D6.2 Annex II

Review of existing persistent unique identification systems

ARROW PLUS Interim report on image identification

Project Acronym: **ARROW Plus**

Grant Agreement number: **270942**

Project Title: **European Registries of Rights Information and Orphan Works**

Revision: 1

Authors:

Sylvie Fodor, Executive Director, CEPIC

Angela Murphy, Consultant to CEPIC

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
PU	Public, for wide dissemination	x
CO	Confidential, only for members of the consortium and the Commission Services	
RE	Restricted to a group specified by the consortium	

Table of Content

2.4. Identifying Creators

The Visual Creator Index

Creator's Index

2.5. Standardised Image Identifiers

The DOI System

The DOI Name

Executive Summary

Background

The proliferation of images in today's digitally-connected world has had a profound impact on the way in which we use, manage, distribute and archive images. Today images are not only reproduced in analogue media, such as books, magazines, marketing literature, newspapers and academic publications, but also in all kinds of digital publications - from e-books to websites. This, in turn, has made the way in which we discover, manage and track rights information immensely complex in a world in which digital copies can be made at the click of a mouse. In this process, picture libraries and agencies have an increasingly important role to play. They not only hold extensive databases containing rich information about both image creators and the digital objects that they create, but they are also experts in researching rights holders and licensing image rights.

Taking photographs at a wedding, Arles, France (Angela Murphy)

Scanning transparencies and slides (Artix)

Scanning transparencies and slides (Artix)

Getty Images Search Scree

Aim of the Project

The Arrow Plus Project was designed as an extension of the preceding Arrow Project whose aim was to find a solution to the mass clearance of publishing rights in books to assist mass digitisation projects in libraries. The Arrow Project devised a system for the mass clearance of book rights but did not include image rights clearance. Arrow Plus is continuing the work of Arrow and is also including for the first time the issue of the clearance of rights in visual works, particularly those in "image-based works". This report is a key task in Work Package 6 of Arrow Plus.

Aim of the Report

The aim of this report is to outline the way in which picture libraries and agencies - as representatives of image rights holders - identify and track the use of images, including the processes for image cataloguing and indexing - and gather and manage rights information. In particular, this report aims to review and investigate the image numbering systems that exist in the picture industry today in order to feed this information into the wider investigation of the Arrow Plus project. This will also clarify the ways in which standard image identifiers might play a part in improving the way in which potential users can discover and clear rights in existing visual works.

One of the tasks proposed by the Arrow project was this review of existing image identification systems with the suggestion that this might include a proposal for standard unique image identification system. Research into existing systems and the way in which image identifiers interact with rights holder information suggests that this would not assist in the identification of the millions of images that already exist. Any such system would be dependent on the creation of an image ID at the point of image creation (ie in the camera in the case of new 'born digital' works). The introduction of such a system would be dependent on the software governing the functionality of cameras. Although there is a strong case to be made for this, such a system would be impractical in the context of the identification of existing images. It does not therefore form a part of this report which focuses on current identification systems.

The report describes both the history and the current practice in the use of image identifiers in the picture industry according to information provided by picture agencies and image technology specialists. The first part of the report deals with the creation and use of image identifiers based on the needs and practices of picture agencies, and includes alternative identification standards such as DOI names and Metadata. The second part of the report focuses on the use of visual recognition technology to find images on the internet.

Key Findings

There has always been a requirement for individuals seeking rights holders to use the information included in publications to aid their research. Professional picture researchers used a wide range of information sources to research right holders and picture libraries and agencies grew to meet the increasing need for representation for photographers, illustrators and artists. Information included in books about rights holders has always been a crucial component in that research. In an analogue world, image identifiers were created and used to facilitate the business requirements of individual image licensing– for example, to track the distribution of images and the allocation of particular kinds of rights to publishers who wanted to reproduce them.

The switch to digital introduced digital identifiers and systems - often based on pre-existing analogue indexing systems. Databases were created to hold this information interacting with databases holding information on rights holders and on licences allocated and fees paid.

Information Held in Publications About the Images Reproduced

- Although rights holders usually require publishers to publish credits and identify image rights holders, this information is not usually displayed in a way that facilitate rights holder discovery as this was not the business of publishers themselves.
- Information in publications has not been designed to facilitate the requirement for this rights information to be made publicly available or for these elements to be automatically indexed
- Rights information in books is often not sufficiently detailed - and rarely includes information about the image identifiers of individual images
- Nevertheless, the information given has generally been sufficient for professional researchers to discover rights holders on an item level basis
- Indeed, the information provided in publications was usually a critical tool for professional picture researchers (ie rights clearance experts) who were looking for illustrative material

Publishers' Information About Image Rights

Publishers do hold detailed rights information about the individual images reproduced in their publications but the rights licensed in an analogue world have been held in analogue formats (eg advice notes, invoices, etc held in folders and files) – and would only be held in databases in summary form to indicate what licences had been negotiated on a contractual basis for that particular work.

Since publishers have no financial incentive to digitize detailed information, they have tended to licence mages in a specific publication, for an agreed term and in specific territories. This information would normally be provided by the rights holder at the time the image was licensed for use.

This project must therefore test the following propositions:

• Whether publications contain sufficient information to identify individual images and rights holders.

• Whether image identifying information can be incorporated into the current ARROW system.

• Whether publishers' information systems include information about the rights holders of individual images

Picture Library Information About Image Rights

- In order to protect the individual rights of rights holders, it is vital that image copyright is dealt with at an item level.
- Any rights clearance system needs to be able to identify the rights holders of individual images and to match this up with the rights required to be licensed
- Picture libraries use an enormous variety of image codes and numbering systems that are deeply embedded in their systems. Picture libraries use these internal systems to identify images; to identify the rights holders of individual images; to identify customers (or licensees); to identify licenses; to track image distribution; and to invoice and pay reproduction and other fees.
- Picture libraries have traditionally licensed their images for very specific uses e.g. reproduction in named publications, within specific territories and for specified terms
- Images have a distinctly separate copyright framework from that of the books that include copies of the images
- Books may be currently either out of copyright, 'orphan works' or "out of commerce" but the images contained within them will usually still be in copyright, will not be 'orphaned' and will usually still be actively "in commerce"
- Picture libraries have many agreements with other libraries for the distribution of their images. Once ingested into the systems of other libraries, it is common for the image identifiers to be changed by the distributing agency. indidual images do not therefore have unique identifiers except within individual library systems.

ARROW PLUS

- One of the purposes of this study is to test the ability of the Arrow System to identify individual images and their rights holders when these image are reproduced in books
- It is therefore important to test its ability to query a database holding current image IDs and to measure its success in enabling this identification.
- The Project also needs to be able to identify contact information about rights holders where they are identified

The Use of Persistent UIDs (Unique Identifiers)

The following propositions need to be tested:

- Whether the introduction of persistent UIDs for individual images would assist image library workflows and the identification of individual images
- Whether persistent UIDs could be retrospectively assigned to existing images – whether born digital or scans of existing analogue images

Other Ways of Identifying Rights Holders

The picture library community has been aware of the necessity and difficulties of identifying, sourcing and tracking images for some time and has therefore invested an important amount of resources in:

- Developing metadata standards so that detailed information about the content and rights f individual images can be shared and exchanged
- Investing in visual recognition software as the most efficient solution to identify images in which they hold rights.

