IMPLEMENTATION OF RDA IN HUNGARY

Szabina Ilácsa*

INTRODUCTION

In 2015, following the decision of the National Library Standardisation Committee, National Széchényi Library launched a project to implement Resource Description and Access (RDA) as the national cataloguing code. This article gives an insight into the process showing how many layers it has. It also provides details about the Hungarian approach to the implementation.

Our position is somewhat special. Our decision to implement RDA happened the same time as RSC decided to launch the 3R project. We were given the instruction not to translate RDA straight away, because it will be rewritten. Instead, we can start with translating the International Cataloging Principles (ICP) and The IFLA Library Reference Model (LRM). The benefit of translating these two documents beforehand is that we could examine the new terminology in a more real-life context. Therefore, our new terms could be used versatilely in sentences. We also translated MARC 21 Bibliographic, Authority and Holdings formats. The terminology of the formats, the two basic documents and RDA was kept in alignment.

RDA AS BIBLIOGRAPHIC FRAMEWORK

Its own definition says the following: "*RDA is a package of data elements, guidelines, and instructions for creating library and cultural heritage metadata that are well-formed according to international models. These metadata are intended to support the discovery and identification of resources in library and other cultural heritage collections.*"[1]

One of the main benefits of "new" RDA is that it is based on IFLA LRM[2] and by implementing it we can take advantage of the Work-Expression-Manifestation-Item level elements, and we can create search and other functions on them. Another advantage is its connection to linked data technology. The element structure and applying certain options give us an opportunity to record our metadata in a way that can be read automatically by computers. The first "edition" of RDA was often criticized for being too traditional and stuck in the past.[3] Post 3R RDA is criticized for going too far. This is not for the cataloguer any more but for machines.[4]

RDA is the successor to Anglo-American Cataloguing Rules, Second Edition (AACR2). Although largely based on AACR2, RDA's instructions are more general to

_____
* National Széchényi Library, Budapest
  e-mail: ilacsa.szabina@oszk.hu

the point that you cannot use it as it is for cataloguing purposes. There are simply too many options and not enough detailed instructions. So, in my opinion, RDA should not be referred to as a cataloguing code but as a bibliographic framework. I personally do not mind this framework feature of RDA. Having to write your own application profile(s) and policies makes you reconsider your practices.

Why do we record this element this way? Does it serve a purpose or is it just a tradition? Is it recorded in the most usable way? And the list of questions goes on.

It is entirely the implementing agency's decision what it wants from RDA. If you want to implement RDA just to take advantage of some of its value vocabularies, you can do that. If you implement it because you want to FRBRize your catalogue, you can do that. If you want your library metadata available as linked data, you can use it for that as well.

This is what I meant when I said that implementing RDA has many layers. I think this is the real gain of the implementation. It makes us rethink our whole cataloguing process (the workflows, the display etc.)

First, let's see those issues we face irrespectively of the level of implementation.

TRANSLATION

Assuming your native language is not English you have to translate RDA at least partially. Partial translation means you only translate the RDA Reference which contains the element sets and value vocabularies.

The RDA reference and the Toolkit itself are written in very technical English. We, in Hungary, chose to conduct a full translation instead of the other possibility, which involves translating the Reference and writing the application profile and the policy in Hungarian. The reason why we chose this option is that we would like to support cataloguers in Hungary who want to have a better understanding of RDA even if they do not speak English very well or they do not speak it at all.

Translating RDA Reference is not an easy task. Sometimes the biggest headaches are caused by other things than those you would expect. The "telegram style" – which is quite typical for standards – does not give you much room to explain things if the target language does not have the same structure or the same semantic field for a word as the original. This is how it can happen that translating a mundane phrase like "over time" can take so long to solve. Hungarian simply does not have a phrase that covers the same meaning namely that it happens at a period of time step-by-step. RDA prefers

BIBLIOGRAFIA

using passive voice over active which is also very problematic for us, because Hungarian lacks passive voice and many times the context does not make it clear which is the subject of the sentence.

REWRITING OUR CURRENT CATALOGUING CODE ACCORDING TO RDA

As I mentioned before, you cannot use RDA for cataloguing as it is because it gives you many options on how to record data but still lacks the detailed instructions cataloguers are accustomed to have. You have to add your own application profiles and policies to the Toolkit in order to use it for everyday cataloguing. An application profile contains what elements to use in the description, which recording method to use and if the value of the element comes from a value vocabulary, meanwhile policy contains the detailed instructions RDA often lacks.

