

UNIFORM RESOURCE IDENTIFIERS
FOR INDUSTRIAL PROPERTY RESOURCES

*This non-paper has been prepared by a WIPO consultant
and does not necessarily represent the views of the International Bureau*

The International Bureau took the liberty of preparing the background information that is reflected in this non-paper in order to facilitate the discussions on Uniform Resource Identifiers (URIs) by the Standards and Documentation Working Group (SDWG).

(See paragraph 20 of document SCIT/SDWG/10/12.)

UNIFORM RESOURCE IDENTIFIERS INFORMATION PAPER

EXECUTIVE SUMMARY

Introduction

1. This paper has been prepared by a WIPO consultant and it does not necessarily represent the views of the International Bureau. The issue of Uniform Resource Identifiers (URIs) was raised by the ST.10/C Task Force at the tenth session of the Standards and Documentation Working Group (SDWG), in November 2008, in relation to the discussions on the revision of WIPO Standard ST.10/C relating to application and priority application numbers (SDWG Task No. 30). The Task Force requested the SDWG to comment and provide guidance on as to whether recommendations regarding URIs should be added to the revised version of WIPO Standard ST.10/C. The SDWG agreed that the Task Force should first focus on finalizing the revision of WIPO Standard ST.10/C in regard to the current recommendations. The SDWG also agreed, however, that further discussions on URIs would be necessary, at its next session in 2009, under an agenda item concerning SDWG Task No. 30. The International Bureau took the liberty of preparing the background information that is reflected in this non-paper in order to facilitate the discussions on URIs by the SDWG. (See paragraphs 15 to 22 of document SCIT/SDWG/10/12.)

2. The current executive summary document explores and summarizes some issues, benefits, and risks associated with the possible development of a scheme to identify industrial property content and to include the identifiers in industrial property documents, particularly, but not only, bibliographic patent data. This executive summary is supported by Annexes I (Industrial Property Identifiers – Supporting Information for Executive Summary), II (Definitions) and III (Informal Examples), which provide detail, illustration, definition, and background.

3. For the purpose of this paper, the term:

(a) Uniform Resource Identifier (URI) consists of a string of characters for identifying an abstract or physical resource. A URI can be further classified as a locator (e.g., Uniform Resource Locator (URL) which is often called an Internet address), a name (e.g., Uniform Resource Name (URN)), or both. URI is a phrase in common usage in the online technical world, but may not be so familiar to many industrial property practitioners;

(b) Digital Object Identifier (DOI) is another implementation of identification for which there is a managed system for persistent (permanent) identification of content-related entities on digital networks;

(c) Patent Object Identifier (POI) is the concept of a prospective patent related identification system loosely based on the DOI system, but dealing specifically with patent identifiers. POI is a term synonymous with IPI, but relates only to patents;

(d) Industrial Property Identifier (IPI) is a hypothetical term, proposed in this discussion paper, to describe an unambiguous, unique, permanent, and reliable identifier suitable for the industrial property world; and

(e) Registration Authority (RA) is a hypothetical term used throughout this paper to signify an entity (an organization) whose role would be to resolve IPI names into document locations. A Registration Agency is a real entity found within the DOI system, which performs the said role of registration and resolution.

4. The term URI is used in this Executive Summary since it is of generic meaning and was used by the ST.10/C Task Force in its request to the SDWG. The annexes to the current paper, however, refer to IPIs, rather than URIs because the term IPI conveys a meaning more specifically aligned with the business and functionality required, than does URI or POI. The term POI was considered less suitable because it pertains only to patent-related identifiers, which at this early discussion phase may be too exclusive. IPIs, rather than POIs, might be the candidate object of the discussions about a possible identifier of industrial property resources. It should be noted, however, that there is no predetermined identifier term at this preliminary stage of the discussions.

5. In the industrial property domain, the original question of URIs pertained to bibliographic patent data. Should other IP types (e.g., trademarks and industrial designs) be included in the scope of these discussions? Or would it be enough that the scheme was extensible to include other forms of intellectual property in the future? It makes sense to consider that POIs will likely need to be extended to all industrial property rights sooner or later.

6. The issue of the introduction of identifiers has been raised by the SDWG ST.10/C Task Force within the framework of the discussions to improve patent priority data. The industrial property world has noted the increasing problems the many and varied users of patent information have been facing in locating appropriate cited patent documents. Through the development of WIPO Standards, the use of consistently applied XML coding, improved data handling, and transfer mechanisms Intellectual Property Offices (IPOs) have tried to provide a reliable means for identifying and locating industrial property. But despite recent efforts, retrieving cited documents remains a problem.

7. The topic of integration of POIs, a prospective form of URIs, into the patent system is being discussed by the IP5 (CN, EP, JP, KR, and US Industrial Property Offices (IPOs)), whose aim is to eliminate unnecessary duplication of work among the IPOs, enhance patent examination efficiency and quality, and guarantee the stability of patent rights.

8. Most of the recommendations given in WIPO Standard ST.10/C refer to patents. However, it is possible that, in the future, similar recommendations dealing with trademark and industrial design are established. These new recommendations could be included, e.g., either in an extended version of current WIPO Standard ST.10/C or in one or more new WIPO standards that should be adopted accordingly. The question regarding whether a URI scheme for industrial property resources should be added to one or more existing WIPO Standards (e.g., Standard ST.10/C) or to a newly created specific standard should be addressed in the course of the discussions, but, probably, not yet at this stage.

Problems to be solved

9. WIPO Standard ST.1 provides recommendations to define the minimum data elements required to uniquely identify all types of patent documents whether published in paper or electronic form. At present, there is no need to revise the said Standard since it is still applicable, the industrial property world, however, lacks a standardized and reliable way to identify and locate specific (authentic) renditions of industrial property documents, which is particularly relevant to documenting and retrieving patent citations in a search report and in the detailed description.

10. Ambiguity can exist when a patent information user attempts to locate a “patent document”. Multiple renditions of a patent document abound. A single patent document typically is available on multiple websites, in different file formats, and in different language versions. Thrown into the mix is the possibility that the original document has been corrected or amended. It is not always evident to the reader of a citation what rendition (or rendering) was referred to by the creator of a citation some years earlier. Nor is it always clear which is the authentic rendition for legal purposes.

11. Furthermore the patent world particularly requires that a patent document, once published, must continue to be available in the public pool of knowledge. But the Internet Uniform Resource Locator (URL) system is prone to instability. Industrial Property Offices (IPOs) are reluctant to rely on non-standardized URL locations for identification, as they are prone (despite careful management) to “link rot.”

12. Not all industrial property content is available online on the Internet. Some authentic (official) patent documents are available on optical disc (e.g., DVD-ROM) and non-digital formats such as paper.

13. There is no common appropriate central standard document identification scheme available for use by IPOs (large and small) and the various types of users of the industrial property system.

14. A new URI scheme for industrial property resources and (partial) WIPO Standard must be easy and inexpensive to adopt, by IPOs, because the more IPOs that adopt and apply the Standards’ guidelines, the more likely further IPOs will participate and the more benefit is gained, by all affected parties.

Need for Uniform Resource Identifiers for industrial property resources

15. A carefully crafted URI scheme would have practical application to ease identification of industrial property content. An individual identifier name could be specified on search report citations, prior art background statements, and other forms of industrial property such as trademark applications. In addition, a URI scheme could be used to fulfill the need to present application numbers (filing and priority numbers) in a clear, unambiguous manner as provided for in WIPO Standard ST.10/C, the originator of this question within the SDWG context.

16. The purpose of a URI would be to provide unambiguous, unique, permanent, and reliable identification of a rendition of a patent or other industrial property document (or part thereof) existing on a network such as the Internet.

URI schemes (models)

17. Schemes (models) that may be suitable (or adapted to be suitable) for the industrial property world include:

(a) The Uniform Resource Locator (URL) is the current typical means of an IPO to indicate where content is located on the worldwide web. And it typically follows a hierarchical “directory like” path.

(b) The Uniform Resource Name (URN) which serves as a persistent, location-independent resource identifier following the rules of the URN framework. A URN (e.g., International Standard Book Number (ISBN)) is like a book's name, while a URL is like the address where it resides online.