These increasingly sophisticated means of tracking images and image rights will be the subject of further research as part of the Arrow Plus Project.

Conclusions

Getty Images made a submission to the UK government after the publication of the Hargreaves Report on 'Orphan Works' [i]., they pointed out that they have over 8.4 million images available online with another 70 million images offline. They also stated that they daily issue to clients some 400,000 licences to reproduce images. Although Getty Images is the largest global provider of images to publishers, there are many more. The scale of this global marketplace shows that the image rights business is alive and well – and that CEPIC represents the largest and most substantive Reproduction Rights Organizations in Europe.

As a member of the ARROW PLUS project, CEPIC now needs to test the ARROW system with real data taken from picture library systems in order to find out whether it has the ability to enable bulk rights clearance in images reproduced in books and other publications. Further research also needs to be carried out to test the various other propositions in the Tasks outlined in Work Package 6. The sheer volume of images and the many ways in which they are brought to market require digital technology to provide a means of solving the problem of identification. Digital problems require digital solutions.

This report is an introduction to the general issue of image identification. Further interim reports will go into more detail, if this is relevant within the context of the Arrow Plus project. A recommendation will be issued in the Feasibility Study to be delivered at the end of the project.

Part I. of this introductory report looks at way picture libraries and agencies identify and track each individual image in the analogue and in the digital world. Part II. looks into existing identification systems not commonly used by the picture community. This second part is not meant to be an exhaustive list but provides contextual information on how identification systems are created and why they are useful or not.

---

i Notes

┐ Getty Images and Picscout submission is available on the CEPIC website/

http://www.cepic.org/news/blog/2011/10/how_image_tracking_technology_may_be_used_ow_legislation

The Context of This Report

Background: Arrow and Arrow Plus

Arrow Plus is an EC-funded Project that aims to continue the work of the Arrow Project (eContentPlus Programme) - in alignment with the ICHSP work programme - to create " a stable, sustainable infrastructure for rights information management, including clearance of rights and a European registry of orphan works". Arrow Plus will extend the system created within the Arrow Project to discover rights holders and clear rights - focusing on the need to facilitate libraries in performing a 'diligent search' for rights holders in books selected for inclusion in digitisation programmes.

Arrow Plus is designed to maximise impact in the areas identified by Objective 24 of the ICTPSP work programme, especially to "enable the retrieval of information on the copyright status of European works and the identification of appropriate rights holders". In broadening the scope of Arrow, the project will also facilitate the enrichment of information sources in European countries in order to include them in the Arrow system. The key concept in this work is interoperability - the prerequisite for information retrieval from sources originally created for different purposes.

The second expected outcome of Arrow Plus is to "establish a registry of orphan works" and this will be natural outcome of the process which is designed to improve the search for rights holders. "Orphan works" will emerge when rights holders cannot be traced and data about these works will be collected to form the basis of an orphan works registry.

To achieve the third outcome - to "facilitate the clearance of rights for out of print works" - Arrow Plus participants include Reproduction Rights Organistions (RROs) - in particular EVA[ii] and CEPIC[iii]. These organisations are in a position to act as clearing centres for out of print works, and can provide immediate access to licensing terms and conditions in accordance with national legislation and rightsholder mandates. Arrow Plus will enhance access to this information by developing new and collaborative tools for rightsholders, providing rights information and, when appropriate, give mandates to RROs to offer licences to libraries.

ii European Visual Artists

iii Coordination of European Picture Agencies - Stock, Press and Heritage

Searching for images (Alamy)

Duplicate slides ready to be destroyed (Alamy)

Engravers at the Graphic Studio, 1882 (From The Graphic)

Visual Material in Arrow Plus

A significant element of Arrow Plus is to facilitate the clearance of rights in the images published in books - so-called “embedded artistic works” - and, where possible, to incorporate this within the Arrow workflow. Accordingly, the Arrow Plus has set three main objectives. Firstly, to test how easily the clearance of visual material might be incorporated into the Arrow system. Secondly, to prepare a feasibility study for a comprehensive project that would review the broader issues of clearing rights in visual material, rather than just the clearance of “images contained in books”. This study will review existing databases containing information on artistic works and their rights holders (e.g. the databases of picture libraries and agencies or the repertoires of Collective Management Organisations (CMOs)) with particular reference to their interoperability and potential for integration into an “Arrow-like” system for rights clearance that might satisfy some of the requirements for ‘diligent search’ set by the HLEG on digital libraries. This report will contribute to this study by reviewing existing persistent unique identification systems and reviewing proposals for a standard unique identification system which might result in a formal standardisation proposal to ISO TC46/SC9.

PART I – IDENTIFICATION OF IMAGES
IN THE PICTURE INDUTRY

1.1. Identifying and Classifying Images

Images, or pictorial representations, have been produced by human beings from the beginnings of pre-history and long before the introduction of writing. Images were used for a variety of reasons, many of which have been lost to the depredations of time. However, we do know that prehistoric man used images not only to illustrate their own existence but also to leave a record of their activities and increasingly frequently to record transactions and to give instructions for future activities. In the course of time, their form expanded from cave paintings through home and burial decorations to the full pantheon of paintings, drawings, wall decorations, and sculpture.

Probably the first identifiers were those used by groups to identify items that they produced and used, such as identifying marks used by potters over 7000 years ago; religious symbols such as the Christian cross, the Judaic Star of David, and the Islamic crescent moon; and royal and aristocratic marks and designs designed to indicate allegiance to a group. Later group identifiers included professional symbols such as physician's signs and guild trademarks. Then and now, such identifiers became key components in trade and daily life.

Sitting for a portrait photograph (From The llustrated News)

Selection of old photographsc 1880-1910

Photographic Studio c 1898 (Wichita-Sedgwick County Historical Museum)

The Use of Image Identifiers in the Picture Business

The need to identify images only appears in any quantity once man began to make multiple copies of unique individual images, such as engravings made to distribute copies of individual works of art and, significantly, with the invention of photography in 1839. Image identifiers were originally used by photographers, photo agencies and picture libraries to organise and classify negatives, prints and transparencies - and to have a reference system that would enable them to both find, track and refile physical images. In the 19th century the most obvious examples of this were the multiple filing systems of studio portrait photographers who usually adopted simple numbering systems to mark the back of 'cartes-de-visites' so that their clients could reorder copies of their portraits when necessary. Similarly, the early records of national museum photo studios usually consisted of hand-written registers listing photographic assignments numerically in chronological order. Finding aids were gradually introduced by means of loose-leaf albums organised by subject. At this stage, illustrations in books were generally very limited and individually inserted into books by mechanical processes; rights were generally licensed on the basis of simple agreements referring to single editions.

All this changed with the invention of the photo-mechanical process and other technological advances in the printing industry at the end of the 19th century. Photographs could now be used in mass-produced books, magazines and newspapers and the first truly commercial picture libraries emerged to support this growing industry. At first, these photographic libraries emerged within traditional news agencies, such as Havas, the first French news agency, created by Charles-Louis Havas in 1835 and later to create Agence France-Presse (AFP). Two of his employees, Paul Julius Reuter and Bernhard Wolff, later set up rival news agencies in London and Berlin respectively. At the same time, the emergence of the illustrated press meant that magazines and newspapers began to build up substantial picture libraries of their own. Significant among these was the picture libraries of the Illustrated London News[iv] in London; Harper's Weekly[v] and the Daily Graphic in the USA; and Ullstein Bilderdienst, the photo archive of Leopold Ullstein's massive newspaper publishing company, founded in 1877.

