

JAAP KAMPS

UNIVERSITY OF AMSTERDAM

TDDL @ TPDL'17, THESSALONIKI, 21 SEP 2017

TAKING TIME SERIOUSLY: DIACHRONIC COLLECTIONS IN DIGITAL LIBRARIES

ABSTRACT

- ▶ Digital library collections may appear static, but usually contain content from many different time periods.
- ▶ What changes when your collection is a diachronic corpus spanning decades or even centuries?
 - ▶ The temporal structure is a lightweight annotation that provides powerful new handles for users to explore the collection, and engage with the material from many different angles.
 - ▶ There are also fundamental consequences for the design of digital library systems and back office systems empowering curators and digital humanities researchers to explore the data at scale.
- ▶ We study this question for collections of parliamentary proceedings spanning multiple centuries in a uniform format.

**LET'S MAKE
THINGS CONCRETE**

Exploratory Political Search (ExPoSe)

[← Naar de lijst van onderzoeksprojecten](#)

Samenvatting

The Parliamentary Proceedings of a country provide a backbone of its textual cultural heritage. It covers centuries, in a standardized format, with highly controlled quality. Its completeness as a longitudinal corpus and its content covering a large part of a nations history, makes the proceedings both of interest in their own right, and as a tool to contextualize other data collections. The Dutch Parliamentary Proceedings are available in digitized format starting from 1576

Mail
 Delen

Print

Kenmerken

Projectnummer
314-99-108

Hoofdaanvrager
Dr. ir. J. Kamps

Hansards (Historians, Data Journalists)

Digitized heritage texts into a connected network of information.
How to exploit the obtained networked structure for interactive exploratory search?

The ExPoSe project aims to bring out the full potential of the Dutch Proceedings by transforming them into a structured XML format. To really function as a backbone, it must be easy to connect other documents deeply into the proceedings. We make this possible by connecting the content of the Proceedings to the Linked Open Data cloud, which we extend with several historically relevant sources (among them the huge newspaper collection of the KB and the WNT). The project will run for 5 years and employ 2 PhD students and one .5 FTE programmer. The consortium contains the complete chain from raw data producer to end-users in the Creative Industry. It consists of 5 public and 2 private partners, contributing 315K and 83 K Euro, respectively.

Afvoer: en
informatiewetenschap

Uitvoerders

/

Looptijd
01/10/2013 tot 30/09/2018

Budget
€ 597.136



Netherlands Organisation for Scientific Research



ExPoSe Consortium



UNIVERSITEIT VAN AMSTERDAM

CWI

Centrum Wiskunde & Informatica



nationaal archief



DIGGING INTO DATA CHALLENGE

Awardees for 2013

TRANSATLANTIC DIGGING INTO DATA CHALLENGE 2013 WINNERS ANNOUNCED

January 15, 2014—Today, ten international research funders from four countries jointly announced the winners of the third Digging into Data Challenge, a competition to develop new insights, tools and skills in innovative humanities and social science research using large-scale data analysis.

[MORE »](#)

Digging into Linked Parliamentary Data

 Search

[Home](#) [About](#) [Blog](#) [Contact](#) [News and events](#)

Parliamentary proceedings reflect our history from centuries ago to the present day. They exist in a common format that has survived the test of time, and reflect any event of significance (through times of war and peace, of economic crisis and prosperity). With carefully curated proceedings becoming available in digital form in many countries, new research opportunities arise to analyse this data, on an unprecedented longitudinal scale, and across different nations, cultures and systems of political representation.

Focusing on the UK, Canada and The Netherlands, this project will deliver a common format for encoding parliamentary proceedings (with an initial focus on 1800 to yesterday); a joint dataset covering all three jurisdictions; a workbench with a range of tools for the comparative, longitudinal study of parliamentary data; and substantive case studies



RECENT POSTS

- [Accelerating and validating ex post facto hypothesis formulation](#)



Standard search focuses on the content exclusively

Essence is not only what is said, but also by who and to whom, and why

Parliamentary Proceedings

NOS.nl



wil ik dat graag rechtgezet hebben, maar als dat wel zo is, wil ik daarvan in de meest felle termen afstand nemen. Als ik de heer De Roon verkeerd citeer, hoor ik het graag.

De heer Wilders (PVV):

Ik houd de minister-president net een uitspraak voor en

aap is. Als dat niet zo is, wil ik graag dat hij dat nu ont-kent, maar zo is het in de media gekomen, zo is het hier gezegd in het Vragenuurtje en ik neem daar in de meest felle termen afstand van. Zo kunnen wij niet praten over de minister-president van een bevriend land.

Parliamentary Proceedings

aap". Het was mijn collega die zei: "Daar komt de islamitische aap uit de mouw, en hij heet..." Dat is de beeld-spraak.

Minister Rutte:

Ach ...

De heer Wilders (PVV):

Wat, ach? Doe eens normaal, man! Wat, ach?

Minister Rutte:

Doe eens normaal, man? Doe lekker zelf normaal! Tsjonge jonge!

De heer Wilders (PVV):

Ja, lezen voordat u wat zegt!

Minister Rutte:

Ik zou zeggen: doet u zelf eens normaal, mijnheer Wilders!

De heer Wilders (PVV):

Nee, doet u eens normaal!

gan". Dat is dus niet het noemen van een aap.

Minister Rutte:

Wat is nu het verschil?

De heer Wilders (PVV):

Dat is zeker een verschil. Het een is een uitdrukking, het ander een directe kwalificatie. Ik blijf erbij dat dat niet zo is gezegd en dat de minister-president de heer De Roon echt onrecht doet als hij hem dat in de ... Ik geloof de heer De Roon als hij tegen mij zegt hoe hij het heeft gezegd. Hij zat er zelf bij.

Minister Rutte:

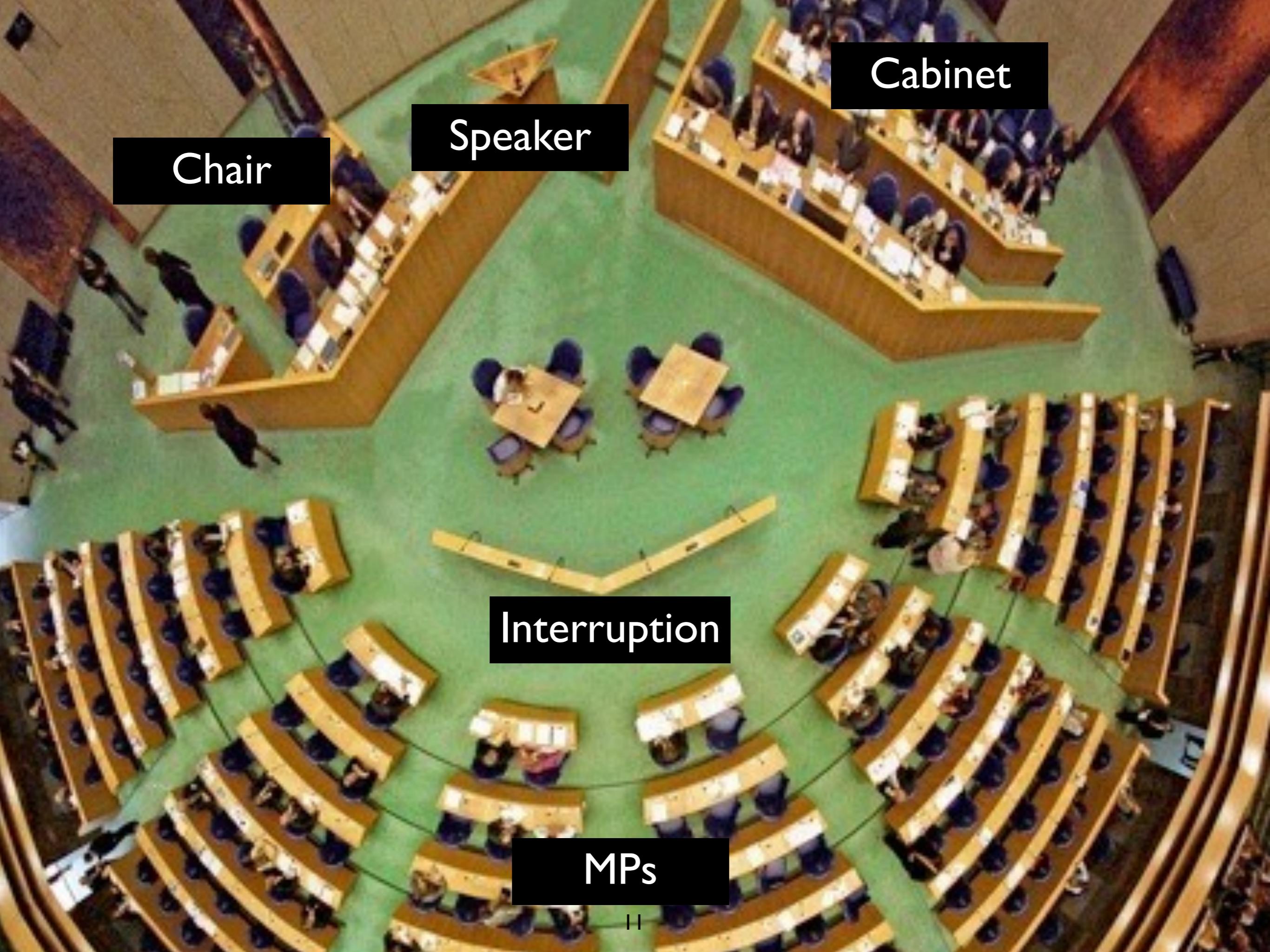
Mag ik even reageren?

De voorzitter:

Dat punt is gemaakt nu. De reactie van de minister-president.

Minister Rutte:

Op dat punt: ik vind ook de uitspraak: "Daar komt de islamitische aap uit de mouw en die heet Erdogan" eerlijk gezegd net zo erg. Dus voorzitter, ik handhaaf mijn op-



Chair

Speaker

Cabinet

Interruption

MPs

wil ik dat graag rechtgezet hebben, maar als dat wel zo is, wil ik daarvan in de meest felle termen afstand nemen. Als ik de heer De Roon verkeerd citeer, hoor ik het graag.

De heer **Wilders** (PVV):

Ik houd de minister-president net even voor mij. Dan zeeg hij: ik reageer niet voordat ik [redactie] Chair [redactie] inderdaad [redactie] zie. [redactie] vertold die zei: "een islamitische aap". Net was mijn collega die zei: "Daar komt de islamitische aap uit de mouw, en hij heet..." Dat is de beeldspraak.

Minister **Rutte**:

Ach ...

De heer **Wilders** (PVV):

Wat, ach? Doe eens normaal, man! Wat, ach?

Minister **Rutte**:

Doe eens normaal, man? Doe lekker zelf normaal! Tsjonge jonge!

De heer **Wilders** (PVV):

Ja, lezen voordat u wat zegt!

Minister **Rutte**:

Ik zou zeggen: doet u zelf eens normaal, mijnheer Wilders!

De heer **Wilders** (PVV):

Nee, doet u eens normaal!

Speaker

Cabinet

De heer **Wilders** (PVV):

Voorzitter. Zo is het niet gezegd. Er is gezegd: "Daar komt de islamitische aap uit de mouw, en hij heet de heer Erdogan". Dat is dus niet het noemen van een aap.

Minister **Rutte**:

Wat is nu het verschil?

De heer **Wilders** (PVV):

Dat is zeker een verschil. Het een is een uitdrukking, het ander een directe kwalificatie. Ik blijf erbij dat dat niet zo is gezegd en dat de minister-president de heer De Roon echt onrecht doet als hij hem dat in de ... Ik geloof de heer De Roon als hij tegen mij zegt hoe hij het heeft gezegd. Hij zat er zelf bij.

Minister **Rutte**:

Mag ik even reageren?

De **voorzitter**:

Dat punt is gemaakt nu. De reactie van de minister-president.

