



Inniúlacht Cartlainne Meán

# Creating a Central and Contiguous Digital Repository of Irish-Language Media Within Library of Congress Standards

[www.ard.ie](http://www.ard.ie)

November 2007

© ICM Ard-Media, 2007 and 2013

ICM, Inniúlacht Cartlainne Meán and Ard-Media are registered as a business with the Companies Registration Office, Dublin, <http://www.cro.ie>. Retrieved online from [http://ard.ie/cartlann/LC\\_2007\\_ath\\_2013\\_en.pdf](http://ard.ie/cartlann/LC_2007_ath_2013_en.pdf)

## **Foreword: Advancements Since 2007**

This document written originally as a proposed starting solutions to media non-contiguity of Gaeilge (ISO 639-3=gle) multimedia material. Since 2007, the Linked Web of Data has become developed from conception at that stage to a relatively mature field of research and real-world development.

The core motivations for this publications have not changed, simply the tools have appeared and are being developed at an ever-increasing pace as the web of data becomes more mature, the web of data becomes greater and greater.

---

ICM Ard Media, May 16, 2013

## Introduction

At the present time, there exists no central repository—physical, digital or otherwise—of Irish media (including books) in print. As a major fundamental infrastructural shortcoming, this poses a major obstacle to the accessibility and communicability of the Irish language. A transparent, easily accessible (open access, needless to mention) library, that can be cross-referenced against a myriad of criteria, is a basic infrastructure to a language that seeks to vanquish its status—alleged or otherwise—as “impoverished.”

Although many libraries contain relatively comprehensive collections of Irish titles, Irish language titles are neither listed nor cataloged with any contiguity or distinction from other general collections. It is proposed here that a central digital repository—accessible online and on an open-access basis—be set up. With an appropriate level of support research and development—something that can be achieved comprehensively, yet cost-effectively—the repository would have no problems becoming the standard inventory that would parallel the Library of Congress, Washington, DC.

The Library of Congress, located in Washington, DC’s Capitol, is considered the foremost repository of books (as well as other media) around the world. Such is its importance, that libraries world-over catalog publications by their “LC Call Number”; all books, including most Irish language titles published since June 2000 (and some previous to that) have a LC call-number. Nevertheless, even the Library of Congress collection of Irish-language titles remains discontinuous and discontiguous—a useful reference source, yet defying comprehensiveness and class-searchability. Ultimately, it is planned to interface this collection to the Library of Congress as a custodial collection of the Congress Library’s catalog, even though it would operate primarily on a “standalone” basis.

So how would this library appear from the outside? Quite simply, there is no need for a physical library of media; on the contrary, this resource should serve as a search engine for anyone, remote or otherwise, to be able to search all Irish titles. This digital library would enable for the first time, the sensible categorization of Irish books.

### **A proposed digital (media) library: “LC na Gaeilge”**

The proposed digital library would preferably be interfaced to an institution that has a central focus of applied Irish-language education/research/development. Acadamh na hOllscolaíochta Gaeilge’s objectives, for example, parallel the objectives of this proposed project. OÉG Carna is in the process of digitizing contemporary and endangered audio and audiovisual publications: to restate the previous point, an absolute parallel of this project. Along with other project “Carna” or “Gaoth Dobhair” or “...” can become, with appropriate patronage, synonymous with an Irish-language Library of Congress: a digital repository for all Irish-language media.

Having recently visited the Library of Congress for a two-week fact-finding research mission, I have ascertained that the resources—extremely useful resources—exist at the present time, a selected sample of which are listed in the bibliography, below. As can be seen from the list, several efforts have been made to catalog Irish-language books, but none approach what can be termed “comprehensive,” and by their associated call numbers, it can be seen that many of these resources are not even to be found in the same wing of a library, when available. This causes huge problems for those attempting to access Irish language media: an efficacy, which can be taken for granted when researching material in other languages. Beyond seasoned graduates and academics, along with a select number of vendors—that specialize in Irish-language titles—who do have a somewhat comprehensive access to most books through a number of accumulated databases (examples: [litriocht.com](http://litriocht.com) and CIC), the normal user who is not professionally expert in such fields is at a serious loss. Additionally, these catalogs, (although several vendors claim to sell “every title in print”) even as a sum of their individual parts will never be fully comprehensive, firstly, and more importantly, openly accessible.

