Building Meaning DH6034 - TOOLS AND METHODOLOGIES ESSAY



What Do I Want?



DÍOSPÓIREACHTAÍ PARLAIMINTE PARLIAMENTARY DEBATES

DÁIL ÉIREANN

TUAIRISC OIFIGIÚIL—Neamhcheartaithe (OFFICIAL REPORT—Unrevised)

Front page of a PDF of the daily proceedings of Dáil Éireann. he Official Report of the Houses of the Oireachtas is a continuously updated corpus comprising a written record of debates occurring in Dáil Éireann, Seanad Éireann and various committees or parliamentary bodies. It is made available primarily through the Debates Office website (<u>https://www.oireachtas.ie/en/debates/find/</u>)¹ and the text of the debates is usually uploaded within several hours of the debate happening and after the report has gone through a multistage editorial process.

The Official Report has a corpus dating back to the First Dáil more than a century ago and large amounts of debate are added with every sitting day of the Houses. It is intended to explore processes whereby the act of extracting meaning from Dáil debates, both current and historical, could be facilitated by using digital tools and alternative ways of reading the text of the Official Report and, if possible, relating these readings to other forms of political discourse.

Getting It.

nce they have gone through the editorial process, debates in the Official Report are published to the web as XML, with the structured data also made available on the Oireachtas open data API (<u>https://</u> <u>data.oireachtas.ie</u>)², along with other structured data in the

Oireachtas, including those

relating to Members,

 Share this page 	Debates
 > Debates > Find a debate > Dáil debates > Seanad debates 	Show: Dáil debates For: specific dates Update list From: 12/04/2021 To: 03/05/2021 Member: Brian Stanley Image: Stanley
 Committee debates Find a vote Dáil votes Seanad votes Committee votes 	Dáil debates for specific dates Search within this list Q Page 1 of 1 Show: 20 per page Thu, 29 Apr 2021 Climate Action and Low Carbon Development (Amendment) Bill 2021: Second Stage (Resumed) View
Parliamentary questionsWritten questions	Wed, 28 Apr 2021 Residential Tenancies (Student Rents and Other Protections) (Covid-19) Bill 2021: Second Stage [Private Members]
 > Oral questions > Search tips > Utterances having an adverse effect 	Page 1 of 1 Show: 20 per page Y

Detail of search page on the Debates Office section of the Houses of the Oireachtas website.

constituencies, votes, etc., for retrieval and reuse outside of the Oireachtas website.

The simplest way to retrieve the text of a debate is to perform a search of the Debates Office website and this

may be useful if dealing with a relatively small portion of debate. Searches can be filtered by House and date, with an option to filter by written or oral parliamentary question or refine by keywords. Once a search is complete, a PDF of the entire day's debate can usually be downloaded in tandem with viewing the returned results as web pages.

KildareStreet

Home	<u>Dáil</u>	<u>Seanad</u>	Committees	Northern Ireland	Help

Source code

KildareStreet.com is a modified installation of the <u>TheyWorkForYou.com source code</u>, which is maintained by the lovely people from <u>MySociety</u>, a a place to start building your own copy of this site for your home country, we **strongly** urge you to go there and start with the 'canonical' version

It is available under a BSD-style license, which you can find here. Roughly, this means you are free to copy, use, modify and redistribute the code Commercial or non-commercial use is allowed. However, both its original authors and ourselves disclaim any warranty, and you are expected not i TheyWorkForYou without permission. You may not use KildareStreet's name without permission either.

We feel it's only right that we share our modifications. You can find a tarball of this site's code as it stood on April 16th, 2009 by clicking on this li

How on earth do I use this code?

Probably best that you ask on the mySociety developers-public email list. But again, don't use our modified version. Start with the original.

About us Contact Us House rules Source code

kildarestreet.com brings advantages ... and disadvantages.

dvanced search queries are difficult or impossible using the Debates Office website, however, and it may also be difficult to work with returned data in the form of either URL or PDF.

