August, 2003 Hiroyuki Yamashita (Content ID Forum)

This paper gives a broad overview of the ID-based contents distribution framework that we are proposing. After briefly reviewing the significant characteristics of digital contents, the various circumstances affecting the way contents are distributed today are described. This will provide a good introduction to the rest of this article, in which the Content ID system is described in detail and the efforts of the Content ID Forum (cIDf) to precede standardization in this area are covered.

■ Characteristics of Digital Content, Role of Metadata

Digital has some distinct advantages over analog, most notably digital files can be copied much more easily with virtually no noise or signal degradation, and this makes the potential for illegal copying or piracy a major issue. Above and beyond simple legal remedies, a number of technological schemes have also been devised to discourage and prevent illegal copying. For example, digital watermarking deters piracy and stronger measures such as encryption and Digital Rights Management (DRM) prevent illegal copying altogether.

Contents

Sent as an analog signal attached to something else and subject deterioration

(C) Copyright 2002 & 2003, cDf

Characteristics of Digital Contents

Digital

Digital

Original source is copied as is

1. Regarding Digital Contents

Aside from the piracy issue, another problem with digital contents is that you don't know what the contents actually contain by just looking at it from the outside. You have to actually open the content up and look at it or play it. This raises the issue, how do you find out about the actual substance of the contents and the rights-related information that protects the contents? Well, we believe that the best answer is to assign every digital content object an ID that

1. Regarding Digital Contents

Technological Protection of Digital Contents

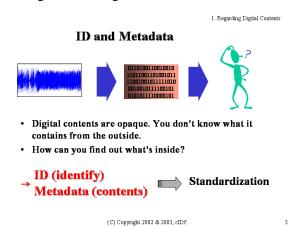
Cost

DRM encrypted

Management function protects against illegal copying
Contents

Digital Valency file (C) Copyright 2002 & 2003, cIDf

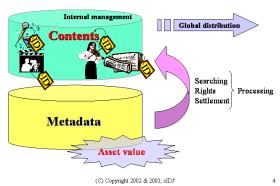
uniquely identifies it. Also, information that is useful to know about the contents can be added in the form of special text called *metadata*—data about data—that can be accessed whenever necessary to find out what the contents actually contains. The primary mission of the Content ID Forum (cIDf) is to promote standardization of the Content ID and ways of attaching the ID to digital contents.



The content itself and associated metadata describing attributes of the content are managed internally, but it is apparent that metadata itself will become increasingly valuable as an asset in its own right, so we want to combine metadata with contents for global distribution. The direction we are moving will provide metadata with a number of important roles to provide better searching for contents, rights-related processing, settlement and payment processing supporting content transactions, and so on.

1. Regarding Digital Content

Potential Benefits of ID and Metadata



A major role of metadata, as I just mentioned, is as a means of searching. For example, there are techniques available that can recognize and search images directly, but in terms of accuracy these methods still have a long ways to go. Describing images with metadata is a better approach for the foreseeable future; searches can be performed very easily and with a high degree of accuracy.

Roles of Metadata

1. Regarding Digital Contents

Metadata as information search tool

In order for software to understand the information content of Web pages and the substance and attributes of multimedia contents, some explanatory text based on standardized conventions and that can be manipulated by computers must be added to each content object. This is the essential meaning of metadata. The Platform for Internet Content Selection (PICS) developed by the World Wide Web Consortium (W3C) is one implementation of metadata.

Metadata for recording content integration

Digital contents are easy to reuse, so it is common for different content objects to be combined as components, and for new objects to be created by editing an original content object. Metadata assumes the role of modulator in these cases by recording the different content objects that are combined and by recording the editing history of the content.

Metadata as tool for recording rights relations

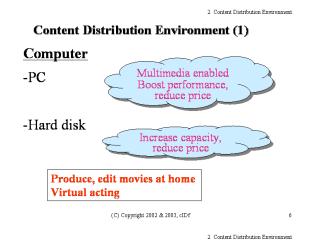
A content object whose copyright attributes are unclear is essentially worthless, because cannot be used. Metadata can be used to record the complex rights-related attributes of content objects.

