Media Impact Measurement System

Data Repository

Technical Overview

Version 1.0

January 2016

Dana Chinn, USC Annenberg

Mike Lee, USC Viterbi

Jonathan Weber, LunaMetrics

USC Viterbi School of Engineering

USC Annenberg School for Communication and Journalism

LunaMetrics

Contents

Purpose

Media organizations face specialized challenges in measuring their **audiences on and off the web** and their **interaction with and impact from media content**. Best practices in e-commerce digital analytics and standard analytical tools such as Google Analytics are often generalized and simplified to work across dissimilar industries, or specialized for industries other than media. The media industry needs a foundation of measurement and analysis derived from e-commerce best practices but which is based on media-centric impact and sustainability objectives.

This document provides a technical overview of the first phase of the Data Repository component of the Media Impact Project's Media Impact Measurement System. It outlines:

- Audience metrics relevant to media organization websites, e-mail newsletters, events and donors, and the structure of the Data Repository MIP is building to store the longitudinal data needed to track a media organization's impact over time.

- The customized, media-centric website metrics MIP is gathering that aren't available through the standard installation of Google Analytics.

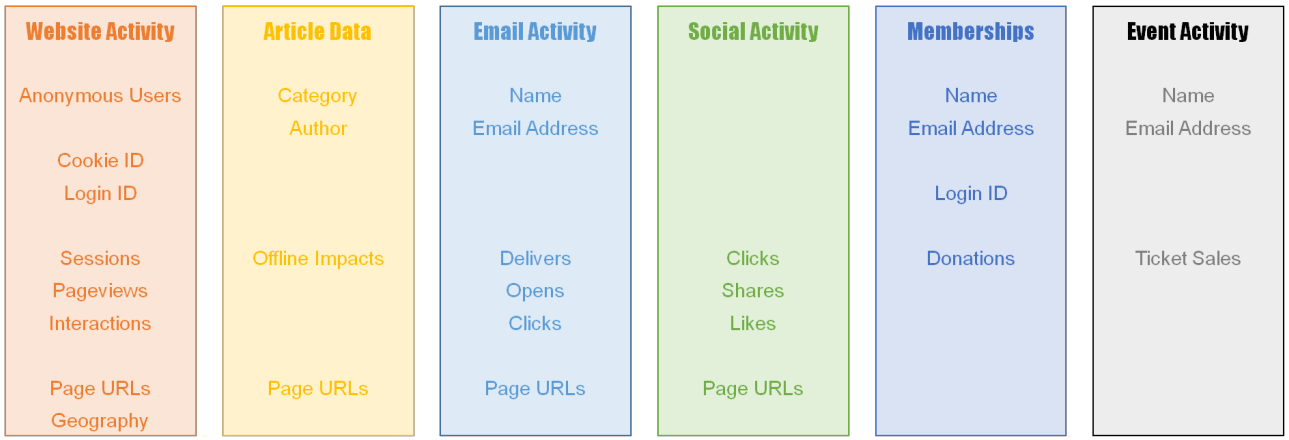
- The methodologies and tools MIP is developing to connect the data from disparate tools to provide a holistic view of how audiences interact with and are impacted by a media organization’s content.

The Texas Tribune and Southern California Public Radio/KPCC are the two pilot media organizations for the MIP Data Repository. This document is for media organization CTOs and information technology professionals who need to assess whether to build their own media metrics Data Repository.

The Necessity of a Data Repository

Media organizations often feel that they are awash in data—website and social media analytics, supporter or customer relationships, and more. Don't they already have everything they need to make assessments of impact?

