



68th IFLA Council and General Conference

August 18-24, 2002

Code Number: 054-133-E
Division Number: IV
Professional Group: *Cataloguing*
Joint Meeting with: -
Meeting Number: 133
Simultaneous Interpretation: Yes

Report on the successful AustLit: Australian Literature Gateway implementation of the FRBR and INDECS event models, and implications for other FRBR implementations

<http://www.austlit.edu.au>

Marie-Louise Ayres

Project Manager, AustLit: Australian Literature Gateway 1999-2002
University of New South Wales at the Australian Defence Force Academy
Project Manager, MusicAustralia, 2002-
National Library of Australia

Kerry Kilner

Project Manager, AustLit 2002-; Content Manager, 1999-2002
University of Queensland

Kent Fitch

Developer
Project Computing Pty Ltd

Annette Scarvell

Content Manager,
University of New South Wales at the Australian Defence Force Academy

Abstract:

This paper discusses the first major implementation of two significant new cataloguing models: IFLA's FRBR (Functional Requirements for Bibliographic Records) and event modeling (INDECS and Harmony). The paper refers briefly to the decision making processes leading to the adoption of these models, and outlines the implementation process, the benefits of the implementation, the practical and conceptual difficulties encountered in this implementation, and some observations on the future of these models in the library and information worlds. IFLA's Functional Requirements for Bibliographic Records was published in 1998, and was widely accepted as providing a sound conceptual model for a new generation of bibliographic records which record and present the publishing history of information resources. The 2000 LC Cataloguing conference included a number of papers on the requirement to add 'event models' into cataloguing. FRBR and Event Modeling are powerful tools for presenting bibliographic and other information in a richly contextual environment. Implementing the models presents significant challenges but is achievable, cost effective, offers many benefits to practitioners and should be considered by a range of information providers.

Keywords:

Australian literature; Functional Requirements for Bibliographic Records; Literary databases; Subject Gateways; XML.

Introduction

The International Federation of Library Associations' *Functional Requirements for Bibliographic Records* (FRBR) model¹ has made a major contribution to theorising bibliographic description, and the ways in which bibliographic description needs to be rethought in the Internet age. The model has also been the subject of considerable and accelerating comment² and suggestions for amendment. This paper describes the implementation experiences of a small group which chose to implement and extend the FRBR model because it most suited a particular set of literature oriented requirements. AustLit: the Australian Literature Gateway, a web-based resource discovery service about Australian writers and writing, is the result of collaboration between eight Australian universities, each of which had developed specialist but non-standards based biographical and bibliographic databases³ and the National Library of Australia⁴, with the Australian Research Council⁵ providing development funding in 2000, 2001 and 2002.⁶ Development of the AustLit technical infrastructure commenced in May 2001, and by October 2002 – less than 18 months later – the AustLit service was released for public use, first as a free public trial, and, from January 2002, as a fully functional subscription service which makes a portion of its author information freely available.⁷

Choice and Extension of Models

As the AustLit development team placed a very high value on representing the publishing histories of works, finding the International Federation of Library Associations' 1998 *Functional Requirements for Bibliographic Records* (FRBR) model was very exciting.⁸

As FRBR aficionados will know, the FRBR model includes the concepts of:

- The Work: an abstract concept (e.g. the idea of the novel *Voss* by Patrick White);
- The Expression: a realisation of the Work (e.g. White's original version of the novel in English or the German translation by John Stickforth);
- The Manifestation: a particular embodiment of the Expression (e.g. the 1958 Kiepenheuer & Witsch publication of Stickforth's translation of the novel *Voss* by Patrick White);

- The Item⁹: the individual item on a Library shelf (e.g. the copy of the 1958 Kiepenheuer & Witsch edition of the John Stickworth translation of the novel *Voss* by Patrick White, held at the National Library of Australia).

