of the specific features, service availability and funding situation of just one interface provider.

7. This document articulates the core issues to be resolved in establishing a coherent set of IPDLs to comprise the global IPDL, as well as the consensus recommendations of the workshop participants. Of central concern is reliable access and discovery using standards and common agreements, all built on each office's basic commitment to keeping its identifiers persistent.

SUMMARY OF THE DISCUSSIONS OF THE WORKSHOP

Policy for Persistent, Reliable Identifiers

8. Every library, digital or otherwise, needs a way to uniquely and permanently identify every object in its holdings. As a condition of inclusion in the global IPDL, we recommend that each local IPDL be required to state, as an organizational value, a commitment to pair each IP object accessible through the digital library with an identifier that will be permanently and uniquely associated with the object, such that the combination of local identifier and IPDL number will be globally unique. This is a service issue, not a technology issue. A short, written statement declaring organizational policy for maintaining these persistent identifiers is required of each IPDL.

Implementation of Resolution Service for Identifiers

9. In order to provide a basic, reliable addressing scheme for the global IPDL environment, the persistent identifier will always redirect to the object whenever it is moved, and will never be reassigned to a different object. This redirection capability amounts to little more than maintaining ordinary URL redirection tables. Individual IPDLs should implement access objects and related services using ARKs. For the global IPDL, these identifiers will have the form

`http://localhostname/ark:/ipdlnumber/localname`

For example,

`http://uspto.gov/ark:/12045/us8479931`

Recognizing that most IP documents already have well-defined identifiers based on WIPO standards, some guidelines for their construction are given in Appendix I.

WIPO IPDL Service Registry

10. In order to establish the Service Registry as specified in the report of the First Workshop on IPDL Standardization, we recommend that WIPO create and maintain a database by collecting machine-readable IPDL summaries from the individual IPDLs. We further recommend that one application of this database should be to automatically create a web-accessible document at a well-known web-based address. This database would become the definitive list of IPDLs and *ipdlnumbers*.

11. We believe that the global IPDL will ultimately need a dynamic, machine-accessible mechanism for discovering services available at individual nodes. Such a mechanism might be constructed following the W3C Web Services Description Language (WSDL) (W3C WSDL). Note that WSDL itself is not standardized by the W3C at this time, but it at least provides an XML-based format for describing network and IPDL services.

12. This registry should contain one record for each IPDL node. Each record will include an ARK that identifies the IPDL itself, the IPDL number, and the kinds-of-documents it serves. Over time, information concerning collections, content and service availability are likely to be added. See Appendix II for additional details. Individual IPDLs are responsible for the accuracy of the information, as they possess the master data record that WIPO periodically harvests. The machine-readable format for registry records is specified in Appendix II.

Partitioning of IP Domains

13. In order to simplify the task of working with the diversity of intellectual property, we adopt the principle of partitioning data into separate intellectual property domains. By considering one domain at a time (e.g., patents, trademarks, industrial designs, etc), we intend to focus on separate areas for which WIPO already has well-known semantics until such time as the results from several domains can be combined.

Network Access Protocols

14. Significant progress has been made in defining the vision and basis for the Global Intellectual Property Digital Library, following the work of the First Workshop on IPDL Standardization. This progress, when coupled with continued prototyping efforts, will demonstrate that implementing a digital library based on internationally recognized standards is an efficient and effective strategy for meeting the integrated information dissemination requirements of the IPDL environment. We recommend: (a) that the Simple Object Access Protocol (SOAP) be explored as the mechanism for the exchange of information between applications at the IPDL nodes (W3C SOAP); (b) that the content exchanged between nodes, as carried by SOAP, be constructed in XML (W3C XML); and (c) that the content exchanged between nodes as XML messages be encoded using Unicode.

Search and Retrieval Concepts

15. The following logical concepts will need to be considered in constructing an XML-based syntax for search and retrieval:

- Service Status
- User Identification and Authentication
- Search
- Retrieval
- Additional Services
- Confidentiality, including confidentiality of user search parameters

These concepts are largely captured in the XML Encoding Rules (XER) for representing search and retrieval data structures. XER is currently an ISO standard (ISO XER).

