stm
INTERNATIONAL ASSOCIATION OF SCIENTIFIC, TECHNICAL & MEDICAL PUBLISHERS
AN AUTONOMOUS ASSOCIATION AFFILIATED TO THE INTERNATIONAL PUBLISHERS ASSOCIATION

INFORMATION IDENTIFICATION

A Report to
STM International Association
of
Scientific, Technical and Medical Publishers

Prepared by
Douglas Armati
Jackson Brevis Limited
Woodbridge, Suffolk
England

Commissioned by the
STM Task Force on Information Identifiers
and
Metering Systems in the Electronic Environment

---

stm

INTERNATIONAL ASSOCIATION OF SCIENTIFIC, TECHNICAL & MEDICAL PUBLISHERS

'AN AUTONOMOUS ASSOCIATION AFFILIATED TO THE INTERNATIONAL PUBLISHERS ASSOCIATION'

INFORMATION IDENTIFICATION

A report to STM

Brief introduction

At the initiative of the STM Chairman, John Dill, a task force on Metering Systems and Information Identifiers in the Electronic Environment was created.

Arnoud de Kemp (Chairman of the STM Innovations Committee) was appointed to chair the Task Force and several members and colleagues have been involved: Maurice Long, Norman Paskin, Charles Clark, Alexis Koutchoumow, Douglas Armati, with John Dill, Herman Pabbruwe, Jean-Manuel Bourgois, Herman Frank and Stephen White as corresponding members.

The task force met twice - on 8 December 1994 in London and on 10 February 1995 at the STM Secretariat in Amersfoort. They commissioned a report to Douglas Armati, which, after an initial discussion and various enhancements, could be presented at the St. Louis Conference on 24 April 1995 and to the Executive Board on 26 April, 1995. Copies were also distributed at the IFRRO board meeting in Copenhagen on April 19-20.

This report and the recommendations made will help STM to start or at least stimulate discussion of an international initiative for multimedia information identifiers and metering systems. The task force is grateful to Douglas Armati for his collective work which has proven to be very comprehensive.

The STM Executive Board accepted Armati's report as a formual document to be made available to the STM members.

Amersfoort, June 1995

SECRETARIAT: MUURHUIZEN 165|3811 EG AMERSFOORT|THE NETHERLANDS|TELEPHONE: +31 33 65 60 60|FAX: +31 33 65 65 38

BANK ACCOUNT ABN AMRO BANK|KONINGSPLEIN|1012 WP AMSTERDAM. ACCOUNT NR. 46.69.77.743

Jean-Manuel Bourgois
38, rue du Tage
75013 Paris
France
Tel: +33 1 45 88 41 88
Fax: +33 1 45 88 55 92

John F. Dill
Mosby-Year Book, Inc.
11830 Westline Industrial Drive
St Louis MO 632416
USA
Tel: +1 314 872 8370
Fax: +1 314 567 0190

Herman Pabbruwe
Wolters Kluwer Academic Publishers Group
Postbus 989
3300 AZ Dordrecht
The Netherlands
Tel: +31 78 33 4283/4
Fax: +31 78 33 4268

Maurice Long
BMJ Publishing Group
BMA House
Tavistock Square
London WC1H 9JR
United Kingdom
Tel: +44 171 387 4499
Fax: +44 171 383 6402

Arnoud de Kemp
Springer-Verlag
Tiergartenstraße 17
D-69121 Heidelberg
Germany
Tel: +49 6221 4870
Fax: +49 6221 413982

Contact Information

Douglas Armati
Jackson Brevis Limited
7a Angel Lane
Woodbridge
Suffolk IP12 4NG
United Kingdom

Telephone: +44 1394 380874
Facsimile: +44 1728 453909
Email: armati@udid.u-net.com

Lex Lefebvre
Secretary
STM
Muurhuizen 165
3811 EG Amersfoort
The Netherlands

Telephone: +31 33 65 60 60
Facsimile: +31 33 65 65 38
Email: lefebvre@stm.nl

**STM International Association
of Scientific, Technical and Medical Publishers
is an autonomous association affiliated to the
International Publishers Association**

Dr Norman Paskin
Elsevier Science
The Boulevard
Langford Lane
Kidlington
Oxford OX5 1GB
United Kingdom
Tel: +441865 843 000
Fax: +44 1865 843 010

Stephen White
Sweet & Maxwell
183 Marsh Wall
London E14 9FT
United Kingdom
Tel: +44 171 538 8686
Fax: +44 171 537 6613

Corresponding member:
Herman Frank
Elsevier Science S.A.
Avenue de la Gare 50
CH-1003 Lausanne
Switzerland
Tel: +41 21 320 7381
Fax: +41 21 323 5444

Contents

Information Identification: Executive Summary

The Opportunities

- The economic value and strategic importance of intellectual property (IP) is growing rapidly
- IP is being traded more frequently in digital, networked environments
- A substantial global market exists for valuable IP delivered via networked digital devices
- Rapid, low cost, interactive access to these assets would be a boon to users
- Exploitation of this market provides growth opportunities for IP rights (IPR) owners as well as for suppliers of networks and digital devices
- An open market in IP assets would potentially add value to the portfolios of all participants
- Solutions enabling cost effective dynamic licensing offer the most promise

The Threats

- Networked digital devices facilitate simple, quick, cheap reproduction of valuable IP assets
- Existing solutions provide limited protection for IPR owners
- Existing standards do not support licensing of data objects smaller than a complete work
- Proprietary identification, security and trading of IPR based assets is expensive
- Proprietary solutions do not allow open network trade in IP assets
- No effective means exist to identify unlicensed uses of IPR in open networked environments
- IPR owners have good reason to be concerned about losing control of their assets in this domain
- They are naturally reluctant to license use of their IP without adequate protection

Action

- Standardize technological solutions for identifying, securing and trading IP assets
- Design dynamic functional license management specifications around common framework
- Within the architectural framework provide an adequate environment for IP use management
- Agree standards that:
  - enable automated dynamic licensing of uses of data objects from mixed sources
  - take account of the multilingual, multiformat, global nature of the marketplace
- Develop identification tools that enable rapid, cost effective links to rights management systems
- Construct a common IP rights data model for use by IPR owners and rights management organizations
- As far as is possible, codify and strengthen the existing legal framework
- Set maximization of value, distribution efficiency, interoperability and open trade in IPR based assets as the goals of the standardization process

Who?

- These standards should be agreed on a pre-competitive basis
- Content, network and digital device suppliers should participate in the standards process
- Representative international associations should be involved, where possible and appropriate
- Open lines of communication should be maintained with other market and public stakeholders
- International standardization agencies should be involved

Identifiers: Under Development

- **Need for identifiers is widely perceived**
  - active development of identification schemes is underway in many industries
  - many projects assume unifying scheme will emerge
  - pilot projects generally use bespoke or existing industry based identification systems
    - » widespread commercial use of this approach will lead to "Tower of Babel" in open markets
  - crucial data object granularity issues not addressed
    - » pilot projects are using file/document level identification and management only

- **Information Distribution Technologies**
  - Universal Data Identification (UDID) (see Appendices 6 & 7) codes should ideally be used in all information distribution transactions involving reporting to or dealing with external third parties
  - Information distribution technologies will only develop from highly bounded proprietary systems to relatively boundless standard open systems when automatic rights control can be exercised
    - » Suitably designed UDID codes will provide the hooks to enable such systems to be built

- **Critical issues**
  - flexible, granular (sub-file level) data object identification
  - identification to survive
    - » movement of data from one operating system to another
    - » movement of data from one application to another
    - » movement of data between different networks
    - » expression of data in different hardware systems
    - » less than ideal network and storage environmental conditions
  - real time identification of rights in a particular data object
    - » enabling differential pricing in different global markets without leakage to higher priced markets
    - » enabling national treatment according to local cultural and legal norms
    - » ensures neutral or positive influence as tool in world intellectual property trade policy
  - real time differentiation between master data object and individual expressions of whole or parts of it
    - » this implies inbuilt replication control systems
  - real time, trivial cost, locally generated data object encoding systems
  - privacy not compromised
  - codification of licensing agreement terms required for automation of rights management systems
    - » this is necessary to enable real time open network rights management
  - legal
    - » matching identification and transaction technologies to differing legal regimes
  - time
    - » leakage minimized by decisive harmonious action
    - » change will take time
      - to encode valuable back catalogue
      - for all content owners to adopt standard coding system
      - for all computer and communications wares vendors to adopt open standard
      - to build supporting rights management and network infrastructures
  - cost
    - » lowering transaction costs while enhancing security is good business
  - recognition of interdependence of interests

Action Agenda

- **Using existing standardization procedures as a model, on a pre-competitive basis, establish small, global, broad based, high level, ad hoc task force of content controllers involving major content organizations from Europe, North America, East Asia and Japan.**
  - To agree need for universal approach to intellectual property identification and commit to implementation of resulting system
  - To set functional specification
    - » establish comprehensive set of questions requiring resolution
    - » delegate working groups to analyse and report on each
    - » drawing strongly on legal and technical advice
    - » integrate with existing industry-based identification systems
    - » communicate openly with participants and non-participants
  - To establish international administrative structure for long term promotion and management of the standards, including small professional support team
    - » stressing global standards, local implementation
  - To represent a combined voice on international policy issues in this domain

- **Main task force to establish small multi-disciplinary task force of technical, legal and commercial leaders from content controllers, computer, communications, software and hardware industries.**
  - To produce a comprehensive universal open standard data identification protocol and requisite agreements on open interchange standards for hardware and software
    - » coordinating role
    - » delegate to working groups to answer specific questions
    - » encourage pre-competitive agreement among computer and communications organisations to facilitate use of standard universal data identification (UDID) codes across operating systems, networks and applications
    - » involve ISO, ITU and other standards bodies as necessary
  - To make recommendations on the logical structure and physical features of a UDID protocol that is
    - » format independent
    - » object type independent
    - » unique to an object
    - » unique to the expression of an object
    - » easy and cheap to generate and use
    - » able to accommodate all known object types
    - » compatible with existing archives of objects
    - » an essential and integral part of the data structure of the object
  - To coordinate activities of competent organizations in development of comprehensive system
  - To advise main task force on choice of standard system
  - To oversee implementation of the standard system
  - To advise on standards, including upgrades, once established

- **Establish cross industry working party to develop and deploy common copyright data model to enable automatic rights management**
  - identifying
    - » Interested Parties (natural and legal)
    - » Works
    - » Unique Data Objects
    - » Codified terms of agreements between interested parties in relation to works and unique data objects

- **The Domain**
  - ***Secure identification of valuable digitized copyright content***
    - From physical packages to digital files to file contents
    - From stand alone to client/server to fully distributed systems
    - Immutable identity for each valuable data object
    - Digital DNA?
      - traceable
      - self-replicating
      - positively related through the generations
    - Codes as agents for the design and assembly of data
      - basic building blocks of all valuable information
  - ***Simple, universal, reliable, flexible, coherent metadata system***
  - ***Silicon, software and society***
    - Identification issues touch every aspect of our social constructs
      - reaching international agreement on core standards
      - building systems that reflect local legal and cultural values
    - We are engaged in committing our legal and cultural values to silicon and software

- **Introduction**
- This report extends the exploratory map of the new planet. This is "the world of digital technology products, services, platforms and media" -- described by Charles Clark in The International Publishers Copyright Council report The Publisher in the Electronic World (Turin, May 1994).

  It results from an initiative taken by STM Chairman, Mr John Dill, following the STM Annual General Assembly in Frankfurt in October, 1994. The STM Task Force was formed and met for the first time in December, 1994, under its Chairman, Mr Arnoud de Kemp. This study was commissioned by the Task Force following that meeting.

  The domain it surveys is secure identification of valuable digitized copyright content.

  The task of identifying the contents of relatively stable physical packages has proven challenging enough. Now owners of valuable copyright based assets are faced with devising new schemes. These will need to allow for simultaneous identification of highly dynamic electronic packets, the contents of those packets, the parties interested in those contents and the rights agreements relating to the contents.

  In the physical world it was enough to identify types of homogeneous content packages. Only commercial manufacturers had the equipment necessary to create these packages together with reasonable control over the distribution of them.

  On the new planet it is possible for anyone with relatively low cost equipment to rapidly multiply perfect clones of electronic packets. In the absence of any control mechanisms it is possible to extract the value from the packet without committing it to local storage devices (except in a fleeting sense) at all. If rights to the valuable content carried in these packets are to remain meaningful, it will be necessary for each valuable element of the contents to have its own immutable identity. This identity will need to be part of the fabric of the content, carried in it at all times.

- **Introduction (cont'd)**

Computer relationships maturing. Once the stand alone machine dominated the industry. Now distributed client/server architecture is becoming the dominant model. Indications are this is a staging post on the way to fully distributed computing.

The required data object identification system may need characteristics very like DNA and RNA to be able to generate codes that are unique, permanent, traceable, self-replicating, positively related through the generations. In order to do their job reliably, it is possible these codes could even be required to be the agents for the design and assembly of data, the basic building blocks of all valuable information.

Once a simple, universal, reliable, flexible, yet coherent data content identification system is agreed, it will then be possible to build rapid access rights transaction and management interchange systems, preferably as a part of the core of computer equipment, operating systems and applications.

This study seeks to address some of the central issues of rights identification in this new domain. It is a complex subject. It is hoped this report in some small way helps to clarify a way forward.

These issues must be addressed before we can sensibly recommend moving highly valuable assets into the more uncharted areas of this burgeoning new realm.

Secure network delivery is only part of the story. Acceptable solutions to permanent identification of valuable content components must be developed, agreed and implemented. Once in place, the promise of the information century may be realised.

Identification issues touch every aspect of our social constructs. It is no wonder, then, that they are seen as an important piece of the information infrastructure jigsaw. Reaching international agreement on core standards is an essential pre-requisite to building systems capable of reflecting local legal and cultural values.

We are engaged in committing our economic and social system to silicon and software. In the process our view of intellectual property may change. Already the fluidity and flexibility afforded by sophisticated communications, applications and operating systems software are blurring the picture.

In the course of this study it has been a privilege to deal with some of the remarkable people who are helping to build the bridges to the new world. They have been generous with their time and contacts. All seem possessed by the same sense of urgency -- to find a solution to this problem as soon as possible.

Douglas Armati
Consultant

on behalf of the
STM Task Force on Information Identifiers and Metering Systems in the
Electronic Environment

March 1995

■ **Relevant International Projects**

Any examination of identification systems in the digital domain is best conducted from a global perspective. Cross-border internetworking of computers makes common international protocols essential. The relatively painless interconnections that provide the impression of a single virtual machine are themselves the result of rigorous standardization and international cooperation in the telecommunications industry.

Following is a review of some of the most relevant projects and products hat will impact on the debate about the most appropriate approach to information identification:

- ***CISAC/BIEM International Numbering Working Group***
  - » International Standard Work Code (ISWC) (see Appendix 2)
  - » The Common Copyright Data Model (see Appendix 3)
- ***CISAC/SUISA***
  - » CAE list extensions to include neighbouring rights
- ***European, US and Japanese IT Association cooperation***
  - » EUROBIT, ITI and JEIDA working towards common position
- ***G7/GII***
  - » Protect Privacy and Personal Data
  - » Increased Information Security
  - » Protect Creativity and Content Provision
  - » Share Experiences on Emerging Applications
  - » Act as Catalyst for the Protection of Research, Applications and Generic Services
  - » Promote Joint Ventures to demonstrate commitments
  - » Pilot projects
    - Global Inventory
    - Global Interoperability of Broadband Networks
    - Cross Cultural Training & Education
    - Electronic Libraries
    - Electronic Museums and Galleries
    - Environment and Natural Resources Management
    - Global Emergency Management
    - Global Healthcare Applications
    - Government Online
    - Global Marketplace for SMEs
    - Maritime Information Systems
- ***International Council of Scientific Unions (ICSU)/UNESCO***
  - » Reviewing the domain

Information Identification: Identifiers: Under Development

- **Relevant International Projects (cont'd)**
  - ***International Federation of Library Associations (IFLA)***
    - » Functional Requirements for Bibliographic Records
    - » Text Encoding Initiative (TEI)
    - » Subject Indexing
    - » Automated Systems for Access to Multilingual and Multiscript Library Systems
    - » Network storage/re-use
  - ***International Federation of the Phonographic Industry (IFPI)***
    - » International Standard Recording Code (ISRC) promotion (see Appendix 1)
    - » Source Identification (SID) code promotion (with Philips)
  - ***International Federation of Reproduction Rights Organisations (IFRRO)***
    - » Metering systems report (October 1994)
  - ***International ISBN Agency (see Appendix 4A)***
    - » Extensions to ISBN system
  - ***International Standardization Organization (ISO)***
    - » International Standard DIS 10164-10: Information Technology - Open Systems Interconnection - Systems Management: Usage Metering Function
      - use in conjunction with
        - CCITT Rec. X.700 | ISO/IEC 7498-4
      - positioned in application layer of CCITT Rec X.200 | ISO/IEC 7498
      - defined according to model provided by ISO/IEC 9545
      - role of systems management functions described in CCITT Rec. X.701 | ISO/IEC 10040
    - » JTC1 standards and TC46 and TC154 also relevant
  - ***International Telecommunications Union (ITU)***
    - » DIS10164-10 (above) is also known as ITU-T Recommendation X.742 Draft
      - X.700 series - Network Management
      - X.900 series - Security
  - ***Internet Engineering Task Force (IETF) (see Appendices 4, 4E & 4F)***
    - » Unique Resource Identifiers (URI), Locators (URL) and numbers (URN)
      - based on URL s developed for locating files on World Wide Web servers
      - work to date has focussed on cataloguing, locating and accessing files
    - » Internet Assigned Fields Authority (IAFA) catalogue like templates
    - » Identification working group beginning to look at granularity issues
      - expect incremental change, led to some extent by proprietary solutions
      - leading players:: strong telecoms transaction orientation
      - not strongly IPR oriented at present

Information Identification: Identifiers: Under Development

- **Relevant International Projects (cont'd)**
  - ***Internet Society***
    - » Overviewing and developing the Internet: interoperability and interconnection
  - ***World Intellectual Property Organization (WIPO)***
    - » Charged with IPR/GII responsibilities by Brussels G7 Ministers meeting
    - » Voluntary numbering project reconvened. New directions?
    - » Conference in Mexico in May: Copyright and the Superhighways
    - » Conference in Italy on The Globalization of Markets for Copyright
  - ***W3 Consortium***
    - » MIT-CERN joint initiative
    - » Developing open standards for the Internet
      - secure transaction issues are high on agenda
      - not currently addressing IPR identification standards

Information Identification:
Identifiers: Under Development

- **Significant International Corporation Activity**
  - ***Adobe***
    - » Acrobat Portable Document Format (PDF) extensions
      - enables common viewing and printing platform for documents originally created in many different applications
        - *including graphic and photographic formats*
      - reader software support for audio and video with release of Version 3 (early 1996)
      - HTML capabilities summer 1995
      - SGML capabilities late winter 1996
      - Open to advice on object identification issues
      - No work yet on fine granularity issues
    - » Acrobat 2.0
      - much tighter use control
        - possible to restrict to viewing only
        - copying, printing, modifying are all preventable
      - *embedded notes can be added without affecting the underlying document*
        - tight version control
  - ***ADONIS***
    - » World's largest holder of electronic copyright
    - » CD-ROM delivery and print metering of journal articles
      - 12,000-15,000 pages indexed, scanned and committed to CD-ROM each week
      - over 600 journals from more than 70 publishers
      - participant in CITED project
      - failed to persuade every vendor to add their article identifier
        - as a result they now use ISSN plus own in house allocated codes to identify articles (see Appendices 4 and 4A)
      - no viewing control
      - *post event invoicing*
    - » Examining online delivery and dynamic transaction management systems
      - dial-up access to Adonis collection will become available summer 1995
        - page images to be faxed back on demand
  - ***Digicash***
    - » Secure, anonymous, network based digital cash software (ecash)
      - mirroring physical cash based economic activity: may provide rights clues
      - involved in anonymous cash aspects of EU's CAFE project

- **Significant International Corporation Activity (cont'd)**
  - ***Elsevier***
    - » Standard Serial Document Identifier (SSDI) (see Appendices 4 and 4B)
      - Also applicable to books
    - » The University Licensing Program (TULIP)
      - started March 1991, in cooperation with nine US universities
      - each site has integrated the program differently, according to budgets, existing hardware and software and information policies
  - ***First Virtual***
    - » Public protocol for Internet Payment System
      - larger publishers can configure their servers using FV software to automatically charge for information the servers distribute
      - small information offerings can be sold through FV's Infohaus server
      - users must first obtain First Virtual Account ID
      - user needs no special software
  - ***IBM***
    - » Many cooperative ventures in digital library domain
      - Researchers at Almaden Research Center (ARC), Poughkeetsie, TJ Watson, Santa Teresa, Germany and elsewhere
        - exploring architectural and software implications of IPR transaction management systems
      - Working closely with publishing industry
      - Keen to get detailed functional requirements established
    - » OpenDoc computing model (Component Integration Laboratories)
      - CLI is an association founded by Apple Computer, IBM and WordPerfect
  - ***ISI***
    - » Online Electronic Library Project
      - in conjunction with:
        - IBM Almaden Research Center (ARC)
        - STM publishers
        - Library community
      - to allow publishers and the library community to test electronic distribution variables
        - Lotus Notes front end
        - full images of articles available for viewing (or other transactions depending on publishers' permissions)
        - client/server architecture
        - relational database data storage model
        - based on Current Contents/Life Sciences set of approximately 1,350 journals from 330 publishers
  - ***Kodak***
    - » Kodak Picture Exchange (KPX) and Shoebox
      - image identification, network delivery and post delivery security vital areas of research