The implementation of the project would necessarily be carried out in conjunction with a graduate student who would be enrolled with a specific project aimed at this very goal. Intuitively, graduate students would require a working-level proficiency in Irish and the project would primarily be Computers/IT directed in both scope and thrust.

Support data included here illustrate just how vulnerable non-European languages—specifically Irish at highest risk—are to miscataloging. In addition to the article included—published in London’s *Guardian* newspaper—it is worth pointing out that miscataloging is not a phenomenon limited to non-Irish libraries. The level of incomplete listings and ASCII compatibility errors is alarmingly high, even in universities with strong Irish-language philosophies and extensive teaching departments, such as NUI, Galway.

Why does a non-contiguity of Irish-language books occur in major collections and libraries? Before confronting this question head-on, it is, perhaps, pertinent to examine some of the catalogs available, where they can be found, and what problems exist with these...

Below is a general listing of subject areas of publication and their associated LC letter-designations. So, for example, a medical student is afforded the comfort of knowing that all of the resources s/he may need will be found in one area: the R catalog. Rather than being filed under “Irish,” Irish-language titles are filed under the appropriate subject-heading letter, as if it were an English-language publication. For this reason, a bibliographical title such as *Apalóga na bhFilí, 1200–1650* [Ó Caithnia, 1984] will be found under the Z class, whereas a novel, *Cois Caoláire* [Ó Cadhain, 1953] will be filed under P, the language and literature class, and *Ceimic Bhunúsach* [ÓCinnéide, 1973] will be found under science class Q.

Notwithstanding this order, older titles, such as Ceimic [Diolún, 1950], published in 1950 has a LC class P, consigning it to the language and literature section! Finally, folkloric titles, which contribute a substantial fraction of the Irish-language collection, have a roughly equal chance of being be filed under P as it is under the correct D class!

*Irish Books in Print & Leabhair Gaeilge i gCló* [Cleary, 1984], *Pamphlets, &c., Printed Wholly, or Partly, in Irish, from the Earliest to 1820* [Dix, 1905], *Apalóga na bhFillí* [Ó Caithnia, 1984], *Catalóg Leabhar Gaeilge: Bilingual Guide to Current Books in Irish* [Bord na Gaeilge, 1978] and *Irish Research: A Guide to Collections in North America, Ireland and Great Britain* [Lester, 1987] represent five very distinct catalogs that exist for Irish language books. Individually they represent a lot of time and effort, time and effort that would be much more effectively harnessed by their amalgamation in a single, understandable, standard format.

Since the last three paragraphs have outlined some of the catalogs available in the Library of Congress (catalogs within a catalog, so to speak). Problems with these cataloging efforts, as well as cataloging, in general, are discussed below.

Although books are categorized as “Irish language” in the Library of Congress, they are not searchable by language in the case of Irish, and as mentioned above, non-contiguous by language, since field of study categorizes them. This is, nonetheless ahead of the OCLC (World Cat, the acronym derived from its original name, the Ohio Cataloging Library Center), which has no category for Irish language. Not a surprise, one should expect, in light of the fact that the “dedicated Irish-language University of Ireland,” the National University of Ireland, Galway has no such class either. This is something that the University itself is now in a position to rectify, along with its dedicated applied-Irish wing, Acadamh na hOllscolaíochta Gaeilge, since media archiving is a new field of data management globally, requiring novel solutions, bespoke per language.

At the present time, all publishers in Ireland automatically apply for LC cataloging before printing a title. With this digital resource in place, it will be of great interest to Irish publishers (and other Irish-language publishers) to have Acadamh-cataloging (if it can be referred to as such, even exemplarily, at this tentative stage). For this reason, ongoing updates to the library would involve minimal administration, with publishing houses sending their titles to the repository administrators.

## Bibliography

Bord na Gaeilge. *Catalóg Leabhar Gaeilge: Bilingual Guide to Current Books in Irish*,  
Baile Átha Cliath: Bord na Gaeilge, 1978.

Cleary, S. & J. *Irish Books in Print & Leabhair Gaeilge i gCló*, Wicklow: S. & J. Cleary.  
c1984.

Diolún, Tomás. *Ceimic*, No publisher included, 1950.

