Solving journalism's hidden problem: Terrible analytics

Tom Rosenstiel

INTRODUCTION

The path toward sustainable journalism, already challenged by a disrupted advertising business model, is also being undermined by something more unexpected—terrible data.

Analytics—another word for audience data or metrics—was supposed to offer the promise that journalists would be able to understand consumers at a deeper level. Journalism would be more connected and relevant as news people could see what audiences really wanted. Handled well, this should have helped journalists pursue what is at its core their fundamental challenge: learning how to make the significant interesting and the interesting more significant.

But a generation into the digital age, the problem associated with analytics isn't the one that some feared—the discovery that audiences only care to be entertained and distracted. The bigger problem is that most web analytics are a mess. Designed for other purposes, the metrics used to understand publishing today offer too little information that is useful to journalists or to publishers on the business side. They mostly measure the wrong things. They also to a large extent measure things that are false or illusory.

As an example, the metric we have taken to call "unique visitors" is not what it sounds. Unique visitors are not different people. Instead, this metric measures devices; the same person who visits a publication on a phone, a tablet, and a computer is counted as three unique visitors. If they clean their cookies they are counted all over again. The traffic to most websites is probably over counted by more than double, perhaps more than triple.

Time spent per article, in contrast, might offer a sense of depth of interest in a particular piece. But by itself it might also mean that someone stopped reading and

Tom Rosenstiel is a nonresident senior fellow at the Brookings Institution. He is the executive director of the American Press Institute and one of the nation's most recognized thinkers on the future of media.

> Designed for other purposes, the metrics used to understand publishing today offer too little information that is useful to journalists or to publishers.

walked away from the computer.

Page views can tell a publisher how many times an individual piece of content was viewed. But views cannot tell the publisher why.

Using conventional analytics, every story is an anecdote. Publishers may look at popular stories and say let's do more like those. But they are largely inferring what "like those" means.

More often, the metrics reveal that the most widely viewed stories are extraordinary by definition—and fairly predictable. I know of one publishing company, for instance, that examined its 50 most popular stories each year to look for patterns. Nearly all of their most popular stories were outliers, stories that revealed no patterns at all.

Knowing that a story about the championship football team or the bizarre tale involving a celebrity was viewed by a lot users offers little help in understanding how to make one's coverage of government or water quality 20 percent more engaging.

Nor does a page view by itself tell whether consumers found that content valuable or an annoying waste of time they were teased into viewing.

In addition to the problem that the metrics used in digital life are often crude and insufficient, there are too many of them. Yet if you pick the wrong ones to focus on at the exclusion of others—page views as an example while ignoring time—it may chase you down the path of ignoring the significant in favor of trivial and hasten the irrelevance of your publication.

Journalists themselves haven't helped. The prospect of suddenly knowing how many people read or watched a given story was frightening. To journalists, the whole was always bigger than the parts. You put the great photo on Page 1, the comics inside, or teased video of the water skiing squirrel coming up, in the hope that people would learn the important news along the way. But if you chose the other path, putting bikini clad celebrities at the top of your web page every day (and some have), you have forfeited your authority in favor of page views to sell banner ads that only a fraction of one percent of those viewers will ever click on.

Journalists also feared metrics because the numbers, particularly in the first decade of the web, looked awful. Eye tracking studies suggested that the average web pages were looked at for just seconds, raising the possibility that digital screens are a terrible place to get people's attention.[1] The web itself, many theorize, may be shrinking attention spans, reducing patience and leading to a loss of deeper thinking.[2] Maybe the only thing people wanted in the digital age when consumers were in charge, was eye candy stories, (Justin Beiber), quick hit items with no depth, or "list-icles" (7 things to know about Donald Trump's hair).

A decade later, it turns out that many of the early hypotheses were oversimplified. For one thing, the early data on digital engagement weren't measuring something inherent to digital screens.

They were monitoring the behavior of people on desktops, mostly at work. With the advent of touch screen mobile devices, we have learned that not only could people pay attention to a screen, they would even read books on their phones.[3]

But for journalism to understand and make use of the new world of metrics, it first must learn how to turn bad data into good.

