From WEMI to WI to WII: FRBR, BIBFRAME and the 21st Century Bibliographic Model

Anne Welsh, Lecturer in Library and Information Studies, University College London

Introduction

This short communication summarises some of the ideas shared at CIG 2016 (Welsh 2016b, Welsh, 2016c) regarding the current state of art in the bibliographic model used by the Anglo-American cataloguing community. It highlights our trajectory from the 'monolithic' catalogue record of the card and dictionary catalogue era, through the hyperlinked computer-based records of the 20th century, towards models that focus on data that can be reused and repurposed within the 21st century linked data environment.

The Four Models

As presented elsewhere (Welsh et al, 2016), in order to fully understand modern cataloguing and meet the needs of employers, cataloguers need to have a working awareness of the four bibliographic models utilised in the Anglo-American cataloguing community (Table 1).

Card / Dictionary Catalogue	Paper-based system; entry points and length of description limited by concerns around physical space (size of card / page and size of catalogue cabinet / shelving unit). Each record describes the item in the cataloguer's hand entirely – its intellectual and physical contents.
ISBD and AACR in MARC	ISBD, AACR and AACR2; built on principles of the card and dictionary catalogue; monolithic record structure in which information in fields is contextualised and explained by reading the entire record; inherited the card / dictionary catalogue's description level, covering both the intellectual and physical contents of the item in hand.
FRBR in RDA	Although much is inherited from ISBD, AACR and AACR2, the principle of RDA is a version of FRBR's WEMI model: Works, Expressions, Manifestations and Items are catalogued separately and linked to each other (and to the appropriate WEMI of other records); moves away from the monolithic record, favouring instead field-level links between records, aiming for the linked data environment. Currently constricted by systems based on the MARC format, which does not accommodate RDA's linked data capacities.
BIBFRAME	Linked data solution proposed by the Library of Congress; based on RDF; data model proposes Work, Instance, and Item, allowing for simplified analysis on the part of cataloguers and, significantly, the construction of RDF triples – a foundation for linked data in the wider web environment. Many triples in BIBFRAME incorporate Annotations – pieces of data that provide information about the Work or Manifestation, such as creator, publisher or ISBN.

Table 1. Cataloguing Models Anglo-American cataloguers learn in 2016, updated from Welsh et al, 2015 to reflect BIBFRAME 2.0

Given the uptake of *Resource Description and Access* (RDA) in the UK, cataloguers are now quite familiar with its implementation of the FRBR model, and even although it cannot be fully exploited in MARC, its concepts of the Work which is realized through an Expression which is embodied in a Manifestation which is exemplified in an Item have fed through into our cataloguer jargon, for better or for worse. Figure 1 presents an entity relationship diagram for Walter de la Mare's (1930) *Desert Islands and Robinson Crusoe* now in the collection of Cambridge University Library (Classmark CCB.49.1), a WEMI example originally discussed in an article in the National Acquisition Group's journal *Taking Stock* (Welsh and Wright, 2016).
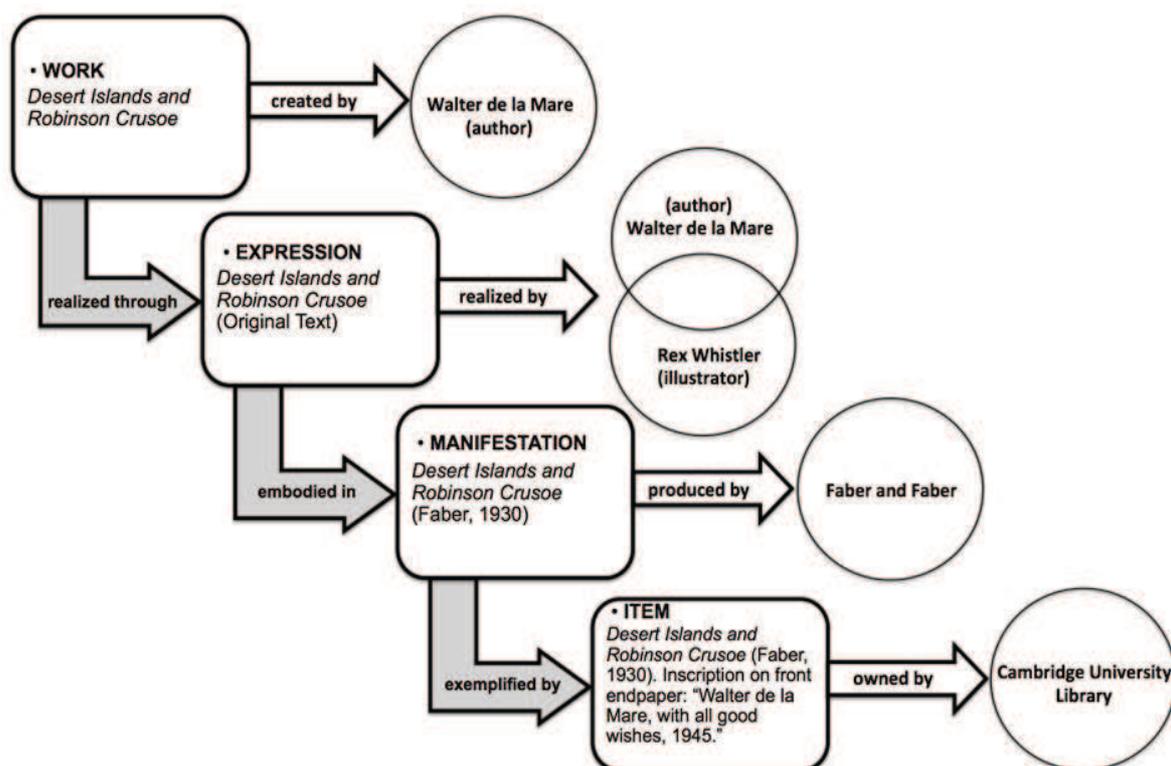