(c) The Digital Object Identifier (DOI) system is a managed system (controlled by the International DOI Foundation (IDF)) for persistent identification of content-related entities on digital networks. DOI is commonly used to describe digital identifiers for NPL, e.g., doi:10.1006/jmbi.1998.2354.

18. The URL model does not seem to meet the minimum requirements for the URIs that we are looking for. The DOI and the URN are two initiatives attempting to define long term identifiers for information resources. These initiatives are related, in that they both try to overcome the limitations of the URL insofar as it is used to 'identify' resources on the Internet. The URL does not provide a stable, long term identifier; it simply provides the current location of the resource (or copy of the resource). If the resource moves, the URL changes.

19. The DOI and URN schemes require a registration authority (RA), preferably independent, to register and resolve names. RAs exist for other disciplines such as financial messaging (swift) and motion pictures (mpeg). There is currently, however, no RA within the DOI or URN framework that specifically caters for industrial property documents.

20. Although both the DOI and URN schemes could be adapted to meet the needs of the industrial property world, the DOI scheme is likely to take less effort, than the URN, to define and implement to meet the likely needs of a URI suitable for industrial property resources.

21. DOIs are already used by industry for the identification of non-patent literature (NPL) and may have advantages over the URN. The problem of content identification is not unique to the industrial property world. Indeed the scientific and technical publishing world has developed the DOI framework to address the similar set of problems for non-patent literature (NPL) that the industrial property world now seeks to address for patent documents, particularly in terms of patent citation provision and retrieval. The DOI system also helps prevent duplication of content and increase content authenticity.

22. Table 1 of Annex III to this paper provides a matrix comparing characteristics that may be required for an URI scheme for industrial property resources, the most important of which are discussed under the headings of Section “Problems to be solved” of Annex I. More information about DOIs and URNs can be found in the section entitled “IPI schemes (models)” in Annex I and the definitions in Annex II.

Elements for consideration

23. The following key characteristics would add great benefit and should be present in a URI scheme for industrial property resources. An indication of schemes that are most likely to provide the stated benefit is indicated in square brackets:

(a) **Permanent and stable names** could be provided by employing a resolution system whereby an RA would allow an issuing authority (an IPO) to register the name and location of an authentic industrial property document. If a user later requests the location for a specified registered name of a URI for industrial property resources, the location is returned to the user. The URI name would remain the same even if the location changes. [URN, DOI]

(b) **Unambiguous unique** identifiers, to adequately identify and locate authentic renditions of patent document content, would be provided by the RA employing an effective and consistent naming strategy according to the guidelines which would be provided in the WIPO Standards. Names could include self explanatory identification WIPO Standard ST.1 components, i.e.: country code, application number, kind-of-document code, and publication date amongst other components as considered necessary. [URN, DOI]

(c) **Authenticity** could be conferred by the attribution of a URI name for industrial property resources to a document.

(d) **Brevity** of a URI name for industrial property resources would be an asset, provided uniqueness was preserved. A brief URI name would minimize the name taking up too much space on the user’s computer screen, and output documents, as well as minimizing hyperlink line break errors that sometimes occur in email messages where the link name is too long.

(e) **Affordable** systems should be developed with the objective that there is no charge to the patent document user to access URI information (including locating the appropriate document.) In addition, the cost to the issuing authority (the IPO) to set up and interact with an URI system should be minimized to encourage maximum IPO participation. [URN, URL]

(f) **Robust, well established** URI systems for industrial property resources would add greater chance of success. [DOI]

(g) **Standardization** provided by a well drafted WIPO Standard and URI scheme for industrial property resources would increase clarity, transparency, and consistency for all industrial property users. [URN, URL, DOI]

(h) **Easy retrieval** of identified content could be achieved by the use of hyperlinks and identifier look-up pages provided by the issuing authority and RA respectively. [URN, URL, DOI]

(i) **Legacy IPO infrastructure** should be supported as much as possible. [URN, DOI]

(j) **Multiple media support** for common patent document platforms, including networked optical discs and the Internet, could be maximized. Paper may be also considered a suitable candidate to be given a URI. [URN, DOI]

(k) **Scalability, extensibility, and interoperability** as with any new initiative are highly valued. [URN, DOI]

24. A wide range of users including: IPO staff, technicians, patent information professionals in the private industry sector, and commercial providers would benefit from clear, consistent, unique, brief and stable identifiers used to streamline XML, citations, correction procedures, as well as bibliographic data elements.

Risk factors and compliance issues associated with the introduction of a URI scheme for industrial property resources

25. Some risks and compliance issues requiring careful monitoring and discussion are listed below, the most risky schemes (for each factor) being indicated in square brackets:

(a) Building a complex elaborate system without using an established framework. [URN, URL]

(b) Lack of a central independent organization, who is not itself an IPO, to fulfill the role of RA. [URN, DOI]

(c) Less affordable set up [URN] and maintenance costs. [DOI]

(d) Increased risk of compromising existing legacy URL naming systems. [URL]

(e) High IPO participation required for sufficient momentum for the scheme to succeed. [URN, DOI]

(f) Increased robot access activity. [URL]

(g) Difficulty in controlling amendments and changes to documents, particularly where an issuing authority did not follow paragraph 10 of WIPO Standard ST.16 (relating to corrected patent documents). [URN, URL]

(h) Difficulty to provide brief but meaningful URI strings. [URN, URL, DOI]

(i) Careful definition of what constitutes a registerable document. [URN, URL, DOI]

26. In addition, careful control should be considered of where in the WIPO Standards infrastructure to place the guidelines regarding URIs for industrial property resources, i.e., whether the recommendations should be in WIPO Standard ST.10/C or elsewhere.

Next steps

27. The question of the use of identifiers for patent documents is not a new one. But the timing for considering the adoption of a uniform identifier is starting to reach a critical mass as the digital IPO world grapples with issues relating to citation references, XML tagging, and correction procedures for which reliably identifying and locating patent documents pose problems for users.

28. Logical next steps to progress this issue of URIs for industrial property resources include:

(a) Considering whether it is convenient to create a task and the corresponding task force to investigate and agree on the scope, purpose, and solutions relating to the use of URIs in the industrial property world.

(b) Creating guidelines within one or more new or existing WIPO Standards.

(c) Determining a suitable URI scheme that can be used, created, or adapted to fulfill the requirements of the industrial property world.

[Annex I follows]

ANNEX I

INDUSTRIAL PROPERTY IDENTIFIERS

SUPPORTING INFORMATION FOR EXECUTIVE SUMMARY

Introduction

1. In the framework of the discussions on the revision of WIPO Standard ST.10/C, the ST.10/C Task Force and the Standards and Documentation Working Group (SDWG) of the Standing Committee on Information Technologies (SCIT) have briefly discussed whether the Uniform Resource Identifier (URI) should be added to the revised version of ST.10/C, particularly to paragraphs 5 to 7 relating to application numbers.
2. The SDWG in its tenth session considered that the discussion of URIs was a matter that would likely be dealt with by a separate task force (from the ST.10/C Task Force). The SDWG agreed that further discussions would be necessary at its next session in 2009 under an agenda item concerning the ST.10/C Task Force (dealing with Task 30). For more information, see document SCIT/SDWG/9/12 paragraph 21, SCIT/SDWG/10/2 paragraph 5 and SCIT/SDWG/10/12 paragraphs 16 and 20.
3. The term URI has been used in the Executive Summary since it is of generic meaning and was used by the ST.10/C Task Force in its request to the SDWG. For clarity and comprehensiveness, this Annex refers to *Industrial Property Identifiers (IPI)*, rather than URIs. The term Industrial Property Identifiers conveys a meaning more closely aligned with the business and functionality required, than does URI. The term *Patent Object Identifiers (POIs)* was considered less suitable because it excludes the notion of other forms of industrial property, which at this early discussion phase may be too exclusive. IPIs, rather than POIs, might likely be the candidate object of the discussions about a possible identifier of industrial property resources. It should be noted, however, that there is no predetermined identifier at this preliminary stage of the discussions.
4. The topic of integration of *Patent Object Identifiers (POIs)*, a prospective form of URIs, into the patent system is being discussed by the IP5 (CN, EP, JP, KR, and US), whose aim is to eliminate unnecessary duplication of work among the IPOs, enhance patent examination efficiency and quality, and guarantee the stability of patent rights.