Some publishing operations are beginning to try. They are attempting to look at the data and apply more journalism values to understanding it. Mic, for instance, is thinking about different metrics that combined equal trust. The audience tracking firm Chartbeat talks a good deal now about time spent in highly engaged activity (recent scrolling).[4] Upworthy tracks what it calls Attention Minutes, which involves both how many people look at a piece of content and for how long and how long a particular unique visitor may spend with any of its content over time.[5] Buzzfeed, one of the most sophisticated publishers when it comes to data, thinks a good deal about what content people are sharing, or what it calls "social lift." If people have decided to share something with their friends, Buzzfeed values it in a way that is measurable. The compay can also triangulate that with topic to identify communities of interest.[6]

What follows is one such effort in trying to create new, more useful analytics. It is based on work at the American Press Institute, where I am executive director, working with a talented digital team and in partnership with some 55 publications. The results may shock some people, but they are based on a unique data set of more than 400,000 stories from these publications and the patterns emerging about what people engage with.

The work also shows that any publication, even small ones, can become a master of rather than a victim of its data.

ONE APPROACH TO BETTER ANALYTICS

This project began when the CEO of a media company asked us a question. Could we help his editors make editorial choices based on something other than just their gut and possibly outmoded conventional wisdom? Is there a way to base editorial choices on something measurable so we know we are making the right ones?

There was some significant risk here. If a chief executive with little newsroom experience began to demand his editors adhere to the wrong set of metrics to justify their editorial choices, it could drive away all his most creative journalists. Metrics, poorly used, could become a Trojan horse for cutting costs and abandoning hard-to-cover journalistic topics. As the axiom goes: not everything you count matters and not everything that matters can be counted.

Yet if there were a way to do this right, it could help propel local publications with limited data skills substantially ahead and begin to connect them to audiences in their communities with whom they had little relevance previously.

Certainly many of the new all-digital companies that are attracting more of people's time are data

driven, place like Facebook and Buzzfeed. If local news operations, which often are the only monitors of local civic life, are going to survive, they need to understand their audiences and their data on a different level.[7-]

How could we build something that was a tool for editors rather than a noose?

The first step, we thought, was to find a way to understand each piece of content for its key journalistic characteristics. Some of these are fairly basic notions that editors already knew, even before the digital age, but never quantified. What is the story about? How much effort went into it? What is the story's relevance to the audience, or in other words, why would someone read it?

> If local news operations, which often are the only monitors of local civic life, are going to survive, they need to understand their audiences and their data on a different level.

Each of these basic tags involves taking the ideas that editors already have in their heads about the content they produce and converting them into structured data that they can look at later and measure to understand what they are covering and how, and also how audiences are engaging with it.

Journalists have long thought about stories according to how much effort went into them. Some stories are a paragraph long, "a brief" in newspaper speak or a "tell story" in television and radio. Other stories are basic daily news stories. Some are longer features. That spectrum of effort can be measured all the way up to the highest level of enterprise—a major exposé or a multi-part series.

Journalists also think about stories based on how they might help the audience. Some pieces are explainers. Others are just for the record to note some incremental development. Some pieces help people solve problems. Others are watchdog journalism.

In journalistic parlance, the "so what" or "significance" paragraph of a story often signals why someone should or would read it. In the nomenclature of Clayton Christensen, whose thinking about disruption and media is illuminating here, this characteristic is what job a particular piece of content is doing for its audience.
Over time we have tested and arrived at nine different types news stories. Conventional analytics do not provide publishers with any of that.

With their current analytics, frankly, most publications do not even know what topics they cover.

Publications have "taxonomies," in which each piece of content is tagged for all the subjects a story may touch on. These taxonomies, however, are designed to help people find stories later in a digital topic index and thus maximize the number of topics a story is linked to. Most stories are "tagged," accordingly, for five, seven, or perhaps nine different topics.

But these taxonomies do not help publications understand how much they cover certain topics or how they spend their time and resources. No story, for instance, is about seven different topics, though it might easily be tagged that way in a conventional taxonomy.