Figure 1. Group 1 and 2 entities and relationships for *Desert Islands and Robinson Crusoe* (Faber, 1930), Cambridge University Library CCB.49.1, originally published in Welsh and Wright, 2016.

The diagram demonstrates the most common relationships between "Group 1 entities" (Works, Expressions, Manifestations and Items) and "Group 2 entities" (people, corporate bodies and families):

- A Work is created by a person, corporate body or family (in this case, the person Walter de la Mare)
- An Expression is realized by a person, corporate body or family (in this case the persons Walter de la Mare and Rex Whistler)
- A Manifestation is produced by a person, corporate body or family (in this case the corporate body Faber and Faber)
- An Item is owned by a person, corporate body or family (in this case Cambridge University Library).

Relationships at Heart

In a bibliographic model which places "relationships at the heart of the catalogue" (Welsh and Batley, 2012, 8), thinking about the people or organisations responsible for what we are describing can help us to gain a firmer idea in our own minds of which Group 1 entity we have.

In this example, Rex Whistler was commissioned to illustrate de la Mare's work, and so his involvement was in its realization – therefore, at the Expression level. However there are times in which a writer and an artist work together to create a Work, or in which the creators are both artists and writers. Figure 2 shows the WEMI relationships for *A Child of Books* (Jeffers and Winston, 2016), whose co-creator Sam Winston was our after-dinner speaker at CIG 2016.
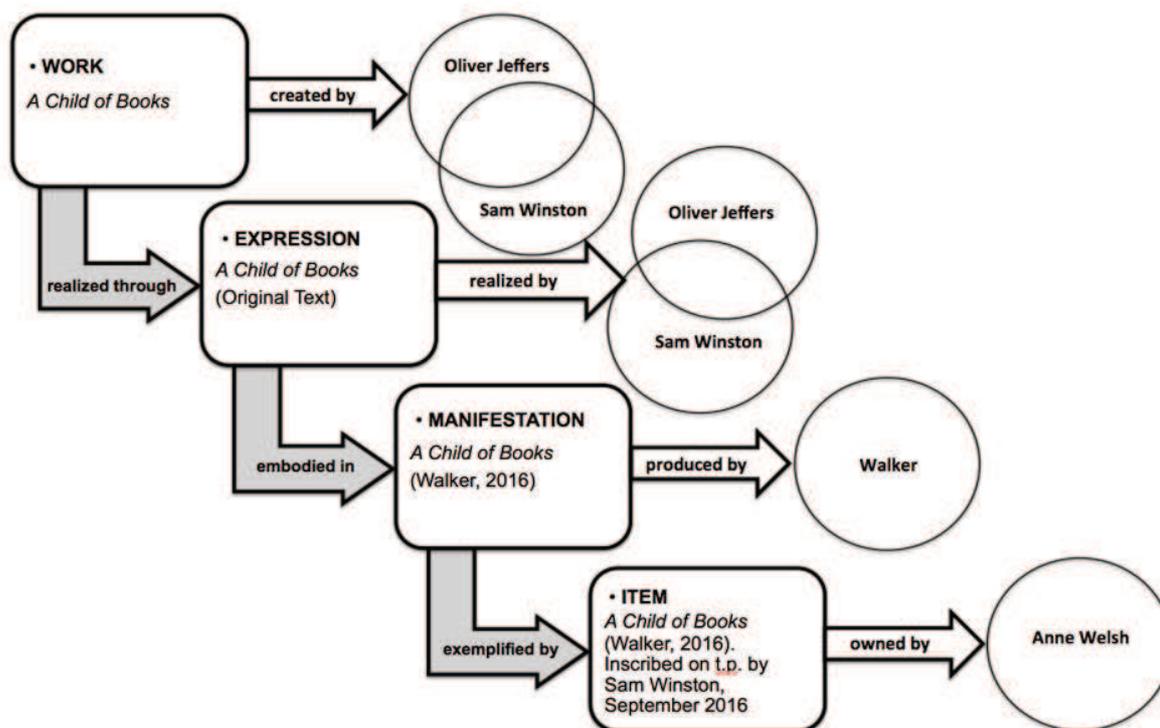