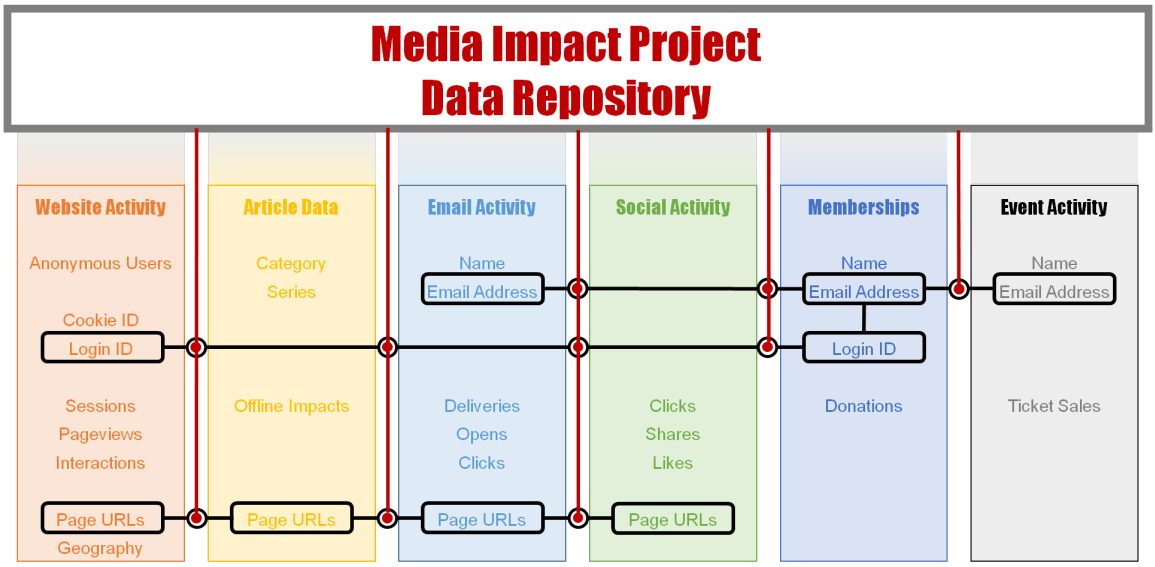
Despite all these data, it remains challenging to construct a full picture of engagement with or impact of media. Although there is a large volume of data, these data sources are almost entirely quantitative and composed of metrics built to inform marketing and advertising pursuits, the main drivers of the development of these tools. Additionally, these sources are largely siloed, disconnected from the data in other tools. Each tool treats audiences as an aggregate block to be sliced, but there is a lack of connection to understand how these audiences correspond between different data sets.



The Media Impact Project Data Repository brings together the data across media products to enable assessments of impact and deep analysis that are not feasible with typical data silos. The repository:

- Joins together disparate data sources to provide content-centric and audience-centric views of data to understand the impact of stories on individuals.

- Gathers and codifies quantitative and qualitative data about impacts and categorical metadata about content to provide fuller context and a broader picture of media impact.



MIP Data Repository Structure

Architectural Overview

The Data Repository fits into MIP's Media Impact Measurement System (MIMS), a broader architectural system that ingests, stores, enriches, visualizes and reports on data to provide organization-specific and industry-wide insights into the impact of media.

The MIP Data Repository is designed to accommodate many types of data into a set of standardized processes for cleaning and joining that data for meaningful analysis for media sites. MIP is developing standardized, media-centric ETL (extract, transform, and load) processes for data from any type of schema so that:

- Data from standardized, e-commerce-oriented tools can be used for more insightful media impact analyses.
- Organizations can change vendors or data sources (e.g., from MailChimp to Constant Contact; from spreadsheets to Salesforce) without sacrificing data integrity for longitudinal tracking.
- The Media Impact Project can gather and analyze data from multiple media organizations to compare and contrast impact across different types of organizations and the media industry overall. A longer-term goal for MIP is to make the Data Repository available for researchers once there are enough organizations to provide anonymity for any published results.