Dix, Ernest Reginald McClintock. *List of Books, Pamphlets, &c., Printed Wholly, or Partly, in Irish, from the Earliest to 1820*, Dublin. 1905.

Lester, Dee Gee. *Irish Research: A Guide to Collections in North America, Ireland and Great Britain*, New York: Greenwood Press. 1987.

Ó Cadhain, Máirtín. *Cois Caoláire*, Baile Átha Cliath: Sáirséal agus Dill, 1953.

Ó Caithnia, Liam P. *Apalóga na bhFilí, 1200–1650*, Baile Átha Cliath: An Chlóchomhar Teoranta, 1984.

Ó Cinnéide, Seán. *Ceimic Bhunúsach*, Baile Átha Cliath: Oifig an tSólathair, 1973.

## Support Data

[I: A newspaper article outlining the dearth of library cataloging and its implications for minority languages.](#)

This communication, a forwarded article printed in *The Guardian* of London, clearly illustrates how publications in specific minority languages are “languishing unfindable” in libraries.

### **A catalogue of errors: Libraries' missing millions**

Marc Abrahams

Tuesday September 26, 2006

The Guardian, London

How many books written in seemingly obscure languages are misfiled and languishing unfindable in libraries? Joyce Flynn's experience at Harvard suggests the answer is: a lot.

Flynn, a researcher in Celtic languages, discovered some common mishaps that no one discusses much.

Sometimes, cataloguers and shelfers did strange things with books written in foreign languages. They mangled the catalogue listings, and tucked the books away on the wrong shelves.

Then later, when libraries converted their paper card catalogues to computerised systems, most of the books with screwed-up paper records stayed or went deeper into library limbo. Even though the books themselves may be sitting on library shelves, hardly anyone will ever be able to find them. In libraries where only the staff are allowed to wander through the book stacks, a mere patron might never even know those books exist. This all happened to foreign-language films and other items, too.

About 25 years ago, as a graduate student, Flynn took a summer job involving Harvard's library collection of audiovisual materials. "I came across goofy mistakes in some main entries in Scottish Gaelic and in Modern Irish. Titles and artists that were plural nouns had been catalogued by 'na' (the equivalent of English 'the') as the first word of the titles or of the performing group's name." Lengths of shelving were packed exclusively with titles that begin with that word "na".

"I tried to track how the same mistake could have happened so frequently. It turned out that a staff cutback, years earlier, had eliminated the library cataloguer familiar with the languages. The library had assigned cataloguing in Celtic to someone else. As a result, book titles beginning with 'na', for something like Na Fir (The Men), had been catalogued under 'na' (the) as the first word in the title. Many items catalogued under 'n' belonged elsewhere."

Imagine if *The Great Gatsby*, *The Sound and the Fury*, and *The Tragedy of Hamlet, Prince of Denmark* were all filed under "T."

Recently, Flynn checked Harvard's less-than-25-year-old computer-based catalogue system, and discovered that many - perhaps most - of the Gaelic and Irish books with Na ... titles are miscatalogued and so, in this odd way, are half-missing. That catalogue system is now the only way the public can access titles in the Harvard College Library collections.

"The issue goes beyond just Harvard's Widener Library," Flynn says. "Because Widener is often the first North American library to acquire and catalogue an obscure foreign language title, Widener's cataloguing data frequently become the standard for libraries that acquire the book later.

"Imagine," Flynn mutters, "a row of titles written in non-global languages, waiting to be checked out for the first time - but invisible to scholars seeking them. Imagine a future in which these books no longer wait for Professor Godot to borrow them, because libraries have discarded them - because their circulation statistics show that they were never requested by readers."

To how many books has this happened, in how many languages, in how many libraries around the world? Nobody knows.

Above article is available online at:

<http://education.guardian.co.uk/egweekly/story/0,,1880490,00.html>

## II: A case in point: *Cill Éinde—Céad Bliain*; no LC-classification

This book is a primary reference material; essentially, this resource is essentially unsearchable (despite having an associated ISBN number) because it will not show up in a literature review in most libraries and digital repositories. The book was published privately and for this reason, Library of Congress Categorization was overlooked. Although many publishing companies such as Cló Iar-Chonnachta do ensure Library of Congress Categorization, there is a vast amount of literature and research reference material (such as the *Cill Éinde* example above) that slips through the categorization net, leaving research in Irish language material somewhat speckled, at best, and unreliable in the broadest sense. Often times, there is a British Library association and no LC number.

