



Legislative
Assembly
of Ontario



Assemblée
législative
de l'Ontario

USING PARSEHUB TO (HELP) STRUCTURE DATA, SELECT RESOURCES, AND BULK CATALOGUE

PRESENTED BY VICKY LUU
LEGISLATIVE ASSEMBLY OF ONTARIO

GOVERNMENT INFORMATION DAY(S)
DECEMBER 13, 2023





AGENDA

Introduction: ParseHub

Examples

- Orders in Council selection list
- Bulk/retrospective cataloguing
- Structuring Hansard transcripts

INTRODUCTION: PARSEHUB

- ParseHub is a web scraping tool that extracts data from websites using machine learning

The screenshot shows the ParseHub web scraping tool interface. On the left, a sidebar contains a 'main_template' section with various extraction rules like 'Select page (NA)', 'Select OIC', 'Extract name', 'Extract url', 'Relative Ministry', 'Relative Act', and 'Relative DateApproved'. Below these is a 'Contents' section with rules for clicking items and navigating. A 'Get Data' button is at the bottom of the sidebar. The main area displays the target website, 'Orders in Council' on the Ontario government website. The website content includes the Ontario coat of arms, a title 'Orders in Council', and a description: 'An Order in Council is a legal order made by the Lieutenant Governor, on the advice of the Premier or a Minister.' Below this is a search bar and a list of search results for 'Orders in Council' with details like 'Order in Council 1737/2023' and 'Ministry of Mines'. At the bottom, there are links for 'API', 'Tutorials', and 'Contact', and a footer with 'Live preview is disabled when working on interactive templates. Use Test Run instead.' and checkboxes for 'Show more data' and 'Visuals enabled (advanced)'.

USING PARSEHUB TO REDUCE REPETITIVE ACTIONS

- Have many pages that need to be checked
- So, lots of clicking and back and forth
- Depending on what needs to be done, also lots of copying and pasting

The screenshot displays the Ontario Government website's 'Orders in Council' page. At the top, there is a navigation bar with 'Home > Government > Orders in Council'. The main heading is 'Orders in Council', followed by a brief explanation: 'An Order in Council is a legal order made by the Lieutenant Governor, on the advice of the Premier or a Minister. Search for digital copies of Orders approved after July 1, 2016 from newest to oldest.' Below this, there are links for 'Contact us' and 'Learn more about Orders in Council'. A search box is visible with the text 'Search by keyword' and '14,035 results'. A list of recent Orders in Council is shown, including 'Order in Council 1737/2023' (Ministry of Mines, Mining Act, November 23, 2023), 'Order in Council 1736/2023' (Ministry of Tourism, Culture and Sport, Niagara Parks Act, November 23, 2023), 'Order in Council 1735/2023' (Ministry of the Solicitor General, Strengthening Safety and Modernizing Justice Act, 2023, Provincial Animal Welfare Services Act, 2019, November 23, 2023), and 'Order in Council 1734/2023' (Ministry of Colleges and Universities, Less Red Tape, Stronger Economy Act, 2023, Private Career Colleges Act, 2005, Employment Standards Act, 2000, Licence Appeal Tribunal Act, 1999, Ministry of Training, Colleges and Universities Act, Occupational Health and Safety Act). Two large blue arrows point from the list to a detailed view of 'Order in Council 1736/2023'. This detailed view shows the full text of the order, including the date (November 23, 2023) and the specific details of the lease agreement with Niagara Adventure Excursions Inc. (NAE) for zip line and aerial adventure course attractions at Niagara Parks.

USING PARSEHUB TO REDUCE REPETITIVE ACTIONS

- Have many pages that need to be checked
- So, lots of clicking and back and forth
- Depending on what needs to be done, also lots of copying and pasting

The House met at 2 p.m.

Prayers.

MINISTER'S RETURN TO HOUSE

Hon. Mr. Wells: Mr. Speaker, just before the business of the day: as has already been noted, it is, I am sure, a pleasure for all of us to see the Minister of Health (Mr. Norton) back.

Hon. Miss Stephenson: Healthy.

Hon. Mr. Wells: He is back healthy, and I am sure he is going to talk to members about nursing homes today.

Mr. Speaker: I hope at not too great a length. It is nice to see you back.

