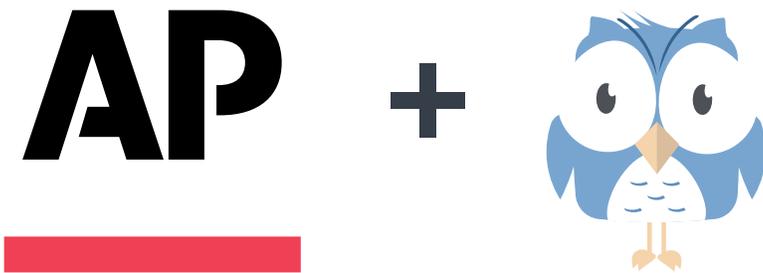


How the Associated Press puts data journalism in the heart of local newsrooms



Meet the Associated Press.

The Associated Press is an independent, not-for-profit news cooperative dedicated to telling the world's stories, from breaking news to investigative reporting, providing its customers with compelling data ready to be turned into localized news stories.

-  Over half the world's population sees AP content every day
-  Teams in 250+ locations, 100+ countries
-  Clients include 3,000+ publishers
-  52 Pulitzers and counting

AP data is more valuable on data.world

“We thought if we could put this data into the hands of our members and customers, they could find stories in the data we'd never see.



Troy Thibodeaux
Editor, Interactive
Newsroom Technology

Challenge

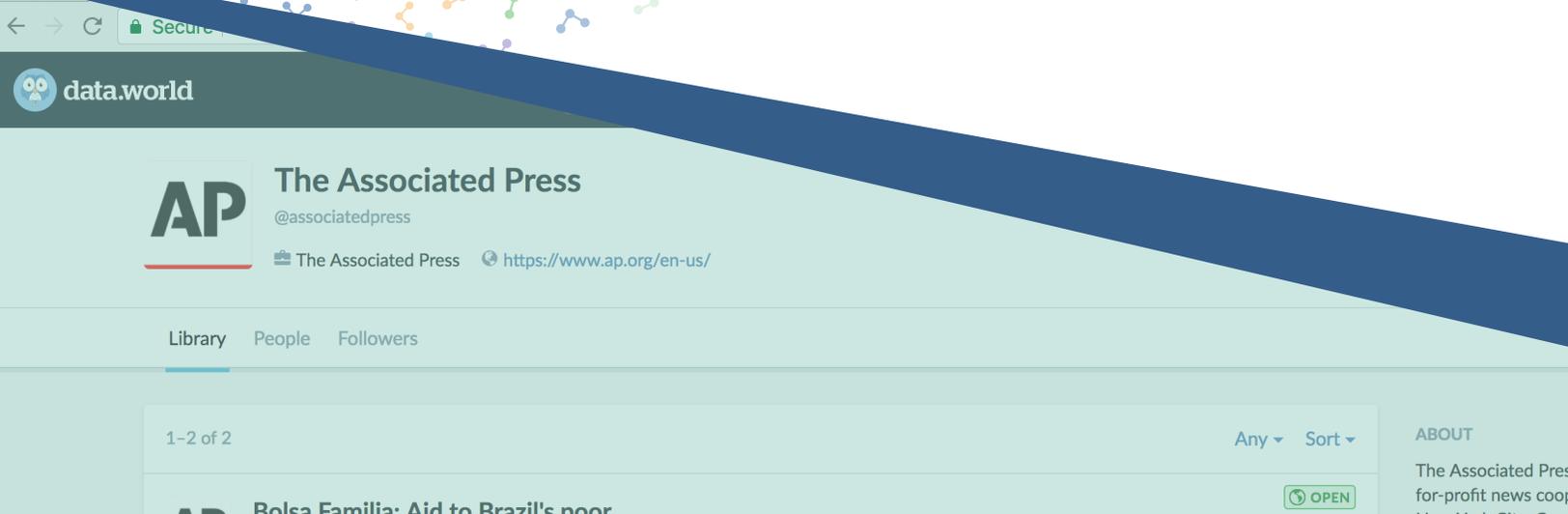
- Members weren't making the most of AP's data: it went to the wrong people in an organization, was tough to connect with the right analytics tools, and was difficult to find.
- AP was investing significant cost and resources building infrastructure to support its data services.