At least one alternative to the Houses of the Oireachtas website exists in <u>https://www.kildarestreet.com</u>³, which

uses <u>TheyWorkForYou.com</u>⁴ source code and the structured data from the Oireachtas open data API to create a repository of parliamentary debates. Boolean searches can be performed but this does not appear to be a complete record of all parliamentary debates. The source code of the site is made available by

the developer, however, along with the original code from MySociety.org, so although the website is dependent on the developer, it may be possible to customise elements for use. The problem of capturing data for anything outside passive use remains, however.

he Oireachtas open data API is a fulfilment of the EU Open Data Directive that encourages the making available of public sector information for reuse. REST queries may be made via Elasticsearch⁵ to return datasets within specific parameters, with the structured data returned as JSON datasets, providing much more

The Houses of the Oireachtas is providing these APIs to allow our datasets to be retrieved and reused as widely as possible. They are intended to be used in conjunction with https://data.oireachtas.ie, from where our datasets can be accessed directly. By using the APIs, users can make metadata queries to identify the specific data they require. New data are available through the API as soon as they are published.

Currently, https://data.oireachtas.ie contains data in XML format from the Official Report of the Houses of the Oireachtas (the "debates") and replies to Parliamentary Questions in XML files complying with the Akoma Ntoso schema, as well data in PDF format for Bills, Acts and other documents published by the Houses of the Oireachtas.

Files can be retrieved from https://data.oireachtas.ie by adding the URI fragment contained in the "formats" fields of the JSON documents returned by these APIs. At the moment only PDF and XML files are available directly from https://data.oireachtas.ie but this will become the endpoint for direct access of all "uri" fields in the data queried through https://api.oireachtas.ie. We will also be making bulk downloads available through https://api.oireachtas.ie. We will also be making bulk downloads available through https://api.oireachtas.ie. We will also be making bulk downloads available through https://api.oireachtas.ie. We will also be making bulk downloads available through https://api.oireachtas.ie. We will also be making bulk downloads available through https://api.oireachtas.ie. We will also be making bulk downloads available through https://api.oireachtas.ie. We will also be making bulk downloads available through https://api.oireachtas.ie.

Please note the APIs are a work in progress. We are working on expanding the range of datasets we publish, and we are interested in hearing about how to make these APIs more useful and wide ranging. For these reasons, we welcome any feedback, suggestions and user stories to <u>open.data@oireachtas.ie</u>

Data published through these APIs are made available under the Oireachtas (Open Data) PSI Licence

<u>Oireachtas Open Data - Website</u> <u>Send email to Oireachtas Open Data</u>

legislation	\sim
GET /legislation API	
debates	\sim
GET /debates List	
constituencies	\sim
GET /constituencies List	
parties	\sim
GET /parties List	
divisions	\checkmark

Detail of Houses of the Oireachtas open data API, including the integrated Swagger UI.

flexibility than website resources outlined above, although some familiarity with javascript may be required in order to work comfortably with JSON files. Details of uniform resource indicators, URIs, also allow the possibility of further customisation or refining of data searches. The more technical approach to search and data retrieval leaves the possibility to yield far more versatile datasets.

tina-iMac:~ David\$ curl -X GET "https://api.oireachtas.ie/v1/debates?chamber id=&date start=2021-04-29&date 'head": { "counts": { "debateCount": 1, "resultCount": 1 "dateRange": { "start": "2021-04-29T00:00:00.000Z", "end": "2021-04-29T00:00:00.000Z" "lang": "mul" . results": ["debateRecord": { "date": "2021-04-29", "debateSections": ["debateSection": {
 "debateType": "statement", "formats": { "pdf": null, "xml": { "uri": "https://data.oireachtas.ie/akn/ie/debateRecord/dail/2021-04-29/debate/mul@/dbsect_2.xml" "showAs": "Covid-19, Mental Health and Older People: Statements", "containsDebate": true, "parentDebateSection": null, counts": { "speakerCount": 31, "speechCount": 63 "speakers": [], "debateSectionId": "dbsect 2". "bill": null, "uri": "https://data.oireachtas.ie/akn/ie/debateRecord/dail/2021-04-29/debate/dbsect_2" "debateSection": { "debateType": "questions", "formats": { "pdf": null, "xml": { "uri": "https://data.oireachtas.ie/akn/ie/debateRecord/dail/2021-04-29/debate/mul@/dbsect_3.xml" "showAs": "Ceisteanna ó Cheannairí — Leaders' Questions", "containsDebate": true, "parentDebateSection": null, counts": { "speakerCount": 6, "speechCount": 18 "speakers": [], "debateSectionId": "dbsect_3", "bill": null. "uri": "https://data.oireachtas.ie/akn/ie/debateRecord/dail/2021-04-29/debate/dbsect_3" "debateSection": { "debateType": "questions", "formats": { "pdf": null, "xml": { "uri": "https://data.oireachtas.ie/akn/ie/debateRecord/dail/2021-04-29/debate/mul@/dbsect_4.xml" "showAs": "Ceisteanna ar Reachtaíocht a Gealladh - Questions on Promised Legislation", "containsDebate": true, "parentDebateSection": null, "counts": { "speakerCount": 22,

ccessing the Houses of the Oireachtas REST API can be done in several ways. Command line scripts can be achieved using cURL and this is at once the most powerful and customisable search strategy for returning JSONs with very specific data. Such a method is unforgiving, however, and a good knowledge of the method is required to return datasets without error.