Information Identification: Identifiers: Under Development

- **Significant International Corporation Activity (cont'd)**
  - ***Lotus***
    - Lotus Notes
      - important front end for most corporate information distribution applications
      - *opportunities to develop deeper granularity in usage tracking system*
  - ***Microsoft***
    - Visa partnership
      - to develop secure transaction technologies
    - Windows 95/ Windows NT Servers
      - much tighter network security control options
    - Microsoft Network
      - *will include secure network transaction facilities*
    - OLE/COM
      - seamless exchange of complex source documents across application and network frontiers
  - ***Netchex***
    - Secure network chequing facility
      - enabling another dimension of electronic commerce
  - ***Netscape Communications Corporation***
    - Netscape Navigator browser has 75% of WWW browsing traffic
    - Agreement with First Data Card Services' Electronic Funds Services (EFS)
      - world's largest credit card payment processor
      - enables banks to provide merchants and users with simple, secure means of Internet electronic commerce
      - real-time online card authorization for transactions with any major credit card
    - Agreement with Mastercard
      - similar rationale to First Data arrangement
    - Netsite secure Commerce Server software
      - designed to enable companies and individuals to easily set up and maintain servers for distributing information and conducting Internet mediated commercial operations
        - Server authentication (thwarting impostors)
        - Privacy using encryption (thwarting eavesdroppers)
        - Data integrity (thwarting vandals)
      - incorporates Netscape implementation of RSA Data Security encryption technology
      - List price of Netsite Commerce Server: $5,000
      - Bank of America, First Interstate and many other US banks already using this system
      - DEC first reseller: available on Digital's Alpha OSF, Alpha NT and Intel NT platforms
      - internetMCI service using it as core technology

Information Identification: Identifiers: Under Development

- **Significant International Corporation Activity (cont'd)**
  - ***Netscape Communications Corporation (cont'd)***
    - » Strategic agreement with Novell, Inc
      - Novell will include Netscape browser in its products from its UNIX Systems Group, Novell Applications Group and Information Access Management Group.
      - First product: WordPerfect Internet Publisher Pro for Windows
    - » OEM Agreement with Silicon Graphics
      - Netscape products to be available on all SGI high-performance workstation and server platforms
    - » Secure Socket Layer (SSL) non-proprietary protocol
      - Internet security proposal
      - freely available now in Netscape Navigator
    - » Committed to open standards process
      - will, for example, support EIT/Terisa's S-HTTP protocol when available
  - ***Object Management Group***
    - » Reusable software object technology to promote integration and interoperability between computer applications
      - enabling identification, linkage and seamless, automatic exchange of objects across networks
      - fast and flexible developments of distributed applications
      - likely to become core client/server computing technology
    - » Common Object Request Broker Architecture (CORBA) standards
      - supported by 450 corporate members, including IBM, Hewlett-Packard, Sun Microsystems and Digital Equipment
      - considered superior to Microsoft's Common Object Model (COM)
        - Microsoft continuing to develop COM as an opposing standard despite its membership of OMG
        - Microsoft is not enamoured of the standards process generally
        - It does not like programmer's having access to its products source code for re-use in other modules
    - » Some reconciliation of the two positions is vital for fully open two way client/server computing between the products operating under one of the two standards
      - functional interoperability between the standards is the likely outcome in the short term
      - longer term an area to be watched carefully, especially for global information distribution systems
  - ***Penobscot Development Corporation***
    - » KALA server
      - radically different data storage approach
        - "monads" rather than files or database tables: monad can be as small as a single bit
        - each monad has its own immutable, unique 62 bit identifier
        - fully controllable access systems
      - their approach may become increasingly important

Information Identification: Identifiers: Under Development

- **Significant International Corporation Activity (cont'd)**
  - ***Springer-Verlag***
    - » Red Sage project
      - in cooperation with AT&T Bell Labs and the University of California, San Francisco
      - this is now maturing as it moves from a proprietary interface to enable broader access to the information base via WWW etc
      - prospect of expansion to other UC campuses
      - also possible extension in cooperation with Pacific Bell and Genentech to delivery via high speed ATM networks
    - » Right Pages
      - AT&T developed client/server software for online information distribution: used in the Red Sage project
  - ***The Centre for Exploitation of Science and Technology (CEST)***
    - » Funded by core consortium of over 30 companies and government departments
      - Electronic Article Surveillance (EAS) source tagging project
        - de-activatable, single component, magnetic tag: small, discrete, robust
        - developing universal , generic source tagging solutions: to replace physical bar-codes
        - universal data identifiers could be used in this domain also
  - ***UMI***
    - » Advanced Document Delivery System (ADDS)
      - online access into UMI information collection
      - viewing articles on-screen
      - articles then printed or faxed to designated locations
      - information collection based on more than 19,000 periodicals, 7,000 newspapers and over 1.2 million dissertations
  - ***WAIS Inc***
    - » Wide Area Information Systems
      - WAISserver 2.0: commercial large scale network publishing server
        - enables literal and natural language database searches
        - tracks usage and document status
        - custom modules for user registration, transaction-based and subscription based billing, personalized invoicing, archived searching for back issues, automatic content expiration and new content alerting
  - ***Xerox***
    - » glyph technology developments
    - » open document standards
    - » digital: analogue paper interface

Information Identification: Identifiers: Under Development

- **Significant International Corporation Activity (cont'd)**
  - ***Some examples of commercial interface software***
    - » Reviewers/browsers
      - Booklink Technologies Windows based reviewer/browser (Beta testing)
      - Netscape Windows based browser (v1.1b)
      - WinWeb browser
      - Amadeus
      - many others in development
  - ***Some examples of data protection and security software and hardware***
    - » Hardware based access control systems
      - Infosafe Systems
    - » Encryption, metering and access control systems
      - See IFFRO report
        - CD-MAX; Wave Systems
      - Ç-Dilla (UK) - Gold Safe for CD-R
      - AT&T/VLSI IVES chip and operating system
      - DES encryption
        - US government standard, now almost twenty years old
        - largely used in private networks
      - PGP (Pretty Good Privacy) encryption
      - RSA Data Security
        - dominant player
        - licensing their system for many open network based commercial transaction solutions
        - only useful for delivery and authentication
        - cannot prevent reuse
        - US prohibits export of strongest form
        - treated as armaments
        - provides considerable competitive advantage in secure delivery system design
    - » Electronic Publishing Resources: Electronic Copyright Management System
      - see IFFRO report
      - commercial system later in 1995
    - » The Research Libraries Group, Inc (RLG)
      - Ariel software
        - document transmission system
        - based on commercial hardware plus Ariel software
        - users can scan articles, photos and similar documents and transmit resulting electronic images over the Internet to each other's Ariel workstations
        - higher resolution and cheaper than fax

Information Identification: Identifiers: Under Development

- **Significant International Corporation Activity (cont'd)**
  - ***Some examples of data protection and security software and hardware (cont'd)***
    - » AT&T: Firewall
    - » Bellcore: BETSI: Bellcore's Trusted Software Integrity System
      - secure software certification service
    - » Softlock
    - » Check Point Software Technologies, Inc
      - Firewall-1 Network Security software
    - » Trusted Information Systems, Inc.
      - supplier to US National Security Agency
      - cryptographic filters; trusted operating systems; network security and Internet firewalls
      - trust engineering; certification and accreditation
    - » DEC Digital Consulting
      - secure systems
    - » Sun Microsystems
      - secure client/server solutions
    - » InfoLogic Software, Inc
      - software "envelope" permanently holding basket of valuable components
      - embedded digital signature and envelope software provide for secure use tracking and authentication
      - interfacing to conventional EDI and other transaction management solutions
  - ***Passive usage monitoring systems***
    - » Cyphertech
      - only suitable for content containing audio track
      - proprietary database holdings
    - » BDS
      - database of audio signal signatures
    - » RCS
    - » Eurodat Digital Audio Technology
      - hardware and software solutions
    - » Denon
      - ISRC Logger (CD player add-on)
  - ***Information Retrieval/ Intelligent Agents***
    - » General Magic: Magic Cap and Telescript
  - ***Information Summarizers***
    - » BT Labs

Information Identification: Identifiers: Under Development

- **US based projects**
  - ***Access control models***
    - » Kerberos/Athena (MIT)
      - widely applied protocols for authentication and secure client workstation access to server in relatively insecure network environment
    - » DYAD: physically secure co-processors (J.D. Tygar - Carnegie Mellon)
      - "protecting integrity of publicly accessible workstations
      - tamper-proof accounting /audit trails
      - copy protection
      - electronic currency without centralized servers"
  - ***ANSI/NISO***
    - » Serial Item and Contribution Identifier (SICI - ANSI/NISO z39.56 - 199x) (see Appendices 4, 4A & 4C)
      - ISSN in or out? Latest decisions pending.
      - Universally applicable?
    - » Other relevant standards activities (being revised):
      - ANSI z39.5-1985 ISO4 Abbreviations of Titles of Publications
      - ANSI z39.14 (R1987) ISO 214 SC AG Writing abstracts
      - ANSI/NISO z39.21-1988 ISO 2108 Book Numbering (ISBN)
      - ANSI z39.53 - 1994 ISO/CD 639/2 Codes for the representation of Languages for Information Interchange
  - ***Association of American Publishers, Inc (AAP)***
    - » Identifying critical new media needs of publishers
      - copyright issues
      - secure transaction networks
      - file format standards
      - business models
      - technology survey
      - reporting April 1995
  - ***Center for Electronic Markets***
    - » George Mason Program on Social and Organizational Learning
      - Ryoichi Mori and Masaji Mawahara (SUPERDISTRIBUTION)and Japan Electronics Industry Development Association are related to this project
      - seeking to establish standards for information distribution and security
      - designing systems based on "usage" of intellectual property rather than "owned copies" model
        - this enables fundamentally new operating system architectures
      - Superdistribution implies intellectual property is stored across the network, not stored on a local hard drive

Information Identification: Identifiers: Under Development

- **US Projects (cont'd)**
  - ***Center for Technology Policy and Industrial Development in MIT***
    - » Digital Open High Resolution Systems (DOHRS) project
      - established 1991, sponsored by Advanced Research Projects Agency (ARPA), DEC and Apple Computer
      - studying technical, economic and policy challenges to open interfaces for high resolution systems and broadband networks
        - especially to facilitate cross border, intergovernment agency and multidisciplinary academic dialogue
      - Mainly focussing on imaging, TV, video, audio, communication and computing
        - transporting applications,
        - sharing generic components
        - network technologies to improve access and IPR protection and authentication
  - ***CommerceNet Consortium***
    - » Non profit corporation
      - founders BBN Barrnet, Enterprise Integration Technologies (EIT) and Stanford University's Center for Information Technology (CIT)
      - matching funded by US government Technology Reinvestment Project (TRP)
      - TRP sponsored by Defense Department's Advanced Research Projects Agency (ARPA), the Department of Commerce National Institute of Standards and Technology (NIST), the National Science Foundation (NSF), the Department of Energy (DOE) and the National Aeronautics and Space Administration (NASA)
    - » Bounded commercial trading environment
      - participants include General Electric, Spry, Amdahl, Bank of America, Citibank, Digital Equipment Corporation, Financial Services Technology Consortium, Network Computing Devices and Tandem Computers
    - » Secure transactions
    - » Client-server architecture
    - » Transaction identification
    - » No content identification protocol
  - ***Copyright Clearance Center (CCC)***
    - » Ongoing survey of electronic copyright management technologies
    - » Focus of US electronic rights management expectations
    - » Strongly involved in IFFRO activities

Information Identification: Identifiers: Under Development

- **US Based Projects (cont'd)**
  - ***Corporation for National Research Initiatives (CNRI)***
    - » Handle generators
      - data object identifiers (Unique authority | Unique Object | Notes)
        - cheap, scaleable
      - value adding metaobjects:: collections of handles
      - a subset of URLs
    - » Library of Congress
      - system for deposit of digital works
    - » Electronic Copyright Management System
      - work underway
      - for very large digital libraries
        - billions of users, billions of documents
        - use handles to build layers of information
        - totally flexible
        - scaled infinitely
    - » Metering: "very limited enthusiasm"
    - » Relationship between digital and paper objects becoming priority
    - » Preservation and archiving of digital works
  - ***Digital Libraries***
    - » Many projects, most already documented elsewhere
    - » Case Western Reserve University Library Collection Services (LCS) & IBM
      - Royalty Manager System
        - explores "fundamental library functions relating to the acquisition, mounting, access, distribution, and use of intellectual properties in electronic form."
        - including applications "to control and monitor the access and use of licensed intellectual property"
        - pre-digitization license agreement between rights holder and digitizer setting out legal language of agreement and list of rules for storage, use and access to the digitized intellectual property, together with use/royalty relationships.
    - » Identification issues similar to physical library world
      - cataloguing
      - archiving
      - indexing
      - bibliographic control
    - » Protocols on project by project basis
    - » No universal identification schema currently proposed
    - » Michael Jensen, University of Nebraska Press
      - ease off use
      - IP stealing to be necessarily intentional
      - header-based security proposal (Journal of the IMA IP Project January 1994)

Information Identification: Identifiers: Under Development

- **US Based Projects (cont'd)**
  - ***Harvard University John F. Kennedy School of Government***
    - » Information Infrastructure international policy issues
      - conference planned for September 1995
  - ***Hypermedia Projects***
    - » Project Xanadu
      - Ted Nelson's project
      - IPR identification: another hypermedia protocol layer?
    - » Internet Engineering Task Force (IETF) - see above
      - No work (yet) on IPR identification issues
    - » Corporation for National Research Initiatives (see above)
      - Expressions not copies
      - Same object contents in different format generates different handle
    - » University of Colorado at Boulder
      - Harvest project
  - ***Information Infrastructure Task Force (IITF)***
    - » Intellectual Property and the National Information Infrastructure - Lehman Report 1994
      - access control (server and file level)
      - use control
      - authentication
      - rights management
      - standards development: "some level of interconnection, interoperability, telecommunications, computer, wireless, satellite, broadcast and cable TV technologies and networks may be essential"
  - ***Interactive Multimedia Association (IMA)***
    - » Identifying and managing intellectual property
      - differing industry requirements
      - subject of proposed Interactivity Convergence Forum planned for October 1995
  - ***Joint ventures: information and entertainment on demand projects***
    - » Proprietary transaction management protocols generally used
    - » Also proprietary or single industry identification systems
    - » Need for cross-industry identification initiative
  - ***National Institute of Standards and Technology (NIST) and National Science Foundation (NSF)***
    - » Several relevant projects, but specific identification issue not currently being addressed except in joint activity

Information Identification: Identifiers: Under Development

- **US Based projects (cont'd)**
  - ***Other Identification Proposals***
    - » Extension of CUPID project architecture to include permissions and payment servers
  - ***Relevant academic proposals***
    - » Netbill (Marvin Sirbu et al, Carnegie Mellon)
      - "Business model, set of protocols and software implementation for supporting commerce in information goods and other network delivered services."
      - low transaction cost for micropayments (1 cent on a 10 cent transaction)
      - certified delivery mechanism: delivery only if customer has paid
      - pre-commercial development and trial partnership with VISA announced February '95
  - ***Society of Motion Picture and Television Engineers***
    - » ASN.1 modifications
      - primarily aimed at creating open standard for transport of audiovisual data
      - enabling high speed broadband interoperability and interconnections
  - ***The Coalition for Networked Information***
    - » Established by:
      - Association of Research Libraries
      - CAUSE
      - EDUCOM
    - » 1990: Task Force formed
      - now involves about 170 institutions and organizations pursuing a shared vision of information management.
      - includes higher education, publishers, network service providers, computer wares and systems companies and library interests
  - ***The Image and Information Standards Initiative***
    - » sponsored by the Getty Art History Information Program
      - identifying issues in imaging that require collective solutions and standard approaches
        - developing guidelines for access to intellectual property for the "ever-growing universe of digitization projects" is a priority
        - protecting intellectual property within open access systems

Information Identification: Identifiers: Under Development

- **US Based projects (cont'd)**
  - ***The Media Laboratory (MIT)***
    - Steganography
      - data hiding in image and audio signals
      - like the Imperial College work below
      - novel approaches being explored
    - dtype standardization
      - simple (eg integers and floating point values)
      - complex (eg strings, lists and packets of raw data bytes)
      - dtype type fixed at creation
      - dtype can only contain values of the creation type
      - may contain only one value at a time
      - once data is in dtype form any dtype routines may be applied to that data (eg writing dtypes to disk or transmitting dtypes across a network)
    - The Media Bank
      - "distributed storage and computing facilities for audiovisual information via a set of ATM linked workstations
      - testbed for personalized information infrastructures
      - current work reports on universal content-location systems and startup protocols
      - Media Browser: automatic page generation is used to create files accessible by Mosaic that depict current media bank content and object location."
      - supports browsing through the full repertoire and access to data elements allowing interactive perusal of distributed visual archives
  - ***US Multimedia Clearinghouse Proposal***
    - Fred Greguras, Michael R .Egger (Fenwick & West, Palo Alto, CA) and Sandy J. Wong (MPAHSM, Portola Valley, CA)
      - basic data elements:
        - name of work
        - licensing conditions
        - royalty fees
        - contact information
      - highlights need for coalition support
      - CCC as possible administrator?
      - no identification protocol contained in proposal

Information Identification: Identifiers: Under Development

- **European projects**
  - ***CITED***
    - » Final Report December 1994
      - tools could be used to perform identification and metering
      - problems coping with "cut and paste" reuse: fixed version of multimedia document does not exist
      - useful as benchmark model for electronic IPR management systems
      - three layers
        - legal
        - human/functional (CITED triple: USER-USAGE-INFORMATION)
        - process
      - "Identification of objects has to be highly flexible and open. It is the purpose of public bodies or copyright management organizations to propose to standardization a common identification numbering system."
  - ***COPICAT***
    - » In early stages
      - nothing yet on granular identification, but plans to devise system of tracking digital material to enable royalty payment. Access security and and negotiation to be covered also.
      - considers target user to have an IBM-compatible PC running Windows connected to the Internet.
        - Protected material must run, and be protected, in such an environment
        - working versions will be based on single layer ownership of rights (with multiple layers in model)
  - ***New Proposals (still to receive funding)***
    - » Imprimatur
      - Authors' Licensing & Collecting Society (ALCS) is marshalling resources and participants
      - granular identification one of the issues to be addressed
      - goal: "To agree, within a forum representing the widest range of information industry organisations, a swift and practical response based on a mutual understanding of the problems arising at the interface between IT, telecommunications and IPRs; and to identify and develop a set of tools to address those problems in the business, technological standards and legal areas."
      - present list of interested parties predominantly European
    - » The European Information Network Services (E.I.N.S.)
      - European Space Agency concept to provide a uniform, unrestricted infrastructure and graphic user interface platform (GUIP) for access to electronic information services
    - » Computer Security Technologies (Sweden)
      - Welcome (Worldwide Electronic Commerce Environment)
        - Secure Business electronic Transactions System based on X.500 + WWW+ Mosaic + TTPs + Smart Cards
        - European equivalent of CommerceNet

Information Identification: Identifiers: Under Development

- **European projects (cont'd)**
  - ***RROs: paper to electronic reproduction: a common problem***
    - » e.g. Copyright Licensing Agency (CLA) in the UK
      - CLARCS (their CLA Rapid Clearance System)
        - currently licenses photocopying from hundreds of thousands of serials plus Books in Print
        - licensing transaction by phone, fax or on-line
      - automating the rights clearance process will become more pressing
        - a common data element identification system will make this far more viable
      - involved with the British Library and others in the COPICAT project
  - ***Strategic Analysis in Science and Technology (SAST)***
    - » EU DG XII Science Research and Development
      - Standards, technical regulation and quality assurance theme
      - Cooperation between EU countries in science and technology
  - ***Useful public domain technologies***
    - » Imperial College of Science Technology & Medicine, London
      - Digital Data Security System
        - International Patent Application Number PCT/GB89/00293
      - Signal Security System
        - International Patent Application Number PCT/GB90/00268
- **Other relevant projects**
  - ***British Film Institute***
    - » Summary of Information in Film and Television (SIFT)
      - database of information on over 500,000 films and TV programmes
      - unique ID on input to system
  - ***M.O.R. Limited***
    - » Digital Image Identification
      - "FBI: Fingerprinted Bitmapped Identification"
      - patent applied for
      -
  - ***The Audio-Visual File Index (AFIndex)***
    - product of Joint Cue Sheet Working Group formed in 1991/2 by ASCAP, BMI, PRS, SESAC, SOCAN and APRA.
    - last edition of AFIndex contained 729,169 titles
    - pilot index set up in Scandinavian Societies (KODA, STIM, TEOSTO and TONO in January 1994)
    - Data items in AF Index are: Society Code, Audio-Visual Number, Title of series or feature film, Episode Title, Production Company, Director, Year of Production, Country of Origin, Total Duration (minutes), Category of AV; Other Information in a total record length of 256 bytes.