Solution

- AP brings together clear, comprehensive data on a foundation designed to scale.
- Processes and production are structured to match today's 24/7 news cycle: faster, more efficient, more transparent.
- Now, 300+ organizations instantly gain access to vetted, story-rich data.
- Increased collaboration powers a more engaged, self-supporting network of AP customers.

Results

- Doubled data production
- More than doubled customer reach
- Single point of contact to engage customers
- Data becomes more valuable as people work with it
- Consolidated, streamlined tech stack
- Less time cleaning and managing data



Clear results

Grow AP's business

People are hungry for data access. With data.world, AP more than doubled customer reach.



Reduce operational costs

AP eliminated redundant tools and avoided spending significant cost on building a tech solution from the ground up. With data.world, AP **got to market faster and more cost effectively.**



Get more done

AP **doubled its average data output and increased speed to publication.** Less time managing data and supporting members means more time is spent developing stories that influence the world.



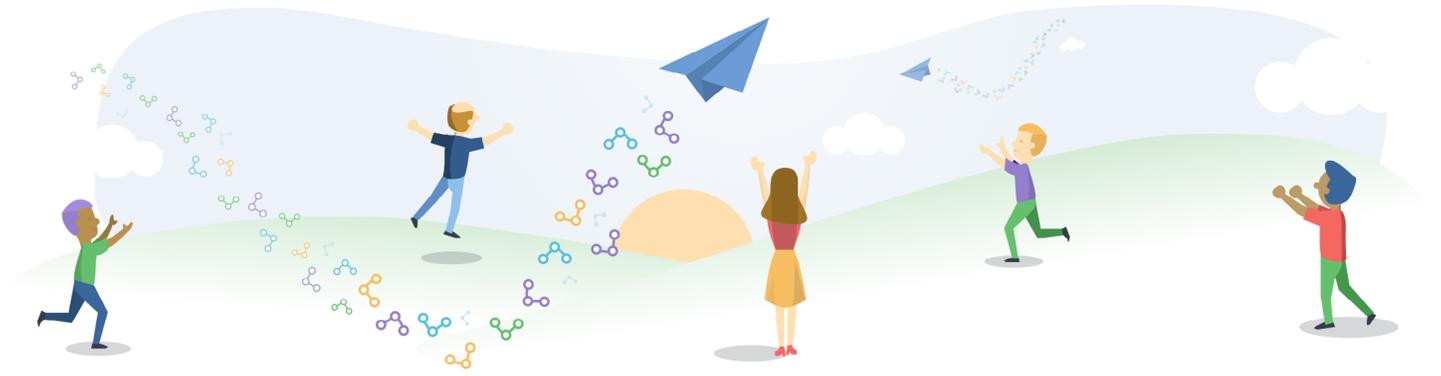
Give customers a better experience

AP **spends less time on support** and more time deepening client relationships. Unlike past solutions, AP gets fewer questions about how to use data.world. People just intuitively know how to use it.



The challenges of data dissemination

Providing the right data to the right people in AP's broad network comes with extensive operational needs.



The Associated Press provides quick and reliable access to compelling content that's ready to be turned into local news stories. But getting data from those stories into the right hands in local newsrooms is no small task.

Previously, data would reach the wrong people, get distributed at times that did not match user needs, and get lost in inboxes far from those who could leverage it. Often, using data required more time and work than busy journalists had. Advanced data professionals wished they could easily bring AP data into their favorite analysis tools while non-technical users felt overwhelmed.

Before data.world:

- Data was scattered, lost
- Data didn't reach the right people
- Redundant, expensive tools
- Too much time spent disseminating resources
- Diminishing engagement and value-add for users

Data should be easier to discover

Customers weren't making the most of AP's data because it wasn't reaching the right individuals within newsrooms. Individuals who needed AP data had no way to find it.

Build a better foundation

Organize scattered data. Bring together every asset and user—of any type—into a single, powerful platform.

AP began by bringing together strengths they already had: rigor and passion for data and an extensive customer base hungry for insights. They replaced disparate tools with data.world to provide members with a unified, robust, integration-ready platform that made it easy to find and use data.

For AP, data.world was the best option to have immediate impact on the data subscription service. Other tools needed extensive infrastructure investments, lacked key capabilities, or wouldn't scale with the program's success.