The Oireachtas open data portal has a bundled interface description language interface for making REST queries in <u>Swagger/OpenAPI⁶</u>. This has the effect of combining the ability to make intricate search queries within a reasonably intuitive GUI that does not require the exactness of cURL, with a JSON returned within the interface that can be copied to file. This interface has many, although not all, the advantages of the command line method but constraints arise because of limited customisation opportunities.

Accessing the Oireachtas API through the command line with cURL.

built-in Swagger

interface on the



Increased functionality (without the command line expertise) with a dedicated API application like Postman.

Oireachtas open data portal. It also has the advantage of being well-known and popular, meaning support should be readily available if necessary, and being stand-alone should lend itself to designing custom workflows where necessary rather than relying on the interface on the Oireachtas website, with a trade-off of being slightly less user-friendly.

formats xml uri	writtens pdf uri	chamber showAs	chamber uri	questionCount	billCount	contributorCo	divisionCount	debateSectionCo	house showAs	chamberType	committeeCode	house houseCode	houseNo	house uri	debateRecord uri	contextDate
https://data.oirea	cht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	ht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	ht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	ht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	cht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	ht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	ht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	ht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	cht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	cht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	cht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	ht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	ht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	ht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	ht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	ht null	Seanad Éireann	https://data.oireacht	C	1	42	0	16	26th Seanad	house	null	seanad	26	https://data.oireacht	https://data.oireacht	2021-04-30
https://data.oirea	ht null	Dáil Éireann	https://data.oireacht	C	1	130	0	13	33rd Dáil	house	null	dail	33	https://data.oireacht	https://data.oireacht	2021-04-29
https://data.oirea	ht null	Dáil Éireann	https://data.oireacht	C	1	130	0	13	33rd Dáil	house	null	dail	33	https://data.oireacht	https://data.oireacht	2021-04-29
https://data.oirea	ht null	Dáil Éireann	https://data.oireacht	C	1	130	0	13	33rd Dáil	house	null	dail	33	https://data.oireacht	https://data.oireacht	2021-04-29
https://data.oirea	cht null	Dáil Éireann	https://data.oireacht	C	1	130	0	13	33rd Dáil	house	null	dail	33	https://data.oireacht	https://data.oireacht	2021-04-29
https://data.oirea	ht null	Dáil Éireann	https://data.oireacht	C	1	130	0	13	33rd Dáil	house	null	dail	33	https://data.oireacht	https://data.oireacht	2021-04-29
https://data.oirea	ht null	Dáil Éireann	https://data.oireacht	C	1	130	0	13	33rd Dáil	house	null	dail	33	https://data.oireacht	https://data.oireacht	2021-04-29
https://data.oirea	ht null	Dáil Éireann	https://data.oireacht	C	1	130	0	13	33rd Dáil	house	null	dail	33	https://data.oireacht	https://data.oireacht	2021-04-29
https://data.oirea	ht null	Dáil Éireann	https://data.oireacht	C	1	130	0	13	33rd Dáil	house	null	dail	33	https://data.oireacht	https://data.oireacht	2021-04-29
https://data.oirea	ht null	Dáil Éireann	https://data.oireacht	C	1	130	0	13	33rd Dáil	house	null	dail	33	https://data.oireacht	https://data.oireacht	2021-04-29

Tableau Desktop, with some use, might be the most seamless end-to-end experience but it comes with the cost of a greatly sandboxed workflow.

nother option may be proprietary data visualisation software, such as <u>Tableau Desktop</u>⁸, which has the ability to plug directly into the Oireachtas API to make REST queries with third-party web connector software. This has the potential to greatly simplify data queries, albeit at the expense of using proprietary software where outputs might not be easily used or shared outside the software's ecosystem. It is not completely closed, as there is a free version and data can be hosted on Tableau Public, but neither work on the data nor outputs can be worked with easily outside the ecosystem.