AustLit augmented the FRBR bibliographic description model with 'event modelling':

- Works have a **creation** event
- Expressions have a **realisation** event
- Manifestations have an **embodiment** event

Works can be expressed one or many times, Expressions can be manifested one or many times,¹⁰ and manifestations can result in one or many items. In the AustLit model, Works, Expressions and Manifestations all have attributes, and Creation, Realisation and Embodiment **events** all have **attributes**. AustLit has also augmented the model by incorporating the concept of SuperWork, as suggested by a number of FRBR commentators.¹¹

Perhaps AustLit's most significant extension¹² of the FRBR model lies in its representation of agents (authors and organization). While the FRBR model and its subsequent commentators have been at pains to stress the need for agent role information in relation to works, expressions, manifestations and items, AustLit also includes:

- Birth and death (or creation and cessation) events for authors and organizations, and date and place attributes for those events;
- Award events (drawn from both agent and work records, with both displayed on the agent record) and award name, date and place attributes;
- Gender, nationality and self-claimed cultural heritage attributes;
- Arrived in Australia events and associated date attributes;
- Uses alternative name attributes (for navigation of pseudonyms and other multiple names);
- Archival holdings attributes.¹³

Implementation: building the database

Once desired functionality had been specified, it was clear that we would need to build, rather than buy, a system: there are currently no commercial systems which support all the data models, or which support the complex relationship concepts of Topic Mapping¹⁴ in database design. All AustLit entities, including events and attributes, are **topics**, and relationships between those entities are also topics: the AustLit Gateway includes more than 3.3 million topics. The basic design documents relating to our custom built system are publicly available at the website.¹⁵

Although the topics and their relationships are stored in conventional (but unusually highly normalised) relational database tables, the system converts the data into a common XML format at an early stage of output processing. From this common XML format, information is transformed into the desired final output format (typically HTML) using XSL (eXtensible Stylesheet Language). The XML representation contains enough information to generate alternative encodings such as MARC or to augment the HTML with Dublin Core or RDF metadata.¹⁶ With the exception of the Oracle database – which our University licence made available to us – all other software used is open source. AustLit runs on a Sun Microsystems Blade-1000 workstation, under the Solaris operating system.

At the outset of the implementation phase, we believed that the major risks lay in the complexity of designing a database to accommodate the *FRBR*, *INDECS* and *Harmony* models along with all the multitudinous relationships we had mapped out, and the likely performance of a highly normalised (i.e. consisting of some millions of 'topics') database. As it turned out, these were not the major hurdles we had envisaged, and the development team has been very pleased with the design outcome and database

performance. Database design work commenced in July 2000, and by March 2001 the database had been designed, most data had been converted, the essential elements of the maintenance interface had been developed, and the AustLit staff from eight institutions around Australia were trained and commenced work in the new system.

Implementation: Converting the data

Having said that, there certainly have been issues in implementing the *FRBR* model and the other elements of the AustLit model which intersect with it, such as representations of events and agents. The major implementation issues had little to do with the models we chose: most of these issues would have arisen whatever data models and standards were chosen. We substantially under-anticipated the risk which lay in migrating a range of existing non-standards based databases to the new structure. Every new database brought new problems and we were not able to reapply previous conversion solutions!

We also encountered significant issues relating to interpretation of the *FRBR* and the pragmatics of implementation. The model was clearly written with a 'whole monograph' emphasis (although the model demonstrates that it can be used for other types of works, such as performances). AustLit's implementation was complicated because only a small portion of AustLit records fit this model, as AustLit includes a wide variety of individual non-monograph items (eg. individual poems, reviews and articles), and represents complex clusters of items such as poem sequences and author series.