Hon. Mr. Norton: I have a lot of catching up to do.

Mr. Speaker: You have, indeed.

...er, at the distance to my statement than simply
...er, at the ... like to extend my thanks to the
... who during ... months have expressed to me in
... nt at a time ... assure them it was both very
... rders on the ... who co-operated with our
... the time for ... of the Ministry of Health,
... py if they had ... harder nosed about it and our
... es.

... under ... what I mean and the sincerity of it when I
... ment House leader who, in addition to his already
... er of Intergovernment Affairs assumed the
... tence, and acted as Minister of Health for a period of two

Was this page helpful? Yes No

USING PARSEHUB TO REDUCE REPETITIVE ACTIONS

- Have many pages that need to be checked
- So, lots of clicking and back and forth
- Depending on what needs to be done, also lots of copying and pasting

The screenshot shows the Ontario Newsroom website interface. At the top, there is a navigation bar with 'Ontario' and 'Newsroom' labels, along with search and menu icons. Below this, the 'Natural Resources and Forestry' section is highlighted. A 'Filter Items' sidebar on the left allows users to filter by 'News Type' (set to 'News Release') and 'Topic'. A '1629 Results Found' box is visible. The main content area displays a list of news releases, with the top one titled 'Ontario Protecting the Economy and Environment by Taking Action Against Invasive Species'. A large blue arrow points from the search filter area to a detailed view of this news release. The detailed view includes a 'Table of Contents' with links to 'Content', 'Quick Facts', 'Quotes', 'Additional Resources', and 'Related Topics'. It also features a 'Share' section with social media icons and a 'Media Contacts' section with contact information for the Media Desk and Minister's Office.

EXAMPLES



Compile inventory list of Orders in Council to aid selection



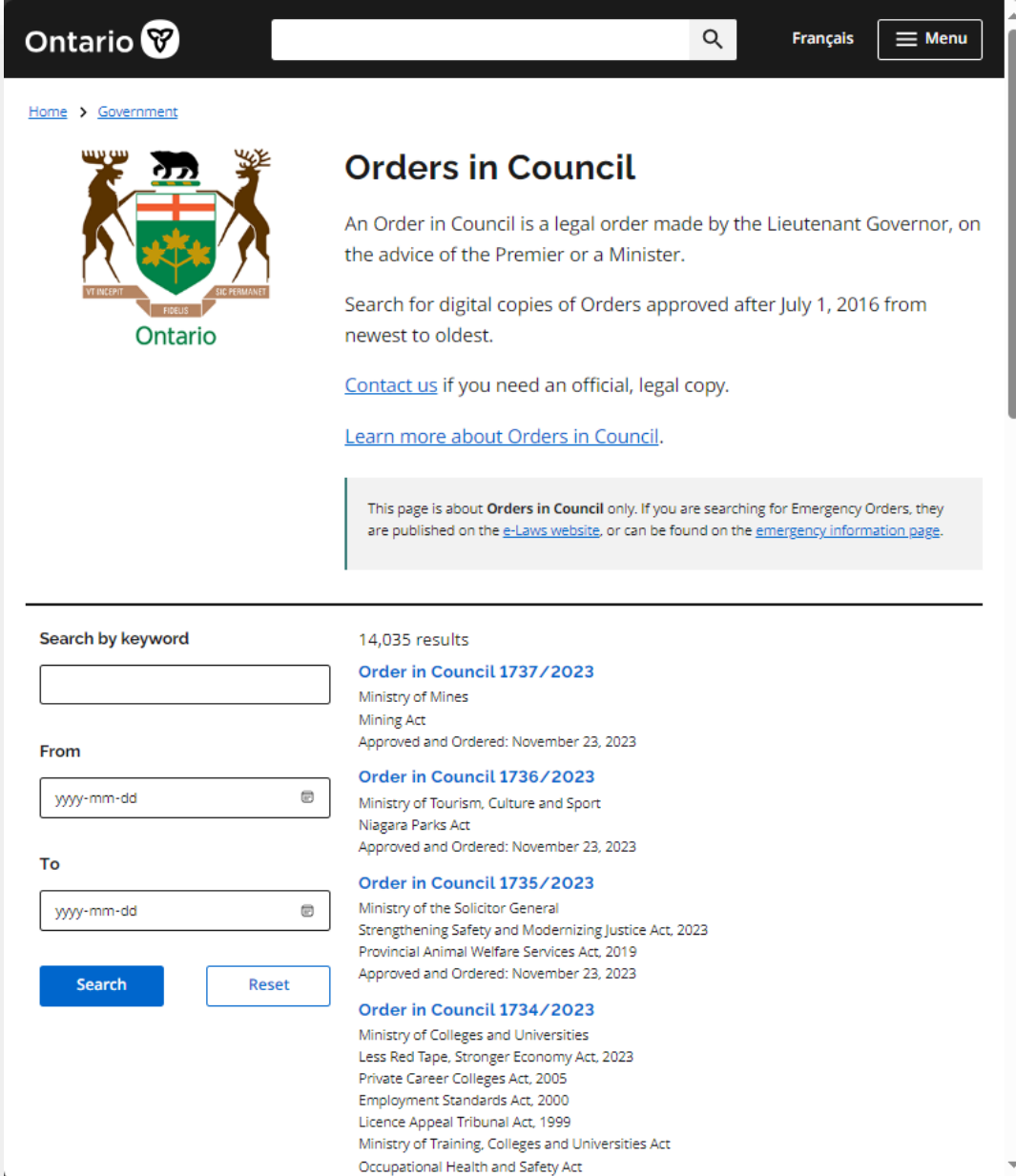
Bulk catalogue resources (retrospective cataloguing)