Traditionally, all these picture archives indexed their photographic negatives numerically (often using numerical sequences within chronological years e.g. 1935/001 etc) but accessed the pictures through a visual subject index organised with separate alphabetical indexes of portraits and a general subject index.

iv The Illustrated London News was produced between 1842 and 2003 by which time it only appeared twice yearly

v Harper's Weekly was produced between 1857 and 1916

Franklin County Historical Society negatives

Photographer's Studio Register (Franklin County Historical Society)

Scientist with miniature camera

By the 1920s and 1930s, the use of photographs was clearly established in the press and for advertising and this era saw a rapid growth in independent photo agencies supplying images for these kinds of uses. Some were extensions of traditional news agencies, such as Associated Press, and others were exclusively devoted to photography, such as the Keystone Agency in New York started by Bert Garai and Photographic Advertising Ltd in London, an agency that existed solely to provide images for advertisements.

At this time, picture agencies used image identifiers, usually known simply as picture numbers or codes, to identify individual negatives or slides/transparencies for the following reasons:

1. To indicate the physical location of the negative or master transparency for filing - and re-filing after making prints or du plicate transparencies
2. To label prints (made from unique negatives) or duplicate slides which were to be sent on loan to clients
3. To indicate which image was being licensed on the invoice to the client
4. To identify which image was licensed in the sales report to the author or provider
5. To provide a link to other features related to the image, such as model or property releases

Photographic prints and duplicate transparencies were filed by subject in filing cabinets. Users who wanted to find image for use in publications or other media would either ask the library researcher to carry out a subject search - or they could visit the picture libraries to carry out searches themselves. Libraries would develop their own subject indexes as there were no pre-existing taxonomies specifically for picture libraries - and these would reflect the specialisation of the library, when relevant. For example, travel photo libraries would have a large geographical index; nature picture libraries would have detailed indexes based around natural history categories; and

so on. Picture numbers were generally hand-written on prints or slides to cross-refer the images to the original negative or master transparency. Library staff were also an important component of each library as their detailed knowledge of the content enabled them to find unusual subjects in various locations.

Sources of Images

Historical picture libraries, such as Mary Evans and Mansell in London or the Bettmann Archive in New York, had often been started as private collections of prints and engravings that became commercial entities during the rise in popularity of the illustrated book from the 1950s on. Others had their roots in their role as internal picture libraries for newspapers and magazines, like the Illustrated News Picture Library and the Hulton Picture Library. Press photo agencies employed staff photographers to cover newsworthy events and also accepted news photography from stringers (i.e. freelancers) and from other agencies abroad.

General photographic agencies also proliferated from the 1950s onwards and many originated to represent the stock photography of groups of professional photographers. Agencies like this ranged from the famous photographic cooperative 'Magnum' and the library started by photographic agent, Susan Griggs to specialist agencies started by individual photographers like Robert Harding's travel photography library and the Bridgeman Art Library, which started as a collaborative venture with art photographer A. C. Cooper. as time went on these libraries entered into agreements with agencies in different countries.

Science and Society Picture Library, Science Museum London, 2000[vi] (Angela Murphy)

Traditional newspaper photo archive and filing system, Zimbabwe, 2010. [vii]

Selecting pictures from a photo shoot for publication in Photo magazine, 1980. (Robin Scagell)

vi Photograph by Angela Murphy

vii Chronicle and Sunday News, Zimbabwe, 2010. http://mediahelpingmedia.posterous.com/newspapers-making-do-with-old-technology

Digital Image Identifiers in Picture Libraries

The proliferation of duplicates and the introduction of digital imagery - with multiple digital versions of images - increased the need for more complex identification systems which did not require physical filing systems. Instead an enormous variety of image management systems were developed in the 1990s to cope with the issues arising from the growth of first, analogue, and later, digital image management.[viii] Once picture libraries started to combine images from different agencies, they had to develop more sophisticated numbering systems to ensure that their image identifiers were unique. The most common way of doing this was to allocate the home agency's own alphanumeric codes to these images while keeping track of the identifiers of the original source. Alternatively, some agencies allocated unique codes to their sources - and maintained that source's numbering system. as this system was not foolproof, the software driving the underlying database would generally assign a unique computer-generated identifier to each record and image - although these identifiers were hidden from both internal and external users.

viii Boyd-Barrett, Oliver, ed. (2010). News Agencies in the Turbulent Era of the Internet. Generalitat de Catalunya. ISBN 978-84-393-8303-1

Examples of Picture Library Codes

When assigning unique identifiers to images in their collections, picture libraries operate in a variety of different ways. Whereas some libraries allocate a simple numerical sequence to all their images regardless of their source (see Getty Images below), others will base their numbering system on the provider's file number or other factors. For example, the same picture may be allocated a number based on the identification system developed at a time when all the pictures were analogue (negatives, prints, transparencies etc). The following example shows the code numbering system used by an international picture agency to differentiate between pictures indexed for particular usages:

FOT_Código Agencia	**FOT_Código**	**FOT_Código Catalogo**
SDP1473	1473-79L11	SDP1473
Image code allocated by the supplier	New image code created by the picture agency	Image code as displayed in the agency's picture catalogue

In this case, the first column shows the filename supplied to the agency by the source agency; the second columns shows the number allocated by the host picture agency (ie the distributor) for general use; and the third indicates that the picture was used in the picture library printed catalogue. All of them retain the numerical element of the provider's original image. However, the same picture library uses only one type of picture code for all the images that were "born digital".

The following table lists various codes allocated by picture libraries throughout Europe.

Type of License	Provider_Name	Collection_Name	Final_Image_Number
RM	The Fine Art Society	Bridgeman Art Library	FAS 28325
RF	ALAMY- ROYALTY FREE	KAKIMAGE	B0RR97
RF	David Ramkalawon	David Ramkalawon / Alamy	A1WCX9
RM	GETTY IMAGES	GETTY IMAGES NEWS	#114730371
RM	GAMMA RAPHO	GAMMA	GA702463_60

From even a cursory examination of the image numbering systems used by picture libraries, it can be seen that every stock agency, photographer and image provider uses a rainbow of different coding systems. Consequently, it can be seen that there is no image code standard in the picture industry.

Sarah editing transparencies (SSPL)

Slide scanning equipment (SSPL)

Science Museum curator selecting images (SSPL)

Global Distribution

Most picture libraries enter into agreements with picture libraries based in different countries so that they can achieve local representation for their images and some of the major picture libraries, such as Getty Images, age fotostock, Alamy and Corbis act as global agents not just for photographers but also for other picture libraries. For example, age fotostock has created THP - a hub of technological services based upon a search engine and hosting platform - which offers over 17 million images sourced from over 2000 age fotostock individual photographers and c 500 stock agency collections. Migrating large quantities of image data necessitates the use of a single standard for identifying the images and their sources. Each provider, whether agency or photographer, is allocated a unique three letter code that is added to the original image code. The original image code is kept for reporting purposes, since without it stock agencies in the THP platform would not be able to report sales to the photographers that they represent. Similar solutions are adopted by o0ther large agencies, such as Alamy, Getty, or Corbis. However, even these major suppliers - licensing tens of millions of images - have not been able to implement a common standard amongst their suppliers.