However, as would be expected of any catalogue or index, the overwhelming majority of AustLit records have one to one relationships between work, expression and manifestation, with conversion of these records being relatively simple. Our conversion methods evolved as we worked, with quite a number of mistakes made along the way, none of them irretrievable. Our conversion methodology was (roughly):

- All records which appeared to fall into the one-to-one group (via checking author, title and publication details) were handled using a series of stylesheet passes to convert the data into the AustLit XML schema;
- All records which unambiguously contained translation relationship data were automatically converted to expression and manifestation level of the relevant work;
- Any records which appeared to be possible multiple expressions of a work (via checking author, title and publication details) were quarantined. These were then inspected by librarians and indexers trained in the *FRBR* model, and a series of very efficient web tools were developed allowing staff to merge multiple records into a single work record, to create new expressions within work records, and to merge expression and manifestation information where this was duplicated.

Implementation: maintenance interface and retraining the staff

The AustLit maintenance interface tightly couples the various model elements of work, expression and manifestation with interface elements: staff work within a single but highly customisable 'record' which visually mimics the 'enclosures' inherent in the model: eg these particular manifestations belong inside this expression; these expressions belong inside this work. The maintenance interface makes extensive use of the scripting facilities and Document Object Model (DOM) interface provided by Internet Explorer version 5.5 (or above). This means that AustLit maintainers require no client software, that start up costs are minimal (all that is needed is a reasonable PC, IE5.5 or above and access to a network), and that staff have great flexibility in choosing which record level, events and attributes they wish to work with. As the number of events and attributes which staff can include is considerable, separate start-up 'templates' are available to staff which include the events and attributes mostly commonly associated with particular worktypes, forms and genres (only the 'poetry' template, for instance, automatically presents the field for the work attribute 'first line of verse').

Retraining AustLit staff to work within the FRBR model was a high priority for the development team. Once they were familiar with the model, staff became very appreciative of the opportunity to represent works in a rich context. They enjoy the maintenance interface which gives them many choices about how to describe works and authors – in many cases recording information which had always been ‘to hand’ when describing items, but simply could not be represented in previous data models.

AustLit also has a very effective review interface. The need to review work should decrease over time¹⁷, but the interface still provides an excellent opportunity to ensure that records are as consistent as possible – especially in those areas where FRBR type decisions need to be made. The fundamentals of the model are easily understood by professional staff. It must be said, however, that distinguishing between new expressions and new manifestations of works can pose significant challenges. The application of the model to the ‘real world’ of describing real items in hand involves considerable ongoing discussion among the AustLit staff, and requires both regular guidance from content managers and thoughtful revision and enrichment of the manual. Of course, inconsistency of cataloguing practice is not confined to FRBR description,¹⁸ and it is likely that acceptance of some level of inconsistency will occur in future large-scale implementations.

Given the emphasis on providing effectively modelled, coded and navigable relationships between entities, the maintenance interface reflects AustLit’s concentration on the use of authority files. The interface requires far more ‘selection’ of authority-defined events and attributes than is the case in a standard cataloguing interface. An example of this extra authority orientation is that all place data – whether this is place of birth for authors, subject or setting for works, or place of publication - must be selected from the place section of AustLit’s thesaurus. The topic map basis of AustLit’s thesaurus means that it is possible to retrieve, for example, all authors born in the Gippsland region of Victoria (the author’s actual town of birth is recorded in the author record, but the topic map ‘Gippsland’ gathers them together). While the development team certainly heard negative reactions from non-AustLit cataloguers to this requirement to work extensively by selection from authority files, rather than by entering self-generated data, those working within the system do not appear to resent this ‘direction’ of their work. This is perhaps because all the staff were very aware of the difficulties which the relative lack of effective authority files in the pre-existing databases caused during the conversion process.

It must be said, however, that most of the professional librarians and bibliographers working on AustLit are true specialists, were already working in non-catalogue environments, and have a deep knowledge of the subject matter. It is therefore difficult to know whether this happy adoption of the model and interface would hold true for other groups of cataloguers. However, in the conversion and cleaning up phase, a number of professional librarians from outside the AustLit ‘circle’ were employed on contract, and experienced no difficulties in quickly learning to work within the FRBR environment and to use the AustLit interface.