(C) Copyright 2002 & 2003, cIDf

Just viewing digital contents does not in any way exhaust the full potential and capabilities of the medium, for the same contents can be reused in any number of places, and using portions of digital materials to create all new digital contents will be more and more commonplace in the years ahead. Anticipating these developments, a record of changes such as combining or editing digital contents is also preserved in metadata. And metadata is not just used for searching, although this is an important application. used record also to rights-related information—that is, who owns the rights to every portion of the content object—and a history of how the content was modified or edited in the past.

■ Content Distribution Environment and Requirements

Now with this next series of slides, let us consider the environment in which content distribution is now starting to evolve as a business, and highlight some of the technological developments and social changes that are occurring.



Content Distribution Environment (2)

Terminal

- Digital camera, video
- Cell phone, car navigation
- Game equipment
- ent Improved functionality, performance Reduce cost
- Set top box (STB)
- High-efficiency voltage technologies
- DRM technologies

(C) Copyright 2002 & 2003, cIDf

Improve quality

Increased bandwidth

Reduced

communications cost

Infrastructure support Multi-channel upgrade

2. Content Distribution Environment

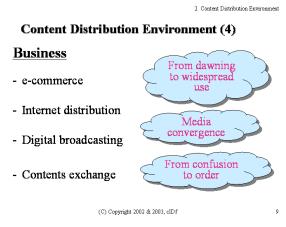
Content Distribution Environment (3)

Network

- Broadband
- Mobile (iMODE, next generation)
- FTTH/xDSL/CATV
- Always-on connection
- Internet mansion
- Individual broadcasting station
- Convergence of communications and broadcasting

(C) Copyright 2002 & 2003, cIDf

First and foremost, the cost of computer engines themselves and memories for processing digital contents continues to fall and this is happening even as the performance of computers continues to improve. Ever more compact and smaller terminal equipment is being developed that performs better and costs less than last year's model. And even networks are very different from what they were even just a year ago as more and more broadband delivery capability is brought on line.



2. Content Distribution Environmer

Increase

public

awarenes

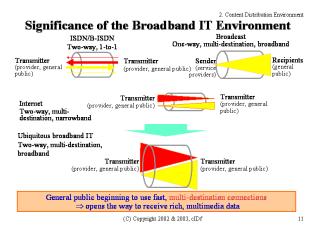
Content Distribution Environment (5) Social institutions

- IT-based strategy (IT Basic Law)
- Copyright and other administrative
- Discussion in the World Intellectual Property Organization (WIPO)
- Movement to revise copyright laws
- Provider responsibilities
- e-Japan 2002 and later Program (IT strategy)

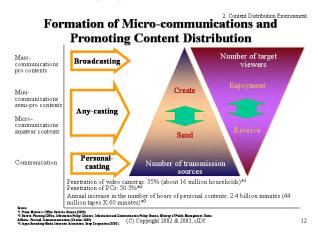
(C) Copyright 2002 & 2003, cIDf

Indeed, network-based content delivery businesses are really starting to take off with (a) the rollout of digital broadcasting a few years ago, (b) the government's stepped-up commitment to IT-based public initiatives that are now starting to be implemented, (c) and a growing public understanding and awareness of intellectual property and copyright-related issues.

Highlighting just the enormous impact of broadband, until recently data has been moving over narrow capacity lines that provided just enough throughputs for broadcasting stations to deliver contents one-way to general subscribers. Further progress saw the advent of narrowband Internet which opened the way for back-and-forth two-way delivery of substantially greater amounts of data. Now jumping to the present, networks are being upgraded to broadband that provides more than enough capacity to send large amounts of data back and forth between users. With broadband capacity, ordinary users will have the ability for the first time in history to deliver rich multimedia contents over networks, a development of enormous potential significance.



To put it another way, the situation up to now has been such that a small number of broadcasters could send contents to a large number passive recipients, or a large number of people could set up Web sites which any number of people could visit via the Internet.



Now with the greater capacity of broadband, we are embarking on a new era in which ordinary users will be able to send data to larger community-sized groups of recipients.



Also, "enrichment of network contents" has been and also is one of the five key pillars of the government's IT-based public initiative announced at the end of June 2001. This involves not only fostering the human

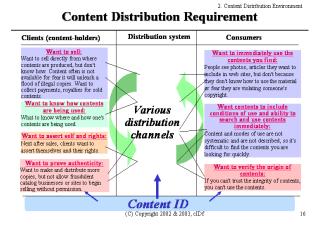
resources to create contents, but also creating the kind of contents that will promote more widespread and robust distribution. In other words, all of our efforts to deploy broadband delivery capabilities will be meaningless if the contents themselves remain mediocre and uninspired.

