



Inniúlacht Cartlainne Meán

# Comhtheagmháil Aircíve Gàidhlig & Gaeilge

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

Nollaig 2010

© 2010 & 2013 ICM Ard-Media

# Réamhfhocal

Scríobhadh an tuairisc seo i 2010 agus tá sé á chur ar fáil arís i 2013 gan mórán athrú.

Déantar na nótaí seo mar sin féin:

Tá an cód inchuardaithe atá leagtha amach anseo nóiseanúil; tá difríochtaí thábhachtacha a fhoilseofar amach anseo sa leagan deiridh. Tá an liosta séimhithe agus urraithe do Ghaeilge na hÉireann agus na hAlban neamh-chríochnaitheach freisin, agus mar gheall ar sin, is treoir nóiseanúil atá sa sonraíocht ag an staid seo. (Tá an t-inneall féin críochnaitheach, mar sin féin.

Ar ndóigh, tá dul chun cinn déanta idir an dá linn. Tuilleadh amach anseo ar <http://blog.ard.ie>. Bí isteach, bí i dteagmháil!

Tá an cód inchuardaithe a dearadh anseo mar chuid den tairseach ilmheáin <http://www.ard.ie>.

# Buíochas

## Colmcille:

Tá ICM Ard-Media fíorbhuíoch as an gcabhair a thug Colmcille, <http://www.colmcille.net> leis an togra seo a chur i gcrích. Go sonrúil, buíochas le

Maolcholaim Scott  
Noeleen Ní Cholla

## Taobh Amuigh de Cholmcille:

Ronan Rafferty (Dearadh Gréasáin, Thran <http://www.thran.co.uk> a bhí mar réamhtheachtaí ar an gcaineál ilmheáin Ard-Media, <http://www.ard.ie>).

Bhí daoine ar leith a thug comhairle úsáideach dom freisin idir an comhad inchoardaithe agus leis an OAI PMH/cainéal Europeana:

Dr.Ciarán Ó Duibhín (Sabhal Mòr Ostaig)  
Valentine Charles (Europeana)  
Annette Friberg (Europeana)

## Nóta ar an ngníomhaíocht ICM

Tá an ghníomhaíocht ICM cláraithe leis an Oifig um Chlárú Cuideachtaí, Baile Átha Cliath, <http://www.cro.ie>, faoin ainm Seirbhísí Cartlannaíochta Sna Meáin Nua, Uimhir Cláraithe Gnó 388153.

Inniúlacht Chartlainne Meán (le séimhiú) is gnáthach ach leis an réiltín Inniúlacht \* Chartlainne Meán atá ar an áisíneacht.

## Achoimre

Leagann an tuairisc chinn bhliana seo amach an obair a baineadh amach le tacaíocht ó Cholmcille le bliain anuas sa togra CAGG, Comhtheagmháil Aircíve Gàidhlig & Gaeilge.

Go hachomair, scríobhadh agus cuireadh i bhfeidhm go réaláíoch uirlis ionchuardaithe a chuireann cuarduithe cartlanna ar fáil ar bhun neamhchasta. Leis an réiteach ionchuardaithe seo, is féidir ábhar cartlainne a chuardach (Gàidhlig & Gaeilge) ag leibhéal eochairfhocail ina n-aithneoidh an t-inneall cuardaigh focail ionann leis an eochairfhocal cuardaithe a bhfuil séimhiú nó urú orthu—bac cartlannaíochta go dtí seo mar go gcuireann na meicníochtaí seo na focail seo i bhfolach ón gcuardach.

Ag an staid seo, is féidir a chur in iúl go mbeidh an tuairisc chinn bhliana seo mar chuid d'éileamh úr chuig Colmcille don bhliain 2011. Is córas iomlán cartlannaíochta atá á thairiscint don bhliain seo chugainn.

Tugtar léargas anseo ar an obair a comhlíonadh sa togra CAGG i gcomhthéacs na móroibre atá sa tsiúl. San obair atá sa tsiúl, tá garspriocanna ríthábhachtacha bainte amach ag an staid seo; go deimhin, is féidir a rá go bhfuil an bunchreatlach infreastruchtúir cruthaithe anois chun treabhadh chuig an gcéad chéim eile—an chéim chríochnaitheach.