Information Identification: Identifiers: Under Development

- **Japanese Projects**
  - ***Advanced Telecommunications Research Institute (Kyoto)***
    - » Work on computer biodiversity
      - "viruses "may be "civilized" and eventually prove to be useful tools for permanent data generation of data and cross-generational data identification
  - ***Institute of Intellectual Property***
    - » Multimedia subcommittee
      - commissioned by MITI to study multimedia intellectual property issues
      - reported February 1994
      - proposes central collective administration (digital information center)
      - voluntary registration
      - no identification protocol proposed
  - ***Kyoto Comparative Law Center***
    - » Copymart project
      - new commercial partners
      - support from National Institute for Research Advancement
      - early stages
  - ***Ministry of International Trade and Industry (MITI) and Ministry of Posts and Telecommunications (MPT)***
    - » Information infrastructure development
  - ***National Center for Science and Information Systems***
    - » Interfaces between computers, terminals and other network components
  - ***Nihon Chosakuken Kyogikai (Japan Copyright Council)***
    - » Subcommittee on Multimedia
      - first reported November 1993
      - proposed Copyright Rights-Information Centralization Organization for rights clearance in pre-existing works when used as multimedia product content
      - latest report due February 1995
        - no specific identification agenda
  - ***Tsukuba University***
    - » Superdistribution chip
      - mainly used for software

- **Identifiers: Under Development: A Summary**
  - ***Need for identifiers widely perceived***
    - active development of identification schemes in many industries
      - based to a greater or lesser extent on existing systems
    - many projects assume unifying level without resolving "how to"
    - pilot projects use bespoke or industry based identification
      - wide use of this approach will lead to "tower of Babel"
    - file/document level identification and management
      - granularity issues not addressed
    - critical issues:
      - flexible granular (data object) identification schemes
      - providing coherent cross-operating system pathways
      - providing coherent cross-application file format pathways
      - providing cross-hardware standard pathways
      - universal top level protocols enabling real time identification of rights in both vertical and horizontal channels and links to rights management (RM) databases
      - simultaneously identifying master and individual expression of a work
        - this implies inbuilt replication control mechanisms
      - real time, trivial cost, locally generated encoding systems
      - privacy
      - codification of licensing agreements required in RM systems
      - neutral or positive influence on and outcomes in world trade and policy

- **Music**
  - ***Interlocking system***
    - » Work (proposed International Standard Work Code - See Appendix 2)
    - » Composer/author (CAE Number)
    - » Active Works (ASCAP World List)
    - » Rights management by territory (CAE List)
      - extension to Rights Ownership Database mooted
    - » Primary expression (International Standard Music Number)
    - » Audio or video recording (International Standard Recording Code - See Appendix 1)
      - facilitates passive monitoring systems
    - » Physical CD carrier and manufacturer (Source Identification Code)
    - » Physical package of one or many recordings (UPC/EAN bar codes)
    - » Serial Copy Management System (in domestic hardware: prevents serial copying)
    - » Proposed Common Copyright Data Model (see Appendix 3)
  - ***Features***
    - » High level cooperation among international representatives of various sectors of industry and ISO
    - » Concerted work towards a common copyright data model to rationalise links to rights management systems
    - » Pursuing overall lowering of industry's transaction costs
  - ***Ready for open internetworking?***
    - » Even this sophisticated system is not capable of surviving the rigours of open internetworking.
    - » ISRC can be stripped from the audio signal as it is carried non-invasively
      - i.e. it is not in the main signal, but resides in a non-audible sub-signal
      - stripping sub-signal (including the ISRC identifier) has no effect on audio quality
  - ***Cross-industry standards agreement required regarding:***
    - » high level, non-industry specific commonly structured object identifier
      - to provide common automatic link to application, operating system, network and back office rights management systems
    - » technical means of "inserting" and "reading" such identifiers
    - » the various wares required to organise transaction systems based on the identification protocols

Information Identification:
Principles for Data Object Identification

- **Digital Object Identification Scheme (DOIS)**

  (See Appendix 4 for an overview of document identification systems)

  - ***To facilitate efficient vertical and horizontal distribution***
  - ***Standardise identifier structure across all object sources***
    - » existing industry based standards are too industry specific to provide universal solution
    - » DOIS will need to be devised from scratch, drawing on all possible existing identification knowledge
      - *scheme must be general*
    - » universal codes can be used as unique identifiers across all media, referring back, where and for as long as is necessary, to industry based codes such as SSDI, SICI, ISSN (see Appendices 4 & 4A) and so on.
  - ***Integrating industry based schemes*** (see Appendices 4A to 4F for an outline of some specifics of the existing publishing industry systems)
    - » common universal data identification coding protocol requirements (with thanks to Dr Norman Paskin)
      - general, neutral structure
        - common start/stop bits
        - common structure for unique authority/data generator section
        - information on data generator's identifier sets held in unique authority's database
      - format independent
        - application or operating system information held in database
      - object type independent
        - data type held in database
        - see MIT Media Laboratory's dtype
      - unique to an object
        - cannot be accidentally created identically from two sources
        - theoretically no duplication should occur
        - an identifier never used twice
      - unique to the expression of an object
        - each new expression requires a new identifier
        - none of the existing industry codes obeys this principle
        - may involve allowing for the dynamic generation of a date/time stamp, machine identity code or other unique event identifier related to the expression
      - easy to generate and use
        - alphanumeric only
        - algorithmically explicable
      - not restrictive
        - able to accommodate many object types
      - serves only one purpose
        - no compulsory explicit meaning
      - compatible with existing archives of objects
      - generated by originator of published digital object
        - identifier set issued to each data object generator
        - then no further reliance on external validation
        - can be automatically allocated as part of production process

- **Music**
  - ***Interlocking system***
    - » Work (proposed International Standard Work Code - See Appendix 2)
    - » Composer/author (CAE Number)
    - » Active Works (ASCAP World List)
    - » Rights management by territory (CAE List)
      - extension to Rights Ownership Database mooted
    - » Primary expression (International Standard Music Number)
    - » Audio or video recording (International Standard Recording Code - See Appendix 1)
      - facilitates passive monitoring systems
    - » Physical CD carrier and manufacturer (Source Identification Code)
    - » Physical package of one or many recordings (UPC/EAN bar codes)
    - » Serial Copy Management System (in domestic hardware: prevents serial copying)
    - » Proposed Common Copyright Data Model (see Appendix 3)
  - ***Features***
    - » High level cooperation among international representatives of various sectors of industry and ISO
    - » Concerted work towards a common copyright data model to rationalise links to rights management systems
    - » Pursuing overall lowering of industry's transaction costs
  - ***Ready for open internetworking?***
    - » Even this sophisticated system is not capable of surviving the rigours of open internetworking.
    - » ISRC can be stripped from the audio signal as it is carried non-invasively
      - i.e. it is not in the main signal, but resides in a non-audible sub-signal
      - stripping sub-signal (including the ISRC identifier) has no effect on audio quality
  - ***Cross-industry standards agreement required regarding:***
    - » high level, non-industry specific commonly structured object identifier
      - to provide common automatic link to application, operating system, network and back office rights management systems
    - » technical means of "inserting" and "reading" such identifiers
    - » the various wares required to organise transaction systems based on the identification protocols

Information Identification: Principles for Data Object Identification

- **Digital Object Identification Scheme (DOIS)**

  (See Appendix 4 for an overview of document identification systems)

  - ***To facilitate efficient vertical and horizontal distribution***
  - ***Standardise identifier structure across all object sources***
    - » existing industry based standards are too industry specific to provide universal solution
    - » DOIS will need to be devised from scratch, drawing on all possible existing identification knowledge
      - *scheme must be general*
    - » universal codes can be used as unique identifiers across all media, referring back, where and for as long as is necessary, to industry based codes such as SSDI, SICI, ISSN (see Appendices 4 & 4A) and so on.
  - ***Integrating industry based schemes*** (see Appendices 4A to 4F for an outline of some specifics of the existing publishing industry systems)
    - » common universal data identification coding protocol requirements (with thanks to Dr Norman Paskin)
      - general, neutral structure
        - common start/stop bits
        - common structure for unique authority/data generator section
        - information on data generator's identifier sets held in unique authority's database
      - format independent
        - application or operating system information held in database
      - object type independent
        - data type held in database
        - see MIT Media Laboratory's dtype
      - unique to an object
        - cannot be accidentally created identically from two sources
        - theoretically no duplication should occur
        - an identifier never used twice
      - unique to the expression of an object
        - each new expression requires a new identifier
        - none of the existing industry codes obeys this principle
        - may involve allowing for the dynamic generation of a date/time stamp, machine identity code or other unique event identifier related to the expression
      - easy to generate and use
        - alphanumeric only
        - algorithmically explicable
      - not restrictive
        - able to accommodate many object types
      - serves only one purpose
        - no compulsory explicit meaning
      - compatible with existing archives of objects
      - generated by originator of published digital object
        - identifier set issued to each data object generator
        - then no further reliance on external validation
        - can be automatically allocated as part of production process

Information Identification: Common Content Classifiers

- **Serendipity facilitators**
  - ***Pre-competitive or competitive?***
    - » a value adding opportunity built around needs of individuals in identified markets
      - attractive to perform natural language searches rather than complex boolean queries, providing the outcome is similarly selective
    - » industry wide agreement on broad subject classification categories may be useful
      - incorporated into "back office" database design, not part of primary object identifier
      - providing a high level search field for intelligent agents
  - ***House policy and market relationships more likely to determine approach to classification***
    - » individual tailoring of information will become an important market feature
      - in this context the common content classifiers may not meet these precise needs well enough
  - ***A secondary level tool***
    - » primary linkage via universal identifiers
    - » see CNRI metaobjects concept

Information Identification: What Needs to be Identified?

- **People. Works. Agreements. Licensing Schemes.**
  - The effort in the music industry is directed to identifying the people, works, agreements and common industry licensing schemes
  - If each industry involved in a universal data identification project were to build similar databases, the basic "back office" data would then be in place.
  - The STM Task Force could provide the ideal vehicle to focus the standardization of this model in the stm world.
  - It would be enormously helpful if the structure of this supporting data could be stored in a common copyright data model (such as is being proposed by CISAC - see Appendix 3.) Not all fields may relate to each industry, but different data entry views of the underlying structure would overcome these differences.
  - These needs must be determined precisely.
  - Effort should be directed to gathering and storing commonly structured data about:
- **Interested parties (natural and legal) (IPs)**
  - ***Every person (natural or legal) who has an interest in any of the items below will need a code.***
    - » Details to enable positive, rapid identification.
    - » This would be supported by information regarding their interests and the management of these in particular territories.
    - » Existing industry files (e.g. the CAE file) may provide a starting point.
- **Works**
  - ***From which Unique Data Objects (UDOs) are derived***
  - ***Works exist irrespective of whether they are published, performed or recorded.***
    - » they enjoy rights.
    - » it is important to be able to differentiate works accurately and quickly.
    - » Each of these works needs a code which is separate from the code it carries when published, performed or recorded.
    - » This code does not currently exist in most industries.
    - » The ISWC is the first major push in this direction.
    - » Using such a system it becomes much easier to link the interested parties to the work and the work to various uses/expressions of it.
- **Unique Data Objects**
  - ***Valuable part or whole of***
    - » any expression of work (however defined)
  - ***Every data object will need a code.***
    - » this is the proposed universal data identification code
    - » this presently does not exist

- ***A data object, like a work, may exist in its own right without having to be recorded, published or performed.***
  - » Supporting data is required to make this code relational
- ***Unique object ID for each expression***
  - » no two objects the same
  - » different object when different
    - format
    - medium
    - space
    - time
    - interested parties
    - rights

- **Codified agreements between interested parties in relation to works and UDOs**
- **Standard industry licensing agreements and schemes**

Other elements requiring ongoing identification:

- **Publications**
  - ***Each publication requires its own unique code***
    - » for whole works this is generally well handled by existing systems
    - » constituent sections of publications are still not coped with coherently - hence the ongoing debate about the z39_56 SICI code and so on.
    - » this still requires further work. Minutes of latest deliberations expected shortly.

- **Recordings**
  - ***Each recording requires its own unique code***
    - » the ISRC is extremely useful here to define the whole recording
    - » although the ISRC code can be contained in small samples of audio material, the fact that it is strippable from the audio signal makes it facultative rather than definitive as a method of identification
    - » uniform data object identifiers embedded in the audio signal would be a more thorough and far more flexible solution to handling the sampling problem

- **Films & Television**
  - ***Each film or television work needs its own unique code***
    - » there are many different systems in use in the film and television industries and they may be the last group to perceive the need to collaborate in this effort
    - » the main reason for this is that only when sufficiently bandwidth and/or cheap CD-R storage is available will these industries be seriously threatened to the same extent as less bandwidth/storage intensive industries
    - » some major efforts are underway at the British Film Institute (the SIFT project) and elsewhere

Information Identification: What Needs to be Identified?

- **Performances**
  - ***Each performance (however this is defined) needs its own unique code***
    - » although appropriate societies are working towards the goal, a great deal of work still needs to be done to devise a suitable uniform system

- **Formats**
  - ***Each format may require its own unique code***
    - » this needs to be explored further
    - » it follows on from the notion expressed by John Garrett at the CNRI in the US that there is no such thing as a "copy" in the digital world, only another expression or manifestation
    - » such manifestations have a different form depending on the operating system, video card, application software and so on

- **Rights Ownership**
  - ***This ties together the interested parties with the data objects and the works and sets out the rights arrangements relating to the various uses of them***
    - » the ownership of the rights in each data object needs to be identified and codified
    - » these rights may attach to the work as a whole (when the work is a whole data object) or be directly related to a data object (which itself may be only a designated part of a work)
    - » a common, codified way of storing the data about rights holders and agreements relating to their rights is required
    - » if this can be made reliable enough it may form the basis of a "rights titles" agency
    - » as economic value is increasingly stored in intellectual property rather than real assets it will become necessary to "securitize" these assets in the same way as real assets. It will become desirable to raise financing by mortgaging them, to put caveats and liens on them and so on
    - » for such a system to operate it must be possible to guarantee title to a defined asset
    - » a definitive international "title by registration" system would facilitate the development of such instruments and markets.

Information Identification: What Needs to be Identified?

Issues:

- **Securitizing IP assets**
  - throughout the value chain
  - cost effectively

- **An international agency to facilitate a highly distributed system**
  - on neutral ground politically and industrially
  - may be an appropriate next step in WIPO programme
  - focussing the policy agenda in a disciplined manner
  - working across all hemispheres
  - encouraging pre-competitive cross-industry agreements
  - negotiating the grey areas
  - early agreements to implement final collective decisions
  - open lines of communication and thorough information dissemination
    - » especially close working relationship with wares providers

- **Industry based action essential**
  - focussing the identification effort
  - unifying industry approach by internal pre-competitive agreement (see Herman Frank: *Pre-competitive stm standards*, August 1994)
  - interweaving industry needs into global project
  - referring to other successful identification models

- **Company based action**
  - harmonizing company identification programme with industry and global standards
    - » enhancing shareholder value

- **Time**
  - leakage will be minimized by decisive harmonious action

- **Initiatives**
  - each industry is still addressing its own identification system. In most sectors there is some high level activity. It is beginning to dawn on information professionals that a unified top level system may have many attractions
  - during this study no other all-industry initiative has been discovered, or even hinted at. Should stm and its publishing industry associates decide to take the work further, a useful next step may be to conduct a formal survey to determine all-industry support for a unified top level approach

Changing relationships

Will uniform identification change control? Instant universal support should not be anticipated. Some early detractors can be anticipated.

There are suggestions from some quarters that the identification debate is being driven by collective rights management organisations for their own advancement.

The Motion Picture Association of America (MPAA), for example, sees this as a highly political issue. Its Senior Vice-President and General Counsel, Mr. William Billick, expressed the Association's view that they would only be interested in looking at uniform identifiers once the European threat of compulsory licenses for use of the works of their members was off the agenda.

The MPAA is concerned not to find itself having to share part of its rightful revenue stream with a collective copyright organisation not of its own choosing. He could see, however, if such a threat were removed, the Association may well be prepared to look at the issue, providing it was in the context of providing its members with the opportunity to realize otherwise unobtainable trading arrangements and anti-piracy protection. Mr. Billick voiced his hope the outcomes of negotiations at the G7 ministers meeting in Halifax, Canada in June, 1995 would make the position clear.

These issues need to be addressed squarely.

There is a natural fear of the loss of control, no matter what the situation.

Any changing relationships must be seen to be to the advantage of those presently in control or they will tend not to accede to the course of action being proposed.

Uniform identification, in and of itself, will not change the existing control structures. Every effort must be made to make the system as neutral as possible. Indeed it will only be useful if it is content neutral. This will only be achieved if the current framework is reflected as closely as possible in the electronic trading model that is being designed to complement it.

Over time there may well be changes to the economic relationships. These should not, however, be ascribable solely to the introduction of uniform identification. The goal should be to empower the existing rights owners, not to earn a slice of their revenues by a technopolitical sleight of hand.

The major players will only play if they can see a way of adding shareholder value.

Information Identification: Threats, Opportunities & Action

- **Threats**
  - ***A combination of:***
    - Ubiquitous digitizing tools
    - Trivial storage cost
    - Cheap broadband communications
    - Processing power: supercomputers on the desktop
    - Many to many networked interconnectivity
    - Attitudes to intellectual property
- **Opportunities**
  - ***Developing online, open systems, based around a common identification system, which facilitate:***
    - Monitoring
    - Control
    - Compensation
  - ***Metadata***
    - Adding value to existing resources by providing definitive collections of information identifiers that provide users with their desired level of interaction with the global knowledge base on a particular area of interest
      - open, universal data identification standards will facilitate the emergence of this market
        - the performance of intelligent agents and other information retrieval software tools will be considerably enhanced
  - ***For long life information, secure, controlled access, controlled facility CD-ROM carriers offer the most viable current solution (see the IFFRO report, October 1994)***
    - Tolerate proprietary identification solutions
      - this should not obscure the rapidly emerging need for identification capable of operating successfully in open systems
- **Action**
  - ***Content, software and silicon cooperation***
    - interoperability and interconnectivity
    - open information exchange capability the goal
    - will require involvement of all major wares companies
      - may involve modifying operating system kernels
        - if so, lead time for next generation of systems is three to five years
      - must take account of emerging hardware standards also
    - at least three years to devise and agree a comprehensive set of identification protocols
      - Thereafter there will be a lengthy period for rights owners and their agents to encode their content and build back office systems that smoothly interface with the new technology

- **Real World Issues**
  - ***An open system must reconcile issues surrounding:***
    - keyboards and other input devices
    - operating systems
    - applications
    - file formats
    - object interlinking and re-use
    - compression
      - especially the lossy varieties
    - multilingual scripts
    - new tools e.g. summarisers
    - culture
    - privacy
    - analogue:digital interface
    - identification code awareness and use throughout value chain
  - ***Many similarities with telecommunications standardization***
    - global vision
    - local implementation
    - simplifying access
    - fierce competition in the committee room and in the market

- **1995: Secure CD-ROM delivery and metering systems**
  - In this channel the contract will be based on either subscription and/or volume. In the encryption/decryption for payment model, it is quite feasible to use non-standard identifiers as the decrypted data should be subject to metered use only on that single system or boundary controlled network. Universal Data Identifiers (UDIDs) are not mandatory in order to effect protection, unless the variant of the system used reports on the data to a third party rights clearing house rather than directly to the rights owner.
  - Any time there is an external third party involved the UDID code should preferably be used. The name of this game is automation. Anything that may make that process unnecessarily less efficient is to be avoided, especially as bespoke systems imply higher overall costs.
  - Also, if the same data may be reused either in the same combination or in new combinations in other marketing channels either concurrently or in the future, then it seems prudent to use a UDID as house policy.
  - The secure CD-ROM systems are a comparatively narrow market and may be rapidly eclipsed by the next generation of copyright management systems, exemplified by the Electronic Publishing Resources system (see IFFRO report October 1994). This system relies heavily on third party reporting. As a result, a UDID should be routinely used.
  - As these UDIDs do not currently exist those early adopters of these systems may have to weigh the advantages of early adoption of the system against the cost of later recoding the source material once a UDID code is available and a standard encoding technique has been agreed.

- **1996: Licensed, controlled access digital libraries**
  - The next distribution model of significance will probably be the licensed controlled access digital library model.
  - Most of the major IT companies have a stake in this market.
  - These projects are still at the pilot stage, but over the next two years will become more mature and begin to play a significant role in the professional and academic information market.
  - Here is it likely there will be a combination of subscription, volume and content based contracts.
  - All would eventually be best serviced using a UDID code rather than any proprietary or publishing industry code.
  - In the short term, however, the industry based ISBN/ISSN and SICI codes or the SSDI (Standard Serials Document Identifier) used by Elsevier Science with a SSDI adaptation for books may be serviceable interim solutions (See Norman Paskin's review of these systems at Appendix 4). They are unlikely to fulfil all requirements in the longer term.