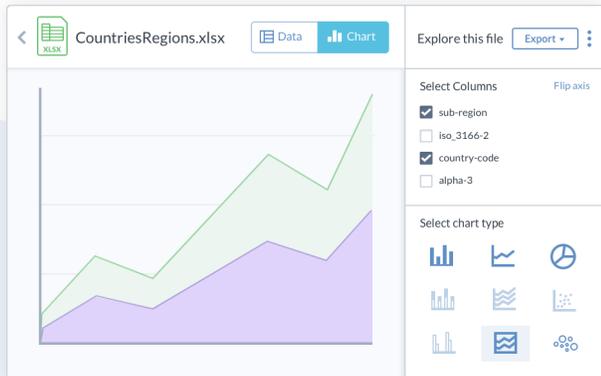


Impact

- Consolidated, streamlined tech stack
- Less time cleaning and managing data
- Single point of contact to engage customers

Make insights accessible

data.world helps AP support the full spectrum of newsroom data literacy



AP wanted to help everyone get the data they need. Insights make it simple for non-technical users to gather key takeaways.

Those who wish to dig deeper get direct access to clean, consistent underlying data. Managing editors no longer need to serve as the only data access point. With data.world, new data resources reach everyone that can use them.

name	alpha-2	alpha-3
Afghanistan	AF	AFG
Aland Islands	AX	ALA
Albania	AL	ALB
Algeria	DZ	DZA
American Samoa	AS	ASM
Andorra	AD	AND
Angola	AO	AGO
Antarctica	AQ	AIA
Antigua and Barbuda	AG	ATA



@dash wants you to contribute to the dataset Power Grid Data

Subject: Files were added to a dataset

Hi @bgadoci,

@dash added the following files to your dataset Power Grid Data

- eiaserviceterritories2014.csv
An index of electricity providers nationwide & the counties...
- eia_reliability_2014.xls
Reports of power outages & other emergency incidents...
- cleaning.ipynb
Python notebook used to clean the data.

Alerts instantly notify users of new assets, questions, and insights. Stories publish faster.

Data should be easier to explore

Data workers rely on highly customized toolchains. AP needed a tool-agnostic data offering so it could provide “something for everyone,” regardless of tool configuration, workflow, or persona.



data.world

Customers get what they need, faster

“ I often say that most of the work in a data project is caught up in that **unglamorous 80%**—finding the data, vetting it, cleaning it, coming to understand its limitations.

We were doing that for all of these stories anyway, so why not let our members benefit from that head start?



Troy Thibodeaux
Editor, Interactive Newsroom
Technology

AP collects, prepares, and packages the data so its members can easily discover local stories.

Every dataset ships with ready-to-fire SQL queries and simple instructions to guide less-technical members to easily segment and explore the data. Data-savvy journalists start with powerful building blocks for sophisticated analysis.

Centralizing data prep **substantially reduces time-to-publication for member newsrooms.**



Impact

- 300+ AP members gained instant access to a pilot program with reliable, pre-prepared data
- Increased activity: AP's datasets have been queried over 5,000 times

The screenshot shows a data visualization interface. At the top, there is a SQL query: `SELECT year, month, sum(low_income_families) as total_lowincome, sum(total_bf_funding) as total_bf_funding, (sum(low_income_families)/sum(total_bf_funding) * 100) as pct_served_by_bf FROM bolsa_familia_by_region GROUP BY year, month`. Below the query, there are two tabs: "Results table" (selected) and "Chart". The "Results table" tab displays a table with 77 query results. The table has columns for year, month, total_bf_families, total_bf_funding, and total_lowincome. The data is grouped by year (2011) and month (1-5).

	year	# month	# total_bf_families	# total_bf_funding	# total_lowincome
1	2011	1	12,851,345	1,210,230,621	
2	2011	2	12,946,313	1,224,704,614	
3	2011	3	12,944,676	1,224,307,029	
4	2011	4	13,058,476	1,471,589,861	
5	2011	5	12,986,870	1,467,768,035	

Data work that helps everyone

data.world extends the technical capabilities of AP customers' teams to everyone else. Technical users quickly create and share queries without leaving the platform or spinning up a database. Results can be exported in common formats so anyone can dig in.

Everyone reaps the benefits: customers no longer have to start from scratch by creating their own queries; they can use and adapt what AP provides centrally.



