Study Supreme Court Decision Making with Linked Data

Xian Li
Rensselaer Polytechnic Institute
110 8th Street, Troy
New York 12180
USA
lix15@rpi.edu

Li Ding
Rensselaer Polytechnic Institute
110 8th Street, Troy
New York 12180
USA
dingl@cs.rpi.edu

James A. Hendler
Rensselaer Polytechnic Institute
110 8th Street, Troy
New York 12180
USA
hendler@cs.rpi.edu

ABSTRACT
Current social science researchers depend heavily on data to perform analytical studies. The multitude of problems studied by researchers today is so pervasive that it is difficult for scholars to collect needed information independently, format it in an easily processable way, check the accuracy of that information, and publish and maintain it in a reusable form. Moreover, as the complexity of analytic problems grows, data from different sources are needed to provide multiple perspectives to gain insights into the problems being studied.

In this paper, we show how linked data and other Semantic Web technologies can enhance the study of U.S. Supreme Court justices' decision-making. We first describe challenges in current Supreme Court judicial research. Then the linked data technology is introduced, followed by how it addresses those challenges. Last, by illustrating our method with several examples, we demonstrate the advantages of this approach and identify future work.

Keywords
Linked Data, Semantic Web, Web Science, Supreme Court, Decision Making.

1 INTRODUCTION
The multitude of problems existing in current society is so pervasive that researchers are increasingly putting requirements on the amount and quality of data they need to perform analyses from multiple perspectives. A "mashup" of data from different sources, in different areas, can lead to insights on certain specific problems. For example, data about future contracts on orange juice can be used to predict weather conditions in Florida [9]; individual investors' investment returns can be analyzed with data about their residential regions [5]. However, it is difficult for scholars to independently identify relevant data resources, integrate all needed information, format them in an easily processable way, check the accuracy of the information, and publish and maintain them in a reusable state.

As the amount of data keeps growing, and more and more data become openly accessible on the Web, structured data integrated with the Web can help tackle this difficulty. The decentralized nature of the Web allows data from heterogeneous sources to be combined together to solve a specific problem from different dimensions. Semantic Web technologies such as linked data [1] enable meanings to be associated with data so that relevant data can be identified and linked together automatically.

This paper investigates U.S. Supreme Court justices' decision-making as a representative study of social science with data, and shows how linked data and other Semantic Web technologies can be used to enhance current judicial research.

The organization of the paper is: in Section 2, we review related works about U.S. Supreme Court justices' decision making and summarize existing challenges. In Section 3, we analyze the advantages of linked data over existing methods, and then propose approaches describing how it can be combined with other Semantic Web technologies to address these challenges. The results are summarized and verified in Section 4 by providing three examples that illustrate enhancements over existing work using a prototype application we have developed. We conclude in Section 5 with a discussion of future work.

2 SUPREME COURT DECISION MAKING

2.1 Related Work
Studying Supreme Court justices as policy-minded decision makers has existed for many years [7][8][10][11], utilizing large amounts of data encoding dockets, cases and justices' votes to analyze judicial decisions. Therefore, the core data in use has been curated and maintained as reusable resources for other researchers. The U.S. Supreme Court Database (SCDB)[1], built primarily by Harold J. Spaeth, has encoded a variety of aspects of justices' votes since 1953 [15]. It is regarded as a core dataset for studying the Supreme Court by providing basic details from both the case and vote levels.

Based on data from SCDB, various analyses about judicial decision-making were done, such as precedential/preferential behavior [12], proportion of liberalism voting in civil rights cases [12], and changes of political preference over years [4].

Besides the SCDB datasets, studies on justices' decision-making behavior have relied on variables independent from judicial votes. For instance, a personal attribute theory of liberalism in the U.S. Supreme Court was presented including political, social cleavage, family origins, and career socialization variables [17] [18]. These variables were a compilation of justices' personal attributes manually coded from readily available biographical directories (e.g. Who's Who in America) and data collections of other scholars in the field.

The influence of public opinion on the Supreme Court's rulings has also been studied. In these cases, a purposefully and carefully prepared newspaper content analysis [13][14] had been the primary approach used by judicial researchers to analyze the ideological values of the justices. The measures of these variables

Web Science Conf. 2010, April 26-27, 2010, Raleigh, NC, USA.