- When one confronts the implications of Ken Dowlin's startling statistic that only 16% of the holdings in the San Francisco City Library are books and one reads of their plans for the New Main Library (opening early 1996) to be totally wired with over 600 multimedia, multi-lingual workstations, the shrinking relevance of industry based identifiers begins to make sense. Speaking at the 1993 IFLA Satellite meeting in Madrid, Dowlin said: "These workstations will provide access to the on-line catalog in all languages contained in the library (44 in 1993). It will include the general indexes to most of the journals in the collection, will connect directly into the Internet, and will provide electronic document delivery from the Library. A goal is to have electronic connection to every home, school, and office in the City of San Francisco by the year 2000, and for the connection to provide for multimedia and multilingual capability. In addition, the Library intends to provide the gateway for all these microcomputers to be connected to the Internet."
- In this monomedium, a UDID system must eventually be the sensible approach, albeit these UDID codes may eventually lead back into databases that use these industry identifiers at some other level in the transaction management system. In some instances this may provide a bridge to a more unified future, in other cases it may remain the norm for a long time to come.

■ **1997: Comprehensive, robust and flexible electronic copyright management systems**

- We are yet to see the exact shape of these systems, as they are still largely either under non-disclosure agreements or on the drawing board. There are hints of many proposals, including the Imprimatur project presently being assembled by the Authors Licensing & Collecting Society in London.
- It seems feasible that any model identification system developed under the scheme proposed in this report may be suitable for testing in systems like the planned Imprimatur engine once its shape develops. Indeed, all such projects should be given the chance to trial UDID protocols before they finally go "live".
- There will be plenty of chance for this, as any system developed over the next two years will take even longer to get to market. By the time they are ready to launch on a commercial footing the UDID protocols should be nearing their official launch also.
- Ideally these two concepts should be launched simultaneously so that coding of works to be managed using these systems is done once and done right.

■ **1998: Licensed, controlled access commercial internetworks**

- Looking out even three years on the new planet is truly star gazing.
- This looks, however, to be one of the stronger contenders, given the furious activity at the Internet Engineering Task Force, Microsoft and elsewhere.
- Licensing may well become an issue during this period as rights holders look for more solid legal comfort in this new domain.
- This suggests a CommerceNet or NetSite/CommerceServer (see reference to these models in earlier sections) type of approach, where commercial activity (as opposed to free exchange) is isolated in a special marketplace (as in the medieval village).
- UDID coding beginning.

- **2000: Licensed, controlled general access internetworks**
  - This implies there are solutions available which enable commercial traffic to mix freely with non-commercial with adequate, affordable, internationally enforceable safeguards for the buyers and sellers.
  - This also assumes some in-built mechanisms for automatically regulating the international re-transmission of data.
  - Licensed trading and controlled access may still be required in order to control the interactions.
  - UDID coding continuing. Many UDID aware devices and systems coming into the marketplace.
  - Strong vertical relationships between IPR owners and users still being maintained.

- **2002?: Licensed, open access networks**
  - The greatest disciplines are the prerequisites of the greatest freedoms.
  - A comprehensive global, entirely robust end-to-end architecture gradually being deployed, readying global networks to meet all the previously expressed legitimate concerns of IPR owners and others about asset security.
  - Even then it is possible traders will need to be licensed, especially if they are dealing with rights databases.
  - As more material is released and back catalogue material is UDID coded, trade in these assets becomes simpler. IPR owners still need vertical control in order to protect their interests.
  - Serious work begins on creating an international rights titles registry.

- **Beyond?**
  - Bandwidth increasing - high resolution image and video assets become more mobile as a result.
  - Around 2012 a sufficiently robust network in place, with sufficient UDID coded material available to create a genuinely open marketplace in IPR based assets.
  - IPR asset owners begin to encourage re-use of their content, certain in the knowledge the marketplace has the in-built mechanisms to guarantee they will be paid.
  - An international rights title registry becomes fully operational.
  - Low cost securitization of rights based assets becomes possible.

■ **The future of information commerce.**

It is worth considering the identification implications of the five scenarios proposed by Northeast Consulting Resources, Inc in the January 1995 issue of Mapping the Future of Information Commerce Newsletter (p3 of which is reproduced at Appendix 5).

A. Superdistribution

» The redistribution aspect of this model (i.e. beyond the initial parties to the transaction) points to the desirability of a universal identification system.

B. InfoCommerce Fast Forward

» Incompatible computing and telecommunications technologies make rights management extremely difficult. A Tower of Babel. Universal identification protocols may still be useful in providing information linkage to rights databases from the various virtual worlds created by multiple incompatible superhighway systems, albeit the "format" database may need to reflect the particular superhighway model used for a particular manifestation or performance.

C. Global Network Utopia

» This scenario of information utilities using very fast integrated networks maps well to universal systems of all kinds, identification included. Such a scenario would make devising means of identifying the rights in the various elements of the integrated data stream crucial. Determining and identifying the rights resulting from interactive communication becomes an issue also.

D. An Interactive Archipelago

» The data from some CD-ROMs will need to be re-usable under some arrangement between the parties. If this is the case then any re-use which is for release to third parties will need to be identified. Again, a universal approach makes this considerably easier to manage.

E. Knowledge Refining

» This leads on from the kinds of problems suggested by the Text Summariser mentioned above. If these systems are driven by artificial intelligence engines there will be major intellectual property issues to resolve. Once the decisions are clear on this front, however, it should be possible to use the universal codes to identify the input to and the output from these engines. Interface customization presents the same kinds of issues and similar solutions.

All these are likely to manifest to some degree. Universal data identification standards would add value in all cases.

- **Reaching cross-industry pre-competitive agreements.**
  - One important function of peak international industry associations is pre-competitive agreements.
  - Whatever work is done in this field, of course, requires the greatest of care to ensure at no time do the parties take any action which could be considered anti-competitive.
  - To this end, the widest possible support and involvement should be constantly sought and, as noted, the flow of information should continue to non-supporters also.
  - International industry associations have grown in stature over the past decade, as more industries and individual organisations have become truly global.

- **Getting the databases to relate.**
  - Rational cross-linking of databases is an essential foundation to open network trade in intellectual property assets.
  - Achieving this should become one of the key objectives of all those international associations whose mission calls for fostering and coordinating efforts towards a more integrated information economy.

- **Getting the identifiers right.**
  - There are no second prizes. Once committed, the implications are far from trivial. The chosen system must be suitable for open internetworking, but ready as soon as possible. Tomorrow's solution today.
  - Once a standard electronic object identifier (EOI) has been agreed the cost of each insertion may be trivial, but the cost of reinventing the system would be enormous.
  - Future proofing is always a tricky game. The future has a habit of arriving before we have finished with the present! That said, every effort must be made to predict the future identification needs while constantly bearing in mind the requirement that their deployment must be sustainable now, both economically and technically.
  - Several future scenarios (such as those above) should be carefully considered and, if possible, the chosen UDID system should cope with them all.

- **Are the identifiers "numbers" or is "digital DNA" possible and feasible?**
  - Initially there should not be any presumptions made about the exact nature of the universal identifiers. It is good to remember the phenomenal identification power of DNA which uses just four "letters" to refer to four different bases. When combined in special ways in long enough chains bundled into every cell, these have the power to code for life! The parallels are worth exploring closely.
  - Computer virus research, such as the biological diversity models being explored at the Advanced Telecommunications Research Institute in Japan, may provide the answers.

- **Choosing technically competent organisations.**
  - By using the networks of the peak industry bodies it should be possible to identify all the likely contenders capable of offering solutions.
  - In order to bring a universal data identification system to fruition it will be necessary to involve at some time or another all the major players in computing and communications.
  - Amongst them there is already great interest in trying to solve these problems and coordinated input from the rights industries will only assist the process of reaching mutually beneficial conclusions.
  - It is not too early to be opening formal discussions with them about ways in which they may be both willing and able to assist.
  - Over the course of the first phases of the project those organizations that are both willing and able to assist will become clear.
  - It should be stressed from the outset that for them this, too, should become a pre-competitive issue. The foundations of the information economy will only be strengthened by this initiative and there needs to be broad agreement on this.

Information Identification: Outcomes

- **Open standards.**
  - The protocols that are finally agreed should become open standards, controlled by a nominated agency, but essentially in the public domain.
  - All potential players should be made aware of this from the beginning of the project.

- **Real time**
  - To be useful in open nets, the universal data object identifier must link to databases which facilitate real time rights clearance and transaction management.

- **Transaction mechanisms**
  - These transaction mechanisms also need to operate extremely rapidly. Their design will need to remove all possible data communications bottle-necks.

- **Lowest possible cost**
  - The overall goal is to provide the lowest possible transaction cost for delivery of the desired right to a particular piece of core intellectual property value.

- **Systems ready**
  - The goal is an end-to-end and wall-to-wall system that is aware of the embedded codes. If the solutions are built on an open standard, the unit cost of doing this should be very low and it would be relatively simple (if enabled in the initial design) to declare a cut off date beyond which only UDID aware equipment and software would be able to use UDID encoded data. Upgrades should be made available at nominal cost.

- **Applications ready**
  - As mentioned earlier, the development of software applications often requires many elapsed years. If a decision were taken immediately, depending on the complexity of the application, it could take up to five years to filter through into release versions.

- **Networks ready**
  - Networks will play a vital part in maximizing transaction speed, routing efficiencies, use tracking and boundary controls.
  - The designers of all networks from LANs to GANs need to be involved as closely as possible in the process to enable them to tackle the upgrading of their systems to make them UDID aware.

- **Minimise processing**
  - As the goal is high speed data communications, every effort must be made to minimise the processing required at each stage in the rights process.

Information Identification: Outcomes

- **Maximise user friendliness**
  - Keeping the customer satisfied. Convenience. Ease of use. Perceived value. Robust user interface. Every effort should be made to make the data identification a background issue for those going about their normal business.
  - Control of the primary coding of the data objects should be as largely as possible as close as possible to the source.
  - Simple, GUI interface software based on common operating system platforms may need to be devised to facilitate this process.

- **Education**
  - This may sound unimportant, but there are those who are conducting what they call the Subversive Agenda, such as Steven Harnard at Southampton University, Alan Ginsparg at Los Alamos and Andrew Odlyzko who propose that information wants to be free. Some members of the Board of the Electronic Frontier Foundation and others with a shareware or public domain software background take a similar line.
  - Those whose asset portfolio is based in IPRs may need in time to promote their own case more vigorously.
  - From an identification perspective, the Information for Free proponents may choose to participate or not in identifying their free journals and the data objects within them.
  - Even if they do identify them (perhaps for authentication and archival purposes) there is nothing in-built in the UDID type codes which insists that there be payment for use! This will only occur if there is a legitimate registered interest on a rights database that is triggered by the code.

- **Privacy**
  - Any widespread use of identifiers to track the individual uses of information will need to be preceded by careful work in the privacy area.
  - It is essential the transaction is a private affair between the buyer/s and seller/s of the rights to the object and thereafter the same privacy provisions apply as in any other data based transaction. Only if the buyer and seller both actively express the wish for the data about the transaction to be stored beyond the period needed to complete it should any residual data be held in a third party database.

- **Legal**
  - This proposal may eventually impact on the international legal system. Early advice should be sought on this issue to allow any parallel legal work to continue while the technical work is underway.

- **Who will pay?**
  - Clearly there are many players who will benefit if this project can be bought to a successful conclusion. Those who contribute to it both practically and financially will have the greatest chance to shape the outcomes.

- It is for STM and its publishing associates to determine whether this is a project worthy of support. They should, however, only be called upon to provide their share of the funding until other industries can be brought into the project.
- From the start the international bodies representing these industries must be invited to participate.
- A clear statement of the goals of the project need to be drafted and approved by the widest possible groupings.

- **Economics**
  - The points made by the STM Task Force members at the meeting in London have been born out by other opinion freely expressed by those interviewed for this study: the economic benefits are clear if success is achieved.
  - It is a matter then of testing this opinion further by seeing the level of investment the various affected industries are prepared to make to see the project succeed.

- **Some questions:**
  - ***Independent long term administration***
    - » WIPO?
    - » Another existing international agency?
    - » A purpose-specific all-industry council?
  - ***International legal issues***
    - » Who will address these?
    - » What will be the legal status of the body?
    - » Defunct database managers
    - » Who would deal with transfer to other database managers?
  - ***Licensing and monitoring of networks***
    - » Who would be responsible for this?
    - » What might be the terms of such licenses?
  - ***Licensing of transaction organisations***
    - » Do these organisations need to be treated like banks?
      - Fiduciary responsibilities?
      - Capital adequacy provisions?
      - Insurance?

Information Identification: Conclusions and An Action Agenda

Conclusions

1. STM publishers and others should further strengthen identification programmes within the industry.
2. Encourage WIPO to continue with its project on voluntary numbering, with revised objectives.
3. Encourage the IPA, FEP, AAP and other appropriate associations within the publishing sector to join STM in providing funds to establish a cross-industry task force.
4. Encourage the practical and financial participation of all other affected industries in the membership and ongoing work of this task force.
5. The brief of the proposed task force is to establish cross-industry consensus on a comprehensive set of universally applicable identifier protocols by the end of 1997. To achieve this result it will need to begin its work immediately.
6. Subsequently it is to coordinate the programme needed for this set of protocols to be adopted as international (ISO) standards.
7. Upon the achievement of ISO standards, this task force may
   a) cease functioning
   b) seek to become incorporated in its own right and be appointed as the International Registration Authority for these standards.. It may also then oversee the appointment of as many subsidiary registration authorities as necessary or
   c) continue with an ongoing independent task force programme in *information identification issues.*

   note: b) and c) may be undertaken simultaneously.

An Action Agenda

- ***Using existing standardization procedures as a model, on a pre-competitive basis:***
  - » Establish small, global, broad based, high level, ad hoc task force of content controllers
    - To set functional specification
      - establish comprehensive set of questions requiring resolution
      - delegate working groups to analyse and report on each
      - drawing strongly on legal and technical advice
      - integrate with existing industry-based identification systems
    - To establish international administrative structure for long term promotion and management of the standards
    - global standards, local implementation
    - To represent a combined voice on international policy issues in this domain
  - » Establish small technical task force of content controllers, computer and communications software and silicon companies
    - To produce a comprehensive universal open standard data identification protocol and requisite agreements on open interchange standards for hardware and software
      - coordinating role
      - delegate working groups
    - To coordinate efforts to produce a fully deployable standard system
    - To advise on standards, including upgrades, once established

Information Identification: Conclusions and An Action Agenda

- **An Action Agenda (cont'd)**
  - ***All standardization activity to be multilateral***
    - » Global approach, globally funded
    - » Open information at all stages
    - » Inclusive, not exclusive
    - » Facilitating all possible business models
    - » Agree core international values to be reflected in fixed elements of the system
    - » Only it necessary set rules by treaty and national legislation
    - » Local variants accommodated without jeopardizing whole
    - » Recognition of interdependence of interests
    - » Functional administration at lowest possible levels in the system
    - » Peak level to ensure ongoing cohesion and equity of application
  - ***Small core administration***
    - » Task forces and working groups all drawn from and funded by existing organisations
    - » Professional support team to facilitate and coordinate action agenda
  - ***Develop rapidly***
  - ***Deploy incrementally***
  - ***Limit Leakage***
  - ***Mixture of Top Down and Bottom Up solutions***
    - » Many fronts need to be covered simultaneously
  - ***Keep emerging horizontal distribution systems constantly in mind***

Information Identification

The Opportunities

- The economic value and strategic importance of intellectual property (IP) is growing rapidly
- IP is being traded more frequently in digital, networked environments
- A substantial global market exists for valuable IP delivered via networked digital devices
- Rapid, low cost, interactive access to these assets would be a boon to users
- Exploitation of this market provides growth opportunities for IP owners as well as for suppliers of networks and digital devices
- An open market in IP assets would potentially add value to the portfolios of all participants
- Solutions enabling cost effective dynamic licensing offer the most promise

The Threats

- Networked digital devices facilitate simple, quick, cheap reproduction of valuable IP assets
- Existing solutions provide limited protection for IP owners
- Existing standards do not support licensing of data objects smaller than a complete work
- Proprietary identification, security and trading of IP assets is expensive
- Proprietary solutions do not allow open network trade in IP assets
- No effective means exist to identify unlicensed uses of IP in open networked environments
- IP owners have good reason to be concerned about losing control of their assets in this domain
- They are naturally reluctant to license use of their IP without adequate protection

Action

- Standardized technological solutions for identifying, securing and trading IP assets are required
- Dynamic functional license management specifications designed around common framework
- The architectural framework must provide an adequate environment for IP use management
- Standards to enable automated dynamic licensing of uses of data objects from mixed sources
- They need to take account of the multilingual, multiformat, global nature of the marketplace
- Identification tools must enable rapid, cost effective links to rights management systems
- Rights management organizations may need to construct a common IP rights data model
- The existing legal framework may also need codification and enhancement
- The goal of the standardization process should be to maximise value and distribution efficiency
- The standards should seek to enable open trade in IP assets

Who?

- These standards should be agreed on a pre-competitive basis
- Content, network and digital device suppliers should participate in the standards process
- Representative international associations should be involved, where possible and appropriate
- Open lines of communication should be maintained with other market and public stakeholders
- International standardization agencies should be involved

Information Identification

Identifiers: Under Development

- Need for identifiers is widely perceived
  - active development of identification schemes is underway in many industries
  - many projects assume unifying scheme will emerge
  - pilot projects generally use bespoke or existing industry based identification systems
    - » widespread commercial use of this approach will lead to "Tower of Babel" in open markets
  - crucial data object granularity issues not addressed
    - » pilot projects are using file/document level identification and management only
- Information Distribution Technologies
  - Universal Data Identification (UDID) codes should ideally be used in all information distribution transactions involving reporting to or dealing with external third parties
  - Information distribution technologies will only develop from highly bounded proprietary systems to relatively boundless standard open systems when automatic rights control can be exercised
    - » Suitably designed UDID codes will provide the hooks to enable such systems to be built
- Critical issues
  - flexible, granular (sub-file level) data object identification
  - identification to survive
    - » movement of data from one operating system to another
    - » movement of data from one application to another
    - » movement of data between different networks
    - » expression of data in different hardware systems
    - » less than ideal network and storage environmental conditions
  - real time identification of rights in a particular data object
    - » enabling differential pricing in different global markets without leakage to higher priced markets
    - » enabling national treatment according to local cultural and legal norms
    - » ensures neutral or positive influence as tool in world intellectual property trade policy
  - real time differentiation between master data object and individual expressions of whole or parts of it
    - » this implies inbuilt replication control systems
  - real time, trivial cost, locally generated data object encoding systems
  - privacy not compromised
  - codification of licensing agreement terms required for automation of rights management systems
    - » this is necessary to enable real time open network rights management
  - legal
    - » matching identification and transaction technologies to differing legal regimes
  - time
    - » leakage minimized by decisive harmonious action
    - » change will take time
      - to encode valuable back catalogue
      - for all content owners to adopt standard coding system
      - for all computer and communications wares vendors to adopt open standard
      - to build supporting rights management and network infrastructures
  - cost
    - » lowering transaction costs while enhancing security is good business
  - recognition of interdependence of interests

Information Identification

Action Agenda

- Using existing standardization procedures as a model, on a pre-competitive basis, establish small, global, broad based, high level, ad hoc task force of content controllers involving major content organizations from Europe, North America, East Asia and Japan.
  - To agree need for universal approach to intellectual property identification and commit to implementation of resulting system
  - To set functional specification
    - » establish comprehensive set of questions requiring resolution
    - » delegate working groups to analyse and report on each
    - » drawing strongly on legal and technical advice
    - » integrate with existing industry-based identification systems
    - » communicate openly with participants and non-participants
  - To establish international administrative structure for long term promotion and management of the standards, including small professional support team
    - » stressing global standards, local implementation
  - To represent a combined voice on international policy issues in this domain
- Main task force to establish small multi-disciplinary task force of technical, legal and commercial leaders from content controllers, computer, communications, software and hardware industries.
  - To produce a comprehensive universal open standard data identification protocol and requisite agreements on open interchange standards for hardware and software
    - » coordinating role
    - » delegate to working groups to answer specific questions
    - » encourage pre-competitive agreement among computer and communications organisations to facilitate use of standard universal data identification (UDID) codes across operating systems, networks and applications
    - » involve ISO, ITU and other standards bodies as necessary
  - To make recommendations on the logical structure and physical features of a UDID protocol that is
    - » format independent
    - » object type independent
    - » unique to an object
    - » unique to the expression of an object
    - » easy and cheap to generate and use
    - » able to accommodate all known object types
    - » compatible with existing archives of objects
    - » an essential and integral part of the data structure of the object
  - To coordinate activities of competent organizations in development of comprehensive system
  - To advise main task force on choice of standard system
  - To oversee implementation of the standard system
  - To advise on standards, including upgrades, once established
- Establish cross industry working party to develop and deploy common copyright data model to enable automatic rights management
  - identifying
    - » Interested Parties (natural and legal)
    - » Works
    - » Unique Data Objects
    - » Codified terms of agreements between interested parties in relation to works and unique data objects

Appendix 1

The International Standard Recording Code

Practical Guide

IFPI

International ISRC Agency, London 1994

THE INTERNATIONAL STANDARD
RECORDING CODE

PRACTICAL GUIDE

(2nd Edition)

INTERNATIONAL ISRC AGENCY
LONDON 1994

BACKGROUND 1

The International Standard Recording Code (ISRC) was developed by the International Organisation for Standardisation (ISO) as a means of identifying (sound and audio-visual) recordings and is known as International Standard ISO 3901.