From a journalistic standpoint, most stories are about one or two, or at most three topics. Consider, for example, these stories and topics: city council met and talked about the new highway plan. There was a fire on Main Street. The new House speaker intends to approach the president's agenda differently from his predecessor.

So in our system, we began to have editors tag stories by what they were about—not in a way to help them be found later.

In addition to topic, story type, and relevance, as mentioned above, we built our system to let editors tag content for a menu of other characteristics that might matter to them.

They can identify what triggered a story in the first place. Is a piece reacting to breaking news, a planned news event or is this story the fruit of your own idea and initiative? What style of writing is used, traditional news style or more subjective and informal? Does the story have non-narrative elements, including audio and video? How much social promotion did it have? And more. Indeed, different publishers testing different ideas work with us to develop temporary tags that help them answer hypotheses they might have.

It's also possible to know the geographic focus of the story, by town, region, or even neighborhood.

We piloted our new system with the company that initially approached us, the Pioneer News group headquartered in Seattle. The editors and technical people at Pioneer, some skeptical and some intrigued, became our beta partners. Inside the app, we flowed whatever analytics a given publisher already had, and layered those with our new journalism tags.

This new layer of tagging suddenly converts conventional analytics into journalism analytics. A story is not just a piece of content with a certain number of page views. One can see how explanatory stories are doing and how that varies by topic. Or whether stories that help people save money and solve problems perform differently than stories that are covering breaking news.

Within a few months, the publications began to learn things about their output they didn't know before. What percentage of content they produced was devoted to what topic? How journalistically are they approaching each of their beats, and what are their audiences responding to and not responding to?

The Pioneer group also asked if we could twin this new structured data with audience research about each of their markets. It is one thing to know what content is working and not. But could we find out if there was more potential reader interest in the content that wasn't working?

With emails of subscribers and hopefully non-subscribers, conducting audience research, if done properly, is far easier than it once was. The key here is asking people about their lives—their passion, their worries and their community—not asking them about their view of the publication or

asking them, artificially, to play the role of city editor or TV executive producer. ('Would you like to see more stories about...' is not a useful or meaningful question.)

BETTER MEANS DIFFERENT, NOT MORE

Once a publisher has data about what content they are producing and what's working and not working journalistically, the next step is try to change what they are doing to do more work that readers value inside key coverage areas and spend less effort on what isn't working.

This has several implications. The first is not all topics are equally popular. Covering water quality is not going to be as popular as covering a championship sports team, and the responsibility of a publication is not to cover only popular topics. The tagging of content for its journalism character-istics allows a publisher to compare how engagement in a topic is changing over time. The goal, then, may be to increase engagement with government coverage by 20 percent, not to give up on government coverage because it will never be as popular as something else.

The second implication is that covering something more effectively does not necessarily mean cov-ering it more. Usually it means covering it differently. Building structured data about one's journal-istic characteristics empowers a publisher to know how much enterprise coverage they are doing, or how much of their coverage is explanatory versus simply for the record. A publisher can change the character of their coverage without covering it more. In some cases it may mean improving coverage by pulling back on things that are not useful.

SOLVING THE DATA GLUT

Before long, one of our partners, the *Dallas Morning News*, began to talk about something that became a second phase of the project. In their efforts to get better metrics, they began to drown in data. Rather than clarifying anything, the data glut encouraged arguments that were never settled about what metrics mattered. It led other people to ignore metrics altogether. Too many conflicting numbers just seemed like more noise.

The challenge of our program, they told us, was that yet another layer of metrics might only make the problem worse.

Once you look outside news, a solution isn't far away. Researchers routinely cope with large data sets by creating indexes. The government, for instance, measures the growth of the economy by blending scores of different numbers into the Gross Domestic Product, the price of many items into the Consumer Price Index, or the direction of the economy in the Index of Leading Economic Indicators.

We began testing a similar notion of indexing with conventional journalism metrics publishers already had, layered with our new journalism tags.

With indexing, one can balance the breadth of page views with the depth of reading times, and

blend in various social metrics as well. Loyal readers can be given more weight if a publisher wants. So can local users, without ignoring entirely those out of market.

Part of the value of indexing, indeed, is the argument that an individual publisher can have in designing it.