Documentation Objective

5. This Annex provides discussion and detailed information to support the executive summary – the main document.
6. The discussion summary and accompanying Annexes provide a starting point for discussions during the SDWG eleventh session and beyond. The points raised are not proposed as set decision points for adoption by the SDWG, because they have not been fully discussed either by a designated task force or by the SDWG delegates.

Brief Definitions

7. An Industrial Property Identifier (IPI) is a hypothetical term, proposed in this discussion paper, to describe an unambiguous, unique, permanent, and reliable identifier suitable for the industrial property world.
8. A Uniform Resource Identifier (URI) consists of a string of characters for identifying an abstract or physical resource. A URI can be further classified as a locator (e.g., Uniform Resource Locator (URL) which is often called an Internet address), a name (e.g., Uniform Resource Name (URN)), or both. URI is a phrase in common usage in the online technical world, but may not be so familiar to many industrial property practitioners.
9. A Digital Object Identifier (DOI) is another implementation of identification for which there is a managed system for persistent (permanent) identification of content-related entities on digital networks.
10. A Patent Object Identifier (POI) is the concept of a prospective patent related identification system loosely based on the DOI system, but dealing specifically with patent identifiers. POI is a term synonymous with IPI, but relates only to patents.
11. A Registration **Authority** (RA) is a hypothetical term used throughout this document to signify an entity (an organization) whose role would be to resolve IPI names into document locations. A Registration **Agency** is a real entity found within the DOI system, which performs the said role of registration and resolution.
12. For more information about these and other related terms, please refer to Annex II.

Problems to be solved

13. Table 1 of Annex III to this paper provides a matrix comparing characteristics that may be required for an IPI scheme, the most important of which are discussed in paragraphs 16 to 30 under the italicized headings below.
14. WIPO Standard ST.1 provides recommendations to define the minimum data elements required to uniquely identify all types of patent documents whether published in paper or electronic form. The industrial property world, however, lacks a standardized and reliable way to identify and locate specific (authentic) renditions of industrial property documents, which is particularly relevant to documenting and retrieving patent citations in a search report and in the detailed description.

Ambiguity and Lack of Uniqueness

15. In terms of the current patent environment, components that can add ambiguity to identifying and locating patent documents include:
 - (a) similar numbered documents but from different countries or organizations (if no country code is married to the number),
 - (b) similar numbered documents from the same country but with different kind-of-document codes (e.g., A1 and B2),
 - (c) similar numbered documents from the same country, the same kind-of-document codes but relating to different renditions / file formats (e.g., PDF and HTML),

DRAFT

SCIT/SDWG/11- Ad Agenda item 5 - Industrial Property Identifiers

Annex I, page 3

(d) similar numbered documents from the same country, the same kind-of-document codes but for different language versions, and

(e) amended or corrected patent documents, with or without different kind-of-document codes, e.g., A1 and A9 (correction and republication of a published patent application).

16. As noted in the paragraph above, there are many variations of patent documents, including multiple document variants of the same document. Not only that, but the same document content can be published and available on multiple Internet sites. One motivation for duplicating content on multiple websites is to bolster the search capacity of an IPO. For example, a single WO patent application can be found on several IPO Internet sites (WO, EP, and SG) as well as many commercial provider sites. In addition, most Internet sites provide several format versions (renditions), e.g., HTML and PDF, for the same patent document.

17. Such a plethora of renditions of the same patent document, while often conveniently placed for the user, can lead to confusion as to which is the authentic document. In the case of WO patent applications, the PDF document on the WO web site is the authentic rendition and location. Authenticity is not always an issue for the user, but it does play a part when it comes to citations (particularly in legal proceedings) when ensuring that the creator and the reader of a citation are referring not only to the same document, but also to the same material within the patent document (e.g., is page 2 line 3 in one rendition on one site the same content as page 2 line 3 in a different rendition on an alternative site?) The multiple versions of equivalent or similar patent content relates to a lack of uniqueness (a kind of ambiguity).

18. Citations can be found in a variety of places including:

- (a) search report documents (kind-of-document code A3),
- (b) prior art statements by the applicant within the detailed description of a patent document,
- (c) bibliographic patent data on Internet summaries and first pages of a patent document,
- (d) within IPO internal databases, as well as
- (e) in private industry.

19. Increasingly, patent information users are asking themselves the following questions when they are monitoring, citing, and retrieving cited patent documents:

- (a) Does the content occur in more than one location (IPO site)?
- (b) Is more than one identifier reference used to identify the content, e.g., a permalink versus a longer URL?
- (c) Are there different versions of the patent document number (e.g., the original versus amended claims)?
- (d) Are there different renditions of the patent document, e.g., HTML versus PDF?
- (e) Which is the authentic version and rendition of the content?
- (f) Can the identifier, e.g., the URL, relate to more than one piece of content, e.g., the URL may be a list of multiple documents (A1, B2, A9, etc.) in different formats?
- (g) Where is the content located, if no URL is provided?

DRAFT

SCIT/SDWG/11- Ad Agenda item 5 - Industrial Property Identifiers Annex I, page 4

20. To be unique, an IPI should denote one and only one industrial property document. And ideally, one patent document should have only one formal IPI, for consistency.

Instability of identifiers over time

21. Persistence (permanence) is the requirement that content should always be available from the point of publication onwards, although not necessarily at the same location, unless there is a specific reason (e.g., a legal reason) why the content should be withdrawn or overwritten.

22. Published patent documents have a long life cycle. A patent document, once published, must continue to be available in the public pool of knowledge. This requirement of a patent document to be “open to public inspection” forever is a founding principle for the entire patent system and is likely to remain so.

23. Publication models are changing. It is an increasing trend for IPOs to switch their publication platform from physical means such as “paper” to digital publication, e.g., by use of URLs on the Internet.

24. An unfortunate trend of the Internet in general, however, is the instability of URLs. URLs are apt to change as documents and IPO systems evolve. Although URLs in the industrial property world are rather more stable (through careful management) than some other worlds, e.g., online newspapers, the use of URLs do sometimes attract “404 page not found” error messages, sometimes called “link rot.”

25. On the one hand the requirement of the patent world to publish patent documents available forever stands. But on the other hand, the mechanisms by which the inventions are published and kept available are changing. The provision of standardized IPIs allows an opportunity to stabilize and make consistent the identifiers for industrial property documents.

Resource limitations and compliance costs within IPOs

26. To formally introduce IPIs to bibliographic patent data (and elsewhere) would require human resource, time, and money, not only at the investigation and discussion stages, but also at the implementation stages and beyond. IPO resources are already stretched, so to participate in this new initiative, a substantially positive benefit needs to be evident for each IPO and its clients.

27. As with other WIPO Standards that aim to make the industrial property world more consistent and seamless, the more IPOs that adopt and apply the Standards’ guidelines, the more benefit that is gained, by affected parties. The more and the larger IPOs that ‘buy into’ a new standard, the greater the critical mass and increased likelihood of still further IPOs participating.

Content existing in various media types

28. Industrial property content can be found on various kinds of media (such as paper, DVD-ROM, and the Internet), and is likely to continue doing so for the foreseeable future. Although the trend for IPOs is to move away from the physical towards online digital media, an IPI system should optimally allow for a piece of content to exist in one of a variety of media types (e.g., DVD-ROM for patent document X but the Internet for document Y.) The industrial property community may consider that an IPI could be applicable to a paper document that was not on a network, as a kind of a placeholder, e.g., where the location of the

authentic paper document was recorded against the IPI name until such time as the patent document (collection) could be digitized and made available on a patent document network.

Control of the IPI scheme

29. Management of identifiers (including registration and resolution) are best controlled by a central, and preferably independent, RA. To have each IPO administer its own IPI scheme is possible, but not necessarily efficient.

Other important issues

30. The following characteristics, typical for new implementations in the digital online world require little explanation but should not be forgotten in our considerations:

(a) Interoperability - the ability of the system to work with other systems such as IPO networks and non-patent literature (NPL) identifiers without special effort on the part of the user,

(b) Easily retrievable – hyperlinking that is clear, quick, and locatable using minimal human effort,

(c) Scalability - identifiers that can be assigned to any resource that might conceivably be available on the network, in future years,

(d) Extensibility - permitting future extensions to the scheme,

(e) Legacy support - permitting the support of existing legacy naming systems, insofar as they satisfy other specified requirements.