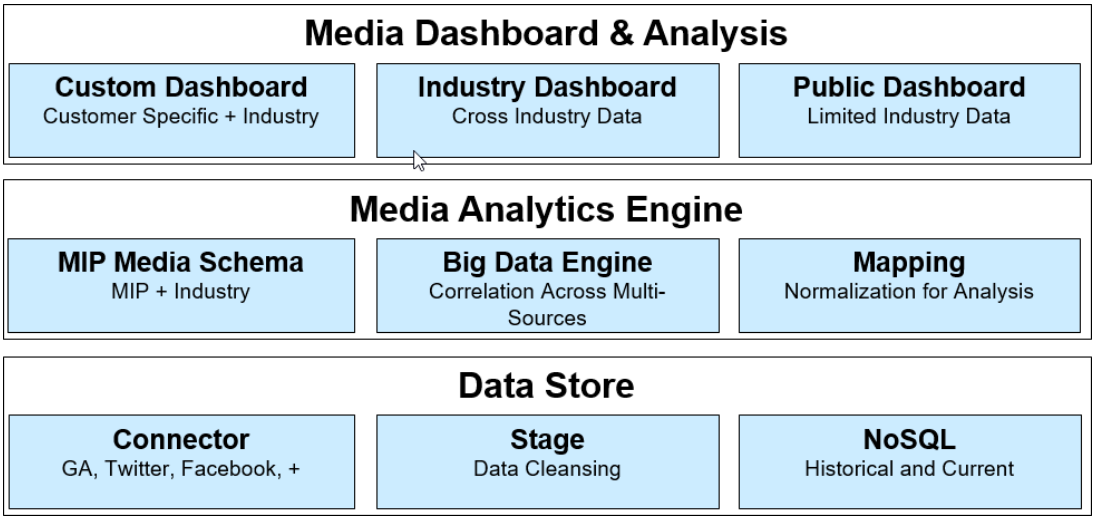


Figure 1. MIMS Data Repository Architectural Overview Diagram

Data Types

The data to be incorporated in the Data Repository include a wide variety of data types that measure the behavior, engagement, and attitude of the audience as well as the nature of the media content itself. Each of these data types can come from one or more sources (Connectors) that are brought together in the repository. (The list of Connectors in the Data Repository will be posted and updated at mediaimpactproject.org/connectors.

1. Website Activity

Website user activity from MIP's media-centric implementation of Google Analytics includes several sub-categories of data, including the following:

- Detailed interaction with content, including views of articles and other pages, scrolling activity, time spent reading articles, and interactions with multimedia such as video, audio, or image galleries.
- Responses to content, such as clicks on options to post about articles on social media, emailing or printing an article, leaving comments, or clicking on links to related articles.
- Information about the users themselves and their repeat use of the website over time. Users are distinguished through a persistent cookie and potentially identified by logging in, signing up for a newsletter using an email address, or making a donation or subscription via the website.

2. Article Data

Website data (described above) includes information about the interactions with articles on the website, as well as some limited metadata on articles beyond the URL (such as title and author, typically). However, additional data can be gathered about the content and its impact beyond website activity:

- Additional metadata about the content to provide context, such as taxonomic classifications, articles in a series, and other descriptive information.
- Data about offline impacts that cannot be measured in website activity.

Sources of data for these types of information can include content management systems, RSS feeds, or tools that enable gathering additional data such as NewsLynx or the Center for Investigative Reporting's Impact Tracker.

3. Email Activity

Media organizations use email to communicate with their audiences. This can include email summaries or alerts about news topics, solicitations for donation, and other communications.

Organizations use a variety of tools for these email communications, including MailChimp (Texas Tribune), Oracle Eloqua (KPCC), and many others. Typically such tools include the following kinds of data:

- Emails sent to a list of email addresses.
- Personal information linked with email addresses, such as name, company, postal code, or other user data.
- Activity data related to the emails sent, such as when the emails were open and links were clicked.

4. Social Activity

Online content for media sites takes a life beyond its publication on a website through social media platforms. This can include:

- Posts by the organization about its own content on social media ("owned social media"), as well as data about of the organization's reach and interactions with its audience on these social platforms (shares, comments, reposts, etc.).
- Posts by others of the organization's content ("earned social media") that arise organically, including mentions by other media organizations or influencers and content that spreads virally.

Social media activity is also connected to website activity in several ways:

- Likes, shares, and other social actions that originate from a media organization's website (e.g, the user clicks the "Share this article" button).
- Website users who arrive by following a link they found on a social media site.

The details of data available and the exact formats and labels vary from platform to platform. For example, Facebook has likes, shares, and comments; Twitter has favorites, retweets, and replies. The MIP repository has the flexibility to develop connectors for a variety of social networks and incorporate these data sets into its database.

5. Memberships, Subscriptions or Donations

Memberships or donations (for nonprofit media) or paid subscriptions (for for-profit media) can be a major funding source for media organizations. Typically organizations use a customer relationship management (CRM) tool such as SalesForce (Texas Tribune) or Roundcause (KPCC) to track and solicit current, potential, or lapsed members/subscribers.

These systems may include data such as the following:

- Personal information such as name, company, postal code, or other user data.
- Membership/subscription history.
- Contact activity over time for solicitations or renewals (by email, phone, direct mail, etc.).