Figure 2. Group 1 and 2 entities and relationships for *A Child of Books* by Oliver Jeffers and Sam Winston, inscribed by Sam Winston to Anne Welsh.

As cataloguers, we have to take what appears on the title page as our chief source of information, and in the case of *A Child of Books*, this is easy:

245 12 $aA child of books /$cOliver Jeffers, Sam Winston

Similarly for *Desert Islands and Robinson Crusoe*:

245 10 $aDesert islands and Robinson Crusoe /$cby Walter de la Mare ; with decorations by Rex Whistler

So, in the purely practical terms of what we "do" with the statement of responsibility, both are equally straightforward. However, when we think about the WEMI, there are clear differences. Rex Whistler was a well-known artist and book illustrator, and his work with de la Mare came under his practice in illustration – that is taking the words that someone else (in this case, de la Mare) had written already and providing artwork to accompany them. We could, of course, go through the publisher archives to discover the exact nature of the contribution of each person, but, at base level, the Work consists of a text by de la Mare, and Whistler's decorations form part of its Expression. We can think of other Works that exist in this way – Michael Bond's character Paddington Bear was inspired by a teddy bear he bought for his then wife, famously illustrated by Peggy Fortnum in the original books in the 1950s and 1960s, then by Fred Bandery in the Young Readers series from the 1970s, before being transformed into a cartoon by Ivor Wood, who also created the puppet used in the 70s TV series.

Later illustrators include David McKee and Barry Macey in the 1980s and RW Alley from the 1990s. For most of the publications featuring Paddington, at Work level Michael Bond is responsible, and then the illustrators become involved in its realization through different Expressions. Paddington could, in fact, be the subject of an entire workshop and article on his own. For those interested in a quick summary of the famous bear's illustrated evolution, *The Guardian* shared a gallery of artwork to celebrate Bond's 90th birthday in January 2016 ('Paddington through the Ages,' 2016).

Relationships between People, not Solely Resources

Jeffers's and Winston's collaboration is interesting for artistic reasons, but also because it is well documented through interviews with them. As an article in The Telegraph (Calkin, 2016) outlined, "When Jeffers and Winston were introduced … they decided to collaborate, despite – or perhaps because of – their different styles. They recognised something in each other: both have a love of literature rendered as art." Jeffers is quoted as saying, "I thought something interesting would happen if we put my drawings together with Sam's landscapes," while Winston is reported as sharing that, "It was a dreamy kind of idea to start with … It could be a book, a sculpture, a set of paintings … we were playing around - on the shoreline where ideas come from."

Here, we can really see that in thinking about how the bibliographic entities of WEMI and creators relate to each other, we are not solely thinking as people tasked with data entry for a book. We are also engaging with ideas around the relationships between people and the intellectual and artistic outputs they are creating. As we move "down" the WEMI model through the Manifestation level, we are also thinking about and documenting the relationship between the book and its publisher, and, through the book, between the creators and *their* publisher – in this case, Jeffers's and Winston's Work is realized *by them* in its Expression, which is embodied in a Manifestation produced by *their publisher, Walker*. In the final level of WEMI, another relationship is represented – Jeffers's, Winston's and Walker's Manifestation is exemplified by an Item that is owned by me.