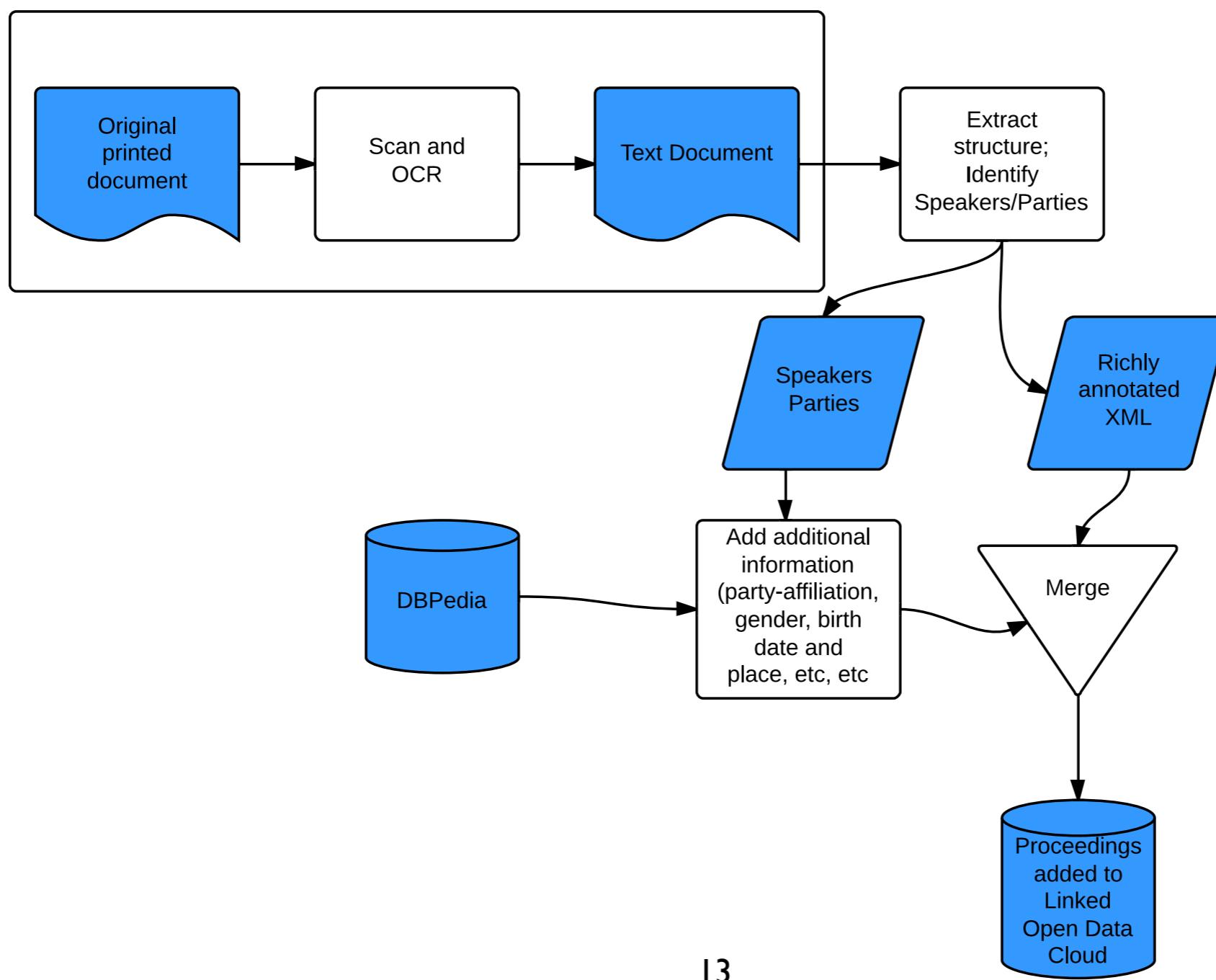
Interruption

Minister **Rutte**:

Op dat punt: ik vind ook de uitspraak: "Daar komt de islamitische aap uit de mouw en die heet Erdogan" eerlijk gezegd net zo erg. Dus voorzitter, ik handhaaf mijn op-

PoliticalMashUp

Earlier project gathered all data (up to yesterday)



50STE VERGADERING

VERGADERING VAN DINSDAG 9 FEBRUARI 1971

(Bijeenroepingsuur 14.00)

Ingekomen stukken. – Verlof verleend tot het houden van een interpellatie aan de heer Imkamp betreffende de kanalisatieplannen voor de Dommel. – Regeling van de

Voorzitter

Van Lier, Van der Louw, Maenen, Masman, Van der Mei, Mellema, Meijer, Mommersteeg, De Mooij, Nederhorst, Notenboom, Nuijens, Nypels, mevrouw Padt-Jansen, de heren Van der Peijl, Peijnenburg, Portheine, Rietkerk, Roethof, Van Rossum, Schakel, Schmelzer, J. N. Scholten, W. Scholten, Schuite-maker, Van der Spek, Stemerdink, Van der Stoel, Van Thiel, Van Thijn, Tilanus, Tolman, Tuijnman, Den Uyl, mevrouw Veder-Smit, de heren Veerman, Vellenga, Verlaan, Visser, Vondeling, Vonhoff, Voogd, Vrolijk, Walburg, Westerterp, Weijers, Weijters, Wiebenga, Wiegel, Wieldraaijer, Wierda, Wierenga, Ter Woorst, jonkvrouwe Wtewaal van Stoetwegen, de heer

Original hansards (1815-1995 printed, 1995-digital)

Wet Premie Kerkenbouw. – Behandeling en verwerping van de motie van de heer Franssen c.s., betreffende mogelijkheden van gemeentelijke subsidie voor de bouw van kerken en bezinningsgebouwen. – Behandeling en aanneming van de wetsontwerpen Wijziging van de artikelen 1637*s* en 1638*r* van het Burgerlijk Wetboek. – Nadere regels met betrekking tot het afleggen van eed of belofte (Eedswet 1969). – Behandeling van het voorstel van wet van de heren Westerterp, Boersma, Van Mierlo, Van Thijn, Wiegel en De Boo tot het in overweging nemen van een voorstel tot verandering in artikel 110 van de Grondwet (Vervroeging Prinsjesdag). – Regeling van de werkzaamheden. – Behandeling en aanneming van het ontwerp van rijkswet Goedkeuring van het op 31 maart 1953 te New York gesloten Verdrag betreffende de politieke rechten van de vrouw (*Trb.* 1968, 92). – Indiening en behandeling van een motie van mevrouw Padt-Jansen c.s., betreffende de wenselijkheid van wijziging van artikel 11 van de Grondwet in die zin dat bij erfopvolging vrouwen en mannen gelijkgerechtigd zijn. – Behandeling en aanneming van

Zaken, Schut, Minister van Volkshuisvesting en Ruimtelijke Ordening, Roolvink, Minister van Sociale Zaken en Volksgezondheid, en Kruisinga, Staatssecretaris van Sociale Zaken en Volksgezondheid.

De Voorzitter: Ik deel aan de Kamer mede, dat zijn ingekomen berichten van verhindering tot het bijwonen van de vergadering van:

de heren C. J. van der Ploeg, Boersma, Bos en Oele, wegens het bijwonen van de vergadering van het Europese Parlement, te Straatsburg, de gehele week;

de heer Van Harselaar, wegens familie-omstandigheden, de gehele week;

de heer S. van der Ploeg, wegens het bijwonen van de begrafenis van zijn moeder;

de heer Van Mierlo, wegens ziekte;

de heer Vonhoff, wegens bezigheden elders;

mevrouw Van Veenendaal-Van Meggelen, wegens verplichtingen elders, de gehele week;

de heer De Boo, wegens verblijf in het buitenland.

Deze berichten worden voor kennisgeving aangenomen.

5QSTE VERGADERING

50ste vergadering -9 februari '71 2661

VERGADERING VAN DINSDAG 9 FEBRUARI 1971 (Bijeenroepingsuur 14.00)

Ingekomen stukken. -Verlof verleend tot het houden van een interpellatie aan de heer Imkamp betreffende de kanalisatieplannen voor de Dommel. -Regeling van de werkzaamheden. -Behandeling en aanneming van het wetsontwerp Wet houdende verlenging werkingsduur Wet Premie Kerkenbouw. -Behandeling en verwerping van de motie van de heer Franssen c.s., betreffende mogelijkheden van gemeentelijke subsidie voor de bouw van kerken en bezinningsgebouwen. -Behandeling en aanneming van de wetsontwerpen Wijziging van de artikelen 1637S en 1638r van het Burgerlijk Wetboek. -Nadere regels met betrekking tot het afleggen van eed of belofte (Eedswet 1969). -Behandeling van het voorstel van wet van de heren Westertern, Boersma, Van Mierlo, Van Thijn, Wiegel en De Boo tot het in overweging nemen van een voorstel tot verandering in artikel

OCR'ed content (1815-1995)

behandeling van een motie van mevrouw Padt-Jansen c.s., betreffende de wenselijkheid van wijziging van artikel 11 van de Grond-wet in die zin dat bij erfopvolging vrouwen en mannen gelijkgerechtigd zijn. " Behandeling en aanneming van het wetsontwerp Regeling met betrekking tot de arbeid ten behoeve van de volkshuishouding, de landsverdediging en de overheidsdienst voor het geval van oorlog, oorlogsgevaar, daaraan verwante of daarmede verband houdende buitengewone omstandigheden (Noodwet Ar-beidsvoorziening). -Aanneming van wetsontwerpen. -Behandeling en aanneming van het wetsontwerp Wijziging van de wet van 10 september 1853, Stb. 102, tot regeling van het toezicht op de onderscheidene kerk-genootschappen. -Ingekomen stukken.