31. Preferably IPI reference names should also be available to the issuing authority before publication date, so that the IPO can include the IPI within the bibliographic patent data, ahead of the publication schedule.

32. It is anticipated that there will be no cost to the user of an IPI.

Purpose of an industrial property identifier

33. When looking at the context of paragraphs 5 to 7 of WIPO Standard ST.10/C, a carefully specified URI (IPI) could be used to fulfill the need to present application numbers in a clear, unambiguous manner. The need for clarity and unambiguity applies equally to all presentations of application numbers of patent documents whether it be an application filing number or a priority document number.

34. An IPI would be useful, not only for application numbers appearing in bibliographic patent data (as prescribed by paragraphs 5 to 7 of WIPO Standard ST.10/C), but also appears to have practical application to other locations within the industrial property world including search report citations, prior art background statements, and other forms of industrial property such as trademark applications.

35. For a IPI to be a truly useful addition to the industrial property world, not only should the IPI be clear and unambiguous, the IPI should also be unique, easily retrievable, and continue to be available over time on a variety of platforms and networks including the Internet.

36. The purpose of an IPI in relation to industrial property-related content, particularly bibliographic patent data is suggested as a starting point for discussions: *The purpose of an*

DRAFT

SCIT/SDWG/11- Ad Agenda item 5 - Industrial Property Identifiers
Annex I, page 6

IPI is to provide unambiguous, unique, permanent, and reliable identification of a rendition of a patent or other industrial property document (or part thereof) existing on a network such as the Internet.

IPI schemes (models)

37. Schemes (models) that appear to be suitable (or adapted to be suitable) for industrial property include:

(a) The Uniform Resource Name (URN) which serves as a persistent, location-independent resource identifier following the rules of the URN framework. A URN (e.g., International Standard Serial Number (ISSN)) is like the journal name, while a URL is like the address where it resides online.

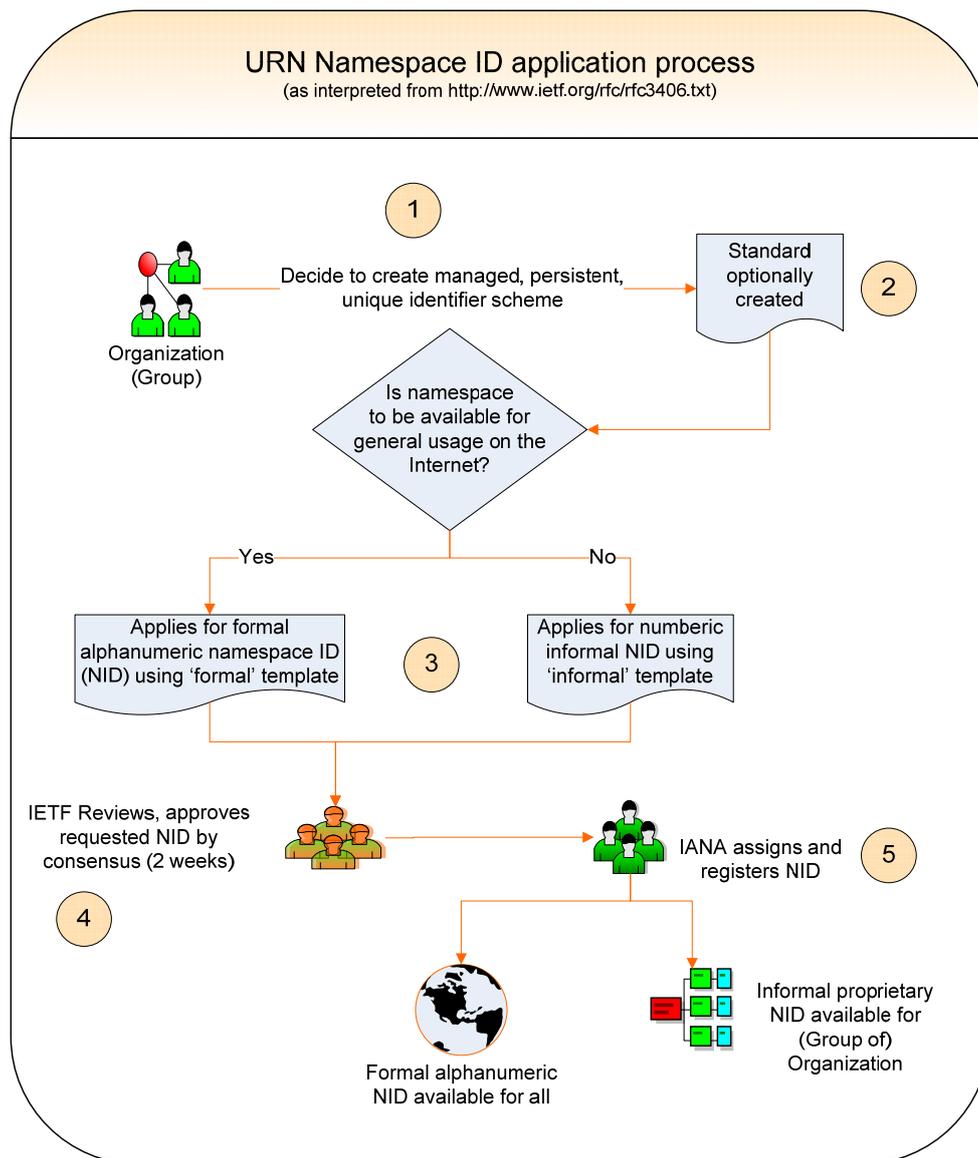
(b) The Uniform Resource Locator (URL) is the current typical means of an IPO to indicate where content is located on the worldwide web. And it typically follows a hierarchical “directory like” path.

(c) The Digital Object Identifier (DOI) system is a managed system (controlled by the International DOI Foundation (IDF)) for persistent identification of content-related entities on digital networks. DOI is commonly used to describe digital identifiers for NPL. The DOI identifier is a name not a location, so the DOI redirects the reader to the current location. Such redirection is reliant on RA mechanisms in place, in partnership with the IPO registrant, to ensure currency of the location of the content over time.

38. The DOI and the URN are two initiatives attempting to define long term identifiers for information resources. These initiatives are related, in that they both try to overcome the limitations of the URL insofar as it is used to 'identify' resources on the Internet. The URL does not provide a stable, long term identifier; it simply provides the current location of the resource (or copy of the resource). If the resource moves, the URL changes. No industrial property flavored identifier scheme exists within either the URN or DOI frameworks.

39. The URN scheme was created in 1992 as a standard for naming and identifying objects in a persistent, location-independent manner. The URN scheme was designed to keep the cost for providing gateways and using URNs as low as possible (comparable to URLs). The URN scheme allows an organization such as the International Standards Organization (ISO) or a group of organizations (such as an association of publishers) to register a formal or informal namespace ID (NID). A simple diagram depicts the general NID application process on the next page.

40. Formal NID application requirements, because they are not limited to users in specific communities or networks, and instead must be functional on and within the global Internet, are more stringent than for an informal NID application. For example, a formal application must show how the requested NID will benefit the wider Internet community. Although the applicant can request a specific NID for either the formal or informal NID, the informal NID can only be numeric.



41. Once a NID has been vetted by the Internet Engineering Task Force (IETF) and registered with the Internet Assigned Numbers Authority (IANA), a NID will form part of the URN syntax as can be seen in the following examples:

(a) *urn:ietf:std:50*

where *urn* denotes the scheme name, *ietf* is the NID, which in turn is followed by the namespace specific string *std:50*

(b) Another globally unique string representing an ISO related identifier is *urn:oid:1.0.3166.1* which represents the coding system published in the ISO 3166-1 Standard for country codes under the NID called Object Identifier (OID).

(c) The International Standard Book Number (ISBN) identifies an edition of a monographic work such as a book. On the right is depicted is a unique ISBN commercial book identifier barcode which can be resolved at <http://isbndb.com/>.



For more information about the URN namespace process see <http://www.ietf.org/rfc/rfc3406.txt>.

