The ISSN and the Turing machines: the history of a relationship

Pierre Godefroy
Project manager and Information Technology Coordinator of ISSN International Centre.
E-mail: godefroy@issn.org

Abstract

On the unending path towards a better integration of the ISSN and its accompanying metadata in the semantic web, the new horizon is that of linked data, so as to make sure that ISSN data is fully linkable for universal computing. However, while striving to feed hungry Turing machines with ever more data, thus fulfilling in a way Garfield's visions outlined sixty years ago, one should not forget about so called end users, the ghosts behind the machine.

Keywords: ISSN. Semantic web. Turing machines.

O ISSN e as máquinas de Turing: a história de um relacionamento

Resumo

No caminho sem fim para uma melhor integração do ISSN e seus metadados na Web semântica, o novo horizonte e o de dados vinculados, de forma a certificar que os dados do ISSN são totalmente passíveis de ligação na computação universal. No entanto, enquanto há o esforço para alimentar com cada vez mais dados as famintas máquinas de Turing, cumprindo assim de certa forma as visões que Garfield esboçou 60 anos atrás, os chamados usuários finais não devem ser esquecidos, os fantasmas por trás da máquina.

Palavras-chave*: ISSN. Web semântica. Tecnologia.*

El ISSN y las máquinas de Turing: la historia de un relacionamiento

Resumen

En el camino interminable para una mejor integración del ISSN y sus metadatos en la Web semántica, el nuevo horizonte son los datos vinculados, que certifican que los datos del ISSN son totalmente posibles de conectar en la computación universal. Sin embargo, mientras nos enforzamos por alimentar con cada vez más datos las hambrientas máquinas de Turing, cumpliendo de cierta forma las visiones que Garfield esbozó 60 años atrás, no debemos olvidarnos de los llamados usuarios finales, los fantasmas detrás de la máquina.

Palabras clave*: ISSN. Web semántica. Tecnología.*

INTRODUCTION

From its inception, the ISSN has been linked to the world of computers. As a number, it is an abstraction which can be compared to what is called a *pointer* in the framework of computer science (a pointer is a programming language object whose value points to another value stored elsewhere in the computer memory). An ISSN is in fact a *pointer* to a specific journal (or *continuingresource*), itself identified by a set of metadata elements.

Such a pointer had become a practical necessity: after the second world war, the universe of continuously expanding published scientific knowledge, stored in streams of articles bundled under specific titles, could be compared to a kind of giant virtual memory which had to be rationally organized. The ISSN was in fact part of this endeavor and its practical implementation was based on the availability of emerging computing power and networking.

The aim of this article is to outline the successive steps of the history of ISSN mass registration of serial publications in the real world of *Turing machines* which made it possible.

THE BIRTH OF A CONCEPT

When it first appeared (GARFIELD, 1955), the concept of a universal identifier for periodicals as one of the cogwheels for organizing the management of the virtual web of scholarly citations was intrinsically linked to the possibility of using *mechanical devices of high speed*:

> Some time ago I became concerned with the problem of developing a citation code for science. This was necessary for the efficient manipulation by mechanical devices of entries to scientific indexes. In the course of this research I developed a very simple system for identifying an individual scientific article that had appeared in the periodical press. The resulting numerical code consisted of two parts. The first part was a serial number, used instead of an abbreviation, to identify each periodical; it was similar to the serial numbers employed in the World List of Scientific Periodicals, by no means a new idea. For example, *Die Bibliographie der fremdsprachigen Zeitschriften Literatur* has for many years used such a system to save space.
>
> A citation index to science would have the following main characteristics. First there would be a complete alphabetic listing of all periodicals covered, in addition to the code number for each periodical. This list would be similar to the World List, but without the library holdings information. The main portion of the citation index would list in straight numerical order the code numbers for all the articles covered. Under each code number, for example, 3001-6789, there would be listed other code numbers representing articles that had referred to the article in question, together with an indication of whether the citing source was an original article, review, abstract, review article, patent, or translation, and so forth.