Feargal Ó Béarra, An Moins. Breandán Ó Ceallaigh, Beairtle Ó Conaire, Maidhc P. Ó Conaola, Póilín Mhic Chonaonaigh, Conall Mac Cuinneagáin, Timín Ó Curraidhin, Eoin Ó Droighneáin, Joe Steve Ó Neachtain, Máire Ní Neachtain, *Cill Éinde—Céad Bliain*. Eag: Eoin Ó Droighneáin, An Spidéal: Comhlacht Forbartha an Spidéil Teoranta, 2007. ISBN: 978-0-9557500-0-7

III: A case in point: ASCII corruption of non-Roman (accented) characters

The title, below, reproduces verbatim, the title Ó Cinnéide 1975 as it appears in the Library of Congress digital catalog, online at <http://catalog.loc.gov/>. Notice the ubiquity of corrupted or non-native character accents. This especially features in the listing of many older titles.

**LC Control No.: 78354130**

**Type of Material: Book (Print, Microform, Electronic, etc.)**

**Personal Name: Oí Cinnéide, Seán.**

**Main Title: Ceimic bhunúsach / Seán Oí Cinnéide.**

**Published/Created: Baile Átha Cliath : Oifig an tSoláthair : [sold by] Oifig Dhíolta Foilseacháin Rialtais, c1975-**

**Description: v. : ill. (some col.) ; 25 cm.**

IV: An Illustration of Irish-Language-Specific Search Parameters:  
Non-correlation between singular/plural/nominative/genitive, along with  
nomenclature

### 1. Personal Names

What's in a name? Whatever is *in* a name can be outwardly represented by many representations of the same name. Consider a common Conamara (and Irish) name, such as “Mícheál Ó Conaola,” *versus* its antecedent, yet contemporary, “Mícheál Ó Conghaile.” Though Conghaile and Conaola are not simultaneously adopted (people will identify by only one version or the other), both forms cannot be differentiated aurally.

In addition, the person that is known in many facets of his life as Mícheál Ó Conaola, may not only be known in his locality as Mícil Chóilín Mhaitiú, but this may be how he is published...case in point, broadcasters Máirtín Tom Sheáinín Mac or Máirtín Jamesie Ó Flaithearta.

Below, is a brief discussion of what is probably the single most aspect of Irish-language material, as it pertains to archiving: genitive possession and the plural. As it turns out, the names Conghaile, Conaola and Mícheál are all mutated by genitive possession, raising the number of permutations and combinations in which the name can appear in any media.

### 2. Normal Genitive and Plural

Plurals are more complicated than in English, German and Spanish, for example, where the addition of *-s*, *-en*, and *-s*, respectively, yields a plural. Nevertheless, it is quite straightforward to feed a search capability every form of a word necessary, thereby uniting as a single entity. Genitive possession exists where some word denotes possession or pertinence to another word (there are only seven grammatical conditions that place a noun in the genitive tense, and even in computer terms, is reasonably tractable). Irish K-12<sup>1</sup> education does not ascribe to teaching language elements such as genitive possession and so school-leavers can find real-world spoken Irish alien to the language they have learned.

### 3. Search patterns for keyword altering lenition and eclipsis

see CAGG, Comhtheagmháil Aircíve Gàidhlig agus Gaeilge on <http://ard.ie>.

---

<sup>1</sup> K-12 (pronounced “Kay-thru-twelve”) is a byword, in American educational parlance for post pre-school through to pre-college-level (i.e. primary and post-primary education). Since it succinctly describes primary and post-primary education in three alphanumeric characters, it is used here.

Litreacha a dtéann séimhiú i bhfeidhm orthu	Litreacha a dtéann urú i bhfeidhm orthu	Litreacha a úsáidtear chun urú a dhéanamh
b	b	b
c	c	bh
D	d	d
f	f	g
g	g	m
p	p	n
t	s	t
a	t	
e		
i		
o		
u		
Séimhiú	bean → an <u>b</u> hean	
Urú	carr → ar an <u>g</u> carr	
<b>Fíor I</b> Litreacha a fheidhmíonn i meicníochtaí séimhiú agus urú		