REPRINTING SCHOLARLY WORKS AS E-BOOKS FOR LESS-RESOURCED LANGUAGES

Delyth Prys, Mared Roberts, Gruffudd Prys
Language Technologies Unit, Canolfan Bedwyr, Bangor University

2nd Workshop on Collaboration and Computing for Under-Resourced Languages, 23 May 2016, LREC Portorož
Context: Wales and Welsh

Wales: Population 2,955,841*

19% speak Welsh* (562,016)

*2011 Census

maps by freevectormaps.com
Context: Book publishing in Wales

- Subsidies and grants are used to make publishing books commercially viable
- The Welsh Books Council is a national body, funded by the Welsh Government
- 367 Welsh language books were published in 2015 (excluding textbooks)
- Few grants to cover reprinting costs
Context: Coleg Cymraeg Cenedlaethol

Establishment of Welsh-medium virtual national college in 2011
Challenges

• New demand for higher education resources through the medium of Welsh
• Many important scholarly books out of print
• Reprinting on paper is expensive and books will become out of print again
• Welsh Government’s Language Strategy lists e-books as an important tool to strengthen infrastructure for Welsh
Solution

E-publishing project: DEChE – Digitization, E-publishing and Electronic Corpus

- Digitizing academic books that are out of print, creating free open access e-books.
- Additional benefits for users. They can:
  - work anywhere
  - search text and create notes
  - listen with Welsh text-to-speech software
- Create Welsh corpus of the e-books’ content to support independent research
- Work undertaken by Bangor University’s Language Technologies Unit
Selecting and Prioritizing Books

- Nomination process arranged by Coleg staff
- Criteria for e-publication:
  1) core material for Welsh-medium module
  2) on students’ reading lists
  3) key research material
- 189 titles nominated by 41 different academics spanning over 20 subjects.
Copyright issues

• Permission sought for works still within copyright period
• Small fee offered
• Publishers paid to help trace copyright holders
• Goodwill towards project within publishing houses. All but one of publisher consented to work with project
• The project had no intention to undermine the Welsh publishing industry, if a nominated book is commercially viable
OmniPage Nuance Ultimate OCR program

Phaedon

Suspect word:

Emrys Evans o'i ragarweidiad i'w gyfeithiad Cymraeg o Phaedon (xxi-xxii): 'Yr hyn

Suggestions: (Click or press Ctrl+Number to select)

1) Phaedon

User dictionary: cymraeg

Page Ready Document Ready Close
<body>
  <p class="poem_number">XVI</p>
  <p class="poem_title">Gwedddi am bresenoldeb Duw.</p>
  <p><br/></p>
  <div class="poem">
    <p>Pam y caiff bwystfilod rheipus<a href="../Text/Nodiadau.html#16" id="nol[16]">[16]</a></p>
    <p class="indent">Dorri'r egin mân i lawr?</p>
    <p>Pam caiff blodau peraidd ifainc</p>
  </div>
</body>
Calibre

Open source e-book management program to convert EPUB file to MOBI format (for Kindle) and PDF format
Llyfrgell Adnoddau (Resource Library)
https://llyfrgell.porth.ac.uk
Y Traddodiad Rhyddiaith yr Oesau Canol – Geraint Bowen (gol.)

Gan Coleg Cymraeg Cenedlaethol

ePub, MOBI, PDF (gwel y 'Atodiadau').

Cyfres y benodau yn trafod traddodiad rhyddiaith Cymru yr oesau Canol, pan oedd y Gymraeg yr un o 'eithoedd llenyddol pwsiaf Europs'. Cer y mirdiniaethau ar y chwedlau, gan gynnwys Pedwr Caen a Mabinogion, yr Rhamantau a phenodd bwsig Daiyd Glyn Jones, 'Breuddwyd Rhonabwy'. Trafodir y bheddelau, rhyddiaith grefydol a chylfiethiadu cyfresu yr Gymraeg o ysgol yr cyfrifiau cymr.

Tagiau: barddoniaeth, cyfrws ddigidiol y Coleg Cymraeg, cymraeg, dwrwyddiaeth, hanes Cymru, llenyddiaeth Gymraeg, oesodd canol.
Within EPUB and MOBI:

- Bookmark
- Search text
- Write notes
- Toggle between endnotes and main text
610 results in 44 documents

rhydd (free), rhydd (rhoddi/to give), rydd (mutated form), rhyddid, rhyddiaith, rhyddhau, rhyddymofyn, Cymru Rydd, canu rhydd, Eglwys Rydd, ysgol rydd
Marketing Tweets, posters, leaflets, slides, radio interview, presentations and articles.
Total Unique Hits
(per quarter, Oct 2014 to March 2016)
Conclusions

• Successful method of republishing academic books
• Weakness in uptake and visibility being addressed
• Important works sustainable in long term
• Stimulate original scholarly discourse
• Status in digital media for minority language
• Marketing tool for Coleg to wider audience
• Language corpus as secondary benefit
Diolch yn fawr!

Y Coleg ar y Bryn