1.2. The Uses of Metadata

Background

The most common way to identify a digital object, such as music, books, videos or images, is through their metadata. Metadata describesother data. It is especially important in the area of digital photography. It provides information on the image, what the photo represents and where, when and how the image was taken. Photo metadata is essential for the identification and the preservation of the Image. Photo metadata is commonly held in the databases of individual photographers, agencies and picture libraries and has always been a vital adjunct to images. However, the rapid increase in digital imagery highlighted the importance of including a subsection of core information within the digital image itself. The introduction of photo metadata standards for this use was pioneered by IPTC and later championed by CEPIC. who joined IPTC in 2005 and was a founding member of its “photo metadata” working group. In the context of ARROW PLUS, the question is how this textual information can be used to support the system.

The Identification of Digital Objects

Digital objects can be extremely complex and may be comprised of multiple rights layers, e.g.scanned digital books. Once created, they will also be distributed and exchanged in a wide variety of ways. Facilitating this has been an enduring issue for digital object creators, administrators and users in every branch of society accept that:

- Using established standards facilitates exchange
- Complex objects require more metadata than analogue for their management and use
- Metadata for physical and digital objects needs to be integrated, e.g. in cultural heritage organisations

The Uses of Metadata

Metadata is used to facilitate:

- Search and Discovery
- Management
- IP rights control
- Identification and the certification of authenticity
- Marking of content structure
- Indication of status
- Description of processes

Types of Metadata Used to Describe Digital Objects

Metadata is structured data which describes different aspects of an item. Metadata provides basic information about a digital object, such as details of the author, title, date of release, ownership and rights information Thus rich metadata makes an image easier to find. There are six common types of metadata that can be applied to digital objects, including images. These are:

- Descriptive
- Administrative
- Technical
- Digital provenance/events
- Rights/Terms and conditions
- Structural

The full complexity of this metadata as it applies to the business of trading in image rights can only be accommodated in an image management system specifically designed for image libraries. However, the core information required to identify and describe images can also be replicated in individual image headers through the medium of IPTC and Exif information standards.

The IPTC Standards

Starting in the 1970s, the Photo Metadata Working Group of the IPTC (The International Press Telecommunications Council) has proposed several standards to hold and organize metadata within images. Today the labeling technology used for that purpose is Adobe's XMP (Extensible Metadata Platform). IPTC have designed a number of data entry fields that include caption, keyword, event and copyright information. Adding this essential information to the image's file header assists the interchange of news data between different databases. This set of defined fields makes up the IPTC metadata schema, first released on 21 March 2005. [ix]

The advantages of embedding metadata in images are significant: It helps to reduce the number of Orphan Works, improves information retrieval, and represents an advancement of search quality for the author. Besides this, the fact that the IPTC's schema works on Adobe's XMP considerably increases the chances of this schema becoming a vehicle for a standard image identifier. If this is not the case yet, it is because picture agencies do not use XMP fields in a standard way. It is also because several libraries make use of non-standard keywords when building their vocabulary.

As a member of the IPTC since 2005, CEPIC has been working resolutely to promote and establish a uniform use of XMP fields, as well as the correct use of keywords for building an image vocabulary. To achieve these goals, CEPIC has been represented since 2006 in the IPTC Photo Metadata Working Group by Image Technology and Metadata Consultant, Sarah Saunders. CEPIC has also organised a Metadata Conference every year since 2007 - one of the main events of the CEPIC Congress. To support the uniform use of metadata for images, CEPIC launched the CEPIC/IPTC

Metadata Handbook in May 2011, a guide setting out some of the principles that need to be applied when planning metadata. This publication can be downloaded from the CEPIC website. In December 2011, IPTC launched the Embedded Metadata Initiative (http://embeddedmetadata.org/index.php) supported by photographers associations and other organisations involved with the identification of images [x]

Since the subject of metadata is extensive and this report intends to be succinct, we will deal with IPTC standards in more detail in a separate report. We would also like to refer to further metadata and databank information sources. The first - giving an overview of metadata standards - is **Metadata Made Simple**[xi], an article written by Image Management and Rights Clearance consultant, Angela Murphy, which can be downloaded from the CEPIC website. The second is the website of **Controlled Vocabulary**[xii], a photographer's guide to building controlled vocabulary lists, thesaurii, and keyword hierarchies for describing images in databases. The web site has been created by professional photographer and metadata expert, David Riecks, and the platform has played an active role in developing industry metadata standards.[xiii]

As has been shown in the first part of this report, the picture industry does not have a common standard for a unique image identifier. The current enormous diversity of image identifiers is the result of the anarchical origin of photography itself, which allow photographers, even today, to catalogue their pictures using any system they want. Today, considering the fact that each digital image is given an identification code by the camera the moment it is shot, one might think that camera manufacturers could take on the responsibility of creating or agreeing on a unique image identifying system, the DOI System, for example. However, as commercial rivals in a tough economic climate, picture libraries do not always want to invest the time required unless they see a very direct ROI (return on investment) for their work.

N.B. It is also important to note that attempts to standardize image identifiers with a top-down approach, such as the DOI system, failed right at the creation of the Internet.

The PLUS Standards

The standardization of Rights metadata has been the subject of intensive research in the USA over the last few years and, in 2008, resulted in the creation of a universally-accepted standard called 'PLUS' (www.plus.org). . The Picture Licensing Universal System, "provides an integrated set of standards for communicating rights and ownership information associated with commissioned and existing images". [xiv]

These standards have been developed by an international, non-profit organization called the PLUS Coalition which "approves and maintains a set of standards for licensing language and formats. It serves as an umbrella association representing publishers, designers, advertising agencies, photographers, illustrators, stock image distributors, artist representatives, museums, libraries and standards bodies, such as UPDIG, IPTC, IDEAlliance and others. " Clear guidelines relating to the use of these standards by both image maker and image receivers are available on the UPDIG web-

site and many plug-ins have been written to incorporate these standards in the most widely available image management software.

The incorporation and persistence of rights metadata is vital to rights discovery and should be integral to any photographic or digitsiation programme. The PLUS website offers free tools to embed and read PLUS licenses using an XMP metadata format. A license string can reside in IPTC or XMP metadata, with potential for direct embedding in an image file. CEPIC will deal with rights standards in greater detail in a separate report.

1.3. Image Recognition Technology

Background

Technology experts have been working to perfect image recognition technology for over ten years and picture libraries have often looked to this field (i.e. computer vision) as a potential solution to their more problematical issues. However, it is only in the last few years that developers have created a suite of inter-related technologies that could have a profound impact on the picture business. These technologies include CBIR (content-based image retrieval), object recognition, image processing and analysis, and fingerprint technology. This technology was originally marketed to picture libraries on a prohibitively-expensive basis but now the service businesses that have developed these technologies are beginning to see a business advantage to making these technologies more widely available.