6. Event Activity

Organizations may hold events ranging from live speakers to donor fundraising to online streaming, generating additional audience data that is often separated in a registration or ticketing system, such as EventBrite (Texas Tribune) or TicketLeap (KPCC). Much like member or subscriber data in a CRM, such tools contain identifying information as well as additional data types including registration and attendance history.

Data Repository Stages

The Data Repository handles data in three stages:

1. Collects data from Connectors
2. Cleans and processes the data
3. Stores the data for use in reporting & analysis

The remaining sections of this document look at each of these stages in detail.

Data Collection

There are three modes of data collection employed for the Data Repository:

- Real-time streaming of website data using Google Analytics
- Automated batch import of data from Connectors using APIs
- Manually exported batch import of data from Connectors

With each data type, MIP will assess whether to develop an API or to ingest the data manually. Developing an API is a long-term resource commitment given the frequency in which vendors change both in the data they collect and the way they make the data available. MIP will publish its schemas and APIs after they're developed and tested.

Website Data via Google Analytics

Google Analytics (GA) is a common, free website analytics tool. It is in widespread use on media websites and across many other industries.

GA does have limitations, which we seek to overcome by collecting this website data into the Data Repository. The advantages of the repository include the following:

- Overcome limitations of reporting in the GA interface, such as sampling for large volumes of data and providing individual user data.
- Join website with other data in the repository for a more complete view of interactions with the audience.
- Customize the metrics for reporting and analysis to focus on measurements most relevant to media.

GA does provide a robust data collection scheme that can be used to collect and validate website data in the Data Repository. The following sections examine the challenges and the process for collecting this data in the repository.

Avoiding Sampling and Aggregation

Google Analytics has many pre-aggregated reports available. It also provides tools for exploring and analyzing data such as custom reports and segments, but for these customized presentations of data, the results are not pre-aggregated. As a result, GA imposes limitations on the amount of data that is

used to calculate these results. For example, if the data set includes more than 500,000 sessions, GA employs **sampling** to calculate custom reports, segments, and other custom requests. It takes a sample of the data (fewer than 500,000 sessions), calculates the results for that sample, and then scales the results to the size of the original data set. As a result, metrics computed in such a way are only estimations. (More information about the conditions under which sampling applies can be found in the GA documentation.[1])

Even when sampling does not apply, the reports in Google Analytics are largely **aggregated**, meaning that they show totals and averages for a variety of metrics and can be sliced by a number of categories, but do not allow an easy view or export of behavior of individual users and all of their activity. For example, pages per session is an average that masks the distribution of the number of sessions that involved one page vs. those that had 20.

One solution to these challenges is to subscribe to Google Analytics Premium, a paid offering of Google Analytics that raises the sampling limits much higher and also allows channels to export the session-level data for analysis. However, GA Premium is relatively expensive, starting at $150,000 annually.

To address these challenges, the Data Repository collects a duplicate of all data sent to GA by the website (see the section Parallel Data Collection below). Since the repository collects the raw data, it can compute reports for any data set desired, without being subject to the limitations of sampling and aggregation.

Data Validation

The Data Repository data undergoes an ongoing process of verification to ensure that its data processing tasks are executed correctly, comparing the counts of records provided in Connectors, ingested in the repository, and provided in exports.

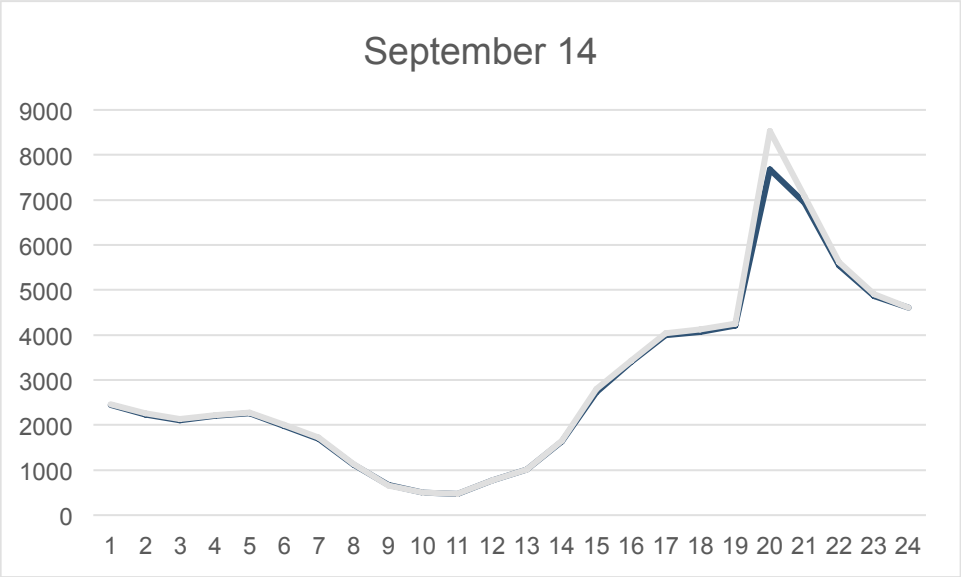
In addition, to evaluate the design of the repository processes, MIP performed an extensive evaluation of metrics in the Connectors compared to the same metrics reported through the Data Repository, with especial emphasis on the website activity data, which is collected in parallel in the Data Repository and in Google Analytics. Early analysis of the differences led to refinements in the collection process for