Implementation: the user interface

Throughout the development of the AustLit database and user interface, the team worried about how to present this new concept of works, expressions and manifestations to users. This seemed to be a very complex notion to try to convey through a web interface, especially given our own need to keep drawing diagrams and verbalising relationships for our own benefit. The development of the final user interface was deliberately left until very late in process – we did not allow interface needs to ‘drive’ our modelling. Of course, as the AustLit staff began using the maintenance interface and it in turn required a basic ‘user view’ interface for their use, the iteration of the basic user interface elements occurred over a series of months, and partly in response to staff requests.

However, we were still very concerned about representing the FRBR relationships in our final graphic design, feeling that we really needed to highlight the groupings of expressions and manifestations. As we

started working with the graphic designer, we came up with all sorts of heavy visual ‘clues’ about these groupings. We variously tried having expressions appear in different coloured ‘blocks’; using obviously table formats with ‘cells’ drawn around expressions and various manifestations within those expressions; coloured bars down the sides to draw together expressions; different forms of words. The interesting thing about this process was that when we did our very first external user testing¹⁹, we found that users did not require such obvious clues about the relationships between works, expressions and manifestations. In the end, we chose to use ‘light’ visual clues such as dot points and separator lines, and simple prose statements such as ‘This work has appeared in x different versions’ and ‘This version of this work has been published x times’. While we have been unable to do any sophisticated user testing yet²⁰, there appears to be good user acceptance of these methods.

Like the maintenance interface, the AustLit user interface tightly couples the FRBR model with the presentation layer. As all AustLit data is output as XML, the interface uses an XSL stylesheet to present the data to users: this stylesheet is readily changeable. Once users proceed beyond summary data, all expression and manifestation information is viewed – users do not have the choice of looking at only one expression record. While this seems to have worked very well for AustLit’s purposes, we note with interest a variety of visualisations for FRBR data, including the ‘card catalogue’ and Windows like ‘directory tree’ concepts sketched by Knut Hegna and Eeva Murtomaa²¹. As more FRBR databases are developed, perhaps an optimal ‘OPAC’ representation will be developed. With increasing use of XML and XSL stylesheets, however, individual database owners have the ability to change their presentation layers for local audiences – or even to generate multiple presentation formats for different audience segments – without affecting underlying models or data integrity.

Scalability

Development of the AustLit Gateway required a large number of people from different professions (academics, librarians, bibliographers, programmers, web specialists and graphic designers), and from nine different institutions in two sectors (tertiary education and government) to work together towards quite a ‘grand’ vision. It would be fair to say that all those involved, and their home institutions, and important funding bodies such as the Australian Research Council, regard the Gateway as a major success on a number of fronts. But does AustLit’s success in implementing an FRBR based system mean that other, larger information spaces can be confident about moving forward in this arena? In answering this question, the following factors need to be considered:

- In terms of specialist databases, AustLit is quite large, describing more than 60 000 agents and nearly 400 000 works.
- National bibliographies and large commercial databases often run to millions, tens of millions or even hundreds of millions of records.
- AustLit is a single database, with a single ‘entry point’ for addition and maintenance of data. As a non-holdings database, AustLit does not need to consider the myriad of issues arising from **use** of items.
- National bibliographies already face considerable complexity in facilitating addition of holdings data from individual database owners – a function that is crucial to both efficient use of collecting resources, and inter-library loan functions. This complexity is likely to be multiplied if national bibliographies or union catalogues also need to facilitate addition of expression, manifestation and holdings data to existing ‘work’ records.
- While AustLit’s staff is scattered across a large country, it is relatively small, relatively cohesive and highly knowledgeable about AustLit subject matter.
- The library profession as a whole must ensure cohesiveness of descriptive standards, and are justifiably concerned about the level of ‘variation’ which might appear in more complex, FRBR systems in situations where there is a much lower capacity to enforce uniformity of practice.
- The various sets of legacy data which formed the amalgamated AustLit database were not

standards based, and did not conform to a single set of rules and encodings. This data **had** to be converted in order to continue being used at all, providing both a high incentive and perhaps a unique opportunity to justify the expenditure of significant resources on the conversion process.