Because there is an inscription, my relationship to the book is documented: it is clear that I have met Sam Winston at some point because he has inscribed the title page, whereas Oliver Jeffers has not. There are limits to the level of revelation this provenance evidence provides. There is a date, but nothing to indicate what we know at CIG – that Sam Winston signed the book at our conference, at which he spoke after dinner and where Walker had a stall. Nor can there be anything to indicate that Karen Pierce and I had, in fact, met Winston when he provided a keynote at the Livres d'artistes conference in Cardiff the previous December and that we suggested to the CIG conference committee, of which Karen was part, that they might invite him to speak after our conference dinner. However, the recording of the basic provenance information could set a scholar on the road to such a discovery, should anyone ever wish to trace, for example, the history of the Cataloguing and Indexing Group conferences in the 20th and 21st centuries – a suggestion that does not seem as flippant as it first sounds when I look at some of the work I am currently doing around the institutional history of UCL Department of Information Studies as it approaches its centenary in 2019.

Here we can see that final relationship, chronologically, between the researcher and the materials they seek – often in pursuit of information about a literary or artistic creator – is, as argued vociferously elsewhere (Welsh, 2016d), facilitated by the intellectual work of the cataloguer.

From Records to Data

In recording information about relationships between people and resources and through this between people and other people, we are meeting the challenge of going beyond what FRBR has called "the elementary uses that are made of the data by the user" (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998). In establishing intellectual links of the kind WEMI forces us to consider, we are also moving beyond what Wilson (1968) termed "descriptive power" to "exploitative power," defined by Smiraglia (2008) as "the power of a scholar to make the best possible use of recorded knowledge,"

and enabling us to meet what I have described as "the needs of researchers who want to engage with our data not [solely] as a route through to 'the real' objects of their research – full-text files, books, the item for which the catalog data is a surrogate – but as an integral part of their own research" (Welsh, 2016d).

Linked data is powered by relationships. As Godby and Smith-Yoshimura (2017, 18) put it in their recent practical article on 'Managing the Transition from Legacy Library Metadata to Linked Data,' "To make linked data work, the library community needs good data that is structured, unambiguous and published in a format that enables linking with data produced by other communities. Library data also needs to be more about the *Things* or the people, organizations, places and topics that users care about and the library community has something to say about" (authors' emphasis). In short, in entering the linked data environment, we are moving from the monolithic catalogue record, in which most fields and sub-fields are human-readable and dependent on being read within the context of the record as a whole, to models in which we present our data in ways that can be combined and recombined; in which each field is not dependent on another field in order to be understood; and in which many of the key pieces of data are not human readable, but machine-readable. In the Internet of Things, each Thing has a Unique Resource Identifier (URI) to which those who wish to reuse that data can point using semantic web specifications such as RDF (the Resource Description Framework) in order to describe relationships between that Thing and any other Thing with a URI. RDF is formed using 'triples' in the format shown in Figure 3.
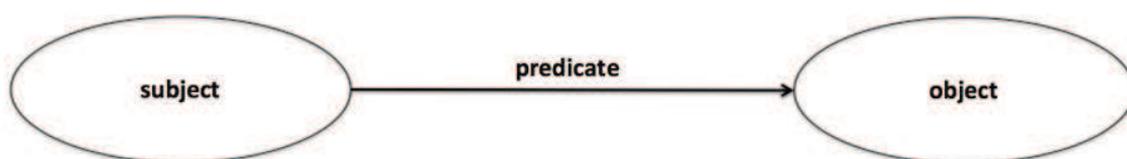

Figure 3. Standard format of a triple in RDF

BIBFRAME Triples

The format being developed by the Library of Congress to replace MARC is BIBFRAME (the Bibliographic Framework Initiative). Built on RDF, it describes relationships between data using RDF triples in the format exemplified in Figure 4. Where RDA follows FRBR in using the WEMI model, BIBFRAME has a simplified version, based initially on only two levels, Work and Instance, and from April 2016, when BIBFRAME 2.0 was announced, three levels, Work, Instance and Item. The easiest way to visualize the mapping between the FRBR in RDA and BIBFRAME models is to think of their Work and Item levels as being equivalent and the Instance level of BIBFRAME equating to both Expression and Manifestation levels in FRBR in RDA. Figure 4 shows the BIBFRAME triple that links data at the Work level to data at the Instance level. Figures 5 and 6 show this triple 'in action' with regard to *Desert Islands and Robinson Crusoe* and *A Child of Books*.