RECOMMENDATIONS

16. The participants in the Second Workshop on IPDL Standardization made the following five recommendations.

(i) *Commitment to Persistent Identification and Access for Core Collection Objects.* The distinguished factor between a digital library and a website is a service commitment to manage identifiers with enough care that a core collection object, once found and saved, can then reliably be retrieved using that same identifier indefinitely into the future. Such identifiers are also the basis for reliable, automated IPDL interoperation. As a condition of inclusion in the global IPDL, each local IPDL shall adopt, as an organizational value, a commitment to make each such object retrievable by users with a unique, permanently assigned identifier. Each IPDL shall maintain and publish an explicit statement of that commitment on its IPDL.

(ii) *Mechanics and Form of Persistent Identifiers.* Each IPDL shall establish procedures to ensure that each persistent identifier is never reassigned to a different object and is always correctly resolved no matter how often the object is moved. The technological burden of this requirement is low, but example techniques are discussed below. Each IPDL shall make its core objects available via URLs that conform to the Archival Resource Key (ARK) persistent identifiers scheme. The International Bureau should create and maintain a small, proof-of-concept centralized identifier resolution service that takes an incoming identifier request and forwards it to the appropriate IPDL.

(iii) *Registry of IPDLs in the Global IPDL System.* In order to create a formal, ongoing registry of IPDLs and services, each IPDL, during its lifetime, will make available a permanent *IPDL summary* at a fixed web-based address. This summary will always return an up-to-date, machine-readable description of the IPDL and its services. It will include at least the IPDL name, its assigned *ipdl number*, and the kinds of documents it serves. Over time, other elements will be incorporated to support additional inter-IPDL functionality. The IPDL registry itself is formed and updated by periodically harvesting all the IPDL summaries and collecting them in one place (e.g., as a dynamic document or a database). This establishes the most basic piece of infrastructure called for in the first workshop report.

(iv) *Prototype Search and Retrieval.* In order to successfully adapt existing standards for basic search and retrieval in the context of the global IPDL, the International Bureau and offices of Member States, possibly through a SCIT Task Force, must eventually support the development of a core set of common IPDL data elements and standards to support full IPDL interoperability. Prototyping cross-collection search is essential to this development. There is no relevant standard that fully meets the requirements for IPDL search and retrieval at this time, but the basic concepts are stated in the next section.

(v) *Standards Validation.* Fully analyzing a standard such as that envisioned for IPDL interoperation is a complex undertaking. To provide a measure of confidence without undue effort, we recommend adopting the validation threshold used by the Internet Engineering Task Force (IETF), namely, the successful interoperation of three systems developed independently from the written standard.

17. Information on definitions, naming considerations, IPDL descriptions and bibliography are contained in Appendices I to IV.

[Appendices follow]

APPENDIX I

DEFINITIONS

Archival Resource Key (ARK) - a scheme for creating identifiers that support persistence: access to the object, to metadata, and to policy statements from an identified provider. ARKs have the additional property of being valid URLs.

Digital Library (DL) - an electronic, network-based repository of objects, combined with defined mechanisms for locating, navigating and retrieving individual objects. Digital libraries agree to adopt a certain level of administrative commitment to offer and maintain data and services through the library. SCIT/1/5 defines a Digital Library as "an organized collection of electronic information disseminated to a designated community through network technologies providing easy access to data."

Distributed System - an electronic, network-based collection of sites distributed via the Internet, all of which are able to interoperate using commonly agreed-upon protocols and message formats.

Extensible Markup Language (XML) - a markup language for representing structured data.

Hypertext Transfer Protocol (HTTP) - the communication protocol implemented by World Wide Web servers and clients (or browsers).

Intellectual Property Digital Library (IPDL) - a digital library in which the content consists of IP objects. SCIT/1/5 states that a digital library provides "access to timely and complete collections of intellectual property records maintained by other intellectual property offices."

Internet Engineering Task Force (IETF) - the body responsible for Internet standards.