II. Measures in Specified Areas Content Distribution Environment During 2002, school lessons utilizing retworks will be enriched. To this end, schools will be provided with high-speed Interret access and a wide variety of educational content that utilizes retworks will be eveloped and their widespread use promoted. During 2002, school lessons utilizing retworks will be enriched. To this end, schools will be provided with high-speed Interret access and a wide variety of educational content that utilizes retworks will be developed and their widespread use promoted. Digitation of School Bidestion The utilization of School Bidestion of the utilization of th

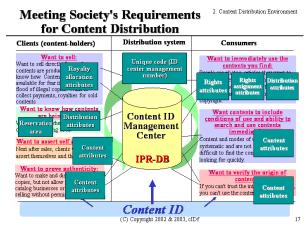
Specific objectives of the government initiative are to promote creation and use of good content for education and learning, and to promote protection and widespread distribution of contents for broadcasting. A number of guidelines and policies to promote these objectives are already starting to be implemented.



The different parties involved in the distribution of contents want different things. For example, the content producers and owners are eager to sell their contents, they want to know how and to what extent their contents are being used, and of course they want to draw attention to their products. They also need a sure way to prove that what they offer is the authentic article, the genuine content object that they created. Consumers on the other hand have different needs. When they find something they want to keep or use, they need to know who to contact and what the price or conditions are for using the contents. They also want to be able to find what they are looking for quickly and easily, and to determine that the person in possession of the contents is the legitimate owner or creator.



We advocate the Content ID as compatible with the diverse methods of content distribution, yet also satisfying the diverse needs of both content-holders and content-consumers. We observed earlier that the Content ID includes various attribute data relating to digital contents that are stored as metadata in a logical database. By accessing this information, this provides a robust framework that satisfies all requirements for virtually any content distribution situation.



Content ID Requirements

Now let's take a closer look at the Content ID itself. Very simply, the Content ID is a unique identifier that can specify and distinguish any kind of digital content that is distributed. As a unique code attached to a content object, the Content ID serves well enough as an identifier, but actually it is much more than just that.

3. What is Content ID

18

What is Content ID?

A unique code assigned to a content object that uniquely identifies the content.

The Content ID (cid) uniquely identifies the content itself, and the various attribute data (metadata) associated with the content.

(C) Copyright 2002 & 2003, cIDf

It is also the key to a complete set of attribute information about a content object stored as metadata including the nature of the contents, rights-related information, information about distribution, and more. The Content ID provides the key enabling metadata to be uniquely associated with a particular digital object.

A unique code or ID must satisfy five basic requirements. First and most obviously, there can't be more than one instance of the number floating around in the global environment. If the same number were assigned to different contents that were distributed over the Internet, then the number clearly could not be for distribution management and other processing tasks. Second, the ID must be permanent. Once a content object with an assigned number is put out on the Net, it could be out there is cyber space for a long time. It would negate the whole purpose if the numbering system were to suddenly become unavailable, so the numbering system must be permanent. Third, there must be a system in place like the Domain Name System (DNS) that allows you to access an address through the ID to find out the metadata-based attribute information on a content object that has been made available on the Net. This is called location resolvability. Forth, the ID must be inseparable from its contents. Here again, the purpose of the system would be defeated if the contents could be separated from its ID number after the content was distributed over the Net, so the scheme must ensure that the number cannot be separated or removed from the content. Fifth, there are already a lot of numbering schemes in existence besides the one proposed by the cIDf so it would be almost impossible at this point for everyone to unify around a single scheme. This means that the Content ID must be compatible and inter-work with other numbering schemes.

3. What is Content ID?

Content ID Requirements

→ Unique

No duplicate in the global environment

→ Permanent

Continue to exist based on established rules of inheritance

→ Location Resolvability

Management and other data can be accessed using the number

→ Inseparable from Contents

ID cannot be separated from contents

→ Interoperable

Compatible with other numbering schemes

(C) Copyright 2002 & 2003, cIDf

The cIDf settled on a Content ID system that satisfies these conditions. Basically, the format consists of two parts: the unique number part that I mentioned earlier, and a content attribute part expressed using metadata. The entire Content ID number is stored in a database in an ID Management Center. The unique ID number portion is then embedded in the content object by digital watermarking and the contents are distributed.

