NAME NUMBERING
FOR THE UK BOOK INDUSTRY

**A study conducted by
Book Industry Communication
on behalf of the
British National Bibliography Research Fund**

David Martin and Liz Potter

**Library and Information Commission Research Report 45
British National Bibliographic Research Fund Report 98**

April 2000

**BOOK INDUSTRY
COMMUNICATION**

CONTENTS

ACKNOWLEDGEMENTS

1. BACKGROUND

1.1 WHY NAME NUMBERING?

In brief, name numbering is seen as the only way in which a computer system can match two personal or corporate names with complete certainty.

In the book trade as it is, name numbering has three main areas of application:

(a)   Personal or corporate names as authors: is this book *by* the same Jack Smith as that one?

(b)   Personal or corporate names as subjects: find all the available books *about* this particular Jack Smith.

(c)   Personal or corporate names as publishers and imprints: who publishes this book?

The fourth, and eventually the most important, area of application:

(d)   Personal or corporate names as parties to an electronic rights transaction

is not yet sufficiently widely recognised to have become a general priority, but it is on the horizon.

1.2 ORIGINS, SCOPE AND OBJECTIVES OF THE STUDY

The origins of this study lie in two separate initiatives from bodies within or linked to the UK book trade. In late 1997, in the context of the development of daily EDI updating from publishers' distribution centres to the trade bibliographic agencies, Book Data proposed to a meeting of the Book Industry Communication (BIC)[1] EDI Implementation Clinic that it would be helpful to establish a standard coded identification for publishers and their imprints so that product information updates sent as EDI messages, particular from distributors handling the output of many different publishers, could be correctly assigned to their "owners". The BIC Secretariat was already aware of the developments in name numbering which were taking place internationally among the copyright societies, and suggested that these might provide a way forward.

Early in 1998, Mark Bide & Associates, acting as consultants to the Publishers Licensing Society (PLS), drew attention to the potential benefits for new systems being introduced at the UK Copyright Licensing Agency (CLA) if both authors and publishers could be uniquely numbered using the IPI scheme (see Section 1.4.1 and Section 6).

Combining both of these initiatives, BIC and the Authors Licensing & Collecting Society (ALCS) organised an ad-hoc meeting in July 1998, bringing together representatives of the bibliographic agencies, the copyright societies and The British Library (the last-named in view of their central role in the management of library name authority systems). Its purpose was both to brief all parties on the IPI system, and to discuss the UK book sector's requirements and determine what actions might be taken to progress them.

The proposal to the British National Bibliography Research Fund (BNBRF) for the present study was a consequence of this meeting.

Meanwhile the **<in*d*ecs>** Project[2] was being planned, and support was successfully sought from the European Union. **<in*d*ecs>** is concerned with developing a schema for the interoperability of data to support electronic commerce in intellectual property in a global network environment. Within the

---

[1] We have tried to follow the principle that each organisational acronym used in this report is spelt out in full on its first appearance, but not thereafter. There is also a full list of acronyms in *Appendix A*.

[2] See *http://www.indecs.org*

overall schema, an essential element is the unambiguous identification of personal and corporate names and their relationships to natural and legal persons who may hold rights and/or participate in transactions. Because of a substantial overlap between the bodies and individuals involved in **<in*d*ecs>** and those who were to work on or be consulted for the purposes of this study, it was agreed that the two pieces of work should proceed as far as possible collaboratively, and that the BNBRF study should be seen partly as providing detailed input to **<in*d*ecs>** in relation to the special needs of the UK book trade. This was particularly appropriate and timely since the music industry, a major participant in **<in*d*ecs>**, was already relatively well advanced in understanding the issues relating to the control of rightsholders' names; the book sector is much less so. **<in*d*ecs>** recent thinking and proposals for further work on name numbering will be found in *Directory of Parties: outline specification* by Mark Bide.[3]

The original aims of the BIC/BNBRF study could be summarised as being to answer the following questions:

(a)   Might it be feasible to adopt the IPI system as a single common standard across the book industry, without compromising its primary purpose, ie accounting for rights usage?

(b)   If it appeared to be feasible and beneficial to adopt the IPI system as a common standard, what steps would be needed in order to introduce it across the book industry?

(c)   If it did not appear to be feasible to adopt the IPI system, but the requirement for a standard name numbering system has been validated, what other approaches should be explored?

As is often the case, one of the results of the study has been that some of these questions as originally phrased no longer seem entirely appropriate. It is probably wrong to expect that the IPI system on its own could or should be the complete solution to a wider set of requirements. It would certainly be wrong to contemplate setting up a common book trade name numbering system which was not integrated with the systems used in the relevant copyright societies.

1.3 TERMINOLOGY USED IN THIS REPORT

To avoid tedious repetition, we will use the following terms as defined below, except where the context makes it clear that they are to be interpreted more specifically:

Author: any contributor to a published work (author, editor, illustrator, etc).

Publisher: a publisher or publisher's distributor.

Bookseller: a book wholesaler, retailer, or specialist library bookseller.

Personal names: names which by their form appear to identify real, fictional or pseudonymous natural persons.

Corporate names: names which by their form appear to identify real, fictional or pseudonymous corporate bodies or groups of natural persons.

1.4 EXISTING AND PROPOSED NAME IDENTIFICATION SYSTEMS

In this section we review a number of existing and proposed systems for the identification of natural persons and/or corporate bodies in contexts which are relevant to the book trade. Some of them may be properly regarded as identifiers of trading locations rather than of the entities which operate through those locations; but this is a distinction which is sometimes likely to be blurred.

---

[3] *http://www.indecs.org/results/persons.htm*

1.4.1 The CISAC IPI system

(Information on the CISAC[4] Interested Party Information (IPI) system contained within this report has been derived from unpublished documentation kindly made available by permission of SUISA, the coordinating agency for the system.)

The IPI system is an extension of the CAE (Compositeur-Auteur-Editeur) name numbering system which is used by a substantial group of major national intellectual property licensing societies. The new international system is being rolled out in 1999, and will gradually be adopted by the participating societies as their own internal system developments permit.