[1] http://scdb.wustl.edu

were to encode newspaper editors' assessments of justices' ideological values gathered from representative newspapers with liberal stances (e.g. the *New York Times*) and conservative ones (e.g. the *Los Angeles Times*).

2.2 Challenges

Challenges associated with the existing judicial research described above could be summarized as follows:

First, as to the widely used SCDB datasets, it was the judicial scholars themselves who collected and coded the data, performed reliability checks and published the data sets. Therefore, although it is periodically released in formats accepted by most statistical applications such as CSV, DTA, POR, SAV, and XLSX, it could not be automatically connected to other data sources. Moreover, focusing on statistical aspects of these isolated data may result in a limited set of views and analytical directions, with missed opportunities to gain different perspectives and insights on existing data.

Second, while analyzing personal attributes models, due to the large amounts of data involved, the manual encoding method is time consuming and its limited visibility minimizes the chances for others to correct mistakes. The independence of each data source may also have limited insight into aggregate relationships that could be investigated.

Third, the newspaper content analysis method is said to be susceptible to changes in the social, political and economic environments that generate the comments in newspapers [13]. Therefore, additional data about the macro environments would be helpful.

3 LINKED DATA AS A SOLUTION

Linked data technology can provide significant support in addressing these challenges. First of all, the Linked Open Data (LOD) cloud [2] (such as *DBpedia* [3] and *Freebase* [4]) covers significant factual information about justices' personal attributes as well as career histories, and is in a ready-to-use state. More importantly, because these encoding are data-oriented, all the available properties can be automatically generated given the identification of a justice. Furthermore, since linked data is contributed and maintained in a social community, the quality of the information can potentially be superior to that collected by individual scholars in terms of bias and errors.

3.1 From Structured Data to RDF

Since SCDB is the core source, to make these data interact with other linked data, we converted the CSV datasets to RDF/XML encoding following the methodology proposed by [3]. The advantage of this solution is the ease of implementation without losing the "meaning" of the data. For example, in the justice-centric raw CSV files each row represents a vote along with justice information and case information encoded in the columns. During conversion, each non-header table row (entry) is assigned a URI, and the column names (on the header row) are mapped into RDF properties with the namespace tied to the dataset.

As a result, the converted SCDB data can be made Web accessible, and easily consumed by parsers for RDF and XML. Further filtering on the datasets is easily done by SPARQL queries or XQuery.

3.2 Make Data Linkable

After transforming the SCDB data to linkable data, and linking it with the Linked Open Data cloud, the variables for judicial research mentioned in Section 2 can be found as linked data, which enlarge data from current research, as shown in Table 1. To achieve this linkage we established connections via justice names. In the SCDB data, justice names are some abbreviations only used within SCDB, such as "HLBlack" for Justice Hugo Black. As such abbreviation is not recognizable among other linked data, we use an instantiation of the Semantic MediaWiki[5] to solve this problem from several aspects, as we describe below.

Table 1. Linked data of U.S. Supreme Court justices

Justice Attributes	Exist in Current Research	Linked Data Source
Justice name	Yes	Data-gov Wiki
Votes from SCDB	Yes	TWC triple store
Birth, Upbring, Education	Yes	DBpedia, Freebase
Age and Tenture	Yes	DBpedia, Freebase
Ethnicity	No	Freebase
Gender	No	Freebase
Career Characteristics	Yes	DBpedia, Freebase
Partisanship	Yes	DBpedia, Freebase
Families	No	DBpedia
Related Cases	No	DBpedia
Successor and Predecessor	No	DBpedia, Freebase
Courts dissent with	No	DBpedia
Participated inauguration of president	No	DBpedia
Quotes	No	Freebase
Awards	No	Freebase
Newspaper Comments	Hardcopy newspaper	New York Times linked data

3.2.1 Dereferenceable Properties for Each Justice

We use the Semantic MediaWiki to enable collaboratively adding resources linking justices with properties from well-known ontologies. For example, we used *dcterms:identifier* to capture the abbreviated name in SCDB, *owl:sameAs* to indicate the DBpedia resource link for justice, and *foaf:name* to refer to justice's full name which appears in DBpedia, as shown in Figure 1[6]. Thus, justices' voting data retrieved from converted SCDB RDF triples, is connected with open linked data such as DBpedia, and they can be queried together.