- Very large database owners (eg national bibliographies) are very concerned about their huge investment in their legacy data, and whether resources to convert this data are either justified or securable.
- AustLit began with a set of highly articulated research needs, which drove the concepts of providing a single space in which to provide data pathways to, and interface exposure of a set of complex relationships between a range of entities much broader than that encountered in the typical library catalogue. AustLit was also operating within very short, externally imposed timelines: the opportunity to achieve a large vision needed to be grasped.
- The much larger world of the international library profession must necessarily track a slower path in the transition from the traditional card catalogue, to online representations of these catalogues, through to much more sophisticated, navigation oriented structures.

Conclusion

The experience of this relatively small project will certainly not convince large and necessarily conservative organizations to effect such a radical change in their data models and standards. What the AustLit experience does show, however, is that:

- The FRBR model meets a number of sophisticated information needs, especially in subject areas where there is a high need to understand work contexts;
- That database designs to accommodate the model are implementable;
- That a large portion of legacy data can be converted programmatically;
- That legacy data which requires human decision-making can be converted efficiently provided the right tools are provided to staff;
- That professional librarians, indexer and bibliographers can be readily retrained to work within the model, and embrace the model enthusiastically when they can see its benefits;
- That FRBR databases can be fast and responsive;
- That user interfaces can be readily implemented.

and perhaps most importantly, that users of this particular FRBR database find the presentation of information about related works to be both useful and comprehensible.

¹ IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report*, approved by the Standing Committee of the IFLA Section on Cataloguing in 1997, and published K.G. Saur, Munich, 1998. The model is available at <http://www.ifla.org/VII/s13/FRBR/FRBR.pdf> (full model) and <http://www.ifla.org/VII/s13/FRBR/FRBR.htm> (no tables or figures).

² Patrick le Boeuf of the Bibliotheque Nationale, has recently been engaged in creating an FRBR bibliography, which will be a very valuable resource to practitioners. In his recent article 'FRBR and Further', *Cataloging and Classification Quarterly*, Vol. 32 (4) 2001, he notes that by 2001 there were a substantial number of FRBR oriented documents available on the internet, and that a full scale monograph was published on the subject in the same year. FRBR issues have also been a key feature of several major conferences, including the Bicentennial Conference on Bibliographic Control in the New Millennium, hosted by the Library of Congress in November 2000, and the very recent European Library Automation Group's 2002 conference (see <http://www.ifnet.it/elag2002/papers.html>).

³ The University of New South Wales at the Australian Defence Force Academy (lead), University of Queensland, University of Sydney, University of Western Australia, University of Canberra and Monash, Flinders and Deakin Universities.