## An Togra

Leagtar amach an cód ionchuardaithe ina iomláine, Aguisín I. Is é seo luach an togra, agus mar atá aontaithe mar choinníoll leis an maoiniú, tá cead ag Colmcille athúsáid a bhaint as agus fáilte.

Déanann Aguisín II cur síos i bhfoirm Tháblaí, i bhfocail eile, inléite don ghnáthléitheoir—léargas ar an dearadh atá ar chód Aguisín I.

In Aguisín III, déantar athscríobh díreach ar alt le Marc Abrams as an Guardian i 2006, “A catalogue of errors: Libraries' missing millions,” a thugann léargas lom ar an gcontúirt do mhionteangacha—teangacha Gaelacha luaite go sonrúil—nuair nach ndéantar cartlannú fhreagrach nó inniúil orthu.

Tá na garspriocanna atá bainte amach ag ICM le gairid, le hobair leanúnach san áireamh, leagtha amach in Aguisín IV. Go hachomair, seo:

- An t-inneall cuardaigh teanga–dhírithe
- Comhpháirtíocht, soláthar ábhar agus cruthú ardán ar Leabharlann Digiteach na hEorpa, Europeana
- Maoiniú bainte amach don togra iTunes U trí hOllscoil na hÉireann, Gaillimh, Apple agus an Chomhairle um Oideachais Gaeltachta agus Gaeilge. Áiríonn sé seo maoiniú suntasach don togra infreastruchtúir chartlannaíochta do 2011
- Seoladh leagan 0.1 den togra, “Fealsún: Samhail foghraíochta, bunstruchtúr teanga agus áis fhoghlama do thosaitheoirí in inmhe: treoir léamhéisteachta.” Thosaigh an togra seo i 2007 le deontas beag ón gCrannchur Náisiúnta tríd An Roinn Gnóthaí Pobail, Tuaithe agus Gaeltachta

Tá foilseacháin bhainteacha leagtha amach in Aguisín V. Tosaíonn an liosta roimh an am a rinne mé obair chomhairleach chartlainne d’Acadamh na hOllscolaíochta Gaeilge go dtí an t-am i láthair, mar a bhfuil mé i gceannas ar m’áisiúneacht phearsanta chartlannaíochta, ICM. Tá péire atá bainteach leis an togra, ach CAGG, go sonrúil,

- alt a dhéanann cur síos ar an obair agus cuspóirí ICM, a d’fhoilsigh Cumann Ríomh–Oideachais na hÉireann agus atá ar fáil ag <http://www.cesi.ie/digiteach-icm>
- an staidéar féidearthachta agus feidhmithe “Iniúchadh ar Fhéidearthachtaí agus Cur Chuige Bunaithe iTunes University (iTunes U)” atá mar fhasach don togra iTunes atá maoinithe don bhliain 2011

San aguisín deireanach, Aguisín VI, tá achoimre ar na cruinnithe agus teagmhálacha bainteacha a bhí agam le blianta beaga anuas.

## Clabhsúr

Tá mé iontach buíoch, arís, den eagraíocht Colmcille as a suim sa togra seo. Aithníonn a suim tuiscint ar an dearcadh fad-bhreathnaíoch a theastaíonn chun infreastruchtúir theanga a chruthú. Tá súil agam gur léirigh an togra seo go bhfuil go leor acmhainní gur féidir a fhorbairt go comhthreomhar do Ghaeilge na hÉireann agus na hAlban.

Farraor, is é modus operandi cartlannaíochta na Gaeilge go dtí seo ná breathnú ar na meicníochtaí teanga seo mar mhístuaim theanga nuair is amhlaidh gur místuaim chartlannaíochta atá i gcóras ar mian leis ábhar Gaeilge a chartlannú ach nach dtugann deis an t-ábhar a chuardach, agus dá bhrí sin a chur ar fáil, go héifeachtach.

Sa ré dhigiteach, is léir gur féidir ábhar iomlán a chuardach, ní hamháin na hinnéacsanna/meiteashonraí mar a bhí amhlaidh sa ré réamhdhigiteach (samhlaigh cuardach Google Books ina n-aithnítear gach staid ina dtagann eocharfhocail cuardaigh aníos i leabhar).