Figure 4. Example of a BIBFRAME triple showing the subject (WORK) with predicate (hasInstance) and object (INSTANCE)

Figure 5. Example of a BIBFRAME triple in which the subject is the Work *Desert Islands and Robinson Crusoe* and has the predicate "hasInstance" and the object is the Instance published by Faber in 1930.
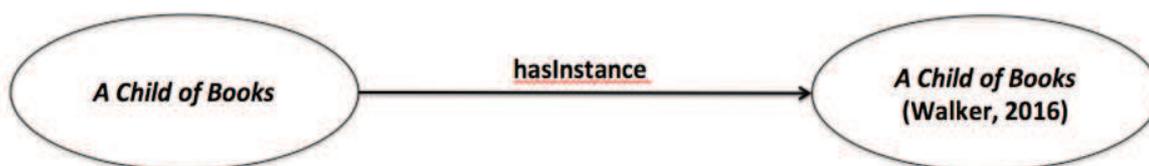

Figure 6. Example of a BIBFRAME triple in which the subject is the Work *A Child of Books* has the predicate "hasInstance" and the object is the Instance published by Walker in 2016.

Using triples defined by BIBFRAME, it is possible to model a web of bibliographic data, in which we can map the relationships between people and resources. One way to create such a web quickly and easily with minimum use of technology is to work in small groups and build it from post-it notes. Figure 7 shows part of such an exercise undertaken by the Advanced Cataloguing class at UCL in February 2017, using the same Jane Austen and Fay Weldon examples we used in workshops at CIG 2016. The UCL example is being shown in preference to the CIG one only because the post-its I was able to acquire were narrow and so could indicate a vague form of direction and had different coloured edges allowing for a loose form of colour coding, so that even though it is difficult to read the writing in the photograph here, it is possible to gain a sense of the web of connections that is forming.
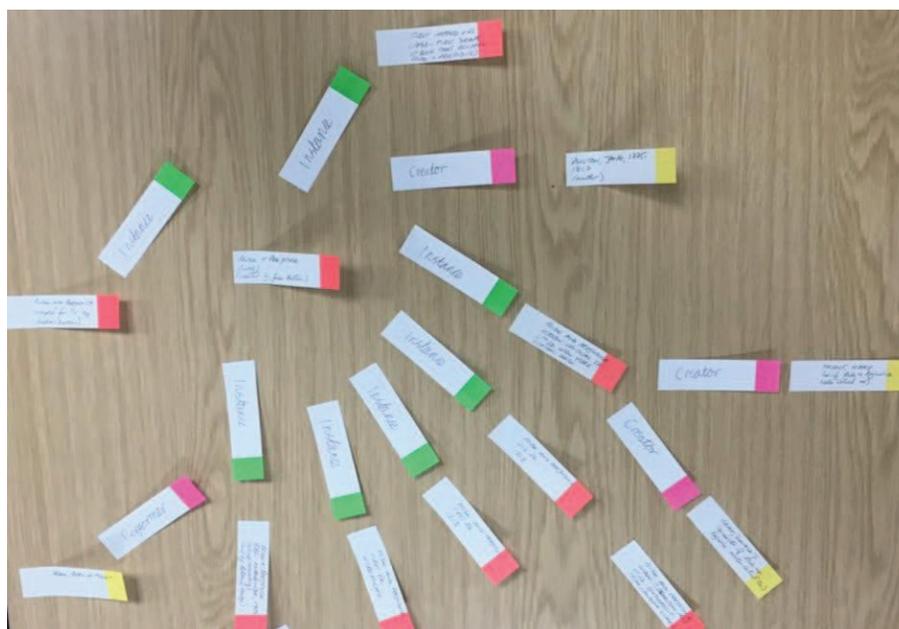

Figure 7. Post-its arranged by students in INSTG005 Advanced Cataloguing in 2017 to plot relationships between some Works, Expressions, Manifestations and Items by Jane Austen and Fay Weldon.