Then What?

he text of the Official Report is structured with XML metatags using the <u>Akoma Ntoso</u>⁹ schemas, specialised for parliamentary and legislative documents. This means both the JSON data returned from the Oireachtas API and the XML documents within are structured so as to allow data collation or query based on specific terms, including date, Chamber, debate type and speaker, among others, with entities such as party, speaker and House having URIs to eliminate duplication or confusion, while at the same time making it easier to relate entities via these URIs.

As noted, JSONs can be viewed many different ways, including but not limited to online viewers, the REST query apps used previously or even the most basic text and code editors. For the purposes of extracting meaning from debate text, it should not be necessary to change the data much, if at all. However, should it be necessary, a code editor such as <u>SubEthaEdit</u>¹⁰ or <u>Sublime Text</u>¹¹ is probably most appropriate and has built-in features to check code and reduce the risk of introducing syntax errors in large datasets.

Dealing with a volume of data means it will probably be necessary to clean the data and eliminate extraneous information. As the JSON datasets can be edited in a code editors, they would be the most straightforward way to make minor changes. If converted to a CSV file, the data could see more efficient cleaning even with the likes of spreadsheet software such as <u>Google Sheets</u>¹² or <u>MS</u> <u>Excel</u>¹³, although it is important to maintain the veracity of the data during any conversions.

WHAT DO I DO WITH THE DATA?

211		- Anno - Califordia (California) - E Anto Alifornia - Service (California) - Service (California) - E Anto Mandella - Service (California) - Service (California) - Service (California) - Service (California) - Service (California) - Service (California) - Service
212	<pre><doclitle>PARLIAMENTARY DEBATES</doclitle></pre>	anger inde Beerkender Weckensen Freisensetzensen Anger inder Berger in der Staten in Anger inder Berger in der Staten in Anger in der Berger in der Berger in der Berger in der Staten in Anger in der Berger in der Berger in der Berger in der Berger in Anger in der Berger
213		
214	<pre><block name="proponent_ga"></block></pre>	 A second s
215	<pre>docProponent>DAIL EIREANN</pre>	
216	<th></th>	
21/	<pre><block name="proponent_en"></block></pre>	Set of the set of t
218	<pre><docproponent>DAIL EIREANN</docproponent></pre>	
219		
220	<pre><block name="status_ga"></block></pre>	
221	<pre><docstatus>TUAIRISC 0IFIGIUIL</docstatus></pre>	
222		
223	<pre><block name="status_en"></block></pre>	
224	<pre><docstatus>(OFFICIAL REPORT)</docstatus></pre>	and a second
225		And the second second second second seco
226	<pre><block name="date_ga"></block></pre>	
227	<pre><docdate date="2021-04-29">Deardaoin, 29 Albrean 2021</docdate></pre>	
228		
229	<pre><block name="date_en"></block></pre>	
230	<pre><docdate date="2021-04-29">Thursday, 29 April 2021</docdate></pre>	
231		
232	<pre><block name="volume"></block></pre>	De la contra y de la contra
233	<pre><docnumber refersto="#vol_1006">Vol. 1006</docnumber></pre>	
234		
235	<pre><block name="number"></block></pre>	
236	<pre><docnumber refersto="#no_3">No. 3</docnumber></pre>	
237		
238	<pre><block name="version_en" refersto="#unrevised"></block></pre>	
239	<pre><docstatus>Unrevised</docstatus></pre>	
240		i per al pela provinsi dan ber ana den pela pela de ser de ana de ser de la pela pela de ser de ser de ser de l'ana de la pela de ser de s
241	<pre><block name="version_ga" refersto="#unrevised"></block></pre>	
242	<pre><docstatus>Neamhcheartaithe</docstatus></pre>	
243		A set was the "an address and was address and the descent of the set of the transition of the set o
244		i new all the little state is set of the first of the state of the state of the state of the state of the state

Using Sublime Text to interact directly with debate XML file (above) and JSON dataset from the Oireachtas API.

