

Integrating Data Journalism into Newsrooms

by Michael Zanchelli with Sandra Crucianelli

**KNIGHT INTERNATIONAL
JOURNALISM FELLOWSHIPS**



International Center
for Journalists

Advancing Quality Journalism Worldwide

EXECUTIVE SUMMARY

This report aims to create a snapshot of the structure, process, incentives and resources behind well-established data-journalism teams at six major U.S. and British media outlets and one Latin American newspaper. The goal: to identify the common threads that are essential for data-journalism teams to work effectively in newsrooms.

Our findings are based on phone interviews and email exchanges with leaders and members of four data-journalism teams in the United States and one in Latin America. We also used online sources including the [Data Journalism Handbook](#), news articles, conference presentations and blog posts for additional examples of data-journalism teams, advice from experts on building teams, lessons learned and data-journalism stories that have an impact.

We found four key factors behind the creation of successful data-journalism teams:

1. Locating the data-journalism team close to the news desk: Members of data-journalism teams emphasized that being close to the news desk gives them critical access to editors and reporters as they and the data team develop and plan stories based on data. Proximity to the news desk also means opportunities to build relationships with reporters and editors, generating and improving collaboration between news desks and data-journalism teams.

2. Encouraging reporters and developers to work together to come up with ideas for data-driven stories: Because developers and reporters often have specialized skill sets, it's important to bring them together to brainstorm story ideas. This increases the likelihood of identifying unique and significant angles for data-driven stories.

3. Recruiting reporters and developers who bridge the skills gap:

Data-journalism team members who have backgrounds and skills in journalism and data mining, development and/or coding are valuable assets to data-journalism teams because they can sort and analyze data and then identify news and trends that resonate with their audience.

4. Producing stories that show what data mean and why the audience should care:

Data-driven stories about topics affecting the lives of news consumers produce impact and drive web traffic, underscoring the need for greater investment in data-journalism teams.

Key Findings

We examined data-journalism teams at the BBC, *The New York Times*, *The Guardian*, *USA TODAY*, *The Washington Post*, *O Estado de São Paulo* and the *Chicago Tribune* to find out how the seeds for data awareness were planted, how each data-journalism team is structured, common challenges faced by data-journalism teams and the process of creating a data-driven story. The data-journalism teams at the six media organizations from the U.S. and Great Britain and one Latin American newspaper have produced [award-winning](#) work. Some have been working for decades, adapting to a changing landscape of journalism. These teams also are noted for their thought leadership in data journalism through regular participation in [conferences](#) and frequent [publication](#) of information about data journalism.

As part of our research, we reviewed online sources including the [Data Journalism Handbook](#), news articles, conference presentations and blog posts, which we searched for advice from experts on building data-journalism teams, lessons learned in the process and examples of data-journalism stories that spurred action.

Ultimately, we found four key factors that underlie the creation and successful operation of data-journalism teams. Each of these factors is described in detail below, using examples from the data-journalism teams we researched. A list at the bottom of the report has specific information on the structure of each of the seven data-journalism teams, the process and workflow used to produce data-driven stories, the type of data-related stories produced (i.e., visualizations, maps, long-form, short-form) and the format of these stories (online, print or both).

Locating the data-journalism team close to the news desk

Paul Overberg, database editor at *USA TODAY*, says, “Being in the newsroom really matters.” This is the most-common refrain of the people who have been leaders in data journalism. Simon Rogers, editor of *The Guardian’s Datablog*, [reinforces this claim](#): “News organizations are all about geography—and proximity to the news desk. If you’re close, it’s easy to suggest stories and become part of the process; conversely, out of sight is literally out of mind.” Serdar Tumgoren, news applications developer at *The Washington Post*, explains that a team of “embedded developers” sits next to news desks to offer support for data analysis and to create apps that make it easier for journalists to use data in their reporting. He says the close proximity of the embedded developers to the news desks is a key factor in being able to collaborate effectively on data-driven stories.