"The purpose of the IPI system is the global unique identification of a natural person or a legal entity acting across multiple creation classes, roles and rights."

The essential feature of the IPI system is that it distinguishes between an IP Base Number and an IP Name Number. An IP Base Number identifies an underlying entity. An IP Name Number identifies a name by which the entity is or has been known, or in which the entity participates together with others.

In the UK, ALCS is committed to using the IPI system, and already uses its predecessor, the CAE number, for natural persons who hold authorship rights in literary works. ALCS is an assignment agency in this particular context.

1.4.2 The IPDA IPN system

The International Performers Database Association (IPDA), linked to SCAPR[5], has recently developed its own International Performer's Numbering system, managed from Sweden. Its scope includes actors, conductors, dancers, musicians and singers. The IPN is an eight digit "dumb" number designed to be used for the administration of performers' rights in any country which has a member association, and for reporting across national boundaries. Since many performers are also writers or composers, this means that the same person may be identified by different numbers depending on her/his role in connection with a particular work and performance - possibly by two different numbers in connection with a single recorded performance, say.

The IPN system is a single-level system - that is, it does not separately number the underlying identity and the names which the performer uses - but it does record multiple pseudonyms when so required.

1.4.3 Library name authority systems

The library community at local and national level devotes considerable effort and expenditure to the maintenance and use of name authority systems for both personal and corporate names as authors or subjects of published works, but has not extended those systems to embrace the names of publishers and imprints. There have been significant developments at international level towards the online sharing of name authority databases across the Internet, but these are still far short of any full implementation.

There are two key differences between library name authorities and the licensing societies' name numbering schemes: first, library name authorities do not (yet) use any visible ID number, but aim to establish an "authorised form" of each name, which is then used as an entry point to a library catalogue; and second, library name authority systems are not (yet) interested in whether the names which they cover are or are not rightsholders, or potential rightsholders.

---

[4] CISAC: the International Confederation of Authors and Composers Societies

[5] The Societies' Council for the Administration of Performers' Rights

1.4.4 The ISBN system

The International Standard Book Number (ISBN) includes a segment (the "ISBN prefix") which in principle identifies the publisher. When the ISBN system was first introduced, it was perceived as something which would identify both producer and product, and this aspiration is still sometimes reflected both in the attitudes of some of the ISBN agencies and in the assumptions which have been made in systems design for book-trade-related computer applications.

In reality, the ISBN prefix is a fairly safe indicator of the publisher responsible for the item at the date when the ISBN was first assigned (which is usually several months and may be several years before publication). It is certainly not a safe indicator of the current publisher, given the number and frequency of the occasions when individual titles or whole lists are sold by one publisher to another. It is even less likely to be of any help in identifying a publisher's imprints. Typically, a publisher will use the same ISBN prefix across most or all imprints, and may or may not choose to use some locally-defined internal structure to distinguish them.

For these reasons, the ISBN prefix cannot be considered as a candidate to be used for the identification of publishers, let alone of their imprints or trading names.

1.4.5 EAN Location Numbers

EAN International and its affiliated organisations in most countries of the world administer a system by which a 13-digit number can be assigned to represent a trading location, particularly for use in EDI and barcoded labels. For example, an EAN Location Number might represent a branch delivery address or a network mailbox to which EDI invoices must be sent. These numbers have been adopted for the identification of book trade locations in EDI communication in the UK and in some other countries. They are not regarded as a candidate for a more generalised name numbering system.

1.4.6 Book Trade Standard Address Numbers

In the UK and North America, separate but parallel systems of Standard Address Numbers (SANs) for trading locations in or connected to the book trade have been in existence for some time. In Germany, the Börsenverein administers a national book trade location number known as the Verkehrsnummer.

The North American SAN system is widely used and extends to all types of entity involved in the book trade as suppliers or customers. The UK SAN system has been applied comprehensively to booksellers, but not to most other trading entities, and is, in principle if not yet entirely in practice, being replaced by EAN location numbers for EDI purposes. In each case, the SAN is a six-digit number.

Like the EAN Location Number, the book trade SAN is not regarded as a candidate name number.

1.4.7 Proposed International Standard Identifier for Libraries

A draft ISO standard[6] which would establish a format for an International Standard Identifier for Libraries (ISIL) is out for ballot at the time of writing. The ISIL has two components: an ISO 3166 two-letter country code, and a variable-length code of up to ten characters which identifies a library in accordance with conventions defined and administered within the country concerned.

While the draft recognises that there are many reasons why a standard identifier for a library may be useful, a prime objective is to facilitate access to library catalogues and holdings information. The

---

[6] ISO/DIS 15511:1999(E) *Information and documentation - International Standard Identifier for Libraries and Related Organisations (ISIL)*

preamble states that "Any library or related organisation, any library administrative unit, service point or subordinate unit, acting autonomously, which is engaged in international library or book trade communication should be assigned an ISIL." It is envisaged that "agencies doing business with libraries, such as suppliers and publishers" would use the ISIL to identify their library customers.

Apart from the country code, any meaning or hierarchy within the code is left open for national code assignment agencies, with a recommendation, however, that newly assigned identifiers should use ISO 3166-2 geographic subdivision codes. At the same time, it is expected that existing identifiers which do not use these codes will be incorporated into the ISIL as they stand. Globally, therefore, the code would have to be treated as unstructured.

For the purposes of this study, the ISIL is seen as being "more like" a trading location identifier (using trading in the widest possible sense) than a name identification system, but this is a case where the distinction is certainly unclear.

2. BOOK TRADE

In Section 2 we summarise the present position within the UK book trade, and the levels of awareness of and interest in the potential application of name numbering which we found in a series of interviews with relevant companies or organisations.

2.1 PUBLISHERS

None of the publishers consulted as part of this study had either a fully effective internal system for identification of personal or corporate names as authors or subjects, or any means of communicating such names in a controlled manner to the outside world (in particular, as part of the metadata describing their products).