We encourage each publication to engage in a robust debate as it works through the index it wants. To begin to use data in journalism effectively, indeed, a publication needs to settle the question of what kind of engagement they value in advance.

That uncertainty over what engagement to value is one of the problems of bad metrics.

The index also doesn't make the individual numbers disappear. One can always look inside an index to see the components.

WHAT DO AUDIENCES ACTUALLY VALUE?

We have been working with this program for a little over two years. More than 55 publications are using the tagging app, which is constantly being refined and updated. The data set includes more than a quarter million stories or pieces of content. It is growing all the time, and some publishers keep coming up with new theories to test with structured data.

It is useful to note here that the tagging is done by the section editors at the partner publications. These are the people who have edited the stories for publication—not the individual reporter or producer who created them.

They also do the tagging outside of their regular work flow and inside an app we have built at API. We recommend they do it a day later. This has several advantages. People tag more quickly and more accurately if they do several stories at once, rather than doing it inside their content management system as they are posting stories. They will also believe the data if they created it themselves.

What have we found so far?

The answers vary by publication, but there are some findings that have tended to hold true across all of them, which I discuss below.

For the purposes of this analysis, I am using a generic Engagement Index that we developed at API. While it has more than a dozen metrics that go into it, the basic blend is 35 percent page views (which is a blend of loyal, local and overall), 40 percent time spent per article (again a blend of different types of users), and 25 percent a mix of different sharing metrics (sharing, social entry and more).

MAJOR ENTERPRISE PAYS

The single biggest change publishers can make in general is to produce more high-value major enterprise journalism.

Major enterprise stories score 48 percent better than others in our measure of overall engagement. And it holds true across various submetrics. Major enterprise stories, for instance, generate 83 percent more page views, 39 percent more time spent per article, and 103 percent more sharing activity.

The only problem is there isn't nearly as much major enterprise work being done as publishers probably think or want. Even according to the tagging that the editors did themselves, just one percent of all content is major enterprise.

> There isn't nearly as much major enterprise work being done as publishers probably think or want. … Just one percent of all content is major enterprise.

What doesn't help is the middle ground story—something in between a simple daily news story but less than major enterprise. These "light enterprise" stories make up about 13 percent of all content across our partners. They take more effort than a traditional piece. But by and large, audiences don't engage with them any more than they do basic news stories (though there are a couple beats that deviate from that rule).

One reason may be the nature of the web. In the old way of producing news, journalists were trained to add whatever new angles and dimensions they had gathered in their reporting and pile them into their stories each day. After all, the paper, or the newscast, vanished at the end of the news cycle. The next day, you produced a new version. You advanced stories inch by inch. In print and in newscasts, there was a clear signal to audiences if a story was a major piece of content, based on whether it was on the front page or the top of the newscast.

In digital form, the majority of traffic never sees a home page. Most people come into the side door, to a particular story, increasingly through social media. On mobile, there isn't really a home page at all, but sections, of which top stories may be just one.

Thus the data may reflect that in digital form, it may be more engaging to do a daily quickly and then, if a journalist wants to do a more substantial analytical story, to produce that separately. In a similar vein, it may make more sense to collect the small interesting angles and insights that might have been lumped into "today's story" into a separate piece. Particularly for those reading on a mobile device.

In other words, journalists now need to send signals about significance and value to readers differently. Rather than by placement on a home page or a newscast, or by making a story simply a lot longer, the story needs to signal to the audience how it will help them through the headlines, graphics, and other elements. Simply piling more into a single story each day is obsolete. (The digital *New York Times* I think has embraced this notion of breaking what once would have been a

single long story into clearer smaller pieces divided by their specific point.)

STORIES TRIGGERED BY NEWSROOM INITIATIVE PAY OFF

Initiative also drives engagement. Initiative stories, using our analytical approach, are those that the newsroom dreams up. They answer a question journalists think audiences might be wondering, explore some new unexplored angle to a running story, or were triggered by something the journalist uncovered. It could be a new angle on a running news event, an analytical piece, a non-narrative graphic work, a sidebar, or feature. The original event being considered might be external—a speech by the mayor, a shooting, new crime statistics. But the story itself was a new idea born from that.