At the outset, it was proposed that each community should develop its own application profile, but EURIG (*European RDA Interest Group*) was of the opinion that it would be more beneficial if a common or base application profile were developed.

An application profile typically specifies:
- the elements to be recorded as a metadata description set for an entity,

- the mandatory and repeatability status of each element,
- the vocabulary encoding scheme to be used as a source of data for an element,
- the string encoding scheme to be used to assemble or derive the data for an element.

An RDA application profile may specify additional parameters that are unique to RDA:
- the recording method to be used for an element, where a choice is available,
- the optional instruction to be applied to an element,
- the policy statement to be applied to an element[5].

If you visit any of the element pages in the Toolkit you will see a definition for the element. The prerecording section gives you additional information about what that element includes or excludes. Then you have to make your first choice. What recording method to use. Our choice of recording methods will limit what we can do with the data later on. So, it is a very important decision.

At this stage, you have a crystal clear vision of how you want to implement RDA. Otherwise, you will not be able to make the right decisions. By right decisions, I mean the option which serves your concept the best.

There are 4 recording methods: unstructured description, structured description, identifier or IRI.
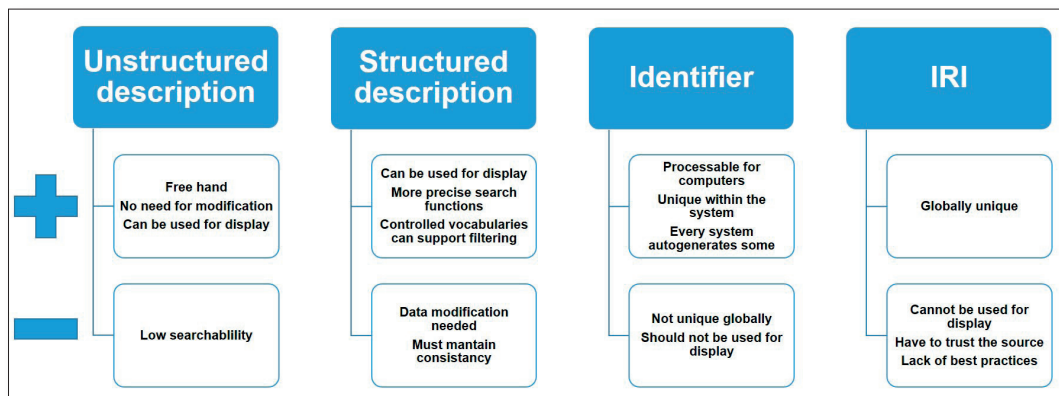

Figure 1

AN UNSTRUCTURED DESCRIPTION

This term is basically just a fancy new name for "record how the resource represents itself". An unstructured description of an RDA entity is a string that is a kind of Nomen entity. The recording always involves some kind of transcription.

Kinds of unstructured description include:
- a manifestation statement,
- an unstructured note,
- a name or title in direct order, as it appears in sources of information,

- an uncontrolled term for a concept[6].

Advantages of this recording method:
- recoding how the resource represents itself helps the user to identify and/or select a resource,
- it gives you a free hand; certain elements have too diverse values to record any other way,
- you can record the data without any modification,
- you can use the data for display…

Shortcomings of this recording method:
- …in fact it is the only thing you can use it for; searchability is provided only at the keyword level.

A STRUCTURED DESCRIPTION

This method is where „*the result is a string, but it is not a direct transcription from the resource but it involves some kind of manipulation*". This manipulation can be instructions on how to formulate the string (in RDA terminology: using a string encoding scheme) or choose a term from a controlled vocabulary (in RDA terminology: using a vocabulary encoding scheme).

Kinds of structured description include:
■ an access point,
■ a structured note,
■ a name or title taken from an authority control system,
■ a term for a concept taken from a controlled vocabulary,
■ a value associated with a structured data type, including numbers, dates, and times[7].

Advantages of this recording method:
■ you can use the data for display,
■ it supports more precise search functions than the unstructured description,
■ if the data value is chosen from a controlled vocabulary it can be used for filtering.

Shortcomings of this recording method:
■ you have to modify the data to some extent,
■ you have to make an effort to maintain consistency.