Structure data for Advanced Hansard Search Modernization project

USING PARSEHUB TO SELECT RESOURCES (EX. ORDERS IN COUNCIL)

- Metadata Services monitors the Orders in Council (OICs) database
- OICs that meet the Library's selection criteria are requested from Cabinet Office



Ontario

Home > Government

Orders in Council

An Order in Council is a legal order made by the Lieutenant Governor, on the advice of the Premier or a Minister.

Search for digital copies of Orders approved after July 1, 2016 from newest to oldest.

[Contact us](#) if you need an official, legal copy.

[Learn more about Orders in Council.](#)

This page is about **Orders in Council** only. If you are searching for Emergency Orders, they are published on the [e-Laws website](#), or can be found on the [emergency information page](#).

Search by keyword

14,035 results

Order in Council 1737/2023
Ministry of Mines
Mining Act
Approved and Ordered: November 23, 2023

Order in Council 1736/2023
Ministry of Tourism, Culture and Sport
Niagara Parks Act
Approved and Ordered: November 23, 2023

Order in Council 1735/2023
Ministry of the Solicitor General
Strengthening Safety and Modernizing Justice Act, 2023
Provincial Animal Welfare Services Act, 2019
Approved and Ordered: November 23, 2023

Order in Council 1734/2023
Ministry of Colleges and Universities
Less Red Tape, Stronger Economy Act, 2023
Private Career Colleges Act, 2005
Employment Standards Act, 2000
Licence Appeal Tribunal Act, 1999
Ministry of Training, Colleges and Universities Act
Occupational Health and Safety Act

From: yyyy-mm-dd

To: yyyy-mm-dd

Search Reset

SETTING UP PARSEHUB

- Takes time, but was easy✂ to set up
 - Point and click for the most part

✂ If I run into something I don't know how to do, I can usually find a solution using ParseHub's support page or by contacting them.

The screenshot shows the ParseHub search interface. On the left, there is a search form with a 'Search by keyword' field, a 'From' date range selector (Year, Month, Day), and a 'To' date range selector (Year, Month, Day). A blue 'Search' button is located below the form. On the right, the search results are displayed, showing 10,616 results. The first five results are highlighted with orange boxes and include the following information:

- Order in Council 192/2022**
Ministry of Finance
Made by Royal Prerogative
Approved and Ordered: February 03, 2022
- Order in Council 191/2022**
Ministry of Finance
Made by Royal Prerogative
Approved and Ordered: February 03, 2022
- Order in Council 190/2022**
Ministry of the Environment, Conservation and Parks
Legislation Act, 2006
Approved and Ordered: February 03, 2022
- Order in Council 189/2022**
Ministry of Labour, Training and Skills Development
Workplace Safety and Insurance Act, 1997
Approved and Ordered: February 03, 2022
- Order in Council 188/2022**
Ministry of the Attorney General

The screenshot shows a text input field labeled 'Expression' containing the following text: `$e.text.toLowerCase().contains("next")`. A small grid icon is visible in the bottom right corner of the field.