Extensibility and reusability are the main advantages of this solution. We fully respect the decentralized and social nature of current Web. Any further identifiers could be easily added by

[2] http://linkeddata.org

[3] http://dbpedia.org

[4] http://www.freebase.com

[5] http://data-gov.tw.rpi.edu

[6] http://data-gov.tw.rpi.edu/wiki/Hugo_Black

others through skos:altLabel property. Once mapping resources are created, they can be refined over time, while still remaining readily consumable and reusable by other applications.

Figure 1. Dereferenceable properties for justice

3.2.2 *Dynamic RDF Graphs for All Justices*

As described in Section 2, U.S. Supreme Court justices have been systematically studied across different periods. We use the Semantic MediaWiki to capture these categorizations and produce an instance of such knowledge as linked data. Figure 2[7] shows the automatically generated wiki page for U.S. Supreme Court justices, and the corresponding dynamically generated RDF graphs are shown in Figure 3[8].

Figure 2. Categorization of Supreme Court justices

In summary, both of the above approaches use the Semantic MediaWiki to provide extensible and reusable ways to implement linkages between existing SCDB datasets and the more open linked data on the Web.

[7] http://data-gov.tw.rpi.edu/wiki/Supreme_Court_Justices_of_United_States

[8] http://data-gov.tw.rpi.edu/wikidata/Supreme_Court_Justices_of_United_States

Figure 3. Dynamic RDF graph for all justice

3.3 Linking to the Linked Data Cloud

3.3.1 *Linking with DBpedia for Personal Attributes*

Based on the approaches described above, we can generate SPARQL queries against the DBpedia endpoint to retrieve justices' personal attributes to be used for analysis about their decision-making. An example query illustrating how we connect the SCDB data with DBpedia is shown in Figure 4.

```
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX scdb:    <http://data-gov.tw.rpi.edu/vocab/p/10016/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT ?name count(?o) ?dbpediauri
WHERE {
  GRAPH <http://data-gov.tw.rpi.edu/vocab/Dataset_SCDB> {
  ?o scdb:justicename ?jname1.
  ?o scdb:direction "2"^^<http://www.w3.org/2001/XMLSchema#integer>.
  }
  GROUP BY ?jname1.
  GRAPH <http://data-gov.tw.rpi.edu/wikidata/Supreme_Court_Justices_of_United_States> {
  ?s dcterms:itentifier ?jname2.
  ?s owl:sameAs ?dbpediauri.
  ?s foaf:name ?name.
  }
  FILTER (?jname1 = ?jname2)
}
```

Figure 4. Query for DBpedia URI

With the DBpedia URI for each justice generated as above, we can go on to query a variety of personal attributes that are used in existing research, including *dbpprop:almaMater* for the justice's education, *dbpprop:nominator* for the president who nominated the justice, *dbpprop:birthPlace*, *dbpprop:dateOfBirth*, *dbpprop:termstart*, and so on. For example, Figure 5 shows querying for justice's nominator party and date of birth, given the DBpedia URI *<http://dbpedia.org/resource/Ruth_Bader_Ginsburg>* returned by the query above.

```
PREFIX dbprop: <http://dbpedia.org/property/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name ?birthdate ?partyname
WHERE {
  <http://dbpedia.org/resource/Ruth_Bader_Ginsburg> foaf:name ?name.
  OPTIONAL{
    <http://dbpedia.org/resource/Ruth_Bader_Ginsburg>  dbprop:nominator ?nominator.
    ?nominator <http://dbpedia.org/ontology/party> ?party.
    ?party  dbprop:partyName ?partyname.
  }
  OPTIONAL{
    <http://dbpedia.org/resource/Ruth_Bader_Ginsburg>  dbprop:dateOfBirth ?birthdate.
  }
}
```

Figure 5. Query DBpedia for personal attributes

3.3.2 *Linking with Other Open Data*

With the linked data technology, scholars also have a simplified means of access to all available linked government data [2] containing facts about various aspects of society relevant to a given study. With an abundance of data resources, scholars can gain more insight into justices' decisions-making from multiple perspectives, for example, budgets for the court, annual GDP, national unemployment rate and so on.

Media coverage has also been an indispensable source of public opinions towards decisions from the Supreme Court justices. The New York Times has recently provided linked data about People, Organization and Location[9], and each justice can be found in their People data ready for SPARQL query.