This distinguishes an initiative story from one that was produced in reaction to an event, either a breaking news incident or a planned event such as a press conference.

An initiative story, in other words, is triggered by an idea (though it may be an idea about an event that itself started elsewhere).

By their nature, then, initiative stories are unique. They are stories audiences cannot find everywhere. They are also proactive rather than reactive. (One caveat in our system is against imitation. A trend story, for instance, that was first seen elsewhere and then copied by a news organization might be idea driven, but it wasn't your idea.)

Stories triggered by newsroom initiative score 30 percent better than others in our measure of overall engagement index. They generate 62 percent more page views, 11 percent more time spent per article and 101 percent more shares.

Initiative, as we measure it, is different than enterprise. Enterprise tracks how much time and resources went into producing the story—on a scale ranging from a brief to a major enterprise piece. Major enterprise pieces are ones that people spent several days on, or longer. Initiative measures simply where the impetus or trigger for the story came from.

Put another way, all major enterprise stories are usually initiative stories, unless it is really replicating what someone else did first. But not all initiative stories are major enterprise. They may take much less time. They may be something produced in a few hours.

The problem, once again, is there aren't enough initiative stories either. While there are five times as many as there are major enterprise stories, still only 5 percent of the more than quarter million stories tagged were initiative stories. We have some publications where that number is closer to 15 percent. But more initiative helps.

PEOPLE LIKE LONG STORIES

This more granular approach defining and measuring engagement also reveals something about attention spans and story length. People consuming journalism digitally like long stories. Even on their mobile phones. The conventional wisdom that writing for the web needs to be short and fast,

and that writing for mobile even more abbreviated, simply is not true.

> People like quality and depth, and will stick with a well-told story.

"Long form" stories, those averaging 1,200 words, drive 23 percent more engagement in our index, and lift various metrics such as page views (up 11 percent), sharing (by 45 percent) and reading time (by 36 percent). People like quality and depth, and will stick with a well-told story.

TRADITIONAL NEWS STYLE WORKS FOR SOME TOPICS, OPINION WORKS FOR OTHERS

Another myth is that a story has to have an opinion these days for people to read it. Overall, stories that were written in a traditional news voice, whether they were straight news stories or feature pieces but still written a conventional reportorial voice, score 21 percent higher on our basic engagement index. They generated 52 percent more views, and 71 percent more shares than stories written with a more subjective voice.

But this finding about voice doesn't hold true for all topics. A more analytic, more personal and subjective style or presentation works far better for sports, food and dining and, at some publications but not all, for lifestyle generally.

These are content areas where journalists can be tastemakers and where people may be looking to publications for cues, argument and opinion more than just facts about what happened.

THE POWER OF PHOTOS, AUDIO, AND VIDEO

The data set also provides powerful evidence of the influence of non-narrative content online.

The numbers themselves are fairly compelling. Stories presented with a photo score 19 percent higher in engagement than stories without photos. Stories with multiple photos score 43 percent higher.

And here, too, topic makes a difference. Stories about government that have a photo with them score 75 percent higher than those that do not (more on this later). That is even higher than sports stories, which get a 43 percent boost. Food and dining stories, interestingly, do not benefit.

It's also not just photos. Add audio and video and you can do the same. In the data, the presence of either audio or video correlates with 36 percent more overall engagement—in particular 81 percent more views and 84 percent more shares.

Video and audio also create new ad inventory and cross platform ad opportunities.

These results are topline generalizations we have seen across partners. But one of the things the data tell us is you shouldn't cover every topic the same way. It's worth taking a minute to walk

through some general findings by different beat or topic areas.

COVERING DIFFERENT BEATS DIFFERENTLY

GOVERNMENT

No topic is more central to the basic mission of journalism than covering government. And we see a lot of it in our partners, most of whom are general interest newspapers. In that universe, government is the second-most covered topic, after sports. Fully 16 percent of all stories tagged were government stories.

> One of the things the data tell us is you shouldn't cover every topic in the same way.

The data also belies the idea that government is an anathema for readership. Across this universe of nearly 400,000 pieces of content, stories about government perform about average.