The Picture Library Business

A recent report on web statistics[xv] has shown that by the end of 2010 the number of worldwide internet users has grown to 1.97 billion with the number of websites created exceeding 255 million. This growth is expected to continue at the rate of some 20 million web sites every year - each rich in imagery. Furthermore, the statistics relating to images are astounding. By September 2010, Flickr was hosting about five billion images with over 3000 additional photographs being uploaded every minute. Furthermore, over three billion photos are uploaded to Facebook every month.

In the face of such statistics, it seems unlikely that any agreement on standardising image identifiers is likely to have much of an impact. Businesses and cultural organisations must all find a cost-effective way to manage their images and users must have increasingly intuitive ways of searching for images. Image search and recognition technology seems to offer that opportunity and many picture libraries have already turned to these technologies in their attempts to detect and track illegal uses of their images on the internet.

---

ix CEPIC/IPTC Metadata Handbook:

http://www.cepic.org/news/industry_press_releases/2011/05/cepic_and_iptc_release_metadata_handbook

x Following associations support the initiative :

- AIPA - Advertising and Illustrative Photographers Association (NZ)
- ACMP - Australian Commercial and Media Photographers Association (AU)
- AIPP - Australian Institute of Professional Photography (AU)
- AOP - The Association of Photographers (GB)
- APA - American Photographic Artists (US)
- Artists Bill of Rights (GB)
- ASMP - American Society of Media Photographers (US)
- ASPP - American Society of Picture Professionals (US)
- BAPLA - British Association of Picture Libraries and Agencies (GB)
- British Photographic Council (GB)

- BVPA - Bundesverband Pressebild-Agenturen und Bildarchive (DE)
- CAPIC - Canadian Association of Professional Image Creators (CA)
- CEPIC - European picture agency federation (EU)
- ControlledVocabulary.com (US)
- NPPA - National Press Photographers Association (US)
- NZIPP - New Zealand Institute of Professional Photography (NZ)
- PACA - Picture Archive Council of America (US)
- Photometadata.org (US)
- PLUS - Coalition for communication and management of image rights (Int'l)
- Pro-Imaging.org (GB)
  Pyramide Europe - representing visual artists in the European Union
- Stop43 - stop Commercial Orphan Works Exploitation (UK)
- UPDIG - Universal Photographic Digital Imaging Guidelines Coalition (Intl'l)
- WAN-ifra - World Association of Newspapers and News Publishers (Int'l)

xi Metadata Made Simple by Angela Murphy: http://www.cepic.org/issues/industry_issues_downloads/2008/06/metadata_made_simple

xii Controlled Vocabulary by David Riecks: http://www.controlledvocabulary.com/

xiii Also the Metadata Working Group (MWG) organization, created in 2007 with Apple, Adobe, Canon, Microsoft, Nokia and Sony issues guidance for the use of metadata on images.
http://www.metadataworkinggroup.org/pdf/mwg_guidance.pdf

xiv The PLUS standard is described in detail on the website of the PLUS Voalition (www.plus.org) and described in its wider context on the UPDIG website (http://www.updig.org/guidelines/ir_metadata.html). UPDIG is a global coalition of photographic associations and publish guidelines on all aspects of photographic creation, submission and distribution.

xv Internet statistics for 2010 gathered by pingdom.com http://royal.pingdom.com/2011/01/12/internet-2010-in-numbers/

An illustration from "Roger Fry: A Biography" by Virginia Woolf, 1940

Selecting images for digitisation (Ullstein Bilderdienst

Image Recognition Technology and Orphan Works

As part of the CEPIC Congress 2011 in Istanbul[xvi], the Stakeholder Meeting on Orphan Works included the participation of the three leader companies in visual recognition technology: PixTrakk, PicScout, and Corrigon. These technologies work well for picture agencies. The companies had been selected for a presentation at the CEPIC Congress 2011 Istanbul on the ground that they all have investigated in technologies fit for tracking "orphan works". In the context of ARROW PLUS, the question is how these technologies may be adapted to a Librarian or/ and a Publishers 'environment[xvii]. This point necessitates further investigation and will be part of the Feasibility Study.

The following is a summary of their presentations and services:

PixTrakk

PixTrakk director Kelli Grant introduced the PixTrakk service which allow photo agencies to automatically track photos in print (e.g. from scans of printed pages or PDFs) and online publications through the use of image recognition software from LTU technologies. This tracking system uses finger-printing technology, unique patterns that are part of the image itself, to mark pictures. After the images have been marked, the technology sends out web spiders to crawl the Internet looking for them.

The PixTrakk service then compares photos from the agencies to those in a database of publications that are automatically analyzed by LTU's image-matching technology. The image-matching technology is resistant to image transformations and can not only detect identical images, but also identical images that have been edited in some way. The usage information, updated in real-time, is made immediately available via the web interface of the PIXTRAKK system. The client may search and sort these images by filename, date or publication, as well as generate reports (CSV files) that can be imported into their invoicing system. As a result, agencies are able to reduce billing time and to track the fraudulent use of their photos. The system allows you to zoom into the images found and check them for credits, helping to determine if the picture you are looking at came from your agency or another collection.

Kelli Grant pointed out that books need to be scanned and a constant flow of PDFs must be guaranteed in order to determine the location of the images they contain. PixTrakk currently covers 2500 publications on the French market and 9000 publications in the US market.

PicScout

PicScout has been offering its image tracking services for eight years. Based in Israel and the United States, in 2007 they were chosen to testify before the House Subcommittee on Intellectual Property at a Hearing on Orphan Works Legislation in the United States.

PicScout's ImageTracker offers an end-to-end service which uses PicScout's fingerprinting technology to compare images on the web to the images in its database. When an image usage is found, a screen shot is taken along with other relevant image information. This information is then presented to clients through a compliance management web application that enables them to compare these usages to their database to determine if images are being used legitimately. Agencies and licensors can take action upon receiving PicScout usage reports. Like PixTrakk , PicScout fingerprinting technology allows clients to identify images wherever they appear on the Internet, even if the image has been altered, colorized, edited, cropped, rotated or embedded into another image. While picture agencies limit their searches to catalogues or collections, companies like PicScout search the entire Internet.

During the Stakeholder Meeting, Offir Gutelzon, PicSocuts CEO and co-founder, stressed the importance of "digital fingerprints" in finding images, as they can be created every time a photograph is taken .In addition, fingerprinting information cannot be separated from the original digital

image. Many proposals for digital registries, he added, had ignored this vital component when formulating their registry systems.

PicScout has developed a registry that holds over 100 million images, 60 million of which have already been indexed. Their registry can be used by anybody to find a visual match of images, making it possible to connect image users with content owners. Content providers – including cultural institutions and commercial agencies – are free to register their images in the PicScout registry.

Offir Gutelzon pointed out that when looking at orphan works the scope should not be confined to pictures from the past stored in historical archives. On the contrary, all images created today and uploaded to the Internet without metadata are potential orphan images.

Corrigon

Like Pixtrakk and PicScout, Corrigon uses its own content tracking and visual search system. It offers companies a way to control and monitor web content related to their content and products, in order to understand customer needs, identify infringements, protect their corporate reputation online and embed a positive and supportive public awareness. The user's interface is web-based with no software installation needed; only a browser is required for loading the images and start using the service.

