STANDING COMMITTEE ON INFORMATION TECHNO LOGIES

STANDARDS AND DOCUME NTATION WORKING GROU P

Second Session
Geneva, December 2 to 6, 2002

IPDL STANDARDS DEVEL OPMENT

Document prepared by the Secretariat

1.   The SCIT Standards and Documentation Working Group (SDWG) has, as one of its ongoing tasks, the development of standards to enable all Member States to participate in a globally cohesive Intellectual Property Digital Library (IPDL) system (see document SCIT/SDWG/1/9, Annex III, Task No. 10). The SDWG also instructed the WIPO Secretariat to establish a SCIT Task Force to ensure the involvement of Member States in the expedition of these tasks. 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.   In accordance with the above -mentioned decision by the SDWG, the SCIT IPDL Task Force was established in July 2002 with the Secretariat as the Task Leader. The reports of the two informal workshops were presented to the Task Force for consideration. The report of the Task Force is found i n the Annex. For reference, the Report of the First Workshop is available online at   http://www.wipo.int/scit/en/ipdl/workshop/1/pdf/ipdlws01_06.pdf in the SCIT Documentation area, and the Report of the Second Workshop may be found in SCIT/7/6, Annex I.

3. *TheSDWGisinvited:*

(a)   toconsiderandapprovetheReportof theSCITIPDLTaskForcereproducedinthe Annextothisdocument;

(b)   toconsiderandadoptRecommendations 1through7intheReportoftheSCITIPDL TaskForcealsoreprodu   cedintheAnnexto thisdocument.

[Annexfollows]

ANNEX

INTELLECTUAL PROPERTY DIGITAL LIBRARY STANDARDS

A Report by the SCIT IPDL Task Force

Introduction

Two informal workshops concerning Intellectual Property Digital Libraries (IPDL) have been held by the International Bureau during the previous twelve months. The First Workshop considered the mechanism for developing and adopting IPDL standards, and the Second Workshop considered standards should be evaluated as a matter of priority.

This paper documents the results of deliberations of the SCIT IPDL Task Force concerning these reports.

Recommendations of the IPDL Task Force

The First and Second Workshops made a number of recommendations, as outlined in their reports. In general, the IPDL Task Force supports the findings of these workshops, and recommends that the SDWG accept these recommendations and where necessary, forward them to the SCIT for approval. In a few areas, the Task Force has found it necessary to make certain changes to these recommendations and expand upon them slightly. The recommendations of the Task Force, taking into consideration changes deemed necessary by the Task Force, are:

1. Establishment of an IPDL standards program as outlined in the First Workshop Report is recommended.

The First Workshop made the following recommendations:

(a) the existing WIPO Intellectual Property Digital Library (IPDL) program (aside from systems now moving into production) needs to be reoriented to focus on the development and coordination of standards and related infrastructure to complement and support Member State IPDLs;

(b) under the guidance and leadership of the International Bureau, a standardization program should move forward immediately for the development of IPDL-related standards, building on the existing body of standards already developed by the intellectual property community; and

(c) as a complement to the standardization program, modest-scale work should also proceed on proof of concept and standards validation demonstration projects and on infrastructure components to support these standards.

The fundamental goal of the proposed standards development effort is therefore to provide a basis for IP Offices to develop national IPDLs that meet national needs while also enabling these IPDLs to be integrated into the worldwide distributed intellectual property digital library. The standards should be sufficiently simple that they do not add significantly to new IPDL system complexity or cost.

The Task Force concurs in these recommendations. The Task Force agrees with the general level of funding and support indicated in paragraphs 32 to 34 of the First Workshop report, but considers that some development work in addition to prototyping should fall within the mandate

of this program. Such development work would include centralized services in support of a global IPDL environment (such as the IPDL Registry and centralized components of a persistent resource identifier system), and the development, collection, acquisition, and/or documentation of packages that implement standards agreed upon for use by participant IPDLs.