But there are things publications can do in covering government that are more successful than others.

In government, initiative—again, this means telling stories no one else has—is the single biggest driver of engagement. Stories that journalists conceived, that answered a new question or explored a new angle—generated 52 percent more engagement than government stories that were produced in reaction to an event.

Major enterprise stories were almost as effective. Going deep generated 46 percent more engagement.

Watchdog stories also helped—performing 25 percent higher in engagement—than government coverage in general, though the data made clear that you don't have to have an exposé of wrongdoing to engage people here. This 25 percent boost, by the way, is the same boost watchdog coverage gets on most topics.

As noted above, one interesting finding is the power that photographs, even a single photograph in a story, has on audiences engaging with government news. Government stories that had a photo, again, scored 75 percent higher in engagement than government stories that did not.

Government is not a topic about which you might imagine highly dramatic photos, a collision at second base captured at just the right moment for instance. Government related photos might be a person at a lectern or a protester with a sign.

Our inference is that even these photos do something more subtle—since their power is greater on government content than most others. They put a human face to what otherwise might be abstract issues.

Publishers can also do some things in covering government that will likely depress engagement, according to the data. One of those is produce too many stories that are about incremental chang-

es, particularly about government process. These "for the record" stories perform less well than other government stories. And there are far too many of them. In all, 48 percent of the government stories tagged, more than 30,000 out of 64,000, were simple for the record stories.

BUSINESS AND ECONOMY

What has proven effective in coverage of business and economics, the data suggest, is a little different than for government. Initiative still helps; it boosted overall engagement by a third (33 percent).

But there was a practical, news-you-can use dimension to business and economic coverage that we don't see in most other topics. Across all topics, stories that help people solve problems, save them money, or help them figure how to spend their time, tend to perform a little worse than average. In business and economic stories they help, including locally helping people find out about job fairs and openings. These kinds of "things to do" stories boosted engagement 20 percent.

Watchdog coverage also helped, but a good deal less than in general. They scored 15 percent higher in engagement than the average business story, while they score 27 percent higher in engagement in general.

Stories that were tagged as also being ones that editors thought would be heavily shared, which the system calls "conversation starters," also did well in business, boosting engagement by a similar 19 percent. Major enterprise also matters, a 25 percent boost. Like with other topics, there isn't much major enterprise being produced; it accounts for about one percent of all business coverage.

CRIME AND PUBLIC SAFETY

If government is one staple of local journalism, crime is another. Overall, across the more than 50 papers studied, crime ranks with food and dining as the topic audiences engaged with most. But in a digital world, what works best is somewhat different than it may have been previously.

Online, crime briefs—the classic police blotter of small incidents—do not perform well, particularly by an engagement index that takes into account sharing and depth of interest rather than just number of eyeballs. Crime briefs, indeed, perform 36 percent less well than crime coverage in general. But it's not just time. Crime briefs online also generate fewer page views than other crime coverage. The police blotter, in other words, does not translate well to digital by any metric.

> The police blotter does not translate well to digital by any metric.

What helps audiences engage with crime and public safety news, as we found in so many topics, is covering a more proactive and enterprising approach that helps put crime and public safety in context. Crime stories triggered by newsroom initiative perform 54 percent better than those that were not. Crime stories that involved major enterprise over many days perform 34 percent better than other stories. Crime stories that are explanatory in nature perform 33 percent better. And sto-

ries tagged as watchdog, in the sense that they exposed wrongdoing or some problem, performed 26 percent better.

Interestingly, crime is one of the few topics where light enterprise, turning a daily into something just a bit more or something of a light feature, does help. These stories generate 23 percent more engagement than crime on average.

Crime blogs are a useful form (they perform 64 percent better than the average crime story), perhaps because they allow some analysis in covering crime—without being simply opinion. And photos help, but far less than in some other topics (a boost of 17 percent).

SPORTS

Sports present an interesting challenge in journalism today for several reasons. First, it was the original disrupted topic. Sports events began to be broadcast on radio in the 1920s. They've been on television since the 1950s.