DRAFT

SCIT/SDWG/11- Ad Agenda item 5 - Industrial Property Identifiers Annex I, page 8

42. URN and DOI schemes require an RA, preferably independent, to register and resolve names, e.g., CrossRef is a popular registration agency for DOIs. An RA would need to be chosen and set up to take on the mantle of being the designated RA for IPIs. The RA would sit either within an existing (DOI) or a newly established (URN) framework. There is currently no Registration Agency within the DOI framework that specifically addresses the specific needs of industrial property documents.
43. DOIs can be resolved at <http://dx.doi.org/>. And URNs can be found by using a lookup tool provided by a registration authority that has registered the URN, e.g., you can look up an ISBN URN at <http://isbndb.com/>. An IPI scheme would require a look-up page which might relate only to industrial property or might be integrated with other disciplines (such as NPL). More information can be found in Annex II.
44. Although both the URN and DOI schemes could be adapted to meet the needs of the industrial property world, the DOI scheme is likely to take less effort to define and implement to meet the likely needs of an IPI. The DOI is a less general (than URN) and carefully defined scheme that works well for the scientific and technical NPL publishing world which has similar needs (to the industrial property world) in terms of citation provision and retrieval.

Benefits of using an IPI scheme

45. Table 1 of Annex III provides an informal comparison of values attributed to about 20 characteristics in relation to the three abovementioned possible IPI schemes.
46. IPIs would add clarity, consistency, brevity, and stability in the identification of digital industrial property resources. This benefits a wide range of users including: IPO staff, technicians, patent information professionals in the private industry sector, and commercial providers.
47. Of the readily available models, the DOI scheme is possibly the most pertinent, targeted existing model to solve the kind of problems, such as instability and ambiguity that the industrial property world faces. The DOI model is a well established scheme to identify items existing in other types of publications, such as journal articles (NPL) within the scientific and technical publication world. The IDF who control the DOI scheme would be willing to work closely with the industrial property world to set up an IPI registration agency within the DOI framework.
48. The DOI framework has a variety of associated monetary cost including the need to pay a small (a few cents) administration fee to the central IDF for each document registered. There are also usually costs payable to the Registration Agency, e.g., CrossRef, which typically are an annual licensing fee (per issuing authority, i.e., IPO) and / or individual document registration fees. Resolution for the user costs nothing. The URN model could be devised to include many of the positive attributes of the DOI system, but without the monetary costs. An industrial property URN would likely require more effort to design and implement than a DOI scheme.
49. For either the URN or the DOI, an RA would collect metadata information about each document suitable to manage and resolve data efficiently. Metadata that might be required would include the issuing authority (publisher) name, the location where the object can be found, and the date of publication amongst other parameters. While Table 2 of Annex III gives an indication of the kinds of parameters and examples that might be required, the

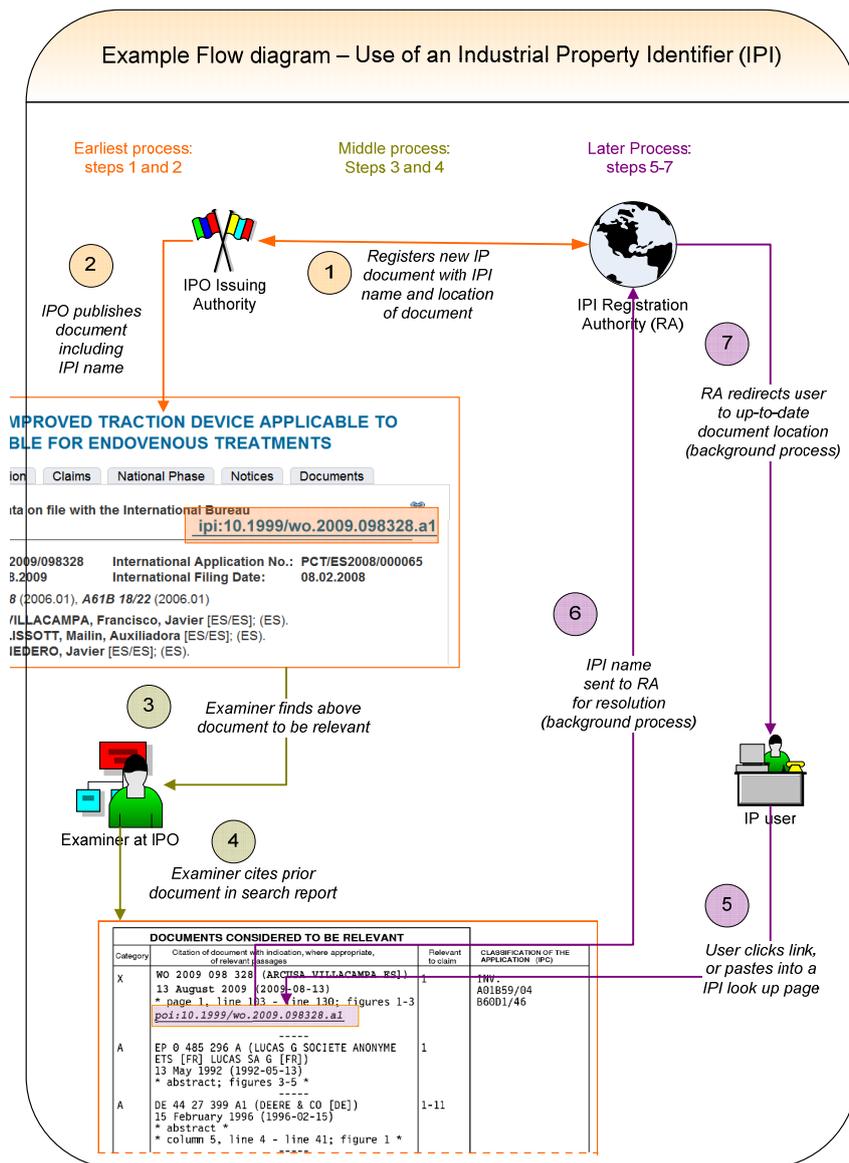
information presented in the table is only intended to give a flavor, rather than define the actual parameters to be used.

50. Platform independence could be maximized if the URN or DOI schemes were to be chosen, thus allowing CD-ROM, an Internet site, open web service or any other media type existing on the user's network to be referenced. Practically speaking, the referencing of an object (e.g., a patent document) on a publicly available Internet site is likely to be the most common scenario, now and in the medium-term future.

51. Defining IPIs can help the streamline or specification of other WIPO Standards aspects relating to XML, citations, correction procedures, and bibliographic data elements.

52. At a practical level a suitable IPI scheme could indicate an authentic source for a patent document, in a similar way that a DOI currently does for NPL. This could help obviate the citation difficulties that IPOs have identified (at least in the citation practices survey) in knowing which is the authentic rendition of a cited document, e.g., is it the PDF rendition on site A, or the html rendition on site B, or is it the paper copy on the examiners desk.

53. The following diagram illustrates how the mechanism of an IPI might work. The example is equally relevant for the URN and DOI schemes.



Risk factors and compliance issues associated with IPIs

Increased robot access

54. The provision of an IPI name according to specified predictable logical rules (e.g., if the country and number form the basis of the name) can increase the risk of robot access to patent information, if no mechanism such as *Captcha* is used to screen robots out. Robot access can cause bottlenecks in Internet traffic which can affect human user's access and document download times.

Organizational issues

55. Building an industrial property resolution system from nothing, i.e., without using an established framework such as DOI, is likely to be complicated. If URN were considered the appropriate model, a complete framework would need to be defined and set up. The URN 'green fields', start from the beginning, option would require substantial expertise, particularly during the definition and setup phases. IPOs may lack the human resources to staff the development.

56. It is unlikely that there currently exists an independent organization, who is not itself an IPO, which could fulfill the role of RA for an IPI system.

Affordability and likely uptake by IPOs

57. Significant investment is already put into systems by IPOs. The investment applies not only to systems currently in use, but also to ensure legacy system information is integrated from previous systems and in planning for future systems and modifications.

58. Changes, while they can bring considerable benefits, do require:

- (a) financial cost commitment,
- (b) human resource availability,
- (c) integration with interfacing systems,
- (d) marketing to and educating new and existing clients, and
- (e) quality management.

59. In addition to the points listed above, it is likely there will be an initial cost in terms of human and possibly financial resource to set up or modify an IPI system and WIPO Standard in terms of thrashing out the details, implementing, and persuading IPOs to participate in a new scheme and implement (part of) a WIPO Standard. The size, solvency, and agility of an IPO can have a bearing on the speed and fervor with which an IPO can implement a new (partial) system. The likely uptake (use) by (a substantial number) of IPOs of a new IPI system should be considered carefully before imposing a new system on overstretched IPO resources. Perhaps the industrial property community may consider there is a minimum critical number threshold of IPOs that intend to apply a new IPI system.