ISO is a worldwide federation of national standard bodies. ISRC was adopted as an international standard in 1986. The International Standard ISO 3901 was prepared and is administered by the Technical Committee ISO/TC 46, Documentation.

The purpose of this International Standard is to define and promote the use of standard code ISRC for the unique identification of recordings.

The International Federation of the Phonographic Industry (IFPI) recommended to its member companies in 1988 that the International Standard Recording Code system (ISRC) should be adopted as soon as possible as an international means of identification of the recordings on short form music videos.

In 1989 IFPI was appointed the International Registration Authority for ISRC by the International Organisation for Standardisation (ISO).

In the course of 1989, national registration authorities for ISRC were appointed in all European countries for the system to be launched in August 1989 for short form music videos.

With the successful introduction of ISRC for music videos IFPI was now actively considering a recommendation to its members to insert an ISRC into the sub-code of all digital sound recordings.

In the meantime the Japanese recording industry had already decided to press ahead with ISRC coding and issued guidelines for its implementation to all member companies of the Recording Industry Association of Japan (RIAJ) in November 1989.

At its meeting in Washington in March 1990, the Board of IFPI gave the IFPI/RIAA Working Group the task of finding a system to identify tracks of sound recordings which would be acceptable to the worldwide industry. The working group made the following recommendations:

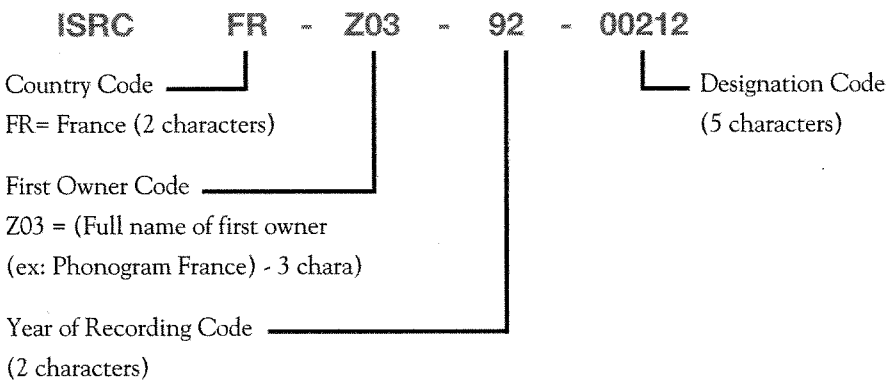
(a) Member companies should assign an ISRC to each track of digital sound recordings;

(b) The ISRC should be encoded in the sub-code for all digital sound carriers at the time the tape-master is prepared, together with a copy flag (refer to IEC 908 for CDs) and a relevant POS (Point of Sale) code.

The IFPI Board approved the recommendations of the Working Group on 19 March 1991 with a view to the IFPI Secretariat issuing detailed guidelines to the members of IFPI for implementation with effect from 1 January 1992.

For visual presentation an ISRC code should always be preceded by ISRC.

The structure of the ISRC is shown in the following example:



2. Functions and contents of code

2.1 Country Code

The Country Code identifies the country of residence of the first owner of the recording. It consists of two letters which have been allocated to each country by ISO. The International Registration Authority (IFPI) will hand over to the National ISRC Agencies a list of alpha-2 codes in accordance with ISO 3166 (code for the representation of names and countries). National ISRC Agencies (see section VII) shall inform all producers of the national Country Code.

Example: FR = France.

2.2 First Owner Code (Company Code)

The first owner of the recording is the producer of the recording (the (P) notice on a recording to be used as a guide to the correct allocation and use of company codes).

If the producer of a recording sells the recording with all rights before its ISRC is assigned, the acquirer should be considered as the first owner for the purposes of the ISRC.

The first owner of the recording is responsible for allocating the ISRC to the recording and informing the national agency of the ISRC details allocated.

The First Owner Code is alphanumeric and consists of three characters. This provides a maximum capacity of 46,655 (if 000 is not allowed) First Owner Codes. It is assigned by National ISRC Agencies.

Example: Z03 = Phonogram France.

2.3 Year of Recording Code

The Year of Recording Code shows the year in which the recording process is completed. The code consists of the last two digits of the year and is assigned by the first owner.

PRINCIPLES AND PROCEDURES TO BE OBSERVED BY PRODUCERS

IV

1. The ISRC should be included in all relevant documentation concerning a recording.
2. A producer must ensure that a competent person is responsible for the assignment of ISRC and the application of the pertinent regulations.
3. A producer will be assigned a First Owner Code (see III.2.2) by the National ISRC Agency.
4. A producer is responsible for assigning the Designation Code (see III.2.4).
5. A producer must keep a register of all ISRCs which that have been assigned.
6. The following chart provides information about the standard format for exchange of information by means of Electronic Data Interchange (EDI). This is the format intended for the first notification of an ISRC code recording.

The following layout is for an export file which is likely to be retrieved from an already existing database.

The line-numbering should be as indicated below.

WORK INFORMATION

NO	TYPE OF INFORMATION	NUMBER OF CHARACTERS	IDENTIFICATION LETTER (meaning of the abbreviation)
1	ISRC	12	
10	Title	120	
11	Composer(s)	120	
12	Arranger(s)	120	
13	Text writer(s)	120	
14	Language	20	
15	Publisher	80	
23	Pop/Classical	1	P (Pop)/C (Classical)

APPLICATION OF ISRC

V

1. General

A separate ISRC must be assigned to every different track of a recording but not to an unchanged track when it is reused on a new album of recordings.

1.1. Examples of the application of an ISRC

Example 1. New recording:

For an album comprising ten recordings taken from a master tape completed in 1991.

ISRC FR-Z03-91-01231 (Recording No 1)

ISRC FR-Z03-91-01232 (Recording No 2)

ISRC FR-Z03-91-01233 (Recording No 3)

" " " " " "

ISRC FR-Z03-91-01240 (Recording No 10)

Example 2. Compilation:

For compiling a new album using previously released recordings

Examples of using complete recordings in their entirety

ISRC FR-Z03-91-01234 (Recording No 1)

ISRC FR-Z03-90-02345 (Recording No 2)

ISRC FR-Z03-89-03456 (Recording No 3)

" " " " " "

ISRC FR-Z03-88-06789 (Recording No 10)

2. Experiences from daily use

The following are some of the day-to-day problems of the application of the ISRC.

2.1 Re-mix

If multiple recordings are produced in the same recording session without any change in orchestration, arrangement, or artist and if they are preserved or turned into commercial products, each recording shall be encoded with a new ISRC.

A new ISRC will be given to a new recording produced through re-mixing.

It is recommended that the ISRC numbers of the original recordings used in the remixing be kept on file by the producer.

2.2 Changes in the playing time

The playing time of a recording is an important characteristic as it is used for product design and also

ENCODING OF THE ISRC

VI

The ISRC is encoded in digital sound carriers in the pre-mastering process in accordance with each system specification. An edited Master Tape for mass production and the corresponding ISRC information must be prepared at this stage.

Therefore, it is desirable that ISRC is allocated in each recording when a Master Tape is completed or the decision on the release of a recording has been taken.

U-Matic tapes or Exabyte or CD Rom or CD MO can be used as carriers for CD pre-mastering.

The ISRC's encoding, together with the PQ-data inserted in the U-MATIC tape for CD, are encoded in the disc sub-code (Q channel) in the disk mastering process. For this reason, ISRCs must be encoded for each track in the Master Tape for CD. The ISRC codes, together with the SCMS copy inhibitor, and the relevant point of sale bar-code (EAN/UPC) should be inserted on Master Tapes for CD in the pre-mastering process for making Master Tapes for CD from the original Master Tape. A PQ-editor is used for this.

This Master Tape for CD can be used as a Master Tape for producing DCC and MD. However, since it is mandatory to insert text information for DCC, attaching text data with a diskette, etc. is necessary.

Music videograms (analogue format) should carry an ISRC number on the time-clock and, it is recommended, on the label on the outside of the box.

ADMINISTRATION OF THE ISRC SYSTEM

VII

The administration of the ISRC system is carried on at two levels:

1. International administration

The international administration of the system is in the hands of the International ISRC Agency which has an Advisory Board representing the ISO, the national agencies and producers.

The address of the International Agency is:

INTERNATIONAL STANDARD RECORDING CODE
International Federation of the Phonographic Industry (IFPI)
IFPI Secretariat, 54, Regent Street
London W1R 5PJ, United Kingdom
Telephone: (4471) 434 3521, Facsimile: (4471) 439 9166

a)	Country Code	First Owner Code	First Owner (name & address)	Department or person to be contacted (if necessary)
b)	First Owner (name & address)	Country Code	First Owner Code	Department or person to be contacted (if necessary)

(8) act as controller and arbitrator for all ISRC matters in its country;

(9) report periodically, at least once a year to the International Registration Authority (IFPI) on the implementation of ISRC in their respective countries and consult with IFPI before any new code is used to identify a specific recording format (see section III 2.5).

THE BENEFITS OF ISRC

In view of technological trends in the field of consumer electronics, information, broadcasting and telecommunication technologies, the industry needs to prepare itself for the time when phonograms and videograms will eventually be distributed directly to the consumer's home by electronic means.

With the development of Digital Audio Broadcast (DAB) and electronic delivery systems of sound recordings, the phonographic industry is faced with a challenge to keep control of the use made of its works. Music disseminated in electronic form will no longer by identifiable as a tangible property. Technical methods need to be developed to enable the industry to collect remuneration or prevent unauthorised use.

In order to prepare itself to meet this new challenge the industry has been working on programmes aimed at music track identification. The implementation of a worldwide system for the identification of music will have the following advantages:

(a) It will enable the use of copyright protected works to be controlled;

(b) It will facilitate the distribution and collection of royalties (performances, private copying) as appropriate;

(c) It will assist in the fight against piracy.

In view of the industry's requirements, the adoption of the ISRC system by the industry has the

NOTES

International Federation of the Phonographic Industry
IFPI Secretariat, 54 Regent Street, London W1R 5PJ United Kingdom
Telephone: (4471) 434 3521, Facsimile: (4471) 439 9166

Published by IFPI © 1994 Designed and printed by The Advertising Business, London W1V 5LR. Telephone: 071-439 6722

Appendix 2

Proposal for An International Standard Work Code (ISWC)

Godfrey Rust, MCPS
CISAC/BIEM International Numbering Working Group

Version 5

London, December 3 1994

Appendix C6

Proposal for an International Standard Work Code (ISWC)

The attached document is a draft for a proposed standard for an International Standard Work Code (ISWC). The document draws heavily on the existing ISO standard 10957 for the International Standard Music Number (ISMN) recently adopted for sheet music.

Version 4 has been produced following three meetings of the ISSC Numbering Group and consultation with the ISMN agency in Berlin and representatives of the ICMP (International Confedation of Music Publishers).

Godfrey Rust, MCPS
CISAC/BIEM International Numbering Working Group
Version 5
London December 3 1994

ISSC/IN/2.5

Proposal for an International Standard

INTERNATIONAL STANDARD WORK CODE (ISWC)

Version 5/December 3 1994

1 SCOPE

This International Standard specifies a means of uniquely identifying musical works. It standardises and promotes internationally the use of numbers for this type of intellectual property so that musical works can be uniquely distinguished from one another within computer databases and related documentation by means of an international standard code.

2 DEFINITIONS

For the purposes of this International Standard, the following definitions apply

2.1 Musical Work

[To be supplied]

2.2 Allocator

Any person, company or organisation who is authorised to assign ISWCs. An Allocator must be the owner or administrator of copyright(s) in one or more music works by virtue of being the creator of the work or his authorised administrator (for example, a composer, writer, publisher or collecting society).

2.3 Check digit

Added digit which may be used to verify the accuracy of a standard number through a mathematical relationship to the digits contained in that number [Adapted from ISO 7064].

3 CONSTRUCTION OF AN ISWC

An International Standard Work Code consists of the letter T followed by eight digits and a numeric check digit. Specifically the elements are:

- The letter T as a prefix — 1 character (T)
- An administrator identifier — } together being
- A work identifier — } eight digits
- A check digit — 1 digit

When an ISWC is written or printed it shall be preceded by the letters ISWC and each element shall be separated by a space or a hyphen as in these examples:

ISWC T-01-637478-4
ISWC T 9586 3002 4

3.1 **Letter T prefix**
The first element of the ISWC shall be the letter T. The function of the T prefix is to permit ISWC to be distinguished from ISMN and other similarly constructed numbers in situations, such as some computer applications, where the letters ISWC are not included with the number.

3.2 **Allocator identifier**
The second element of the ISWC identifies the allocator who has allocated the ISWC. It varies in length from allocator to allocator according to the output of each allocator, and an allocator may have more than one identifier.

3.3 **Work identifier**
The third element of the ISWC shall be the work identifier. This element is allocated by the allocator to identify the specific unique music work. The length of the work identifier is determined by the length of the allocator identifier which precedes it.

3.4 **Check digit**
The fourth element of the ISWC shall be the check digit. The check digit is calculated on a weighted modulus *[modulus to be agreed].*

4 **ADMINISTRATION**
The ISWC system shall be administered by the international registration agency appointed for the purpose; by other appropriate national or regional agencies appointed by the international agency; and by allocators appointed by national or regional agencies.

Annex A (normative)
GUIDELINES FOR THE USE OF ISWC

A.1 WORKS FOR WHICH ISWCS MAY BE ASSIGNED

A.1.1 Eligible repertoire
ISWCs may be assigned to any musical work, newly-created or already existing, irrespective of copyright status.

A.1.2 Versions
A new version of a work (for example, an arrangement, lyircal adaptation or translation) shall be assigned a new ISWC.

A.1.3 Excerpts
An ISWC may be assigned to any recognised excerpt from another musical work (eg an aria from an opera or a movement from a symphony) where such an excerpt requires identification in its own right.

A.1.4 Composite works
An ISWC may be assigned to any recognised composite of other musical works (such as a cycle of operas or an arrangement of works in the form of a medley) where such a composite requires identification in its own right.

A.2 ADMINISTRATION OF ISWCS

A.2.1 Central authority
The central registration authority for the ISWC shall be the International ISWC Agency. This agency shall appoint local agencies.

A.2.2 Local agencies
There shall be national or regional agencies for the administration of the ISWC (normally a national collecting society). These agencies shall authorise individual Allocators.

A.2.3 Authorised Allocator
The Allocator of an ISWC for a musical work will be its creator (eg the composer or author) or his authorised agent (eg publisher or society).

A.2.4 Musical works with more than one allocator ("split copyrights")
Where a work has more than one authorised allocator (for example, being co-written or co-owned), where possible the ISWC will be assigned by one authorised allocator by prior agreement. Where an ISWC is assigned for a work of shared authorship it is the responsibility of the registering national

agency to ensure that all other interested parties are notified of the ISWC at the earliest opportunity to prevent duplication.

A.2.5 **Public domain works**
The authorised allocator for assigning ISWCs for works in the public domain shall be the appropriate national agency for the creator or ethnic origin of the music work.

A.2.6 **Duplicate ISWCs**
The same ISWC cannot be assigned to more than one work. On the other hand, if one work has more than one ISWC assigned to it, they may remain in circulation.

A.2.7 **Re-use of ISWCs**
Once assigned ISWCs shall never be used again, even if found to have been issued in error.

A.2.8 **Supporting data**
Allocators of the ISWC must capture the essential supporting data for the work and make it available in the agreed international standard format. The supporting data includes, at minimum, the following:

- One **title** of the work
- All **creators** of the work (composers, authors, arrangers, translators etc) identified by their international CAE number where one exists
- The **ISWC** of the work
- In the case of a **version** or **excerpt**, a code indicating this status
- In the case of a version or excerpt, the **ISWC of the parent work** (or the title where no ISWC exists)

The full specification of the international copyright property data standard is available from the national or international ISWC agency.

A.3 APPLICATION OF ISWC

Details of the application of the ISWC will be explained in a user's manual available from the ISWC Agencies.

A.4 RELATED INTERNATIONAL CODES

A.4.1 **CAE**
The CAE (Compositeur, Auteur, Editeur) number is a unique international identifier of authors, composers and publishers administered by the Swiss copyright protection society SUISA.

A.4.2 **ISRC**

The International Standard Recording Code (ISRC) is the standard international identifier for a sound or video recording. The linking of the ISRC to the ISWC(s) of the work(s) recorded is a key to effective international administration of musical copyrights, mechanical, performing and other.

A.4.3 **ISMN**

The International Standard Music Number (ISMN) is the standard international identifier for printed music editions. Its number structure differs from ISWC only in the substitution of the letter "M" in place of "T" and in the method of calculation of the check digit. However, there is no structural connection between an ISMN and an ISWC for a related work.

Appendix C6

Proposal for an International Standard Work Code (ISWC)

The attached document is a draft for a proposed standard for an International Standard Work Code (ISWC). The document draws heavily on the existing ISO standard 10957 for the International Standard Music Number (ISMN) recently adopted for sheet music.

Version 4 has been produced following three meetings of the ISSC Numbering Group and consultation with the ISMN agency in Berlin and representatives of the ICMP (International Confedation of Music Publishers).

Godfrey Rust, MCPS
CISAC/BIEM International Numbering Working Group
Version 5
London December 3 1994

ISSC/IN/2.5

Proposal for an International Standard

INTERNATIONAL STANDARD WORK CODE (ISWC)

Version 5/December 3 1994

1 SCOPE

This International Standard specifies a means of uniquely identifying musical works. It standardises and promotes internationally the use of numbers for this type of intellectual property so that musical works can be uniquely distinguished from one another within computer databases and related documentation by means of an international standard code.

2 DEFINITIONS

For the purposes of this International Standard, the following definitions apply

2.1 Musical Work

[To be supplied]

2.2 Allocator

Any person, company or organisation who is authorised to assign ISWCs. An Allocator must be the owner or administrator of copyright(s) in one or more music works by virtue of being the creator of the work or his authorised administrator (for example, a composer, writer, publisher or collecting society).

2.3 Check digit

Added digit which may be used to verify the accuracy of a standard number through a mathematical relationship to the digits contained in that number [Adapted from ISO 7064].

3 CONSTRUCTION OF AN ISWC

An International Standard Work Code consists of the letter T followed by eight digits and a numeric check digit. Specifically the elements are:

- The letter T as a prefix — 1 character (T)
- An administrator identifier — } together being
- A work identifier — } eight digits
- A check digit — 1 digit

When an ISWC is written or printed it shall be preceded by the letters ISWC and each element shall be separated by a space or a hyphen as in these examples:

ISWC T-01-637478-4
ISWC T 9586 3002 4

3.1 **Letter T prefix**
The first element of the ISWC shall be the letter T. The function of the T prefix is to permit ISWC to be distinguished from ISMN and other similarly constructed numbers in situations, such as some computer applications, where the letters ISWC are not included with the number.

3.2 **Allocator identifier**
The second element of the ISWC identifies the allocator who has allocated the ISWC. It varies in length from allocator to allocator according to the output of each allocator, and an allocator may have more than one identifier.

3.3 **Work identifier**
The third element of the ISWC shall be the work identifier. This element is allocated by the allocator to identify the specific unique music work. The length of the work identifier is determined by the length of the allocator identifier which precedes it.

3.4 **Check digit**
The fourth element of the ISWC shall be the check digit. The check digit is calculated on a weighted modulus *[modulus to be agreed].*

4 **ADMINISTRATION**
The ISWC system shall be administered by the international registration agency appointed for the purpose; by other appropriate national or regional agencies appointed by the international agency; and by allocators appointed by national or regional agencies.

Annex A (normative)
GUIDELINES FOR THE USE OF ISWC

A.1 WORKS FOR WHICH ISWCS MAY BE ASSIGNED

A.1.1 Eligible repertoire
ISWCs may be assigned to any musical work, newly-created or already existing, irrespective of copyright status.

A.1.2 Versions
A new version of a work (for example, an arrangement, lyircal adaptation or translation) shall be assigned a new ISWC.

A.1.3 Excerpts
An ISWC may be assigned to any recognised excerpt from another musical work (eg an aria from an opera or a movement from a symphony) where such an excerpt requires identification in its own right.

A.1.4 Composite works
An ISWC may be assigned to any recognised composite of other musical works (such as a cycle of operas or an arrangement of works in the form of a medley) where such a composite requires identification in its own right.

A.2 ADMINISTRATION OF ISWCS

A.2.1 Central authority
The central registration authority for the ISWC shall be the International ISWC Agency. This agency shall appoint local agencies.

A.2.2 Local agencies
There shall be national or regional agencies for the administration of the ISWC (normally a national collecting society). These agencies shall authorise individual Allocators.