The screenshot shows a text input field labeled 'Expression' containing the following text: `$e.text.toLowerCase().contains("français")`. A small grid icon is visible in the bottom right corner of the field.

RAW EXPORT – CSV OR JSON file

Raw CSV file

A	B	C	D	E	F
OIC_name	OIC_url	OIC_Ministry	OIC_DateApproved	OIC_Description	OIC_PostDate
Order in Council 192/2022	https://www.ontario.ca/orders-in-council/192/2022	Ministry of Finance	03-Feb-22	On the recommendation of the undersigned,	Published: February 17, 2022
Order in Council 191/2022	https://www.ontario.ca/orders-in-council/191/2022	Ministry of Finance	03-Feb-22	On the recommendation of the undersigned,	Published: February 17, 2022
Order in Council 190/2022	https://www.ontario.ca/orders-in-council/190/2022	Ministry of the Environment, Conservation and Forestry	03-Feb-22	On the recommendation of the undersigned,	Published: February 17, 2022
Order in Council 189/2022	https://www.ontario.ca/orders-in-council/189/2022	Ministry of Labour, Training and Skills Development	03-Feb-22	On the recommendation of the undersigned,	Published: February 17, 2022
Order in Council 188/2022	https://www.ontario.ca/orders-in-council/188/2022	Ministry of the Attorney General	03-Feb-22	On the recommendation of the undersigned,	Published: February 17, 2022
Order in Council 187/2022	https://www.ontario.ca/orders-in-council/187/2022	Premier and President of the Queen's Privy Council for Canada	03-Feb-22	On the recommendation of the undersigned,	Published: February 17, 2022
Order in Council 186/2022	https://www.ontario.ca/orders-in-council/186/2022	Ministry of the Solicitor General	03-Feb-22	On the recommendation of the undersigned,	Published: February 17, 2022
Order in Council 185/2022	https://www.ontario.ca/orders-in-council/185/2022	Ministry of the Solicitor General	03-Feb-22	On the recommendation of the undersigned,	Published: February 17, 2022
Order in Council 184/2022	https://www.ontario.ca/orders-in-council/184/2022	Ministry of the Solicitor General	03-Feb-22	On the recommendation of the undersigned,	Published: February 17, 2022
Order in Council 183/2022	https://www.ontario.ca/orders-in-council/183/2022	Ministry of Health	03-Feb-22	On the recommendation of the undersigned,	Published: February 17, 2022
Order in Council 182/2022	https://www.ontario.ca/orders-in-council/182/2022	Ministry of Health	03-Feb-22	On the recommendation of the undersigned,	Published: February 17, 2022
Order in Council 181/2022	https://www.ontario.ca/orders-in-council/181/2022	Ministry of Health	03-Feb-22	On the recommendation of the undersigned,	Published: February 17, 2022
Order in Council 180/2022	https://www.ontario.ca/orders-in-council/180/2022	Ministry of Health	03-Feb-22	On the recommendation of the undersigned,	Published: February 17, 2022

CLEANING THE DATA

Clean data in Power Query

The screenshot displays the Microsoft Power Query Editor interface. The ribbon at the top includes the following groups: Query (Close, Refresh Preview, Properties, Advanced Editor), Manage Columns (Choose Columns, Remove Columns), Reduce Rows (Keep Rows, Remove Rows), Sort (Remove Duplicates, Remove Errors), Transform (Split Column, Group By, Data Type: Text, Use First Row As Headers, Replace Values), Combine (Merge Queries, Append Queries, Combine Binaries), and New Query (New Source, Recent Sources).

The formula bar contains the following M code: `= Table.TransformColumnTypes(#"Split Column by Delimiter",{{"OIC_name.2 - Copy.1", Int64.Type}, {"OIC_name.2 - Copy.2", Int64.Type}})`

The data table below shows the following columns: OIC_name.2, OIC_url_856uOnline_resource, OIC_Ministry, OIC_DateApproved, OIC_Description.2, OIC_PostDate.2, OIC_na..., and OIC_na... The table contains 14 rows of data.