Voorzitter: de heer Van Thiel

Tegenwoordig zijn 118 leden, te weten: de heren Aantjes, Aarden, Abma, Assmann, Bakker, me-juffrouw Barendregt, de heren De Beer, De Bekker, Van den Bergh, Biesheuvel, Boot, Van de Brake, mevrouw Brautigam, de heren Cornelissen, Dankert, mevrouw Dettmeijer-Labberton, de heren Diepenhorst, Dirx, Van Dis, Van den Doel, Dolman, Dusarduijn, Van Dijk, Dijkstra, Egas, Engels, mevrouw Epema-Brugman, de heren Fiévez, Franssen, Geelkerken, Geertsema, Geurtsen, De Goede, Goedhart, Gortzak, mejuffrouw Goud-smit, de heer Goudzwaard, mevrouw Groensmit-Vander Kallen, mejuffrouw Haars, de heren Harmsen, Hermsen, me-vrouw Heroma-Meilink, de heren Hutschemaekers, Imkamp, Jansen, Janssen, Joekes, Jongeling, mejuffrouw Kessel, de heren Kieft, Kikkert, Koekoek, Van Koeverden, mejuffrouw Kok, de heren Kolfschoten, Koudijs, Krosse, Van Laak, Laban, Lam-berts, mejuffrouw Van Leeuwen, de heren Van der Lek, Lems,

Voorzitter Van Lier, Van der Louw, Maenen, Masman, Van der Mei, Mellema, Meijer, Mommersteeg, De Mooij, Nederhorst, Noten-boom, Nijjens, Nypels, mevrouw Padt-Jansen, de heren Van der Peijl, Peijnenburg, Portheine, Rietkerk, Roethof, Van Ros-sum, Schakel, Schmeizer, J. N. Scholten, W. Scholten, Schuite-maker, Van der Spek, Stemerdink, Van der Stoel, Van Thiel, Van Thijn, Tilanus, Tolman, Tuijnman, Den Uyl, mevrouw Veder-Smit, de heren Veerman, Vellenga, Verlaan, Visser, Von-delting, Vonhoff, Voogd, Vrolijk, Walburg, Westerterp, Weijers, Weijters, Wiebenga, Wiegel, Wieldraaijer, Wierda, Wierenga, Ter Woorst, jonkvrouwe Wttewaall van Stoetwegen, de heer Zegering Hadders, en de heren Luns, Minister van Buitenlandse Zaken, Polak, Minister van Justitie, Beernink, Minister van Binnenlandse Zaken, Schut, Minister van

Detecting debate speaker, party, role, ...



The screenshot shows a web browser window with the address bar containing a URL starting with <speech xmlns="http://www.politicalmashup.nl" xmlns:pm="http://www.policalm...". The main content area displays an XML document representing a speech. The XML code is color-coded for syntax highlighting:

```
<?xml version="1.0" encoding="UTF-8"?>
<speech xmlns="http://www.politicalmashup.nl" xmlns:pm="http://www.policalm...
  pm:speaker="van Thiel" pm:function="voorzitter"
  pm:role="chair" pm:party-ref="nl.p.kvp" pm:member-ref="nl.m.01355" pm:id=
  "nl.proc.sgd.d.197019710000647.1.2">
  <p pm:id="nl.proc.sgd.d.197019710000647.1.2.1">Ik deel aan de Kamer mede, dat
    zijn ingekomen berichten van verhindering tot het bijwonen van de
    vergadering van:</p>
  <p pm:id="nl.proc.sgd.d.197019710000647.1.2.2">de heren C. J. van der Ploeg,
    Boersma, Bos en Oele, wegens het bijwonen van de vergadering van het
    Europese Parlement, te Straatsburg, de gehele week; de heer Van Harselaar,
    wegens familie-omstandigheden, de gehele week; de heer S. van der Ploeg,
    wegens het bijwonen van de begrafenis van zijn moeder; de heer Van Mierlo,
    wegens ziekte; de heer Vonhoff, wegens bezigheden elders; mevrouw Van
    Veenendaal-Van Meggelen, wegens verplichtingen elders, de gehele week; de
    heer De Boo, wegens verblijf in het buitenland.</p>
  <p pm:id="nl.proc.sgd.d.197019710000647.1.2.3">Deze berichten worden voor
    kennisgeving aangenomen.</p>
</speech>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<stage-direction xmlns="http://www.politicalmashup.nl" xmlns:pm="http://
www.politicalmashup.nl" pm:id="nl.proc.sgd.d.197019710000647.1.1">
<stage-direction pm:type="pagebreak" pm:id="nl.proc.sgd.d.
197019710000647.1.1.1">
<pagebreak pm:originalpagenr="2661" pm:source="http://resolver.kb.nl/
resolve?urn=sgd:mpeg21:19701971:0000647:0001:image"/>
</stage-direction>
<stage-direction pm:type="header" pm:id="nl.proc.sgd.d.197019710000647.1.1.2">
<p pm:id="nl.proc.sgd.d.197019710000647.1.1.2.1">Interpellatieverzoek van de
heer Imkamp inz. de Dommel 50ste vergadering - 9 februari '71 2661</p>
</stage-direction>
<p pm:id="nl.proc.sgd.d.197019710000647.1.1.3">50STE VERGADERIN G VERGADERING
VAN DINSDAG 9 FEBRUARI 1971 (Bijeenroepingsuur 14.00) Ingekomen stukken. -
Verlof verleend tot het houden van een interpellatie aan de heer Imkamp
betreffende de kanalisatieplannen voor de Dommel. - Regeling van de
werkzaamheden. - Behandeling en aanneming van het wetsontwerp Wet houdende
verlening werkingsduur Wet Premie Kerkenhouw. - Behandeling en verwerving
```

Flat text turned into richly annotated XML

Behandeling en aanneming van de wetsontwerpen Wijziging van de artikelen 1637\$ en 1638r van het Burgerlijk Wetboek. Nadere regels met betrekking tot het afleggen van eed of belofte (Eedswet 1969). - Behandeling van het voorstel van wet van de heren Westerterp, Boersma, Van Mierlo, Van Thijn, Wiegel en De Boo tot het in overweging nemen van een voorstel tot verandering in artikel 110 van de Grondwet (Vervroeging Prinsjesdag). Regeling van de werkzaamheden. - Behandeling en aanneming van het ontwerp van rijkswet Goedkeuring van het op 31 maart 1953 te New York gesloten Verdrag betreffende de politieke rechten van de vrouw (<tagged pm:type="reference">

<tagged-entity pm:type="reference" pm:sub-type="dossierref"
pm:reference="19701971:0000647:K1968"/>Trb. 1968, 92</tagged>). - Indiening en behandeling van een motie van mevrouw Padt-Jansen c.s., betreffende de wenselijkheid van wijziging van artikel 11 van de Grondwet in die zin dat bij erfpvolging vrouwen en mannen

About: http://nl.dbpedia.org/resource/Frans-Jozef_van_Thiel

An Entity of Type : [persoon](#), from Named Graph : <http://nl.dbpedia.org>, within Data Space : [nl.dbpedia.org](#)

Property

Value

- Frans Joseph Frits Maria (Frans-Jozef) van Thiel was in de zestiger jaren een bekende en populaire voorzitter van de Tweede Kamer. Hij zat onder meer voor tijdens de Nacht van Schmelzer en de debatten over de drie van Breda. De KVP-voorman vervulde dit ambt met veel Brabantse humor, maar moderniseerde ook de werkwijze. Hetgeen gebeurde in de Tweede Kamer vanaf 1965. Na de val van de KVP in 1972 volgde hij Louis Beel als voorzitter van de Tweede Kamer op. Hij stopte in 1972.