A.2.3 Authorised Allocator
The Allocator of an ISWC for a musical work will be its creator (eg the composer or author) or his authorised agent (eg publisher or society).

A.2.4 Musical works with more than one allocator ("split copyrights")
Where a work has more than one authorised allocator (for example, being co-written or co-owned), where possible the ISWC will be assigned by one authorised allocator by prior agreement. Where an ISWC is assigned for a work of shared authorship it is the responsibility of the registering national

agency to ensure that all other interested parties are notified of the ISWC at the earliest opportunity to prevent duplication.

A.2.5 **Public domain works**
The authorised allocator for assigning ISWCs for works in the public domain shall be the appropriate national agency for the creator or ethnic origin of the music work.

A.2.6 **Duplicate ISWCs**
The same ISWC cannot be assigned to more than one work. On the other hand, if one work has more than one ISWC assigned to it, they may remain in circulation.

A.2.7 **Re-use of ISWCs**
Once assigned ISWCs shall never be used again, even if found to have been issued in error.

A.2.8 **Supporting data**
Allocators of the ISWC must capture the essential supporting data for the work and make it available in the agreed international standard format. The supporting data includes, at minimum, the following:

- One **title** of the work
- All **creators** of the work (composers, authors, arrangers, translators etc) identified by their international CAE number where one exists
- The **ISWC** of the work
- In the case of a **version** or **excerpt**, a code indicating this status
- In the case of a version or excerpt, the **ISWC of the parent work** (or the title where no ISWC exists)

The full specification of the international copyright property data standard is available from the national or international ISWC agency.

A.3 APPLICATION OF ISWC

Details of the application of the ISWC will be explained in a user's manual available from the ISWC Agencies.

A.4 RELATED INTERNATIONAL CODES

A.4.1 **CAE**
The CAE (Compositeur, Auteur, Editeur) number is a unique international identifier of authors, composers and publishers administered by the Swiss copyright protection society SUISA.

A.4.2 ISRC

The International Standard Recording Code (ISRC) is the standard international identifier for a sound or video recording. The linking of the ISRC to the ISWC(s) of the work(s) recorded is a key to effective international administration of musical copyrights, mechanical, performing and other.

A.4.3 **ISMN**

The International Standard Music Number (ISMN) is the standard international identifier for printed music editions. Its number structure differs from ISWC only in the substitution of the letter "M" in place of "T" and in the method of calculation of the check digit. However, there is no structural connection between an ISMN and an ISWC for a related work.

Appendix 3

The Common Copyright Data Model
Draft November 1994
G Rust MCPS

Appendix C1

The Common Copyright Data Model

Draft/November 1994/G Rust MCPS

The electronic distribution of digitised copyright material - whether music, words, pictures or software, alone or in combination - will turn the world of copyright administration inside out within the next decade.

The need for a simple, effective, common means of identifying copyright material and its ownership is self-evident.

Unique international numbering of copyright material is advancing, with coding systems like ISBN (books), EAN (records) , ISRC (recordings) and the proposed ISWC (musical works).

However, numbers are only part of what is needed. Things which are numbered also need to be described. There is no international standard method of describing copyright material, or for storing and communicating this descriptive information to other parties.

Traditional cataloguing methods such as MARC are inappropriate for the computer age. However, something like the "MARC catalogue record" is needed as a universal format for copyright data recognition. This might be described as an "EDI" format.

As well as this, many organisations are involved in re-engineering their computer systems to deal with changing times, and in their analysis are covering the same ground over and over again.

The **common copyright data model** has been developed to meet this need. In itself it is neither a computer file specification, nor an EDI format: it provides the underlying data model for either of these applications.

The model works for any kind of copyright material including music, film, literary texts and computer software.

Common copyright data model

1 Table of entities

There are four data entity types:

Data Entity Type	Definition	Examples
Interested Party (IP)	A person (natural or legal) who has contributed in some way to the creation of a property or acquired rights in it.	Composers, authors, performers, producers, publishers, record companies, film companies, copyright societies
Work	An intellectual creation which is, was or may be protected by law on authors' or neighbouring rights. Works may or may not take on a physical form.	Musical works, sound recordings, audiovisual work, literary works, photographs, books, videocassettes, sheet music, computer software, multimedia products
Agreement	An agreement between IPs which determine rights to works	Agreements between authors and publishers, performers and record companies, publishers and sub-publishers
Licensing Scheme	A scheme or agreement under which copyrights are used	Society licensing schemes, publisher licences, industry agreements

Note that all kinds of recordings and products are grouped under the Work entity type.

Common copyright data model

2 Data entity structure: IPs

Data element	Description	Structure
Name(s)	Name(s) or title(s) of IP	Name-Type parameter code, plus alpha-numeric data of at least 50 characters plus
Number(s)	Number(s) used to identify IP	Number-Type parameter code, plus alpha-numeric data of at most 20 characters
Class(es)	Classification(s), role(s) of IP	Class-Type parameter code, plusdata in the form of parameter Value
Event(s)	Events in the IP's life	Event-Type parameter code, plus Date or date range, plus Place

Each of these data fields may have multiple occurrences. IPs may also have links to other IPs.

Common copyright data model

3 Data entity structure: Works

Data item	Description	Structure
Name(s)	Name(s) or title(s) of work	Name-Type parameter code, plus alpha-numeric data of at least 50 characters plus
Number(s)	Number(s) used to identify work	Number-Type parameter code, plus alpha-numeric data of at most 20 characters
Class(es)	Classification(s), format(s) of work	Class-Type parameter code, plusdata in the form of parameter Value
Event(s)	Events in the work's life	Event-Type parameter code, plus Date or date range, plus Place
Extent(s)	Measured extent of work (size, duration)	Extent-Type parameter code, plus numeric data
Contributor(s)	Contributors/IPs	Contributor-role parameter code, plus Contributor IP-ID
Contents	Works contained within this work	Entity-Type parameter code, plus Entity-ID

Each of these data fields may have multiple occurrences. Properties may also have links to other properties of the same type.

Common copyright data model

4 Data entity structure:Agreements *(unfinished)*

Data element	Description	Structure
IP(s)	IP(s) who are parties to the agreement	IP number
Shares	Shares allocated to each IP	%
Agreement Number(s)	Number(s) used to identify agreement	Agreement number
Agreement Type	Classification(s) of agreement	Agreement-type code
Right(s)	Right(s) covered by the agreement	Right code(s)
Entitlement(s)	Entitlement(s) covered by the agreement	Entitlement(s)
Duration	Period of time covered by the agreement	Date range
Territorie(s)	Places covered by the agreement	Territory code(s)
Schedule(s)	Schedules of works covered by this agreement	Schedule number(s)

Each of these data fields may have multiple occurrences.

Examples of IPs (1)

Data item	*Par*	*Seq*	*Data*	*Code Explanation*
Name(s)	AUT	1	Geoff Shattock	*Authorised name*
	BOR	1	Geoffrey Paul Shattock	*Born name*
Number(s)	ID	1	1234567890	*IP ID*
	CAE	1	3452364622	*CAE Number*
Class(es)	ENT	1	IND	*IP Type=Individual*
	ROL	1	LYR	*Role=Lyricist*
	ROL	2	C	*Role=Composer*
	ROL	3	VOC	*Role=Vocalist*
	ROL	4	GUI	*Role=Guitar*
	GEN	1	CCM	*Genre=Cont. Christian*
Event(s)	BOR	1	1957.03.11/Leicester, England	*Born*

Examples of properties (1)

Audio LP

Data item	*Par*	*Seq*	*Data*		*Code Explanation*
Name(s)	AUT	1	Diamond Life		*Authorised name*
Number(s)	ID	1	1234567890		*Property-ID*
	CAT	1	EPC26044		*Catalogue No*
	CAT	2	26044		*Catalogue No*
	BAR	1	5099702604444		*Barcode*
	LAB	1	CB481		*Label Code*
Class(es)	ENT	1	AP		*Property Type=Audio*
	FOR	1	12V		*Format=12inch Vinyl*
	FOR	2	ALB		*Format=Album*
	GEN	1	POP		*Genre=Pop*
Event(s)	REC	1	1983/Power Plant		*Recorded*
	PUB	1	1984/UK		*Published*
Extent(s)	DUR	1	00.51.30		*Duration*
	DIA	1	12		*Size*
IP(s)	*HDR*	*1*		*Group members*	*Group header*
	GRP	1	1234567890	Sade	*Group*
	VOC	1	9874563743	Sade Adu	*Vocals*
	SAX	1	8347348347	Stuart Matthewman	*Sax*
	GUI	1	8347348347	Stuart Matthewman	*Guitar*
	KEY	1	3627222222	Andrew Hale	*Keyboards*
	EBS	1	2558869561	Paul S Denham	*Bass*
	HDR	*2*		*Supporting musicians*	*Group header*
	PER	2	9385933444	Martin Ditcham	*Percussion*
	DRU	2	3434343422	Paul Cooke	*Drums*
	TRU	2	7755123455	Terry Bailey	*Trumpet*
	TRU	2	3463473743	CordonMatthewman	*Trumpet*
	HDR	*3*		*Production*	*Group header*
	PRO	3	2345678901	Robin Millar	*Producer*
	ENG	3	7890123455	Mike Pela	*Engineer*
	ENG	3	3436473433	Ben Rogan	*Engineer*
	HDR	*4*		*Package design*	*Group header*
	SLD	4	3456789012	Graham Smith	*Sleeve design*
	MKP	4	0339449434	Paul Gobal	*Makeup*
	PHO	4	4567890123	Chris Roberts	*Photographer*
	HDR	*5*		*First publication*	*Group header*
	LAB	5	0343438439	Epic	*Label*
	PUB	5	4374374633	Sony Music UK Ltd	*First publisher***
Contents	REC	1	3456789001		*Recording-ID**
	REC	2	3456789002		*Recording-ID*
	REC	3	3456789003		*Recording-ID*
	REC	4	3456789004		*Recording-ID*
	REC	5	3456789005		*Recording-ID*
	REC	6	3456789006		*Recording-ID*
	REC	7	3456789007		*Recording-ID*
	REC	8	3456789008		*Recording-ID*
	REC	9	3456789009		*Recording-ID*
	REC	10	3456789010		*Recording-ID*

Note: there is a minimum necessary amount of information for any property. In this case one Name, one Number (excluding internal IDs), two Classes (one property type and one format) and one IP. The Sound Carrier can be accurately but incompletely described without any contents.

**Note: "publisher" here refers to the (P) publisher of the sound carrier, not the publisher of any sound recordings or musical works.
*Note: recording-IDs can be internal system IDs or ISRCs.

Examples of properties (2)

Sound Recording

Data item	*Par*	*Seq*	*Data*		*Code Explanation*
Name(s)	AUT	1	Kind of Magic		*Authorised name*
Number(s)	ID	1	1234567890		*Property ID*
	ISR	1	238237283123		*ISRC*
	MAS	1	232323		*Master tape no*
Class(es)	ENT	1	AR		*Prop. Type=Audio Rec.*
	FOR	1	ADD		*Format=ADD*
	GEN	1	CCM		*Genre=Cont. Christian*
Event(s)	REC	1	1994.07.12-26/ICC Studios		*Recorded*
	MIX	1	1994.09.13-15/ICC Studios		*Mixed*
	P	1	1994/UK		*Published*
Extent(s)	DUR	1	04.35		*Duration*
IP(s)	*HDR*	1		*Group members*	*Group header*
	GRP	1	1234567890	Shattock & Smith	*Group*
	VOC	1	2633646362	Geoff Shattock	*Vocals*
	GUI	1	2633646362	Geoff Shattock	*Guitar*
	BVC	1	2362632636	John Smith	*Backing vocals*
	GUI	1	2362632636	John Smith	*Guitar*
	HDR	2		*Supporting musicians*	*Group header*
	BVC	2	2387881111	Jenny Legg	*Backing vocals*
	PIA	2	3434634663	Nick Lacey	*Keyboards*
	TSX	2	1234623626	Dave Fitzgerald	*Tenor saxophone*
	EBS	2	3473473744	Paul Kimber	*Electric bass*
	DRU	2	2362362366	Tony Marsh	*Drums*
	PER	2	2372372737	Dave Engel	*Percussion*
	EGT	2	2374643643	Ian Smith	*Electric guitar*
	HDR	3		*Production*	*Group header*
	PRO	3	2384734352	John Pantry	*Producer*
	ENG	3	2385735112	Paul Freeman	*Engineer*
	ENG	3	0594753754	Bruce Pont	*Engineer*
Contents	MWK	1	T236263626		*ISWC*

Examples of properties (3)

Musical Work

Data type	*Par*	*Seq*	*Data*	*Code Explanation*
Name(s)	AUT	1	Symphony No50 in D major	*Authorised name*
	POP	1	Saturn	*Popular name*
Number(s)	ID	1	1234567890	*Property ID*
	CAT	1	K1234	*Catalogue No*
	ISW	1	T12912919	*ISWC*
Class(es)	ENT	1	MWK	*Property Type=Mus. Work*
	GEN	1	CLA	*Genre=Classical*
	FRM	1	SYM	*Form=Symphony*
Event(s)	COM	1	1777/Vienna	*Composed*
	PER	1	1778/Vienna	*First Performed*
Extent(s)	MVM	1	4	*Movements*
IP Link(s)	C	1	1234567890 Wolfgang A Mozart	*Composer*
Contents	*none*			

Appendix 4

Notes for Document Identification discussion
Dr Norman Paskin, Elsevier Science
9 December 1994

A.
From: Serial Publications: Guidelines for Good practice in Publishing Printed Journals and other Serial Publications; UK serials Group, 1994
ISBN 0 906148 10 3

B.
Specification of Elsevier Science SSDI

C.
Notes on SICI prepared by Elsevier Science 8 September 1994

D.
CODEN: Introductory pages from the International Coden Directory

E.
"A Standard Book Number for the Internet": pp 560-561 of "The Internet for Everyone", R.W. Wiggins, McGraw-Hill 1994
ISBN 0 07 067018 8

F.
Proposed IAFA template for documents

Notes for Document Identification discussion

These notes represent suggested starting points for discussion, and are not guaranteed to be comprehensive.

1. Suggested principles for document identification

- Format (presentation) independent. Relates to semantic content.
  → does not include reference to page numbers (cf SGML coding and independence of content from presentation).
  - SICI does not obey this principle.

- Unique to a document.
  → a number cannot be accidentally created identically from two sources.
  - basing on a recognised unique identifier for serials (e.g. ISSN, CODEN) or books (ISBN) ensures that duplication can only occur if the issuing publishing house duplicates.

- Easy to generate and use.
  → alphanumeric only.
  → algorithmically explicable.

- Not restrictive. Able to accomodate many article types.
  → e.g. book chapters as well as serial articles
  → e.g. all article types in a journal

- Serves only one purpose. Does not carry any "compulsory" explicit meaning other than that of unique identification.
  → e.g. if it contains "1995" this should not be taken to mean "year of publication"; information of this sort belongs within the article.
  - (explicit meaning may exist for the originator but not the user; e.g. one publisher may choose to use 123456 for volume 123, issue 45 article 6, but this is not a requirement and not publicised to users).

- Compatible with (does not cause problems for) any existing archives of articles.
  → possibly more of a problem? Is this a real issue? I am thinking in particular of CAS.

- Generated by the originator of the published item (i.e. the publisher)
  → not reliant on validation by an external body, can be assigned instantly as part of the production process.

Norman Paskin Elsevier Science 9 December, 1994

2. Relevant existing standards

The following is a summary of standards which I am aware of. For more information on some of these, a good summary is found in Serial Publications: Guidelines for Good practice in Publishing Printed Journals and other Serial Publications; UK serials Group, 1994 ISBN 0 906148 10 3: I have attached a copy of some relevant pages (Attachment A).

SSDI Standard Serials Document Identifier

Used by Elsevier Science (implemented 1994 onwards, and being extended to all journals; will appear as a printed number on the article). See Attachment B.
The SSDI as currently defined covers serial journal articles; however an extension to books is easily possible:

Current SSDI has form **xxxxxxxxyyddddc**

x	8 digits, ISSN
y	2 digits, year assigned
d	5 digits, sequence num
c	1 digit, check character

Possible extension to books: **x,y** replaced by ISBN
add prefix e.g. S for serial item, B for books

(This possible extension is not yet implemented)

ADONIS number

Essentially same as SSDI. Elsevier refined the check digit algorithm, as the early version was inaccurate.

SICI Serial Item and Contribution Identifier

Proposed by SISAC and adopted as NISO Z 39-56. Not format independent (the SICI uses page numbers); can be derived from physical manifestation at any time.
Under revision now (due Dec 1994?).
A number of stm publishers print the code (in its bar coded form) on their journals; SICI appears to have less following in EU than USA.
See attachment C.

Biblid

The official ISO standard (ISO 9115-1987), . Not widely implemented. Essentially similar to SICI. It is not yet clear whether the revision of SICI will address the issue of incompatibility with Biblid.

ISSN

Serial (not document) identifier. Widely used/accepted. *Note: there seems to be some confusion in the STM community at present over the allocation of ISSNs to electronic journals: ISSNs have appeared on some electronic-only titles (e.g. Complexity International); when a journal appears in both printed and electronic*

forms, should the two forms have the same ISSN (I would argue yes, see principles above, but am told that others argue no); can recipients of this note comment?

- *ISSN/EAN bar code*

  The implementation of ISSN as part of the European Article Numbering standard bar coding; widely used for periodicals sold through the retail trade (magazines) but not implemented in the stm world.

- **CODEN**

  Like ISSN, a serial (not document) identifier. Not as widely used as ISSN. In theory can be extended to article level (see attachment D),but I am not aware of any activities in this regard. Based on ASTM standards initiative.

- **Internet URL: uniform resource locator**

  Identifies location not document (one document at two locations = two URLs). Universal on WWW implementations; little correspondence to other media. Essentially a UNIX file name.

- **Internet URN: uniform resource name and IAFA: Internet Assigned Fields Authority**

  Not yet accepted standards, currently under draft discussion, URI (Uniform Resource Identifier) is a generalisation of the concept of URL, which will also embrace URN. See attachment E. Original aim was to make URN into an "ISBN" for the Net, but seems to be heading away from document standards. (e.g. separate URN for ASCII and PDF versions?)
  IAFA is a proposal for templates, essentially for catalogue-like entries on the Internet. See attachment E. Whilst the IAFA is not a document identification number, the possibility of such a catalogue entry existing and being capable of algorithmic conversion to a number is relevant.

  Proposed WWW consortium(s) may take initative in developing these concepts? What is clear is the the URL/URN/IAFA identification discussion is coming froma very different direction to that of traditional publishers.

Attachments:

A. From: Serial Publications: Guidelines for Good practice in Publishing Printed Journals and other Serial Publications; UK serials Group, 1994 ISBN 0 906148 10 3

B: Specification of Elsevier Science SSDI

C: Notes on SICI prepared by Elsevier Science 8 Sept 1994

D: CODENS: Introductory pages from International Coden Directory

E: "A Standard Book Number for the Internet": pp 560-561 of "The Internet for Everyone", R.W. Wiggins, McGraw-Hill 1994, ISBN 0 07 067018 8

F: Proposed IAFA template for documents.

Norman Paskin Elsevier Science 9 December, 1994

Copyright Clearance Center
27 Congress Street
Salem
Massachusetts 01970
USA
Tel: 617 744 3350

35. Abstracting and indexing services

35.1 Abstracting and indexing services provide the chief means of access to individual articles in periodicals, since libraries do not usually catalogue articles in periodical issues (see also section 24 on abstracts and summaries). Abstracting and indexing services are provided by various types of organization in print, online and on CD-ROM.

35.2 It is usual for abstracting and indexing services to cover not only periodical articles but also books, reports, conference proceedings and so on. Subject areas covered by individual services may be broad, as in *Chemical Abstracts* and *Humanities Index*, or narrow and more specialized, as in *Soyabean Abstracts* and *Bibliography of Old Norse-Icelandic Studies*. Some services simply reproduce contents pages of periodicals while others list titles of articles, with or without abstracts, and index them by author, subject, keyword and so on.

35.3 It is unlikely that any user of serials has access to all titles relevant to their interests in their local library, no matter how well endowed that library may be. Even if a wide range of serials is accessible it is doubtful whether anyone has sufficient time to scan them all regularly. Many users of serials, therefore, rely on abstracting and indexing services either to search for articles relevant to their research or simply to keep pace with what is happening in their area of interest. Coverage by relevant services is not only time saving for the user but also an effective means of promoting a title. Searches in abstracting and indexing services will often result in requests for copies of articles which will not only generate income from reproduction fees but may also generate subscriptions. Systematic single article supply to end users by document delivery services is increasing significantly and more and more of them are paying fees to rights' owners.

35.4 A publisher who would like a periodical title covered by abstracting and indexing services should first establish if there are services appropriate for the title. *Ulrich's International Periodicals Directory*, among others, lists abstracting and indexing services separately for most of the subjects it covers. Other sources of information to try are local libraries with extensive reference collections and libraries specializing in relevant subjects. Once having established which services are appropriate then one or more recent issues of the periodical should be sent with a request that it be included in the service. The launching of a new journal will often prompt requests from various abstracting and indexing services for free subscriptions. The cost of satisfying these requests should be weighed carefully against the benefit they might provide. For many journals adequate coverage will be provided by the major services.