The folks in the press box, in other words, lost their monopoly over telling people what happened in the game long ago. Sports journalists have been trying to figure out what added value they can bring to news that people already know for more than three generations.

The data make clear that there are certain characteristics online that make this content more engaging. Not everyone is doing them.

To a large degree, straight news game stories flop online. Writing with a strong subjective or analytic voice and demonstrating initiative and enterprise pays off. Stories with a strong analytic voice perform 34 percent better than other sports stories. Major enterprise sports pieces do 90 percent better. Initiative helps some, but less than elsewhere. It provides a boost of 23 percent.

But stories that mark things just for the record perform eight percent less well than other sports. Dailies score six percent below average. Briefs perform 35 percent less well than other sports content. But we found generally too much of what doesn't help. Fully half the sports stories we analyzed were dailies, and half were for the record. Sports also benefits less than other topics from photos, interestingly (a boost of 26 percent). That may be because so much has already been seen.

Sports is the topic that gets the most coverage among our local news partners. It's not the most read topic overall, but it attracts a highly loyal audience. In fact, it's the third most popular topic among the most loyal readers (and mobile readers), behind crime and food and dining. Nearly 60 percent of all the views to sports content come from those who visit the site at least three times a month. And, about 45 percent of all views are done on mobile devices.

FOOD AND DINING

Finally, it's worth making a point about food and dining journalism.

Just as food channels and celebrity chefs have proven a magnet on television, food and dining is an

important category for local journalism, tied in most of the cities we have examined with crime for the most popular topic.

But food journalism is more than recipes. In fact, they may be the least of it.

Restaurant openings, reviews and, interestingly, also health inspections do really well, along with simple things like school menus and school lunches.

> Food and dining stories have a long shelf life. They are twice as likely to be viewed a week after publication than content overall.

Food and dining works online, in part, because stories also have a long shelf life, something that wasn't relevant in the days when journalism was limited to print, radio, or television and vanished either instantly or the next day. Food and dining stories are twice as likely to be viewed a week after publication than content overall.

We found one other challenge to the conventional thinking about food journalism in the data. Photos don't make a huge difference in food and dining coverage, which contradicts the days when newspapers had elaborate photo studios for their dining sections. You get about the same amount of engagement whether you include a photo or not.

A NEW MODEL OF PUBLISHING: FRANCHISE TOPICS

Understanding data at this level encourages publishers to think about a different model for building their brand. The traditional model to news publishing, particularly for local publishers, might be thought of as the General Store. You should read our local newspaper, or watch our local TV newscast, because we have a little of everything, in the same way that the general store in a western town was the place in town for most of your store bought goods. If they didn't have it, you probably didn't get it. Not right away, at least.

The model worked in a more monopolistic era, when one needed a license or printing press to be a publisher.

The web, however, encourages specialization. Online, a better source on any subject is just a click away.

On top of that, at a time when most newsrooms are shrinking, trying to do all things equally probably translates into many newsrooms getting a worse at everything.

It makes more sense, we believe, for publishers to build their brand around a core of franchise topics that offer people a reason to come to them in the first place. This doesn't imply that newspapers or TV news outlets become niche publishers. But there has to be a reason for people to come to them besides convenience. The web itself has supplanted that. And the claim by most local publishers that they have more quality doesn't stand up to scrutiny on all topics.

A more modern and more sustainable approach is for new publishers to determine what topics they can excel at and have those become magnet topics. Audiences may discover other things when they arrive. But simply being the source in town that is on your side, or that you can trust, or that is "the local paper," is no longer sufficient.

The API tagging system encourages this notion of franchise topics because it allows publishers to see deeply into what works with different beats.

MAKING CHANGES IN COVERAGE

It takes more than a year for a publication to work the process of learning the app, generating enough content to begin to learn meaningful things, make changes based on this data, and then assess whether those changes are working. About a dozen publications, so far, have gone through the cycle long enough that they have begun to make changes they can evaluate.

The results suggest that, to begin with, journalists can make their content more engaging by becoming more quantitatively aware of their journalistic characteristics.

The findings also suggest that this notion of franchise topics can actually lift engagement with all topics.