The one area of the business where person identification had been systematically addressed was in royalty calculation and distribution. Royalty systems are likely to use an internal number as a unique identification of either an author or an author/agent combination to whom royalties for particular products must be attributed. Some of the complexities which were reported in this area were:

(a)   A requirement to distinguish between the same author under the same name but using different agents for different genres of work, or simply for different works.

(b)   A requirement to link different pseudonyms to the same author.

(c)   A requirement to link a single pseudonym to two or more authors writing collaboratively under an identity which appears to be a single person.

Present royalty systems are not linked to any other internal systems which carry authors' names.

There was general support for a trade-wide system of name numbering: very enthusiastic support in some cases, and no more than lukewarm in others. Those who were less enthusiastic seemed to be influenced as much by scepticism that such a system could be made to work as by any doubt that it would in principle be a good thing.

It was clear that any name numbering system which was to be effective for publishers would have to be:

(a)   International: at the very least to the extent that from the outset it must cover the UK and the USA, as the two dominant English-language publishing industries.

(b)   Immediate: in the sense that registration of new names must be handled quickly, as and when required.

2.2 TRADE BIBLIOGRAPHIC AGENCIES

Consultation with Whitaker and Book Data (see Section 3.2 below for Bibliographic Data Services) indicated quite strong support tempered with concern about practical issues. Whitaker maintains the UK Standard Book Numbering Agency, whose staff initially questioned the need to introduce another identifier for publishers alongside the ISBN prefix, but that view would not be shared by Whitaker Bibliographic Services and TeleOrdering, who are very much aware of the problems outlined in Section 1.4.4.

Whitaker uses a system of personal and corporate name control based on but not always precisely following the British Library Name Authority system (Section 3.1), delivering controlled name forms rather than name numbers. Name numbering would definitely be beneficial. Publisher and imprint names are separately controlled against a database of some 70,000 names, of which about half have at least one title in print.

Book Data has no formal author name authority control yet, but is working towards its introduction. A standard name number would be supported provided the industry as a whole was willing to adopt it. Publishers and imprints on the Book Data database are controlled by an internal numbering system which links titles to imprints and imprints to publishers. In principle, it would be helpful to have a "public" standard, provided that publishers could apply the codes consistently at imprint level.

2.3 BOOKSELLERS

The Booksellers Association, as the compilers and publishers of printed and electronic directories of publishers and distributors, would support an industry standard name number. High street and Internet booksellers were not specifically consulted, since it was felt that the benefits which they would see from name numbering would largely be visible as greater consistency in access to and presentation of product data by name. This was perhaps a gap in the study which should be filled in the first stages of any subsequent work.

2.4 SUMMARY: BOOK TRADE

During the period of this study, there has been continuing pressure from some BIC members for early implementation of a UK name numbering system for publishers and imprints. The need originally advanced in 1998 for a scheme which could be used in communications between publishers, distributors, bibliographic agencies and others is still strongly felt. It would be untrue to say that there is any equivalent pressure for the general numbering of persons and corporate bodies as authors or subjects.

It would also be untrue to say that the book trade is yet widely conscious of the need to prepare itself for electronic rights trading, so that the issues which are driving the **<in*d*ecs>** project's proposals on name numbering were barely mentioned by interviewees. There would probably have been a higher level of awareness among STM publishers.

3. LIBRARIES

The use of name authority control is a long-established part of library practice, with two fundamental aims:

(a) To "collocate" within the library catalogue all items which share a common authorship, personal or corporate, and

(b) To distinguish within the library catalogue between items which carry superficially identical authorship, but which are in fact by different persons.

Thus the catalogue user should be able to find all items by author X, without any ambiguity as to whether the list includes items by another author with the same name.

This has traditionally been achieved, and continues to be achieved, by establishing an "authorised form" of each name known to the catalogue. Names are not numbered, except to the extent that there may be an internal system control number associated with each name authority record; but this number is not used in bibliographic records as a surrogate for the name.

Library name authority control is *not* concerned with issues of ownership or rights: its purpose is to give users the best possible access to library collections, and nothing else.

It is also characteristic of library name authority systems that they do not usually have separate (linked) records for each name by which an entity is known, as does the IPI system. Instead, they have a single record carrying both the authorised name and any other names; and where there is a complex pattern of pseudonyms or variant names, it is likely to be explained by a notes field requiring human interpretation rather than by a set of formal links which are suitable for machine processing.

3.1 THE BRITISH LIBRARY

The British Library is both the national library and the operator of the UK national bibliographic service (NBS), providing among other things authoritative catalogue records for other libraries at home and overseas.

The British Library maintains a Name Authority List (BLNAL) which therefore serves two purposes. First, and perhaps primarily, it must support the catalogues of the Library's own collections, providing consistent access to personal and corporate names in those catalogues. Secondly, by publishing its name authorities in microfiche and computer-readable form, the British Library NBS provides a resource which is available to be used by other UK libraries to help maintain their own local authority control.

In recent years there has been increasingly close cooperation between the British Library and the Library of Congress (LC), and this has been especially fruitful in the area of name authority control. In 1994 the British Library joined the Name Authorities Cooperative (NACO), a group of major libraries which all contribute to the development of the US Name Authority File at LC.   From that beginning has emerged a new Anglo-American Authority File (AAAF) which should in the long term replace BLNAL. It has already largely eliminated duplication of work between LC and the British Library on the establishment of new personal name authorities.

The cost of maintaining name authorities, and the importance which is attached to them, can be seen in the fact that the British Library has a staff of some sixteen cataloguers dedicated exclusively to this task.

Statistically, the total British Library authority file contains some 750,000[7] entries, of which 63,000 form the new AAAF.

The table below summarises the data elements which are used in the combined BL/Library of Congress name authority file:

Data elements used in BLNAL / AAAF name authority records	
Record identifier	
LC Control Number	In records derived from the Library of Congress
Cataloguing source	The source from which the authority was created
Cataloguing application	The BL collection for which the authority was created
Personal name	Authorised form of personal name. May include: (a) Dates (b) Additions to names other than dates (c) Both initials and spelled out form of name, when the person is customarily known by his/her initials
Corporate name	Authorised form of corporate name.
Personal name "see" reference	Non-used form of a personal name used to generate a "see" reference to the authorised form.
Corporate name "see" reference	Non-used form of a corporate name used to generate a "see" reference to the authorised form.
Personal name related heading tracing	Other authorised personal name, used to generate a "see also" reference.
Corporate name related heading tracing	Other authorised corporate name, used to generate a "see also" reference.
Title and source of data found note	May be either the title of the work in which the name was first encountered, or a reference source used to establish the name.
Data not found note	Details of sources checked without success.