35.5 When a title is abstracted and indexed the services which cover the title should be listed in the masthead (see section 8).

36. International Standard Serial Number (ISSN)

36.1 ISSN are unique standard numbers assigned to serial titles by the International Serials Data System's (ISDS) worldwide network. There is an international centre in Paris and about fifty national centres most of which are in national libraries. The United Kingdom's centre, the ISSN UK Centre is in the British Library.

36.2 ISSN are used as control numbers in library cataloguing systems, in ordering, claiming, check-in and circulation systems in libraries, in document delivery, royalty payments, copyright clearance, information retrieval, etc. In the United States they are used as the equivalents of postal service numbers for favourable mailing rates. They are also cited in bibliographies and lists of serials and in subscription agents' lists. A growing important use is as integral parts of the numbers from which bar codes are constructed (see section 38). They are of value in a wide variety of computer applications in libraries, subscription agents, document supply centres, abstracting and indexing services and publishers.

36.3 It is recommended that the ISSN is printed as two blocks of four digits separated by a hyphen, for example:

ISSN 0953-0460
(the ISSN for *Serials: the Journal of the United Kingdom Serials Group)*

36.4 The international standard also recommends, as do these guidelines, that the ISSN should be printed in a prominent position on or in each issue of a serial (front cover, title page, masthead, etc.). On a periodical the preferred location is in the top right hand corner of the front cover. It should not be obscured by other printed matter and the type size should be sufficiently large to make it distinctive. It is also recommended that ISSN are used in promotional material.

36.5 The ISSN of a series title or of a serial published annually or less frequently should appear at least on the back of the title page of each part of the series or of each issue of the annual publication. It should be printed together with the ISBN of each part or issue and the two standard numbers should be distinguished by their own prefixes (see figure 6).

36.6 ISSN can be acquired in the United Kingdom from the ISSN UK Centre in the British Library. An ISSN should be requested before publication of the first issue of a new serial. ISSN can be assigned to an existing serial and on occasion publishers may be asked to use an ISSN assigned by the ISSN UK Centre even though not requested by the publisher. There is no charge for the assignment.

36.7 The ISSN assigned to a particular title should *never* be used for any other serial title. If it is proposed to change a title, no matter how apparently minor the change may be, or to merge or split existing titles, the ISSN UK Centre should be informed.

36.8 Different regional and language editions of a serial each require their own ISSN.

36.9 The ISSN UK Centre will ask for a copy of the first issue of a serial or some other acceptable documentary evidence (such as a photocopy of the cover, title page, masthead, etc.) to be sent as proof that the serial has been published. Sending a first issue of a new serial or of a first issue under a new title should not be confused with the publisher's legal obligation to deposit a copy of each issue as required by the Copyright Act 1911, as amended by The British Library Act 1972, and by the Irish Copyright Act 1963 (see section 41).

36.10 The assignment of an ISSN and the registration of a title with ISDS does not confer copyright in that title.

36.11 Enquiries about ISSN should be addressed to:

ISSN UK Centre
The British Library
Boston Spa
Wetherby
West Yorkshire LS23 7BQ

Tel: 0937 54 6959/6958
Fax: 0937 54 6979

37. CODEN

37.1 While ISSN is recommended as the unique identifier to be used on serials there are other identifiers the most familiar of which is CODEN. It is available for use in all subject fields and for books as well as serials. Its use is most common, however, in the United States in the fields of science, technology and medicine. It is emphasised, however, that ISSN is the preferred identifier and if other identifiers, such as CODEN, are used they should be used in *addition* to and not *instead* of ISSN.

37.2 To obtain a CODEN for a publication, contact:

International CODEN Service
Chemical Abstracts Service
2540 Olentangy River Road
P O Box 3012
Columbus, Ohio 43210, USA

38. Issue and article identifiers

38.1 There are several systems for identifying specific issues of serials and particular articles within a serial. They are at different stages of development. Some are associated with bar codes.

SISAC

38.2 In the United States the Serials Industry Systems Advisory Committee (SISAC) was formed specifically to deal with issues of developing standards applicable to the needs of electronic data

transmission systems for serials. One of its objectives was to develop a standardized code for unique identification of serial issues and articles.

38.3 The SISAC code identifies serials at the issue and article level. Starting with the ISSN, the code adds information about the date, volume and issue, page number, and serial identification code or title, and a check character. The resulting numerical citations are displayed in bar code format and in eye-readable form. The issue identifier is displayed on the front cover of the issue and the article identifier on the first page of the article.

38.4 SISAC recognizes that in order for the code to be used effectively it must be widely used. For the publishing industry the code will speed up the processing of orders. The use of the article level code should promote the direct ordering of article reprints through article ordering services. As new technologies develop to take advantage of the SISAC code the bar code could be scanned at the copying stage in order to credit royalties automatically to publishers' accounts. There is considerable potential for use of the code in library circulation and serial check-in and claiming systems.

38.5 Instructions for constructing the issue and article identifiers in machine-readable bar code form are given in *Serial Item and Contribution Identification: Code and Symbol Guidelines* obtainable from: Book Industry Study Group, 160 Fifth Avenue, New York, NY 10010.

38.6 The American National Standards Institute (ANSI) approved the Serial Item and Contribution Identifier (SICI) in July 1991 as ANSI/NISO Z39.56-1991. A number of international journal publishers now print the bar code symbol on all their journals and for commercial reasons, if for no other, it is advisable to use this code in academic and scholarly journals in the United Kingdom, particularly those which have or expect to have, a market in the United States.

38.7 The Faxon Company has been designated as the maintenance agency for the standard and any enquiry concerning SICI should be addressed to:

The Faxon Company
15 Southwest Park
Westwood
MA 02090, USA
Tel: 617-329-3350
Fax: 617-461-1862

ISSN/EAN Bar Code

38.8 The ISSN/EAN bar code will be familiar from periodicals sold through retail outlets. The European Article Numbering Association (EAN) covers Europe and the rest of the world except North America. The Periodicals Barcoding Association is the body responsible for the administration and development of the system and publishes a manual and explanatory publicity for periodical publishers, technical specifications for film-master manufacturers for the production of ISSN/EAN bar codes, etc. Further information is available from:

Periodicals Barcoding Association
Imperial House
15/19 Kingsway
London
WC2B 6UN

Tel: 071 379 6268
Fax: 071 379 5561

38.9 Publishers intending to use this bar code on issues of their serial publications must first acquire an ISSN for each title.

38.10 In the UK the principal reason for the introduction of EAN bar coding was for the benefit of the retail trade, especially those outlets where goods other than periodicals and books are sold and which goods also carry EAN bar codes, such as supermarket chains, retail groups, etc. The benefits to publishers are faster and more accurate sales data, indications of market trends, stock and distribution control and, of course, re-ordering. Wholesalers gain information on sales situations in individual outlets. Retailers gain a faster response to market trends and stock control to maximise potential sales.

38.11 There is little evidence of ISSN/EAN bar codes being used by publishers of academic, scholarly and specialist journals of the kind not normally sold in the mass retail market, that is the kind which is normally sold largely by subscription and would otherwise only be sold through a limited number of specialist outlets. It seems unlikely, in their present state of development and for commercial reasons, that these bar codes will be used by publishers of academic and scholarly journals.

ADONIS number

38.12 ADONIS is a document delivery service that supplies the contents of selected biomedical journals from a CD-ROM database. Each article is identified by a unique 16 digit number (which incorporates the journal's ISSN). The usefulness of the ADONIS number is being compared with the SISAC number and the International Organization for Standardization's Biblid system. If the ADONIS number proves useful, it should be possible to ask journal publishers to print it on each article, etc., and for abstracting services to include it so that it is available for document ordering. Its usefulness, however, is confined to journals participating in the ADONIS system.

Biblid

38.13 The international standard ISO 9115-1987: *Documentation - Bibliographic Identification (Biblid) of Contributions in Serials and Books* is intended to facilitate the identification of contributions in serial publications, and contributions in books containing separate works by different authors, by means of a standard code. It is designed for computer and manual applications and permits transcription into machine readable systems.

38.14 While in many respects similar to the SISAC code there are differences which need to be resolved. SISAC will probably predominate, although it is claimed that the SISAC serial issue and article identifier is compatible with Biblid; but Biblid, so far, has had very limited acceptance by publishers.

39. International Standard Book Number (ISBN)

39.1 An ISBN is a 10 figure number which identifies a particular edition of a work issued by a specific publisher and is unique to that edition. It is assigned either by the publisher (with the authority of the Standard Book Numbering Agency) or directly by the Agency. Its relevance to serials is that it is customary to assign ISBN to issues of serials published annually or less frequently (such as directories, yearbooks, annual reports, etc.), to the individual works within a series and to supplements and special issues of journals which are marketed and sold separately outside the journal's usual subscription. It is correct procedure in these instances to have an ISSN assigned to the *title* of the serial and to have an ISBN assigned to the individual issue or part. Information about ISBN can be obtained from:

Standard Book Numbering Agency
12 Dyott Street
London WC1A 1DF

Tel: 071 836 8911
Fax: 071 836 2909

39.2 Information on ISBN based bar codes is obtainable from:

Book Industry Communications
39/41 North Road
London N7 9DP

Tel: 071 607 0021
Fax: 071 607 0415

40. Cataloguing in Publication

40.1 In the UK the Cataloguing in Publication (CIP) Programme is run by the British Library's National Bibliographic Service to provide libraries with details of publications in advance of publication so that they are aware that a basic and timely record is available from the British Library's machine-readable tape service for use in their catalogues. J Whitaker and Sons Limited are solely responsible for creating CIP records for all titles on behalf of the British Library. Other national libraries, notably the Library of Congress in the United States, also have CIP programmes.

40.2 This provision of advance information is also a positive sales tool for publishers. The service is free of charge and the information appears in the relevant printed, microform, CD-ROM and online products of both organizations. These include the *British National Bibliography*, BLAISE-LINE, *BNB on CD-ROM, BOOKBANK CD-ROM, The Bookseller*, and *Whitaker's Books in Print - on microfiche.*

40.3 It is unusual for serials to carry CIP information, either in the form of a CIP Print Block or as an acknowledgement that a record is available from the British Library. Serials, however, are *not* excluded from the programme. There is no reason why a new serial title (or a changed title) should not have CIP information. While uncommon in journals or periodicals they are used in serials pub-

(5)

1 Purpose

To describe the use and structure of a standard code for the unique identification of serial publication items.

2 Introduction

A publication item is any of the autonomous items that together form the content of a publication. Examples of publication item types are: full article, abstract, editorial, publishers note. For a more detailed description of publication item see Handbook PAS, chapters 41.11.01 and 41.11.05.

To be able to identify a publication item for retrieval or other purposes, it has been decided to introduce the serial publication item identifier SSDI: Standard Serial Document Identifier. The SSDI will be printed on the first page of each publication item.

3 Structure of SSDI

Format: xxxx-xxxxyyddddc (17 characters)

where:

* xxxx-xxxx: ISSN of serial publication to which the publication item has been primarily assigned; ISSN is the identifier of the journal; it includes the hyphen between the code parts and also the check digit. In some cases the ISSN of the journal in which the publication item is published may differ from the one in the SSDI.
* yy: year of SSDI allocation for publication item.
* ddddd: sequential number for publication item within serial identifier and within year.
* c: check digit for the code formed by the concatenation of yy and ddddd (parenthesis not included). The check digit is calculated with a Modulo 11 check digit algorithm.

4 Presentation of SSDI

The SSDI will be presented in print in full format, including hyphens and parentheses, and will be preceded by the text "SSDI". An example SSDI will be printed as:

SSDI 0001-5806(94)00001-8

5 Check digit algorithm

The check digit is calculated with a Modulo 11 check digit algorithm over the last 7 digits using weights 8 to 2, i.e. the first digit gets weight 8, the second 7, etc. This algorithm is the same that is used for calculating the check digits in ISSN.

6 Example

The first article in 1994 in the journal with ISSN 0001-5806 has as its last 7 digits in its SSDI without check digit:

'9' '4' '0' '0' '0' '0' '1'.

They have weights

8 7 6 5 4 3 2

Multiplication gives

72 28 0 0 0 0 2

with sum 102. Division by 11 gives rest value 3. Substract the remainder from 11:

11 - 3 = 8. The SSDI will therefore be: 0001-5806(94)00001-8

If the remainder is 0 then the check digit will be 0. If the remainder is 10 the check digit will be 'X' (capital).

Notes on SICI

(taken from draft on standard Z39.56-199x, published March 1991)

1 Goals of SICI

SICI: Serials Item and Contribution Identifier.

Goals of SICI are:

(1) To limit the standard to a code for unique identification of serial items and contributions.

(2) To cover the broadest possible range of serials, for example, scholarly, trade and popular, and both domestic and foreign, in both print and nonprint media.

(3) To allow creation of the SICI code from either a citation or the serial itself, regardless of whether the serial is currently published and regardless of whether the publisher has printed the identifier on the serial.

(4) To provide the briefest possible code consistent with unique identification.

(5) To maintain consistency with other NISO standards.

2 Area of use

The SICI has two main areas of use:

(1) identification of a serial item: volume/issue;

(2) identification of a publication item (ES term) or article.

3 Global structure

(1) Serial identification:

(2) Serial item identification (volume/issue):

(3) Serial contribution identification (article):

Part 1 identifies a journal, part 1+2 together an issue of a journal and part 1+2+3 identifies an article within a journal volume/issue.

4 More detailed structure

In this section only the main line is discussed. More details can be found in the description of the standard.

The structure of SICI is as follows:

(1) Serial identification:

(a) ISSN

Example of a serial identification:

8756-2324

(2) Serial item identification (volume/issue):

(a) Chronology

Identifies a specific date. i.e. the cover date of an item. Format (YYYYMMDD): YYYY is year, MM is month an DD is day.

Example:

May 1994= (199405)

June 12, 1994= (19940612)

Exceptions:

Months:

21= Spring

22= Summer

23= Fall

24= Winter

31= 1st quarter

32= 2nd quarter

33= 3rd quarter

34= 4th quarter

Combined chronology is possible. May/June 1994 = (199405/06)

(b) Enumeration

Identifies a certain item of a serial. First and subsequent order designators are used: volume, issue etc. Example volume 5, issue 3= 5:3 (vol/issue information separated by colon). Combined numbering, alternatives schemes are also possible.

(c) Standard version number

Is used to identify the version of the SICI standard. Will be omitted when this item identifier part is used for a serial contribution identifier.

(d) Check digit

Modulo 37 is used. Omitted when part of a serial contribution identifier.

Example of a serial item identification:

8756-2324(19860305)65:2;1-2 = volume 65, issue 2 published 5th March 1986 in journal 8756-2324. Note: standard version number 1 is used; the check digit is not correct in this example.

(3) Serial contribution identification (article):

(a) Location number

Character L followed by a period and the location of the initial contribution (article) as it appears in the contribution. Page number is used for that purpose.

Example: L.27 for an article starting at page 27.

(b) Title code (optional)

Must be used if no location number (page number) is available or if more than one article starts on the same page. Only the first letter of the first four words with more than three characters are used for the title code.

(c) Standard version number

(d) Check digit

Modulo 37 is used.

Examples of article identifiers:

0277-0288(198606)6:6L.4:CP;1-3 = article in journal with ISSN is 0277-0288, cover date June 1986, volume 6, issue 6, starting at page 4, title "The Compaq Portable".

0185-125X(19860612)8:4L.12;1-3 = article in journal with ISSN is 0277-0288, cover date 12 June 1986, volume 8, issue 4, starting on page 12.

G. Roza

8 Sep 94

(D)

INTERNATIONAL CODEN DIRECTORY

INTRODUCTION

This is the 1990 Supplement to the *International CODEN Directory* 1989. Included are all new CODEN assigned, CODEN deleted, and changes to existing CODEN made during the period 13 December 1989 to 12 December 1990.

The *International CODEN Directory* is a single microfiche file, the content of which was derived from a number of different data bases. The Directory brings together publication citations for all CODEN published by the American Society for Testing and Materials through 1974[1], as well as those assigned by the International CODEN Service since 1975. The Directory is issued every five years. A cumulative annual supplement is issued in each of the intervening years.

The purpose of the Directory and its supplements is to provide to editors, publishers, librarians, abstracting and indexing services, and others engaged in information activities a comprehensive source of CODEN for the titles of both serial and nonserial publications.

The *International CODEN Directory* contains three sections:

CODEN Index
Title Index
KWOC Index

The **CODEN Index** alphabetically lists all CODEN assigned to serials and then alphanumerically lists all CODEN assigned to nonserials. Each entry in this index consists of the CODEN, the publication title associated with it, and any references to known predecessor and/or successor CODEN and titles. Also listed are CODEN that have been deleted from active use.

The primary use of this index is to identify a publication's title when only a CODEN is known. It also can be used as a source of CODEN check-characters for users whose files contain only five-character CODEN. In addition, this index may be used to determine alternate forms of a publication's title since these alternate forms will all have the same CODEN. The preferred title is listed first in a CODEN entry.

The **Title Index** provides a single, alphabetic listing of the titles of both serials and nonserials to which CODEN have been assigned. Each entry in this index includes a title and its associated CODEN as well as references to predecessor and successor titles and title translations. When a publication has more than one known version of its title, e.g., when the title appears in more than one language or in more than one format, an entry for each version is included in this section.

The alphabetic order of the titles is affected by Arabic numerals and punctuation. After alphabetic characters, the numeric, and then punctuation characters are ordered. In a few hyphenated words, the word following the hyphen may occur with or without capitalization. The lowercase version alphabetizes before the uppercase version. The following example illustrates the filing sequence used in the Directory:

Byproducts
Byzxum
By-laws
By-products
B.A.
B.-bulletin
B-flokkur
B'lgarski
Caaso

The Title Index can be used to determine the CODEN of a publication when only the title, or some alternate form of the title, is known. It also can be used as a source of information about predecessor and successor titles and their CODEN and whether translated versions of publications exist.

The **KWOC (Keyword-Out-of-Context) Index** provides an alphabetical listing of all significant words contained in the titles to which CODEN have been assigned. All forms of each title are indexed, i.e., the preferred title as well as variant titles. Below each keyword are the titles containing that keyword. These titles are in alphabetical order according to the characters and punctuation of the title.

Creation of the KWOC Index includes suppression of nonsignificant words. Included in this group are articles, prepositions, and conjunctions, as well as such generic words as journal, proceedings, bulletin, report, and their non-English language equivalents. A copy of the list of suppressed words is included as an appendix to this Introduction. All words of three or fewer characters are suppressed. However, any title which is composed completely of suppressed words is included in the KWOC Index, listed under the first word of the title. The result is occasional listings of titles under such keywords as journal, oil, etc. Very long keywords that duplicate the first part of a previously listed keyword, but which differ in their endings, may not be listed. However, the titles containing these long keywords will be listed in the proper sequence.

The KWOC Index is used to locate a title and its assigned CODEN when the precise wording of the title is unknown, or the exact word order is uncertain. Because of differences in title selection rules, several entries in the KWOC Index should be searched for a given title before deciding that the given title does not have an assigned CODEN. The large file of suppressed keywords also makes this approach advisable.

The KWOC Index is also a subject index to the publications for which CODEN have been assigned. Once the CODEN is known for any publication, use of the CODEN Index will show if there are any predecessor or successor titles or translated titles.

For all three indexes, each microfiche contains 325 frames arranged in 13 horizontal rows and 25 vertical columns. Each frame contains two columns of information. The eye-readable heading strip contains information identifying the particular index, the number of fiche within the index, and the characters of the first index entry on the first frame of the fiche.

SCOPE OF THE INTERNATIONAL CODEN DIRECTORY

When CODEN were first introduced in the early 1950's by Dr. Charles Bishop,[2] their assignments were limited to scientific, technical, and medical journals. After the American Society for Testing and Materials (ASTM) assumed responsibility for CODEN in 1961, the scope of assignment was broadened first to include the full range of science and

technology, and later was expanded to include all subject areas. Neither Dr. Bishop nor the ASTM restricted CODEN assignments on the basis of language or period in history during which the publication was issued.

CODEN which contain only letters (A-Z) in the first five positions have been assigned to serial publications. CODEN for patents also have only letters in the first five positions, but the third and fourth positions are always the letter "X".

CODEN assigned to nonserial publications have numbers (0-9) in each of the first two positions. Generally, the publications to which these CODEN have been assigned are volumes of papers emanating from conferences, symposia, institutes, lectures, etc. CODEN beginning with numbers also are used for monographs, most of which are edited collections of individually authored papers.

INFORMATION ON ENTRIES

Publication Titles

Most publication titles found in the *International CODEN Directory* reflect the order of the words contained in the title as it appears on the publication. However, because in the early history of CODEN assignment some CODEN were assigned to serial titles listed in bibliographies, the form of the title found in the Directory may be that used in these bibliographies. Where this condition occurs, several forms of the title may be included in the directory, each with an identical CODEN.