27	"showAs": "Business and Covid-19: Statements",	ine m Wasanner 12
28	"containsDebate": true,	
29	"parentDebateSection": null,	
30	"counts": {	And the set of the set
31	"speakerCount": 22,	
32	"speechCount": 33	The second
33		And
34	"sneakers": [].	
35	"debateSectionId": "dbsect 2".	
36	"bill": pull.	And the structure of th
30	"uril" "https://data_oireachtas_ie/akn/ie/dehateRecord/seanad/2021_04_30/dehate/dhsert_2"	
38		Territoria No successione concernances
30		And An Article and An Article and Article
10		Files Pase
40	udahataSastian", J	
41		internet internet
42		The second secon
45	Initials (Elizari Elizari
44		an a
45		The new second s
40	"url": "nttps://data.olreacntas.ie/akn/le/debatekecord/seanad/2021-04-30/debate/mul@/dbsect_3.xml"	And
47		1079029/209-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-
48		BCDD and an and a second secon
49	"showAS": "Messages from Joint Committees",	
50	"containsDebate": true,	
51	"parentDebateSection": null,	And and a second
52	"counts": {	, Enterna / Filipine source Million of the second
53	"speakerCount": 1,	-
54	"speechCount": 1	1855 Marcoller of the construction of the cons
	· · · · · · · · · · · · · · · · · · ·	1997 - Hans are arrent to all or should be
	"speakers": [],	and became and a construction of the second
57	"debateSectionId": "dbsect_3",	197 Marrienn chainmeanna "Milling ann
58	"bill": null,	New Terrar
	Ilumille Illetter //data aircaster is/a/s/is/datataDasard/2021_04_20/datata/dasart 21	10203/071 10203

WHAT DO I DO WITH THE DATA?

or more extensive work a specialised cleaning application may be beneficial; a great advantage of such applications is the availability of nondestructive editing, where

changes can be made while leaving the original dataset in situ.

Tableau Desktop, if used to connect directly to the Oireachtas API for data, can export data for cleaning to

<u>Tableau Prep Builder</u>¹⁴, which is specifically for cleaning and

OpenRefine

A power tool for working with messy data.

Create Project Create a project by importing data Open Project TSV, CSV, *SV, Excel (.xls and .xlsx) Import Project Get data from Language Settings This Computer Web Addresses (URLs) Clipboard

OpenRefine offers an open source alternative as a specialised data cleaning tool, easing worries about risks to sustainability.

third-party software with no guarantees of continued operation.

OpenRefine¹⁵ is an open source alternative to Tableau Prep Builder, with similar capabilities at the cost of a slightly less intuitive GUI for working with larger files or combining datasets.

Data from the Debates Office should not need much alteration because of its excellent structure so any required data cleaning will

combining datasets in a non-destructive way. It may be the easiest way of combining large datasets from the Oireachtas where it is necessary, although the same caveat of caution applies, as Tableau uses proprietary formats and data can only be used with software outside the Tableau ecosystem if exported as CSV. The process may also rely on

probably take the form of removing extraneous data and combining datasets where necessary. Nonetheless, in order to maintain data integrity, a specialised open source and sustainable data cleaning application such as OpenRefine would minimise risks to data integrity and sustainability.

What's the Goal?

nce data from the Debates Office has been selected and prepared for analysis, there is more scope for divergence in decision-making. The stated goal is to extract meaning from the voluminous corpus of text generated both historically and on a continuing basis but Members of the Dáil since 1919 number in the thousands and there have been many hundreds of thousands of speeches made in the Dáil Chamber; if it is impossible to comprehend the meaning of the entire corpus at once, it may be worth exploring more than one aspect of the ways in which people derive meaning from political debate in this sphere. It is intended to evaluate four approaches in this respect:

32nd Dáil



Detail of chord diagram of parties and constituencies in the 32nd Dáil, via everviz.com.

- ➡ Close reading of selected speeches aided by digital tools.
- ➡ Distant reading of part of the corpus aided by digital tools.
- ➡ Network analysis.
- ➡ A combination of the above, where appropriate, and related digital toolsets.

Close Reading

he most natural way of extracting meaning from text remains close reading and although it is not possible to give a close reading to all the corpus of Dáil debates, it is possible to use the structured nature of the Oireachtas datasets to aid such reading of a smaller section of debates.