Bella Hurrell, specials editor at the BBC News Online, [explains the BBC’s approach](#): “We have found that proximity really is important to the success of projects. Although we have done this for a while, increasingly other organizations are reorganizing along these lines after coming to realize the benefits of breaking down silos and co-locating people with different skill sets can produce more innovative solutions at a faster pace.” Putting developers and journalists together is particularly important when team members specialize in different disciplines and when journalists’ data skills and developers’ journalism skills are weak. Brian Boyer, former news apps editor at the *Chicago Tribune* and news apps editor at NPR, says this is the case at the newspaper, where the [news apps team](#) is comprised of only developers: “It’s important that we sit in the newsroom. We usually find work via face-to-face conversations with reporters. They

KEY FINDINGS

Locating the data-journalism team close to the news desk

know that we're happy to help write a screen scraper for a crummy government website, tear up a stack of PDFs, or otherwise turn non-data into something you can analyze." The *Chicago Tribune's* news apps team supports editors and reporters, helping them to research and report data-driven stories, visualize stories online and develop web resources for news consumers.

Media in developing countries have not always been able to replicate the successes of their counterparts in developed countries when it comes to integrating investigative journalism, computer-assisted reporting and more recently, data journalism into their newsrooms. In the developing world, media are hampered by a lack of human and technical resources, technological problems, and in some cases the way journalists are educated at universities. The *sine qua non* for an effective data-journalism team—whether it is in the developed or developing world—is the integration of that team into the newsroom. At *La Nación* newspaper in Costa Rica, Giannina Segnini leads a team of data journalists that works in the newsroom. The team consists of three journalists, including Segnini, and two developers. One of the programmers is in charge of data mining (cleaning, loading and data processing). The other is tasked with developing interactive applications. The team also has access to the professional services of other departments including the Geographic Information Systems (GIS) department for geography and map-related data and visualizations and audiovisual departments.

José Roberto de Toledo, columnist and blogger at *O Estado de São Paulo* newspaper, coordinates a professional team dedicated to working with data. "I prefer

to call it 'the Core,'" he says. "The group includes a political journalist, an expert in Excel who also is a journalist, an expert in data scraping and a developer who has expertise in data visualization. In fact, all group members have developed several skills." Regarding the data journalism teams' placement in the newsroom, Toledo notes that "there's no other way to do it." According to Toledo, it is vital for "the Core" to be in the newsroom regularly producing content related to the newspaper's agenda and to interact with all the newspaper's sections. Using a soccer team as a metaphor, Toledo explains that each position brings different expertise to the game, all while working toward a common purpose.

Encouraging reporters and developers to work together to come up with ideas for data-driven stories

Putting developers and reporters in the same area of the newsroom also facilitates another component of these data-journalism teams: ideas for stories that come from either the developers/data miners or the journalists. At *USA TODAY*, for example, reporters and editors generate the story ideas. In fact, the members of the data team are journalists who have programming and other technical skills. As they consider possible data-driven stories, the data team keeps in mind the topics that are critical to the paper (i.e., politics, sports, airlines and travel). At *The Guardian*, the Datablog team, which includes a researcher, a part-time journalist and developers, originates story ideas. At the BBC, data-driven stories or visualizations come from all angles, originating with reporters, editors or developers. The process for producing data-related stories in each newsroom is described in more detail in the "Data Journalism Teams" section and lists, below.



"The Core" O Estado: Eduardo Malpeli and Amanda Rossi (front), José Roberto de Toledo and Daniel Bramatti (back) ([source](#))

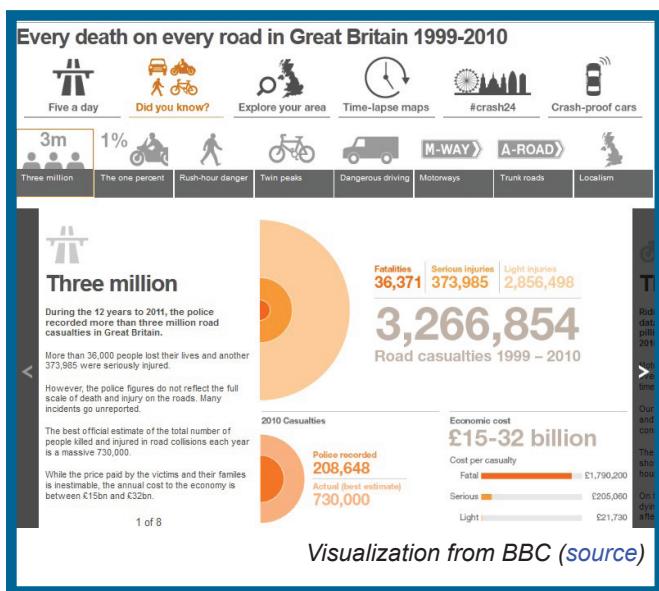
"The sine qua non for an effective data-journalism team – whether it is in the developed or developing world – is the integration of that team into the newsroom."