2. *A basic, non-binding set of policies and standards should be developed and recommended to Member States desiring to participate in a global IPDL environment*

A non-intrusive set of voluntary policies that regulate interactions between participating IPDLs is highly desirable. The rationale for such a set of policies and guidelines is clear; a standardized approach to the implementation of technical standards is as equally important as the technical standards themselves. These policies should outline a series of stepped levels of compliance which may be followed by participating offices. These steps could, for example, range from supporting access to data in standardized data formats, to supporting standardized search and retrieval protocols. Comprehensive degrees of compliance can be strongly encouraged by the SCIT, but are ultimately to be decided by individual offices. A basic level of compliance, for instance, could be required to certify a site as "IPDL compliant".

3. *A commitment to persistent identification and access for core IPDL collection objects is recommended*

The distinguishing factor between a digital library and a web site 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. This is of critical importance for collaborative work-sharing efforts involving the referencing of on-line data. Such identifiers are also the basis for reliable, automated IPDL interoperation. As a condition of inclusion in the global IPDL, the Task Force recommends that 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. The Task Force recommends that the SDWG request the development of an example statement for use by interested parties.

4. *A standard specifying the mechanics and form of persistent identifiers should be adopted.*

The Task Force recommends that each IPDL should 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 multiple schemes for implementing these procedures exist. The Second Workshop recommended that each IPDL should make its core objects available via URLs that conform to the Archival Resource Key (ARK) persistent identifier scheme. The International Bureau created a small prototype implementing this particular scheme and found it suitable for use within the IPDL environment. Despite this, the Task Force recommends that a more detailed review of at least two competing potential models for permanent resource identification be reviewed before a decision is made. It is thus recommended that the SDWG request that a simple prototype of the ARK scheme and the Digital Object Identifiers scheme be established by the International Bureau, and made available for review by the IPDL Task Force as quickly as possible, with a report to be presented to the SDWG at its 2003 meeting at the latest.

5. Establishment of a registry of IPDLs in the global IPDL system.

In order to create a formal, ongoing registry of IPDLs and services, it is recommended that each IPDL, during its lifetime, 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 *ipdlnumber*, 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.

6. 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 SCIT IPDL Task Force, must 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, and support for this activity should be provided through the mechanism outlined in Recommendation 1. There is no relevant standard that fully meets these core data element, attribute, and protocol requirements for IPDL search and retrieval at this time, but the basic concepts are stated in the Second Workshop report.

7. Establishment of a standards validation process.

Fully analyzing a standard such as that envisioned for IPDL interoperation is a complex undertaking. To provide a measure of confidence without undue effort, the Task Force recommends 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. Given the realities of the development environment facing the intellectual property community, a model wherein three separate implementations are developed using tools and standards developed by a single entity and approved as a standard implementation by the SDWG could also be effective.

Persistent Resource Identifiers

Given the recommendation of the Task Force for further research, some background discussion of persistent resource identifiers is in order. Persistent resource identifiers provide a potential solution to a problem facing intellectual property offices and consumers; how to "permanently" (that is, within the scope of a service commitment made by each office) provide electronic linkage to data collection objects. Document retrieval by permanent, standardized reference will permit the smooth integration of multiple IPDL's into a transparent global data store without imposing upon the desires of individual offices with respect to user interfaces, national language support, search engine selection, or participation in regional data publication systems. Recommendations 3 and 4, above, form the core of interoperability between IPDL's within the global IPDL environment and are of the utmost importance to future IPDL interoperation.

Current practice (e.g. the use of URL's) provides a transient solution at best, which is not suitable for long term record retrieval of the sort required for work sharing and long-term document access. Furthermore, the reality of the current situation within the IPDL environment dictates several requirements:

1. Implementation of a permanent identifier scheme must require no additional tools or training at the user level.

2. No office should be required to make changes to current IPDL systems. Changes may be *desirable* but cannot be *mandatory*.