At the time of writing of this prescient intuition, Alan Turing (1912-1954) had just died, computers were alluded to, but the approach was more of a pragmatic nature than a theoretical one. The new field of *documentation* was in its infancy (it would be known as *information science* in the 1960s) and the sheer perspective of organizing human knowledge through appropriate standardizing tools was in itself more attractive than the somewhat crude calculators of the time which were using programming languages more adapted to the management of numbers than to the handling of text strings.

Although the pioneering ISSN *concept* had in fact been clearly exposed by Garfield, it took a number of years to put it into universal practice, at least at a truly international level. The individual technical components necessary for a practical

implementation of the idea were most certainly already present by the beginning of the 1960s but the major obstacle, apart from their crudeness, was most certainly the lack of political will to tackle the issue at the appropriate level.

However, the CODEN, a six character, alphanumeric bibliographic code already provided for the identification of titles of periodicals and non-serial publications, from the end of the 1950s. It was one of the first experiments in the field of machine readable identification systems for periodicals.

When the ISSN was finally designed in 1972, it could already rely, apart from the previous CODEN experience, on two pillars which started to disseminate in the 1960s:

– The standardization of bibliographical metadata in the form of the *marc record* (machine readable cataloguing).

– The *wider* availability of mainframes for computational operations.

THE INITIAL DATA AGGREGATION

The initial partners in this endeavor were institutions already familiar with the management of these tools: national libraries, institutions managing scientific information, abstracting and indexing databases.Following the somewhat crude pioneering steps where the very first ISSN had been assigned *mechanically* to already existing serials lists (Bowker'sNew Serial Titles), the initial challenge of what was called at the time the ISDS (International Serials Data System) was to merge together different heterogeneous sources already identifying serials titles.

The initial work was done in cooperation with ICSU AB (the International Council of Scientific Unions Abstracting Board created in 1949 by Unesco, it changed its structure in 1984 and became the International Council for Scientific and Technical Information: ICSTI). ICSU AB had launched a *numbering* project for periodicals present in abstracting databases in 1970 and was keen to pragmatically piggy back on the newly emerging ISSN, in parallel with the existing CODEN.

Slawek Rozenfeld, the computer specialist who was hired for the task at the time and who was to manage the ISSN database from its inception for many years, remembers a series of successive painstaking processes designed to combine several source files in order to obtain the Holy Grail of an *aggregate list of periodicals*. It should be noted here that the order of magnitude of the list at the time (30,000 titles) is not very different from what can still be considered as a core collection of STM and scholarly titles today. It also seems that the common ISSN/ICSU AB initial project was the scene of a *clash of cultures* which probably had its source in different views as to the nature of the endeavor: on the one hand ICSU AB had a more or less finite goal which could be achieved in a relatively limited amount of time whereas the ISSN was obviously more ambitious and universal in its coverage and, being more library based, relied on more detailed identification metadata which were to be carefully checked, thus delaying in some cases the identification priorities of the other party. Scientists were furthermore at pains to understand

why on earth librarians seemed to attach as much importance to obscure local news bulletins as to their valuable journals.

THE ISSN AND THE EARLY DAYS OF COMPUTING

The initial programs for the management of the ISSN had been written and designed by engineers from the École des Mines (France). Initial tests were positive but when the machines were loaded with greater quantity of data, the internal chaining logic of the system tended to collapse. This was due to the lack of experience with the handling of text strings. COBOL was the programming language of choice at the time but it had been designed for management and accountancy purposes (as shown by its acronym: common business-oriented language). The small IT team of the ISSN Centre had to switch the initial routines to a more promising and efficient language, i.e. PL/1 developed by IBM.

Until 1983, all data processing was based on *time sharing* rented on external mainframes, i.e. there was no hardware on the premises of the ISSN International Centre. The IT staff often had to spend a lot of time, including cheaper evening hours, by the mainframes themselves, far from the ISSN office. Punched cards were still being used in order to submit the programs to the machines with the risk of inadvertently losing or inverting one… Time sharing also meant that one had to be careful with expensive time management: a program which was found endlessly *looping* in the morning, clogging the system after a whole night of unnecessary processing could potentially mean the loss of a full year of planned computer expenses.

For quite a number of years, data exchanges with big institutions were based on magnetic tapes the size of a wedding cake. The IT team even had to process punched paper tape from the Soviet Union. On the other hand daily production at the ISSN International Centre (and from many other centres) was manually written on worksheets (exceptionally typewritten) and given to an external company in order to be fed into the data processing per se. This process was particularly error prone due to the international nature of the data with many similar looking diacritics being used.

