Building a database into the heart of our newsroom
Goal: Make Sacramento comprehensible

- 931 bills passed
- 80 state assembly people
- $63B spent on education annually
- 2,343 registered lobbyists
- 120 state representatives
- $300M spent by lobbyists
- $254B state spending plan
- 384 state agencies
- $108B general fund budget
- $198B unfunded pension liability
- 40 state senators
- $428M raised for 2014 state campaigns

CA
Money isn’t the only story. It’s the fuel, but the legislature is the engine.

- Where they come from (district characteristics)
- Interest groups and lobbyists
- Personal characteristics
- Campaign financing and independent expenditures
Data as a catalyst, not a roadblock

Focus on the storytelling and distribution challenges

Make it easy for all reporters to independently run queries

Ability to scale
A model built around the legislator

- Financial Assets
- Salary, Travel, Car, etc
- Personal Wealth/Net Worth (Investments, Property, Businesses, etc)

- Term in Office
- Office
- Public Offices
- Office History
- Elections
- Votes / Candidates

- Lobbyists
- Profile
- For / Against

- Bills
- Authors & Opponents
- Yes / No, Pass / Fail

- Vote
- Vote & Attendance

- Person
- Elected Official
- SIG Payments to Lobbyists

- Candidates
- Affiliation

- Campaigns
- Contributions & Independent Expenditures

- Political Parties
- Contributions

- Financial Expenditures
- SIG Expenditures

- LEGISLATOR
- Profile
- Members / Roles
- Attendance

- Committees
- Committee Hearings

- Districts
- District Data
- Economy / Demographics / Voting

- Interest Groups
- SIG Ratings & Endorsements
- Representing

- INDUSTRY CATEGORIES

- People / Organizations
- Legislative Activity
- Public Office & Elections
- Money / Donations

Support / Oppose
Data sources

- VoteSmart
- Sunlight
- Maplight
- Center for Jobs
- CAL Access
- FPPC
- …

Configurable loader with fuzzy matching

Front end search UX and Python APIs
A simple search interface...
... That can run sophisticated queries
Learnings

• You can’t rely on anyone (from a data source perspective)
• Fuzzy matching worked really badly, and then really well

Next Steps

• Adding several more data points (including IEs)
• Test with partners