IP Domain - a specific type of intellectual property, including patents, trademarks, copyrights, industrial designs, traditional knowledge and others.

Global IPDL - the coherent collection of individual IPDLs which interoperate through common protocols, messages and content.

Identifier - An association between a name and an object which may be used to unambiguously refer to an IP object within the scope of the global IPDL, manifested by metadata that characterizes the object.

IP Object - the logical abstract notion of an intellectual property document, e.g., a patent, trademark document, industrial design, etc., the object is considered to be distinct from a particular representation of the document or the individual components of the document.

Metadata - Data which describe an IP object, and which might be used as a basis for locating or characterizing the IP object. Metadata might be included in the document itself, but not all metadata associated with an IP object will necessarily be contained in the document.

Protocol - any communications specification for the exchange of information between computers or computer applications.

Simple Object Access Protocol (SOAP) - a protocols specification for invoking methods on servers, services, components and objects.

Universal Resource Locator (URL) -the addressingscheme for the World Wide Web.

Web Services Description Language (WSDL) -an XML -based specificationschema for describing the operational information of a Webservice.

XML Encoding Rules (XER) -the ISO standard XML -based specification for representing the information in ASN.1 encoded data structures. The Z39.50 protocol is described in ASN.1, so XER gives an XML representation for the Z39.50 data units.

Z39.50 -the NISO standard communications protocol for search and retrieval.

[Appendix II follows]

APPENDIX II

NAMING CONSIDERATIONS

This Appendix discusses some of the issues that IPDLs may need to address in constructing ARK identifier names for IP objects.

Although there are benefits in terms of permanence that an ARK identifier gains from being semantically opaque, within the specialized realm of IP it may be worthwhile to trade it for limited semantic transparency. In particular, ARK identifiers that disclose, in their composition, certain key information elements to individual IPDL support offices can bring many advantages to system support staff whom may already be burdened by the commitment to persistence. This semantic transparency described next is not suitable for all providers, since it can be a help in some instances and a burden in others. Alternatively, there may be value in adopting identifiers from already existing standards (e.g., Uniform Resource Names [URN], Digital Object Identifiers [DOI]) and applying ARK functionality to them.

The following identifier components are based on WIPO Standards and common established practice in IP offices, most of which have individual collections limited to a single IP domain (such as patents). Within these collections, documents are generally identified and organized by means of the fields associated with INID Codes as described in WIPO Standards ST.9, ST.60, and ST.80. WIPO Standard ST.1 [ST.1] specifies the four ST.9 INID Codes [ST.9], which are both necessary and sufficient to uniquely identify a patent publication (ST.1 does not explicitly assign an order to these four fields). Thus, to create a transparent ARK containing these fields, the suggested form for a patent publication is of the form:

<http://ark.wipo.int/ark:/12345/pt/oc/kc/nnnnnnnnnn/ccyyymmdd>

where:

- 12345 is the IPDL Number (NAAN¹, in ARK terms),
- pt is the literal string "pt",
- oc is the ST.3 code of the issuing office or organization,
- kc is the ST.16 kind-of-document code,
- nnnnnnnnnn is the 1 to 10 numeric digit ST.6 publication number (ST.6 is currently under revision and the number of digits may be expanded in the future),
- and
- ccyyymmdd is the ST.2 date of publication.

For trademarks and industrial designs, and using fields described in ST.60 [ST.60] and ST.80 [ST.80], the recommended transparent ARK is of the form:

<http://ark.wipo.int/ark:/12345/domain/oc/nnnnnnnnnn/ccyyymmdd>

¹Name Assigning Authority Number

where:

- 12345 is the Name Assigning Authority Number (N AAN) of the IPDL,
- *domaii* is the literal string "tm" or "id",
- *oc* is the ST.3 code of the issuing office or organization,
- *nnnnnnnnnn* is the registration, application, or design document number, and
- *ccyymmdd* is the ST.2 date of publication.