60. Nevertheless, experience has shown that inter-IPO task forces (such as the SDWG task forces) do work. The more delegates from member states that participate in the task force, the more robust the result is likely to be, and the more adherence to any pursuant standard (or part of a standard) there will be.

Ease of amending content

61. Existing identifier schemes such as DOI rely on content remaining stable over time. For example, a journal article once published is less likely to require amendment than a patent application. A patent application is subject to amendment and correction, sometimes more than once.

62. Any IPI scheme that is chosen or modified for the industrial property world will need to carefully determine what constitutes a new version, how to identify the changes, the potential need to charge change administration costs, and the interrelationship between the versions of content.

Risk factors associated with specific IPI schemes (not already covered above)

63. URLs are already widely used by IPOs. Many IPO online systems generate long URL strings which are difficult for humans to work with when generating citations, reports, and communications. Some IPOs have started to provide shortened “permalinks” which can be easily bookmarked and used by humans.

64. If URLs are chosen, consideration would need to be given as to whether the IPOs’ existing URLs would need to be modified or replaced. If so, this could place a burden on IPOs who have invested heavily in their existing URL management system.

65. If the URN or DOI was considered the appropriate model for IPIs, an appropriate RA and governing parameters would need to be defined and set up within the respective framework.

66. Additional financial costs are likely to be experienced if a new or existing centralized Registration Agency is set up or modified within the DOI framework to cater for industrial property-flavored identifiers. The DOI system does require that each document registered will attract a small administrative fee. It is likely that if an independent Registration Agency (that is not itself an IPO) was to be set up or modified, the administrative fee would be higher than if the Registration Agency was (controlled by) one or more IPOs.

Where can IPIs fit within the WIPO Standards system?

Is WIPO Standard 10/C the appropriate place for guidelines for IPI inclusion?

67. WIPO Standard ST.10/C deals with bibliographic data elements including the application number in paragraphs 5 to 7. A bibliographic data element in current praxis is used primarily as a synonym for "elements on the first page" of a published application. These days an HTML summary page is substantially the Internet equivalent of a first page traditionally published on paper.

68. Referring to SCIT/SDWG/10/2 Annex paragraph 11 page 3, the Task Force notes the following negative responses to the question “Should we include URI into WIPO Standard ST.10/C in the future?”

(a) Not all IPOs official patent publications are on the Internet, some official publications are in, e.g., optical disc format which do not lend themselves to the provision of IPIs.

(b) Officially, at the time of the preparation of the Annex to SCIT/SDWG/10/2 in October 2008, and noting that WIPO Standard ST.10/C is dedicated to bibliographic data

DRAFT

SCIT/SDWG/11- Ad Agenda item 5 - Industrial Property Identifiers Annex I, page 12

components of patent documents it was observed that no IPO includes a URI (IPI) as a bibliographic data component [to represent an application number]. Is WIPO Standard ST.10/C the appropriate place for IPI provision?

Interfacing with other WIPO Standards e.g. Citations

69. While URIs (IPIs) do not currently appear in bibliographic data as an application number. DOIs can, however, sometimes appear within citation references, particularly to relevant NPL publications. There is already an XML element <doi> available for this purpose.

70. As occurs in other WIPO Standards guidelines: cross-referencing between WIPO Standards and IPO practices (Part 7 of the WIPO Handbook) is commonplace. Whichever WIPO Standard is chosen or written to accommodate IPIs, the following interrelationships between identifiers and related WIPO Standards are anticipated: minimum data elements (ST.1), bibliographic data elements including Internationally agreed Numbers for the Identification of (bibliographic) Data (INID) codes (ST.9, ST.60, ST.80), citations (ST.14), kind of patent documents (ST.16), XML (ST.36, ST.66, and ST.86), and correction procedures (ST.50).

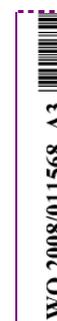
71. XML tags and INID codes may need to be modified or established to accommodate identifiers.

Other considerations concerning the scope and format of IPIs

72. The original question of URIs (IPIs) pertained to bibliographic patent data. Should other IP types be included in the scheme? And if so what should the scope include: e.g., trademarks, industrial designs, and copyright? Or would it be enough that the scheme was extensible to include other forms of intellectual property in the future?

73. If the guidelines for IPI provision are to be contained within the boundaries of WIPO Standard ST.10/C, what might be the characteristics that an IPO should use to present an IPI within the bibliographic context? For example, should the IPI name be hyperlinked to the correlating location? Should an INID code be created to precede the IPI or should the IPI be presented within the context of an existing INID code?

74. Bar codes are already used to identify patent applications, as can be seen in the following examples. Barcodes and IPI names could readily be combined.

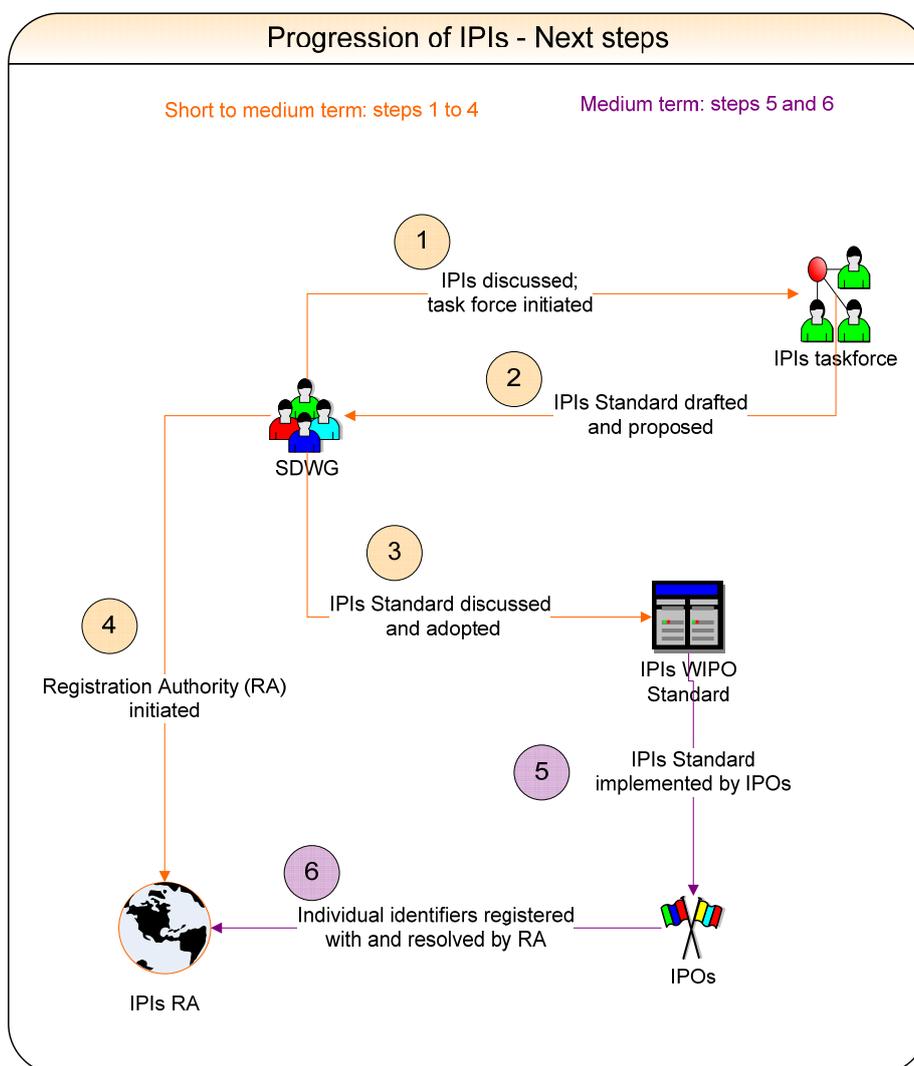


75. For global unique identification of a piece of industrial property content the following parameters should be considered for inclusion within the identifying name (or supporting parameters): issuing authority country code, kind of document, IP type (patent, utility model, trademark...), number, publication date, language of publication, and file type (PDF, HTML). A fine balance of including enough identification information in the IPI string while keeping the string as brief as possible should be sought.