AN IDENTIFIER

An identifier is still a string but unlike the case of the unstructured description, it does not need to be meaningful to a human being. It can only be associated with a vocabulary encoding scheme.

Kinds of identifier include:
■ an identifier assigned by an independent – external agent,
■ an identifier assigned by a local agent,
■ a notation for a concept taken from a controlled vocabulary[8].

Advantages of this recording method:
■ an identifier is more processable for computers than a string,
■ it is and will be unique within the system (in contrast with an access point to which additional information might be needed to add later on in order to maintain its uniqueness within the system),
■ your system automatically generates some identifiers.

Shortcomings of this recording method:
■ although it is unique within the system, it is not unique globally,
■ it should not be used for display, but sometimes we do it anyway, for example, the MARC code for languages.

IRI

IRI (*Internationalized Resource Identifier*) is an identifier based on Semantic Web technologies used as the referent of an entity or controlled term in linked open data using Resource Description Framework.[9]
Advantages of this recording method:
■ it is unique globally.

Shortcomings of this recording method:
■ it is created and maintained outside your system, you do not have control over it, you have to rely on the source,
■ recording IRI is still new to the library world; there are no many best practices for it,
■ it cannot be used for display.

To illustrate how it looks like in practice, let's see how we would record Edinburgh as a place of publication using the different recording methods.

Recording methods - Review
RDA element: place of publication

Data Source	Recording Method	Example
Unmediated transcription	unstructured	"EDINBVRGH"
Mediated transcription	unstructured	"Edinburgh"
Authorized access point	structured	"Edinburgh, Scotland"
Getty Thesaurus of Geographic Names Identifier	identifier	"7009546"
LC/NACO Authority File Control Number	identifier	"n79053770"
DBPedia URI	IRI	http://dbpedia.org/resource/Edinburgh
GeoNames URI	IRI	http://sws.geonames.org/3333229/

Figure 2[10]

BIBLIOGRAFIA

MARC

So far we have had the guidelines in the national language and have had the rules. But we also need an exchange/communication format which also supports storing our metadata in a sufficient way. That, in the current library world, means some flavour of the MARC format.

If you decide to read about the future of the MARC format, the titles of the articles will add up to a quite disturbing search history. Let's walk through the creepy-titled articles to illustrate how our attitude towards MARC changed in this century.

1. "MARC must die"
In 2002 Roy Tennant in his article in the *Library Journal listed* several reasons why "*the very nature of a MARC record is, to some degree, an anachronism*" and it "*outlived its usefulness*".[11]

2. "MARC isn't dying fast enough"[12]
This one is from more than 10 years later. There were some promising initiatives but everything happened so slowly. Although FRBR had been first published in 1998 it did not have much effect on the practical level of cataloguing until the early 2010s when several promising initiative started. For example, the publication of the FRBR-based cataloguing code RDA, the semantic web conversion of the Swedish National Library, the development of BIBFRAME started.[13]

3. "If MARC dies, it will be through obesity"
A quote is by Sally McCallum, she said it at the 2017 Midwinter OCLC Linked Data Roundtable. She referred to the process of the immense modification of the MARC 21 format so it would be able to be used to properly encode RDA elements. Now it is possible to record URIs in a MARC 21 record.[14]

4. "Still Waiting for That Funeral"[15]
The authors are of the opinion that the shortcomings of MARC 21 are too often overgeneralized to be the shortcomings of all the MARC formats. Quoting their words: "… *if we have been waiting so long for MARC's funeral, it may be because it is not altogether dead right now.*" The article is about the creation of the "Next-Gen INTERMARC". It is possible to switch from records to linked, reusable, and trustworthy data by increasing the portability of MARC records to a finer-grained level — all the while fitting into the ISO 2709 formalism. In my opinion, one of their thoughts deserves to have such a "career" as "MARC must die" had. That is: Technology in itself is neither necessary nor sufficient to bring forth meaningful changes if it does not come with a paradigm shift in mentalities.

Our position is neatly phrased by Roy Tennant who 10 years after his iconic article tweeted: "It's not so much that 'MARC must die' as it is "*MARC must let us date other standards.*"[16] To this point, a stable encoding standard that has all of the functionality but none or less of the limitations than MARC has not emerged. Meaning today a well-functioned library system could not overlook the MARC format. But we must be aware of the shortcomings of MARC.