From *Pride and Prejudice* to *The Cloning of Joanna May*

I wrote examples of BIBFRAME Works, Instances and Items on the orange-tipped post-its. The students used the green-tipped post-its to indicate Instance relationships (BIBFRAME's hasInstance predicate) linking these entities. I wrote the names of people involved in the creation or production of the Works, Instances and Items on the yellow-tipped post-its and the students indicated the type of relationship (e.g. BIBFRAME's creator) on the pink-tipped post-its. The post-it diagram we created covered two small desks, and Figure 7 shows an excerpt from it.

Having created the diagram, it was possible for students (and participants at CIG 2016 before them) to see how one could trace a journey through different relationships from the Work *Pride and Prejudice* to the unrelated Work *The Cloning of Joanna May*. In the set of data I had provided on the post-its, the point at which a link appeared was through the actor Patricia Hodge, who read both the HarperCollins Audio recording of *The Cloning of Joanna May* (Weldon, 1996) and the BBC Worldwide recording of *Pride and Prejudice* (Austen, 1999). Clearly, there is no direct relationship between the two Works, but it is fairly obvious how BIBFRAME (or another RDF schema) allows us to plot relationships between people and objects that ultimately do form links. It is also clear that someone looking for material could enter the web we created at different points and still go on a journey – they might, for example, come in with a specific known item search for the Norton Critical Edition of *Pride and Prejudice* (Austen, 2016) and leave having also discovered by following the linked data that Austen's first draft of the book was actually called *First Impressions* (Austen, 1797). Having been reputedly disposed of by the publisher Caddell to whom Austen's father had sent it, the 1797 manuscript of *First Impressions* is of interest within our bibliographic model in that it exists only as a Work that is related to the Work *Pride and Prejudice*. It does not have an Instance, nor, indeed, an Item.

Items Matter

One question that I anticipated being asked by colleagues at CIG 2016 who have been following the development of BIBFRAME, and that I was indeed asked, was why BIBFRAME 2.0 was needed and what impact it might have on our practice. My poster on 'Provenance in the Bibliographic Model' (Welsh, 2016c) considered one of the major improvements made to BIBFRAME by the new version released in April – the introduction of the Item level and the predicate custodialHistory that can be used to form a triple in the form shown in Figure 8.

Figure 8. Example of a BIBFRAME triple showing the subject (ITEM) with predicate (custodialHistory) and object (custodian).

Figure 9 shows this triple for Cambridge University Library's copy of *Desert Islands and Robinson Crusoe*.

Figure 9. Example of a BIBFRAME triple showing the subject (the Item *Desert Islands and Robinson Crusoe* (Faber, 1930) inscribed on front endpaper: "Walter de la Mare, with all good wishes, 1945") with predicate (custodialHistory) and object (Cambridge University Library).

Figure 10 shows this triple within a web of some of the triples that could be created in BIBFRAME in order to describe the resource. You can see that in BIBFRAME both the author (de la Mare) and the illustrator (Whistler) are indicated by "agent" but that de la Mare's relationship is at the Work level, while Whistler's is at the Instance level.