3.2 LIBRARY BIBLIOGRAPHIC AGENCIES

Apart from the National Bibliographic Service, the only UK bibliographic agency specialising in library-quality data is Bibliographic Data Services Limited (BDS), of Dumfries.

For personal and corporate names as authors or subjects, BDS uses the British Library Name Authority File to select an authorised form where it exists, or, for new names, enters the name in accordance with AACR2 principles, but flagged as "not authorised".

For imprint and publisher names, interestingly, BDS use an internal authority file, in which the preferred names have been created according to AACR2. The authority file is accessed by ISBN prefix. Thus, for example, when a cataloguer enters 013 in their input system, s/he is presented with all the names which are or have been associated with that number, in this case Prentice Hall, Random House, Egmont Children's Books, etc.   The correct name is then selected according to what is on the book or in the publisher's advance information.

---

[7] Figures supplied by Andrew MacEwan of the British Library, 5 March 1999.

BDS saw themselves clearly as part of the library community, and wedded to the British Library name authority system. One of its weaknesses, however, was that a new name was not properly established until the book-in-hand was catalogued, which could be long after publication. It was suggested that there would be particular benefits for the library community if a book trade name control system included a procedure by which the publisher would supply information allowing an authorised form of a new name to be established in advance of publication.

3.3 INDIVIDUAL LIBRARIES AND LIBRARY NETWORKS

In principle, name authority control in an individual library is a local matter, and each library's catalogue embodies its own choices as to the preferred forms of names, made often over a considerable period of time. In practice, there is an increasing likelihood that authorised names supplied by the National Bibliographic Service will be used provided there is no conflict with a previously-established local form; but there is no overall consistency from one library to another. This is being recognised increasingly as a problem when users try to search the catalogues of a number of libraries over the Internet, particularly through Z39.50 client/server links.

Talis Information Ltd (formerly BLCMP), as library system suppliers and managers of a central database resource for customer libraries, does not maintain a collective name authority system on its central database. Each individual customer system applies its own authority control to personal and corporate names as authors and subjects, using facilities in the Talis software which are modelled substantially on the US Library of Congress authority system. Consequently, there is no assurance of consistency between libraries. Characteristically of library systems, authority control is not extended across publishers and imprints. Talis foresaw that a cross-industry name numbering system could reduce the very high cost of authority control in libraries and increase consistency at the local level; but it would be important that all major trade and library bibliographic record suppliers should adopt the system.

LASER (London And South Eastern Region, the regional library interlending network operator) places very strong emphasis on authority control in its database, and, while basing its work on the BL Name Authority system, adds its own headings where necessary, eg for foreign authors. LASER strongly supported the idea of an overall name numbering system, but felt that it would be a massive task to set it up, and stressed that it must be done in collaboration with the British Library. In the absence of any control of publisher and imprint names in current library databases, these were a "nightmare", and might be the most beneficial place to start with name numbering.

3.4 INTERNATIONAL DEVELOPMENTS

Library name authority control has so far been practised at local library level and in national libraries and national bibliographic services. IFLA, the International Federation of Library Associations and Institutions, has long pursued a goal of Universal Bibliographic Control, within which it was seen as a fundamental premise that each country should be responsible for establishing name authorities for its own national authors, and that these should be made open to, and should be used by, other countries.

More recently it has been realised that this premise was seriously flawed: there are in fact very good reasons - of language, alphabets and culture - which make it both proper and inevitable that different countries should use different forms of name for the same person.

At the same time, the international library community has become increasingly interested in the sharing of name authority databases, partly because it is felt that much effort and cost could be saved if libraries in different countries could access and use the name authorities established by their colleagues elsewhere, and partly because online name searching across different catalogues on the Internet cannot be fully effective unless different forms of the same name can be automatically or semi-automatically linked.

IFLA recently published a report[8], under the auspices of its Universal Bibliographic Control and International MARC (UBCIM) programme, on the minimum content which might be required for an internationally shared name authority record. (In fact, the report covers authority records for names of natural persons, for names of corporate bodies and conferences, and for titles.)

A fundamental assumption behind the report is that sooner or later different authority records for the same entity will be linked by a shared ISADN (International Standard Authority Data Number). Each country or region would then be able to use its own preferred form of any name, with an automatic link to the preferred form of the same name in another country or region. But to date there has been little sign of progress towards the development of the proposed ISADN, and there has been considerable concern that the task would be prohibitively expensive. Meanwhile, links must remain less than wholly automatic, but there would still be great benefit from shared online access to authority files.

In parallel with this work, a group of national libraries and associated bodies in Europe has reported on Project AUTHOR[9], a research study funded by the European Union as part of the CoBRA programme under DGXIII. The aim of this project was to study the feasibility of transnational exchange of name authority files for personal and corporate names, and to set up a demonstrator system using tools developed under other related EU projects.

Project AUTHOR has not created a working system, but it has established a set of principles, wholly consistent with the IFLA UBCIM report, for the provision of a practical and useful level of shared access to authority files, short of their interconnection by a common numbering scheme.

3.5 SUMMARY: LIBRARIES

Libraries at many levels attach great importance and devote considerable effort and expenditure to name authority control. The approach is often "we do as much as we possibly can" rather than "we have to achieve this standard, and it costs us x in order to do so". In spite of this, the level of control which is achieved is incomplete and inconsistent.

It was noted that publishers often hold the kind of information about authors which would make the task of matching a name to an existing authority or of creating a new entry easier and more reliable. Such information may be included in their CIP data submission, but it is not usually to hand when the British Library Name Authority section processes the book.