76. Is it necessary for a user to recognize the nature of the content by following the logic of the IPI string? Or to avoid robots, and programmatic file downloads, perhaps a random illogical string may be an approach to minimize unwarranted patent document downloads.

Timing for implementation of IPIs

77. The question of the use of identifiers for patent documents is not a new one. But the timing for considering the adoption of a uniform identifier is starting to reach a critical mass as the digital IPO world grapples with issues relating to citation references, XML tagging, and correction procedures for which reliably identifying and locating patent documents remains a problem.



Questions for consideration for the IP World

78. In the framework of previous SDWG discussions it has been suggested that a separate task force be initiated to examine the role URIs (IPIs) can have to play in the world of discrete entities relating to industrial property particularly in relation to bibliographic patent data.

79. Such a task force could begin by asking themselves the following “who, what, why, how, when, and where-type questions:

- (a) What should be the overarching purpose for creating and using IPIs?
- (b) Who will and how will they investigate and make recommendations about the inclusion of IPIs in the industrial property system?
- (c) Is now the right time to be investigating this issue?
- (d) What problems (e.g., persistence, ambiguity, lack of uniqueness, platform independence, pointing to an authentic rendition of a document) are expected to be solved by IPIs?
- (e) What costs (financial, human resource) can be borne by IPOs in the endeavor to set up, maintain, and register industrial property documents within an IPI system?
- (f) Are there one or more appropriate places to include IPIs within the WIPO Standards?
- (g) Which type of IPI scheme (URN, URL, or DOI) would be the best model to fulfill the needs of identifying industrial property documents, particularly on the Internet network?
- (h) Which organization(s) would fill the role of RA to resolve identifier names (to document location), if needed?

References consulted

The following references were consulted during the period July to September 2009 in the preparation of the discussion paper and series of supporting Annexes. References are ordered by ascending alphabetical order of the URL Internet address.

- | | |
|--|--|
| 1. http://en.wikipedia.org/wiki/ | Definitions on Wikipedia |
| 2. http://searchsoa.techtarget.com/ | Definitions for technical terms |
| 3. http://www.dlib.org/dlib/february96/02arms.html | URNs - A Progress Report |
| 4. http://www.doi.org/ particularly
http://www.doi.org/overview/DOI-ELIS-Paskin.pdf | Information about DOIs |
| 5. http://www.hl7.org/oid/ | About ISO Object Identifier (OID)
Definitions |
| 6. http://www.iana.org/assignments/urn-namespaces/ | URN namespaces |
| 7. http://www.ietf.org/rfc/rfc3406.txt | URN Namespace Definition Mechanisms |
| 8. http://www.persistent-identifier.de/?link=204&lang=en | Persistent identifiers |
| 9. http://www.tm-xml.org/TM-XML/TM-XML_impl/TM-Search-TradeMarkList.xml | Example of an XML instance of a URL |
| 10. http://www.w3.org/TR/uri-clarification/ | URIs, URLs, and URNs: Clarifications and Recommendations |

[Annex II follows]

DRAFT
SCIT/SDWG/11 –Industrial Property Identifiers

ANNEX II

DEFINITIONS RELATING TO INDUSTRIAL PROPERTY IDENTIFIERS

The following definitions and acronyms are provided here in Annex II as supporting information for the main document and Annex I dealing with discussion points in the use of Industrial Property Identifiers (IPIs) and their possible relationship to and/or inclusion in patent bibliographic data or other industrial property. The information below has been separated from the main discussion points in order that experts in this field need not wade through material already known to them.

CAPTCHA

1. A *Captcha* is a type of challenge-response test used in computing to ensure that the response is not generated by a computer. The process usually involves one computer (a server) asking a user (via an Internet page) to complete a simple test. Because other computers are unable to solve the *Captcha*, any user entering a correct solution is presumed to be human.



A common type of CAPTCHA requires that the user type letters or digits from a distorted image that appears on the screen as illustrated above.

DOI (DIGITAL OBJECT IDENTIFIER)

2. The DOI system is a managed system for persistent (permanent) identification of content-related entities on digital networks, and is commonly used to describe digital identifiers for non-patent literature (NPL). The DOI is useful because if the object's location (e.g., Internet address location) changes, users will be redirected to the new address.
3. In the DOI model, a Registrant such as an IPO can submit a DOI to a centrally-managed directory and then use the address of that directory plus the DOI instead of a regular Internet address. Essentially, the DOI system is a scheme for Web page redirection by a central manager.
4. The DOI is a "digital identifier of an object", rather than an "identifier of a digital object". The DOI System is not solely designed for use on the World Wide Web; the same functionality can be made available through any digital network and protocol, but the Web demonstrates its advantages well. The DOI can be used e.g. for most scientific publications. The DOI can be resolved via HTTP (transformed into an URL) by appending <http://dx.doi.org/> or <http://hdl.handle.net/> in front. The syntax for a DOI is *doi:10.<publisher number>/<suffix>*, for example: *doi:10.1000/182*. The non-profit organization International DOI Foundation (IDF) governs the DOI system. For more information see http://www.doi.org/handbook_2000/intro.html.
5. A DOI name prefix (for example, 10.1000/) enables a registrant to assign many DOI names, by building on the prefix to construct a range of unique identifiers (10.1000/abc, etc). To obtain a prefix, one needs to work either with a Registration Agency or, for experimental or prototype purposes, with the International DOI Foundation.

IDF (INTERNATIONAL DOI FOUNDATION)

6. The Digital Object Identifier system is managed by the International DOI Foundation (IDF), an open membership consortium including both commercial and non-commercial partners, and has recently been accepted for standardization within ISO. Approximately 40 million DOI names have been assigned by DOI System Registration Agencies in the US, Australasia, and Europe.

INTEROPERABILITY

7. Interoperability refers to the ability to use an identifier in services outside the direct control of the issuing assigner: identifiers assigned in one context may be encountered in another place or time without consulting the assigner. For example, a customer may order a book from a bookseller or a library system by quoting its ISBN, without consulting the publisher who assigned the number.

ISBN (INTERNATIONAL STANDARD BOOK NUMBER)

8. The International Standard Book Number (ISBN) is a unique numeric commercial book identifier assigned to each edition and variation (except reprintings) of a book.



LINK ROT

9. Link rot (or linkrot) is the process by which links on a website gradually become irrelevant or broken as time goes on, because websites that they link to disappear, change their content, or move to new locations.

DRAFT
SCIT/SDWG/11 - Industrial Property Identifiers
Annex II, page 2

NPL (NON-PATENT LITERATURE)

10. NPL (Non-patent literature) refers to any documents that are not patents, typically when citing content in a patent search report. NPL includes web site pages, technical journal articles, databases, textbooks, and conference proceedings.

OPTICAL DISC

11. An optical disc is an electronic data storage medium that can be written to and read using a low-powered laser beam. Optical disc formats include CD-ROM and DVD-ROM.

PERMALINK

[See also the definition for [Link Rot](#)]

12. A permalink, or permanent link, is a URL that points to a specific Internet page (or content) after it has passed from the front page to the archives. Because a permalink remains unchanged indefinitely, it is less susceptible to link rot.

PERSISTENCE

13. Persistence is the requirement that once assigned, an identifier denotes the same material indefinitely. For example, ISBNs, once assigned, are managed so as to reference the same book always (and not be reassigned). Persistence can be considered to be "interoperability with the future." Persistence is also known as stability and permanence.

RA (REGISTRATION AUTHORITY)

14. In the context of the current documentation a registration authority (RA) is a hypothetical term used throughout this document to signify an entity (an organization) whose role would be to resolve IPI names into document locations.

REGISTRATION AGENCY

15. A Registration Agency, is a real entity within the DOI system, performing the said role of registration and resolution, where DOI names are registered by clients (Registrants) via a Registration Agency, such as CrossRef, which specializes in web citations of scholarly non-patent literature (NPL).
16. The primary role of a Registration Agency is to provide services to Registrants - allocating DOI name prefixes, registering DOI names and providing the necessary infrastructure to allow Registrants to declare and maintain metadata and state data. Registration Agencies are free to set fees for assigning DOI names and then in turn forward a portion (of the order of a few cents per DOI name) to support the central activities of the International DOI Foundation (IDF).