4 RESULT ANALYSIS

4.1 Summary

As a result of the use of linked data, the information that is automatically made available to judicial researchers is greatly enlarged and diversified, as shown schematically in Figure 6.

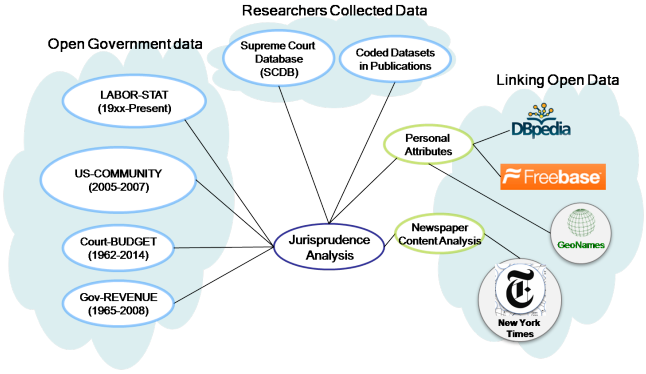

Figure 6. Linked data for studying Supreme Court justices

So far, the three primary challenges described in Section 2 are addressed with the proposed approaches.

First, by converting the core datasets SCDB into RDF data, it is possible to connect them with linked data from other areas with different perspectives while not losing the statistical features of the data.

Second, the Semantic MediaWiki is used to bridge SCDB data and the Linked Open Data cloud, so that those essential variables used by personal attribute models can be automatically retrieved instead of requiring the traditional manual encoding (and as the information in LOD is live linked to dynamic sources, the currency of the data can be maintained). The quality of these variables is improved due to the collaborative work from numerous communities of interest which help maintain the various sources.

Third, with the linked data released by the New York Times, newspaper content analysis can be integrated with other research models based on linked data. The macro environment impacting the newspaper content analysis can potentially be analyzed by linking it with open government data, (We are currently exploring this approach).

On the other hand, there are shortcomings and challenges that are introduced. First, data from the Linked Open Data cloud may be incomplete. For example, the education variables for several justices are missing in DBpedia and Freebase. Second, since data from heterogeneous sources are integrated to present analytical results, it is possible that there are contradictions among them or disagreements with users' prior knowledge, which could threaten the trustworthiness of the analyses developed.

4.2 Application

A prototype application[10] has been developed to demonstrate the approaches described in this paper. The data underpinning the application come from the converted SCDB RDF data, DBpedia and the Semantic MediaWiki that establishes connections between them as described above. Combined with the visualization technologies of Exhibit[11] and Google Motion Chart[12], this application can be used to perform analysis over Supreme Court justices including the replication of existing research for validation. Here are three examples in response to the main challenges described in Section 2:

- *Dynamic and systematic study of justices' voting behaviors*

  Existing research [4] based on SCDB data can be replicated by this application. Moreover, by visualizing all voting data, existing research is extended by incorporating data about more justices and dynamic views of the evolution of their values over years can be presented as shown in Figure 7.

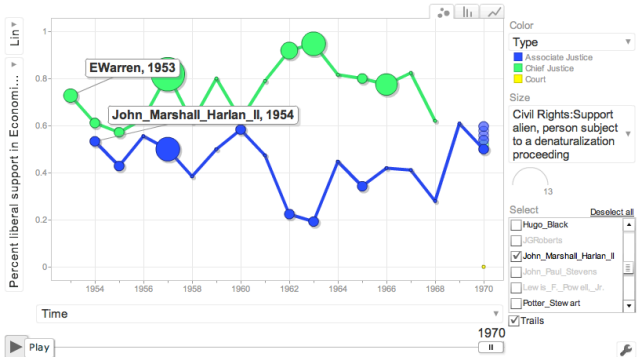

Figure 7. Trend of Supreme Court justices' liberalism

- *Personal attributes study*

  The application displays personal attribute variables which are queried from DBpedia, using features such as facet browsing, search, filtering, and Google maps, as shown in Figure 8. Compared with the mere numbers displayed in existing research, this visualization over linked data provides more flexible visualization over those parameters, allowing the exploration of the interactions in numerous ways. An example about voting directions and nominated political parties is shown in Figure 9.

- *Insights from newspaper and open government data*

  Newspaper comments from the New York Times are associated with each justice, and a macro analysis with open government data is provided, as shown in Figure 10.