Libraries would benefit substantially if publishers and bibliographic agencies adopted a reliable name numbering system which was linked to national and eventually international Name Authority schemes. Failing that, there would also be benefits in ensuring that information which publishers hold which would help to "disambiguate" author names was made available as a matter of course when the book was processed through the national Name Authority system.

The library sector also holds by far the largest existing UK database of standardised names.

---

[8] *Mandatory data elements for internationally shared resource authority records: Report of the IFLA UBCIM Working Group on Minimal Level Authority Records and ISADN,* Barbara B Tillett et al, IFLA Universal Bibliographic Control and International MARC Programme, 1998

[9] *Project AUTHOR Final Report,* Sonia Zillhardt / Francoise Bourdon, 1998

Table of IFLA UBCIM data elements

From the IFLA UBCIM report cited on the preceding page, the table below has been extracted to show the proposed set of mandatory or highly recommended data elements (data elements shaded at the foot of the table are recommended; others are mandatory):

IFLA UBCIM data elements for shared authority records	
Record status	New, modified, deleted
Type of record	Authority record
Entity category	Name of natural person, name of corporate body etc
Encoding level	Minimum level record, or more complete
Record ID	Unique record identifier in the source system
ISADN	International Standard Authority Data Number *This number should be capable of being carried as an element in any other book trade authority database.*
Date entered on file	
Version ID	Date and time of latest update
Language	Code(s) for the language(s) for which the authority record is valid
Character sets	Character set(s) used in the authority record
Script	Code(s) for the script(s) used in the authority record
Descriptive rules	Code for the descriptive cataloguing rules used
Source library/agency	Code for the library or bibliographic agency responsible
Differentiated/ undifferentiated	If the record is for the name of a natural person, a code to show whether the name is differentiated by the inclusion of added data, eg a date of birth.
Authorised heading	The preferred form of the name, in full, in plain text, including added elements of differentiation where applicable.
Nationality of the entity	Indication of the nationality(ies) of the person or corporate body.
Variant forms of the name	Text of each variant form for "see" references.
Related names	Text of other authorised names related to the entity described.
Source citation(s)	Citations for consulted sources, or the item catalogued, which provided information used to establish the authorised name.
Biographical, historical or other information	
Sources consulted - no data	Citations for sources which were consulted but in which the name was not found.
General notes	

4. LICENSING SOCIETIES

The two UK licensing societies which are primarily concerned with book industry copyright management are the Authors Licensing and Collecting Society (ALCS) and the Publishers Licensing Society (PLS). Their data collection and much of their processing operations are carried out by the Copyright Licensing Agency (CLA), which is jointly owned by the two societies.

The ALCS maintains a substantial database of its membership and of other non-member author names. As a participant in the CISAC system it uses and contributes to the IPI name number database at SUISA (at the time of the study, ALCS was in fact preparing to make the transition from the established CAE numbers to the new IP name number structure). While "corporate authors" can sometimes claim royalties through ALCS, they cannot be members of the society. Consequently ALCS does not ordinarily enter any UK corporate names into the database, and even a limited inspection shows that while it has wide coverage of UK personal names from both the book and music sectors, there are very few UK book trade corporate names - and those which do appear are usually there because they are also music publishers.

Unlike the ALCS, the PLS "subcontracts" its publisher database maintenance to the CLA, where the PLS member database can be regarded as a subset of a wider publisher file which CLA maintains as a necessary part of its operations. PLS has not joined the IPI name number system.

The adoption of a standard publisher and imprint numbering scheme in the UK book trade could provide substantial benefits to the CLA and PLS, for whom the tracking of ownership of book titles after their publication is a major problem. Realising those benefits would depend on the efficiency with which publishers updated their own databases and communicated changes electronically through the bibliographic agencies or direct to CLA.

The adoption by publishers and bibliographic agencies of standard name numbering for authors could bring slightly different, but not insubstantial, benefits to ALCS if those numbers were readily and consistently congruent with IP name numbers, so that usage data collected by CLA could be matched directly to the ALCS database.

5. CRITERIA FOR A FULL NAME NUMBERING SYSTEM

We conclude from the discussions summarised in preceding sections that a comprehensive name numbering system for the UK book sector as a whole should have (at least) the following characteristics:

(a)   The system must be *international*, covering the major English-speaking countries as a minimum, and allowing names from all parts of the world to be registered and numbered when they are cited in connection with works published in the UK. To be useful for the UK book trade, a name ID system must not be limited to the UK book trade.

(b)   The system must be *responsive*, enabling any legitimate user to obtain and use any existing name ID when needed, and to register a new name and obtain at least a provisional name ID when needed. The trade will not use a numbering system unless numbers can be provided promptly when the need arises. In practice, this implies an online interactive enquiry and registration process, provided that reasonable criteria - eg the provision of a minimal set of data elements for a new name registration - are met.

(c)   The system should cover names of natural persons and corporate bodies *in a single numbering scheme*. It would not be impossible to treat personal and corporate names separately, and there are some significant differences in the way that they may need to be handled. However, both the IPI system and library name authority systems have chosen to treat personal and corporate names within a single scheme. There is one clear benefit from doing so, from the point of view of name matching and identification: many corporate names are in fact also personal names, and the user may not always know *a priori* whether the "John Smith" in question is, say, the late leader of the Labour Party or the Glasgow bookselling company.

(d)   The system must cover, and the database must be coded to distinguish between, *a number of primary types of names*. These would include, but are not limited to:

   Real names identifying real natural persons, eg *Charles Dodgson*

   Pseudonyms identifying real natural persons, eg *Lewis Carroll*

   Names identifying groups of real natural persons, eg *The Beatles*

   Names identifying fictitious natural persons, eg *Sherlock Holmes*

   Registered names identifying real corporate bodies, eg *Penguin Books Ltd*

   Names used by real corporate bodies, such as brand or trading names, eg *Puffin*

(e)   The name number database must provide excellent support for *disambiguation*, ie the differentiation of identical names which belong to different persons. This requires that it should be able to carry additional data such as dates of birth and death, titles of works for which the person is known to have had some specified responsibility, or statements about other known activities.