However, it would be very inconvenient if you had to access an ID Management Center every time any processing was done, and of course there are also situations where one is not always connected to the network. This led us to devise a scheme in which a subset of attribute data called the *Distributed Content Descriptor* (DCD), is bound to the contents and stays with the contents wherever it is distributed. The Distributed Content Descriptor is written in Extensible Markup Language (XML) and includes data that, once set, cannot be altered including the *title* of the content and the *creator* of the content. The DCD also contains information about conditions for using the content.

ID (Descriptor) and Forms of Metadata

Content

Metadata

Content

Metadata

Content

Metadata

Content

Content

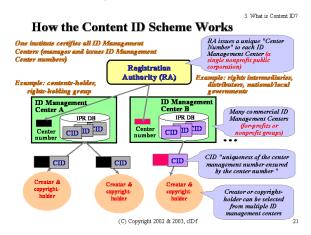
Richts
arributes

Arribu

■ Content ID Management

We realize that there are plenty of content-holding individuals and businesses out in the world that are currently managing their content resources based on their own proprietary numbering systems. Now many of these groups have set up ID Management Centers

to administer their content holdings and attribute information, but as soon as these internal numbers are released without modification into the wide world beyond, their uniqueness is lost.



We addressed this problem by setting up a Registration Authority (RA) as the one and only public certification and registration authority that issues a unique Center Number to each of the ID Management Centers with content holdings. By combining this Center Number with the ID Management Center's own internal number, a unique code is created that can be embedded in the content objects that are put out on the Net.

Now I will briefly describe the various kinds of attribute data that are included in the Content ID format defined by the cIDf. First Content Attributes help reveal the nature of the contents and include the title, the author or creator, keywords, and a brief abstract. Next, Rights Attributes provide rights-related information pertaining to the contents. Note that these Content and Rights Attributes are set when the content object is first created. By contrast, the Rights Assignment Attributes convey information about the contractual or licensing terms when the rights to the contents are assigned to another party and are only added to the Contact ID when the contents are actually distributed. Next, the Distribution Attributes specify the purposes and range of uses of a content object when it is distributed, and conditions on use such as whether the content can be copied. The Royalty Allocation Attributes defines how any income derived from the sale of contents would be allocated among the rights-holders to the contents. In addition, the Content ID format also includes a Free Area and area for System Information. All together, more than 200 items of information have been standardized as meaningful attributes of content objects by the cIDf. Let me also emphasize that only about 20 of these attribute items are essential, and the rest are used only for certain types of contents or at certain times when necessary.



- Unique Code: Number uniquely identifying a content object consisting of "ID management center number," "ID management center internal number," "type (version information)," etc.
- Content Attributes: Information about content (= work, product) including creator, substance, type, category, etc.
- · Rights Attributes: Rights-related information, and contact location
- Rights Assignment Attributes: Information regarding licensing, assignment, conveyance of rights. Terms and conditions of licensing agreement.
- Distribution Attributes: Content usage rules, distribution information.
- Royalty Allocation Attributes: Information about allocation of profits from content sales.
 Free Area: For use by the ID Management Center. Could show distribution (= sales)
- history or other data enhancing the Content ID.

 (System Information): Closed system information needed by the ID Management Center. E.g., digital signature, watermarking information, check digit, content hash value.

(C) Copyright 2002 & 2003, cIDf

Using the metadata defined by the cIDf, you can accurately identify the various metadata components that have been combined to make more complex content objects, and even trace back through the editing history to see how that complex content object was created. The distribution chain of any complex content object will involve a succession of content ID numbers from other systems. Since the inherited history of these other content ID numbers is included in the Content ID metadata, you can access the metadata of other systems.