During the Stakeholder Meeting Corrigon's Managing Partner Dennis van Bodegom explained that their tracking technology adds contextual information to the image search, i.e. the metadata around the published image. By paying attention at the context in which an image is being used, he said, it is possible to better ascertain the rights status of images that are missing IPTC information.

When asked if image tracking technologies and image registries should be owned by one single company or publicly funded, both PicScout and Corrigon expressed that there is public interest in these systems being owned and managed publicly. However, Offir Gutelzon pointed out that initiatives linked to technological innovations normally depend on the private sector. He then added that even more important than the question whether this initiative should be private or public is that it should be open and the API available.

He remarked that the continuity of an image tracking and registry system depends strongly on the needs of image buyers and researchers rather than finding the owner of the content. All three technology speakers agreed that if the search is not made efficient for image buyers and researchers, the best system will remain unused.

Dennis van Bodegom pointed out that tracking technology can now provide accurate results; however, sometimes the intervention of the "human" component will be indispensable to find the correct right owner of an image.

For analogue images to be stored in databases and tracked in the web they need to be scanned, a costly process for several picture agencies. Others fear not to be able to store the new created data. In this respect all agreed that, since digital fingerprints are small, a low resolution image (600x400) is sufficient for storing and tracking.

Finally, all three technology speaker shared the opinion that, ideally, and regardless of which visual recognition technology is being used, one central database should be developed and maintained to store all the references; in this framework, linking multiple databases should also be taken into consideration as it is technologically possible.PixTrakk director Kelli Grant introduced the PixTrakk service which allow photo agencies to automatically track photos in print (e.g. from scans of printed pages or PDFs) and online publications through the use of image recognition software from LTU technologies. This tracking system uses finger-printing technology, unique patterns that are part of the image itself, to mark pictures. After the images have been marked, the technology sends out web spiders to crawl the Internet looking for them.

Other Image Tracking Technology

There are several other companies proposing visual search technology. Systems have different features, approaches and price structure. CEPIC will establish a benchmark analysis of existing systems in a separate interim report. One question will be whether they may support Arrow Plus and at which cost.

---

xvi All films are available at http://www.vimeo.com/album/1628615 (Password is CEPICNM22011)

xvii Speech at the Frankfurt Bookfair 2011:

http://www.cepic.org/sites/cepic/assets/Finding%20the%20Author%20is%20not%20a%20Lost%20Case.pdf

PART II – DEVELOPING STANDARDS AND IDENTIFICATION SYSTEMS

2.1. Developing Standards for Digital Media

Background

Throughout the 1990s rapid technological changes in computing and digital imaging and the growth of the internet led to a massive global expansion of the art and craft of photography. The fantastic growth in digital imaging throughout the 1990s also led to a plethora of industry problems. In the analogue film era, differences between hardware did not impact unduly on the consumer experience, but in the digital era issues arising from the lack of compatibility or interoperability between products provided both manufacturers and consumers with an increasing list of issues at a rapidly-rising rate.

Manufacturers have always know that they have a mutual interest in resolving problems by promoting research and information exchange, setting standards through trade and industry associations and reducing proprietary formats. In the pre-digital age, these organisations could proceed at a relatively leisurely pace and would often confine their activities to regional concerns, however the dawn of the digital age demanded a much faster, and more collaborative global approach as they struggled to keep up with the fast pace of technological change.

As a result, the regional trade and industrial associations that existed in Japan (JEIDA and EIAJ), the United States (SEMTEC) and Europe (ECMA) to promote common standards in the areas of Information and Communications Technology (ICT) and Consumer Electronics (CE) have all expanded into global organisations. As JEITA said in their 2000 press release, "The digital revolution is now global in scale, and new systems and products are being born from the convergence of information, communications and images. To keep apace of this trend, activities are required irrespective of existing paradigms.

Industrial associations are no exception. They must respond quickly to changes in circumstances around them and cover newly converging high-technology areas. Furthermore, they must strengthen their ability to propose new policies and strategies as organizations at the core of information technology (IT) advancements..."

"Tools like digital photography, PDF and XML have dramatically transformed the way we communicate, opening up a world of possibility for the media. Minutes after it happens, we can read online about an earthquake hitting a remote part of the world. Today, anyone can witness an event, capture it on video, and post it on the Web to be seen by millions.

Technology is not only tearing down geographical barriers and speeding up communication, but also giving access to a whole new set of players. But none of this would be possible without the International Standards that behind the scenes promote the dissemination of knowledge and facilitate market access to innovative technologies and which, most importantly, ensure interoperability – key for the uptake and use of new technology in a world of continuous and competing developments."

ISO Focus - April 2009 'ISO and the Media'

By way of illustration, the number of published standards produced by the world's largest developer and publisher of International Standards - the ISO (International Organization for Standardization)[xviii] has grown from just

26 when the ISO was founded to over 18500 today. These standards include guidance on everything from technical processes to environmental impacts and are backed up by governmental edicts. ISO have published a magazine devoted to the standards that apply to the media - ISO Focus - April 2009 'ISO and the Media' [xix]

Camera Manufacturers

In 2002 organisations "engaged in the development, production or sale of standard film cameras, digital cameras, and related devices, instruments and software" got together to form the Camera & Imaging Products Association [CIPA] [xx]. CIPA now includes all the major industry players and its current Board includes Canon, Fujifilm, Olympus, Panasonic, Seiko Epson, Sony and Nikon.

In 2007 Cepic played a vital role in encouraging visual equipment manufacturers to recognise the problems faced by the picture industry - and advancing some of the ways in which they might contribute to their solutions. Following the first international conference on photo metadata that CEPIC hosted at theirCongress in Florence, CIPA established the Image File Format Study Ad hoc-Working Group to look at the requirements and problems relating to the format and use of Exif/DCF and its applications such as standards, software and solutions. In particular, photographers and picture agencies were keen to encourage camera manufacturers to introduce a way of marking images with globally unique identifiers (GUID) that might assist them in protecting the copyright of images.

The following year at the 2008 CEPIC Congress in Malta, the second IPTC Photo Metadata conference attracted speakers from some of the most significant players in digital media, including fascinating presentations from Gunnar Peklis of Adobe, Hiroshi Maeno from Canon, Josh Weisberg from Microsoft, Peter Stig from Hasselblad and Josh Schorr of Apple.[xxi]

Software Manufacturers

From the early 1990s, picture libraries started to install picture library management software and systems to cater for the complex functions needed to manage picture library management. Picture library management systems managed the interplay between multiple databases. These databases contained image records, photographer and source agency records, client records, delivery note records, and invoice records all of which were linked by the persistent unique identifiers allocated to images. Initially these records related to analogue images, such as negatives and slides and larger agencies produced barcode labels for their images which would assist in tracking images.

However, once advances in digital image and storage technology had enabled high value digital photography, picture library software was adapted to reference digital imagery alongside analogue images. The use of these computerised systems enabled machines to allocate numbers for the first time - thereby reducing human error and guaranteeing that each number would be unique within each system.