greater accuracy. For example, in some instances there were small but significant difference in daily event counts. A closer analysis of hourly data revealed that during spikes in website activity the Connector could become overwhelmed by traffic and drop data about some hits.

For example, the following represents a 24-hour section of website data for one site. The overall difference between Google Analytics and the Data Repository was 2.09 percent, slightly more than the tolerance limit. However, looking at the data by hour, we can see that the difference was due to

[1] https://support.google.com/analytics/answer/2637192?hl=en
[2] https://developers.google.com/analytics/devguides/collection/analyticsjs/tasks
[3] https://developers.google.com/analytics/devguides/collection/protocol/v1/parameters
[4] Note that Google Analytics' Terms of Service prohibit collecting personally identifiable information within GA

significant 9.96 percent difference during one hour. (Further analysis reveals that traffic peaked at this time because of a particular article gaining viral popularity on Facebook.)



As a result of this analysis, the Connector was redesigned to accommodate such spikes in traffic load, and daily event totals fall well within the defined tolerable differences (1.5%).

Website activity data gathered into the Data Repository compared to Google Analytics retains small variances for two identified reasons:

- Google Analytics stops counting hits in a single session at 500. It's rare in practice for human visitors to a website to exceed this limit, but it happens in some cases. The Data Repository is more accurate than Google Analytics in these cases.
- The MIMS collector uses an AJAX-style POST request to collect data. Google Analytics uses a simple image beacon (using GET) for requests with fewer than 2000 characters, and POST for requests over that limit. The image beacon GET request is compatible with some older internet browsers (Internet Explorer 7 and older). In an analysis of the participating sites, the volume of traffic from users of such browsers was found to be quite small (<1%). As the number of devices and users with older internet browsers continues to decrease over time, this was deemed an acceptable variance. Future upgrades to the MIMS collector could allow it to receive GET and POST requests if a client had an audience with a larger percentage of users visiting the site from the older internet browsers.

An Implementation Specialized for Media Sites

The standard GA tracking simply captures **Pageviews** (when a page loads in the browser). However, GA supports other types of tracking called **Events** and **social interactions** that can measure activity within a page. Social interactions typically measure an interaction with a social sharing feature (Like this on

Facebook, Tweet this, etc.). Events can be used for any type of interaction within the browser window, including clicking and scrolling, playing video or audio, and more.

The MIP implementation for the Texas Tribune and KPCC through Google Tag Manager (GTM) provides measurement of a number of Events and social interactions that are specialized for media websites. These include:

- Clicks on links within articles and in lists of related articles
- Scrolling activity and time spent on pages
- Interactions with multimedia elements such as video, audio, and image galleries
- User identifiers such as IDs based on login or email address

These customizations are made based on a measurement model developed for each organization that maps metrics to business objectives, and can yield metrics that are specialized to media (see more in "Data Reporting & Analysis" below).

Using Google Tag Manager

Collecting data with Google Analytics requires that JavaScript tracking code is added to each page of the website to be measured. Additional types of data, such as measuring link clicks, scrolling activity, interactions with a video player, and so on, require additional JavaScript to add event listeners to the page to track interactions within the browser.

