KrehbielLetter

Dear Friend,

Welcome to the inaugural edition of *The Krehbiel Letter*, in which I intend to provide **useful thoughts on publishing technology**, with a focus on Customer Data. Please pass this along to anyone who might find it useful, and if you would like to be added to (or removed from) the distribution list, please contact me at greg@krehbielgroup.com.

Best,

Greg Krehbiel

The myth of the single customer record

Sellers of Customer Data Platforms (CDPs) promise their software will gather data from various applications, and assemble it into a single-source-of-truth "golden record" for each customer.

It's a lovely vision, **but rarely achieved**. *And that's perfectly okay.* The trick is to press on anyway.

Let's use this common CDP use case to illustrate the complexity: **Identifying customers among the hoards of anonymous visitors to your website**.

It's a challenge. Anonymity was central to the internet's design. And while there are lots of ways to identify anonymous website visitors, they all have their limitations.

Let's imagine a fellow named Robert Williams, a swing dance aficionado, who interacts with Ella, the publisher of (the fictitious, I believe) *Ella's Swing Dance Magazine*.

Robert meets Ella on his commute to work, and she tells him he ought to read her magazine. On his lunch break, Robert searches for the magazine website on the desktop he uses at the office. When Robert's web browser makes a request to *Ella's Swing Dance Magazine* website, Ella's CDP puts a cookie on that device and creates a user profile. The profile includes the following information:

Profile 1

IP address: 25.23.108.5
User-Agent: Mozilla/5.0 (Linux NT 10.0)
Referrer: https://www.google.com

The record might also include what pages were visited, and what type of content the visitor seems to prefer. **The visitor is still anonymous to Ella's CDP**. The profile is one of the millions of unknown visitors.

When Robert gets home that evening, he types the URL of Ella's website into his iPad. Her CDP dutifully puts a cookie on that device and creates a new profile. But on this visit, Robert decides to sign up for Ella's free e-newsletter with one of his junk email addresses. The CDP captures the email address from the form submission and creates a second profile, which has more information than the first.

Profile 2

IP address: 32.12.100.21
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_6)
Referrer: [blank]
Email: bob2387@hotmail.com
Name: Bob Williams

Nothing in this second record enables Ella's CDP to conclude the records are tied to the same individual. The records were created on different devices at different times, and share no information identifying Robert.

Two weeks later, Robert and Ella are jitterbugging at Mobtown Ballroom in Baltimore. Ella has a few copies of her magazine, and Robert takes one home. He signs up for a print subscription using one of the blow-in cards. Ella's fulfillment service dutifully records this new subscriber data, which is then imported into the CDP, creating **Robert's third profile** with still more information:

Profile 3

Name: Robert Williams
Address: 123 Main Street
City: Bowie
State: Maryland
Zip: 20715
Phone: (301) 555-1212
Email: me@robertwilliams.com

This profile has valuable information, including a new email address. But **this profile has no data from online activity**, so it doesn't help with online ad targeting or customer journey data.

Robert now has three profiles in Ella's CDP. There's no way to merge any of them. We know they're all Robert. The CDP doesn't.

Fortunately, Ella's magazine has the good sense to include some special online content for print subscribers as a way to **link offline and online behavior**. A QR code printed in the magazine allows Robert to view a video on the website about the Travelling Charleston. Robert scans the QR code with his iPad. That takes him to the website, where the CDP recognizes the cookie it put on that device earlier.

Bingo! **Now Ella's CDP can merge the iPad profile** (#2) **with the subscription information** (#3). Several good things happen as a result:

- Robert's three profiles have been consolidated into two
- Robert has become a known user in Ella's CDP
- Ella's CDP knows that Robert uses two different email addresses
- Robert's subscription information (offline behavior) and the profile created when he accessed Ella's site from his iPad (online behavior) are now linked.

The record created from Robert's desktop remains anonymous.

Note that, in this scenario, Ella's CDP has been configured to accept **multiple emails in a customer's profile**. Some companies designate the email address as a unique field – allowing only one per profile. In that case, the records would not merge, and Robert's subscription information would remain in its own profile, not connected to any online activity.

Will Ella's CDP ever be able to attach Robert's work computer to his online profile? Maybe. For example, if Robert opens one of Ella's e-newsletters on his work computer, the CDP might (depending on how strict it is about such things) recognize that as Robert and merge the profiles.