Because RDA is introducing a lot of new elements to the bibliographic description, it is very likely that the MARC you use is not suitable for recording those elements in a way that will benefit the user.

There are two solutions to this problem. First one is that we review and amend the MARC we use enabling to record these new elements. Or we opt for the other version of MARC in which these changes have already been made. Both have their own benefits and problems. If you choose to change MARC you have that means that it will be more tailored to your needs and practices. However, it is a lot of work in itself, not even counting the time and effort you need if you also want to have linked data application, in which case you have to do those mappings on your own.

You can also choose a format which is widely used and has already undergone the necessary changes. It is an advantage that an already prepared format is at your disposal. However, you might have to change some of your practices if the format does not support them the way yours did.

We have chosen the latter alternative. We are going to change from our national MARC (HUNMARC) to MARC 21. We have chosen MARC 21 because it is developed by the same agencies as RDA so it is expected that their changes are going to be parallel.
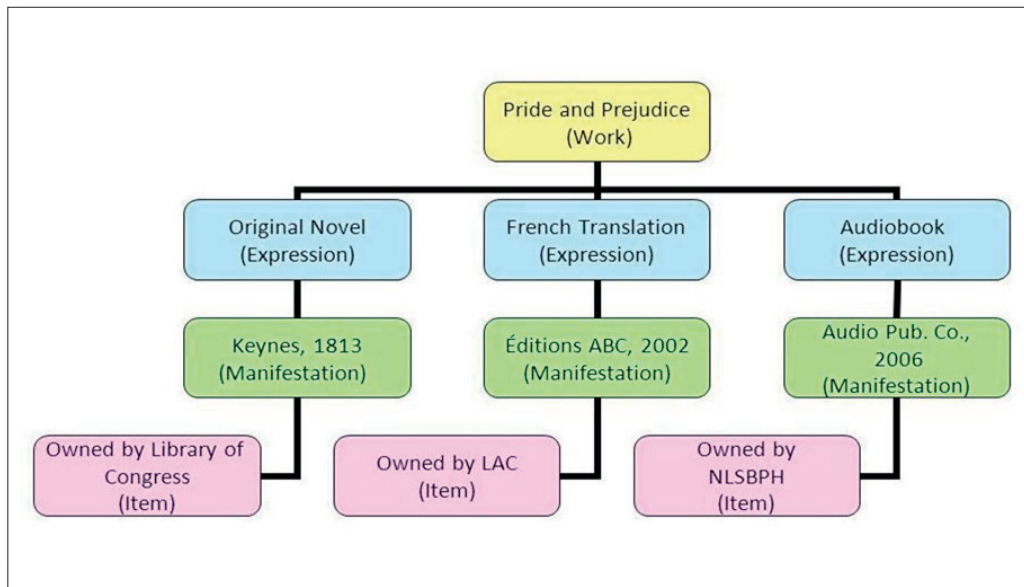
WEMI

To this point, I talked about things that have to be done. Our vision for implementing RDA consists of implementing all 4 WEMI levels and we also want to publish our metadata as linked data. I start with the WEMI levels.

In Functional Requirements for Bibliographic Records IFLA introduced the concept of the Work/Expression/Manifestation/Item levels of library metadata. The model underwent a revision, then it was combined with its related models into a single model now known as the Library Reference Model. FRBR was ahead of its time. But now, a little more than 20 years later, you can see examples of at least partially FRBRized library catalogues.

Some experiments were done to create a FRBRized cataloguing code but RDA became the most known and used among them. And also, currently, the only one that is based on LRM.

One of the reasons why the real-life use of the WEMI levels took so long is that the model intentionally ignored the limitations of the MARC environment. Especially implementing expression level is not so easy. Some of the expression level elements are so deep-rooted part of the bibliographic description that the description cannot be considered to be whole without them, but not all of its elements can be recorded in the MARC bibliographic record.

Figure 3[17]

You have to decide which WEMI levels you want to have and what exactly you want from them. You cannot build a search function for work and expression levels if you do not make records for those levels consistently. On the other hand, creating the description for 2 more levels requires a lot of work.