-
- ⁴ AustLit interoperates with a number of National Library of Australia services, including the National Bibliographic Database for holdings data, PictureAustralia for author photographs and the Register of Australian Archives and Manuscripts for archival holdings.
- ⁵ Funding from the Australian Research Council provided approximately one-third of the development funding, with the university partners providing another one-third in dedicated cash, and all partners providing the remaining third as in-kind resources.
- ⁶ A full description of the genesis of the project, the decision to adopt the FRBR model and the initial service outcomes is available elsewhere, most readily via the papers of the Digital Resources for Research in the Humanities conference held in Sydney, Australia in October 2001: <http://setis.library.usyd.edu.au/drrh2001/papers/ayres.pdf>
- ⁷ See <http://www.austlit.edu.au/browse> for free author information on 1500+ AustLit authors.
- ⁸ AustLit will always be indebted to Dr Judith Pearce, Director of Web Services at the National Library, for pointing us in this direction.
- ⁹ It should be noted that as a subject Gateway, not a holding institution, AustLit has a minimal interest in the 'Item' level of the FRBR model. While the infrastructure does allow staff to record information about, and locations of unique or rare items, this is a rarity. In AustLit, item level information is principally available through the holdings search of the National Bibliographic Database.
- ¹⁰ Recent commentary has highlighted the need to recognize that expressions can give rise to other expressions. This is perhaps most clearly seen in the field of music, but it is possible, for example, that a translation expression of a work could give rise to a second translation expression without the original work ever being consulted. Patrick le Boeuf summarises these suggestions in his excellent 'FRBR and Further', *Cataloging and Classification Quarterly*, Vol. 32 (4) 2001, pp. 15-52, see especially pp. 19-20. The author's own recent recent experience in conceptualizing relationships between pieces of sheet music (often divergent expressions of works) and audio recordings relating to those pieces of music has highlighted this issue.
- ¹¹ See le Boeuf's summary of these calls, p. 24. The AustLit superWork encompassing the novel *Voss* and the opera *Voss* is an example.
- ¹² In addition to extensions described here, AustLit classes works according a a three-tiered typology (workType, formType and genreType), and attributes award relationships to works, expressions and manifestations.
- ¹³ See the free browse pages for Ruby Langford Ginibi at [http://www.austlit.edu.au/run?ex=ShowAgent&agentId=A\(C2\)](http://www.austlit.edu.au/run?ex=ShowAgent&agentId=A(C2)), and Patrick White at [http://www.austlit.edu.au/run?ex=ShowAgent&agentId=A\)\]m](http://www.austlit.edu.au/run?ex=ShowAgent&agentId=A)]m), for examples of these agent pages.
- ¹⁴ See <http://www.infoloom.com/tmsample/bie0.htm>
- ¹⁵ See <http://www.austlit.edu.au:7777/design/index.html>. Our underlying database design is not available to the public. The AustLit Staff Manual is freely available at <http://www.austlit.edu.au/common/manual/WorksContents.html> and the AustLit Thesaurus is available at: <http://www.austlit.edu.au/run?ex=ShowThes>.
- ¹⁶ At the time of writing, the encodings available through the AustLit interface include the HTML default, the XML schema in all its complex glory; a plain text representation, and an encoding designed for rapid export of complex AustLit records into simple and 'flat' Endnote bibliographic databases.
- ¹⁷ And, it must be acknowledged, since quality management is the role of AustLit's Content Managers, who are also heavily involved in many other aspects of AustLit's development, checking of records has not been nearly as extensive as desired.
- ¹⁸ See Hegna, Knut and Murtooma, Eeva 2002, 'Data mining MARC to find: FRBR?', available at <http://folk.uio.no/knuthe/dok/frbr/datamining.pdf>, for vivid descriptions of the differential cataloguing practice they encountered in their study of key literary works!
- ¹⁹ Using the tried and tested method of trying it out on friendly family members first.
- ²⁰ As the AustLit team anticipated a need for significant usability testing, funding applications included requests for testing resources. While the funding body declined to fund usability testing, the AustLit consortia aims to undertake such testing using its own resources in the near future.
- ²¹ Hegna, Knut and Murtooma, Eeva 2002, 'Data mining MARC to find: FRBR?', available at <http://folk.uio.no/knuthe/dok/frbr/datamining.pdf>, pp. 31-34. A number of presentation possibilities are also canvassed in the Library of Congress' 'Displays for Multiple Versions from MARC 21 and FRBR', (see <http://www.loc.gov/marc/marc-functional-analysis/multiple-versions.html>). This work follows on from Tom Delsey's monumental *Functional Analysis of the MARC 21 Bibliographic and Holdings Formats*, published in January 2002 and revised March 2002 (See <http://www.loc.gov/marc/marc-functional-analysis/home.html>).