Boscribing Content Structure with Metadata Example Musts Video Image Text Content attributes Root content 1 Root content 1 Root content 3 Root content 4 Root content attributes Content 15 Root content attributes Content attributes Content 15 Root content attributes Content 15 Root content attributes Content 15 Root content attributes Content attributes Content 15 Root content attributes Content 15 Root

Also, it is common for multiple rights-related items and multiple rights-holders to be associated with a single content object, and the distribution of content objects is based on the licensing agreements between the various rights-holders and licensees. For dealing with these complex contractual requirements, the cIDf metadata is quite capable of representing all necessary data associated with multiple licensing agreements. This rights-related information can be accessed quickly and efficiently using the Content ID as a key.

(C) Copyright 2002 & 2003, cIDf



Units that can be Assigned a Content ID

- Work that is copyrighted or claimed to possess originality (self-asserted work)
- Work created to sell (a product)
- Combined or assembled work: published work, movie (work as a single unit)
- The minimal work beyond which the object is not recognized as a unique work: 1 page, 1 paragraph, 1 cut, 1 sentence, 1 scene, etc. (work as a minimal unit)
- New ID for edited work (edited work, composite work)
- Acquisition of an ID depends on the intent of the copyrightholder

(C) Copyright 2002 & 2003, cIDf 2

For example, a single ID number can be assigned to an entire movie, or separate IDs can be assigned to an entire movie, or separate IDs can be assigned to any user-defined portion of the movie: separate scenes, cuts, and so on. Or IDs can be assigned to parts of an

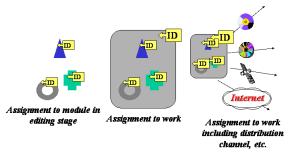
object, and a different ID assigned to the larger whole content object that encompasses those parts.

Also, remember that there are a variety of channels or ways in which contents can be distributed. For example, contents may be delivered as a DVD or some similar package format, or it might be broadcast or downloaded over the Internet. Since the distribution conditions will likely vary depending on the distribution channel, different ID numbers could be assigned to the very same object depending on the distribution channel. This ability to assign different ID numbers for different modes of distribution is one of the features that distinguish our Content ID system from other ID numbering schemes.

Indeed, this particular feature of the Content ID—the ability to assign different Content ID numbers to the same basic content object that are distributed over different channels—was also taken up for discussion by the Society of Motion Picture and Television Engineers, the SMPTE.

3. What is Content ID

Method of Assigning ID to Digital Contents



(C) Copyright 2002 & 2003, cIDf

28

Representing Rights with Metadata: Initial Agreement | Initial agreement | Rights licensing agreement | Licensee | Licen

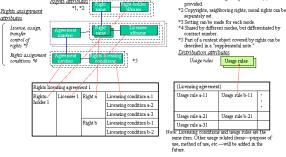
Content 1

[Licensing agreement)

| Usage rule -11 | Usage rule -21 | Usage rule -31 | Usag

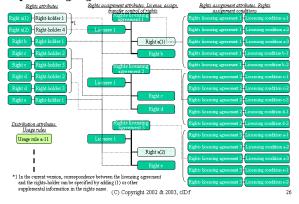
3. What is Content IT

Representing Rights with Metadata: Agreement to Metadata



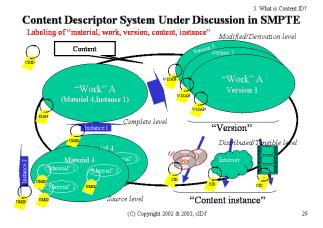
(C) Copyright 2002 & 2003, cIDf

Representing Rights with Metadata: Metadata Details

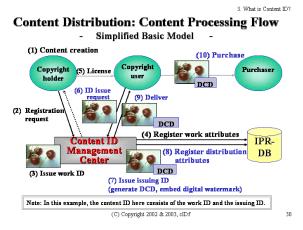


■ Assigning ID and Resolution

The Forum is essentially neutral on the question of what unit of content the ID can be attached to. The creator or distributor of the content object is free to assign IDs to whatever unit of content he or she wants to administer and track.

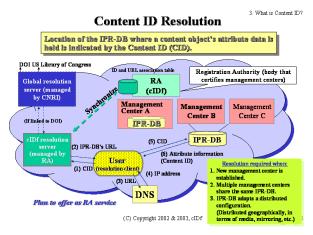


Now let's take a look at an example showing the actual flow involved in the distribution of digital contents using the Content ID. This is a very simple, basic example in which the unique Content ID code consists of only two sub-codes: the first part is a unique *Work ID*, a number representing the work itself, and the second part is an *Issuing ID* that varies for different channels or modes of distribution.