It soon appeared, after the first paper updates of the ISDS Register, quickly interrupted, that practical data dissemination of the *ISDS Register* could only be performed through magnetic tapes and microforms (a bulky set of several hundred cards including an ISSN index and a title index). Soon enough it was discovered that the production cost of the microforms could be significantly reduced by avoiding the output of a wholly new data set of hundreds of cards for each update: only the indexes were reissued, together with additional data cards for created and amended records.

The first computer mainframe appeared in the premises of the ISSN International Centre in 1982: it was a bulky Siemens which needed two offices to set in. It was replaced in 1988 by a more manageable 9370/40 IBM computer which only needed one room. It also meant the arrival of terminals available to cataloguers, who were still unable to catalogue directly *online* but who could now type records in batch mode with results of the loading being made available on the next morning. The final step in this *miniaturization process* took place in 1997, with

a switch to an IBM P390 (with an OS2 interface layer) which had the size of a big suitcase: the PC size was almost reached.

The first CD-ROM edition (introduced in 1992 as *ISSN Compact*) was the first true view of the richness of the accumulated data only a glimpse of which had previously been made available through the existing crude carriers. The initial DOS version (which had its limitations for a proper display of special characters) was further enhanced a few years later through a transition to the brave new world of Windows. The CD-ROM version of the ISSN Register was available for a total of some 15 years during which the *postal dissemination* of the data was much easier. It should be noted that although it is now rightly considered that the CD-ROM is an obsolete carrier, the speed of the index browsing it offered has still no real equivalent over the Internet or the web. Furthermore, thanks some complex indexing routines, the CD-Rom version of the ISSN Register offered an *index filtering* option (allowing for *national* versions of each browseable index), a special feature which is not availablein most current implementations of faceted searching.

In spite of all the progress made in networking speeds, it is obviously difficult to compete with the computing efficiency of a full database which is directly available on your own device with all its indexing structure. Maybe, when the time comes, both in terms of memory availability and transmission speeds, we shall see fully self contained*apps* on mobile phones which will be the equivalent of the CD-ROMs of yesterday, with underlying automated updates? We would once again be able to truly use the catch phrase "the ISSN at your fingertips".

OSIRIS: THE PC MEETS THE ISSN

With the introduction of the PC on the mass market in1980s, ISSN data creation and management could begin to escape the rule of the mainframe and its clumsy data handling extensions. The Osiris software (Online Serials Information Registration and Inquiry System), based on the popular Micro-ISIS software developed and promoted by Unesco, became a tool of choice for small or medium sized ISSN national Centres which started sending floppy disks to Paris (instead of paper worksheets) as from its implementation in 1991. In fact, the planning process for the new software had already begun in in 1986 in Budapest, at the occasion of an ISSN Directors Meeting: it was a symbol of the new possibilities offered by the bludgeoning East West cooperation. The two software engineers who designed the specific ISSN layer on top of the Micro CDS/ISIS core, Péter Jacsó and András Szűcswere both from Hungary. The ISSN Centres choosing to use Osiris were also often from central or eastern European countries.

Osiris had the advantage of incorporating the whole set of strict controlling rules implemented for the ISSN Register. Thanks to these locally implemented rules, the data files exported from the local Osiris databases and sent to the ISSN International Centre usually contained almost no rejected record when being fed to the database, with exception of title duplicates which could only be detected during the matching with the ISSN Register as a whole. The small and medium sized ISSN Centres concerned had potentially a full control over their local national ISSN Database which could also contain data fields not directly

related to ISSN registration. The only drawback was that each national implementation of Osiris had to be fully customized (in particular for *special characters*, sorting rules etc.) which made the updating process of the software a bit difficult.

It should also be stressed that Osiris was in its special way a prefiguration of what could be achieved through the dynamics of open source: an access to computing processes made easy and efficient through the possibility of cooperative work.