	OIC_name.2	OIC_url_856uOnline_resource	OIC_Ministry	OIC_DateApproved	OIC_Description.2	OIC_PostDate.2	OIC_na...	OIC_na...
1	87/2022	https://www.ontario.ca/orders-in-council/	Ministry of Labour, Training and Skills Devel	2022-01-19	Whereas Brian Oâ€™Byrne has been appointec	2022-02-02	87	202:
2	86/2022	https://www.ontario.ca/orders-in-council/	Ministry of the Solicitor General	2022-01-19	Pursuant to section 27 of the Police Services Ac	2022-02-02	86	202:
3	85/2022	https://www.ontario.ca/orders-in-council/	Ministry of Labour, Training and Skills Devel	2022-01-19	Pursuant to subsection 110(2) of the Labour Re	2022-02-02	85	202:
4	84/2022	https://www.ontario.ca/orders-in-council/	Premier and President of the Council	2022-01-19	Pursuant to section 4 of the Royal Ontario Mus	2022-02-02	84	202:
5	83/2022	https://www.ontario.ca/orders-in-council/	Premier and President of the Council	2022-01-19	Pursuant to section 3 of the McMichael Canadi	2022-02-02	83	202:
6	82/2022	https://www.ontario.ca/orders-in-council/	Premier and President of the Council	2022-01-19	Pursuant to subsections 2 (2), 5 (1) and 8 (1) of	2022-01-27	82	202:
7	81/2022	https://www.ontario.ca/orders-in-council/	Premier and President of the Council	2022-01-13	Pursuant to subsections 5(1) and 5(4) of the To	2022-01-27	81	202:
8	80/2022	https://www.ontario.ca/orders-in-council/	Premier and President of the Council	2022-01-13	Pursuant to section 5 of the Ontario Heritage A	2022-01-27	80	202:
9	79/2022	https://www.ontario.ca/orders-in-council/	Premier and President of the Council	2022-01-13	Pursuant to section 6 of the Liquor Control Boa	2022-01-27	79	202:
10	78/2022	https://www.ontario.ca/orders-in-council/	Premier and President of the Council	2022-01-13	Pursuant to section 6 of the Liquor Control Boa	2022-01-27	78	202:
11	77/2022	https://www.ontario.ca/orders-in-council/	Premier and President of the Council	2022-01-13	Pursuant to section 6 of the Liquor Control Boa	2022-01-27	77	202:
12	76/2022	https://www.ontario.ca/orders-in-council/	Premier and President of the Council	2022-01-13	Pursuant to section 6 of the Liquor Control Boa	2022-01-27	76	202:
13	75/2022	https://www.ontario.ca/orders-in-council/	Premier and President of the Council	2022-01-13	Pursuant to section 2 of the Financial Services	2022-01-27	75	202:
14	74/2022	https://www.ontario.ca/orders-in-council/	Premier and President of the Council	2022-01-13	Pursuant to section 2 of the Financial Services	2022-01-27	74	202:

INVENTORY LIST OF ORDERS IN COUNCIL

Added columns

	A	B	C	D		G	H	I	J	K	
	OIC	Coll	Rev	URL	Ministry	Date Approved	Acq(s) / OIC Contents	Date Post	OIC N	Gap?	Year
3	82/2022			https://www.ontario.ca/orders-in-council	Premier and President of the Council	2022-01-13	Pursuant to subsections 2 (2), 5 (1) and 8	2022-01-27	82		2022
4	83/2022			https://www.ontario.ca/orders-in-council	Premier and President of the Council	2022-01-19	Pursuant to section 3 of the McMichael C	2022-02-02	83		2022
5	84/2022			https://www.ontario.ca/orders-in-council	Premier and President of the Council	2022-01-19	Pursuant to section 4 of the Royal Ontar	2022-02-02	84		2022
6	85/2022			https://www.ontario.ca/orders-in-council	Ministry of Labour, Training and Skills	2022-01-19	Pursuant to subsection 110(2) of the Lab	2022-02-02	85		2022
7	86/2022			https://www.ontario.ca/orders-in-council	Ministry of the Solicitor General	2022-01-19	Pursuant to section 27 of the Police Ser	2022-02-02	86		2022
8	87/2022			https://www.ontario.ca/orders-in-council	Ministry of Labour, Training and Skills	2022-01-19	Whereas Brian Oâ€™Byrne has been ap	2022-02-02	87	Missing	2022
9	90/2022			https://www.ontario.ca/orders-in-council	Minister for Seniors and Accessibility	2022-01-27	Whereas subsection 41 (5) of the Access	2022-02-10	90		2022
0	91/2022			https://www.ontario.ca/orders-in-council	Ministry of Seniors and Accessibility	2022-01-27	On the recommendation of the undersig	2022-02-10	91		2022
1	92/2022			https://www.ontario.ca/orders-in-council	Ministry of Economic Development, J	2022-01-27	Whereas eight positions of Special Advi	2022-02-10	92		2022
2	93/2022			https://www.ontario.ca/orders-in-council	Ministry of the Attorney General	2022-01-27	Pursuant to subsection 45.14(2) of the H	2022-02-10	93		2022
3	94/2022			https://www.ontario.ca/orders-in-council	Ministry of the Attorney General	2022-01-27	Pursuant to subsection 45.14(2) of the H	2022-02-10	94		2022
4	95/2022			https://www.ontario.ca/orders-in-council	Ministry of the Attorney General	2022-01-27	Pursuant to subsection 45.14(2) of the H	2022-02-10	95		2022
5	96/2022			https://www.ontario.ca/orders-in-council	Ministry of the Attorney General	2022-01-27	Pursuant to subsection 45.14(2) of the H	2022-02-10	96		2022
6	97/2022			https://www.ontario.ca/orders-in-council	Ministry of the Attorney General	2022-01-27	Pursuant to subsection 2(3) of the Licen	2022-02-10	97		2022
7	98/2022			https://www.ontario.ca/orders-in-council	Ministry of the Attorney General	2022-01-27	Pursuant to subsection 2(3) of the Licen	2022-02-10	98		2022
8	99/2022			https://www.ontario.ca/orders-in-council	Ministry of the Attorney General	2022-01-27	Pursuant to subsection 2(3) of the Licen	2022-02-10	99		2022
9	100/2022			https://www.ontario.ca/orders-in-council	Ministry of Economic Development, J	2022-01-27	Whereas by Order in Council 956/2019, €	2022-02-10	100		2022
0	101/2022			https://www.ontario.ca/orders-in-council	Ministry of Economic Development, J	2022-01-27	Whereas by Order in Council 956/2019, €	2022-02-10	101		2022
1	102/2022			https://www.ontario.ca/orders-in-council	Ministry of the Environment, Conserv	2022-01-27	Pursuant to section 3 of Ontario Regulat	2022-02-10	102		2022
2	103/2022			https://www.ontario.ca/orders-in-council	Ministry of the Environment, Conserv	2022-01-27	Pursuant to section 3 of Ontario Regulat	2022-02-10	103		2022

USING PARSEHUB TO STRUCTURE DATA (EX. HANSARD TRANSCRIPTS)

- The Hansard transcript is the official record of House debates
- Hansards usually include:
 - table of contents,
 - transcript of what is said,
 - votes, etc.
- There are approximately **4000** HTML pages on ola.org

The House met at 0900.

The Speaker (Hon. Dave Levac): Please join me in prayer.

Prayers.

ORDERS OF THE DAY

ACCEPTING SCHOOLS ACT, 2012 / LOI DE 2012 POUR DES ÉCOLES TOLÉRANTES

Resuming the debate adjourned on March 26, 2012, on the motion for second reading of Bill 13, An Act to amend the Education Act with respect to bullying and other matters / Projet de loi 13, Loi modifiant la Loi sur l'éducation en ce qui a trait à l'intimidation et à d'autres questions.

The Speaker (Hon. Dave Levac): Further debate?

Ms. Cheri DiNovo: It's an honour and a privilege to stand today to speak about Bill 13.

Quite frankly, this is not a partisan issue. There is, of course, sentiment all around the House that what we want to do and what we stand together on, shoulder to shoulder, as it were, is to keep our children safe, to prevent teen suicide, the kind of occurrence that happened with our friend Jamie and others who have killed themselves because of bullying in their schools. We are speaking really together around this issue.