(f)   The database must be highly *reliable*, but, recognising that no database is ever without some incidence of error, it is also important that it should record the authority on which key data was entered, both to give an indication of likely reliability and to enable the administrators to go back to a source in order to resolve conflicting input or correct an error.

(g)   The database structure and its administration must support an appropriate level of *confidentiality* of data, particularly about living persons, at the same time as allowing wide access to those parts of the content which are needed for effective use of the scheme.

6. THE CISAC IPI SYSTEM

Much of the detail of the CISAC IPI system relates to the management of licensing society membership details, rights information, and the collection and distribution of revenues based on usage in different countries. This is completely outside the scope of the present study.

In this section we focus on the IP Base Number and Name Number, and the database structure and administrative processes which are being or have been set up to support them, keeping in mind the criteria set out in Section 5.

6.1 DATABASE CONTENT

6.1.1 IP Base Number

An IP Base Number identifies a natural person or a real legal entity. The number has a fixed length of 13 characters: I-999999999-9 (the last digit being a check character). There is understood to be no meaningful structure in the body of the number.

An IP Base Number is accompanied by certain Base Data:

Interested Party Type	Mandatory: N = natural person, L = legal entity.
Sex	Optional, if IP Type = N: F = female, M = male.
Nationality	Optional. Multiple nationalities may be held, with start and end dates where an individual has changed nationality.
Date of birth or foundation	Optional.
Date of death or dissolution	Optional.
Place of birth or foundation	Optional: country, state/province/department, city/town.

An IP Base Number record does not include a name, but it must have at least one linked IP Name Number record, of type PA, showing the current primary name for the person or legal entity identified by the Base Number.

6.1.2 IP Name Number

An IP Name Number identifies a name which is associated with a natural person or a legal entity. The number has a fixed length of 11 digits: 99999999999 (the last two digits being check characters). There is understood to be no meaningful structure in the body of the number, and no necessary link between the structure of a Name Number and its associated Base Number.

An IP Name Number is accompanied by a name and a name type code, whose values depend on whether the name refers to a natural person or a legal entity.

IP Name Number details for natural persons and for legal entities are shown in the two tables on the next page.

Person name format conventions include: no full stops after initials; no titles; prefixes such as "VAN" and "DE" to be entered as part of surnames, not as first names; and identical names may have distinguishing suffixes (examples given are JUN, SEN, though JR, SR would be more normal). No special format conventions are specified for names of legal entities, beyond the recommendation that they should be based on legal documents such as a companies register.

IP Name Number detail for natural persons:

Name	Mandatory: surname(s) up to a maximum of 45 characters
First name(s)	Optional, up to a maximum of 45 characters
Name type	Mandatory PA = patronym: the primary current name by which the person is known PP = pseudonym (pseudonyms which the owner wishes to keep secret are apparently treated as separate patronyms) PG = pseudonym group: a pseudonym used collectively by a number of persons, eg the name of a pop group MO = modification: a former patronym changed eg by marriage (not automatically treated as a pseudonym - a new PP must be created if so desired) DF = different spelling: entered by a participating agency ST = standardised spelling: a variant form entered by the IPI centre to facilitate retrieval

IP Name Number detail for legal entities:

Name	Mandatory: up to a maximum of 90 characters
Name type	Mandatory PA = patronym: the primary current name by which the legal entity is known OR = other reference: eg a publisher's imprint or a record company label HR = holding reference: a group to which the legal entity belongs but which itself is not a registered rights owner MO = modification: a former patronym changed eg by change of company name DF = different spelling: entered by a participating agency ST = standardised spelling: a variant form entered by the IPI centre to facilitate retrieval

6.1.3 General

The IPI database character set is limited to basic upper case ASCII, with special characters and diacritics transliterated.

6.2 ADMINISTRATION

The administrative model for the IPI system is based on a single international coordinating centre (SUISA) which holds the IPI database, and performs a number of crucial management functions.

Participants in the IPI system (the copyright societies) can use the database either online, or on periodically issued CD-ROMs, or on a locally-held copy maintained by regular distribution of updates in a proprietary EDI format. They can contribute updates to the database either by direct online entry, or by uploads in the same proprietary EDI format, or as paper listings. Input provided on paper may take some time to process at the IPI centre.

Direct online entry by a participating society is restricted to data on "its own members". It appears, however, that other forms of data entry may include details which affect other societies, in which case the other society is informed and must respond within 30 days if it does not accept the proposed addition or amendment.

SUISA, as the IPI centre, is responsible for maintaining certain elements of the database, notably the added names of type ST which systematically provide additional look-up points for certain specified forms of name.

The costs of the coordinating centre are in principle covered by contributions from the participating societies, but understandably there is no information on the expected cost levels for the new system.

SUISA is required to follow "guidelines and standards ... established by CISAC". The formal ownership of the database was not clearly established during the present study.

6.3 COMMENTS

The capacity of the IPI Base Number and Name Number formats (allowing up to 999,999,999 unique numbers) is probably not a practical limitation in the foreseeable future, but it would be important that a global scheme should be readily capable of expansion.

(b)   The limitation of names to a maximum of 90 characters, the very basic structure of personal names, and the restricted (upper case ASCII) character set would all appear inadequate to support a realistic book trade name identification system, let alone a global system operating across "multiple creation classes, roles and rights", and of course across all languages.

(c)   The database design quite properly reflects the fact that the IPI System is first and foremost a switching system between an international family of IP rights licensing societies. In most respects, this in itself does not mean that the database is unsuitable to support a more generalised role, but there are some details which reflect a society-centred view - for example, a group name identified by name type "HR" is defined as referring to an entity incapable of concluding "membership agreements with a copyright society", but in a more general system there might be every reason to treat such an entity as a legal person in its own right, with its own IP Base Number.

(d)   The administrative operation also reflects a society-centred view, and it is not easy to see how it could support a requirement for (say) immediate registration, on request from a publisher in London, of the name of a new author from an overseas country whose work is about to be published in the UK in translation, and who is not currently listed on the database.