Linked data in DBpedia about MPs (full bio's)

name metaalindustrie en textielindustrie. Zijn vader heette Frans van Thiel en zijn moeder Henriette Hoppenbrouwers. Hij trouwde op 17 augustus 1931 met Carla Beatrice Maria Rudolphina Hoijng in De Bilt. Hij bleef, ondanks zijn drukke politieke carrière in Helmond wonen. Zijn weduwe overleed 16 jaar na haar man op 101-jarige leeftijd te Helmond. Het echtpaar Van Thiel-Hoijng had 4 zonen en 2 dochters, van wie 1 dochter vóór haar moeder overleed. Zoals in vele vooraanstaande families zijn de kinderen, klein- en achterkleinkinderen uitgezworven over de wereld. Vier van de nog levende kinderen wonen evenwel nog in hun geboorteregio.

[dbpedia-owl:birthDate](#)

- 1906-12-19 (xsd:date)

[dbpedia-owl:country](#)

- [dbpedia-nl:Nederland](#)

[dbpedia-owl:deathDate](#)

- 1993-06-02 (xsd:date)

[dbpedia-owl:longName](#)

- Frans Joseph Frits Maria (Frans-Jozef) van Thiel

[dbpedia-owl:predecessor](#)

- [dbpedia-nl:Louis_Beel](#)
- [dbpedia-nl:Rad_Kortenhorst](#)

[dbpedia-owl:relatedFunctions](#)

- [dbpedia-nl:Lijst_van_voorzitters_van_de_Tweede_Kamer](#)
- [dbpedia-nl:Lijst_van_Nederlandse_ministers_van_Volksgezondheid,_Welzijn_en_Sport](#)

[dbpedia-owl:subtitle](#)

- Afscheid van kamervoorzitter F.J.F.M. van Thiel. Den Haag, Nederland, 6 december 1972.

[dbpedia-owl:successor](#)

- [dbpedia-nl:Marga_Klompé](#)
- [dbpedia-nl:Anne_Vondeling](#)

[dbpedia-owl:wikiPageExternalLink](#)

- <http://www.parlement.com/9291000/biof/01355>
- <http://www.inghist.nl/Onderzoek/Projecten/BWN/lemmata/bwn6/thiel>

[dbpedia-owl:wikiPageID](#)

- 232517 (xsd:integer)

[dbpedia-owl:wikiPageRevisionID](#)

- 38793727 (xsd:integer)

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:dbpediaOntology="http://dbpedia.org/ontology/" xmlns:mime="http://
purl.org/NET/mediatypes/" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:dc=
"http://purl.org/dc/elements/1.1/" xmlns:rdfs="http://www.w3.org/2000/01/rdf-
schema#" xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:rdf="http://www.w3.org/
1999/02/22-rdf-syntax-ns#" xmlns:parlipro="http://purl.org/vocab/parlipro#"
xmlns:pm="http://www.politicalmashup.nl" xmlns:bio="http://purl.org/vocab/bio/
0.1/" xmlns:owl="http://www.w3.org/2002/07/owl#" xml:base="http://
resolver.politicalmashup.nl/">
<rdf:Description rdf:about="nl.proc.sgd.d.197019710000647">
    <rdf:type>
```

200 years of hansards in RDF (each speech/action)