RENDITION

17. There may be multiple published (and unpublished) file formats / renderings / renditions of the same version of a patent document. For example, many IPOs provide the descriptions and claims of patent documents in machine-readable full text (HTML, XML) as well as page-based image formats (PDF, TIFF). Each instance is a rendition.

RESOLUTION

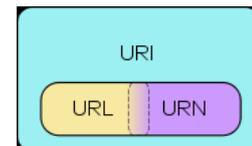
18. Resolution is the process in which an identifier sent to a resolution service to receive in return specific output of one or more pieces of current information (typically location) related to the identified entity. A familiar example of resolution is the Internet Domain Name System (DNS) which resolves a domain name address (URL) to a file residing on a specific host server machine. Another example is an ISBN bar code on a book which is scanned by a bar code reader in a bookshop which resolves to some point of sale information, such as title and price.

URI (UNIFORM RESOURCE INDICATOR)

[See also the definition for [URN](#)]

19. A Uniform Resource Identifier (URI) consists of a string of characters used to identify or name a resource on the Internet. Such identification enables interaction with representations of the resource over a network (typically the World Wide Web) using specific protocols. Schemes specifying a specific syntax and associated protocols define each URI. See http://en.wikipedia.org/wiki/Uniform_Resource_Identifier for more information.
20. A URI includes both Uniform Resource Locators (URLs) and Uniform Resource Name (URN).

A URN is like a person's name, while a URL is like their street address. For example using a topical, but fictitious example, "you can find urn:wipo:st:10c (URN) over at <http://www.wipo.int/standards/en/pdf/03-10-c.pdf> (URL)." You can translate (resolve) a URN into a URL using a URN resolver (tool) such as provided at <http://nbn-resolving.de/> for content in the National Library (NBN) registries. See http://en.wikipedia.org/wiki/Uniform_Resource_Name for more information.



21. The different components (that is the syntax) of a URI has four parts:
<scheme name> : <hierarchical part> [? <query>] [# <fragment>]
See http://en.wikipedia.org/wiki/URI_scheme#Generic_syntax. In a real URI example <http://www.wipo.int/pctdb/en/wo.jsp?WO=2009098328> taken from the WIPO PCT patent database:

- (a) The <scheme name> is *http*. *http*, in this case, is also the type of protocol used. Other examples are *ftp*, *mailto*, and *urn*. A colon : separates the scheme name from the hierarchical part.

DRAFT
SCIT/SDWG/11 - Industrial Property Identifiers
Annex II, page 3

- (b) The <hierarchical part> of the URI holds identification information hierarchical in nature. In the example //www.wipo.int/pctdb/en/ is the entire hierarchical part. Usually this part begins with a double forward slash ("//"), followed by an authority part and an optional path.
- i) The www.wipo.int represents the host name of the <authority> part. Other authority parts can optionally include optional <user information> and an optional <port number>.
 - ii) The /pctdb/en/ is the <path> part which is a sequence of segments (conceptually similar to directories) separated by a forward slash ("/")
- (c) The wo.jsp?WO=2009098328 is an optional <query> part where the <key> wo.jsp and <value> WO=2009098328 are separated by a question mark. The query contains additional identification information which is not hierarchical in nature.
- (d) The <fragment> is an optional part separated from the front parts by a hash ("#"). It holds additional identifying information that provides direction to a secondary resource, e.g. a section heading in an article identified by the remainder of the URI. There is no fragment in the example given.

22. Some URI schemes are registered with the Internet Assigned Numbers Authority (IANA), e.g., http, https, ftp, ldap, and urn. Other URIs are not registered with the IANA but are in common usage, e.g., doi, skype, and sms.

URL (UNIFORM RESOURCE LOCATOR)

[See also the definition for [URI](#)]

23. A URL (Uniform Resource Locator, previously called Universal Resource Locator) is the unique address for a file that is accessible on the Internet. A common way to get to a Web site is to enter the URL of its home page file in your Web browser's address line.

URN (UNIFORM RESOURCE NAME)

[See also the definition for [URI](#)]

24. A URN (Uniform Resource Name) is an Internet resource with a name that, unlike a URL, has persistent significance - that is, the owner of the URN can expect that someone else (or a program) will always be able to find the resource. A frequent problem in using the Web is that Web content is sometimes moved to a new site or a new page on the same site. Since links are made using Uniform Resource Locators (URLs), they no longer work when content is moved.

[Annex III follows]

DRAFT
SCIT/SDWG/11 - Industrial Property Identifiers

ANNEX III

Table 1: INFORMAL COMPARISON OF IDENTIFIERS SCHEMES

The following indicative matrix roughly compares characteristics of three possible identifier schemes that could form the basis for a patent or industrial property flavored IPI scheme for identifying patent and possibly other industrial property content. Each scheme would require at least some definition and adaptation before implementation. The estimated values attributed to each characteristic of each scheme type have been colored green to represent the most desirable value, red the least desirable, orange in the middle, and blue is not known or not applicable. Italicized values have either been estimated or will depend on the decisions made later during the process of defining a suitable scheme.

	Identifier scheme type / characteristic	URN namespace	structured URL	DOI - patent variant
1	persistent - content available even if material is moved, rearranged, or bookmarked	TRUE	NO	TRUE
2	unique - one string denotes one and only one entity, e.g., specific patent application rendition	TRUE	Possibly	TRUE
3	ongoing general affordability	Medium	Most affordable	Least affordable
4	likely uptake by large IPOs	Highest	Medium	Highest
5	likely uptake by smaller IPOs	Medium	Highest	Low
6	legacy support - the scheme must permit the support of existing legacy naming systems	TRUE	Probably not	Possibly
7	platform independent , e.g., can be used for Internet or DVD-ROM	Not known	NO	TRUE
8	compatible with XML	TRUE	TRUE	TRUE
9	original document easily amended , e.g., to remove a patent claim after acceptance	Probably	TRUE	Possibly
10	administrative costs to set up scheme	Highest	Low	Medium
11	requires resolution (redirection) mechanism by an independent organization	TRUE	NO	TRUE
12	suitable registration agencies already exist	NO	Not required	NO
13	interoperability with other data from other sources, e.g., non-patent literature	Possibly	Possibly	TRUE
14	likelihood of needing to restructure existing naming conventions	Low	Highest	Low
15	clear and unambiguous identifiers	TRUE	Possibly	TRUE
16	extensible - by adding new features and services	TRUE	Possibly	TRUE
17	scalable - permit future extensions to the scheme	TRUE	Probably	TRUE
18	supports granularity , e.g., a whole document or a single table in a patent document	TRUE	TRUE	TRUE
19	supports functional granularity , e.g., English or French version (or, e.g., HTML or PDF) of the same document / creation	TRUE	TRUE	TRUE
20	subject to robot attack	Probably not	TRUE	NO
21	identifier comprehensible to human	Possibly	Probably	Possibly

DRAFT
 SCIT/SDWG/11 - Industrial Property Identifiers
 Annex III, page 2

Table 2: INFORMAL EXAMPLES OF POSSIBLE METADATA REQUIREMENTS FOR A RESOLUTION SYSTEM

The following indicative table provides some possible metadata examples to illustrate how identifiers might be described within a resolution system. The column with a blue background represents fictional values

Meta data type	Real DOI Example	Theoretical future example	What is the meta data parameter describing?
identifier	doi:10.1006/jmbi.1998.2354	ipi:10.1999/wo.2009.098328.a1	By what unique names is the object known?
description	Journal of Molecular Biology, Volume 285, Issue 1, 8 January 1999, Pages 1-32	WO/2009/098328 (A1)	How is it described?
location	http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6WK7-45KNCTG-...	http://www.wipo.int/pctdb/images1/PCT-PAGES/2009/332009/09098328/09098328.pdf	Where can it be found?
issuing authority	ScienceDirect	WIPO	Who published it?
date	8 January 1999	13 August 2009	What was the original publication date?
format	html abstract	pdf online	What kind of file format is it?
language	English	Spanish	What language was it published in?
category	Journal article	patent application	What type of object is it?
context	Original publication	A1 kind of document	What has happened to it?
modified date	8 January 1999	13 August 2009	When did the content last change?
registration agency	CrossRef	Industrial Property Identifier Registration Agency	What is the central organization who maintains the list of identifiers?

[End of Annex and of documentation]