One Florida partner, for instance, started this program in July 2014. They had already embraced the notion of franchise topics before we began working with them. The program has helped them improve engagement not just with the franchise topics they already had, but across the board.

Comparing the first four months of the program versus the same four months a year later, the paper has seen an 121 percent increase in page views of their coverage of the environment—one of their four franchises—and 200 percent growth in shares. On their franchise focus on the economic future of the region, they saw 73 percent increase in views and 150 percent in shares. And on lifestyle, a third franchise, a growth of 33 percent in views and 115 percent in shares.

But the benefit has also been seen across the board. All content is up 40 percent in views, 77 percent in time spent and 115 percent in shares.

At another publisher in Pennsylvania, the program told them that crime was a key issue in the community, but their coverage of it needed more enterprise, fewer columns and briefs, and more explanatory coverage and initiative. A year later, they have seen a 231 percent increase in pages views of their crime coverage and 251 percent increase in shares.

Another partner paper in the Northwest worked to slowly roll out franchise topics over the course of a year, starting in January 2014. Their franchise topics include business, outdoors, family, and high school sports. In an eight month span, they saw a growth of 150 percent in views across their franchise topics and 50 percent in shares. But they saw a 287 percent increase in page views across all their content, and a 51 percent increase in shares.

Another partner used to cover business largely through columns. After looking at all their data, they decided to redesign their coverage to focus more on living in the local economy—covering small business and providing local economic indicators. Since January of last year, this has boosted engagement of their business and economic coverage with 65 percent more views and 124 percent more shares.

> Journalists can make their content more engaging by becoming more quantitatively aware of their journalistic characteristics.

Indeed, though not on all topics, on average across all the partners that have reached the point of being able to make and evaluate changes based on the data, the program has increased engagement.

The data, and the comparisons of engagement within beats, offers one other lesson. Some topics are always going to be inherently more popular than others. As noted above, by breaking down coverage by topic, publishers can begin to avoid the trap of click bait—trying to make every story a breakout hit. The key becomes a matter of trying to figure out, instead, how to boost the coverage of whatever you deem important against its prior performance.

This system is not the only approach one could take to reinventing metrics or trying to understand coverage. The point of the findings is something more basic than the idea that we have developed *the* answer at API.

A generation into the digital age, journalism has suffered from trying to manage metrics that were designed for something else and on their own are unhelpful to making journalism better. That, however, need not be the case. By creating new structured data from their content, publishers have the capacity to create new and better metrics than the ones they have inherited. But they need to take this step before it is too late.

The audience is now in charge of the future of news. The publishers that thrive in the future will be those that understand this audience at a deeper level.

ENDNOTES

1. There are various studies on short reading times, from early eye-tracking studies conducted by the Poynter Institute to data by Chartbeat suggesting that half of all readers bail out of stories before finishing 100 words. http://www.slate.com/articles/technology/technology/2013/06/how_people_read_online_why_you_won_t_finish_this_article.html

2. http://www.elon.edu/e-web/predictions/expertsurveys/2012survey/default.xhtml

3. The Pew Project for Excellence in Journalism, "The Tablet Revolution," October 25, 2011 http://www.journalism.org/2011/10/25/tablet/ and "Future of Mobile News," October 1, 2012 http://www.journalism.org/2012/10/01/indepth-news-reading-tablet/

4. The author has heard Chartbeat CEO Tony Haile present on this issue and there are several articles describing the same concept. The following is one of the more detailed examples: http://blog.hubspot.com/marketing/chartbeat-website-engagement-data-nj

5. http://blog.upworthy.com/post/75795679502/what-uniques-and-pageviews-leave-out-and-why

6. The author has heard Buzzfeed founder Jonah Peretti and others Buzzfeed executives discuss the concept of social lift at different events. The following Buzzfeed document also explains the concept: http://s3-ak.buzzfed.com/static/downloads/2012_Social_Lift_Benchmarks.pdf

7. Tom Rosenstiel, "News As Collaborative Intelligence: Correcting Myths About News in the Digital Age."

8. http://www.brookings.edu/research/papers/2015/06/30-collaborative-intelligence-myths-digital-age-rosenstiel