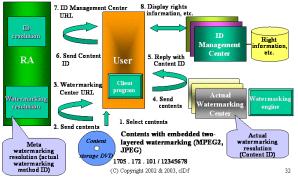
First, the creator of the content registers his or her own work at an ID Management Center. At this point, the first part of the unique code, the Work ID, is issued. In other words, all the attribute data pertaining to the work, including the Content Attributes and the Rights Attributes, are recorded in the database. Now let us assume that some time later a publishing company obtains the right to use the material, and wants the contents to be delivered. At this point, the publishing company sends notification to the Content ID Management Center where the work is registered requesting that the complete Content ID number be issued. Then the additional attribute data regarding the licensing and distribution conditions are added to the database, the full unique code combining Work ID and Issuing ID is attached to the content, and if necessary digital watermarking is embedded in the contents. Later, a required subset of attributes is extracted from the attribute data to generate a DCD (Distributed Content Descriptor), and the publishing company

makes the content publicly available over the Net. This is the basic model that we have in mind. In a real-world environment, there could be many more complicated elements, and of course the model would vary depending on the nature of the business.



I observed earlier that you can find out all kinds of attribute information regarding contents from the Content ID. We call this Content ID resolution. Assume for a moment that a content object with an assigned ID embedded by digital watermarking is distributed over the Internet and a user in cyberspace gets hold of the contents in some form. Assume further that the subset of attribute information in the DCD is not enough. You want to know all the information that's available for the content, but you don't know which database or which ID Management Center to ask for the information. You do however know the ID number, so in this case you can send the ID number to the Resolution Server. The Resolution Server will then return the URL of the database where the attribute data is stored, and the full range of attribute information on the content can be obtained by accessing the URL. To evaluate this approach, we enlisted some 50 companies back in the fall of 2001 for a series of content distribution trials using the Content ID and including the resolution framework described. The trials were conducted in February 2002, and confirmed that the Resolution Server-based scheme worked very well.

2002 CID Registration Authority Trials (Watermarking Resolution)

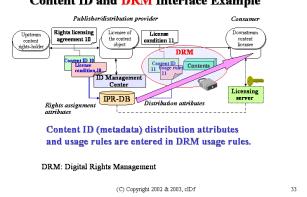


■ Content ID Application

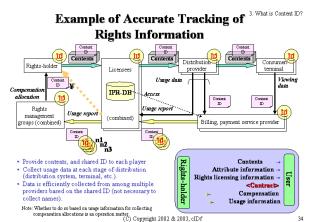
Today all kinds of content distribution businesses are springing up, and the number one worry among the content-holders—as you might expect—is how easily digital contents can be illegally copied and stolen. The next biggest concern is that these operators want to be able to accurately determine how much a content object is actually used, and whether the content has been paid for.

This is why so much interest is now focused on Digital Rights Management that provides the technology needed to protect digital contents from piracy. Contents are actually distributed based on distribution conditions within the scope of the terms and conditions of the licensing agreement with the rights-holder, and all this rights-related information—licensing conditions, distribution conditions, and so on-is managed as Content ID attribute data through an Intellectual Property Rights Database (IPR-DB). These distribution conditions are used by entered them in some form in the DRM used by each company.

Content ID and DRM Interface Example

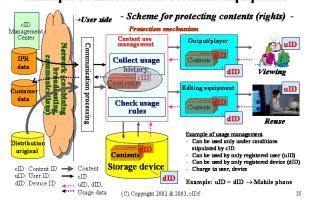


To track the status of content use, the ID is attached to the contents in much the same way as a barcode on a product. One can thus easily gather all kinds of data corresponding to each ID number, and another major benefit is that accurate content usage reports can be quickly and easily generated.



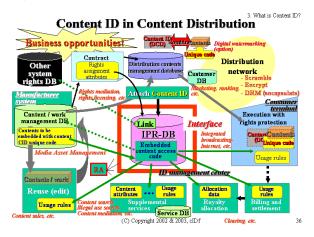
Another approach can be implemented by assigning IDs to terminal equipment, disks, and other kinds of hardware devices (dID). Using this approach, one can check to see if a content object can really be opened, viewed, or used simply by comparing the Content ID with the ID of the disk, the terminal, the player, or the editing equipment. We envision that this information—whether the contents were actually viewed, used, or saved—could be stored on a terminal, and then forwarded at regular intervals to a management server along with the Content ID.