As an example, one may wish to conduct a feminist review of contributions by women to early Dáileanna. The Oireachtas API makes it possible to compile a dataset of historical Members, and although gender is not indicated on metadata, it is possible to clean the sets of superfluous metadata while adding other pertinent data, such as gender. From this investigation, the structured nature of the Oireachtas data can be further leveraged to gain information on speeches made by a particular female Member, such as in the example below, where a REST query of debates that includes the URI for Mary MacSwiney, a nationalist Member of early Dáileanna, returns a JSON of all the debates to which she contributed, within which there are URIs for other sources, including debate titles, headings, debate XML files etc.

curl -X GET "https://api.oireachtas.ie/v1/debates?chamber_id=&date_start=1900-01-01&date_end=2099-01-01&limit=5000&member_id=https%3A%2F%2Fdata.oireachtas.ie%2Fie%2Foireachtas%2Fmember%2Fid%2FMary-MacSwiney.D.1921-08-16" -H "accept: application/json"

"head": {
 "counts": {
 "debateCount": 44,
 "resultCount": 44

Mrs. MacSwiney was an abstentionist and did not always take her seat on election, which explains the relatively low number of debates to which she contributed. Nevertheless, this indicates how URIs in datasets can bring efficiencies in seeking appropriate text for close meaning investigation.

Distant Reading

onversely, it is possible to leverage the structured nature of the Oireachtas debates data for more distant reading. For example, oral and written parliamentary questions are given appropriate subject headings by Debates Office staff and metadata such as this can be combined with other information, such as the Department to which the question is directed, the Dáil Deputy asking the question, the Deputy's constituency and political party etc.



Leveraging the structured data of headings, Departments, Deputies dates etc. in Tableau Desktop to visualise large tracts of debate and gain a bird's eye view, but at what cost to detail?

The structured nature of the Oireachtas debate data lends itself to being combined with other structured data, such as geoJSON data for constituencies. This opens the possibility of visualisation with numerous tools, including online tools such as <u>RAW Graphs</u>¹⁶ or <u>Flourish</u>¹⁷, or more specialised Javascript tools such as <u>d3.js</u>¹⁸ and <u>Observable</u>¹⁹.

It is vital, however, to avoid being tool-led when dealing with such large datasets because of the risk of nuance being lost.

Network Analysis

olitical text may lend itself to network analysis because of the hierarchical nature of political entities; for example, political parties in the Dáil normally operate on a hierarchical nature according to number of seats held in Parliament and whether a party is in government. Furthermore, the hierarchies extend to geographical representation, with constituencies being represented according to population. Last of all, parliamentary political parties in the Dáil can be seen as the sum of the Deputies elected in the country's constituencies.



Detail of interactive chord dependency diagram of how political parties related to constituencies in the 32nd Dáil, via <u>everviz.com</u>²⁰

Network Graphing

Such network data, compiled with <u>Gephi</u>²¹ or <u>Cytoscape</u>²², could also be tied to other URIs or structured data, such as geographical constituency information, gender metadata etc. to provide richer insights. Ultimately, however, such data and visualisations would tend to inform the investigation of meaning from the debates text rather than extract such meaning, unless visualisations of the text can be incorporated into the network graphing.

Olivia Mitchell Nuala Fennell



Dublin-Central

Kildare-North

May

Detail of online and interactive network graph of all Members of the Dáil since 1919, including name, gender, party affiliation and constituency, via Gephi.

EXTRACTING MEANING.

Combination

t is possible that in order to extract the most meaning from the debates corpus, a combination of some or all of the above, or variations of such techniques, would be best employed. For example, close reading of a text can be supported with digital tools like the <u>Google</u> <u>Cloud Natural Language API²³ or Meaningcloud²⁴ for</u> targeted analysis of a text corpus. Networks could be further enriched by using Member URIs to connect to social network analysis with tools such as <u>NodeXL²⁵</u>. There are also myriad text analysis tools, with <u>Voyant²⁶</u> perhaps

providing the most convenient "one-stop shop" for several methods of text analysis. Individual tools for specific operations, such as <u>CasualConc</u>²⁷, can be used but unless and until specific or custom features are required, Voyant appears the most versatile choice for analysing and comparing text. Despite limitations, it can be a useful method in conjunction with other analytical strategies, including sentiment analysis.



Making Meaning? Building Meaning? Are They Different?

aking meaning from the large corpus of political text that is the Official Report of the Houses of the Oireachtas can never be a simple process. The body of text is so large and the intricacies of political debate so great, it is essentially impossible, even with digital tools and visualisations, to distil its complicated meaning.