De bhuíochas na bhforbairtí céanna, tá eagrais in Éirinn agus thar lear ar mhaith leo seilbh, go minic, seilbh eisiach, a ghabháil ar ábhar cartlainne Gaeilge ach nach bhfuil d'acmhainn nó de shuim acu cartlannaíocht (soláthar) freagrach nó inniúil a chur ar fáil. Sna cásanna is tromchúisí (ina mbíonn raon leathan agus uathúil ábhar ach gan aon suim acu an t-ábhar a chur ar fáil mar ábhar inúsáidte agus athúsáidte comhaimseartha), tá sé léirithe ag eagrais faoi leith i.e., Control Vocabulary Lists<sup>1</sup> gur féidir ábhar cartlainne a bheith cláraithe in ainbhíos den saol.

Féach na samplaí thíos d'instidiúidí atá agus a bheidh ag bainistiú ábhair Ghaelacha:

<http://www.universityofcalifornia.edu/news/article/16838>

“...Daniel Melia, co-principal investigator for the new project ... as well as an authority on Irish Gaelic...”

níl aon úsáid sa togra seo d'úsáideoirí comhaimseartha na teanga

---

<sup>1</sup> Cruthaíonn an Control Vocabulary List é seo, áit a mbíonn cartlannaíocht á déanamh ar theangacha nach gcloistear fúthu. Samhlaigh go mbeadh an Library of Congress nó an OCLC ina bhun seo mar a dhéanann siad raon uileleathan leabhar a chartlannú, ach ní heagrais iad atá freagrach as na bunchóipeanna atá sna tíortha dúchais.



<http://ecai.org/neh2007/>

Níl tagairt ar ábhar nó teanga na Gaeilge anseo, cé go ndéanann sé é féin a chur i láthair mar thogra, “Electronic Cultural Atlas Initiative.”

Leabharlann Náisiúnta na hÉireann, féach [Ó Conaire agus Ó Tuairisg 2010]

“The Department of Manuscripts has approximately one million items in its collections spanning nearly a thousand years. They are mostly manuscripts relating to the island of Ireland or written by Irish people in Ireland or overseas....”

<http://www.nli.ie/>

“...classifying material that is in Latin, Irish and English is **complicated** (mo bhéim) and that the National Library of Ireland has not distinguished its collection as so”

Comhfhreagras leis an NIL

Mar sin, gan forbairt den chineál a moltar anseo, éireoidh le hiarrachtaí bac a chur ar theangacha na Gaeilge infreastruchtúr uiledhigiteach a bhaint amach agus teangacha eile mar iad san áireamh.

Chun go mbeadh Gaeilge na hÉireann agus na hAlban comhtheagmhálach, is amhlaidh ar chúiseanna de chuile chineál go mbeadh infreastruchtúir láidre, údarásacha, inúsáidte teanga-lárnaithe cruthaithe i dtosach.

# Aguisín I—An Cód ionchuardaithe CAGG, Comhtheagmháil Aircíve Gàidhlig & Gaeilge

```
//-----
if(preg_match('/^[a,e,i,o,u]/', $queryStringNAME))
{
$var = substr_replace($queryStringNAME, 't', 0, 0);
$var2 = substr_replace($queryStringNAME, 't-', 0, 0);
}
//-----

if(preg_match('/^[A,E,I,O,U]/', $queryStringNAME))
{
$var3 = substr_replace($queryStringNAME, 't', 0, 0);
}

//-----
if(preg_match('/^[A,E,I,O,U]/', $queryStringNAME))
{
$var5 = substr_replace($queryStringNAME, 'h-', 0, 0);
}
//-----
if(preg_match('/^[a,e,i,o,u]/', $queryStringNAME))
{
$var6 = substr_replace($queryStringNAME, 'h-', 0, 0);
}

}

//-----
if(preg_match('/^[a]/', $queryStringNAME))
{
//$var7 = substr_replace($queryStringNAME, "a", 0, 0);
$var7 = substr_replace($queryStringNAME, "&#39;", 1, 0);
}

//-----
if(preg_match('/^[a]/', $queryStringNAME))
{
$var9 = substr_replace($queryStringNAME, '&#146;', 1, 0);
}

if(preg_match('/^[a]/', $queryStringNAME))
{
$var11 = substr_replace($queryStringNAME, '&#146;', 1, 0);
}

//-----
```

```

if(preg_match('/^[a,e,i,o,u]/', $queryStringNAME))
{
$var12 = substr_replace($queryStringNAME, 'n-', 0, 0);

}
//-----
if(preg_match('/^[A,E,I,O,U]/', $queryStringNAME))
{
$var13 = substr_replace($queryStringNAME, 'n-', 0, 0);
}
//-----

if ( preg_match('/^[f,a,e,i,o,u]/', $queryStringNAME) )
{
$other = substr_replace($queryStringNAME, 'dh&#146;', 0, 0);
}
//if(preg_match("/^b/", $queryStringNAME))

if ( preg_match('/^[a,e,i,o,u]/', $queryStringNAME) )
{
$var4 = substr_replace($queryStringNAME, 'dh&#146;', 0, 0);
$var4 = substr_replace($var4, '-', 9, 0);
}

//-----

if (preg_match('/^[f]/', $queryStringNAME))
{
$var14 = substr_replace($queryStringNAME, 'dh&#146;', 0, 0);
$var14 = substr_replace($var14, 'h', 9, 0);
}

//-----
if ( preg_match('/^[S]/', $queryStringNAME) )
{
$uru9 = substr_replace($queryStringNAME, 't-', 0, 0);
}