[9] http://data.nytimes.com/

[10] http://data-gov.tw.rpi.edu/wiki/Demo:_Supreme_Court_Justices_Decision_Making

[11] http://www.simile-widgets.org/exhibit/

[12] http://code.google.com/apis/visualization/documentation/gallery/motionchart.html

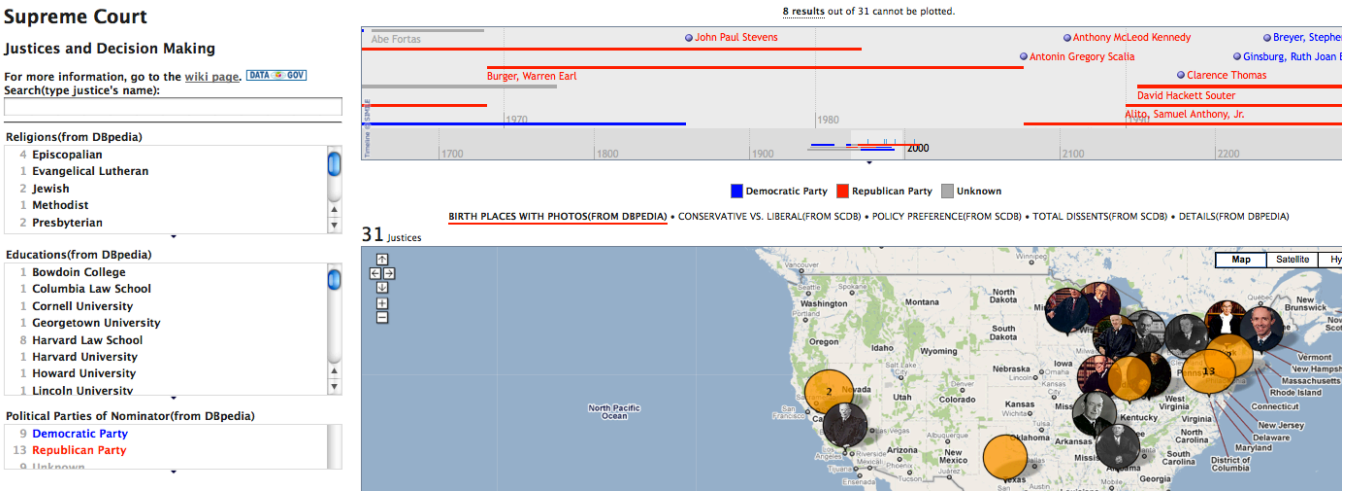

Figure 8 Personal attributes data from DBpedia

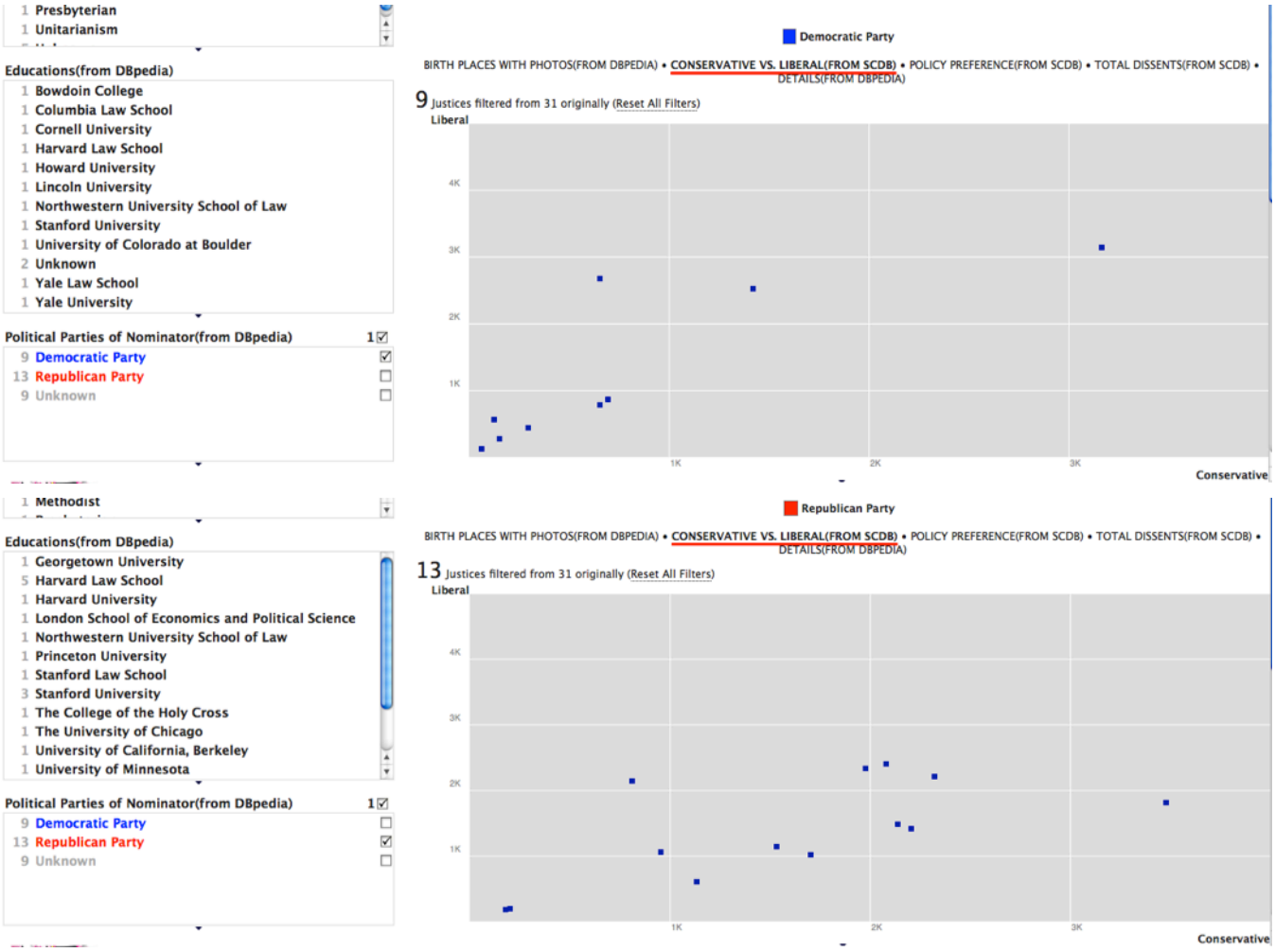

Figure 9 Voting directions and political parties

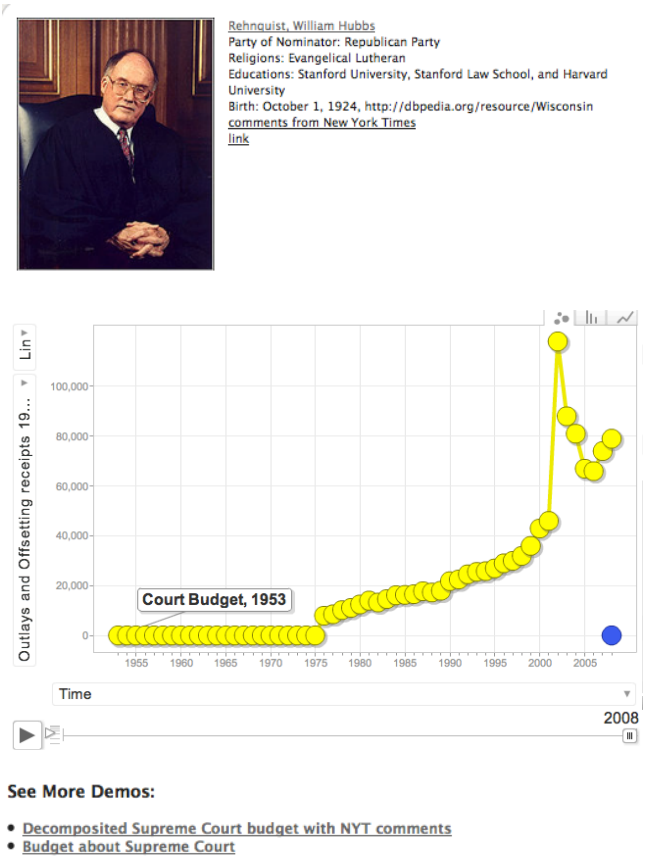

Figure 10 Insights from newspaper comments and government data

5 CONCLUSION

This work shows how semantic technology can make it possible to leverage large and disparate data sets for social science researchers and legal scholars by addressing some of their primary challenges. We show approaches to transforming existing core data to linked data, and connecting them with broader linked data on the Web. The prototype application highlights the fact that the available open data on the Web can help researchers to approach problems in multiple dimensions, and can provide further potential improvement by the use of data analytics and visualization technology. Beyond the judicial research field, the ideas described above are clearly applicable to other forms of research in numerous fields. From the perspective of web science itself, this work will create specific research problems such as knowledge provenance to guarantee the accuracy of data automatically retrieved from linked data cloud, and exploring "community intelligence"'s contribution to scientific research.