The significance of the introduction of BIBFRAME's custodialHistory relationship is that it should be possible to record the chain of custody of objects within collections. Figure 11 shows how it might be possible to show the known history of the Cambridge University Library copy of *Desert Islands and Robinson Crusoe*. The copy clearly started in the hands of de la Mare, as indicated by his inscription. At this point, we don't know if it was a gift to Phyllis T.M. Davies or to another person and Davies acquired it through purchase, but we do know that she was the person who owned it before Cambridge University Library obtained it as part of her collection, and we know that the copy is still sitting in the tower at the UL.
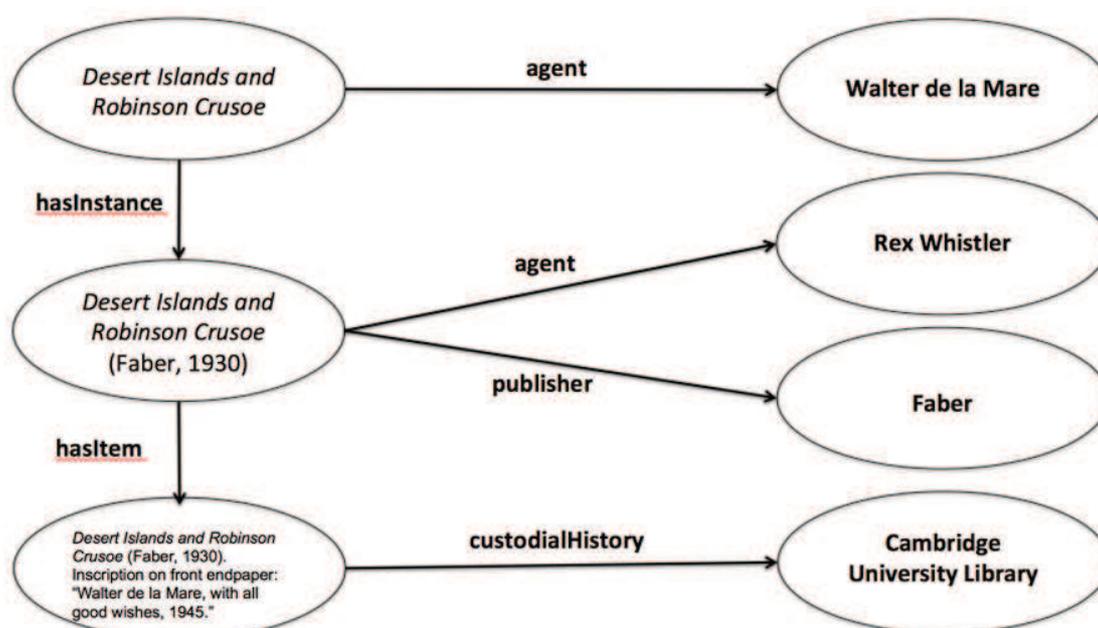

Figure 10. Example of BIBFRAME triples showing some of the relationships we can trace for Cambridge University Library's copy of *Desert Islands and Robinson Crusoe*.
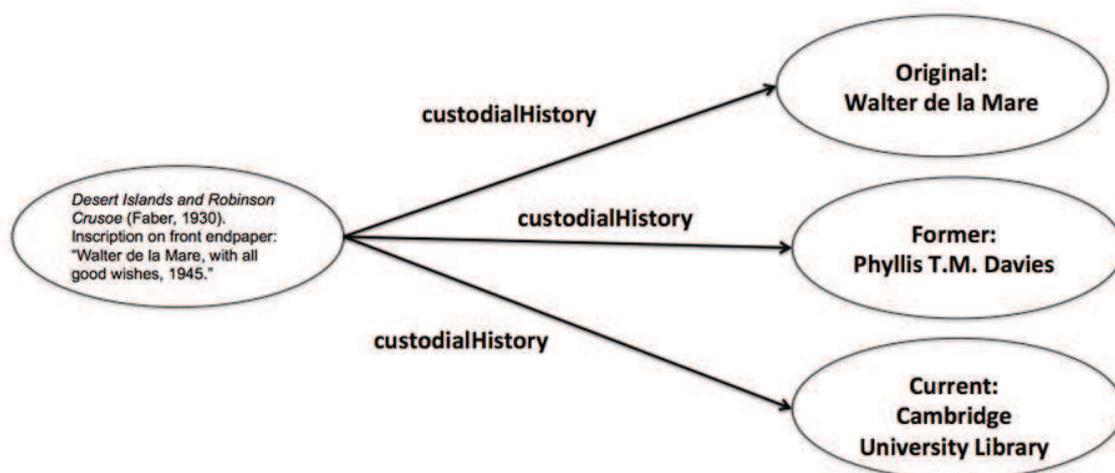
Figure 11. Example of BIBFRAME triples showing some of the chain of custody for Cambridge University Library's copy of *Desert Islands and Robinson Crusoe*.

From this it should be clear that the Library of Congress in creating BIBFRAME is attempting to position library data not only in a way that is more compatible with other forms of heritage data in the linked environment, but that allows cataloguers to record information that we know to be of interest to researcher (such as provenance information) in a way that will be easy to reuse. This element of cataloguing is contributing to the history of the objects within our collections and helping researchers to make connections quickly and easily, based on our intellectual efforts. In this way, a cataloguer can be seen as a "distant research collaborator" (Welsh, 2016d), assisting in the task Wilson (1968) highlights as "exploitative power," making "the best possible use of recorded knowledge" (Smiraglia, 2008).