---

xviii http://www.iso.org/iso

xix http://www.iso.org/iso/iso-focus_2009-04.htm

xx http://www.cipa.jp/english/annai/shushi.html

xxi Photographer Ulrik Sodergren has written comprehensive notes about the meeting in his blog at http://digitalfotografen.se/blog/category/cepic/

2.2. The Creation of Persistent Identifiers

Linked Open Data

The opportunities presented by linking open data sets on the internet has been taken up with particular enthusiasm by the cultural heritage community. As result of engaging with the many issues of combining different data sets, several EC-funded projects have made proposals for creating dereferenceable persistent unique identifiers. For example, the ATHENA project (Access to cultural heritage networks across Europe) had a working group that focused on "Identifying standards and developing recommendations". This group produced a booklet entitled "Persistent Identifiers (PIDs): Recommendations for Institutions"[xxii] intended as a short introduction to why "persistent identifiers" are needed and what systems are currently available. As they stated, "Persistent Identifiers" (PIDs) are necessary in order to ensure that information about an object and the object itself, together with its digital copies, can be related to each other and can be retrieved easily at different points in time and from different places.

Although the booklet is a clear and comprehensive guide to this subject, the group acknowledged the difficulties of creating a central authority that would have sufficient scope to mandate a consistent approach across the sector - even within one country. The fact that mass digitisation projects have been in place for over a decade means that cultural organisations have already created numbering systems for the digital versions of their collections. Nevertheless, this is not to say that best practice should be ignored. It is crucial that individual organisations should act as their own authority and ensure that, at the very least, they apply their own identifiers in a consistent way. For many organisations this will consist of a combination of the domain URL root (since this is usually under the sole control of the organisation) and either the object ID number or an equivalent unique, but locally-generated, identifier. The problem for many cultural organisations is that they have conflated the meaning of an individual object with that of an image of the object.

---

xxii Downloadable at http://www.athenaeurope.org/index.php?en/110/promotional-material/15/14-

Selecting slides for scanning

Guggenheim Museum, Bilbao, Spain (Angela Murphy)

The Society meeting

Other Recommendations for Cultural Organisations

Recommendations were also issued by CIDOC (http://cidoc.mediahost.org/) on behalf of ICOM (http://icom.museum/L/2.html) to assist museums with creating standardised references to the way in which their physical holdings are identified on the web[xxiii]. Key recommendations included

- The museum decides on a base URL that will be extended by the inventory number of the object.
- The base URL could be within the domain name of the main museum website or be a separate distinct domain name. Such a separate domain name provides continuity even if the domain of the museum website is altered in the future.
- For larger collections resolving LOD access requests to object information may cause some server load. This can more easily be balanced when using a second domain name.

Europeana and Image Identifiers

Recent publications that deal directly with the issues of image identifiers have been produced by Europeana in its explication of the new Europeana Data Model (EDM). the EDM has been produced to deal with the many conflicts created when combining multiple data sets for inclusion within the Europeana portal. A large proportion of these data sets take the form of digital images and their accompanying image records, together with other associated information.

xxiii Downloadable at http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/AGM_2011/LoD_For_Museums%20v1.6.pdf

2.3. The Use of Identifiers for Physical and Digital Media

Identifiers

Numbering systems have self-evidently been used for several millenia to organise and classify a wide variety of entities. Numbering schemes exist for chemical compounds, food additives, telephone systems, roads, countries, national identification, vehicles, and diseases - and many others. There are also product numbering schemes for products (barcodes), ISBN codes for books and ISSN codes for periodicals.

Numbering Systems for Cinematographic Works

ISAN International Agency (International Standard Audiovisual Number)

The ISAN (International Standard Audiovisual Number) is a voluntary numbering system for the identification of audiovisual works. It provides a unique, internationally recognized and permanent reference number for each audiovisual work registered in the ISAN system.

The ISAN identifies works, not publications or broadcasts. The ISAN remains the same for an audiovisual work regardless of the various formats in which the work is distributed (e.g. DVD, video recording) or the uses to which it is put.

Numbering Systems for Books and Magazines

A standard numbering system for identifying books and magazines was introduced in the 1960s, driven by the need for a publishers and booksellers to precisely identify individual editions. The International Standard Book Number (ISBN) is a unique numerical book identifier based on a previous 9-digit Standard Book Numbering (SBN) code created in 1966. The current 13 digit ISBN was introduced in 2007 and was developed by the International Organization for Standardization (ISO). The code identifies the country of origin, the publisher and the item number of the book and incorporates a'check digit' to ensure that errors are detected. The allocation of codes is managed by the International ISBN Agency.

Similarly, an International Standard Serial Number (ISSN) is a unique eight-digit number used to identify a print or electronic periodical publication that was also introduced in the 1970s. An ISSN, unlike the ISBN code, is an anonymous identifier associated with a periodical title, containing no information as to the publisher or its location. For this reason a new ISSN is assigned to a periodical each time it undergoes a major title change. ISSNs are allocated by a network of national ISSN centres, usually located in national libraries and coordinated by the intergovernmental organisation - the ISSN International Centre based in Paris.

2.4. Identifying Creators

The Visual Creator Index

The Visual Creator Index Ltd, or VCI, is an independent and not-for-profit Index run on behalf of, and for the benefit of, all creators of visual images in the UK, who have come together to protect their work from unauthorized use, principally in digital media. In 1998, the year Visual Creator Index Ltd. was established in the United Kingdom, the system was considered "a first step towards a world-wide standard for identifying visual creators."[xxiv]

To protect the rights of visual creators in a digital environment, the VCI has set up a Register to provide and assign a standard Creator Code, uniquely identifying an image creator. Any Code can instantly be linked back to the individual creator through a constantly updated index, the VCI Register[xxv], which is available on the web, or by phone or fax through the member organizations of the management committee.

The VCI system is based on digital labels, which work as "invisible ink" on the image itself. The digital label should:

- survive repeated copying
- survive translation into and between other media
- be accessible from even a fragment of the whole image
- survive cropping, manipulation and re-combination
- be secure, not easily removable
- not compromise the image quality

The digital label contains an identifier, a code that identifies the creator of the visual image, the one identifier guaranteed to remain constant and that doesn't need to be updated. The identifier is unique to an individual creator subsisting at least as long as copyright (70 years p.m.a.). To encode information about the creator of the image or its copyright owner VCI uses digital "fingerprinting" or "watermarking". This encoding method adds identification information into the structure of the image itself, at a sub-visual level. Being a part of the image itself it will always travel with it and should survive a good deal of manipulation such as cropping, colour changing, blurring and sharpening. Another advantage of fingerprinting or watermarking is that it can also be easily read by mechanisms built into image processing software (e.g. Photoshop).

Creator's Index

VCI has also developed a public database, an index of creators and their unique labeling codes that allow easy identification. This database runs as a "not-for-profit" resource by and for the benefit of the creators.

In practice how this works is that each visual image-maker working in a particular country would be given a unique ID number on a "Visual Creator's Index" that they could use throughout their everyday work for physically or digitally marking prints, transparencies and digital files. Anybody needing to identify or contact them could use the ID number to consult the Index, either directly on the Web, or by phone. There they can obtain up-to-date details as specified by the individual concerned - this could be a phone number, agent's number or an address.