Bolsa Familia: Aid to Brazil's poor / workspace data.world

TABULAR FILES (1) Hide

- Bolsa_Familia_by_Region.csv CSV

DOCUMENTS (2) Hide

- Dataset summary MD
- Data dictionary MD

QUERIES (5) Hide

- New query ▾

National Totals by Month SQL SHARED

```
1 ##FIND BRAZIL'S TOTAL SPENDING AND RECIPIENTS FOR BOLSA FAMILIA
2 ##Clicking "Run query" will give you the monthly national figures on bolsa familia recipients,
3 ##spending and the percent of low-income families who are served by Bolsa Familia
4 SELECT year, month,
5     sum(families) as total_bf_families,
6     sum(funding) as total_bf_funding,
7     sum(low_income_families) as total_lowincome,
8     sum(families)/sum(low_income_families) *100 as pct_served_by_bf
9 FROM bolsa_familia_by_region
10 GROUP BY year, month
```

Results table Chart 77 query results Download ▾

	year	month	# total_bf_families	# total_bf_funding	# total_lowincome	pct_serve
1	2011	1	12,851,345	1,210,230,621	20,094,955	
2	2011	2	12,946,313	1,224,704,614	20,094,955	
3	2011	3	12,944,676	1,224,307,029	20,094,955	
4	2011	4	13,058,476	1,471,589,861	20,094,955	
5	2011	5	12,986,870	1,467,768,035	20,094,955	
6	2011	6	12,999,560	1,470,915,862	20,094,955	
7	2011	7	12,952,039	1,466,578,595	20,094,955	
8	2011	8	12,805,037	1,482,229,047	20,094,955	
9	2011	9	13,179,472	1,573,687,473	20,094,955	
10	2011	10	13,171,810	1,571,790,011	20,094,955	
11	2011	11	13,306,920	1,594,506,647	20,094,955	
12	2011	12	13,361,495	1,605,970,114	20,094,955	
13	2012	1	13,330,714	1,561,780,652	20,094,955	

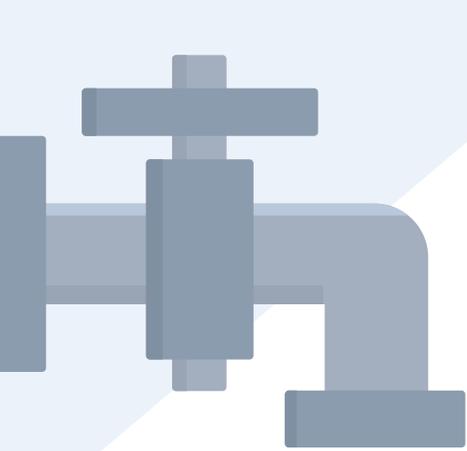
Dataset schema

Search the schema

- ▼ bolsa_familia_by_region
 - year
 - month
 - region
 - families
 - funding
 - low_income_families
 - pct_served
 - yearmonth

Data should be easier to share

AP needed to easily manage access control and permissioning so the right people always have the right level of access to the right data at the right time.