```
</rdf:type>
<dc:identifier>
  <rdf:Description rdf:about="nl.proc.sgd.d.197019710000647"/>
</dc:identifier>
<dc:format>
  <rdf:Description rdf:about="http://purl.org/NET/mediatypes/application/
    rdf+xml"/>
</dc:format>
<dc:contributor>
  <rdf:Description rdf:about="http://www.politicalmashup.nl"/>
</dc:contributor>
<dc:date rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1971-02-09</
  dc:date>
<dc:source>
  <rdf:Description rdf:about="http://resolver.politicalmashup.nl/
    nl.proc.sgd.d.197019710000647?view=xml"/>
</dc:source>
<parlipro:legislativePeriod>1970-1971</parlipro:legislativePeriod>
<dcterms:ISO3166-1>NL</dcterms:ISO3166-1>
```

“Knettergek” in Parliament

Dutch to English: [more detail...](#)

1. knettergek: 

- potty  ; stark staring mad  ; crazy  ; stark raving mad  ; cracked  ; loony  ; bonkers  ; round the bend  ; nuts  ; crackers 



AD

ALGEMEEN DAGBLAD

AFELAY

'Ouders staken zelf huis
met kinderen in brand'

NRC: 'rock bottom' in use of language in parliament

**NEDERLAND 2
SLOWAKIJE 0**

Ibrahim Afellay maakte gis-

Verguisde premier
'Eenzame staatsman'
De Geer was dupe van
roddel en ...

Terugkeerhun-
Univer-

Can he?

ukbeweging

'Abvakabo
ten gaan
er als... PNV

Leze
'Leve
vijft
dan
PNV

Hans Wijers (Natuurmonumenten): 'Ik kan me niet opwinden over de korenwolf p3

**DONDERDAG
31 MEI 2012
VK.NL**

RIDLEY SCOTT MAAKT WEER SCIENCEFICTION
**Terug naar het
succes van 'Alien'**
FILM V2-5

ANALYSE EUROPESE AANBEVELINGEN

Oranje nog een
wankele ploeg
SPORT P32-33

'ORT 'SOEVEREIN HEK' OMHEEN STOND

Telegraaf: Chair says "limit is reached"

Liefst 28 lijstjes met indringende financieel-economische achtergronden presenteerde de Europees Commissie woensdag. Een voor elke EU-lidstaat een daadbevereiging nog een voor die helle economie, alles bij elkaar beginnend verschillend pagina's. Vele hadden zichzelf van de crisis weer een dergelijke herinnering van Brussels mocht niet meer aanbelangden.

hrc

How (un)common is “knettergek” in parliament?



De wegwijzer naar informatie en diensten van alle overheden

 Overheid.nl

nationaal archief

KB

No online proceedings

Only recent (1995-) ones

Everything (1579-) on paper

Digitized (1815–1995) online

Dutch Parliamentary Records

- House of representation: https://www.tweedekamer.nl/kamerstukken/plenaire_verslagen
- National archives: <http://www.gahetna.nl/collectie/archief/ead/index/zoekterm/staten-generaal/aantal/20/eadid/2.02.21.01>
- Government: https://zoek.officielebekendmakingen.nl/zoeken/parlementaire_documenten
- National library: <http://www.statengeneraaldigitaal.nl/>

Elastic Search

- Data:
 - Netherlands 1814-
 - UK 1803-
 - Canada 1901-
 - ...
- Data Format: <http://schema.politicalmashup.nl/>
- Search Engine: <http://search.politicalmashup.nl/>

Parliamentary data: semantic annotation of speeches, full bio of speakers

Parliament Debate Search

About

Add a party

Add a speaker

Rotterdam x Search for...

Search!

Collection Netherlands ▾

Granularity Speech ▾

Advanced search ▼

Date range

Exclude speakers

Exclude parties

Debug

From 01/01/1800 To 01/01/2015

Word clouds ▼

Word cloud

Category word cloud

Significant terms per party

Top terms per party

Significant terms per speaker

Top terms per speaker

Kortenhorst

gerestitueerd

Fock

wijkerslooth

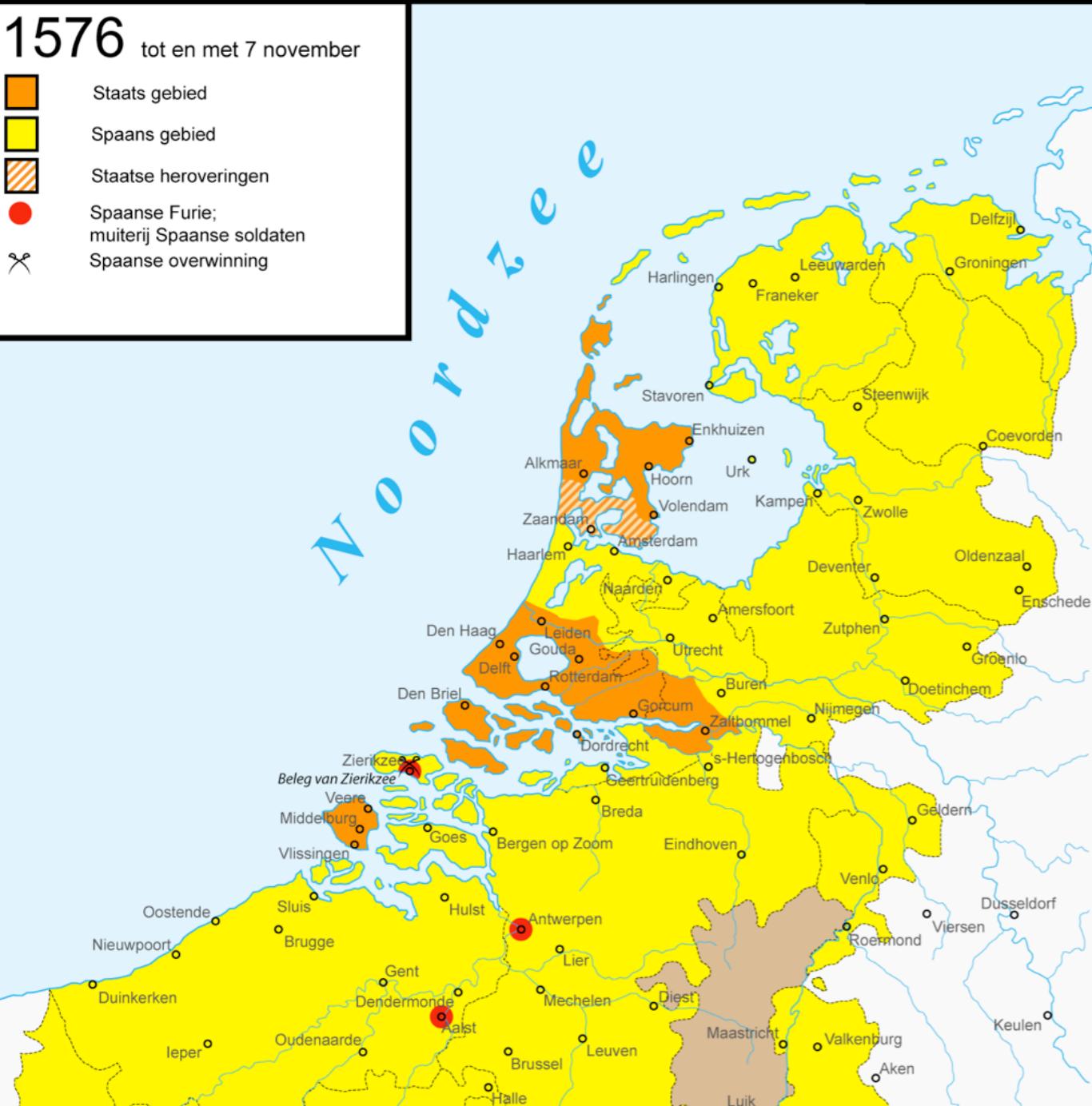
Goeman Borgesius

helsdingen

1576

tot en met 7 november

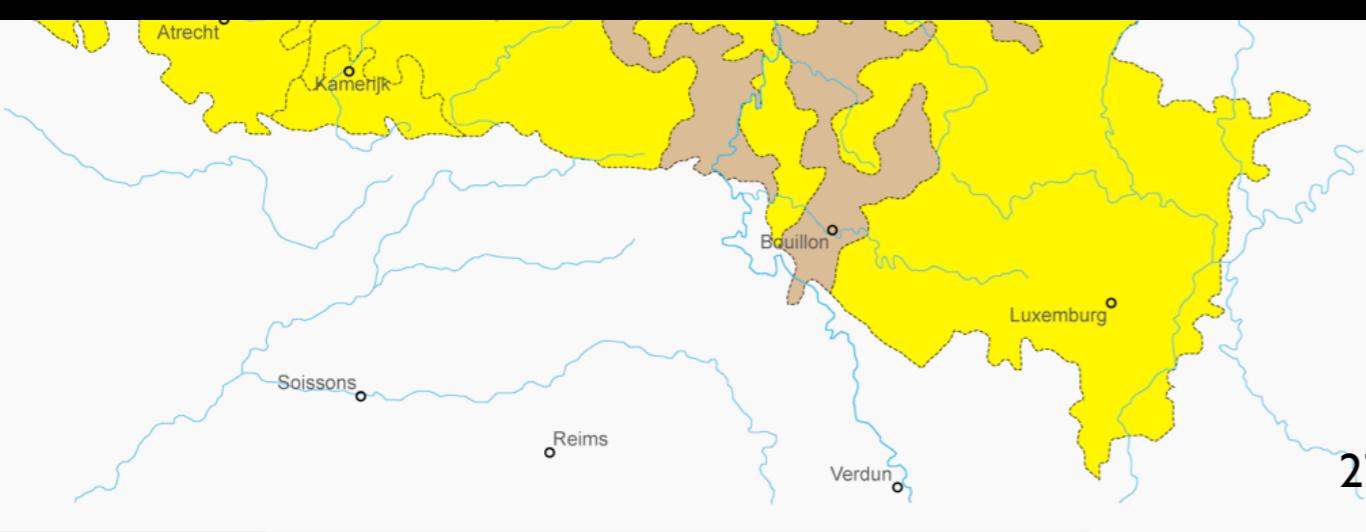
- █ Staats gebied
- █ Spaans gebied
- █ Staatse heroveringen
- Spaanse Furie; muiterij Spaanse soldaten
- x Spaanse overwinning



- 2 Cijferige postcodegrens
- Landsgrafs
- Snelweg
- 2 Cijferige postcode
- 1 Cijferige postcode
- Hoofdstad land
- Hoofdstad Provincie
- Provincie
- Nederland
- België (Vlaanderen)
- België (Wallonië)
- Luxemburg



Proceedings are also a 400+ year longitudinal corpus



Add a speaker

Query

crisis

Search!

Collection

Netherlands ▾

Granularity

Speech ▾

Advanced search >

Impact of people, events, regions, issues, language evolution, ...

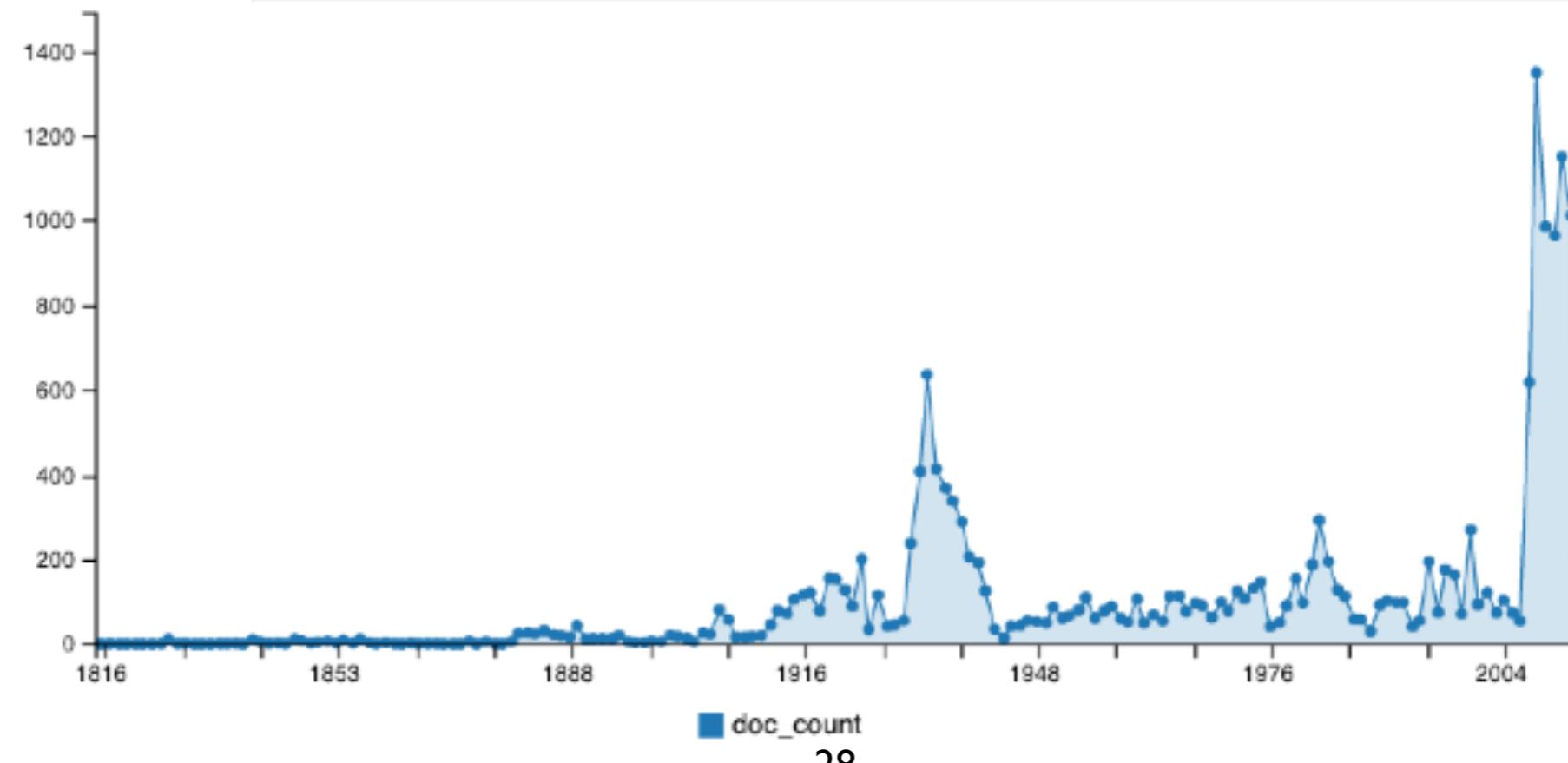
Word clouds

Graphs

Hits over time

Hits over time per party

Hits over time per speaker



LET'S MAKE
THINGS GENERAL

#1 WHAT IS (NOT) TIME

TIME IS A FUNDAMENTAL DIMENSION?

- ▶ Information visualisation clusters into
 - ▶ time
 - ▶ space
 - ▶ topic
- ▶ Also facet classification in libraries (eg., PMEST)

Allard Pierson Museum

[Home](#) [Visitor information](#) [Exhibitions & events](#) [Collections & research](#) [Press](#)

Cabinet of Collectors

Step into the world of some of the most important collectors who helped build the museums collections.



[Het Verzamelaarskabinet](#)