THE ISSN ENTERS THE INTERNET AND SPINS THE WEB

At the beginning of the 1990s, the Internet was still largely a mysterious *terra incognita* restricted to an elite of specialized university departments. The author of these lines was sent as an explorer to investigate its nature and possibilities. Paradoxically, knowledge of this emerging new continent was first explored through a number of books which described it in a very abstract way, comparing it with other existing networking solutions. The ISSN International Centre ended up having its first email account being hosted by one of the very first Internet *virtual communities*, based in California, i.e. *The Well*which offered an easy subscription scheme.

The world wide web itself was also discovered through the *Well* which thus hosted the first ISSN web site which was published in 1996 and which can still be retrieved under an early guise through the invaluable web archivers from archive.org (ISSN, 2015). The web site was initially mirrored by the colleagues from the University of Oslo at http://www.issn.uio.no/issn.

Communicating instantly with the whole world through email or a web site was one thing; it quickly emerged that making the ISSN Register data available *online* on the Internet was the way to go.

It was already felt by the mid-noughties that the ISSN data management module used for the production of the ISSN Register, still based on the original in house developments dating back to the inception of the ISDS, was lagging behind, if compared with the novelty of the possibility of a wide online exposure through a web database and with the possibilities offered by data transfer through the Internet.

The ISSN TAC (Technical Advisory Committee), an expert body convened ad hoc for consultancy on technical strategic decisions, was reactivated in 1994 / 1995 (with the participation of Gianpaolo Del Bigio, the author of the CDS/Microlsis system, Knut Hegna from the University of Oslo, Bernard Martin and Tito Suter from ISSN Argentina) and suggested different possible paths to achieve such a goal while avoiding too much disturbance for current production. One of the main ideas which emerged, partly due to security considerations, was to separate data exposure through the Internet from the data production environment which could be restructured at a slower pace. Concerning the latter, there were obviously some hesitations, as investments costs for what was felt to be basically a brand new information system were deemed not negligible in all cases (both in the case of new in-house developments and in the framework of the purchase of a customized existing proprietary system).

For these reasons, the International Centre chose as a first step to concentrate its efforts on the web

access issue and after a first pioneering test at the University of Oslo, implemented by Knut Hegna, the ISSN International Centre developed its own in house *ISSN Online*web application in 1998, with a Unicode compliance from the very start, due to the truly international nature of the data. For the very first time the ISSN Register was accessible potentially everywhere, all the time (although it was still only available through a commercial subscription). *ISSN Online* offered a regularly updated access to the contents of the ISSN Register.

FINDING THE RIGHT DATA MANAGEMENT SYSTEM

In parallel, the IT team at the ISSN International Centre was busy testing various database tools on the market for the planning of a new data management system. However, combining the requirements of thorough bibliographic indexing and the complex logic of the management of an identifier proved to be a difficult match for *out of the box* products which had various slants. The team finally chose to combine various commercial or open source tools with in house developments. Although significant progress had been made on this path, there was no overall consensus on the viability of the process and it was decided to adopt a new strategy, if possible strictly based on available commercial products designed for the library world. It was felt that *library management systems* were now sufficiently mature to cater for the needs of ISSN.

It was thus finally decided (2003) to try to use for this purpose an existing library system by customizing its cataloguing module. After a wide international call for tender, VTLS, a US based company, was chosen with its Virtua library management software which was then relatively new, as it had been launched in 2000. The ISSN technical specifications were detailed and, in spite of the *out of the box* basic requirement, they finally entailed a thorough customization of the product. VTLSproved to be an efficient and responsive partner, in particular thanks to the pragmatic abilities of its CEO and founder, Dr. Vinod Chachra, who was also the technical designer of Virtua. The goal was to adapt the software wherever this had to be done, while trying to minimize the specific developments. The challenge proved to be not so much *technical* as human; the main issue was to make sure that both teams, at the ISSN International Centre and at VTLS, shared a common understanding of what had to be achieved. For the ISSN team it was a leap into abstraction and for VTLS it was about understanding the requirements of exotic customers who did not have to manage physical items or demanding patrons but a number whose logic had to be protected at all cost in order to maintain the consistency of the database.