Identifying individuals from their online and offline behaviors and creating single records may seem complicated, but it's quite a bit less confusing than what happens in real life. Consider the complexity added when Robert's smartphone and home desktop are added to the equation.

Merging records: deterministic vs. probabilistic method. Which is right for you?

The "golden record" the CDP salesman likes to highlight assumes all these different sources of information can be merged, but **they need to have a field in the record to merge *on*.** What's that going to be?

Most companies opt for an email address as the best piece of personally identifiable information on which to merge records. But as we've seen, people have multiple email addresses. They also change over time.

If you stick with a strictly *deterministic* matching method, you'll need to match a unique field (like an email address or a social media account) across multiple profiles to create your "golden record," and you'll inevitably leave some information behind.

There are other options. Some CDPs use *probabilistic* methods to merge profiles. That method enables you to match records that might otherwise remain distinct. But you risk incorrectly merging profiles and creating a customer experience headache.

You can't create a single record for each customer that covers all the chaos and weird realities of how people behave. What you can do, and what you must do, is decide where that matters.

There are use cases where improperly merged profiles yield very bad customer experience outcomes. Stick with deterministic matching in those cases, even though you're going to lose some of the data on interactions with that customer. You'll have multiple profiles for some individuals, many of which will remain "unknown."

Other use cases are far more forgiving. If you want to create a segment of people who share a particular interest, you don't need to get down to the individual. In these cases, probabilistic methods are sufficient.

In any event, recognize that "golden records" are a nice idea, but you'll never actually get there.   ¶

"Don't believe science until it becomes engineering."
No, not really, but there's a kernel of truth there. You can use the QR code on the left to watch a short video I posted on LinkedIn about the importance of operations. It's all well and good to have ideas, intentions, and plans, but until they're reduced to actionable assets, they're not much use. By "actionable assets" I mean checklists, forms, procedures, workflow requirements, and such. Give the video a look, and please leave a comment.

Maximizing revenue from drive-by visitors

Many customer engagement efforts start with trying to get website visitors to sign up for a free email newsletter. Then they focus on how to get more engagement from the newsletter, or how to sell something.

That's great, but I believe it fails to start at the beginning, and because of this it overlooks the vast majority of a company's website traffic.

Most visitors to most websites are drive-bys. They view one page and they're gone. They're not going to sign up for your email newsletter. They're not even going to read a second article. This means that the "sign up for my free email newsletter" engagement strategy is missing the vast majority of your visitors.

Q: How do you engage these drive-bys?

A: You don't.

Your best strategy for these visitors is to forget about engagement and maximize your ad revenue. Use as much space as you can to get that precious 0.1 cents (which adds up when you have lots of traffic).

Stop trying to get them to do what they're not going to do.

Focus on the people who have stayed for a second pageview (or have come back within a relatively small window of time). For these people, dial back the ads a little and focus on content recommendations. Your customer journey for this group is to encourage them to read a couple more articles.

After someone has read a few articles, they might be a prospect for your email newsletter. But probably not before.

If you have the technology to count pageviews, you can use that data in your copy. E.g., "Hi friend. It's obvious you like our content. You've viewed five pages in the past week. We're happy to send it to you for free. Sign up now."

But what if we don't have the tech?

When a web browser requests a page from your server, it sends something called an http header, which you can easily read with php or whatever programming language you use on your site. That header includes something called a referrer, which is the site the visitor just came from.

To roughly approximate the customer journey suggestions above, have your IT guy write a simple script to replace the email newsletter sign-up form with an ad when the referrer is some other website, and only show it when the referrer is your own website. That is, when the visitor is on a second (or later) pageview.

That's not an ideal solution, but it does allow you to maximize your ad exposure for drive-by visitors. And that's not nothing. ¶

About The Krehbiel Group: I solve technology, strategy, operations, and process problems for publishers. My expertise includes Customer Data Platforms, acquisition and retention, e-commerce, RFPs, fulfillment, and project management. But if you have some other sort of tech problem, I can probably help. Give me a call, or send me an email. 240-687-1230. greg@krehbielgroup.com

Also, please take a look at the articles on my website: https://krehbielgroup.com