Given the nature of the Data Repository gathering data from many disparate media organizations, a solution for deploying and updating this website tracking code should ideally:

- Require minimal intervention by the media partner to deploy
- Enable MIP to make updates at will to data collection without subsequent intervention by the media partner
- Co-exist with whatever website tracking technologies currently in use (including Google Analytics used by the media partner)
- Have minimal impact on website performance
- Provide ability to standardize data collection across many disparate websites, with different designs and underlying technologies

These requirements are handled by using a tag management system. Tag management systems are a class of website tools that are designed to manage these types of tracking code snippets for analytical tools like GA, using a web interface, flexible permissions, and versioning tools to organize them. GTM was chosen for its support of these requirements, its tight integration with GA, and its cost (free).

GTM is deployed through a single JavaScript snippet (the **container code**) in the source code of a website's template, included in each page. The tracking loaded by this snippet is then controlled through

GTM's web interface, with no further changes required to code on the website. This makes it easy to deploy initially by the media partner, and does not require any future changes.

MIP has altered GTM's container code (and the GA tracking deployed through it) to use non-default names for JavaScript functions and variables and the cookie used for GA . Both GTM and GA are widely popular tools, since they are available for free, and these alterations ensure that there will be no conflicts with any existing implementations of GTM or GA on the media partner sites.

GTM also supports export and import of containers. This allows easy duplication of tracking across websites (with additional customization needed for each).

Parallel Data Collection

In addition to the changes described above for compatibility purposes, the GA tracking deployed through GTM is changed in one additional way critical to the operation of the Data Repository: a copy of the data is sent to the MIMS's data collection endpoint, where it is logged and imported to the repository.

The GA tracking is extensible though a system called Tasks,[2] in which the various operations performed by the code can be replaced or augmented. For this purpose, we augment the `sendHitTask`, which sends data to the GA collection endpoint, to send a duplicate to the MIMS collection endpoint. Whatever data the tracking code generates to send to GA, a copy is also sent to MIMS.

Each tracking hit from the website follows a format standardized within GA called the Measurement Protocol.[3] These hits are logged by the MIMS data collection endpoint as a JSON-formatted log file. These logs are imported daily into the repository (see the section Data Storage & Processing below).

Batch Data & API Connections

Aside from the data streamed from the website via GA, the repository is capable of ingesting data from any source (Connector) through batch imports.

Batch imports can occur in one of two ways, depending on the source tool for the data:

- For Connectors with large, complex, or often-changing data, they often offer APIs, where the repository can employ automated scripts scheduled on a daily basis to import data
- For Connectors with data that changes less often, or tools that do not offer APIs, data can be exported manually (in a CSV or other format) and imported to the repository on some regular basis

[2] `https://developers.google.com/analytics/devguides/collection/analyticsjs/tasks`
[3] `https://developers.google.com/analytics/devguides/collection/protocol/v1/parameters`

Data Processing & Storage

The data collected in the Data Repository need to be:

- Processed for cleaning and enhancement from the raw form provided by data collection tools
- Stored in a way that make it easy to retrieve data sets for reporting and analysis

Database

The Data Repository uses big data tools capable of storing and querying large volumes of data. The database supports data in a wide variety of schema, supporting MIP's need to ingest data from many disparate sources, and provides tools for cleaning and processing data in different formats (log files, CSV, etc.) into usable form.

Data files, including logs from GA and batch imports from API or manual processes, are transferred to the database server to a predesignated folder based on the client and data source information. The folder is scanned on a regular basis and new files are imported into the database, which indexes them based on preconfigured and custom formats.

Following an initial import for a new data source format, an analyst configures a new custom import data format for each data source. For easier query writing, human readability, and consistency across data types, field aliasing is set up at this time as well to describe the fields.

Integration Between Data Sources

One of the primary purposes of the Data Repository is to establish connections between data from disparate Connectors, in order to achieve a holistic view of user interaction. There are two connections to be made between data from different Connectors:

- Connections based on individual users, such as connecting a website user in Google Analytics with the recipient of an email in MailChimp and their membership data in SalesForce.
- Connections based on articles, such as connecting an articles data in Google Analytics, the link to the article sent in an email in MailChimp, and metadata about the article from NewsLynx.

In the database import process, the analyst is able to create new extracted fields, including those for analysis of client-specific data formats for user identification.

User Identification

In order to identify users across Connectors, the Data Repository builds a set of lookup tables that join together identifiers from the various data sources. These include the following:

- **From website data:** Google Analytics' persistent cookie identifier, as well as IDs based on a login (if supported by the site) or an email address (from email subscription signups, membership/subscription transactions, etc.).[4]
- **From other sources:** login IDs or email addresses.