According to WorldCat statistics, we can expect that about 78 % of works exist in a single manifestation and 16 % of works exist in a single expression (with multiple manifestations), and only 1 % of the works have more than 8 manifestations, but this 1 % has a disproportionately large number of holdings, and those are the works most often sought by users (6 % complex works – with multiple expressions or realizations of its intellectual content)[18].

Our vision is that we want to implement all 4 levels, but we do not have the human resources to do 2 more levels manually from scratch. So, one of the question our working group has to answer in the near future is what we can automatically generate. Automated record generation can be used in other areas as well. LRM considers digital reproductions separate manifestations. Our digitizing department creates digital reproductions in a rate that our cataloguing department cannot keep up with. We had to figure out a way how we can generate a record for the reproduction from the record of the original print version. And now new records are generated using this method.

The implementation of all 4 WEMI levels could revolutionize how the catalogue looks like. Experimenting with display is important because it is a rarely examined area although this is the first thing our users came in contact. I think all communities should reach out for their users for feedback in this area. Because we cannot know how generally valid the findings of other researchers of the topic are. Do different cultures have different preferences in data display?

LINKED DATA

Linked data is a structured data which is interlinked with other data so that becomes more useful through semantic queries. It builds upon standard Web technologies such as HTTP, RDF and URIs, but it extends them to share information in a way that can be read automatically by computers.

Linked data is relatively new to libraries. We are still at the experimenting phase with it. We have to have a lot of pilot projects to figure out how linked data can serve us. So far we only have had one but we plan to have more. Our aim at this point has been to get a better understanding of how the MARC-BIBFRAME converting process works. We are still at the beginning of the learning curve.

For the first try, we wanted to keep it simple from the cataloguing point of view so it would not distract us from studying the converting process. We chose 16 simple monographs by the same author, created bibliographic records and their accompanying authority records in MARC 21 according to RDA rules. Our record set altogether contained 68 records.[19] But because BIBFRAME has not fulfilled its promises[20], it is very likely that we will choose the path of record URIs in the MARC record. And we will keep an eye on any promising format and technology.

With linked data, we face the same problem as with the WEMI levels. It is only useful if it is consistently used in records. But doing it manually would significantly increase the workload of the cataloguing department. Automatic solutions are needed. Reusing the metadata created in namespaces is one solution. We are not cer-

BIBLIOGRAFIA

tain at the moment about which namespace or authority service to use.

CONCLUSION

There are a lot of unanswered questions on implementing RDA. As I mentioned before in some areas we are still just experimenting and there are not so many best practices to learn from. And our pilot projects often raise even more questions than we originally had. And we try to find some information on how other libraries think of this or that particular problem … and we often find no information. And of course, we get irritated by the fact that we cannot see what the others work on. Then it is time for realization… Wait a second! Do the others find the same information about our project that I am trying to find about theirs? And the answer is sadly no. We are all understaffed and overworked and generally just focused on trying to get the job done in time and the first thing we neglect to do is sharing information with the others. Often because we think: "*I will show my project once I am ready, when I have results.*" But the journey is every bit as interesting to others as the results. Communicating our unanswered questions, our bad turns on the road are very important as well. That is where we can realize that we are not alone with our questions and we can start thinking together. We – and by "we" I mean the RDA core team – decided to make it a priority to communicate more about our project to both the national and the international community and I urge you to do the same.

ENDNOTES

1   RDA Toolkit. About [online]. [Accessed: 2019, October 28]. Available at: https://www.rdatoolkit.org/about

2   RIVA, Pat – LE BŒUF, Patrick – ŽUMER, Maja, 2017. *IFLA Library Reference Model (LRM)* [online]. [Accessed: 2020, February 27]. Available at: https://www.ifla.org/publications/node/11412

3   COYLE, Karen – HILLMANN, Diane, 2007. Resource Description and Access (RDA) – Cataloging for the 20th Century. In: *D-Lib Magazine* [online]. 2007, vol. 13, no. ½. [Accessed: 2020, February 27]. ISSN 1082-9873. Available at: http://www.dlib.org/dlib/january07/coyle/01coyle.html

4   DANCS, Sz., 2019. *Lost in translation – implementing IFLA LRM and RDA in Hungary*. Presented during the 30th National Conference of The Romanian Library Association, Bucharest, 4th September 2019.