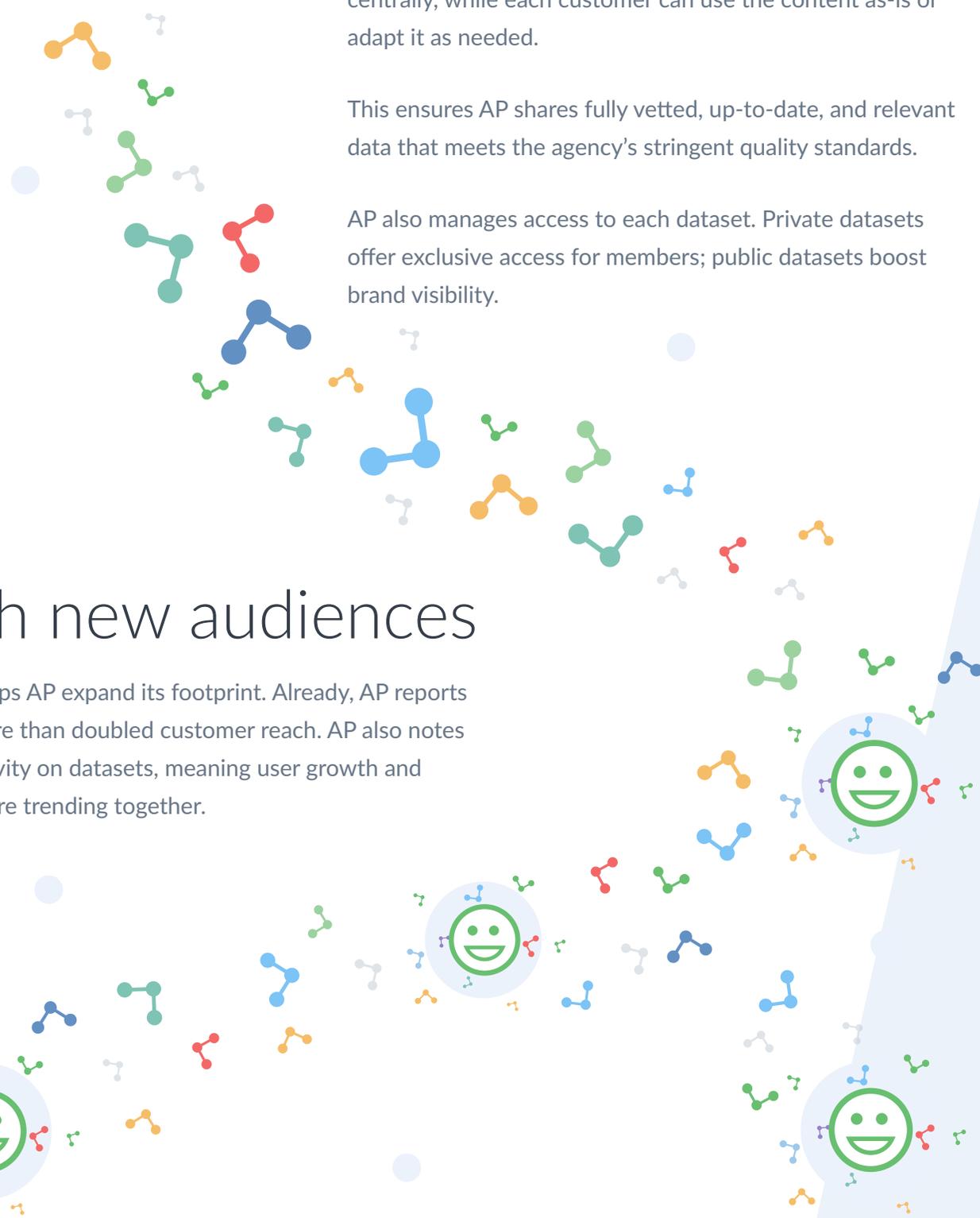
Maintain control

AP manages the data that makes its entire network more effective.

AP's data offering is available to 3,000 publishers and a range of businesses. Assets are always created and maintained centrally, while each customer can use the content as-is or adapt it as needed.

This ensures AP shares fully vetted, up-to-date, and relevant data that meets the agency's stringent quality standards.

AP also manages access to each dataset. Private datasets offer exclusive access for members; public datasets boost brand visibility.



Reach new audiences

data.world helps AP expand its footprint. Already, AP reports that it has more than doubled customer reach. AP also notes increased activity on datasets, meaning user growth and engagement are trending together.

Data should be easier to **scale**

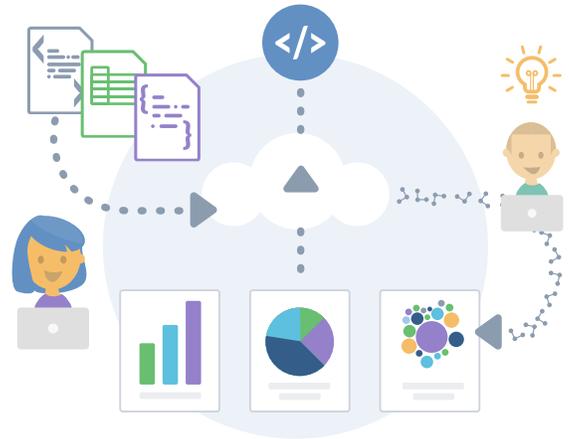
AP needed to spend less time building infrastructure and more time building data products for members. Encouraging open, collaborative discussions could also reduce time spent answering FAQs – and help customers go from data to insights even faster.

Move at the speed of news

Skip starting from scratch.

Creating and sharing an ongoing stream of fresh datasets has allowed AP to sustainably expand its data business. With data.world, AP doubled its data production. More than 20 datasets have been produced in the past six months alone.

Creating a platform for shared insights give data journalists a better starting point, powering more informed, collectively-optimized journalism.



“They just intuitively understand.”

AP reports a reduction in customer support needs. Thanks to annotation in the form of descriptions, sample insights, and data dictionaries, members find it easier to explore, understand, and apply insights from new datasets.