In future work, we will aim to explore more directly how these large amounts of data can help in the process of decision making and explore their effects on various metrics of decision quality.

6 ACKNOWLEDGMENTS

This work is based on the data-gov project at the Tetherless World Constellation at Rensselaer Polytechnic Institute, and is funded in part by a grant from the DARPA Transformational Convergence Technology Office. A number of colleagues have provided generous help with this work. We particularly thank Deborah L. McGuinness, Devin Gaffney, Timothy Lebo and James R. Michaelis for their insightful comments.

7 REFERENCES

[1] Berners-Lee, T. 2007. Linked data. http://www.w3.org/DesignIssues/LinkedData.html

[2] Berners-Lee, T. 2009. Putting government data online. http://www.w3.org/DesignIssue/GovData.html

[3] Ding L., DiFranzo D., Graves A., Michaelis J. R., Li X., McGuinness D. L., and Hendler J. 2009. Data-gov Wiki - Towards Linking Government Data. *AAAI Spring Symposium on Linked Data Meets Artificial Intelligence*.

[4] Epstein, L., Hoekstra, V., Segal J. A., and Spaeth H. J. 1998. Do Political Preferences Change? A Longitudinal Study of U.S. Supreme Court Justices. *The Journal of Politics*, 60, 801-818.

[5] Ivkovic', Z and Weisbenner, S. 2005. Local Does as Local is: Information Content of the Geography of Individual Investors' Common Stock Investments, *Journal of Finance*, 60, 267-306.

[6] Martin A. D., and Quinn K. M. 2001. Bayesian Learning about Ideal Points of U.S. Supreme Court Justices, 1953-1999. *8th Annual Political Methodology Summer Conference*. Atlanta, Georgia.

[7] Pritchett, C. H. 1948. The Roosevelt Court. New York:Macmillan.

[8] Rohde, D. and Spaeth, H. J. 1976. Supreme Court Decision Making. San Francisco, CA:Freeman

[9] Roll, R., 1984. Orange Juice and Weather, *American Economic Review*, 74, 5, 861-880.

[10] Schubert, G. A. 1965. The Judicial Mind. Evanston, IL:Northwestern University Press.

[11] Segal, J. A. and Spaeth, H. J. 1993. The Supreme Court and the Attitudinal Model. New York:Cambridge University Press.

[12] Segal, J. A. and Spaeth, H. J. 2002. The Supreme Court and the Attitudinal Model Revisited. New York:Cambridge University Press.

[13] Segal, J. A., Epstein, L., Cameron, C. M., and Spaeth, H. J. 1995. Ideological Values and the Votes of U.S. Supreme Court Justices Revisited. *The Journal of Politics*, 57, 812-823.

[14] Segal J. A. and Cover A. D. 1989. Ideological Values and the Votes of U.S. Supreme Court Justices. *American Political Science Review*, 83.

[15] Spaeth H. J. 1999. United States Supreme Court Judicial Database, 1953-1998 Terms [Computer File]. *Inter-University Consortium for Political and Social Research, 15th Edition*. Ann Arbor, MI.

[16] Spaeth H. J. and Segal J. A. 2000. The U.S. Supreme Court Judicial Data Base: Providing New Insights into the Court. *Judicature*.

[17] Tate C. N. and Handberg R. 1991. Time Binding and Theory Buiding in Personal Attribute Models of Supreme Court Voting Behavior, 1916-88. *American Journal of Political Science*, 35, 460-480.

[18] Tate N. C. 1981. Personal Attribute Models of the Voting Behavior of U.S. Supreme Court Justices: Liberalism in Civil Liberties and Economics Decisions, 1946-1978. *The American Political Science Review.*