Since many publications carry more than one title, the alternate titles also have been included in the directory, again, each with identical CODEN. However, serials which change titles and subsequently change back to a former title are given different CODEN. In order to distinguish between these otherwise identical titles, the beginning date for each title is added as qualifying information. In cases where a serial has been issued under different titles throughout its publication history, the title changes are noted and cross-references are provided to the CODEN of the predecessor and successor titles. Entries for serials which have translated editions include cross-references to the translations.

Entries for nonserial publications that emanate from meetings, conferences, symposia, etc. include the meeting title, the date and location of the meeting, the name of the editor or compiler, and the name of the publisher of the volume or, in the case of certain USSR publications, the statement "USSR [Non-Sub]," to indicate that the publication is available only through bookstores specializing in Soviet literature.

Entries for nonserial publications that are edited collections of individually authored papers contain the title of the edited collection, its date of publication, and the name of its editor or compiler.

Capitalization

Principal words in the title, subtitle, and other components of the citation are capitalized.

Diacritical Marks and Symbols

All modified letters, including those with diacritical marks of any kind, are printed as their unmodified equivalents with the following exceptions. In German, Danish, Swedish, Norwegian, Finnish, and Icelandic, the letters ä, ö, ø, and ü are transliterated as ae, oe, oe, and ue, respectively.

The symbol & (ampersand) appearing in titles and publisher names is not available in the computer-output-microform character set, and has thus been translated to 'amp'.

Romanization Systems

All entries are printed in the Roman alphabet. Titles that appear on the original publication in the following non-Roman alphabets or ideographs are romanized according to the systems listed below:

Alphabet	*Romanization System*
Chinese	For Chinese language titles from the People's Republic of China (PRC), the Pinyin system[2] is used with cross references using the Wade-Giles system.[4]
	For Chinese language titles from other countries (e.g., Hong Kong, Singapore, and Taiwan) the Wade-Giles system.
Japanese	The modified Hepburn system.[5]
Korean	The modified McCune-Reischauer Romanization scheme developed by the Korean Committee of the Association for Asian Studies.[6]
Cyrillic	The International *System for the Transliteration of Slavic Cyrillic Characters*[7] (modified by the omission of diacritical marks) is used for modern Russian, Bulgarian, Macedonian, Serbo-Croation, Ukrainian, and Belorussian (White Russian).

Translations

Cover-to-cover translations of a publication are assigned their own CODEN and a cross reference is provided to the original publication. A cross reference is also provided at the CODEN of the original publication to alert the user that such a translation exists.

Acronyms, Initials, and Abbreviations

When publications have acronyms, initials, or abbreviations included in the title, an expansion of these shortened word forms is provided, when known, and an additional entry created for the expanded form of the title.

Deleted CODEN

Occasionally, a single title will be identified to which more than one CODEN has been assigned. An example of such an entry in the CODEN Index would appear as:

AAEBAP Alabama, Agricultural Experiment Station. Bulletin

AUXBAM Auburn University, Agricultural Experiment Station. Bulletin

A single title is allowed only one CODEN. Therefore, a preferred CODEN is selected, and the title is deleted from the unpreferred CODEN. The "deleted" CODEN is retained in the CODEN Index so that the user is aware that a CODEN is no longer valid, *and* a cross-reference is added to the incorrect CODEN so that the user can locate the preferred CODEN for a title. An example of the corrected CODEN Index entries from the above example is as follows:

AAEBAP Alabama, Agricultural Experiment Station. Bulletin

AAEBAP Auburn University, Agricultural Experiment Station. Bulletin

AUXBAM Deleted see AAEBAP

Deleted CODEN are listed only in the CODEN Index.

STRUCTURE OF CODEN

CODEN are unique, unambiguous, six character alphabetic and alphanumeric codes assigned to the titles of serial and nonserial publications. The sixth character of each CODEN is the check character. CODEN are used to represent the titles of publications in information processing and library systems that require the unique and unambiguous identification of nonserials and serials to be accomplished through the use of a highly compact code. CODEN as a title identification system are based on the American Society for Testing and Materials *Standard Recommended Practice for Use of CODEN.*[3]

CODEN for Serials

CODEN for serial titles, insofar as it is possible, consist of four alphabetic characters structured into a mnemonic form derived from words within the title. To facilitate the continued use of the mnemonic form, a fifth character, called the grid character, has been added. Use of the grid character theoretically permits up to 26 repetitions of a four-character mnemonic code. For example:

JACSAT Journal of the American Chemical Society

JACSBU Journal of Applied Chemistry, Supplementary Issue

The sixth character of a CODEN is a computer-calculated check-character, the derivation of which is described below.

CODEN for Nonserials

CODEN for nonserial titles consist of two digits followed by two alphabetics, the grid character, and the check character. Such CODEN are assigned sequentially without regard for mnemonics (i.e., 11AAAS, 11ABAV . . . 11ZZAO, 12AAAZ . . .).

CODEN for Patents

The letters XX in the third and fourth positions of a CODEN indicate that the publication is a patent. Only one CODEN is assigned to a patent series.

CODEN Check-Character

The sixth character in each CODEN is an alphabetic or numeric check-character which is added to permit computer verification of the preceding five characters. The check-character, when properly calculated, will detect errors made during the transcription of CODEN.

The initial suggestion for the use of a computer-generated check-character for controlling errors in CODEN came from A. D. Pratt at the School of Library Science, Indiana University, Bloomington. F. E. Hajjar is responsible for reducing the idea to practice for Chemical Abstracts Service. The value of the check character depends upon the value and order of the first five characters of the CODEN. Any error in recording a single character or an inversion of characters will produce an inconsistant check-character and will expose the error.

A check-character is generated as follows:

(a) Each alphabetic character of the CODEN is assigned a value based on its position in the alphabet (A = 1 ... Z = 26); the numeric characters are assigned 27 through 36 (1 = 27 ... 0 = 36).

(b) The equation used to generate the check character is

$$\frac{(11 \times N_1) + (7 \times N_2) + (5 \times N_3) + (3 \times N_4) + (1 \times N_5)}{34} = Q + \frac{X}{34}$$

where N_1, N_2, etc. are the values assigned to the CODEN characters in order of their appearance in the CODEN and Q is a whole number that is discarded.

(c) The "X" value (or "remainder") is converted to a check character by assigning it a character as follows:

X =	1,	2,	. . .	25,	26,	27,	28,	29,	. . .	33,	34
check character =	A,	B,	. . .	Y,	Z,	2,	3,	4,	. . .	8,	9

The numeric check-characters one (1) and zero (0) have been eliminated to avoid confusion with the alphabetic characters I and O.

Sample calculation of a check character for the CODEN JACSA:

N_1 = J = 10
N_2 = A = 01
N_3 = C = 03
N_4 = S = 19
N_5 = A = 01

Substituting these numeric values for the characters in the CODEN into the equation yields a "Q" of 5 and an "X" of 20. The check-character equivalent to 20 is T. Thus the complete CODEN with check-character is JACSAT.

ASSIGNMENT OF CODEN

The *International CODEN Directory* contains entries for all CODEN assigned up to the date of its generation from the International CODEN Service's CODEN data base. Users who are unable to find CODEN they need in this Directory, or who want CODEN assignments made, should contact the International CODEN Service. Depending on the country of residence, requests for CODEN or CODEN assignments should be directed to the following organizations:

Austria, the Federal Republic of Germany, and Switzerland

VCH Verlagsgesellschaft mbH
CA-Vertrieb
Postfach 10 11 61
W-6940 Weinheim
Federal Republic of Germany

Japan

Japan Association for International Chemical Information (JAICI)
Gakkai Center Bldg.
2-4-16 Yayoi
Bunkyo-ku, Tokyo 113
Japan

United States and other countries not named above

International CODEN Service
c/o Chemical Abstracts Service
P.O. Box 3012
Columbus, Ohio 43210
U.S.A.

Telephone: (614) 447-3600 Ext. 3249
TWX: 810-482-1608
Cable: CHEMABS
FAX: (614) 447-3648

No special forms are required. Requests may be transmitted by mail, TWX, FAX, or telephone. Before any request is made, the requester should be certain that the desired

CODEN has not been listed within the latest edition of the Directory or its latest cumulative annual supplement. Requests for CODEN not found in the Directory must include the complete title of the publication and the name of the city and country where it is published.

Mailed requests for CODEN assignment must be accompanied by photocopies of the cover and the title page of the publication. Mailed requests for CODEN assignment for serial publications should also be accompanied by a photocopy of the masthead. Permanent CODEN will be assigned and the requester will be notified as promptly as possible.

Requests for CODEN assignment that are transmitted by TWX or telephone will be fulfilled with *provisional* CODEN. Requesters must then provide *proof* that the publication assigned a provisional CODEN exists by supplying photocopies of the cover, title page, and masthead (for serials). If such proof is not supplied within thirty days, the provisional CODEN will be cancelled.

The International CODEN Service also will, upon request, retrieve CODEN previously assigned to publications. A fee ($5.00) is charged for each CODEN supplied which has already been assigned and published in the Directory or its supplements. No charge is made for assigning CODEN to titles not published in the Directory or its supplements.

REFERENCES

[1] *CODEN for Periodical Titles*. Philadelphia: American Society for Testing and Materials. ASTM Data Series DS23B: 1970: 2v. ASTM Data Series DS23B S1: 1972: 1v. ASTM Data Series DS23B S1: 1974: 1v.

[2] Bishop, Charles. An Integrated Approach to the Documentation Problem. *American Documentation*. 1953: 4: 54–56.

[3] *Pinyin Chinese–English Dictionary*. Beijing: Commercial Press: 1979. 976 p.

[4] Mathews, Robert H. *Mathews' Chinese–English Dictionary*. Rev. American ed. Cambridge: Harvard Univ. Press: 1943 (1966 printing). 1226 p.

[5] Masuda, Koh, ed. *Kenkyusha's New Japanese–English Dictionary*. 4th ed. Tokyo, Japan: Kenkyusha: c1974. 2110p.

[6] *Transactions of the Korean Branch of the Royal Asiatic Society*. 1961. 38: 119–128.

[7] International Organization for Standardization. *International System for the Transliteration of Slavic Cyrillic Characters*. 1968: 3p. (ISO/R9–1968(E)).

[8] American Society for Testing and Materials. *Standard Recommended Practice for Use of CODEN*. Philadelphia: ASTM: 1976: ANSI/ASTM D950–76. 4p.

an essential mechanism for sifting among extremely large collections of documents.

Whether it is WAIS or some other index tool, we will probably reach a point where the normal mode of Internet navigation involves consulting an index as the first step. It is possible to navigate through the collection in a used bookstore or in a small public library by browsing the shelves. It would be unthinkable to try to browse the shelves of a large research library with 3 million volumes. Instead, one always consults the catalog when performing any sort of serious search. An online catalog allows a user to identify a list of candidate documents, chosen based on an understanding of the content of each document. Whether such a tool comes from the Archie developers, the WAIS developers, a consortium of libraries, or some other community, we will need such a tool to support serious, purposeful searching of an Internet with millions of documents and other resources.

A STANDARD BOOK NUMBER FOR THE INTERNET

This book, like all other books from established publishers, has an International Standard Book Number, or ISBN. Periodicals are similarly assigned International Standard Serial Numbers (ISSNs). These are standard "handles" one can use in ordering books, in searching library catalogs, etc.

The Internet needs similar standard names for documents and online resources. The Uniform Resource Locator created for use in the World-Wide Web is a start along these lines. It is a standard way to name a particular document on a particular server.

The Internet Engineering Task Force is working to establish a general standard for a Uniform Resource Identifier that would build upon the URL concept. Under the URI umbrella one would also find Uniform Resource Names, which are roughly analogous to ISBNs. An ISBN identifies a particular title, but does not tell you where you will find copies of the title. For that information you have to interrogate your local library or book store. In the Internet context, the URI engineers envision a global, distributed directory service, analogous to the Domain Name System, that will provide you (or your client program) with a list of URLs that correspond to a given URI. In other words, you tell the service you are looking for a copy of Moby Dick, and the service gives you a list of places where you might find it. Your client, being intelligent, picks a place that is accessible over a short-haul, uncongested link, and you have gained access to your chosen resource.

This scenario sounds sensible in theory, but the devil is in the details of designing a working system. One issue is how URNs will be assigned, and what constitutes a work whose content is "equivalent" to another version. For instance, if a publisher offers an image file in TIFF format, and an archive site offers a copy in JPG format, are the two documents equivalent, therefore meriting the same URN? Who will make this decision? One school of thought

argues that only the publisher can answer that question—but what is the mechanism by which the answer will be promulgated?

The problem of equivalence extends to any document that can take multiple forms. Should the flat ASCII version of a monograph have the same URN as the PostScript version? Certainly not, say some: the PostScript version may contain graphics and formatting that make the two inherently different. Certainly so, say others: the content of the text is what is important.

A similar thorny question is one of how to handle time-specific titles. One would want to be able to say to a client, "Get me the current weather forecast" or "Fetch me a copy of today's *Washington Post.*" If an archive site has not yet received the "current" version of the document you seek, you may get stale data. In any event, the URN for the "current" version of a document will necessarily point to different URLs over time, complicating the delivery process.

Although the task force working on these issues faces daunting questions, their work will no doubt eventually bear fruit. A workable URI scheme will have many advantages:

- Indexing of documents will be enhanced.
- It will be easier for users to include citations in documents they write, without fear that particular copies of the cited documents will disappear. Just as a paper can refer to *Time* magazine by volume and number, without concern as to whether it is archived at a particular library, a future online paper can refer to a particular number of an online journal, without the author worrying about which archive site holds a copy.
- It will be easier to develop technology to support caching of documents, promoting more efficient use of Internet communications links. (See discussion of caching later in this chapter.)

DYNAMIC DOCUMENTS

Print documents are inherently static: A great deal of work may go into document preparation, but once the document is printed, the moving finger has writ. Most online documents are similarly static: An author prepares the text, and it remains as written until a subsequent update phase.

Online document delivery offers the possibility for documents to be updated "on the fly" using automated processes. This could be as simple as a weather server inserting the current local temperature in the text of an explanation of how to use the server. Ever-more elaborate schemes can be envisioned, whereby HTML becomes a dynamic rather than a static medium. For example, researchers at the University of Minnesota offer an interactive geometry service. (See Fig. 25.2.) This server, for instance, allows users to ask for 3-D rendering to be redrawn from the perspective of the user's choice.

Another example of delivery of dynamic information via the Web is offered at Stanford University. A group there has set up a virtual environment that tests

(F)

Members of the Internet Engineering Task Force have devised a scheme called IAFA (for "Internet Assigned Fields Authority") which may provide an answer for how resources are identified by Internet information providers for the sake of automated catalogers like Archie. An Internet resource provider who wants to have his or her resource cataloged would fill out an IAFA template and place the information online. This fits exactly with items 3 and 4 of Dillon's model. In cases were a title merits the labor-intensive effort of human cataloging, the IAFA records could serve as a starting point. Here is the beginning of the proposed IAFA template for documents:

```
Template-Type:          (any one of DOCUMENT, IMAGE or SOUND)
Category:
Title:
Author-Name:
Author-Organization-Name:
Author-Organization-Type:
Author-Work-Phone:
Author-Work-Fax:
Author-Work-Postal:
Author-Job-Title:
Author-Department:
Author-Email:
Author-Handle:
Author-Home-Phone:
Author-Home-Postal:
Author-Home-Fax:
Record-Last-Modified-Date:
Record-Last-Modified-Department:
Record-Last-Modified-Email:
```

Another IETF effort, the work to define standard Uniform Resource Identifiers, may yield both a scheme that fosters cataloging of Internet resources, as well as a standard mechanism to allow access to those resources. One goal of the URI effort is to define a Uniform Resource Name—roughly analogous to an International Standard Book Number—that could be "resolved" into a particular Uniform Resource Locator. As an analogy, consider how a customer might walk into a bookstore armed with the ISBN for a book on theoretical physics. A clerk looks up the book on the store's inventory computer, determines that the book is in stock, and helps the customer fetch the desired title from the shelf. Similarly, Internet users may someday be able to submit Uniform Resource Names to an automated service that locates copies of the work in question, and returns a list of Uniform Resource Locators—i.e., specific point-

Appendix 5

The ABCs of Information Commerce

Mapping the Future of Information Commerce, January 1995, p 3

Northeast Consulting Resources, Inc

The ABCs of Information Commerce

Excerpts From The Latest Endstates

A Superdistribution: Information distribution has become a business of high volume, low cost transactions. Metering technologies incorporated in desktop computers allow users to pay only for information used. The ubiquitous World Wide Web plus metering and content-control technologies allow digital information in cryptograhic envelopes to be duplicated easily and sent simultaneously to many people. This has created new business models in which the down stream redistribution of information, the "multiplier effect," is as important as the initial distribution. Custom publishing now has been widely adopted in business and other information-intensive markets. Some trade publishers are experimenting with print-on-demand technologies (locating printers in super bookstores) in order to reduce distribution and inventory costs. Authors, agents, repackagers, and direct (e)mail marketers have gained at the expense of traditional publishers and on-line services.

B InfoCommerce Fast Forward: Computing and consumer electronics have converged. Inexpensive hardware, ubiquitous wireless, and the MTV generation's immersion in consumer electronics changes information access modes. Cheap, but often incompatible, advanced computing and telecommunications technologies are widely used; home entertainment centers consisting of television, digital audio, and powerful computers are networked together using cable and telco access to the several information superhighways. Advertising is focused on microsegments. Shared on-line information, education, and entertainment environments or "spaces" attract large audiences. On-line virtual reality video games with multiple players have become the rage. Inexpensive hand-held computers connected to information sources via wireless networks have made "just-in-time" information a reality for nomadic users.

C Global Network Utopia: Global information utilities provide low-cost ubiquitous network access to megadata information and entertainment warehouses on the emerging Global Information Infrastructure. Advanced multimedia workstations and very fast networks enable new info- and edutainment products that integrate voice, data, and full motion. full screen video. Consumers demand significantly more graphical, audio and interactive content in education, information, and entertainment products. Electronic newspapers and magazines where readers interact with editors and authors are the rage. The highly visible entertainment production houses who early on acquired or developed the best talent (e.g., Disney, LucasFilm) now have a perceived quality advantage over traditional print publishers. Small, low-cost "multimedia production boutiques" flourish, too.

D An Interactive Archipelago: High speed CD-ROM and multimedia standards are widely adopted together with hardware-based metering to protect digital property. Publishers make significant investments in CD-ROM text and multimedia products. CD-ROM multimedia games, magazines and catalog shopping have gained widespread acceptance. Proliferation of computers in elementary schools and the widespread use of flat-panel hand-helds have led to new interactive reference, education, and entertainment products. Trade publishing remains almost entirely a business of blockbuster novels and "how to" books. The "information superhighway" grew slowly; the Internet remains unreliable.

E Knowledge Refining: Information commerce has become an industry of giants with deep pockets. Knowledge workers demand instant answers to complex questions. The latest innovation is knowledge refineries: sophisticated systems that provide answers more focused than those of traditional index, search, and document delivery systems. To meet market demand, on-line systems have been deployed that qualitatively change the interaction between users and the knowledge base. New visual metaphors (like the "information landscape") make navigating the sea of information easier. Improved ease of use has also enabled significant penetration of the home and education markets by information providers. Tenants of on-line information malls use powerful new interface customization tools to differentiate their infostore from others.

Appendix 6

Preliminary Functional Specification for a Uniform Data Identification (UDID) system

Douglas Armati, March 1994

Appendix 7

Universal Data Identification (UDID) Tests

Douglas Armati, December 1994

Douglas Armati
7a Angel Lane
Woodbridge, Suffolk IP12 4NG England
Phone: +44 1394 380874
Fax: +44 1728 453909
email: armati@udid.u-net.com

Douglas McL. ARMATI

Mr Douglas Armati is an independent researcher, communicator, company director and consultant.

He has been promoting the financial and practical involvement of industry in a programme to specify, develop and implement a uniform global system for the identification of valuable data.

He developed the functional specification for a proposed universal data identification (UDID) system. A practical prototype of the core of this system was designed in collaboration with a team of digital communications specialists at the Imperial College in London.

Mr Armati has been involved in the World Intellectual Property Organization (WIPO) working sessions regarding the establishment of a voluntary international numbering system for certain categories of copyright works and management and protection of copyright works in the digital domain.

He recently completed a report on Information Identification for the STM Task Force on Information Identification and Metering Systems in the Electronic Environment. STM is the international association of scientific technical and medical publishers, affiliated to the International Publishers Association. It represents all major stm publishers.

During the early part of his career Mr Armati worked in Australia with media companies RCA, Paramount Pictures, United Telecasters and News Limited. He was an early advocate of interactive broadband cable and satellite distribution of valuable information, founding Access International in 1975.

For the past fifteen years Mr Armati has been an advisor to senior management in the private and public sectors. This work has focused increasingly on maximising the value of intellectual property assets.

Mr Armati holds a bachelor of Commerce degree with high distinction from Murdoch University, Perth, Western Australia. He is the author of several papers on intellectual property issues, including: "A Uniform System for the Identification of Digitized Copyright Content", STM Newsletter 95, November 1994 and "Copyright: The Fragile Estate", Murdoch University School of Law, 1991.