I want to send some particular notes of respect out to a number of groups: first and foremost, the students themselves—I hosted a presser here with a number of them from our separate school system who came, and I'm going to speak about that in a minute; also to the teachers from all the teachers' unions who have stood up for their students, stood up on behalf of their students very bravely and courageously; and for other organizations like Egale that have been outspoken about this; the Falconer report and others who have pressed this government and pressed all of us to actually take action.

I also want to thank those who have crafted bills; we've seen two of them on this issue. We're looking forward to being in committee to deal with clause-by-clause examination of those bills, because all of this is important. Particular kudos go, I think, to the Ontario English Catholic Teachers' Association, who have been unbelievably brave and courageous in their stance for students.

As I said, I want to mention some students in particular: the students who came to Queen's Park. This was in a presser we just had about a week ago. We had here some incredibly courageous young people. We had Trevor James, who is a straight youth from Peterborough, who "got up at 4:30 a.m. Wednesday and took a two-hour bus ride to Toronto to talk about school clubs for gay students."

SETTING UP PARSEHUB HANSARD TRANSCRIPTS

- Due to slight variations in the HTML code
 - i.e. difference in how h2, h3 headings are coded
- Had to create three different projects

The screenshot displays the ParseHub interface for a project named 'Hansard Transcripts 2011-M'. The left sidebar shows a list of extraction rules for 'main_template' and 'HansardTra...'. The main area shows a preview of the source page with green boxes highlighting extracted content. The extracted content includes:

- DES ÉLECTIONS EN ONTARIO
- FIRE PROTECTION AND PREVENTION AMENDMENT ACT, 2011 / LOI DE 2011 MODIFIANT LA LOI SUR LA PRÉVENTION ET LA PROTECTION CONTRE L'INCENDIE
- The House recessed from 1803 to 1845.
- ORDERS OF THE DAY
- ENSURING INTEGRITY IN ONTARIO ELECTIONS ACT, 2011 / LOI DE 2011 ASSURANT L'INTÉGRITÉ DES ÉLECTIONS EN ONTARIO
- Resuming the debate adjourned on May 30, 2011, on the motion for second reading of Bill 196, An Act to amend the Election Act with respect to certain electoral practices / Projet de loi 196, Loi modifiant la Loi électorale en ce qui concerne certaines manoeuvres électorales.

At the bottom, there is a table showing the extracted data in CSV/Excel format:

urls	p	selection1_...	selection1_...	h2	h3	Date	page
https://www.o.../en/legislative-business/house-documents/parliament-40/session-1/2012-03-29	The House adjourned at 2145.			ORDERS OF THE DAY	FIRE PROTECTION AND PREVENTION AMENDMENT ACT, 2011 / LOI DE 2011 MODIFIANT LA	May 30, 2011	https://www.o.../en/legislative-business/house-documents/parliament-39/session-2/2011-05-30

Below the table, there is a message: "This is a live preview. When you are ready to run your project, click Get Data." and two checkboxes: "Show more data" and "Visuals enabled (advanced)".

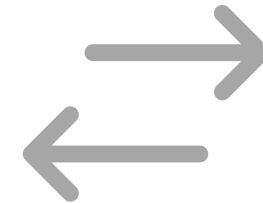
WHAT DID PARSEHUB DO FOR ME?



**Automated clicking
“Next page” and
flipping between pages**



Extracted data from
multiple pages
into a single file



Organize extracted
data into appropriate
columns/fields

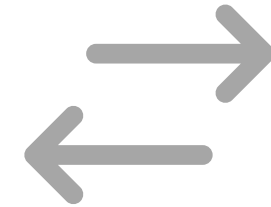
WHAT DID PARSEHUB DO FOR ME?



Automated clicking
“Next page” and
flipping between pages



**Extracted data from
multiple pages
into a single file**



Organize extracted
data into appropriate
columns/fields

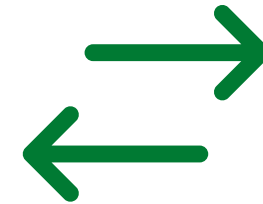
WHAT DID PARSEHUB DO FOR ME?



Automated clicking on
“next page” and flipping
back and forth



Extracted data from
multiple pages
into a single file



**Organize extracted
data into appropriate
columns/fields**

... and more!

Legislative
Assembly
of Ontario



Assemblée
législative
de l'Ontario

THANK YOU MERCI

Questions? Comments?

Email: vluu@ola.org