While one group may bring the initial story idea to the table, each group contributes to the story-creation process. Ultimately, each group adds value and members of each group may have similar skills. Generalizing, developers or data miners tend to specialize in analyzing, interpreting and visualizing data. These skills give developers a knack for understanding how to extract the numbers, see patterns and trends and interpret what those numbers may mean. Reporters tend to bring an understanding of the history and significance of the numbers. They can put the numbers into a broader political, social and economic context and see trends developers might not immediately connect to issues that matter to citizens. Reporters also have storytelling skills.

Without the involvement of both developers and journalists, stories that involve data could fall short. Mc Nelly Torres, co-founder and associate director of the [Florida Center for Investigative Reporting](#), describes some of the issues she faced when only developers generated and developed visualizations and databases. When she worked with a young developer at the *Sun Sentinel*, Torres observed that the developer “didn’t know what the readers wanted to learn. He missed the story, where I could find it in the data.” Torres used her background as an investigative reporter to ask developers for specific data that ultimately helped her to paint a more complete picture for her online audience. Serdar Tumgoren of *The Washington Post* agrees that it’s important for reporters to collaborate closely with developers to help developers understand the context behind the data.

Recruiting reporters and developers who bridge the skills gap

Aron Pilhofer, editor of interactive news at *The New York Times*, says that as long we continue to talk about the skills gap between journalists and developers, the gap will persist. He says investments must be made to [make journalists comfortable](#) with basic data skills (spreadsheets, data analysis), even if it means hiring a strong reporter who is willing to use new tools over a journalist who is a stronger writer. Paul Overberg, database editor at *USA TODAY*, says that news organizations already recognize the need to cultivate talent that bridges the divide in skills. He points to the Medill School of Journalism at Northwestern University, where Rich Gordon, director of digital technology in education, is [recruiting students](#) with a bachelor’s degree in computer science to earn a master’s degree in journalism supported by a [John S. and James L. Knight Foundation](#) scholarship. From 2008 to 2011, [nine students](#)



graduated and landed positions on the *Chicago Tribune's* news apps team, and at the *Bay Citizen* in San Francisco, Public Radio International and NPR. One of these students is Brian Boyer, who has been asked to lead NPR's recently formed news apps team. The Medill School received a new grant from the Knight Foundation to extend [the scholarship program](#) from 2012 to 2015 and fund at least six additional students. That will bring the program's total number of scholarship recipients—developers who have master's degrees in journalism—to 15.

To help traditional reporters learn these skills, Pilhofer suggests starting small by showing journalists who lack the knowledge or resist learning these new skills the fundamentals of data analysis and basic spreadsheet applications. Mc Nelly Torres also recommends starting with the basics for journalists who don't have a data background: “A lot of journalists are so intimidated to learn this stuff [data-driven reporting]. I try and break the intimidation factor—showing how easy it is to use data in everyday reporting. For example, plugging in some numbers on a Google map or showing journalists how to play with Google public data.” At the BCC News Online, [editorial staff and reporters are required](#) to attend conferences and take in-house trainings to learn to use basic spreadsheet applications such as Excel and Google Docs.

For their part, developers, coders and data miners should have an understanding of best journalistic practices: accuracy, sourcing, fact-checking, creating narratives and writing stories. Helping developers focus on data