“We have almost no questions from our members about data.world. Questions focus on the story and how to use the data, and that’s exactly what we want: a tool that gets out of the way and lets us communicate and share.”



Troy Thibodeaux
Editor, Interactive Newsroom
Technology

Congress_Elections.xlsx
Efficiency gap analysis for U.S. House races in 43 states. Seven states have only one U.S. House seat and thus have no efficiency gap. These states are Alaska, Delaware, Montana, North Dakota, South Dakota, Vermont and Wyoming.

[clean data](#)

state	State holding a general legislative election in 2016
# republican_votes	Total votes received by Republican candidates
# democratic_votes	Total votes received by Democratic candidates
# district_average_republican_vote_share	Statewide average of the share of votes that Republican candidates received in each district
# district_average_democratic_vote_share	Statewide average of the share of votes that Democratic candidates received in each district
republican_seats_won	
# republicans_unopposed	
democratic_seats_won	

Congress_Elections.xlsx
[clean data](#) Efficiency gap analysis for U.S. House races in 43 states. Seven states have only one U.S. House seat and thus have no efficiency gap. These states are Alaska, Delaware, Montana, North Dakota, South Dakota, Vermont and Wyoming.

state	# republican_votes	# democratic_votes	# district_average_republican_vote_share
1 Alabama			
2 Alaska			
3 Arizona			
4 Arkansas			
5 California			

TYPE	#	decimal
DESCRIPTION		
Statewide average of the share of votes that Republican candidates received in each district		

Data's value increases over time

Instead of discussion and documentation getting buried and fragmented across emails and in-person conversations, AP's **data becomes more valuable as people work with it.**

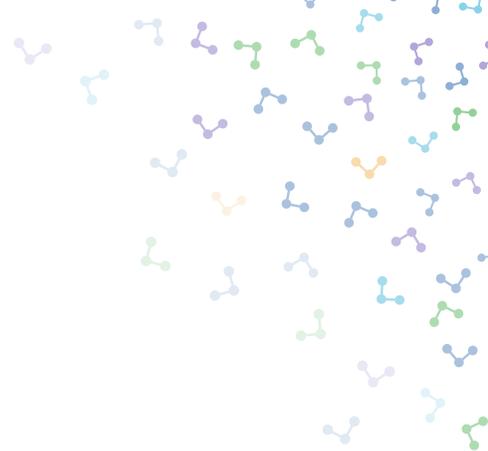
Members now have self-help tools that make it easier to explore the data. Discussions help surface common questions, insights that others can use, documentation needs, and more.





Measure impact

The unified platform gives AP the ability to manage data provenance more effectively.



POSTED IN BEHIND THE NEWS

How data journalism tools revealed segregation in US charter schools

Dec. 8, 2017, by Lauren Easton

In a memo to staff, Deputy Managing Editor Sarah Nordgren described how an AP analysis of charter school enrollment data allowed member news organizations to zero in on segregation in their states:



As journalists use AP datasets in their content, the activity can be attributed and measured.

Source: <https://blog.ap.org/behind-the-news/how-data-journalism-tools-revealed-segregation-in-us-charter-schools>



Maine Focus

White Maine students are least likely in nation to see kids of another race at school

Ashley L. Conti | BDN

An international student listens during English class at John Baptist High School in Bangor, Sept. 21, 2017. In Maine, eight in 10 white students attend schools where at least 90 percent of the total student population is white.



When news organizations use data.world to understand and share localized stories from AP data, the content is accounted for.

Source: <https://bangordailynews.com/2017/12/06/mainefocuswhite-maine-students-are-least-likely-in-nation-to-see-kids-of-another-race-at-school/>





Lead the market

As journalists increasingly tap into AP's data subscription service, it is being recognized for pioneering a new model for data journalism that bridges the resource and data literacy gaps between larger and smaller news organizations.



The new model has earned AP conference speaking opportunities at SXSW as well as the industry's most prestigious events, including the Investigative Reporters and Editors annual NICAR conference and the Online News Association's annual conference. The new model has also been covered in publications like TechCrunch and Poynter.

About data.world

data.world taps into more of your team's collective brainpower to achieve anything with data, faster. The Austin-based Public Benefit Corporation serves the enterprise, journalism, and public sector communities by speeding up data work and creating shared context to preserve and perpetuate essential project knowledge.



To learn more about data.world, please contact support@data.world