Instead of making meaning from the corpus, therefore, it is more appropriate for this project to be about *helping to build meaning*. It is about helping to interpret vast quantities of data - text in the case of the Official Report - that themselves contain the political meaning making of public representatives and the people they represent. To that end, the digital artefact will most likely be a repository of digital and data-driven tools that aim to present political debate in different ways. The tools that will comprise the artefact would preferably be made with open source and sustainable tools, such as the R programming language or the d3.js library, and to work most effectively, they could be presented as complementary to the Oireachtas data. Ideally, they would be presented alongside the Oireachtas data on the official website. Presenting those tools may be a process in itself but they should be portable across platforms; before doing that it is imperative to get the tools right.

This process is not about presenting political meaning making but *re-presenting* it with online tools accessible to members of the public who wish to make use of them. Preliminary data representations are available on <u>changingconversations.net</u>²⁸ under the Change It? menu. Meaning is personal but to build it, the data must be presented, without distortion, in a way that will encourage thought and investigation.

RESOURCE REFERENCES

[1] H. of the Oireachtas, 'Houses of the Oireachtas Debates Office', May 04, 2021. https:// www.oireachtas.ie/en/debates/find (accessed May 04, 2021).

[2] H. of the Oireachtas, 'Houses of the Oireachtas API'. https://api.oireachtas.ie/ (accessed Mar. 19, 2021).

[3] 'KildareStreet.com: Are your TDs and Senators working for you in Ireland's Houses of the Oireachtas?' https://www.kildarestreet.com/ (accessed May 04, 2021).

[4] TheyWorkForYou.com, 'TheyWorkForYou: Hansard and Official Reports for the UK Parliament, Scottish Parliament, and Northern Ireland Assembly - done right', *TheyWorkForYou*. https://www.theyworkforyou.com/ (accessed Mar. 19, 2021).

[5] 'Elasticsearch'. https://www.elastic.co/ (accessed May 04, 2021).

[6] 'Swagger'. https://swagger.io/ (accessed May 04, 2021).

[7] 'Postman | The Collaboration Platform for API Development', *Postman*. https:// www.postman.com/ (accessed May 04, 2021).

[8] 'Tableau Desktop', *Tableau*. https://www.tableau.com/products/desktop (accessed May 04, 2021).

- [9] 'Akoma Ntoso'. http://www.akomantoso.org/ (accessed May 04, 2021).
- [10] 'SubEthaEdit', SubEthaEdit. https://subethaedit.net/ (accessed May 04, 2021).
- [11] 'Sublime Text'. https://www.sublimetext.com/3 (accessed May 04, 2021).
- [12] 'Google Sheets'. https://www.google.com/sheets/about/ (accessed May 04, 2021).

[13] 'Microsoft Excel Spreadsheet Software'. https://www.microsoft.com/en-ie/microsoft-365/excel (accessed May 04, 2021).

[14] 'Tableau Prep', *Tableau*. https://www.tableau.com/products/prep (accessed May 04, 2021).

- [15] 'OpenRefine'. https://openrefine.org/ (accessed May 04, 2021).
- [16] 'Raw Graphs'. https://rawgraphs.io/ (accessed May 04, 2021).
- [17] 'Flourish', Flourish. https://flourish.studio/ (accessed May 04, 2021).
- [18] 'D3.js Data-Driven Documents'. https://d3js.org/ (accessed May 04, 2021).
- [19] 'Observable'. https://observablehq.com/ (accessed May 04, 2021).
- [20] 'Everviz'. https://app.everviz.com (accessed May 04, 2021).
- [21] 'Gephi'. https://gephi.org/ (accessed May 04, 2021).

[22] Morris, 'Cytoscape Automation', *Genome Biol*, vol. 20, no. 1, p. 185, Dec. 2019, doi: 10.1186/s13059-019-1758-4.

[23] 'Cloud Natural Language', *Google Cloud*. https://cloud.google.com/natural-language (accessed May 04, 2021).

[24] 'MeaningCloud', MeaningCloud. https://www.meaningcloud.com/ (accessed May 04, 2021).

[25] 'NodeXL', *Social Media Research Foundation*. https://www.smrfoundation.org/ (accessed May 04, 2021).

- [26] 'Voyant Tools'. https://voyant-tools.org/ (accessed May 04, 2021).
- [27] 'CasualConc'. https://sites.google.com/site/casualconc/ (accessed May 04, 2021).

[28] 'Changing Conversations', *Changing The Conversation*. https://changingconversations.net/ (accessed May 04, 2021).