KEY FINDINGS

Recruiting reporters and developers who bridge the skills gap

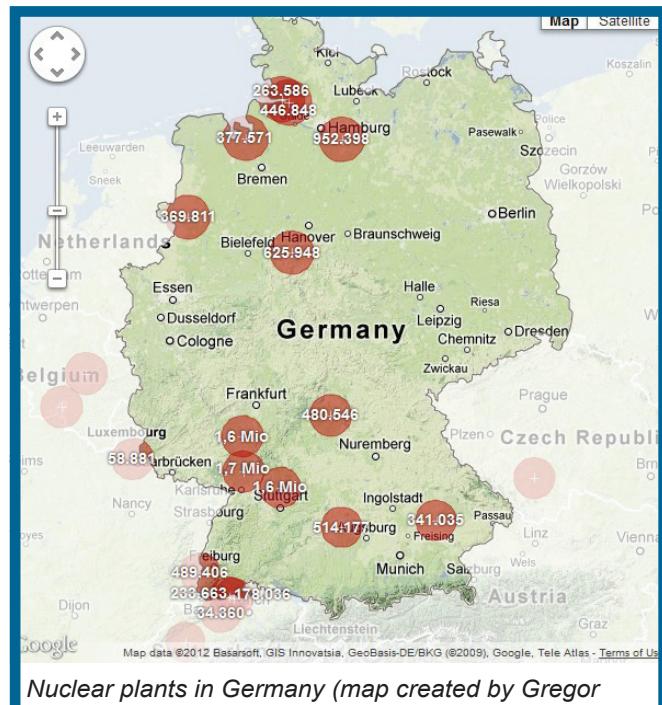
journalism that matters to readers/users also helps to guide their approach to data analysis. They will bring better ideas to the table. Derek Willis, web developer at *The New York Times*, talks about the need for “human-assisted reporting,” or the application of journalistic principles to data. He notes that “reporters are testing out theories all the time, often by calling up sources and putting a question or theory out there. There’s no reason why we can’t enable the same ease of inquiry with data. In fact, it seems possible to do it on a much broader scale and in more precise ways.”

Exacerbating the gap in skills: Often there is a physical divide between data-journalism teams and the news desk. As noted above, the data-journalism teams that have endured and excelled not only bridge the divide in skills, they close the physical gap between developers and journalists by placing them near each other in the newsroom.

Producing stories that show what data mean and why the audience should care

We found that each data-journalism team could point to a story that was a sort of “game-changer” for the organization. These stories were often extremely personal, connecting data with the lives of their audience. Mc Nelly Torres says a database she created for the *Sun Sentinel* changed the way managing editors thought about data-driven journalism. In 2007, Torres worked off hours to put together a database on the results of nail salon inspections in the Miami area. The database, published online with an accompanying [investigative report](#) by Torres, drew hundreds of thousands of hits to the newspaper’s website. Three months later, a story Torres produced on [restaurant inspections](#) followed the same model, drawing millions of hits. Before that story was published, the newspaper did not have a dedicated developer working with reporters to create data-driven stories for online publication. The overwhelming public interest in the nail salon inspection and restaurant inspection databases opened the editors’ eyes; the *Sun Sentinel* hired a developer to exclusively work with investigative reporters with the goal of increasing the volume of data-driven work presented online.

Databases drive the majority of the [web traffic](#) at *The Texas Tribune*, an Austin-based non-profit media organization. To date, the [most-viewed data app](#) (and [database story](#)) focused on government employee salaries, an issue that resonates with Texans as



Nuclear plants in Germany (map created by Gregor Aisch © ZEIT ONLINE)

government spending continues to be a key issue. Another example comes from *Zeit Online*, a German media outlet. During the 2011 nuclear disaster in Fukushima, Japan—the result of a massive tsunami—Zeit appealed to its users by publishing [an interactive map](#) showing the concentrations of Germans living at varying distances from nuclear plants. Sascha Venohr, development editor, says the [result](#) of this project was “lots and lots of traffic” with the map going “viral over the social media sphere.” While these results may lead to buy-in from managing editors and more resources for data journalism, Brian Boyer says “[the motivation should always be impact](#)—on people’s lives, on the law, on holding politicians to account, and so on. The written piece will speak to the trend and humanize it with a few anecdotes.” Boyer’s team at the *Chicago Tribune* produced a number of data-driven reports on important issues using the [Nursing Home Safety Reports](#) and [School Report Card](#) apps. The Nursing Home Safety report sparked the passage of legislation in Illinois that undid policies that had failed to eliminate violence in nursing homes across the state.

Whether it is a searchable database on restaurant inspections, safety reports on nursing homes or reporting on state employee salaries, stories based on data that clearly explain the impact of numbers on citizens can help change policy, drive traffic and help convince media to integrate data journalism across newsrooms.

DATA JOURNALISM TEAMS

Data Journalism Teams



Online only? Yes.

Number on team: The team that produces data-journalism for the BBC News website is [comprised of 18 journalists, designers and developers](#): eight journalists, five developers, five designers, who sit together.

Journalist data skill level/data awareness: BBC does not identify anyone as a “data journalist.” Instead, they [require editorial staff on the team to be proficient using basic spreadsheet applications](#) (Excel and Google Docs) to analyze data.

Structure of team: The team consists of journalists, designers and developers.