VCI envisaged one unique worldwide standard for the format of the identifier code, to allow for a future global database. However, since most of its members are UK creators, they implemented an identification scheme to be used in the UK. The interim standard code - the VCI Code - provides for a "Creator ID" in the form: 00-AA-00

This is an example of a fixed format code supplied to an individual creator by the VCI: 44-ML-01

[xxiv] http://www.vci-uk.com/page1.htm

[xxv] http://www.vci-registry.org/search.asp

2.5. Standardised Image Identifiers

The DOI System

Image libraries are aware that a unique standard image identifier could assist their business and have therefore considered the use of other proposed identification standards. One of these is the Digital Object Identifier System (DOI), a character string used to uniquely identify a digital object such as an electronic document for use on digital networks.

A DOI Name is a means of referencing information about a digital object, such as the object's location on the Internet, its intellectual content, and its metadata. It can also link customers with content suppliers, facilitating electronic commerce, and enabling management of media. By using a DOI to reference a document, the provide and user is able to reference an online document with a more stable link than by simply referring to it by its URL. Thus, if the URL changes, the publisher need only update the metadata for the DOI in order to link to the new URL. However, DOIs also require a separate resolver to turn them into useful information so are not a recommended means of obtaining identifiers.[xxvi]

The DOI system is used mainly for persistent citations in scholarly materials (journal articles, books, etc.) through CrossRef, a consortium of around 3,000 publishers; scientific data sets through DataCite, a consortium of leading research libraries; technical information providers and scientific data centers; and the European Union official publications generated by the EU publications office.

DOI Name

A DOI name takes the form of a character string divided into two parts: a prefix and a suffix. The prefix identifies the registrant of the name, and the suffix is chosen by the registrant and identifies the specific object associated with that DOI. The DOI System works in collaboration with Registration Agencies (RAs), whose main role is to allocate DOI name prefixes, register DOI names and provide the necessary infrastructure to allow Registrants to declare and maintain metadata and state data.

Example of DOI Name: 10.1000/182

Here 10.1000 is the prefix and 182 is the suffix. The "10." part of the prefix identifies the DOI registry, and the characters 1000 in the prefix identify the registrant; in this case the registrant is the International DOI Foundation itself. 182 is the suffix, or item ID, identifying a single object (in this case, the latest version of the DOI Handbook).

When the citation is a hypertext link, users are recommended to embed the link as a URL by concatenating "http://dx.doi.org/" to the DOI name, omitting its "doi:" prefix; e.g., the DOI name doi:10.1000/182 is linked as http://dx.doi.org/10.1000/182. This URL provides the location of an HTTP proxy server which will redirect web access to the correct online location of the linked item.

While there are advantages to a DOI name, the picture industry has never embraced this alternative means of identification. In addition, the DOI system is only open to organizations that

can meet the contractual obligations of the DOI system and are willing to pay to become a member of the system. RAs (Registration Agencies) can assign DOIs.

In 1999 Norman Paskin, then director of DOI, contacted CEPIC with the aim of introducing the DOI System into the image world, with CEPIC acting as a Registration Agency (RA) for images. The proposition failed - partly because the DOI system wasn't ready for full implementation at that time, and because, to comply with DOI's requirements, CEPIC would have had to charge picture agencies exorbitant fees for registering their images.

Allocating a DOI name to an image doesn't represent a real commercial advantage for picture agencies. Today digital images from hundreds of collections are found in online stock picture libraries and searched mainly by the use of keywords i.e. the most appropriate words to describe what the subject matter of the image, rather than a number or code. Once the potential buyer finds the picture that he/she needs, a simple click on the picture will display (on the same platform by accessing IPTC/Exif fields) all the information related to the image, including the author, camera data, keywords, code, collection. In other words, a DOI name does not represent an extra benefit to the business models used for images today.

Appendix: ISO standards

The Creation of an ISO Standard for Image UIDs

ISO is a network of the national standards institutes of 162 countries, one member per country, with a Central Secretariat in Geneva, Switzerland, that coordinates the system.

ISO is a non-governmental organization that forms a bridge between the public and private sectors. On the one hand, many of its member institutes are part of the governmental structure of their countries, or are mandated by their government. On the other hand, other members have their roots uniquely in the private sector, having been set up by national partnerships of industry associations. Therefore, ISO enables a consensus to be reached on solutions that meet both the requirements of business and the broader needs of society.

---

xxvi Guidelines for the creation of persistent identifiers

http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/AGM_2011/LoD_For_Museums%20v1.6.pdf

A) How ISO decides to develop a standard

ISO launches the development of new standards in response to the sectors that express a clearly established need for them. An industry or business sector communicates its requirement for a standard to one of ISO's national members. The latter then proposes the new work item to ISO as a whole. If accepted, the work item is assigned to an existing technical committee. Proposals may also be made to set up technical committees to cover new scopes of activity. At the end of 2010, there were 3 274 technical bodies in the ISO system, including 214 ISO technical committees.

B) How are ISO standards developed?

ISO standards are developed according to the following principles.

Consensus

The views of all interests are taken into account: manufacturers, vendors and users, consumer groups, testing laboratories, governments, engineering professions and research organizations.

Industry wide

Global solutions to satisfy industries and customers worldwide.

Voluntary

International standardization is market driven and therefore based on voluntary involvement of all interests in the market-place.

There are three main phases in the ISO standards development process as follows.

The need for a standard is usually expressed by an industry sector, which communicates this need to a national member body. The latter proposes the new work item to ISO as a whole. Once the need for an International Standard has been recognized and formally agreed, the first phase involves definition of the technical scope of the future standard. This phase is usually carried out in working groups which comprise technical experts from countries interested in the subject matter.

Once agreement has been reached on which technical aspects are to be covered in the standard, a second phase is entered during which countries negotiate the detailed specifications within the standard. This is the consensus-building phase.

The final phase comprises the formal approval of the resulting draft International Standard (the acceptance criteria stipulate approval by two-thirds of the ISO members that have participated actively in the standards development process, and approval by 75% of all members that vote), following which the agreed text is published as an ISO International Standard.

It is also possible to publish interim documents at different stages in the standardization process.

Most standards require periodic revision. Several factors combine to render a standard out of date: technological evolution, new methods and materials, new quality and safety requirements. To take account of these factors, ISO has established the general rule that all ISO standards should be reviewed at intervals of not more than five years. On occasion, it is necessary to revise a standard earlier.

To date, ISO's work has resulted in over 16 000 International Standards, representing more than 620 000 pages in English and French (terminology is often provided in other languages as well). A list of all ISO standards appears in the ISO Catalogue.

N.B. This information has been provided by the ISO and more detail can be found at www.iso.org

Bibliography and Links

Aufderheide, Patricia and Jaszi, Peter, "Untold Stories: Creative Consequences of the Rights Clearance Culture for Documentary Filmmakers". Washington DC: Center for Social Media 2004.[xxvii]

EUROPEAN DIGITAL LIBRARIES AUDIOVISUAL WORKING GROUP. "Orphan Works sector –specific guidelines for the audiovisual sector". http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/orphan

Videos from the Stakeholder Meeting on Orphan Works at the 2011 Cepic Congress, Istanbul - http://www.cepic.org/issues/2011/09/arrow_arrow_plus

Boyd-Barrett, Oliver, ed. (2010). News Agencies in the Turbulent Era of the Internet. Generalitat de Catalunya. ISBN 978-84-393-8303-1

[xxvii] http://www.acsil.org/resources/rights-clearances-1/nps240.tmp.pdf