(e)   The database design does not appear to support any data content for the disambiguation of person names apart from dates of birth and death, and place of birth. Other things which might be required include titles of key works associated with the person name, and statements of the profession or field of endeavour or type of creation with which the person is connected.

(f)   The design also appears not to allow dates to be associated with a name so as to show the period during which the name was in use.

(g)   The IP Base Number is not associated with any name at all. This is not illogical - far from it, it makes perfect sense, assuming that the "patronym" is for all practical purposes the key name with which the Base Data are associated. However, while this approach will work for a database which is used only by a "closed universe" of copyright societies, each honouring its obligations of confidentiality towards its membership, it may be inadequate for a database which must be partly open as well as partly confidential. For example, it may be argued that there is never any reason to allow a corporate body to conceal its base identity - its incorporation *ipso facto* places it in the public domain. But there may be equally compelling arguments for allowing a private person to keep her/his name and other personal details private while also flaunting a public persona as a pseudonymous author or pop star. One way of doing this would be to extend the content of Base Data and define this as a confidential domain for information about natural persons.

(h)   Again no doubt because the IPI system has been designed for a "closed universe", there is no evident provision for allowing a person with two or more pseudonyms to determine whether their common identity is to be made open or concealed.

While the original development of the IPI system was driven by the needs of music licensing, it is now well-established as an essential part of the infrastructure for print authors' royalty collection and distribution in the UK. This surely means that it is inconceivable that any new development could be undertaken without the participation of the IPI system or without a means of linking to IP Name Numbers. However, it appears equally inconceivable that the IPI system in its present form could meet all the perceived needs of the book sector.

7. PROPOSALS FOR A NAME NUMBERING DATABASE DESIGN

7.1 OUTLINE

It is clearly outside the scope of this study to try to specify the detailed design of a database to support a name numbering system. However, the work has stimulated a number of ideas which it seems worthwhile to include in the report.

In this section we will outline the following components:

(a) A numbering scheme which, very much like the IPI System, has what we will call Person Numbers and Name Numbers.

(b) A Person Data Record for a natural person, which, to maintain privacy, is open only to qualified users.

(c) A Person Data Record for a corporate body, which, following the principle that a corporate body has no rights of privacy as to its basic identity, is open to any user.

(d) A Key Name Record, for a name which is distinctively different from any other Key Names for the same natural person or corporate body, ie it is not a variant of another Key Name. Different Key Names for the same natural person may be linked openly by a direct link between the Key Names, or they may be linked confidentially through the Person Data Record.

(e) A Variant Name Record, for a name form which is derived from a Key Name by one or more of the following types of derivation: transliteration from one alphabet to another, variant spelling, *others to be defined.* Links between Variant Names and the Key Name form which they are derived are always open.

Note that at the Person Data level, there is a clear and absolute distinction between natural persons and corporate bodies. At the Name Data level, there is no such absolute distinction: a name may apply to a person, a loose grouping of persons, or a formally incorporated body, or indeed to more than one: for example, the name under which an author writes may be linked both to a record describing him/her as a natural person and to a record describing a company to which he/she assigns copyright and through which royalties are administered.

7.2 NUMBERING SCHEME

Person Numbers and Name Numbers should be expandable to meet any future requirement. They should of course carry check digit validation. They should be purely numeric, in an international system, and they should carry no meaning which would require unnecessary number changes as a result of changes in associated data, or which would disclose links which might sometimes need to be kept confidential.

If there was advantage in doing so, Variant Names could be linked to their associated Key Name by the numbering structure, but this is probably the only information content which either of the numbers should be allowed to carry.

7.3 PERSON DATA

7.3.1 Person Data for natural persons

The table below shows some data elements which could be considered as part of Person Data for a real or fictitious natural person.

Person number	For internal system use only
Entity type	Natural person (with subtypes, eg real or fictitious)
Language / alphabet of entry	
Current real name From date / To date / As at date	One only
Other real name From date / To date / As at date	Repeatable group
Nationality From date / To date / As at date	Repeatable group, for nationalities held concurrently or successively
Address and other contact details From date / To date / As at date	Repeatable group
Date of birth	
Date of death	
Place of birth	
Source	Fictitious persons only
Links to names	"Closed links", only for qualified users

7.3.2 Person Data for corporate bodies

The table below shows some data elements which could be considered as part of Person Data for a corporate body. Since incorporation necessarily brings the corporate identity into the public domain, this is envisaged as an open entry, available to all.

Person number	For internal system use only
Entity type	Legal person (with sub-types?)
Language / alphabet of entry	
Registered name From date / To date / As at date	Repeatable group
Country of incorporation	Not repeatable (requires special provision for international organisations?)
Registered number of incorporation	Where applicable
Date of formation	
Date of dissolution	
Place of formation	
Links to name records	"Open" links, accessible to all users

7.4 NAME DATA

7.4.1 Name data for Key Names

The table below shows some data elements which could be considered for a Key Name record:

Name number	For external use
Entity type	Key Name (with various subtypes, eg name, real name, pseudonym, group name, former name)
Language / alphabet of entry	
Name From date	Only if there is a former name
Date of birth / Date of formation	Only if the name is a real name / group name
Date of death / Date of dissolution	Only if the name is a real name / group name
Title and genre of an associated work / role of name in relation to the work	Repeatable: for disambiguation
Profession, business or field of work	For disambiguation
Other disambiguation note	
Links to variant names	"Open" links, accessible to all users
Direct links to other Key Names	"Open" links
Links to associated Person Data	"Closed" links, only for qualified users if Person Data is for a natural person; "Open" links, if Person Data is for a corporate body.

7.4.2 Name data for variant names

The table below shows some data elements which could be considered for a variant name record:

Name number	For external use
Entity type	Variant name (with subtypes?)
Language / alphabet of entry	
Name From date / To date	Dates for former names only
Link to Key Name	"Open" link, available to all users

8. CONCLUSIONS AND RECOMMENDATIONS

The conclusions and recommendations from this study are in two parts: those that relate to a comprehensive system of name numbering and its possible application in the UK book industry, and those that relate to the very specific issue of improving the control of publisher and imprint names by a targeted subset of what might become a comprehensive system.