Workflow/process: The data-journalism team (“visual journalism team”) often [partners with investigative teams](#) for long-form stories. For basic visualizations, they work internally. Stories originate with the data team or journalists in the newsroom. At the BBC, it’s important that data people are interested in the news not just data. If a story will take longer than a week to produce, there is a kick-off meeting assigning roles (strategy, timeline, etc.). If members of the data team are working with people specializing in different disciplines, it helps to sit in close proximity.

Long- or short-form: Both.

Is there a data section?: Yes, [interactive and graphics section](#). Data-driven stories are also mainstreamed as part of the news.

Incentivizing reporters to use data: Editorial staff members are required to be proficient in basic spreadsheet use and applications. The BBC boosted these skills using in-house or external trainings.



Online only?: No, teams assist with print stories.

Number on team: Four teams doing data-related work with between five and ten developers, graphics people or journalists on each team.

Journalist data skill level/data awareness: Reporters who work on one of the four data-related teams must be proficient in multimedia, interactive news, computer-assisted reporting and graphics.

Structure of team: The four teams work like individual news desks. The editor on each team serves as the gatekeeper and manager of the reports. Editors and journalists on these teams have sound knowledge of tools and data analysis.

Workflow/process: Reports or visualization typically [originate within the four teams or a reporter from a different news desk](#) can approach for help with a story. If a reporter from a different news desk suggests a story, they partner with one of the four data-related teams. If a story is suggested within one of these teams, the editor of that team decides what stories are publishable.

Long- or short-form: Both.

Is there a data section?: Yes, but no section called “Data.” There is a [Multimedia/Photos section](#), which has interactive features, data-driven stories and other graphics. The four data-related teams also contribute to reports from topical news desks (Politics, Business, World, etc.). Also, there is a [Linked Open Data section](#) of the site.

the guardian

Online only?: The [Datablog](#) is only online. The data-journalism team supports print stories, as well.

Number on team: Primarily [three members of the team](#) including Simon Rogers, editor of the Datablog, one researcher and one part-time junior journalist. The team sometimes gets support from developers for more complex visualizations or tools.

Journalist data skill level/data awareness: Wikileaks was pivotal in merging the data people with the journalists, both in terms of physical location and increased collaboration.

Structure of team: Before the Wikileaks cables were released, the Datablog team sat with graphics, located on a different floor than the news desk. Following Wikileaks, [the team moved next to the news desk](#). According to Simon Rogers, the move facilitates the flow of story ideas from data team to the news desk. Reporters have also been more likely to use the data team on stories, now that they are located in close proximity.

Workflow/process: The first step is finding the data. The Datablog team uses a variety of sources including “breaking news stories, government data, journalists’ research and so on.” The team then explores possibilities for analysis (combine with another dataset, show changes over time, etc.). The resulting work is a story, a map, a graphic or a tool. A separate team of developers (“dev team”) produces the more complicated, intricate graphics. More commonly, the Datablog team uses free Google tools to create maps efficiently and easily.

Long- or short-form: Work produced by the Datablog team is short. When collaborating with other news desks, the work is generally more long-form ([example here](#)).

Is there a data section?: Yes, the Guardian’s [Datablog](#).

news app team for support, understanding, analyzing and visualizing.

Workflow/process: The team [supports editors and reporters](#) to help research and report data-driven stories, visualize stories and develop web resources for new consumers. All app ideas come from the reporters and editors in the newsroom. The apps team helps reporters mine and scrape data. Some of the data scraping and mining work leads to the development of a news application. On projects that take longer than a week to produce, the apps team shows their work to reporters, editors and other stakeholders during the process.

Long- or short-form: Both.

Is there a data section: “[Maps & apps](#)” section online.



Online only?: No, data-driven stories are mainstreamed in print in addition to online stories.

Number on team: Five journalists and one editor working on data-journalism team; five reporters on the investigative team (one editor, four full-time reporters).

Journalist data skill level/data awareness: High. The data team and investigative reports team are comprised of journalists with coding, statistics and data backgrounds. One person on the data-journalism team used to be a developer and true database administrator in a former job.

Structure of team: More of a ‘classic data team’ made up of journalists who have gone data. The data team can do programming, scraping, analysis, etc.

Workflow/process: The reporter and the editor are the source of the idea. Reporters and editors strongly guard that. The data team keeps in mind the topics that are critical to the paper (i.e., airlines and travel). The data team works with that reporter and editor and provides a range of options for how to use and visualize the data. Ultimately, the reporters and editors can choose one of these options.