[Appendix III follows]

APPENDIX III

IPDL SUMMARY DESCRIPTIONS

This Appendix proposes a machine-readable IPDL summary record. It will include at least the IPDL's name, its *ipdlnumber*, and the broad domains of coverage. It is retrieved with a URL of this format:

[http://localhostname/ark:/ipdlnumber/?](http://localhostname/ark:/ipdlnumber/)

The returned "document" is a plain text machine-readable record that contains a list of elements.

name:jpo
ipdlnumber:13045
domain:pt|tm|id

The elements are defined as follows. The *name* is a unique abbreviation in the IPDL community (similar to an airport abbreviation). The *ipdlnumber* is a globally unique ARK Name Assigning Authority Number. The *domain* element contains a list of domains covered by the IPO office, in any order; currently, it may assume any of the values

pt=patents
tm=trademarks
id=industrial designs
tk=traditional knowledge
np=non-patent literature
ot=other

Extensions are anticipated that will allow more detailed descriptions of each domain, such as numbers of items, date covered, search capabilities, and system status. For example, detailed information on the patent domain would be available at a URL of the form

<http://localhostname/ark:/ipdlnumber/pt?>

[Appendix IV follows]

APPENDIX IV

BIBLIOGRAPHY

ARK -The Archival Resource Key Persistent Identifier Scheme
(<http://www.ietf.org/internet-drafts/draft-kunze-ark-03.txt>)

DOI -Digital Object Identifier System(<http://www.doi.org>);the DOI Handbook
(http://www.doi.org/handbook_2000/index.html)

IPDLSD -Report of the Workshop on Intellectual Property Digital Libraries (IPDL)
Standardization (Geneva, July 2 to 6, 2001), Draft Version 1.309
(http://www.wipo.int/scit/en/ipdl/workshop/1/ipdlws01_06.htm)

ISOXER –X.693 Information technology –ASN.1 encoding rules:XML encoding rules
(http://www.itu.int/ITU-T/studygroups/com17/languages/X.693_0901.pdf)

SCIT/4/2 -Standing Committee on Information Technologies, Information Technology
Strategic Implementation Plan(http://www.wipo.int/scit/en/meeting/4/2_2.pdf)

SCIT/5/10 -Standing Committee on Information Technologies Plenary Report, Fifth Session
(<http://www.wipo.int/scit/en/meeting/5/10.pdf>)

SCIT/SDWG/1/9 -Standing Committee on Information Technologies, Standards and
Documentation Working Group Report, First Session
(http://www.wipo.int/scit/en/meeting/sdwg/1/pdf/sdwg1_9.pdf)

ST.1 -Recommendation Concerning the Minimum Data Elements Required to Uniquely
Identify a Patent Document(http://www.wipo.int/scit/en/standards/pdf/st_1.pdf)

ST.9 -Bibliographic Data on and Relating to Patents and SPCs
(http://www.wipo.int/scit/en/standards/pdf/st_9.pdf)

ST.60 -Bibliographic Data Relating to Marks
(http://www.wipo.int/scit/en/standards/pdf/st_60.pdf)

ST.80 -Bibliographic Data Relating to Industrial Designs
(http://www.wipo.int/scit/en/standards/pdf/st_80.pdf)

URN -Uniform Resource Name Syntax, RFC 2141, May 1997

W3C SOAP –XML Protocol Abstract Model(<http://www.w3.org/TR/xmlp-abstract/>); SOAP
Version 1.2 Part 0: Primer(<http://www.w3.org/TR/soap12-part0/>); SOAP Version 1.2 Part 1:
Messaging Framework(<http://www.w3c.org/TR/soap12-part1/>); SOAP Version 1.2 Part 2:
Adjuncts(<http://www.w3c.org/TR/soap12-part2/>) W3C XML –Extensible Markup Language
(<http://www.w3c.org/XML/>)

W3CWSDL -WebServicesDescriptionLanguage(WSDL)1.1
(<http://www.w3c.org/TR/wsdl>)

XER XMLEncodingRules(<http://asf.gils.net/xer/>)

ZURL -UniversalResourceLocatorsforZ39.50(<http://www.gils.net/z> -url.txt)

[EndofAppendixIVandofdocument]