5   Application Profiles Working Group, 2020 – 2021 [online]. [Accessed: 2020, February 24]. Available at: http://www.rda-jsc.org/sites/all/files/RSC-Chair-2019-3%20final.pdf

6   RDA Toolkit (beta) / Guidance / Recording methods [online]. [Accessed: 2020, February 27]. Available at: https://beta.rdatoolkit.org/en-US_ala-a06b3aa1-d994-31b1-b961-e4ce37c1a4d3/section_gvh_jsg_nbb

7   RDA Toolkit (beta) / Guidance / Recording methods [online]. [Accessed: 2020, February 27]. Available at: https://beta.rdatoolkit.org/en-US_ala-a06b3aa1-d994-31b1-b961-e4ce37c1a4d3/section_um2_x5h_tcb

8   RDA Toolkit (beta) / Guidance / Recording methods [online]. [Accessed: 2020, February 27]. Available at: https://beta.rdatoolkit.org/en-US_ala-a06b3aa1-d994-31b1-b961-e4ce37c1a4d3/section_uy1_4vh_tcb

9   RDA Toolkit (beta) / Guidance / Recording methods [online]. [Accessed: 2020, February 27]. Available at: https://beta.rdatoolkit.org/en-US_ala-a06b3aa1-d994-31b1-b961-e4ce37c1a4d3/section_mrc_vvh_tcb

10 BARNHART, Linda: RDA Data Recording Methods, Transcription, and Manifestation Statement [online]. [Accessed: 2020, February 27]. Available at: http://www.rda-rsc.org/sites/all/files/Recording%20methods%20transcription%20and%20manifestation%20statements.pdf

11 TENNANT, Roy, 2002. MARC must die. In: *Library Journal*. 2002, vol. 127, no. 17, p. 26 – 28.

12 AYRE, Lori Bowen, 2014. MARC Isn't Dying Fast Enough. In: *Collaborative Librarianship* [online]. 2014, vol. 6, no. 4. [Accessed: 2020, February 27]. Available at: https://digitalcommons.du.edu/cgi/viewcontent.cgi?article=1088&context=collaborativelibrarianship

13 ILÁCSA, Szabina, 2013. A könyvtári katalogizálás megújulásának lehetőségei. In: *Tudásmenedzsment* [online]. 2013, vol. 2., no. 14, p. 98 – 110. [Accessed: 2020, February 27]. Available at: http://epa.oszk.hu/02700/02750/00032/pdf/EPA02750_tudasmenedzsment_2013_02_ksz_098-110.pdf

14 Library of Congress / MARC / Bibliographic / Appendix A – Control Subfields [online]. [Accessed: 2020, February 27]. Available at: http://www.loc.gov/marc/bibliographic/ecbdcntf.html

15 PEYRARD, Sébastien – ROCHE, Mélanie, 2018. Still Waiting for That Funeral: the Challenges and Promises of a Next-Gen INTERMARC [online]. [Accessed: 2020, February 27]. Available at: http://library.ifla.org/2204/. Paper presented at the IFLA WLIC 2018 – Kuala Lumpur, Malaysia – Transform Libraries, Transform Societies in Session 141 – Cataloguing.

16 twitter.com/rtennant [Accessed: 2020, February 27]. Available at: https://twitter.com/rtennant/status/179962550782402560

17 FRBR Overview and Application [online]. [Accessed: 2020, February 27]. Available at: https://slideplayer.com/slide/2292462/

18 SALABA, Athena – MERČUN, Tanja – AALBERG, Trond, 2018. Complexity of Work Families and Entity-Based Visualization Displays. In: *Cataloging & Classification Quarterly*. 2018, vol. 56, no. 7, p. 628 – 652.

19 DANCS, Szabolcs – MOHAY, Anikó – HUBAY, Miklós, 2019. Az RDA és a BIBFRAME hazai implementálása. In: *Tudományos és Műszaki Tájékoztatás* [online]. 2019, vol. 9, p.534 – 539. [Accessed: 2020, February 27]. Available at: http://epa.oszk.hu/03000/03071/00132/pdf/EPA03071_tmt_2019_09_534-539.pdf

20 PARK, Jung-Ran – RICHARDS, Lorraine L. – BRENZA, Andrew, 2019. Benefits and challenges of BIBFRAME – Cataloging special format materials, implementation, and continuing educational resources. In: *Library Hi Tech*. 2019, vol. 37, no. 3, p. 549 – 565.