Online only?: No, data-driven stories are mainstreamed in print in addition to online stories.

Structure of team: The news apps team is based in the newsroom, serving as reporter support. The apps team has built strong personal and professional relationships in the newsroom. When reporters have data, they rely on the

Being in the newsroom really matters. In 1999, each member of the data team was embedded in different content sections (Politics, Business, etc.). They had been attached subject- and content-wise. Presently, the data team sits together as *USA TODAY* moves toward a much more team-based, flatter organization. Data-driven stories originate and are pushed forward in three different ways:

1. Reporting-team-driven data: Sometimes a team leader will say, what kind of data do you have to bring to bear on question?
2. Issue/news-driven data: Sometimes you take a running look at data, like campaign-finance data as it flows in over the next six months or on health care industry money.
3. Data-team-driven data: Data team approaches a topic because it's a hot issue.

Long- or short-form: Both. The data-journalism team works closely with the investigative team for longer work.

Is there a data section?: No, data-driven stories are mainstreamed according to topic (i.e., politics, environment, business, demographics, etc.)

The Washington Post

Online only?: No, data-driven stories are mainstreamed in print in addition to online stories.

Number on team: 12-13 programmers/developers embedded in the newsroom with at least five working on apps. These developers make data available to reporters (legislation, campaign finance, national security) by giving raw access and creating tools to make data more useful on a day-to-day basis. Two people work on computer-assisted reporting focusing on the data analysis itself. Both the computer-assisted reporting experts and embedded developers sit in the newsroom, disbursed across different news desks. There are no journalists on the team, but some members have a journalism background. A handful of journalists at news desks have data skills.

Structure of team: The “Embedded Developers” used to be loosely assigned to news desks by topic, based on specialty (i.e., politics, national security, energy and environment). They have morphed into a general purpose team providing support to reporters, developing the front and back ends of visualizations. The team is part of the IT team, a service department, but sits in the newsroom, separate from the IT team.

Workflow/process: For stories, reporters will approach the embedded developers or computer-assisted reporting experts if they need assistance with data. Embedded developers also work on building news apps for reporters and data visualizations. Due to the high volume of work, the embedded developers have set up a “Pitch Committee” where new app ideas are judged and selected. No reporter has pitched an idea yet, but the committee is in its early stages.

Long- or short-form: Both.

Is there a data section?: No, data-driven stories are mainstreamed by topic.

O ESTADO DE S. PAULO

Online only?: No, data-driven stories are mainstreamed in print in addition to online stories.

Number on team: The team has two journalists and a developer. There is one more journalist from the National Desk who works with the team every day as an informal member, but whose role is just as important as the other, formal members.

Journalist data skill level/data awareness: The journalists are well skilled in spreadsheet software, basic statistics, database applications, Google tools (Google Refine, Google Maps), basic data scraping and precision journalism.

Structure of team: The team includes one coordinator, two journalists and a developer who writes the code for visualizations or apps.

Workflow/process: The coordinator generates the story idea (“the problem” in his words) and the journalists pursue it. The coordinator and two journalists gather the data, structure it, clean it and the developer writes the code. Other journalists or designers at *O Estado* are sometimes used to help find data. All data journalism team members participate in conceptualizing the visualization and app ideas.

Long or short-form?: Both. For example, the *O Estado* team recently published an application, *Basômetro*, an app that displays real-time congressional support for the federal government as measured by voting records on legislation. This app took months to create, needs regular updating and is used by academia to research and understand national politics. During the creation of *Basômetro*, the *O Estado* data journalism team helped the newspaper in many smaller projects on breaking stories of the day.

DATA JOURNALISM TEAMS

Is there a specific data section online?: Yes, a [website](#) where all of the data journalism projects are posted. This site is outside the newspapers' formal website, but linked.

Is there a data section: "Maps & apps" section online.

The International Center for Journalists advances quality journalism worldwide. Our hands-on programs combine the best professional practices with new technologies. We believe that responsible journalism empowers citizens and holds governments accountable.

The Knight International Journalism Fellowships, are designed to instill a culture of news innovation and experimentation worldwide. The goal is to seed new ideas and services that deepen coverage, expand news delivery and engage citizens in the editorial process. The program is funded by the [John S. and James L. Knight Foundation](#).

Michael Zanchelli is a program officer at the International Center for Journalists.

Sandra Crucianelli is a Knight International Journalism Fellow in Argentina.