This process joins together all of the activity where an individual was identified across Connectors, including website activity across multiple devices (as long as the login or email ID can be tied to each device).

Article Data

Articles are typically identified by URL and data is joined between Connectors based on those URLs. In many cases, there may be slight variations of URLs for the same article (with query parameters for tracking tools, to view a second page of comments, etc.). Again, an analyst is able to define URL patterns for extraction to new fields to create client-specific patterns for matching.

Data Reporting & Analysis

Capabilities of the Repository

The Data Repository supports both routine reporting and in-depth analysis of media audience and content data.

The repository is capable of reproducing any report or set of metrics from each of the individual Connectors from which it ingests data. That is, it can provide data such as pageviews from Google Analytics or email opens from MailChimp. The database offers a rich query language to support generating such metrics.

Beyond reproducing the functionality of individual Connectors, it also has the capability of joining together data from multiple Connectors (see the section Integration Between Data Sources above). As a result, MIP can report on individual users and their behavior over time, and segment by that behavior. This allows reporting and analysis such as the following:

- View the website interaction and email activity of identified members/subscribers vs. non-members/subscribers.
- Segment users into categories based on their level of engagement: stories read, frequency of website visits, email subscriptions, membership contributions, events attended, etc.

[4] Note that Google Analytics' Terms of Service prohibit collecting personally identifiable information within GA itself. The MIP repository collects anonymized identifiers based on login or email address that are later matched with data from other tools.

Customized Metrics for Media

The metrics available from common tools are necessarily generic, a one-size-fits-all set of measurements suited to fit websites for many different kinds of organizations. The Data Repository will include customized metrics for media that perhaps will become standard for media organizations.

Such metrics can include those to measure content engagement beyond simplistic metrics such as Pageviews:

- Article Views: rather than Pageviews, counting only articles, not other pages
- Completion Rate: % of Article Views that scrolled to the end of the article text (with similar definitions for audio or video content)
- Improved Bounce Rate: count bounces only if the user didn't scroll and spent less than 30 seconds on the page
- Improved Time on Page: use the more accurate timing events included in the MIMS GTM tracking (rather than page timestamps like the default in GA)
- Related Article Clickthrough Rate: % of Article Views that clicked a link within the article text or a list of related articles
- Category Page Clickthrough Rate: % of Pageviews of a home or category page that clicked through to an article
- Comments View Rate: % of Article Views that scrolled to the comments section
- Comments Submitted: % of Article Views where the user submitted a comment
- Social Shares from Article Page
- Article Prints
- Multimedia Engagement Rate: % of Article Views where the user clicked a multimedia component such as video, audio, or an image gallery

Customized metrics can also be generated to measure repeat user behavior over time, such as Cumulative Lifetime Article Views, Cumulative Lifetime Time on Site, or new metrics for loyalty and recency.

Modeling via R

The Data Repository has also put in place processes to export data to R for analysis. R is an open-source statistical programming language and environment used for modeling and prediction.

With R, sophisticated models of user engagement can be developed. For example, statistical modeling techniques can be used to understand the factors (such as website interaction, email signups, etc.) that are likely to lead to a user signing up for a membership/subscription.

Conclusions and Future Work

The Data Repository is currently operational with a number of Connectors in use, with additional Connectors under development. Its capabilities have been demonstrated and validated for these, and work is ongoing to build upon this foundation for meaningful future analysis of impact in media.

Appendix: Web Metrics for Media Measurement

This document compiles a list of metrics definable in the Data Repository that are relevant to media website measurement.

We have grouped these metrics for the repository into four categories:

- **Group 1:** Metrics that replicate metrics in Google Analytics.
- **Group 2:** Metrics unavailable or difficult to achieve in Google Analytics because of aggregation or sampling.
- **Group 3:** Metrics based on the MIP customizations of measurement with Google Tag Manager.

Metrics in the Data Repository can be defined by the analyst, resulting in great flexibility to define new metrics or adjust the calculation of metrics to reflect the unique configuration of a particular media website.

Group 1

Metrics that replicate metrics in Google Analytics (or simple calculations derived from them).