For current cataloguers, there is no rush to implement these ideas. BIBFRAME is still evolving, and, therefore, is some way from being introduced within the Library Management Systems that we use to create and manage our data. At this stage, the need to engage with these ideas is preparatory – it's a good idea for us to be aware of the changes we will manage in the next few years, and, as always, it is helpful to be able to contribute to discussions with our management teams, our library vendors and with those who are creating the standards, rules and, in the case of BIBFRAME, the exchange formats that we will be expected to use. While computer scientists know more about data models and theorists can expound on the Book (and other formats) and its fundamental role and relationships, it is the cataloguer sitting quietly at their desk creating data day in and day out who knows better than anyone the full range of quirks and joyous oddities that challenge not only their own decision making skills, but the full capacity of our standards and formats created at the highest levels internationally. It is the everyday cataloguer who wields Wilson's (1968) "two kinds of power" – "descriptive" and, increasingly, "exploitative."

Acknowledgements

Works Cited

Austen, J. (1797) *First Impressions.* Unpublished manuscript.

Austen, J. (1999) *Pride and Prejudice*. Sound recording. (London: BBC Worldwide).

Austen, J. (2016) *Pride and Prejudice*. Norton Critical Edition edited by Donald, D.J and Favret, M.A. (New York: Norton).

Calkin, J. (2016) 'A Child of Books: Making its Mark in Children's Literature.' *The Telegraph* (1 September), http://www.telegraph.co.uk/books/childrens-books/a-child-of-books-making-its-mark-in-childrens-literature/

de la Mare, W. (1930) *Desert Islands and Robinson Crusoe* (London: Faber).

Godby, C.J and Smith-Yoshimura, K. (2017) 'From Records to Things: Managing the Transition from Legacy Library Metadata to Linked Data.' *Bulletin of The Association for Information Science and Technology* 43(2): 18-23.

IFLA Study Group on the Functional Requirements for Bibliographic Records (1998) *Functional Requirements for Bibliographic Records Final Report* (Munich: K.G. Saur).

Jeffers, O. and Winston, S. (2016) *A Child of Books* (London: Walker).

'Paddington through the Ages: in Pictures.' (2016) *The Guardian* (13 January), https://www.theguardian.com/childrens-books-site/gallery/2016/jan/13/paddington-through-the-ages-michael-bond

Smiraglia, R.P. (2008) 'Rethinking What We Catalog: Documents as Cultural Artifacts'. *Cataloging and Classification Quarterly* 45: 25-37.

Weldon, F. (1996) *The Cloning of Joanna May*. Sound recording. (London: HarperCollins Audio).

Welsh, A. (2016a) 'Cataloguer as Distant Research Collaborator: Implications of the Use of Catalogue Data in Humanities Research.' Presented at: Emerging Research in Library and Information Science, London, 19 August 2016, https://iris.ucl.ac.uk/iris/publication/1154530/1

Welsh, A. (2016b) 'Getting to Grips with Bibliographic Models.' Presented at: Cataloguing and Indexing Group Conference, Swansea, 31 August – 2 September 2016, http://www.cilip.org.uk/cataloguing-indexing-group/presentations/conference-2016-innovation-discovery

Welsh, A. (2016c) 'Provenance in the Bibliographic Model.' Presented at: Cataloguing and Indexing Group Conference, Swansea, 31 August – 2 September 2016, http://www.cilip.org.uk/cataloguing-indexing-group/presentations/conference-2016-innovation-discovery

Welsh, A. (2016d) 'The Rare Books Catalog and the Scholarly Database.' *Cataloging and Classification Quarterly* 54(5-6): 317-337.

Welsh, A., Bikakis, A, Garea Garcia, N., Mahony, S., Inskip, C. and Vogel, M. (2015) 'Work in Progress: The Linked Open Bibliographic Data Project.' *Catalogue and Index* 178: 16-21.

Welsh, A. and Batley, S. (2012) *Practical Cataloguing: AACR, RDA and MARC 21* (London: Facet).

Welsh, A. and Wright, J. (2016) 'RDA for Acquisitions Staff.' *Taking Stock* 24(2): 14-16.

Wilson, P. (1968) *Two Kinds of Power: An Essay on Bibliographical Control*. Berkeley: University of California Press.