Finally, the reward was up to the expectations:for the first time (2004), it was possible to assign ISSN numbers in real time over the Internet by cataloguing directly into the system, and to make the results of the new identification immediately and universally available. Machines were as happy as humans since Virtua was also by its very nature a Z39.50 server which could serve data to Z39.50 clients: ISSN was available in the world of *machine to machine* interfaces, although the protocol was in practice mainly restricted to the library world.

ISSN cataloguers also had the pleasure to use a WYSIWYG (What you see is what you get) interface

without the clumsiness of special codes for special characters. Thanks to the necessary switch to a widely used bibliographic format (MARC 21) at the occasion of the implementation of the new software, they also had the possibility of entering *original script* entries (in cyrillic, arabic, chinese etc.), alongside entries transliterated into the latin alphabet, thus paving the way for one of the major potential improvements to the ISSN Register, stressing its truly international nature.

During the initial start up year (2005) it was possible for most ISSN centres using the Osiris software to abandon it in favour of the new online cataloguing possibilities offered by Virtua. This entailed a lot of conversion work for the various CDS-ISIS databases concerned as well as the training of all cataloguers concerned; as a result of this initial dynamic, ISSN Virtua is now used for direct cataloguing by more than 50 cataloguers in 35 different countries around the world.

In spite of the detailed initial specifications and of the *out of the box* expectations, it soon emerged in practice that the *data maintenance* of the new Virtua system entailed on the part of the ISSN International Centre an ability to access the underlying relational data structure and in some cases to *massage* directly the bibliographic records both inside and outside the system. This proved to be achievable, thanks to relatively open nature of the software which, although proprietary, did not hide everything under the hood. This was for instance necessary to make sure that ISSN blocks fully maintained their consistency over time. Over the years, a number of Scripts, written in Perl, were added and enhanced in order to make sure that the practical requirements of ISSN data producers and users could be met without costly investments or an excessive dependency on software suppliers.

Thanks to this *customized* approach, it has been possible over the last years to satisfy the diverse needs of various external users who choose to submit their local data (ISSN and/or titles) for checking and enrichment, in the framework of *ISSN premium* services.

SUCCESSIVE IMPROVEMENTS TO THE SYSTEM

The core ISSN Virtua system was itself improved over the years by VTLS at the request of the ISSN International Centre, in particular in order to implement the ISSN-L logic (2007 revision of the ISSN standard) by tying together ISSN records identifying the different medium editions of a given publication. The ISSN-L is designated and stored automatically by the system on the basis of the nature of the bibliographic links entered by cataloguers when identifying a new resource. Human beings should concentrate on text and leave the management of numbers to the machine whenever possible…

Thanks to the use of an underlying relational database management system, in our case Oracle, for the storage and management of the links between ISSN (successor titles, other medium editions, supplements etc.)it was also possible to start exposing the intricacies of this web. *Extended families* of directly linked ISSN records are now regularly computed and made available to users.

In 2010 a new major improvement was made to ISSN Virtua, with the exposure of ISSN data

through the OAI-PMH protocol. For the first time, ISSN users were able to replicate the ISSN Register by updating automatically their own local copies at a chosen rate though automated MarcXML downloads. This new feature was in fact a second *machine to machine* interface which could act as a kind of API (at least for direct ISSN requests). A few years later (2014), the reverse data flow using OAI-PMH was implemented: an ISSN National Centre (Germany) started exposing its ISSN records trough this protocol, thus allowing for a regular harvesting by the ISSN International Centre in order to feed the ISSN Register. One could even imagine that after some testing and tuning, the process could even be fully automated, with human intervention being needed only for the feedback concerning rejected records.

The latest major improvement to the system is the possibility for the ISSN International Centre, added in 2015, of creating and managing its own custom validation rules, expressed as Perl snippets of codes, instead of the hard coded ones initially input into the system, which can only be turned on and off but not directly modified without rewriting some proprietary code. We have here yet another sign showing that open source is slowly but surely expanding its realm in a pragmatic way.

It is now also possible for external pieces of software to load records into the ISSN Virtua system directly over the web, thus opening a whole new range of possibilities.

In fact, one might assert that the overall tendency for the ISSN data system over the years is one of opening up to the external world, obviously at the cost of some needed added complexity.