Example of Execution in Terminal Equipment ${}^{3,\,\text{What is Content ID?}}$



Now to briefly summarize, I have shown that smooth and efficient distribution of digital contents requires the ability to manage and protect copyrights, a billing mechanism, a way to manage distribution, and other capabilities. All this and more can be accomplished with the Content ID framework and other related technologies.

3. What is Content ID?



Functions Needed to Facilitate Content Distribution

- (1) Copyright functions (grasp of copyrights and other rights-related matters, eliminate complexity of licensing procedures, ensure transparency)
- (2) Billing functions (expedite fair compensation to creators, providers)
- (3) Barcode function (promote collection of distribution history, sales data, and other marketing
- (4) Net police functions (ID watcher functions)
- (deter unlawful copying by tracking illegal uses) (5) Portal site functions
- (supports better searching, interoperability using shared ID codes)
- (6) Privacy protection functions (prevent leaking of usage logs and other personal information)
- (7) Direct distribution of content from creation site function (Promote distribution of contents directly from the creator)

(C) Copyright 2002 & 2003, cIDf

■ Content ID Forum (cIDf)

Finally, I would like to say just a few words about the Content ID Forum.

4. Content ID Forum (cIDf)

Content ID Forum (cIDf)

The Forum was established on August 4, 1999 at the urging of the University of Tokyo Professor Hiroshi Yasuda to implement a framework promoting the distribution of digital contents over networks while protecting copyrights by assigning a unique code (Content ID) to each digital object.



The Forum was established in August 1999 at the urging of Professor Hiroshi Yasuda at the University of Tokyo. We realized the first step in promoting distribution of contents was to build consensus behind the idea of allocating unique ID numbers to the full range of different kinds of content objects. And because the distribution of contents must be based on a contractual agreement, the second step is coming up with common terminology for representing the substance of these agreements that can be included as part of the ID. The third step is to enable the communication and other devices to properly execute the copy control and other events represented by this common terminology. Once these three steps are accomplished, we will try to extend and reinforce global distribution of digital contents by promoting the legal standing of the Content ID framework. At this point, we have finally achieved step one and built a pretty good consensus behind the idea of assigning unique Content ID numbers to digital contents.

Steps for Implementing Contents Distribution

Step 1: Consensus behind a shared scheme for specifying attribute information associated with digital content objects (identification by Content ID).

- Step 2: Develop a common terminology to represent ontractual terms and conditions.
- Step 3: Develop hardware (communication equipment, home and office equipment) that correctly executes content distribution-related events specified in the common terminology
- Contract terms based on common terminology will support execution, citizens' rights, and eventually its legal status will be recognized.
- Step 5: Promote distribution of electronic rights. (cIDf Specification, Ver. 1.0, Section 2.2.3)
 (C) Copyright 2002 & 2003, cIDf

Activities of the cIDf mainly concentrate on defining frameworks for assigning ID numbers, and developing and proposing viable open standards in this area. We develop distribution models using the Content ID, conceive concrete methods for applying the ID, and pursue efforts to spread and educate people regarding the importance of the Content ID both in Japan and around the world. While forging ties with related industry groups and official standards organizations in Japan and abroad we make proposals, and try to harmonize cIDf standards with other standards developed by other standards organizations.

4. Content ID Forum (cIDf)

cIDf Activities

- 1. Promote adoption of ID (Content ID) capable of identifying all digital content objects.
- 2. Develop an Intellectual Property Rights Model (IPM) based on the Content ID.
- 3. Conceive concrete ways to use and manage the Content ID
- Promote the adoption and awareness of the Content ID in
- Promote international standardization, and harmonization with other international standards.

(C) Copyright 2002 & 2003, cIDf 40



The Forum has a broad-based membership including some 210 organizations, many corporate members across a variety of different industries, as well as individual members. For more detailed information about cIDf activities or about Content ID specifications, we hope you will visit our Website at http://www.cidf.org/.



Thank you very much for your interest and attention.