Idealism vs Pragmatism

The challenges of collecting social media

Gillian Lee
Alexander Turnbull Library
National Library of New Zealand
Our motivation – gaps in the collection

<table>
<thead>
<tr>
<th>Category</th>
<th>2011</th>
<th>2014</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Website only</td>
<td>88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Website &amp; social media</td>
<td></td>
<td>141</td>
<td>130</td>
</tr>
<tr>
<td>Social media only</td>
<td></td>
<td></td>
<td>253</td>
</tr>
<tr>
<td>No website or social media found</td>
<td></td>
<td>39</td>
<td>135</td>
</tr>
</tbody>
</table>

Statistics taken from websites/social media sites identified by web archiving staff during 3 successive election campaigns 2011-2017.
Which approach to take?

Web Crawl?
Downloaded Twitter account?
The result?
- 1 general election ‘hashtag’ crawl
- 228 candidate accounts captured initially.
- 67 additional accounts captured later
- 23 downloaded Twitter archives received from 180 requests.
Alma - ‘published’ catalogue
Tiaki - ‘unpublished’ catalogue
Online Channel – federated search