THE ISSN AS A LINKINGTOOL

The ROAD project (2014) is the latest milestone in the history of the use made of computing techniques by the ISSN. The ROAD web portal is making freely available ISSN records identifying open access online resources published by scholarly communities. It is meant as a test bed for future improvements which can be brought to the ISSN Register as a whole: faceted searching, records available in MarcXML as well as a linked data version of the data based on the Press00 standard developed by the ISSN International Centre and the Bibliothèque Nationale de France (as an extension of the FRBRoo framework set forward by the CIDOC-CRM).

However, the major feature of this project is the use of the ISSN as a linking bridge between ISSN identification metadata and external resources using the ISSN as an identifier (such as A & I databases, the Directory of Open Access Journals etc.) This allows for a very significant enrichment of the data, the ultimate goal being the *contextualization* of the selected continuing resources so as to help the community of scholarly users with the evaluation of their adequateness.

As an identifier, the ISSN can be used as fundamental linking tool aggregating statements on continuing resources emanating from many heterogeneous sources. The dynamics of the linked data toolbox should of course greatly facilitate this *universal linking* potential in the coming years.

However, one should not forget that the ISSN has already been used for many years in this context, whether it be as a pragmatic obvious *matter of fact* as in the OpenURL framework, or in conscious long

term projects such as *The Keepers* (http://thekeepers.org) aiming at the registration and exposure of long term archival and preservation intent of online continuing resources by various archiving agencies. All this is made possible by the fact that the ISSN is a circulating piece of information which can be *hooked* in different contexts.

In a way, through the semi-spontaneous, semi-conscious dynamics of *ISSN linking*, we might be witnessing a kind of spiraling progress of the ISSN which, having reached an impressive quantitative threshold of almost two million identified resources, will now once again fully live up to the expectations of its early designers and implementors: it should be viewed and used as a simple but elegant cogwheel among the ticking clockworks of the publishing process.

The initial official name of the ISSN network, first known as the *International Serials Data System*, was probably too cryptic and for this reason was abandoned in 1993. The concept however embodied the idea that computing schemes managing assertions about serial publications and their contents (articles etc.)would be linked together thanks to the ISSN as bridge.

On the unending path towards a better integration of the ISSN and its accompanying metadata in a web which is becoming more *semantic*, the new horizon is that of linked data, so as to make sure that ISSN data is fully *linkable* for universal computing.

CONCLUSION

However, while striving to feed hungry linked Turing machines with ever more atomized chunks of data, thus fulfilling in a somewhat unexpected way Garfield's visions already outlined sixty years ago, one should not forget about the so called *end users*:these are the ghosts sitting not inside but outside the machines, who should constantly make sure these tools do not go haywire with junk data.

Both ISSN registration and, at the other end of the chain, ISSN checking, i.e. making sure that one's data is systematically using valid ISSN (these activities being in a way the two faces of the same medal) should be streamlined as much as possible so as to empower information managers as masters of (mostly) true assertions.

The art of computing has many layers, from efficient programming logic to adequate interfaces. The past history of the relationship between ISSN and computers shows that the general trend observed over the last years might be the transition from the careful permanent curation of raw aggregated data (the ISSN Register as a data file) towards the addition of more sophisticated user oriented services such as forms of data mining (faceted searching being a very first step) or the extraction of formerly hidden relationships (*genealogical trees* of serials). The *opening up* of ISSN data through the semantic web, the proper standardization of its metadata elements (for instance using ISNI for the identification of serials issuing bodies) and the linking of the ISSN with other data sources (in particular in order to help solving the issues of long term preservation and archiving of online artifacts) are other aspects of the way ahead.

Computers made the ISSN possible; it is now up to the ISSN to make the most out of computers.

ACKNOWLEDGMENT

I have been a direct witness of the second half of the 40 years ISSN developments, having worked at the ISSN International Centre since 1990. I would like to express my special thanks to Slawek Rozenfeld, head of the ISSN computer section from the very beginning (1974) until 2001 for his invaluable insight into the early years of the system.

REFERENCES

GARFIELD, Eugene. Citation indexes for science: a new dimension in documentation through association of ideas. *Science*. 1955, 122(3159), 108-111.

ISSN. Taming the serials jungle with the issn. [viewed 10 april 2015]. Available from: https://web.archive.org/web/19970129005727/http://www.issn.org/