```

## Aguisín II—Meicníochtaí séimhuithe agus uruithe ó dhearcadh na hionchuardaíochta de (teimpléid don dearadh cóid, Aguisín I

Bun leagan	Séimhiú/athrú	Sampla
a, e, i, o, u A, E, I, O, U an oidhche	h-a, e, i, o, u h-A, E, I, O, U	air feadh na h-oidhche an t-athair An t-Athair
xxx am bòrd	a'xhxx (obsolescent) a' xhxx (An Caighdeán reatha)  Tá dhá chineáil ' ann anois - an ceann ASCII sean-nóis, agus an "smart quote" mar a deir Microsoft.  Áirigh an leagan caighdeánach chomh maith leis an gceann atá dulta as feidhm	air a' bhòrd
xxx f-, a-, e-, i-, o-, u-	dh'xhxx dh'[guta]xx dh'fh-, dh'a-, dh'e-, dh'i-, dh'o-, dh'u-	
nan / nam	Ní bhaineann Is meicníocht urlabra nach scríobhtar atá anseo: seo uirlis chuardaithe scríobhneoireachta!	
t-	t-a, e, i, o, u t-s t-A, E, I, O, U t-S	
n-	n-a, e, i, o, u n-A, E, I, O, U	
	Rémhíreanna Droch- Deagh- Sean-	Ní bheidh aon réiteach le haghaidh rémhíreanna i Leagan I

Litreacha a dtéann séimhiú i bhfeidhm orthu	Litreacha a dtéann urú i bhfeidhm orthu	Uirlisí a úsáidtear chun urú a dhéanamh
b	b	b
c	c	bh
d	d	d
f	f	g
g	g	m
m	p	n
p	s	t
t	t	
s	a	
	e	
	i	
	o	
	u	
Séimhiú	bean --> an bhean	
Urú	carr --> ar an gcarr	
<b>Fíor 1</b> Litreacha a fheidhmíonn i meicníochtaí le séimhiú agus urú orthu		

Litreacha a dtéann urú i bhfeidhm orthu	Foirm uraithe	
b	mb	
c	gc	
d	nd	
f	bhf	
g	ng	
p	bp	
T	Dt	
s	ts	
a	t-a	tA
e	t-e	tE
i	t-i	tI

o	t-o	tO
u	t-u	tU
a	ha	hA
e	he	hE
i	hi	hI
o	ho	hO
u	hu	hU
<b>Fíor 2</b> Teaglaimí ionchuir ceadaithe i gcóir uruithe		

Litreacha a dtéann séimhiú i bhfeidhm orthu	Litrithe séimhithe
b	bh
c	ch
d	dh
f	fh
g	gh
m	mh
p	ph
t	th
S	sh
<b>Fíor 3</b> Teaglaimí ionchuir ceadaithe i gcóir uruithe	

Search String begins with	Add modified search string to original
b	mb
c	gc
d	nd
f	bhf
g	ng
p	bp
t	dt
s	ts

a	t-a, tA, ha
e	t-e, tE, he
i	t-i, tI, hi
o	t-o, tO, ho
u	t-u, tU, hU
b	bh
c	ch
d	dh
f	fh
g	gh
m	mh
p	ph
t	th
s	sh

**Fíor 4** Achoimre iomlán na search strings ag teastáil