- **Pageviews:** the total number of times a page was loaded in a browser.
- **Total Events:** the total number of Event tracking actions.
- **Social Actions:** the total number of social interactions.
- **Users:** the number of unique cookie IDs (based on the Google Analytics cookie).
- **Sessions:** the number of Sessions, where Pageviews, Events, and other interactions are grouped into Sessions by the same User (cookie ID) with less than a 30-minute gap between interactions.
- **Bounces (or Bounce Rate):** the number (or proportion) of Sessions that viewed a single Pageview and had no subsequent Pageviews or interactions Events in the Session.
- **New Sessions (or % New Sessions):** the number (or proportion) of Sessions from a User (cookie ID) that has not previously been to the site. (Likewise, we can define Returning Sessions or % Returning Sessions.)
- **Sessions from Primary Geographic Region (or %):** the number (or proportion) of Sessions from the media organization's primary geographic region (as defined by the analyst).
- **Sessions from Mobile (or %):** the number (or proportion) of Sessions from mobile devices. (Likewise, we can define similar metrics for tablet or desktop categories.)

Group 2

Google Analytics employs **sampling** to calculate custom reports, segments, and other custom requests for sets of data that exceed certain thresholds, resulting in approximations of metrics. Additionally, user-based segmentation in Google Analytics is limited to viewing 60 days of activity at a time. Even when sampling does not apply, the reports in Google Analytics are largely **aggregated**, meaning that they

show totals and averages for a variety of metrics and can be sliced by a number of categories, but do not allow an easy view or export of behavior of individual users and all of their activity.

Group 2 encompasses a number of metrics that can be derived from the Data Repository that are difficult or impossible to achieve in Google Analytics because of these limitations.

- **Time Between Sessions:** the duration of time elapsed between Users' Sessions.
- **Average Weekly Sessions:** Sessions per user per week (or other defined time period).
- **7-Day Active Users:** total Users who have been on the site in a 7-day trailing period (or other defined time period).
- **Cumulative Lifetime metrics:** totals over User lifetimes of metrics such as Article Views, time spent on the website, etc.
- **Identified Users:** Users (cookie IDs) who have been connected to one or more additional data sources, such as email, member/subscriber data, etc.

Additional metrics can be derived from other data sets using the identified Users. Segments can also be defined based on this data in the Data Repository (e.g. subscriber vs. non-subscriber, users who have donated more than $50 in the last year, etc.).

Group 3

MIP has customized measurement using Google Tag Manager to collect data on interactions important to media sites. The metrics in Group 3 encompass those measurements, some of which are improved calculations of existing Google Analytics metrics and others of which are entirely new.

- **Article Views:** like Pageviews, but counting only articles (not, e.g. homepage or category page views). Pagination (splitting an article up across several Ppageviews) can also be accounted for.
- **Category Page Clickthrough Rate:** the proportion of Pageviews of a homepage or category page that resulted in a click to an article.
- **Average Session Duration:** an improved version of the Google Analytics metric of the same name. Google Analytics' metric simply finds the elapsed time between the first and last interaction Events in the Session. The MIP data collection code deployed through Google Tag Manager measures time spent on pages in 15-second increments using Events. If more than 30 seconds elapse without the User moving the mouse or scrolling, or if the page is not in the active tab in the browser, time tracking is paused. Using these Events allows us to achieve a much more accurate measurement of time spent. In one sample, **the traditional calculation yielded an average session duration of 2:18, while the improved metric showed 5:06**, a much more accurate representation of the time Users spent with the site.
- **Time On Page (or Time on Article):** similar to Average Session Duration, but measuring the time spent on each individual page or article.
- **Scroll Depth:** drop-off of scrolling activity at breakpoints throughout a page, including 25% increments, the end of article content, and ancillary content such as comments and footers.
- **Completion Rate:** proportion of views of an article that scrolled to view the entire content of the article. (Similar definitions apply for video or audio content.)

- **Related Article Clickthrough Rate:** proportion of Article Vews that result in a click to a related article through links in the article content or in a list of related content.
- **Comment View Rate:** proportion of Article Views that scrolled into the comment section.
- **Comment Submission Rate:** proportion of Article Views that left a comment.
- **Social Shares from Article Page:** shares using social features on an article.
- **Article Prints:** printing using a print feature on an article.
- **Multimedia Engagement Rate:** proportion of Article Views that clicked on an image gallery, video, or other multimedia content within the article (if present).