CROSSROADS

ArcheoHotspot

Events

[Events calendar](#)

Cabinet of Collectors

PROJECTS WITH MUSEUM OF ARCHEOLOGY

- ▶ Series of EU and other project with the APM
 - ▶ <http://www.allardpiersonmuseum.nl/>
 - ▶ <https://www.uvaerfgoed.nl/beeldbank/en/allardpiersonmuseum>
- ▶ Loads of cool stuff but no timelines
- ▶ Why not?
 - ▶ Only possible for metadata
 - ▶ Data = noisy

DIMENSIONS?

- ▶ Three main dimensions: time, space, topic
- ▶ All three essentially collapse in archeological domain
 - ▶ e.g. Greek department = topic + region + time period
 - ▶ reality more complex (e.g. Greek object from Egypt, exact dating).
- ▶ Similar art museums (e.g., Bauhaus, Stijl, ..) but precise dating is available

WHAT IS (NOT) TIME

- ▶ Time is complexly linked to space/topic/...
- ▶ This is what makes time relevant!
 - ▶ If you have an easy time entry, this relation to space/topic/... is of clear interest
 - ▶ But often you need to go the other way around, e.g., use events or persons or topics to “capture time”

#2 THE RIGHT
TOOL FOR THE JOB

BEYOND PLOTS OVER TIME

- ▶ Temporal plots are cool
 - ▶ In particular when you can slice and dice and see impacts
 - ▶ In particular when connected to the source data
- ▶ But they are only the obvious, first, low hanging fruit
 - ▶ Only scratching the surface
 - ▶ Many, many other use cases (more interesting!)

Cast Your Net Wide: Finding Historical References in Parliamentary Data

Alex Olieman

alex@olieman.net

University of Amsterdam, the Netherlands

Kaspar Beelen

kasparvonbeelen@gmail.com

University of Amsterdam, the Netherlands

Introduction

Analyzing how the past is molded to fit contemporary needs and priorities represents an intensely researched topic among historians, who have scrutinized (among others) the performance of “commemorative practices,” the transmission of “collective memories,” or, in general, “the language of the past.” However, scholars currently lack adequate digital tools to assist them in this endeavor (Likaka, 2009; Piersma and Ribbens, 2013; Wilson, 2016).

The information needs of historians and other humanities scholars are not always adequately captured by traditional full-text search because query terms often serve as a proxy for more abstract concepts, while both items (the word and the concept) sometimes hardly overlap. This technology demonstration focuses on a specific type of concept, namely historical eras or periods.

Because of their complex and heterogeneous character, historical periods are difficult to capture with simple keywords. For example, the historian looking for references to the Renaissance in diachronic corpora will often encounter documents in which the

we introduce “WideNet,” a tool that allows historians and other scholars within the humanities to search for information about specific periods in a multilingual corpus containing the Parliamentary debates from the United Kingdom, Canada, and the Netherlands.

WideNet models historical periods as a container of (hierarchically) related entities. The art historian, interested in how British MPs refer to the Italian Renaissance will, instead of retrieving all occurrences of the term “Renaissance,” be provided with speeches that mention, e.g., composers such as Costanzo Porta, and famous paintings such as The Bacchanal of the Andrians. Whereas earlier work proposed to use such entities as search suggestions (Piersma et al., 2014), we rather prefer a “sculptor’s approach” in which many containers of potentially relevant entities are initially included in the query, and may be deselected based on their empirical relevance.

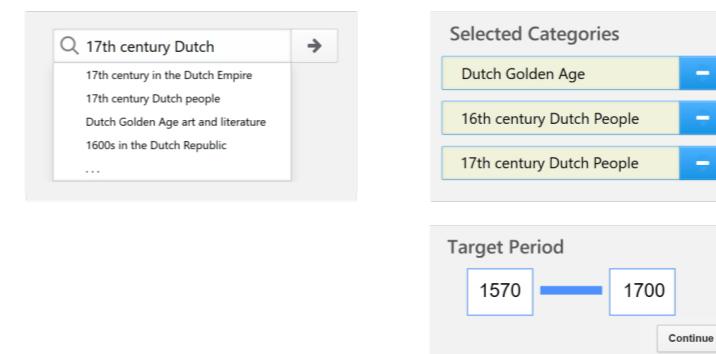


Figure 1: Initial query specification in WideNet

These categories of entities are provided by a knowledge base (KB) that is loaded into the tool in advance. Our demo tool makes use of DBpedia, but any KB that conforms to the SKOS ontology may be loaded. The scholar, using WideNet, starts by selecting one or several root categories from a type-ahead search box (see Figure 1), and can further demarcate the query by selecting a time period, which will be used to prune the underlying entities of the selected categories. WideNet subsequently retrieves the network of narrower categories for each selected root category, and collects the contained entities as potentially relevant query components.

#3 THE RIGHT JOB FOR THE TOOL

ALL CURRENT SEMANTIC ANNOTATION TOOLS ARE RATHER CRAPPY

- ▶ Enriching text (or other modalities) is pretty cool
 - ▶ Example on the slide looks wow
 - ▶ Reality is often rather messy
 - ▶ Fundamental limitations and biases
 - ▶ Many practical reasons for failure