3. Changes that offices desire to implement should be possible at a minimum cost.

With these basic requirements in mind, the Second Workshop recommended (Workshop Report Recommendation 2) the use of Archival Resource Keys (the ARK standard) as a persistent identifier. The Workshop also recommended that the International Bureau create and maintain a small, proof-of-concept centralized identifier resolution service. The IB has prepared this small prototype and has performed a basic evaluation of the ARK scheme using this prototype, and has found it to be both acceptable and suitable for use within the intellectual property environment. Time and lack of support has prevented an equal analysis of the Digital Object Identifier scheme, but should resources be available, it is the opinion of the Task Force that a clear and unbiased report based upon a detailed evaluation of these systems be developed and reviewed prior to a formal recommendation by the SDWG. It is the opinion of this Task Force that such an effort should be expedited to ensure that a report is presented at the following (2003) meeting of the SDWG.

Registry of IPDL's

The development of a registry of IPDLs is the one implementation task immediately recommended by this Task Force. The development of a registry of IPDLs is a very simple and basic task that can provide great benefit in the global IPDL environment. The availability of a simple, machine-readable record that describes the holdings of an individual IPDL node allows a degree of automated query preprocessing and routing to occur without any human intervention. This is important as it allows query systems to route queries only to sites that have a strong possibility of containing data relevant to the query at hand, and which provide search capabilities appropriate to the query presented.

The Second Workshop made a basic recommendation (Recommendation 3) for the structure and contents of this record. The IB has prototyped a retrieval mechanism for such a record using SOAP (Simple Object Access Protocol), an XML-based remote procedure call mechanism suitable for query-passing and data retrieval. The results of this prototyping are:

Such a machine-readable record can be quickly and easily created and published by any IPDL site, using completely free toolkits on a variety of hardware and software platforms. The IB prototyped the record retrieval mechanism using the SOAP-Lite toolkit for Perl, under both

Windows NT and Unix. Other equally simple toolkits are freely available for C, C++, Visual Basic, Delphi, and Java.

The record contents as outlined by report of the Second Workshop requires some modification. The minimum contents of the record require slight modification to make them acceptable for use by existing IPDL nodes. It is recommended that each data publishing office (irregardless of the media of publication) be assigned a standardized "office name" similar to the IPDL name outlined in the report. This office name will be associated with each of the intellectual property domains supported by an IPDL node. This will allow IPDL nodes to indicate that they provide access to data published by other offices.

T   The new record structure would be of the form

```
name:  epo
ipdlnumber:  13045
domain:  epo:pt | uspto:pt | fr:pt
```

This example record indicates that the European Patent Office (as designated by the name epo) provides access to epo patents (EPO:pt), uspto patents (USPTO:pt) and French patents (fr:pt).

1. It is strongly suggested that the SDWG accept and extend the recommendation of the workshop concerning this data record to include additional required information, such as publication date range and searchable fields, and formalize this record structure as a simple standard. A method of identifying "authoritative" data sources should also be indicated. The matter of searchable fields must be addressed concurrently with distributed search and retrieval modeling outlined in Recommendation 4, but not discussed in this report.

2. It is recommended that a Web Services Definition Language (WSDL) specification be developed and published for this SOAP service, and that the service support multiple methods for retrieving individual components of the registry record.

The utility of such a system was graphically illustrated by the simple "Get-A-Patent" demonstration developed by workshop participants during the First Workshop. This demonstration allowed the user to enter a country code and registration number, and retrieve the patent document, without any advance knowledge of where the data was stored.

Summary

The IPDL Task Force in general concurs in the recommendations made by the informal Workshops held by the International Bureau. Differences in the recommendations are noted above. Should the SDWG support these recommendations, this Task Force is prepared to support the continuing processes outlined in this document. Furthermore, the Task Force encourages the SDWG to recommend to the International Bureau that a mechanism be found to permit the evaluation and development work suggested within this document in as short a timeframe as possible, pending the establishment the program outlined in our Recommendation 1, above.

[End of Annex and of document]