8.1 COMPREHENSIVE NAME NUMBERING

8.1.1 The need for comprehensive name numbering

On the evidence of this study, it would be an overstatement to say that the UK book industry sees an immediate and pressing need for the development and implementation of a comprehensive name numbering system covering names as creators, as subjects, and as publishers and imprints. There was, however, an almost unanimous sense that such a system would be beneficial at all levels of the book supply chain (including, very importantly, libraries) if it was sufficiently widely and reliably adopted. There was an equally unanimous sense that it would be a very large and ambitious development.

The absolute need for unique person identifiers for electronic rights trading which has been very persuasively argued in the **<in*d*ecs>** Project (see page 2) is not yet generally recognised, except perhaps in some areas of academic publishing.

There was a clear message that any approach to comprehensive name numbering would have to be international.

8.1.2 Suitability of the IPI system

The brief for this study specifically asked questions about the suitability of the IPI system as the basis of a comprehensive name numbering system for the book industry. Our conclusions are:

(a)   CISAC has unquestionably led the way in identifying and implementing several key elements of a well-designed name numbering system, most notably in drawing a clear distinction between the underlying identity of a natural or corporate person and the names by which the person is known.

(b)   On the basis of the information which was made available to us, the present system falls short of what would be needed, particularly in respect of names of natural persons, to support a comprehensive name numbering scheme serving a much wider community than the copyright licensing societies.

(c)   CISAC and SUISA and/or some of the copyright societies who use the IPI system would, however, be essential participants in any project to develop comprehensive name numbering for the book industry, since IP Name Numbers are already an indispensable part of copyright management and royalty distribution for authors in several countries.

8.1.3 The way forward

In the *Directory of Parties* document already cited[10], the **<in*d*ecs>** Project has outlined the characteristics of a comprehensive system of unique identifiers for person names in electronic commerce. **<in*d*ecs>** is also specifying a possible follow-up project with a view to proving the

---

[10] *http://www.indecs.org/results/persons.htm*

technical and organisational feasibility of implementing such a system. The book industry and the library sector will undoubtedly be represented in any such project, and it would be inappropriate to propose any separate action.

8.2 PUBLISHER AND IMPRINT NAMES

8.2.1 The need for control of publisher and imprint names

Both the study itself and continued demand from within BIC shows that there is a specific need for better control of publisher and imprint names which could be met by a system of name numbering. The requirement is particularly pressing in the UK because of the extent to which book distribution is handled by a relatively small number of large distribution companies.

8.2.2 Feasibility of specific action in this area

The universe of UK publishers is theoretically of the order of 50,000, but in practice real benefits would be obtained if no more than 2,000 to 5,000 active publishers were brought into a well-managed system. While there are arguments for an international approach, it would not be unreasonable for the UK to move ahead either independently or with one or two other English-speaking countries.

The **<in*d*ecs>** model is not a single monolithic name numbering database, but rather a federation of linked systems serving particular communities, with connections which ensure that when a name is numbered in two or more systems a match between the two numbers is set up. A specific publisher and imprint numbering system could be part of this model.

8.2.3 The way forward

If these arguments are accepted, the way forward could be to bring together a group of key participants to confirm whether there is real willingness to commit resources to plan, finance and carry out a project to develop a UK publisher and imprint name numbering scheme within the framework of the wider developments arising from **<in*d*ecs>**. Those key participants would probably include the relevant copyright licensing bodies (PLS, CLA), the trade and library bibliographic agencies (Book Data, Whitaker, BDS, British Library), the trade associations (BA, PA) and some representative publishers and distributors.

We recommend that for this specific need the adoption of IP name numbers should still be seriously considered. The decision would depend not so much on the technical characteristics of the IPI system in terms of database content and structure (which are probably entirely adequate in this particular context) as on the feasibility of agreeing administrative arrangements which would provide the responsiveness, speed of assignment, and open availability needed for general trade use. The alternative of developing a separate numbering scheme and database, and in due course adding IP name numbers if these are required for international copyright management, would surely be more expensive.

In this specific area, we suggest, there is an opportunity both to meet an immediate need and to contribute to the early stages of moves towards comprehensive name numbering.

APPENDIX A: KEY TO ACRONYMS

AAAF	Anglo-American Authority File
AACR2	Anglo-American Cataloguing Rules, 2nd edition
ALCS	Authors Licensing and Collecting Society
BDS	Bibliographic Data Services Ltd
BIC	Book Industry Communication
BLNAL	British Library Name Authority List
BNBRF	British National Bibliography Research Fund
CAE number	Compositeur-Auteur-Editeur number
CIP	Cataloguing-In-Publishing, the procedure by which catalogue records for new books are incorporated into the British National Bibliography before publication.
CIS	Common Information System
CISAC	International Confederation of Societies of Authors and Composers
CLA	Copyright Licensing Agency
EAN	Formerly the European Article Numbering Association, now renamed EAN International
EDI	Electronic Data Interchange, the exchange of trading messages in a standard format between computer systems.
IFLA	International Federation of Library Associations and Institutions
<in*decs*>	INteroperability of Data in E-Commerce Systems
IPD	International Performers Database
IPDA	International Performers Database Association
IPI	Interested Party Information (system)
IPN	International Performer's Number
ISADN	International Standard Authority Data Number
ISIL	International Standard Identifier for Libraries
ISO	International Standards Organisation
LASER	London And South Eastern Region library interlending network
LC	Library of Congress
MARC	MAchine Readable Cataloguing
NACO	Name Authorities Co-operative (USA)
NBS	National Bibliographic Service (UK)
PLS	Publishers Licensing Society
SAN	Standard Address Number (used in US and UK book trade)
SCAPR	Societies' Council for the Administration of Performers' Rights
STM	Scientific, Technical and Medical (of publishing)
SUISA	The society which administers copyright licensing and royalty collection for composers, lyricists and music publishers in Switzerland and Liechtenstein
UBCIM	Universal Bibliographic Control and International MARC (programme) of IFLA