 - ▶ Never ending route to “improve” tools
 - ▶ Rather than improving tools, looks for tasks where current tools are useful

Good Applications for Crummy Entity Linkers? The Case of Corpus Selection in Digital Humanities

Alex Olieman

University of Amsterdam

olieman@uva.nl

Stamkracht BV

alex@stamkracht.com

Kaspar Beelen

University of Amsterdam

k.beelen@uva.nl

Milan van Lange

NIOD Institute for War, Holocaust,
and Genocide Studies

m.van.lange@niod.knaw.nl

Jaap Kamps

University of Amsterdam

kamps@uva.nl

Maarten Marx

University of Amsterdam

maartenmarx@uva.nl

ABSTRACT

Over the last decade we have made great progress in entity linking (EL) systems, but performance may vary depending on the context and, arguably, there are even principled limitations preventing a “perfect” EL system. This also suggests that there may be applications for which current “imperfect” EL is already very useful, and makes finding the “right” application as important as building the “right” EL system. We investigate the Digital Humanities use case, where scholars spend a considerable amount of time selecting relevant source texts. We developed *WideNet*; a semantically-enhanced search tool which leverages the strengths of (imperfect) EL without getting in the way of its expert users. We evaluate this tool in two historical case-studies aiming to collect a set of references to historical periods in parliamentary debates from the last two decades; the first targeted the Dutch Golden Age, and the second World War II. The case-studies conclude with a critical reflection on the utility of *WideNet* for this kind of research, after which we outline how

This limitation of current EL benchmarks, we argue, limits the usefulness of performance statistics for application developers who would like to include EL in their software. From their perspective, the current evaluation paradigm is unsatisfactory because it does not provide clear expectations about how the EL functionality will be perceived by their intended users. The current situation is reminiscent of the discussion of Machine Translation evaluation in the early 1990s, which recognized the importance of identifying good niche applications for state-of-the-art technology, which was arguably crummy at the time [2].

In this paper, we briefly review current-day EL evaluation and discuss the parallels with early work on Machine Translation. We take the position that in order to demonstrate the value of EL for real-world users, a complementary evaluation paradigm based on end-user applications is needed. One particularly high-payoff niche application is found in the area of Digital Humanities, where scholars spend a considerable amount of time selecting relevant

#4 METADATA IS DEAD, LONG LIVE DATA

DATA, METADATA, OR BOTH?

- ▶ Origins of metadata
 - ▶ objects were physical, need for access
 - ▶ description standards emerged to create surrogates for indirect retrieval
- ▶ Historically, we digitised descriptions (metadata) first
 - ▶ and now digitised the objects (data)
- ▶ Revisit the role of metadata
 - ▶ Why needed?
 - ▶ Access role diminishes (even gone?)
 - ▶ Does it increase the “use value”?

DATA, METADATA, OR BOTH?

- ▶ Phases:
 - ▶ data
 - ▶ data + metadata
 - ▶ data + dig. metadata
 - ▶ dig. data + dig. metadata
 - ▶ ...
- ▶ What's next?

FROM ACCESS TO USE

- ▶ Interesting paradox
- ▶ Naive DL view:
 - ▶ dig. data is appendix of dig. metadata
- ▶ Naive User view:
 - ▶ dig. metadata is appendix of dig. data
- ▶ We need to move to a view of: enriched data
 - ▶ what is useful on top of the source data?
 - ▶ useful for use/research (not as access point)

WRAP UP: TAKE TIME SERIOUS!

- ▶ Range of examples of the importance and use of time in a collection covering many years
- ▶ Move from metadata to data, and fruitful combinations
 - ▶ Move from “access” to “use”
 - ▶ Lightweight reliable semantic annotations work!
- ▶ Move from improving tools to improving tasks
 - ▶ Move from “access” to “use”
 - ▶ Availability of “workable” tools, and need to find use cases where imperfect tools “work”

SOME STUFF MENTIONED (SELECTION)

- ▶ Political Mashup project, <http://politicalmashup.nl/>
- ▶ Exploratory Political Search (ExPoSe) project, <http://e.humanities.uva.nl/expose/>
- ▶ Digging into Linked Parliamentary Data (DiLiPad) project, <https://dilipad.history.ac.uk/>
- ▶ Cast Your Net Wide: Finding Historical References in Parliamentary Data (DH'17) <https://dh2017.adho.org/abstracts/579/579.pdf>
- ▶ Good Applications for Crummy Entity Linkers? The Case of Corpus Selection in Digital Humanities (Semantics'17) <https://arxiv.org/abs/1708.01162>